Spatio-temporal Incremental Data Modelling for Multidimensional Environmental Analysis

Doc Thesis
by

Lei Song

Submitted in Partial Fulfillment of the
Requirements of the
D.Comp. Program

Department of Computing and Information Technology
UNITEC Institute of Technology
New Zealand
March 2014

Abstract

A variety of environmental problems increasingly attract academic research in order to protect ecosystems and minimise negative effects on human health. Advanced computational environmental analysis technologies have the potential to detect, monitor and perhaps effectively control these problems. However, computational environmental analysis is a complex and difficult problem to solve.

The problems associated with environmental analysis are the underlying data. These data are collected from sub-optimal positions in urban and rural areas. The limited number of monitoring stations in the network means that we are collecting enough samples over time, however, insufficient data samples are collected across the monitored area.

We determined the difficulty associated with computational analysis of environmental data via a critical review of existing approaches in the literature. Our review confirmed the biggest problem was associated with data collection, including noise introduced into the data stream, the big data problem in the form of an endless data stream, and missing data samples caused by ineffective equipment or poor placement of the monitoring equipment.

In this thesis, we document our research into computational environmental methods for addressing land usage and air quality problems.

The detection of land use change is a process of identifying differences in the state of a phenomenon by observing time-lapsed landscape imagery. Motivated by a simple neural pattern recognition mechanism, we propose a novel "one-step-more" incremental learning change detection method. In this method, an agent discovers knowledge from the first image using pixel-level incremental learning. When we detect changes in the subsequent image, the discovered knowledge model is updated and ready for the next change detection iteration. This is what we have called the "one-step more" incremental learning method. Powered by incremental data modelling techniques, the system demonstrates the capability of continuously detecting time sequenced imagery. Additionally, the method is shown to be computationally inexpensive when initializing and updating the change detection model.

Land encroachment monitoring is essential to assist the economic growth, sustainable resource use and environmental protection of a city. We investigate land encroachment on public parks in the area of Auckland New Zealand, in which the proposed "one-step-more" method employed to analyse 26 Auckland parks. The obtained average region of interest (ROI) detection accuracy is 99.91% on five popular park related objects i.e. fences, houses, parks, trees and roads. The effectiveness of the proposed method is demonstrated on four categories of encroachment: per-

manent land cover & use, and temporary & physical boundary encroachment.

We document two detailed and comprehensive investigations of computational methods for air quality analysis. The first investigation is indoor emission source detection and the second is outdoor air quality prediction. Emission source detection indoors is important when locating the possible origin of pollution that could have a negative effect on the health of occupants. Outdoor air quality prediction is essential when attempting to minimise the negative effect on the health of citizens and the ecosystem.

Addressing indoor emission source detection, we propose a novel inter-pollutant correlation analysis technique. Unlike other documented solutions that analyze merely primary pollution, our method is further enhanced by calculating intra-pollutant correlation coefficients for characterizing and distinguishing emission events. Extensive experiments show that seven major indoor emission sources are identified by the proposed method, including (1) frying canola oil on electric hob, (2) frying olive oil on an electric hob, (3) frying olive oil on a gas hob, (4) spray of household pesticide, (5) lighting a cigarette and allowing it to smoulder, (6) no activities, and (7) using an exhaust or ventilation system. Furthermore, our method improves the detection accuracy by a support vector machine, compared to the classification without data filtering, and with feature extraction of PCA and LDA.

Addressing outdoor air quality predication, we propose a novel spatial data-aided incremental support-vector regression (SaIncSVR). We overcome some of the problems associated with other prediction models. Existing models often demand data to be presented in a more convenient form than real data facilitates, for instance complete data, time-series complete, or a specified data capture process making them inadequate when used for environmental prediction. We conduct extensive experiments of $PM_{2.5}$ prediction over 13 monitoring stations in Auckland, New Zealand. The experimental results indicate that our SaIncSVR performs improved $PM_{2.5}$ prediction. Our method shows the capability to overcome missing data problems, an improvement over many other documented techniques. We compare our method with a local incremental support-vector regression. Further for $PM_{2.5}$ spatial prediction, we conduct experiments on outdoor air quality prediction using a data-driven Gaussian geometry model. This model shows promise on a limited number monitoring stations and warrants further investigations.

We promote incremental data modelling, given its benefits in processing large amounts of data with parallelization capability in many dimensions, as a constructive solution to spatio-temporal computational environmental analysis. This research documents systematic and comprehensive research in the area of computational environmental analysis. For this study we investigated a variety of environmental problems, environmental data, change & emission source detection,

pollution variation & distribution, and data modeling & prediction. The proposed computational methods could be useful for other environmental detection and prediction problems, especially those using large, asynchronous and spatial-temporal data.

Acknowledgments

I would like to thank all people who have helped and inspired me during my doctorate study.

This thesis was made possible by an Unitec Institute of Technology Scholarship. I am grateful to the Department of computing for the scholarship which enabled me to undertake a doctorate program at Unitec Institute of Technology.

I especially want to thank my supervisors, Assoc Prof. Paul Pang and Prof. Abdolhossein Sarrafzadeh, for their guidance during my research and study at Unitec Institute of Technology. Their perpetual energy and enthusiasm for research have motivated all their students, including me. In addition, Assoc Prof. Paul was always accessible and willing to help his students with their research. As a result, research life became smooth and rewarding for me. Many thanks also go in particular to Angelina Chai-Rodgers, for being supportive and helpful whatever the occasion.

My special words of thanks should also go to my industry supervisor, Dr. Ian Longley, for his constant motivation and support during the course of my thesis. I enjoyed the personal discussion with him to explore the idea for our collaborating projects. My sincere thanks to Dr. Gustavo Olivares for being with me as an domain expert. His scientific inputs, personal helps and friendly nature has always made me feel at ease with him. Without the encouragement as well as the access to their expertise and data this work would have not been possible.

All my lab buddies at the Doctorate Lab made it a convivial place to work. In particular, I would like to thank David McCurdy and Yiming Peng for their friendship and help during my thesis. All other folks, including Lei Zhu and Simon Dacey, had inspired me in research and life through our interactions during the long hours in the lab. Thanks.

My deepest gratitude goes to my family for their unflagging love and support throughout my life; this thesis is simply impossible without them. I am indebted to my father, Deguang Song, for his care and love. As a typical father in a Chinese family, he worked industriously to support the family and spared no effort to provide the best possible environment for me to grow up and attend school. He had never complained in spite of all the hardships in his life. I cannot ask for more from my mother, Shuqing Hou, as she is simply perfect. I have no suitable word that can fully describe her everlasting love for me. I remember her constant support when I encountered difficulties and I remember, most of all, her delicious dishes

Last but not least, thanks be to God for my life through all tests in the past years. You have made my life more bountiful. May your name be exalted, honoured, and glorified.

List of Abbreviations

AI	Artificial intelligence
ANN	Artificial neural networks
AR	Autoregressive model
ARIMA	Autoregressive integrated moving average model
ARMA	Autoregressive moving-average model
ART	Adaptive resonance theory
BME	Bayesian maximum entropy
EM	Expectation-maximization
FFT	Fast Fourier transform
GA	Genetic algorithms
GTP	Global temporal prediction
HMM	Hidden Markov models
IOA	Index of agreement
KNN	k-nearest neighbours
LDA	Linear discriminant analysis
LTP	Local temporal prediction
MA	Moving-average model

MAE Mean absolute error

MBE Mean bias error

MLP Multi-layer perceptrons

MRF Markov random field

PCA Principal component analysis

RMSE Root mean square error

RMSE_S Systematic RMSE

RMSE_U Unsystematic RMSE

SaGTP Spatial data aided global prediction

SaIncSVR Spatial data aided incremental SVR

SVM Support vector machine

SVR Support vector regression

Contents

List of Figures

List of Tables

Chapter 1
Introduction

Environmental issues have gained an increasing prominence on a global scale in recent times. As stated in (Fransson & Gärling, 1999), a serious threat to human beings and their environment is the continuous and accelerating overuse and destruction of natural resources. People have made great efforts to protect the environment and preserve their way of life. Unfortunately, many of the environmental problems that have attracted the most public attention, such as global warming, deforestation, and so on are becoming even worse than thought.

Up to now, a numbers of efforts have been made in addressing the environmental issues, chief among them the problems of climate change, global warning, Ozone depletion, and so on. The digitization of our world has been a forward processing now; still, it is surprising to step back and look at the physical change of the environment. People now monitor emission sources in road transport, air quality in urban areas, ground water quality, sea level, satellites images on land usage, and so on by digital sensors. As a result, there is a huge amount of data, returned from monitoring stations, that satisfies compliance obligations and can be mined to provide insight and information about the environment. The database includes satellite photos, cartographic information, and the results of all aspects of environmental studies, which provides a efficient digital library to support environment treatment.

Computational environmental analysis conducts knowledge discovery from data for decision making support towards environmental treatment, which covers specifically pollution detection (monitoring), examination, and prediction. Among them, pollution detection identifies where and what pollution happens to make an intelligent measure of the environmental changes by collecting real-time data on emissions, weather conditions or land covers; pollution examination evaluates the level of pollution and effects of pollutants to investigate the relationship between measures of pollution, such as air quality or proximity and presence of industrial facilities; and pollution prediction estimates the future trend of pollution to identify the best management practices that control pollutant losses. In general, computational environmental analysis presents important information for government environmental agency as well as environmentalists to prevent pollution.

In the field of computational environmental analysis, the focus of the presented

doctorate thesis research lies in:

- Environmental change. Environmental analysis is a systematic process initiated by detecting changes in terms of classification and location of those changes. Given an environmental change has been detected this can trigger the possibility of performing statistical analysis to understand potential causes of that change, thereby yielding a classification (reason and type). The location of pollution (especially in motion) poses problems for environmental change detection and perhaps overlooked by many researchers, owing to the actual and perceived complications of dynamic environmental change detection.

- Air quality. Environmental issues are induced by pollutants in air, soil and water resources. The air is a complex dynamic natural gaseous system that provides essential gas to support life on planet Earth. Pollutants, that are delivered to the surrounding air, have hazard consequences to human and nature. This has been a significant risk factor for multiple health conditions. The health effects results from air pollution include difficulty in breathing, wheezing, coughing, asthma and aggravation of existing respiratory and cardiac conditions. Air pollution, cause enormous external costs; as a consequence the air pollution prevention takes a significant role.

1.1 Background

Environmental problems are some of the most serious problems to affect peoples' lives. Pollution causes unpredictable diseases and disasters; it even brings deadly threats to people. The main forms of pollution are air, water and soil pollution. Pollution is defined as the process of making air, water, soil dangerously dirty and not suitable for people to use, or the state of being dangerously dirty.

1. Air pollution. Air pollutants (WHO, 2005) include many different types, such as sulfur oxides, carbon monoxide, pesticides and fluorides. Also, air pollution has some harmful effects on humans, animals, plants, or other materials in the environment. For example, carbon monoxide could impair the blood's ability to transport oxygen.

2. Water pollution. Water pollutants (West, 2005) include a broad variety of materials, like synthetic organic compounds, human and animal wastes, radioactive materials, heat, acids, sediments, and disease-causing microorganisms. Similar as air pollution, water pollution could also cause damaging issues on humans, animals and environment. For example, the element mercury is used

for producing light switches, air conditioners, fluorescent lights, etc. When such things are made, small amount of mercury metal is likely to escape into the environment and, eventually, into lakes and rivers. As known, mercury is highly toxic to living creatures.

3. Soil pollution/Land pollution. Soil pollution (*Soil Pollution*, 2011) contains the pollutant materials of soils, mostly chemicals. Soil pollution can lead to water pollution if toxic chemicals leach into groundwater, lakes, or oceans. It also gives rise to air pollution, because soil may release volatile compounds into the atmosphere. It is obviously harmful to plants and has destructive effect on living animals.

In order to cope with pollution problems, people have spent lots of money and efforts on establishing an effective strategy. Moving manufacturers to downwind position, recycling waste, educating people the importance of environmental protection to analyzing environmental problems, investigating pollutants, people have done many works on pollution prevention. Apparently, all the above efforts are not sufficient, as we know global warming, ozone hole, melting glaciers and "big smoke" etc are getting worse with the time going on.

1.2 Computational Environmental Analysis and Challenges

Computational environmental analysis uses techniques, derived from fields like computer science and applied mathematics, to simulate and analyze computational models for environmental problems. Statistical analysis and machine learning methods have been widely applied to environmental analysis on account of their advantages on fast and effective calculation (Ratle, Kanevski, Terrettaz-Zufferey, Esseiva, & Ribaux, 2007; Rebenitsch, Owen, Ferrydiansyah, Bohil, & Biocca, 2010; Schwender, Zucknick, Ickstadt, & Bolt, 2004). In addition, remote sensor techniques, which are also widely used, allow people to observe most environmental problems. Therefore, computational environmental analysis is one of the solutions to environmental problems.

Computational environmental analysis includes pollution detection (monitoring), pollution examination and pollution prediction (Raes, Foerstner, & Bork, 2007; Kreinovich et al., 2007; Rangel, Diniz-Filho, & Bini, 2006). Pollution detection identifies where and what pollution happens. Pollution examination evaluates the level of pollution and effects of pollutants. Pollution prediction estimates the future trend

of pollution. Computational environmental analysis presents important information for environmentalists to prevent pollution.

1.2.1 Pollution Detection

Pollution Detection is used to identify the location and types of pollution in the environment. It can be described as:

1. changes in pollutants, increases and decreases in quantity, severity, etc,

2. and introduction of new types of pollutants.

Change detection is also the fundamental tool for pollution analysis, which provides the useful information to pollution examination and pollution prediction.

Environmental change detection has four requirements as detection accuracy, automatic detection, detection sensitivity, and detection continuity. Detection accuracy reports exactly all differences amongst a series of variables for each data collection. Automatic detection dynamically adjusts thresholds of change detection methods. To achieve early warning, the sensitivity of change detection is important for environmental analysis in that an environmental problem cannot happen within one day, there is a procedure of environmental change before the problem becomes serious and widely recognizable (Fries, Ziegler, Kurian, Jacoris, & Pollak, 2005). Environmental changes may be rapid and significant or slow and inconspicuous. Such changes actually continue over time, possibly without end. Thus, continuous change detection with high speed is also important.

Applications

In fact, environmental change detection works as cleaning equipments and air / water indicators for monitoring environmental system (Matsuzawa et al., 1984; Francke, Miessner, & Rudolph, 2000; Carlton & Smith, 2000). Since the advancement of remote sensing technology and improvement of satellite images resolution, it then has been applied on area surveillance / land cover changes detection (Vu, Matsuoka, & Yamazaki, 2004; Haverkamp & Poulsen, 2003; Phua, Tsuyuki, Furuya, & Lee, 2008). Furthermore, for the purpose of ecological restoration, environmental change detection for image data also plays a very important role, because environmental change detection can accurately presents what happened in the target area (Ardli & Wolff, 2009; Fortin et al., 2000).

Current issues

For performance evaluation, typical environmental change detection concerns change detection accuracy and convenient threshold determination. However, the study of sensitivity and continuity of change detection has attracted limited attention so far. Early pollution detection and long term monitoring are the current issues for environmental change detection.

- Detection sensitivity. Sensitivity for change detection refers to the capability for detecting minute changes that are imperceptible and easily neglectable. Note that sensitivity here is different from accuracy, and it can be obtained by just setting narrow thresholds and detecting more changes. It detects where/what changes when it first happens and reports not only percentages, but also the probability of pollution.

  Existing change detection methods consider merely accuracy of the detection. Actually, the high sensitivity is not equivalent to the high accuracy. The sensitivity is limited because existing methods report only one specific variable based on interaction among all variables (Klanova, Kohoutek, Hamplova, Urbanova, & Holoubek, 2006; Martinez, Phillips, & Whitesides, 2008; Fischer, Dryer, & Curran, 2000; Loomis, Castillejos, Gold, McDonnell, & Borja-Aburto, 1999; Dagan, 2000).

  In addition, existing change detection methods enhance the sensitivity via narrowing the thresholds. However, narrowed thresholds not only filter out slight changes but also amplify the effect of noises. This leads to more false alerts.

- Detection continuity. Continuous change detection is to track environmental changes from an uninterrupted environmental variations. A typical example is biological invasions. Biological invasions are defined as breach biogeographic barriers and extend their range. Invasions of nonnative plants, animals, or microbes cause major environmental damage. Though human has attempted to manage biological invasions for long time, but invasions continuously occur among rivers, lakes and canals, even if they are very clean (Vitousek, D'Antonio, Loope, & Westbrooks, 1996). Monitoring environmental changes like biological invasions demands for long term environmental change detection. In general, the difficulties of continuous environmental change detection are given below:

  1. environmental data comes in real time;
  2. environmental change involves huge size of data stream or matrix;
  3. environmental data comes with unknown pollutants;

4. environmental changes are required to be detected instantly without retraining detection model.

It is worth noting that existing change detection methods perform very limited on continuous environmental change detection, because they are not capable to process for large amount of data stream such us huge amount of environmental data. Two limitations of the existing change detection methods to achieve continuity are, 1) the change detection can only be conducted by comparing samples one by one, and 2) the thresholds can only be determined and fixed according to previous experience. Because both re-sampling and resetting thresholds cost much time, a great number of changes may be missed during the process.

1.2.2 Pollution Prediction

Pollution prediction simulates the progress and estimates the future trend of pollution based on the information retrieved by pollution change detection system. It helps stakeholders construct strategies to deal with pollution problems. Analog to most prediction problems, pollution prediction suffers when data lose and special events happen.

Applications

Pollution prediction is important for people to decide strategies on special environmental problems after disasters. Bombay city used to predict concentrations of carbon monoxide (CO) with models like GM, CALINE-3 and HIWAY-2, which reports air pollution for the government to decide traffic system (Luhar & Patil, 1989). A case in Slovenia is to predict SO_2 from coal fired thermal power plants for advising the government to develop new energy and control the quantity of provided power (Boznar, 1997). Global ocean pollution caused by oil spill is one of the most serious problems to people and marines. Hackett, Comerma, Daniel, and Ichikawa (2009) shows the directing applications and strategies for pollution prevention based on oil spill fate forecasting systems.

Current issues

- Data Missing. Data missing problem is a serious issue for all prediction models. The input of prediction models is the most important (Barlow et al., 6 September 2006; Iimura et al., 2009), whereas environmental data analysis has confronted with serious data missing situations, because new pollutants are

continuously presented (Guo, Jia, Pan, Liu, & Wichmann, 2009). Until now, environmentalists still cannot predict Fukushima nuclear leakage 2011, because a great amount of unknown pollutants came up after this disaster and it was the first time that nuclear pollutants exhausted to ocean.

- Special events. Special events sometimes may lead prediction models to produce incorrect forecasts to people. Special events bring new pollutants and new pollution types to environmental analysis without any experience. In other prediction applications such as finance (Pang, Song, & Kasabov, 2011) and medical science (Kanehisa et al., 2008), similar patterns can be extracted from historical data. However, environmental analysis usually confront with pollution types when pollution generated from new technologies. Thus, different from other applications, it is very difficult to perform pollution prediction successfully after special events.

- Spatial consideration. Existing air quality prediction projects focus on applying spatial characteristics to support time series forecasting. Prediction models such as Bayesian Maximum Entropy (BME) method and Hidden Markov Models (HMM) has been applied on spatio-temporal air quality prediction. BME models are often used with ratio prediction such as ratio of $PM_{2.5}/PM_{10}$, which store the ratios for the different locations (Chang & Lee, 2008). HMM applies probabilities on air quality prediction, the regression function can be simulated by Markov chains, which provides more accurate forecasting (Dong et al., 2009). However, HHM models calculate the probabilities with geographical characteristics, it is a way to consider spatial information but not spatial prediction.

- Prediction continuity. Considering the latest change in the air is another hard topic for air quality prediction. Due to changes in the air occurring all the time, a self-updated learning model is required. However, BME models provide a storage for time series prediction models in different location. If a change happen, BME models must be retrained for each location (Chang & Lee, 2008; Christakos & Serre, 2000). HHM models have the same problem as the BME models. If a change happens, all the input probabilities must be re-calculated, which incurs an expensive calculating cost and is time consuming (Schmidler, Liu, & Brutlag, 2000; Dong et al., 2009) .

1.3 Motivation of Incremental Data Modelling

Environmental problems comprise information from both location and time. Integrating data from different spatial and temporal scales is an important component of computational environmental study. Environmental problems, however, which are commonly temporally rich in data, but haven't motivated extensive spatio-temporal environmental studies in literature. This results existing solutions to computational environmental analysis focus on trace of local problems rather than a spatio-temporal analysis. In this research, we consider incremental data modelling for spatio-temporal environmental analysis, owing to the following advantages of incremental learning,

1. capability of keeping all knowledge in memory while learning with protection from forgetting,
2. capability of mining a large amount of data with parallelization on spatial and temporal dimensions.

1.4 Environmental Problems Investigated

Environmental problems are complex, ever changing, and are the subject of many research papers in various disciplines. In this thesis, we address specifically three major environmental problems which are land cover change, indoor air quality and airborne particulate matter.

1.4.1 Land Cover Change

Land cover change is categorized as proximate (direct, or local) and underlying (indirect or root). The proximate change is caused directly by humans operating at the local level (e.g., individual farms, households, or communities), while the underlying change is caused by broader context and fundamental forces underpinning these local actions originating from regional (e.g., districts, provinces, or country). Land cover change effects biotic cover, water quality, hydrological, water flows and biological impacts. Deforestation and encroaching shrubland are two typical phenomenons of land cover change, which may eventually cause flooding and shortage of agricultural land.

1.4.2 Indoor Air Quality Control

People spend about 80% of time indoors, so it is important that we gain a better understanding of the pollutants to which people are exposed indoors. Common in-

door sources of airborne particles include combustion sources (e.g., primarily heating and cooking) and tobacco smoking. Other sources include combustion (e.g., candles, incense, etc), hygiene products (e.g. solvents, pesticides) and activities (e.g. dusting). Identifying the contribution of each source, and exposure to it, is central to the effort to understand health effects and manage risks. The magnitude, frequency and prevalence of these sources are strongly related to individual lifestyles and behaviours.

1.4.3 Airborne Particulate Matter (PM_x)

Air pollutants emitted by pollution sources are usually categorized to suspended particulate matter (PM) (e.g., dusts, fumes, mists, and smokes); gaseous pollutants (e.g., gases and vapors); and odors. PM, one of the major pollutants, is tiny solid and/or liquid particles suspended in the air. These particles can be classified according to total suspended particles: the respirable suspended particle PM_{10} (median aerodynamic diameter less than 10.0 microns), and the fine particles $PM_{2.5}$ (median aerodynamic diameters of less than 2.5 microns). PM_x leads to increased use of medication. The health effects made by PM_x include coughing, wheezing, shortness of breath, aggravated asthma, lung damage (including decreased lung function and lifelong respiratory disease), premature death in individuals with existing heart or lung diseases. Since the size of $PM_{2.5}$ is very tiny, it travels deeper into the lungs. Therefore $PM_{2.5}$ can have worse health effects than the PM_{10}.

1.5 Approach

Existing solutions for spatio-temporal environmental analysis process their data from a huge storage system through a specific batch learning model. In the approach, the learning model not only considers information from both spatial and temporal domains, but also the model updates itself when the latest change occurs without expensive computational costs and high memory usage.

1.5.1 Incremental Learning based Image Series Change Detection

Environmental data such as flooding, deforestation and land use data are monitored by satellite images, they are not a pair of image and only contain a single object. However, existing image change detection methods focus on seeking differences between a pair of images (Im, Jensen, & Tullis, 2008a; Ma, Zheng, Yuan, & Zhang, 2010; Verbesselt, Hyndman, Zeileis, & Culvenor, 2010). For image series change detection, existing methods generally rely on tracking a foreground object

to detect the changes across different images (Tian, Feris, Liu, Hampapur, & Sun, 2011; Borges & Izquierdo, 2010). No existing image series change detection solution offers a general solution for continue and multiple objects detection.

This thesis proposes that one-step more incremental learning for image series change detection, which allows an agent to upgrade its knowledge on the target image by performing incremental learning on top of its current knowledge. By recognizing major objects, the proposed method can learn knowledge from the current image for change detection. As this thesis shows, the one-step more incremental learning model can be self-updated when new knowledge comes, allowing image change detection explicitly to make a trade-off between performance and consistency in the evaluation of accuracy and continuity.

1.5.2 Pollutant Iteration Analysis and Online Emission Source Detection

There are many reasons an emission source detection system should offer an effective knowledge discovery component by which applications can extract the most important features from their data. A detection system supported by a feature extraction system already has an established developer base, making it effective for emission source detection. So far, many pattern recognition models have been used to detect emission sources. These extraction methods are used to magnify the main orthogonal contributions which explain most of the pollutants for an emission source. Therefore it is hard to consider the interaction between pollutants.

The knowledge extraction method in this thesis, on the other hand, instead magnifies the main pollution contributions: the relationships among pollutants. This is a correlation coefficient based approach to support emission sources detection. Furthermore, this approach considers the relationship among pollutants in comparison with other existing feature extraction methods. Therefore the detection results supported by the proposed method are better than the detection results supported by other feature extraction methods.

1.5.3 Incremental Learning based Spatio-temporal Air Quality Prediction

Ideally environmental events prediction model should be able to achieve three criteria: accuracy, spatio-temporal consideration, and real time prediction. Existing prediction models focus on applying geographical characteristics on time series prediction models to improve the forecasting accuracy rather then considering spatial

information. In addition, existing models are not able to carry on the latest change in the air with updating itself only.

This thesis proposes a spatial prediction for $PM_{2.5}$ forecasting and incremental learning models to catch the latest changes in the air. The spatial prediction approach supposes that the concentration of $PM_{2.5}$ gradually decreases from the city center outwards to rural area. We can then provide a set of concentric rings in order to predict $PM_{2.5}$ in a range for related (i.e. locations based on the same ring) locations. In addition, this research applies incremental learning SVR to add new data into the support vector machine by updating SVM model only.

1.6 Contribution

The main contributions of this thesis are as follows:

1. proposes a novel incremental learning based image change detection method capable of detecting sequences of changes over image series,
2. calculate in-between pollutants correlation coefficients for characterizing and distinguishing emission events,
3. applies spatial prediction to analyse the relationship between the strength of $PM_{2.5}$ concentrations and distance to city center,
4. and utilizes incremental learning SVM regression to consider the latest change in the air, which means re-training learning models without high computational cost when the latest data comes.

1.7 Thesis Organization

The rest of this thesis is organized as follows. Chapter 2 explores the data and data collection on different environmental problems and highlights the challenges that we address for computational environmental analysis. Chapter 3 evaluates environmental change detection methods with a focus on the environmental problems addressed by the thesis. Chapter 4 gives a brief introduction to environmental prediction models with a focus on air pollution prediction. Chapter 5 discusses feature extraction researches which have been conducted on land use problems and air pollution. Chapter 6 proposes a one-step more incremental learning approach to image series change detection, and Chapter 7 applies the approach to land encroachment detection of Auckland parks. Addressing air quality problem, Chapter 8 proposes a correlation based feature extraction method to support emission

source detection system. Chapter 9 proposes a spatial prediction for $PM_{2.5}$ concentrations in Auckland region, and measures the performance of incremental learning for spatio-temporal $PM_{2.5}$ concentration forecasting. Finally, Chapter 10 contains the conclusions drawn from this thesis.

Chapter 2

Environmental Data

Environmental data reflects the variation of environmental conditions. We analyse the data for the purpose of understanding environmental behaviors in statistic. Typically, the data is collected by monitoring networks and remote sensors. This chapter explores the data and data collection on different environmental problems and highlights the challenges that we address for computational environmental analysis.

2.1 Introduction

Environmental problems are harmful aspects of human living conditions and biophysical environment. Environmental data describes environmental pressures, the state of the environment, and the impacts on ecosystems. Environmental data collection covers all environmental problems from indoor air pollution to climate migration and from soil nutrition to farmland sandy desertification.

The environmental data and statistics are relevant to city planning. In a city, a monitoring network that consists of a number of distributed stations is operating for monitoring the drink water quality, air pollution pressures and land usage. Such continuous data collection serves a vital role on city environmental management by revealing long-term trends that can lead to new knowledge and understanding, and planning environmental policy.

In practice, environmental data collection is highly influenced by many natural phenomena such as lighting and windy, equipment failure, insufficient sampling or measurement errors. This pose great challenges on computational environmental analysis, to handle noisy, spatio-temporal, full of missing samples, and continuously streaming data.

This chapter is organized as follows: In section 2.2, we present the brief concepts of environmental data collection. In Section 2.3, we discuss the status of environmental data and challenges to computational analysis. Finally, a conclusion is provided in Section 2.4.

2.2 Data Collection

Environmental data is collected for the preparation of environmental impact analysis, as well as in many circumstances in which human activities carry a risk of harmful effects on the natural environment. In all cases the collected data will be reviewed, analysed statistically and published to the public. In a city, the environmental pressures are mostly from energy, population pressures, transportation, and waste. The monitoring items often include air quality, amenity, biodiversity, land, natural hazards and water. The data collection methods generally can be categorised as environmental surveying, auto continuous monitoring, and remote sensing.

2.2.1 Environmental Surveying

Environmental surveying uses surveying techniques to understand the potential impact of environmental factors on humans life. The surveying collects environmental data casually and perhaps irregularly, this includes grab sampling and passive sampling. Environmental surveying is a popular approach used in environmental monitoring, as it is simple and accurate. However, the equipment used for chemical parameter identification can be highly expensive.

To analyse a homogeneous material, for example water, the often used collection method is "grab samples". A very common example of grab samples is to fill a clean bottle with river water. This method provides a good snap-shot view of the quality of the sampled environment at the location of sampling and at the time of sampling. After collection, the sampled water is analysed under chemical environment to produce the level of chemical parameters that have the potential to affect any ecosystem. The list of chemical parameters can be expanded or reduced based on developing knowledge and the outcome of the initial surveys (Felipe-Sotelo, Andrade, Carlosena, & Tauler, 2007). The grab samples method is useful in explaining the inhibition mechanism but they are often expensive and time-consuming (Gece & Bilgiç, 2009). In addition, the data collected cannot be extrapolated to other times or to other parts of the river, lake or ground-water without additional monitoring (Nollet & De Gelder, 2000). Figure 2.1 shows an example of grab samples for water data collection.

"Passive sampling", another environmental surveying, reduces the cost and the need of infrastructure on the sampling location. This method allows for a better cover and more data being collected. Chemcatcher and an air sampling pump are the examples of a passive sampling devices. Figure 2.2 shows the chemcatcher for water/soil quality data collection and an air sampling pump for air quality data collection. Although passive sampling provides clearly the particle phase of the

(a)

(b)

Sites / Parameters	El-Taween Min	El-Taween Max	El Fayoot Min	El Fayoot Max	The Bridge Min	The Bridge Max	El Oura Min	El Oura Max	El Forsan Min	El Forsan Max	Beach clubs Min	Beach clubs Max
Temperature °C	14	29	14	29	14.5	29.5	14.5	29.5	15	31	15	31.5
pH	7.4	8.5	8	8.8	8.6	8.9	7.9	8.4	7.3	7.9	7.2	7.9
Conductivity µmohs cm^{-1}	42.5	52.5	51.5	78	77	85	42	67.1	68	107	79	90.2
Salinity ppt	33.8	41.9	32.2	44.6	33	43.8	31.4	42.7	41.2	44.9	39.9	44.8
Turbidity m	2.6	2.9	2.8	2.9	2.2	2.6	1.8	2.4	0.24	0.45	0.39	0.5
Dissolved Oxygen mg l^{-1}	7.2	8.4	7.3	10.2	6.2	8.9	5.6	9.6	6.3	9.8	7.1	8.8
Nitrate µg l^{-1}	15.6	114.1	29.5	130	53.2	112.2	32	130.4	32.2	98.4	20.2	92
Nitrite µg l-1	2.7	25.2	1.4	19.2	1.3	22.1	1.1	19.1	1	10	1	13
Ammonium µg l^{-1}	118	253	121	188	115.3	159.3	112	157.2	98.3	148	92.2	134.9
Phosphate µg l^{-1}	13.2	20.4	12.2	29.6	13.7	28.5	10.5	26.2	12.2	26.8	6.9	25.9
Silicate mg l^{-1}	3.4	6.6	3.8	6.4	3.3	5.8	3.1	5	2.7	6.7	3.4	6.8
Chlorophyll a µg l^{-1}	0.3	24	2.8	26	3.2	18.1	4.2	16	3.9	17.9	4.7	19.3
N P ratio	1.07	5.59	1.8	5.82	2.8	4.75	2	6.74	1.98	48	2.03	9.45

(c)

Figure 2.1: An example of grab samples: 2.1a filling a clean bottle with river water, 2.1b determining the level of chemical parameters in water, and 2.1c recording the chemical parameters.

atmosphere, water or soil, the data collected cannot be extrapolated to other times or to other place.

2.2.2 Monitoring Networks

As the requirement from the public, many environmental status such as drink water quality and air quality must be reported real-time. Therefore environmental data collection is required to be a continuous work, which involves having an automated collection close to the environment being observed so that data can be monitored in real time. There are a large number of specialized sampling equipments available that can be programmed to take samples at fixed or variable time intervals or in response to an external trigger. For example a sampler can be programmed to start taking samples of air chemical parameters (PM_x, NO_x) of a monitoring station at 1-minute intervals. Such systems are often established to protect important water supplies and air quality monitoring for early warning of potential problems. These systems routinely provide data of chemical parameters such as pH, dissolved oxygen, conductivity, turbidity and colour in the water and SO_2, NO_2, CO, O_3 and PM_x in the air to examine a wide range of potential organic pollutants. Figure 2.3 gives an example of continuous environmental data collection for water and air quality from the national institute of water and atmospheric research (NIWA), New Zealand.

2.2.3 Remote Sensing

The urban sprawl and forest coverage decreasing are national or global problems and happen in a large area. Problems such as climate change, global warming and rainfall also varies over the time and in large areas. Remote sensing is an important technique for monitoring such environmental problems.

For remote sensing, aircraft or satellites are used to monitor the environment by multi-wavelengths sensors. The visible wavelengths (RGB), known as true color images, are often used for land use problems such as area surveillance, ice fields, deforestation, flooding and so on. The invisible wavelength is represented by value of electromagnetic energy and mapped into a false color satellite image. These false color images are often used for monitoring clouds, Earth's vegetation, atmospheric trace gas content, sea state, ocean color. Figure 2.4 shows two examples of remote sensing images for land cover and temperature monitoring in New Zealand.

(a)

	Area of investigation			
	Halifax Marina		Port Soleil Marina	
	Duration of deployment		Duration of deployment	
Organic analytes	7 (days)*	14 (days)*	7 (days)*	14 (days)*
Diuron	6.6±0.5	17.3±1.3 [b]	417.8±8.6	1500.0±11.8 [b]
Atrazine	3.7±0.6	5.9±1.4 [b]	5.2±0.7	6.000±0.4
Lindane	3.0±0.6	3.2±0.5	0.8±0.3	[c]
Fluoranthene	11.7±0.4	7.4±0.7 [b]	10.7±0.8	8.200±0.7
PCB 153	4.6±3.0	2.8±1.6 [b]	2.0±0.5	1.700±0.3
PCB 52	2.7±1.2	5.3±2.6 [b]	1.9±0.4	[c]
Irgarol	[c]	[c]	43.4±1.0	125.00±0.3 [b]

(b)

(c)

Time	Latitude	Longitude	Row_Index	Column_Index	TB_QC_Flag	Heterogeneity_Index	Surface_Type	Soil_Moisture	Veg_Water_Content
4.5533057	-81.82574	-79.282624	584	287	0	5983	257	-9999	-9999
4.5533057	-81.33788	-79.282624	585	287	0	8325	257	-9999	-9999
4.5533057	-81.82574	-79.132224	584	288	0	128	273	-9999	-9999
4.5533057	-81.33788	-79.132224	585	288	0	9643	257	-9999	-9999
4.5533057	-80.551254	-79.87202	583	289	0	8663	273	-9999	-9999
4.5533057	-81.82574	-79.87202	584	289	0	28	273	-9999	-9999
4.5533057	-81.33788	-79.87202	585	289	0	10098	257	-9999	-9999
4.5533057	-80.551254	-79.81172	583	290	0	9997	273	-9999	-9999
4.5533056	-81.82574	-79.81172	584	290	0	421	257	-9999	-9999
4.5533057	-81.33788	-79.81172	585	290	0	9048	257	-9999	-9999
4.5533057	-80.551254	-79.39141	583	291	0	9665	273	-9999	-9999
4.5533056	-81.82574	-79.39141	584	291	0	41	273	-9999	-9999
4.5533057	-81.33788	-79.39141	585	291	0	8779	257	-9999	-9999
4.5533057	-80.551254	-79.09911	583	292	0	8902	273	-9999	-9999
4.5533056	-81.82574	-79.09911	584	292	0	343	257	-9999	-9999
4.5533057	-81.33788	-79.09911	585	292	0	8043	257	-9999	-9999
4.5533057	-80.551254	-77.8308	583	293	0	881	273	-9999	-9999
4.5533056	-81.82574	-77.8308	584	293	0	502	273	-9999	-9999
4.5533057	-81.33788	-77.8308	585	293	0	8418	257	-9999	-9999
4.5533057	-80.551254	-77.5705	583	294	0	373	273	-9999	-9999
4.5533056	-81.82574	-77.5705	584	294	0	0	273	-9999	-9999
4.5533057	-81.33788	-77.5705	585	294	0	944	257	-9999	-9999
4.5533056	-80.551254	-77.3102	583	295	0	547	273	-9999	-9999
4.5533056	-81.82574	-77.3102	584	295	0	412	273	-9999	-9999

(d)

(e)

Feb 15 2010 9:15:50 AM Serial Number [illegible]

	Mode	Flow Rate	Start Date	Start Time	Volume (Litres)	Accum Volume	Duration (dd:hh:mm:ss)	Atm Pressure
76	Hold		19-Jan-10	8:51:44 AM			5:00	28
77	Sleep		19-Jan-10	8:56:44 AM			26d 6:43:26	29
78	Hold		14-Feb-10	3:40:10 PM			0:05	29
79	Run	5.30	14-Feb-10	3:40:15 PM	0.18	43.28	0:02	28
80	Hold		14-Feb-10	3:40:17 PM			0:01	28
81	Run	5.30	14-Feb-10	3:40:18 PM	0.53	43.81	0:06	28

(f)

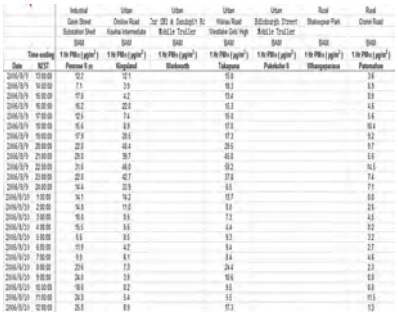
Figure 2.2: An example of passive sampling: 2.2a chemcatcher for water quality data collection, 2.2b data returned from chemcatcher water sampling, 2.2c chemcatcher for soil quality data collection, 2.2d data returned from chemcatcher soil sampling, 2.2e air sampling pump for air quality data collection, and 2.2f air quality data returned from passive sampling.

(a)

2	0	Field3	Field4	Field5
1	10002	0100.0105		LAKE MARVIN AT MIDLAKE
1	10003	0100.0200	07227448	PUNTA DE AGUA CREEK FM 767
1	10004	0100.0300		COLDWATER CREEK AT SH 136
1	10005	0100.0380		PALO DURO RESERVOIR NEAR DA
1	10006	0100.0400		PALO DURO CREEK AT FM 2387
1	10007	0100.0420	07233500	PALO DURO CREEK AT SH 207-1

(b)

(c)



(d)

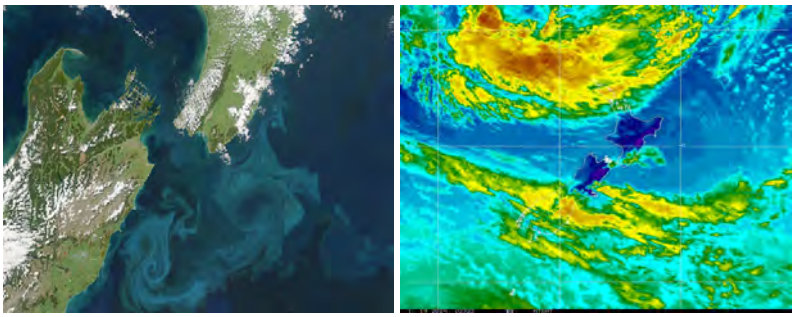
Figure 2.3: An example of continuous data collection: 2.3a real-time water quality monitoring, 2.3b raw data of real-time water quality monitoring, 2.3c real-time PM_x monitoring, and 2.3d raw data of real-time PM_x monitoring.

2.3 Data Problems and Challenges on Data Analysis

Environmental data have problems such as noise, spatio-temporal, endless data stream, and missing data. This problems makes a number of challenges in computational environmental analysis. This section explores the challenges from those problems.

2.3.1 Noise

Noise is often involved in environmental data. Natural ventilation, weather condition and equipment failure is the main causes of noise in air quality data collection (Scrimger, 1985). Raining, hailing and snowing often affects the quality of underwater data (Scrimger, 1985). Lighting, cloudy and weakness of spectral resolution



(a) (b)

Figure 2.4: An example of remote sensing data collection: 2.4a real-time land cover in New Zealand (visible wavelengths), 2.4b the temperature data of New Zealand from satellite (invisible wavelengths).

makes a numbers of speckle noise on satellite images (Celik, 2009). In computational environmental analysis, inclusion of noisy or irrelevant environmental variables can distort the investigation of an environmental problem (McCune, 1997), and most change detection and segmentation methods for land use observation are applied to the raw data domain thus suffer from the inference of speckle noise (Celik, 2009).

2.3.2 Spatio-temporal Characteristics

Environmental data are collected across time as well as space. A monitoring network (e.g., a network of meteorological stations) on which data are collected at regular intervals, say every day or every week. With the introduction of geographical location as a new dimension, spatio-temporal data are gaining popularity as a new form of environmental data representation.

Such complexity appears normally resulting from that environmental data which contains information of locations, time, and state of environmental condition (i.e., pollution level). Fig.2.5 shows an example of landscape image series with changes happening at different levels (e.g., individual, summary) collected periodically at different locations (Rasinmaki, 2003).

Consider the spatio-temperal characteristic of environmental data, the computational analysis has to consider spatial dependence of monitoring stations, and that the observations from each monitoring site typically are not independent but form a time series. The challenges of computational environmental analysis include, (1) The different length of observations, missing data, whether conditions, inevitable

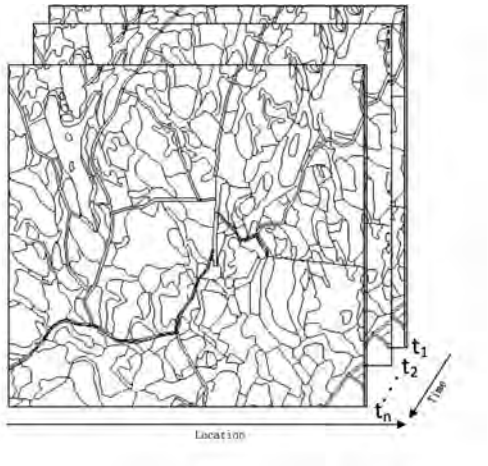


Figure 2.5: An example of spatio-temporal Data.

occasional failures, and open or close a monitoring site through the observation period in the corresponding stations makes huge problems in spatio-temporal data analysis; and (2) Environmental problems are often temporally rich in data, unfortunately sporadic collected spatially. One major short-coming documented in the literature is the ineffective use of spatial data for computational environmental analysis.

2.3.3 Endless Data Stream

The environmental data is collected every hour or every 8-hours or every day depending on the responses from monitor set. The monitoring is conducted continuously possibly without end. For air quality monitoring, the chemical compositions SO_2, NO_2, O_3, PM_x are recorded every day even every hour. For land use observation, the real-time satellites imagery is in a large size with high-resolution. In general, environmental data does not take the form of persistent relations, but rather arrives in multiple, continuous, rapid, time-varying data streams (Babcock, Babu, Datar, Motwani, & Widom, 2002).

In computational environmental analysis, these possibly unpredictable and unbounded streams cause some fundamentally research problems. Traditional database management and analysis models are not designed for rapid and continuous loading of individual data items, and they do not directly support the continuous queries (Terry, Goldberg, Nichols, & Oki, 1992).

In addition, errors and uncertainties are always presented in the latest data, how-

ever, minimizing the real-time uncertainties by recognizing and reducing possible errors rapidly and efficiently is a hard topic (Titov et al., 2005). Therefore, we explore that the challenges on real-time data stream analysis are accuracy, speed and robustness.

2.3.4 Missing Data

In environmental research, missing data often happen, usually due to faults in data acquisition, insufficient sampling or measurement error caused by non-response and break-down of equipment, and external (technical) reasons (Junninen, Niska, Tuppurainen, Ruuskanen, & Kolehmainen, 2004). Problems of missing data arise in most environmental studies, an incomplete data leads to result that are different from those that would have been gained from a complete dataset (Hawthorne & Elliott, 2005).

In practice if missing data exists, standard identification algorithms cannot be applied directly to pollution prediction. The missing data often interrupt the trend of an environmental problem, training models therefore lose the smooth operation unexpectedly (Xue et al., 2005). In addition, few missing data can seriously mislead the interaction between pollutants in training a prediction model (H. Li, Robertson, & Jensen, 2005). Although there are many techniques exist to recover data with missing values, but it has been considered as no one is absolutely better than the others (Graham, 2009).

2.4 Conclusion

In this chapter, we discussed the mechanism of environmental data collection and discover relevant problems on data quality. Environmental surveying provides accurate chemical parameters in water/air/soil, but it is often not possible or impractical to evaluate environmental condition in real time. Monitoring network uses flexible programmable equipment to take samples at fixed or variable time intervals, where the quality of data is highly influenced by many uncontrollable environmental factors. Remote sensing monitors the land cover problems in a large area using high resolution photography, on which noise may appear in data as a consequence of ambient lighting, weather phenomena and limitations of the spectral resolution.

Thus, the quality of environmental data is highly affected by uncontrollable environmental phenomena and unexpected equipment failure. For computational environmental analysis, noise and missing samples may seriously impact the usefulness of analysis model. In addition, environmental problems are all spatio-temporal

characteristic with data collected at different locations and at different time periods. Therefore, the challenge of computational environmental analysis is to handle spatio-temporal data streams with noise and missing samples.

Chapter 3
Environmental Change Detection

Environmental changes with its accumulating effects eventually destroy our agriculture, forest, dike etc., and even cause flooding. The detection of environmental change is a process of identifying differences in the state of an environmental phenomenon by observing it over times. Essentially, it builds the ability for us to quantify temporal effects using multi-temporal datasets, and issue early warning for a potential environmental problem. This chapter discusses environmental change detection with a focus on the environmental problems addressed by the thesis.

3.1 Introduction

Environmental changes include emergence of new pollutant, interaction between pollutants, changes of densities of pollutants and changes of pollution types and level. Well known environmental changes are global temperature changes, drinking water quality changes and land cover changes, which are caused by emission, sewage and storm-water disposal and urban expansion. Many potential issues may be brought in by such environmental changes.

One of those issues is pollution which has already become one of the most serious concerns of humans, since health problems caused by pollution have increased rapidly. Usually, pollution can be categorized as air pollution, water pollution and soil/land pollution. Each of them poses a grave threat to human. For example, World Health Organization (WHO) estimated that air pollution kills at least 2.4 million people every year (WHO, 2002). Guardian (2008) indicates that air pollution from motor vehicles is the most serious factor leading to pneumonia related deaths. Water pollution has been suggested to be a major global problem as it is the leading cause of deaths and diseases worldwide (Larry, 2006). Larry (2006) accounts the 14,000 daily deaths due to water pollution in 2006. Similarly, soil / land pollution emits hazardous substances into the air and groundwater (*Risk Assessment Guidance for Superfund (RAGS)*, 2009), causing a series of issues, such as land surface losing, desertification and deforestation. Therefore, it will be a significant and continuous challenge of controlling pollution.

Every environmental event starts from changes. Environmental changes indicate the beginning of pollutions and examine the effective of environmental protection strategies. For example, melting of the polar ice caps is caused by global temperature changes. Global temperature change is caused by increasing densities of carbon dioxide, methane, chlorofluorocarbons, ozone, and nitrogen oxides. A change of carbon dioxide density is caused by the density changes of burning carbon fuels in the air. Using of natural gas creates methane. Chlorofluorocarbons density changes mean people are using CFCs-MDI. Ozone density changes because of the use of electrical appliances. Nitrogen oxides are from toxic emission. If people are aware of where and what environmental changes happen, many diseases can be avoided. Therefore, environmental change detection is considered as the first step of all environmental protection system.

According to the detection methods for different types of pollution, environmental change detection is categorized as traditional discriminant methods and state-of-the-art methods. Traditional discriminant methods include manual methods and instrumental monitoring methods. State-of-the-art methods use mathematical models to analyze the environmental changes. Recently, air pollution, water pollution, and land pollution are increasingly under observation by satellite images. Thanks to accurate radiometric calibration, satellite images can provide easy-to-use and cost-effective data(Chavez, 1996). To develop change detection methods based on satellite images are more popular. Thus, satellite images change detection plays a very important role in environmental protection.

This chapter is organized as follows: In Section 3.2, we present the brief concepts of existing change detection methods and compare them from the literature. In Section 3.3, we propose new criteria for evaluating change detection methods. In Section 3.4, we introduce some applications of environmental prediction. Finally, a summary is provided in Section 3.5.

3.2 Computational Change Detection Methods

This section reviews change detection methods. Subsection 3.2.1 illustrates traditional discriminant methods. Due to the lack of technology and understanding of pollutants at the early stage, traditional discriminant methods were the major environmental change detection method. Since then our understanding of pollutants including types of pollutants and their complicated interactions has been greatly deepened. Thus statistical methods and mathematical models can be employed on environmental change detection methods, which is discussed in subsection 3.2.2. In the modern era, satellite images provide diversified and accurate information to

researchers, such as globe temperatures, color of sea and land cover changes. Therefore, the satellite images change detection has been widely applied on air, water and land pollution. Satellite image change detection methods are discussed in subsection 3.2.3.

3.2.1 Traditional Discriminant Methods

Sampling prior to statistical applications in environmental monitoring is the most popular method for environmental change detection, especially for air pollutants and water pollutants detection. Before state-of-the-art method, sampling does not only collect data, also detects changes directly. It is used to identify specific abnormal activity patterns in an observation. For example, pondus hydrogenii (pH) value in water represents the acidity or basicity of an aqueous solution (Longsworth, 1964). pH value at 25^oC in pure water is considered as 7.0. pH value is less than 7, water is in acidic condition, oppositely, pH value is greater than 7, water is in alkaline condition. Sampling method uses universal indicator components to evaluate pH value and compare a series pH value in a specific area to detect changes of acidity-alkalinity in water.

Since sampling methods use chemical containers or lighting to detect pollutants, it is categorized into manual methods and instrumental monitoring methods.

Manual methods

Manual methods sample air pollutants, and measure the levels of pollutants in the air/water by a standard. Manual methods for air pollutants detection include passive samplers, paper tape samplers and bubbler systems. It also detects pollutants from water such as porous ceramic cups detection method.

1. Passive samplers (Klanova et al., 2006) collect gaseous air pollutants onto a chemical container.

2. Paper tape samplers (Martinez et al., 2008) analyze gas samples by a special paper. Each pollutants are marked by a series of discrete spots on the paper.

3. Bubbler systems (Anderson et al., 1995) bubble the air by a special designed solution. Laboratory analysts then study the resulting solutions.

4. Porous ceramic cups samplers (E. A. Hansen & Harris, 1975) observe water samples' sorption, leaching, diffusion, and screening of phosphate ions on the cup walls.

Instrumental monitoring methods

Instrumental monitoring methods apply an instrument to measure, detect air/water pollutants through a quantitative chemical reaction.

Non-dispersive infra-red (NDIR), chemiluminescence, flame photometric analyzers and suspended particulate monitoring methods for air pollutants for air pollutants detection and immunofluorescence assay for water pollutants detection are the most popular detection methods at present. liquid chromatography-electrospray tandem mass spectrometry

1. NDIR (Fischer et al., 2000) is employed to measure the absorption peak of the pollutant molecule by light emitted from an infra-red source which is not dispersed.

2. Chemiluminescence (Loomis et al., 1999) determines the pollutant air component, like NO_x, by chemical reactions through light.

3. Flame photometric analyzers (Dagan, 2000) take photos for air pollutants, like sulfur, detected from chemical reactions within a hydrogen-rich flame and narrow light. The photos show capacities and types of air pollutants.

4. Suspended particulate monitoring methods (Pope et al., 2002) apply a high volume air sampler (an air sampling equipment) to collect particles (PM-10) from air, then calculate weight of particles.

5. IFA (Russell, Sampson, Schmid, Wilkinson, & Plikaytis, 1984) visualizes the distribution of the target molecule through samples by the specificity of antibodies to their antigen to target fluorescent dyes to specific biomolecule targets within a cell.

6. Liquid chromatography-electrospray tandem mass spectrometry (Hirsch et al., 1998) combines the physical separation capabilities of liquid chromatography with the mass analysis capabilities of mass spectrometry.

Traditional discriminant methods can be sumaried as (3.1) (Paavola, Ruusunen, & Pirttimaa, 2005)

$$\boldsymbol{\Delta}_t = |\boldsymbol{X}_t - \boldsymbol{X}_{t-1}|, \tag{3.1}$$

where $\boldsymbol{\Delta}_t$ is the difference, $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t-1}$ represents the current value of the parameter and the lagged value respectively. The alarm of change is $\boldsymbol{\Delta}_t \geq ThVal$, $ThVal$ means threshold. Earlier, $ThVal$ was set according to experts experience, it is useful and accurate for some specific pollutants. However, environmental changes are complex, one or few pollutants do not represent an environmental event. In order to

detect environmental changes with a scientific threshold from air / water samples, state-of-the-art methods are employed.

3.2.2 State-of-the-art Methods

Due to the complexity of pollution types and emergence of new pollutants, traditional discriminant methods are not able to detect and analyze environmental changes. In addition, digital computers have continued to increase in speed and capacity, making it possible to approach more complex pollution problems. Thus the change detection is improved by state-of-the-art methods.

State-of-the-art method uses statistical methods and machine learning methods to detect unacceptable or unseen change rather than identify specific abnormal activity patterns. Earlier, state-of-the-art method focuses on setting threshold by statistical analysis, then machine learning methods applies mathematical models in environmental change detection.

Statistical thresholding analysis

Statistical analysis such as sample standard deviation (Norman et al., 2003; Armitage, Moss, Wright, & Furse, 1983), t-test (Glass et al., 2000; Ennaceur & Delacour, 1988) and Pearson correlation coefficient (Y. Li et al., 2006; Robin & Denis, 1999) were popular environmental change detection methods prior to machine learning method. Such methods set change detection thresholds, rather than using fixed threshold, by statistical analysis.

Given observed values of the sample items $\{x_1, x_2, ..., x_N\}$, the standard deviation is calculated as:

$$s_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}. \tag{3.2}$$

where $\bar{x}$ is the mean value of these observations and s is an unbiased estimator. Norman et al. (2003) and Armitage et al. (1983) consider changes when s is far away from average area. Standard deviation method is not only used to detect large samples, it also can be used to evaluate changes in small samples by applying ratio of mean deviation, as defined:

$$s_N = \frac{\sum_{i=1}^{N}|x_i - \bar{x}|}{\left[N\sum_{i=1}^{N}(x_i - \bar{x})\right]^{\frac{1}{2}}}. \tag{3.3}$$

Based on standard deviation, t-test is developed as a thresholding method. Given

observed values of the sample items $\{x_1, x_2, ..., x_N\}$, the t-test is calculated as:

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{N}}}, \tag{3.4}$$

where $\overline{x}$ is the mean value of $\{x_1, x_2, ..., x_N\}$, μ_0 is a specified value and s is the sample standard deviation of the number of values. Brown, Hall, and Westerling (2004) set three level for climate changes, which is low change, no change and high level change.

Based on t-test, another thresholding method Pearson's correlation (K. Pearson, 1897) is employed into environmental change detection. Given time series $\boldsymbol{X} = \{x_1, x_2, \ldots, x_N\}$ and $\boldsymbol{Y} = \{y_1, y_2, \ldots, y_N\}$, the Pearson product-moment correlation coefficient ($\rho_{\boldsymbol{X},\boldsymbol{Y}}$) is calculated as:

$$\rho_{\boldsymbol{X},\boldsymbol{Y}} = \frac{cov(\boldsymbol{X},\boldsymbol{Y})}{\sigma_{\boldsymbol{X}}\sigma_{\boldsymbol{Y}}} = \frac{E((\boldsymbol{X} - \boldsymbol{\mu_X})(\boldsymbol{Y} - \boldsymbol{\mu_Y}))}{\sigma_{\boldsymbol{X}}\sigma_{\boldsymbol{Y}}}, \tag{3.5}$$

where cov is the covariance; $\sigma_{\boldsymbol{X}}$ and $\sigma_{\boldsymbol{Y}}$ are standard deviations; $\boldsymbol{\mu_X}$ and $\boldsymbol{\mu_Y}$ are the expected value; and E is the expected value operator. Practically, except $\rho_{\boldsymbol{X},\boldsymbol{Y}}$, Pearson's correlation returns a probability p-value (p). p-value in statistical hypothesis testing is the probability of obtaining a test statistic at least as extreme as the one that was actually observed ($\boldsymbol{Y}$ to $\boldsymbol{X}$), assuming that the null hypothesis is true. Null hypotheses are typically statements of no difference or effect. The p-values are crucial for their correct interpretation as they are based on this hypothesis. Therefore, a lower p-value or assumption of the null hypothesis can be thought of as the production of a statistically significant result. p is calculated based on (3.6):

$$t = r\sqrt{\frac{n-2}{1-r^2}}, \tag{3.6}$$

where r is sample correlation coefficient, n is number of pairs of data and degrees of freedom equal to $n - 2$.

Consider $\sigma^2_{\boldsymbol{X}} = E[(\boldsymbol{X} - E(\boldsymbol{X}))^2] = E(\boldsymbol{X}^2) - E^2(\boldsymbol{X})$ Due to $\boldsymbol{\mu_X} = E(\boldsymbol{X})$ and likewise for $\boldsymbol{Y}$. Also, $E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{Y} - E(\boldsymbol{Y}))] = E(\boldsymbol{XY}) - E(\boldsymbol{X})E(\boldsymbol{Y})$. (3.5) is often formulated with p as:

$$\begin{aligned} \rho_{\boldsymbol{X},\boldsymbol{Y}} = \quad & \frac{E(\boldsymbol{XY}) - E(\boldsymbol{X})E(\boldsymbol{Y})}{\sqrt{E(\boldsymbol{X}^2) - E^2(\boldsymbol{X})}\sqrt{E(\boldsymbol{Y}^2) - E^2(\boldsymbol{Y})}} \\ subject\ to: \quad & \mathrm{p} < 0.05, \end{aligned} \tag{3.7}$$

$\rho_{\boldsymbol{X},\boldsymbol{Y}}$ is ranged from +1 to -1, which follows that Pearson's correlation includes positive correlation and negative correlation. A positive correlation ($\rho_{\boldsymbol{X},\boldsymbol{Y}} \to 1$) means

that, as one variable/time series ($\boldsymbol{X}$) becomes large, the other ($\boldsymbol{Y}$) also becomes large, and vice versa. $\rho_{\boldsymbol{X},\boldsymbol{Y}} \rightarrow +1$ means a perfect positive linear relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$. In case of negative correlation($\rho_{\boldsymbol{X},\boldsymbol{Y}} \rightarrow -1$), as one variable ($\boldsymbol{X}$) increases the other ($\boldsymbol{Y}$) decreases, and vice versa. Note that Pearson's correlation $\rho_{\boldsymbol{X},\boldsymbol{Y}}$ is statistically significant, only if p is less than 0.05.

For example, Robin and Denis (1999) and Y. Li et al. (2006) recognize water quality and woody plants data changes when $\rho_{\boldsymbol{X},\boldsymbol{Y}} \neq 1$ or $\rho_{\boldsymbol{X},\boldsymbol{Y}} = 1$ and $p \leq 0.05$.

Machine learning method

Machine learning methods have also been employed to environmental change detection. Machine learning method is used to design and develop algorithms that allow computers to evolve behaviors based on empirical data. Different to statistical methods, machine learning methods focus on learning complex patterns and making intelligent decisions based on data (Craven et al., 2000) rather than analyzing, interpreting or explaining, and presenting data (Desrosieres, Naish, & Hunter, 2006). Therefore environmental change detection based on machine learning does not focus on analyzing threshold problem.

For example, multi-layer perceptrons (MLP) require class labels for each different type of data and calculate probabilities for environmental changes over time. It is outperformed than other method applied on oil flow in multiphase pipelines (Bishop, 1994). Another example is adaptive resonance theory (ART). ART can be considered as an alarm that detect changes in normal behavior without previous knowledge of changes (Dasgupta & Forrest, 1995). In addition, decision trees (Sesnie, Gessler, Finegan, & Thessler, 2008; Chu & Zaniolo, 2004), ensemble methods (X. Zhang et al., 2007; Giorgi & Mearns, 2002), k-nearest neighbours (kNN) (Ratle et al., 2007; Rebenitsch et al., 2010) and support vector machine (SVM) (Schwender et al., 2004; Mandalia, 2005) detect new changes and where changes happened in a time series environmental data by accurate classifier function.

3.2.3 Satellite Image Change Detection

Recently, satellite image processing has become an effective technology for detecting all kinds of environmental changes (Chelliah & Ropelewski, 2000; Llewellyn-Jones et al., 2001; Mas, 1999; Lo, 2000; C. Song, Woodcock, Seto, Lenney, & Macomber, 2001; Bruzzone & Prieto, 2002), including change of global temperatures, color of sea and land cover. Especially, small-scaled satellite image may contain plenty of information suitable for change detection in large areas. At early stage, researchers use simple mean methods (simple differencing) to detect changes between two images. Due to the influence from radiometric, atmospheric, solar and topographic

condition of satellite images, simple differencing methods cannot detect changes accurately. Then researchers consider setting thresholds instead of simple calculations in simple differencing methods to deal with noise on satellite images. Thresholds at first are set manually, so called manual thresholding method. After that, in order to setting thresholds conveniently, Gaussian distribution and Markov random field are applied for auto thresholding. In addition, classification methods are able to provide accurate change detection results after a proper training procedure. Thus, this section is divided into simple differencing, manual thresholding, auto thresholding and classification method.

Simple differencing

Simple differencing methods seek changes by computing a simply predefined variation measurement and authenticate the change by simple calculation such as mean difference and ratio of means. Mean difference (R. Jain, 1984) is differencing of pixels on two images (I_1 and I_2),

$$I_D(i,j) = I_2(i,j) - I_1(i,j), \tag{3.8}$$

where I_1 and I_2 are gray-scale multitemporal images with same size $M \times N$, i and j represent positions of pixels on images, $i = [1, 2, \ldots, M]$ and $j = [1, 2, \ldots, N]$. In order to deal with multiplicative noise, ratio of means then is developed. Similar to mean difference, ratio of means uses a ratio instead of the difference.

$$I_R(i,j) = \frac{I_2(i,j)}{I_1(i,j)}, \tag{3.9}$$

In fact, ratio of means is actually used:

$$I_R(i,j) = 1 - min\left(\frac{I_2(i,j)}{I_1(i,j)}, \frac{I_1(i,j)}{I_2(i,j)}\right). \tag{3.10}$$

After that, ratio of means (Yakimovsky, 1976) calculates the mean intensity in the neighborhood of each pixel from both images, denote pixels ($I_2(i,j)$) position as ($I_2(p)$), a block of pixels centered at $I_2(p)$ by Ω_p, the pixel values in the block are denoted $\hat{I}_2(p) = \{I_2(p1)|p1 \in \Omega_p\}$, the block contains B pixels. Then local mean $\mu_1 = \frac{\sum I_2(p1)}{B}$, same as μ_2, then,

$$r = 1 - min\left(\frac{\mu_1}{\mu_2}, \frac{\mu_2}{\mu_1)}\right), \tag{3.11}$$

$r = 0$ means no change and $r \neq 0$ means change.

Manual thresholding

The noise in satellite images may cause by atmospheric conditions, Sun angles, soil moistures and phonological differences in addition to true land-cover change (D. Lu, Mausel, Brondizio, & Moran, 2004), simple differencing methods are not able to deal with such noise. For the purpose of separating areas of change from those of no change in two image with noise, threshold (T) has been employed into change detection methods, which is shown in (3.12) (Singh, 1989).

$$D(i,j) = \begin{cases} 1 & \text{if } I_D(i,j) \geq T, \\ 0 & \text{if } I_D(i,j) < T. \end{cases} \tag{3.12}$$

Threshold can be derived from the histogram of the image (Schowengerdt, 1983), ratio values (Jensen, 1981), standard deviation (Mukai, Sugimura, Watanabe, & Wakamori, 1987), voting via multiple change detection method (Zhan et al., 2000).

Auto thresholding

Manual thresholding is not convenient and threshold has to be set differently in different task, therefore, auto thresholding is expected. Given image pair $\{I_1, I_2\}$ and its difference image $D(x)$ obtained from some simple differencing method described above. Auto thresholding methods describe the difference image $D(x)$ as a mixture distribution of its probability density functions on two competing hypothesizes $P(D(x)|H_0)$ and $P(D(x)|H_1)$ in which H_0 and H_1 denote unchanged and changed respectively, and represent them as a Gaussian distribution (Aach, Kaup, & Mester, 1993; Bazi, Bruzzone, & Melgani, 2005a; Celik & Ma, 2010) or in a Markov random field (Duskunovic, Heene, Philips, & Bruyland, 2000; Aach & Kaup, 1995). To minimize the overall detection error, optimization techniques like expectation-maximization (EM) algorithm (Celik & Ma, 2010) are employed for automatic threshold estimation. The EM algorithm for auto thresholding is shown in (3.13) to (3.24).

$$D(i,j) = |I_2(i,j) - I_1(i,j)|, \tag{3.13}$$

where $1 \leq i \leq M$ and $1 \leq j \leq N$. In order to create two opposite classes H_0 and H_1, they estimate posterior probability density functions $p(D(i,j)|H_1)$ and $p(D(i,j)|H_0)$ as well as priori probabilities $P(H_1)$ and $P(H_0)$ for each classes H_1 and H_0, respectively. The probability density function $p(D(i,j))$ established from the difference image is assumed to be a mixture of two Gaussian distributions,

$$p(D(i,j)) = p(D(i,j)|H_1)P(H_1) + p(D(i,j)|H_0)P(H_0), \tag{3.14}$$

where $P(H_1) + P(H_0) = 1$. Gaussian distribution $p(D(i,j)|H_1)$ and $p(D(i,j)|H_1)$ can be modeled by the mean μ_1 and variance ${\sigma_1}^2$ for the class H_1, same to μ_0 and ${\sigma_0}^2$ for the class H_0. The EM algorithm is an iterative process, which conduct the maximum likelihood estimation to tackle incomplete-data problems. There are two steps in each iteration: expectation step and maximization step. The iteration stops until the convergence is reached. The expectation step is updating the unknown underlying $P(H_1)$ and $P(H_0)$ based on the current estimates of parameters. The parameters are generated from the maximization step at the k^{th} iteration, and it is conditioned by the observations (3.15) to (3.17),

$$ {H_1}^k(i,j) = \frac{p^{(k)}(D(i,j)|H_1)}{p^{(k)}(D(i,j))}, \tag{3.15} $$

$$ P^{(k+1)}(H_1) = \frac{\sum_M^{i=1} \sum_N^{j=1} P^{(k)}(H_1) {H_1}^{(k)}(i,j)}{M \times N}, \tag{3.16} $$

$$ P^{k+1}(H_0) = 1 - P^{(k+1)}(H_1), \tag{3.17} $$

where (k) and (k+1) denote the iteration. The maximization step provides new estimates of parameters for Gaussian distribution (3.18) to (3.22):

$$ {\mu_1}^{(k+1)} = \frac{\sum_M^{i=1} \sum_N^{j=1} H_1^{(k)}(i,j) D(i,j)}{\sum_M^{i=1} \sum_N^{j=1} H_1^{(k)}(i,j)}, \tag{3.18} $$

$$ {\sigma_1}^{2(k+1)} = \frac{\sum_M^{i=1} \sum_N^{j=1} H_1^{(k)}(i,j)(D(i,j) - {\mu_1}^{(k)}}{\sum_M^{i=1} \sum_N^{j=1} H_1^{(k)}(i,j)}, \tag{3.19} $$

$$ {\mu_1}^{(k)} = \frac{p^{(k)}(D(i,j)|H_0)}{p^{(k)}(D(i,j))}, \tag{3.20} $$

$$ {\mu_0}^{(k+1)} = \frac{\sum_M^{i=1} \sum_N^{j=1} H_0^{(k)}(i,j) D(i,j)}{\sum_M^{i=1} \sum_N^{j=1} H_0^{(k)}(i,j)}, \tag{3.21} $$

$$ {\sigma_0}^{2(k+1)} = \frac{\sum_M^{i=1} \sum_N^{j=1} H_0^{(k)}(i,j)(D(i,j) - {\mu_1}^{(k)}}{\sum_M^{i=1} \sum_N^{j=1} H_0^{(k)}(i,j)}, \tag{3.22} $$

Class label therefore can be calculated in two classes according to (3.23),

$$ \begin{aligned} C(i,j) &= argmax_{H_k \in \{H_1, H_0\}} P(H_k|D(i,j)) \\ &= argmax_{H_k \in \{H_1, H_0\}} \frac{P(H_k) p(D(i,j)|H_k)}{p(D(i,j))}, \end{aligned} \tag{3.23} $$

After all calculations, the auto threshold can be calculated as (3.24):

$$D(i,j) = \begin{cases} 1 & \frac{p(D(i,j)|H_1)}{p(D(i,j)|H_0)} \geq \frac{P(H_0)}{P(H_1)}, \\ 0 & \text{otherwise..} \end{cases} \tag{3.24}$$

Also, researchers have considered to synthesize the analysis results from multiple different data sources (Moser, Serpico, & Vernazza, 2007) or models (Melgani & Bazi, 2006) to improve the actual performance of their proposed methods for change detection.

Classification method

Classification methods normally address change detection based on a classification model. Post-classification comparison also known as delta classification is the most widespread classification method for image change detection (Coppin, Jonckheere, Nackaerts, Muys, & Lambin, 2004). It consists of two steps: (1) classify observations (i.e., pixels or blocks) into land-cover conceptual classes (e.g., grass, water, desert) to obtain the conceptual/thematic map of I_i and I_j respectively; and (2) compare the two conceptual maps to identify land-cover changes. An alternative supervised method is to use directly a classification model for pixel change detection. Dengkui et al. (2008) classify pixel pairs into changed and unchanged classes for detecting forest land-cover transformation, where pixel pairs selected manually from changed/unchanged area are used for training. To get rid of the affection of illumination to the change detection system, unimportant changes (e.g., changes due to illumination variation) are counted as unchanged in the training of the classification model.

3.3 Performance Evaluation

This section presents in-depth evaluation of existing change detection methods, where environmental changes detection is measured by change detection accuracy, convenient threshold, detection sensitivity, and detection continuity.

3.3.1 Accuracy

An accurate change detection is required for environmental analysis. Environmental data includes many variables. In order to identify changes in air or water, an environmental change detection is required to report exactly all differences amongst a series of variables for each data collection.

Traditional discriminant methods have been employed into environmental change detection for a long time and it is still been used widely. Accuracy of traditional discriminant methods can be evaluated based on measurement error, which includes random error and systematic error (Keith, 1991).

1. Random error represents the degree to which data randomly selected from replicate or repeat measurements apart from another. It usually uses standard deviation (Paternoster, Brame, Mazerolle, & Piquero, 1998; Boor, Overmars, & van der Stappen, 1999; Robertson, 1959) or relative percent different (Abella & Covington, 2004; Groves, 2006).

2. Systematic error is the difference between limiting average value of measurement series and their true value (Kirchmer, 1987). It estimates the bias in the measurements via comparing the limiting average value analytical result with the true value (Keith, 1988).

The evaluation of state-of-the-art change detection methods calculates their accuracies similar to sampling methods, because they use thresholds to extract useful information in the same way of sampling methods. Recently, machine learning methods represent accurate detection for each pollutant because of advanced classification function. They train mathematical model from previous knowledge and series of data, then classify pollutants even detect new variables (Bishop, 1994; Dasgupta & Forrest, 1995).

Satellite image change detection concentrates more on pixel / a block of pixels detection (R. Radke, Andra, Al-Kofahi, & Roysam, 2005), so the accuracy of the change image can be considered as change / no change detection and from-to change detection level (Mas, 1999). The quantity evaluations of change detection accuracy can be considered in Table 3.1.

Table 3.1: The Quantity Evaluations of Change Detection Accuracy

TP (true positives)	the number of change pixels correctly detected
FP (false positives)	the number of no-change pixels incorrectly detected as change (also known as false alarms)
TN (true negatives)	the number of no-change pixels correctly detected
FN (false negatives)	the number of change pixels incorrectly detected as no-change (also known as misses)

Based on Table 3.1, Rosin and Ioannidis (2003) proposed three other approaches, which are the percentage of correct classification (3.25), jaccard coefficient (3.26) and

yule coefficient (3.27)

$$PCC = \frac{TP + TN}{TP + FP + TN + FN}; \tag{3.25}$$

$$JC = \frac{TP}{TP + FP + FN}; \tag{3.26}$$

$$YC = \left| \frac{TP}{TP + FP} + \frac{TN}{TN + FN} - 1 \right|. \tag{3.27}$$

Those criteria can be used for the evaluation of image change detection. Machine learning methods such as SVM (Bovolo, Bruzzone, & Marconcini, 2008; Nemmour & Chibani, 2006; Camps-Valls, Gomez-Chova, Munoz-Mari, Rojo-Alvarez, & Martinez-Ramon, 2008) is able to detect accurately changes, but have difficulty in collecting data for training (Celik & Ma, 2010).

3.3.2 Thresholding

As discussed above, thresholding methods have been popularly used for detecting environmental changes. Environmental analysis requires the threshold to be dynamically adjusted, because the status of pollutant changes over the time, and pollutants interacts between each others. For example, carbon oxides are considered to be harmful pollutant to human. In fact, CO_2 is not poisonous to human. However, because of the imperfect combustion of fuel, CO emits with CO_2. CO is a pollutant to environment. Nitrogen (N_2) is not a pollutant, however when N_2 interacts with oxygen (O_2) in different temperature, lots of nitrogen oxides are created, all of them are dangerous to human. In addition, everybody needs O_2, but high density of O_2 makes human sick. Those complex interactions in environment make that environmental changes is very hard to evaluate. According to the discussion in section 3.2.2, automatic change detection could play a very effective role in monitoring the environmental changes (Celik & Ma, 2010).

In order to analyze complex environmental changes, existing studies concentrate on setting thresholds automatically for diverse environmental change detection tasks. Traditional discriminant methods use fixed threshold and compare current with previous densities/amount of pollutant, where the threshold is normally determined by previous experiences (Rossi & Braun, 1997; Nelson, Shorter, McManus, & Zahniser, 2002; Lazic, Colao, Fantoni, & Spizzicchino, 2005). Statistical thresholding analysis sets thresholds by standard deviation, t-test and Pearson correlation coefficient for an observation of a group of pollutants. It allows thresholds to be set for different locations and without considering previous experiences (Bazi,

Bruzzone, & Melgani, 2005b; Moser & Serpico, 2009), but considers no variation of environmental changes over time. The threshold setting is difficult for image change detection, because satellite image data includes often noises from conditions of radiometric, sensor, atmospheric, solar, topographic and geometric. For dynamical thresholding, machine learning methods are normally applicable to employ unsupervised learning models, such as MRF (Bruzzone & Prieto, 2000) and EM (Celik & Ma, 2010) to set an adaptive threshold based on Bayesian framework.

3.3.3 Sensitivity

To achieve early warning, it is always significant to detect negative environmental change before any environmental disaster really happens. The sensitivity of change detection is important for environmental analysis in that an environmental problem cannot happen within one day, there is a procedure of environmental change before the problem becomes serious and widely recognizable (Fries et al., 2005). For example, ozone depletion causes global climate changes and health risks, and brings excessive ultraviolet radiation problems to human (de Gruijl et al., 2003). It is much easier to take action at early stage of an environmental problem. Thus, high sensitivity is required for change detection methods.

Sensitivity for change detection refers to the capability for detecting those minute and easily neglected changes. Note that sensitivity here is different from accuracy, and it can be obtained by just setting narrow thresholds and detecting more changes. It detects where/what changes when it first happens and reports not only numbers or percentages, but also the probability of pollution.

Existing change detection methods look merely the detection accuracy. As we know 90% can be considered as a good accuracy, but it doesn't mean the sensitivity for some pollutant. Traditional discriminant methods detect changes with high accuracy. The sensitivity is limited because those methods report only one specific variable based on interaction among all variables (Klanova et al., 2006; Martinez et al., 2008; Fischer et al., 2000; Loomis et al., 1999; Dagan, 2000). For some specific variables, traditional discriminant methods are able to detect changes sensitively, such as nitrogen oxides and hydrides of nitrogen (QI, 2003). However, traditional discriminant methods are not able to sensitively detect systemic change, especially in complex interaction system such as overall water quality.

State-of-the-art methods utilize standard deviation, t-test and Pearson correlation coefficient to set threshold. The only way to enhance the sensitivity in those change detection methods is to narrow the thresholds. However, narrowed thresholds not only extract slight changes but also amplify the effect of noise which leads to more false alarm. Machine learning methods have good potential for sensitive

environmental change detection. The difficulty is training data collection as such methods require often knowledge of larger number of progressive statues machine training.

For satellite image change detection research, sensitivity has been discussed to be evaluated by an equation like (accuracy-change-detection). As seen from (R. Jain, 1984; Yakimovsky, 1976; Bruzzone & Prieto, 2000; Celik & Ma, 2010; Bovolo et al., 2008; Bazi et al., 2005b; Gamba, Dell'Acqua, & Lisini, 2006), both manual and auto thresholding methods provide sensitive results for land cover changes. However, similar to the problem of statistical thresholding method, sensitivity of those methods is being increased at the price of sacrificing accuracy. Also, supervised machine learning has been discussed in (Coppin et al., 2004; Dengkui et al., 2008; Celik & Ma, 2010) for early change detection, whereas they suffer from the difficulty of collecting data for discriminating slight differences.

3.3.4 Continuity

Environmental changes may be rapid and significant or slow and inconspicuous. Also, such changes actually continues over time, possibly without end. Thus, continuity is another criterion to check if a method supports continuous change detection, and if reconstruction/re-training is needed for continuous change detection. A typical example is biological invasions. Though human has attempted to manage biological invasions for long time, but invasions continuously occur among rivers, lakes and canals, even if they are very clean (Vitousek et al., 1996). Monitoring environmental changes like biological invasions, is a challenge of long term environmental change detection. In general, the difficulties of continuous environmental change detection are given below:

1. environmental data comes in real time;
2. environmental change involves huge size of data stream or matrix;
3. environmental data comes with unknown pollutants;
4. environmental changes are required to be detected instantly without retraining detection model.

It is worth noting that existing change detection methods perform very limited on continuous environmental change detection. Traditional discriminant methods are inefficient because they mostly compare samples one by one, thus they support only instant use and they are not able to detect complex environmental changes within a period of time. The limitation of statistical analysis is the thresholding

difficulty, which is, the threshold value is normally fixed by experience for case by case. Machine learning method is able to simulate a procedure of environmental change, which indicates us a good direction for analyzing continuous environmental changes. However, this method requires a huge amount of training data which may not be available in practice. Moreover, the learning models often need to be retrained whenever new observation arrives.

3.3.5 Discussion

For environmental change detection, previous studies focus on improving detecting accuracy and enabling dynamic thresholding. Early detection and continuous pollution detection are considered as additional two requirements for environmental analysis. Table 7.1 summaries the evaluation of each reviewed method according to all four criteria, which include, accuracy, thresholding, sensitivity and continuity. As seen from the table, there is currently none existing method that meets all 4 crite-

Table 3.2: Evaluation of Change Detection Methods According to The Proposed Method Evaluation Criteria

Method	Accuracy	Thresholding	Sensitivity	Continuity
Data Stream change detection				
Traditional discriminant methods	√	×	×	×
State-of-the-art Methods	√	×	×	×
Statistical Thresholding Analysis	√	√	×	√
Machine learning Method	√	Do not need	√	√
Data Matrix change detection (Satellite image change detection)				
Simple differencing	×	×	×	×
Manual thresholding	×	×	×	×
Auto thresholding	√	√	×	√
Classification method	√	Do not need	√	√

ria. Machine learning method has large potential, but it has a number of limitations as discussed in subsection 3.3.3 and 3.3.4. In this sense, we conclude that machine learning method for continuous and accurate change detection is a valuable direction for future research.

3.4 Applications to Environmental Problems

Pollution includes air pollution, water pollution and soil/land pollution. In order to detect different types of pollution for environment monitoring, many change detection applications have been developed. In the early age, environmental change

detection works as cleaning equipments and air/water indicators for monitoring environmental system. Since the advancement of remote sensing technology and improvement of satellite images resolution, it then has been applied on area surveillance/land cover changes detection. Furthermore, for the purpose of ecological restoration, environmental change detection for image data also plays very important role. The following sections will discuss the applications of environmental change detection on condition indicator, area surveillance and ecological restoration respectively.

3.4.1 Pollution Emissions

Cleaning equipment has been widely used in our life, such as house cleaning equipment, field cleaning equipment, purging equipment and decontamination systems. An simple example of house cleaning equipment is dehumidifier, the first step of an automatic dehumidifier is analyzing air humidity, change detection therefore becomes the core part of every automatic dehumidifier (Matsuzawa et al., 1984). Field cleaning equipment is often used to clean contaminated ground water. In order to detect the pollutants in ground water, change detection methods are applied (Francke et al., 2000). To prevent air quality, purging methods are normally used, such as fuel exposure. For the purpose of control emission with purging equipment, change detection methods work as an alarm system to transform information into purging systems and help purging systems to filter emission with high efficiency (Carlton & Smith, 2000). Before decontamination of water and air, the pollutants information is required, detection in decontamination systems is also play a very important role (Wei, Rogers, & Mannan, 2006).

3.4.2 Pollution Measurement

During the pollution monitoring procedure, condition indicators have been widely applied. As change detectors, condition indicators are set for measuring the variation in environmental media, such as air, water and land. In the early stage, the indicator identifies whether a change happened between two situations by checking the correlation coefficient value r, such as $r = 1 \rightarrow$ no change and $r < 1 \rightarrow$ change for desertification detecting (Coiner, 1980). These indicators allow researchers to trace desertification progress in arid and semi-arid regions such as Mali in West Africa. With the improved resolution of satellite images, Percy and Ferretti (2004) developed a new monitoring concepts to monitor forest health problem caused by air pollution. They use a stated probability level (e.g.,$P < 0.05$) to measure the status and degree of change has been integrated in the indicator to enhance its per-

formance. Then, their study is available to understand the nature, clearly express the objectives, define a sampling design tailored to address the objectives, and implement Quality Assurance (QA) procedures.

3.4.3 Land Cover Change

Area surveillance detects new activities and events over a geographic area, such as facility construction/demolition, deforestation, desertification, flooding, etc.

In order to detect damage levels caused by special events (e.g., earth quake), change detection has been employed. For example, Vu et al. (2004) develop a Light Detection And Ranging (LIDAR) based system to detect damaged buildings due to earthquake for updating building inventory database. The LIDAR system detects change of buildings via extracting accurate height information as well as spectral information, and filtering shadows information of skyscrapers by a laser scanner. Through this way, the damage levels of buildings can be obtained by the fast high-level automation LIDAR system with low computational cost. Another example, in order to identify changes in vegetation coverage for analysis construction in urban area, Haverkamp and Poulsen (2003) propose a monitoring approach. Change detection in their approach supplies clearer distinction of change at a high resolution. Thus, the magnitude of changes on contraction can be detected accurately.

There are also many change detection applications on deforestation, like detecting changes among satellite images. Since satellite images provide accuracy record for deforestation, satellite image change detection therefore outperform other monitoring methods. Spectral change detection approach has been employed into multitemporal Landsat satellite images for analyzing deforestation (Phua et al., 2008). Spectral change detection is applied as an alarm. It helps their system to formulate effective protections by tracing the significant events in deforestation in Kinabalu Park, Sabah, Malaysia. For another example, D. Liu, Song, Townshend, and Gong (2008) detect forest coverage change by employing local transition probability models which are based on Markov random field. The goal of their study is to replace existing multi-temporal change detection method and apply change detection to real world spatio-temporal environmental problem. The gain from their method for forest coverage change detection is that the error of classification on individual maps has been reduced. Therefore, the provided analysis data is more accurate than the data given by previous analysis method.

Monitoring desertification is another application of change detection. In order to analyze effect of land use change on desertification in the Iberian Peninsula, Hill, Stellmes, Udelhoven, Roder, and Sommer (2008) proposed an adaptation of the syndrome approach (Sahali et al., 2002) based on multi-temporal satellite image data

to analyze desertification problem for estimating characterized vegetation dynamics. The approach allows researchers to trace important land use change on the Iberian Peninsula by interpreting time series parameters derived from hypertemporal satellite archives (as provided by NOAA-AVHRR, SPOT VEGETATION, MODIS and MERIS). Otindag Sandy Land in China has suffered severe desertification in the last decade, it is one source of sandy dust storm. For achieving the goal of controlling degradation of Otindag Sandy Land, H. Liu, Zhou, Cheng, Long, and Li (2008) format a spatial change detection method, which integrates remote sensing, Geographic Information System (GIS) and field survey to analyze sandy desertification dataset. The results from their method provide accurate information about increasing of fixed sand dunes shrank, semi-fixed and active sand dunes, and decreasing of inter-dune grassland and wetlands. Government then forms project such as Grain-for-Green project with enacted regulations for controlling desertification efficiently.

Change detection methods also have been used in monitoring flooding. For example, in 1998, two floods occurred in Nenjiang and Songhua River Basins. For monitoring these floods and offering necessary information for flooding control and disaster relief, an approach based on knowledge-based RBF neural networks model to extract the dynamic flooding information from AVHRR images was created (Zhou, Luo, Yang, Li, & Wang, 2000). The approach employs few threshold-based segmentation and texture change detection methods on AVHRR images to extract information on flooding duration and depth. Thus, floods monitoring works real-time and in all-weather to provide necessary information for flooding control and disaster relief. Another example, in order to monitor land cover change caused by anthropogenic activities and extreme natural events, Zhan et al. (2002) produce Vegetative Cover Conversion (VCC) detection product. The product was utilized on analyzing floods data in Cambodia and Thailand. Change detection in this product is designed to serve as an alarm for rapid land cover conversion. The date given by VCC product is accurate and necessary for flooding analysis, and then, problems caused by flooding can be analyzed and traced.

3.4.4 Ecological Upset

Except analyzing pollution problems, ecological restoration is another important topic. Environmental change detection methods have been employed to preserve and protect aquatic resources. For example, in order to investigate effects of development of an agricultural area on lagoon and its aquatic resources, Ardli and Wolff (2009) utilized supervised learning change detection to find out changes in Segara Anakan lagoon (SAL), Java, Indonesia (an agricultural area). Their system achieved acceptable accuracy for early detection which is essential to ecolog-

ical restoration. Another example of environmental change detection application on ecological restoration is monitoring ecotones of landscape. For the purpose of ecotones delineating, Fortin et al. (2000) propose an ecotone detection method to detect spatial patterns, statistics to quantify and contrast patterns, and modeling to formulate and predict multivariate interactions. The method provides accurate result of identifying the heterogeneous zone. The long-term monitoring for ecotones of landscape therefore is available.

3.5 Conclusion

Change is the key of environmental analysis. Every type of pollution starts from certain environmental changes. Detecting and studying such environmental changes are the basis of environmental pollution monitoring, examination, and prediction. This chapter reviews existing computational methods for environmental change detection, discussed actual requirements of environmental change analysis, and identified future directions of research for computational environmental change detection.

Existing environmental change detection methods include traditional discriminant methods and state-of-the-art methods. Traditional discriminant methods use simple differencing evaluation, such as a fixed threshold to detect pollution. These methods address merely one single specific pollutant (e.g., NO_x). State-of-the-art methods model a computational hypothesis to examine pollution dynamically. These methods are capable of analyzing the pollution types, amount, and even pollutants interactions. It is worth noting that satellite image change detection is widely researched in computational environmental analysis, in which both simple differencing and state-of-art methods are investigated towards auto thresholding pollution detection and classification.

For performance evaluation, typical environmental change detection concerns change detection accuracy and auto thresholding. In this chapter, we proposed additionally two new measurements, which are sensitivity and continuity of change detection. We argue that evaluation of environmental change detection should follow four criteria, which include accuracy, auto thresholding, sensitivity, and continuity.

Environmental change detection has been widely applied to detect air pollution, water pollution and soil/land pollution. Additionally, emission, vegetation, global temperature, and global sea level are all under observation of environmental change detection. In general, data collected for environmental change detection appears in various forms, including data matrix and data stream. In practice, an environmental

problem involves often data streams for locations (i.e., spatial), time (i.e., temporal), and state of the environment. However, existing applications are merely spatial or temporal study, not the desired spatio-temporal change detection system.

Chapter 4

Environmental Prediction

The future status of pollution is required to be exposed to public, as we need to always keep people free from a contaminated environment. In this concept, environmental prediction plays an important role in our environmental management system. It predicts future pollutant levels, trend, nature response, and environmental behavior for pollution control in specific of prevention of future decline in environmental status. This chapter discusses computational models used for environmental prediction with a focus on air pollution prediction.

4.1 Introduction

Pollution poses a significant threat to human health and the quality of life of millions of people around the world. Gases like sulfur dioxide and nitrogen oxide can cause acid rain (Schmalensee, Joskow, Ellerman, Montero, & Bailey, 1998). Water pollution in terms of Oil spill may lead to death of several wildlife species (Patten, 1992). Also, the decrease of air quality might result in asthma and/or lung cancer of human (Friedman, Powell, Hutwagner, Graham, & Teague, 2001). Thus, the current and future environmental status is always required by the public to make sure they are safe in their living environment.

Comprehensive overall estimates of environmental pollution help people in their everyday lives, and aid decision-making related to pollution control and the implementation of preventive actions to reduce emissions. Pollution prediction provides the possible future pollutant levels, trend, and environmental behavior to the people who take responsibility for pollution prevention and environment management.

Pollution prediction is conducted through numerical weather prediction in which mathematical models are used to generate either short-term pollution forecasts or long-term environmental predictions. In general, computational models for environmental prediction are categorized as statistical models and artificial intelligence models. Statistical models calculate pollutants levels based on pollutant level in history rather than considering other environmental variables. These approaches can only be used for short-term prediction. Artificial intelligence models learn knowl-

edge from all available environmental variables for the purpose of simulating trends of pollutants concentrations that leads to long term prediction availability.

This chapter is organized as follows: In Section 4.2, we present the brief concepts of existing prediction models and compare them from the literature. In Section 4.3, we discuss the performance evaluation of environmental prediction models. In Section 4.4, we introduce the application of environmental prediction on air quality control. Finally, a summary is provided in Section 4.5.

4.2 Computational Environmental Prediction Methods

According to input variables of forecasting models, we categorize environmental prediction into statistical model and artificial intelligence model.

4.2.1 Statistical Models

A statistical model is identical to time-series model as,

$$Y_{t+1} = f(Y_t : t \in T), \tag{4.1}$$

where Y_t is the value of pollution level at time t.

In literature, statistical forecasting models such as autoregressive model(AR), moving-average model (MA), autoregressive moving average model (ARMA) and auto regressive integrated moving average model (ARIMA) have been used in many environmental prediction works.

AR describes certain time-varying processes in time-series problems. The model is defined as

$$Y_{t+1} = c + \sum_{i=1}^{p} \varphi_i Y_{t-i+1} + \varepsilon_{t+1}, \tag{4.2}$$

where $\varphi_i, \ldots, \varphi_p$ are parameters of input model, c is a constant, and ε_t is white noise (a random signal with a constant power spectral density).

Autoregressive model as the fundamental level of prediction work has been applied to environmental problem prediction. Beven and Freer (2001) applied AR model to a rainfall-runoff model to predict annual gross precipitation in Maimai, a M8 (3.8 ha) catchment locate in the Tawhai State Forest, North Westland, South Island, New Zealand. Lichstein, Simons, Shriner, and Franzreb (2002) combined autocorrelation and autoregressive models to predict numbers of three common Neotropical migrant songbirds with breeding habitat in the southern Appalachian Mountains of North Carolina and Tennessee, USA. Elith and Leathwick (2009) used

AR model to predict ecological and evolutionary distributions across landscapes in Australia and New Zealand.

In principle, AR model specifies that the output variable depends linearly on its own previous values. However, environmental problems are mostly nonlinear, thus the usage of AR model for environmental prediction is limited.

MA prediction is a univariate time series (historical data) model formulated as,

$$Y_{t+1} = \varepsilon_{t+1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i+1}, \tag{4.3}$$

where q is a selected window size for observation; $\theta_1 \dots \theta_q$ represent the values of each time instance in the selected window; and ε_t is white noise.

Moving-average model is a simple prediction model that often been applied to air quality prediction. Schwartz and Marcus (1990) predicted SO_2 concentrations by MA model in London for the winters of 1958 - 1972 to analyse the relation between air pollution and mortality. Touloumi, Pocock, Katsouyanni, and Trichopoulos (1994) applied MA prediction model to estimate short-term effects of air pollution on daily mortality in Athens by analyzing 8-hour moving average and 24-hour moving average of CO concentrations. Based on moving average model, Perez and Reyes (2002) embedded MA prediction in a multilayer neural network to predict 24-h average of PM_{10} concentrations in Santiago, Chile.

Compared to AR model, fitting an MA model is more complicated in that the lagged error term of (4.3) are not observable. MA model is useful in the environmental predictions that can be modelled by low-order polynomials, however it is incapable of providing an effective solution to a complex non-linear environmental problem.

ARMA is a prediction model which integrates the p autoregressive (4.2) and q moving-average (4.3). The prediction by ARMA is calculated as

$$Y_{t+1} = c + \varepsilon_{t+1} + \sum_{i=1}^{p} \varphi_i Y_{t-i+1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i+1}. \tag{4.4}$$

In the field of environmental prediction, ARMA has been used to predict earthquake motions, wind speed and level of particulate matter. Polhemus and Cakmak (1981) proposed a fitting stationary ARMA model to estimate earthquake ground motions by site-specific characteristics. Taylor, McSharry, and Buizza (2009) predicted wind power density in five U.K. wind farm locations by ARMA model in 2009 . Goyal, Chan, and Jaiswal (2006) used the ARMA model for the prediction of the respirable suspended particulate matter in Hong Kong and Delhi.

The ARMA integrates linear prediction model 'AR' with simple non-linear prediction model 'MA'. The model is capable of predicting environmental events with

advantages of simplicity, cost effectiveness and accuracy for timely forecasting without complex high-order parametric computations. However, ARMA is unable to provide solutions to high-order polynomials, which is required in pollution variations estimation.

ARIMA model, the most general class of models for forecasting a time series, is a generalization of an autoregressive moving average (ARMA) model. To predict future points in the series, ARIMA model is applied in some cases where data show evidence of non-stationarity.

The model includes three non-negative integer parameters p,d and q, which refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. If one of the three parameters is "0", then the model drops to "AR", "I" or "MA". For instance, ARIMA(0,1,0) is I(1), and ARIMA(0,0,1) is MA(1).

The ARIMA model is formed by two models. The first is non-stationary:

$$Y_{t+1} = (1 - L)^d Y_t, \tag{4.5}$$

and the second is wide-sense stationary:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) Y_{t+1} = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_{t+1}. \tag{4.6}$$

The forecast for the process Y_{t+1} uses a generalization of the method of autoregressive forecasting.

In computational environmental analysis, the ARIMA model has been applied in ambient pollutants prediction. Kumar and Jain (2010) predicted O_3, NO, NO_2 and CO respectively corresponding to their own historical time series in Delhi, India. T.-C. Yu, Lin, Lee, Tseng, and Liu (2012) applied ARIMA on predication of CO_2 concentrations in an indoor environment based on the historical CO_2 time series collected in many places. Kumar and Jain (2010) used ARIMA model to forecast daily mean PM_{10} and daily mean NO_2 for 15 stations for the duration 1-Mar-07 to 31-Dec-07 over the domain Belgium.

ARIMA models solve the non-stationary problem, but they do not solve the nonlinear problem associated with environmental analysis.

4.2.2 Artificial Intelligence Models

Artificial intelligence (AI) model learns knowledge from massive amounts of multidimensional data to predict future environmental events. Typically, an AI model for prediction can be formulated as

$$Y_t = f(\boldsymbol{X}_t), \tag{4.7}$$

where Y_t is the observation value at time instance t and $\boldsymbol{X}_t$ is a collection of environmental variables. Note that although statistical model uses only single variable as input, AI model computes on multiple variables.

Well-known prediction models in previous literature include Hidden Markov Models (HMM), Artificial Neural Network (ANN) and Support Vector Machine Regression (SVR).

A hidden Markov model (HMM) can be considered as the simplest dynamic Bayesian network for a stochastic process on optimal nonlinear filtering problems (Baum & Petrie, 1966; Baum & Eagon, 1967; Baum & Sell, 1968). The model uses generally a probabilistic framework to model a time series of multivariate observations. The inputs of HMM model include sequence $\boldsymbol{X}^{R\times T}$, transition probabilities and initial probabilities.

Given data $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t, \ldots, \boldsymbol{x}_T\}$, all $\boldsymbol{x}_t$ is generated by a hidden state s_t. The states follow a Markov chain. The future event is independent of the past,

$$P(s_{t+1}|s_t, s_{t-1}, \ldots, s_0) = P(s_{t+1}|s_t). \tag{4.8}$$

The transition probabilities is given as,

$$\alpha_{k,I} = P(s_{t+1} = I|s_t = k), \tag{4.9}$$

where $k, I = 1, 2, \ldots, M$ and M is the total number of states. The initial probabilities of states can be represented as π_k,

$$\sum_{I=1}^{M} a_{k,I} = 1 \text{for any } k, \sum_{k=1}^{M} \pi_k = 1. \tag{4.10}$$

Then, transition probabilities

$$\begin{aligned} P(s_1, s_2, \ldots, s_T) = &\quad P(s_1)P(s_2|s_1)P(s_3|s_2)\ldots P(s_T|s_T - 1) \\ = &\quad \pi_{s_1}\alpha_{s_1,s_2}\ldots\alpha_{s_{T-1},s_T} \end{aligned} \tag{4.11}$$

is finalized.

Next, we estimate the transition probabilities. Given two conditional probabilities $L_k(t)$ and $H_{k,I}(t)$. Firstly, in state k at position t, entry the observed sequence $\boldsymbol{X}$, let

$$L_k(t) = P(s_t = k|u) = \sum_s P(s|u)I(s_t = k), \tag{4.12}$$

then, in the state k at position t and in state I at position $t + 1$, after inputting $\boldsymbol{X}$, we have

$$\begin{aligned} H_{k,I}(t) &= P(s_t = k, s_{t+1} = I|u) \\ &= \textstyle\sum_s P(s|u)I(s_t = k)I(s_t + 1 = I), \end{aligned} \tag{4.13}$$

In the end, the state of $\boldsymbol{X}_{T+1}$ can be predicted by the estimated transition probabilities and estimating probabilities (instead of initial probabilities). A predicting model are trained, and Y can be calculated by the state of $\boldsymbol{X}_{T+1}$.

HMM in practice, has been applied to many air pollution problem prediction works with the ability in providing probabilities for the future states. Couvreur, Fontaine, Gaunard, and Mubikangiey (1998) find that HMM-based approach performs well on predicting the average spectrum of noise from emission events car, truck, moped, aircraft and train from the MADRAS database. For predicting $PM_{2.5}$ concentrations at O'Hare airport in Chiago based on a mass spectrometry database, Dong et al. (2009) apply HMM on data with temperature, pressure, cloudiness, wind speed, solar radiation, dewpoint and humidity. In U.K., Senawongse, Dalby, and Yang (2005) predict phosphorylation sites by HMM based on a data with biological contents.

HHM models can combine a variety of knowledge sources knowledge sources into a single HMM to constract a knowledge library. By representing all possible knowledge sources as HMMs, the prediction task becomes a library search in an enormous HMM. However, $P(y)$ must be independent of $P(x)$, which means the knowledge from historical trend is not able to be learned.

Artificial Neural Network (ANN), also known as "Neural Network" (NN) is a mathematical model or computational model that simulates the structure and/or functional aspects of biological neural networks. There are interconnected groups of artificial neurons and process information, which use the connectionist approach for computation inside an ANN model. Mostly, ANN can be seen as an adaptive system which changes its structure based on external or internal information that flows through the network during the learning phase. It is a non-linear statistical data modeling tool and can be used to model complex relationships between inputs and outputs or to find patterns in data. (Mitchell, 1999)

The structure of an ANN as shown in Fig.4.1, is an interconnected group of nodes. ANN time series prediction uses a group of interconnected functions to calculate $\hat{\boldsymbol{x}}(t+\Delta_t)$ by analyzing $\boldsymbol{x}$ within t period. Suppose an ANN has n composition functions. The ANN function $f(\boldsymbol{x})$ is defined by other composition functions $g_i(\boldsymbol{x})$. Thus $f(\boldsymbol{x}) = (g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), ...g_n(\boldsymbol{x}))$. The commonly used type of composition is the nonlinear weighted sum shown in (4.14)

$$f(\boldsymbol{x}) = K\left(\sum_i w_i g_i(\boldsymbol{x})\right), \tag{4.14}$$

where K is some predefined function, such as the hyperbolic tangent. It will be convenient to refer to a collection of functions g_i as simply a vector $g = (g_1, g_2, \ldots, g_n)$.

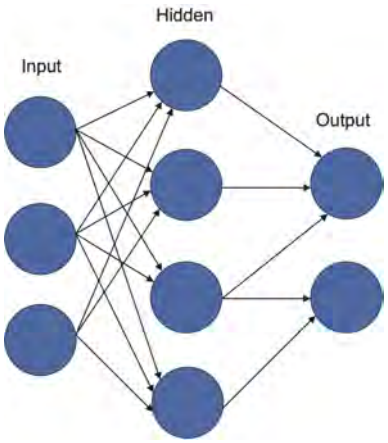


Figure 4.1: A neural network is an interconnected group of nodes.

The process of ANN is shown in Fig.4.2. The learning in an ANN takes place when

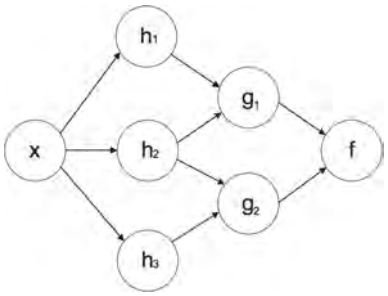


Figure 4.2: The input $\boldsymbol{x}$ is transformed into a 3-dimensional vector $\boldsymbol{h}$, which is then transformed into a 2-dimensional vector $\boldsymbol{g}$, which is finally transformed into f.

it is by given a specific task to solve, and a class of functions F, using a set of observations to find $f^* \in F$ which solves the task in some optimal sense.

Cost function $C : F \to \mathbb{R}$, is an important concept in learning, as it is a measurement of how far away a particular solution is from an optimal solution to the problem to be solved. Learning algorithms search through the solution space to find a function that has the smallest possible cost. For the optimal solution f^* , $C(f^*) \leq C(f) \forall f \in F$. No solution has a cost less than the cost of the optimal solution.

For applications where the solution is dependent on some observed data, the cost must necessarily be a function of the observations, otherwise we would not model anything related to the data. It is frequently defined as a statistic approach to which only approximations can be made. As a simple example consider the problem of finding the model f which minimizes $C = E\left[(f(\boldsymbol{x}) - y)^2\right]$, for data pairs $(\boldsymbol{x}, y)$ drawn from some distribution $\boldsymbol{\mathcal{D}}$. In practical situations, we would only have N samples from $\boldsymbol{\mathcal{D}}$ and thus, for the above example, we would only minimize $\hat{\boldsymbol{C}} = \frac{1}{N}\sum_{i=1}^{N}(f(\boldsymbol{x}_i) - y_i)^2$. Thus, the cost is minimized over a sample of the data rather than the entire data set.

Neural network models as a very popular prediction method have been applied to many air quality forecasting works with abilities of collaborating multi-learning models. Gardner and Dorling (1999) propose an approach based on Multi Layer Perceptron (MLP) NN models to predict hourly NO_x and NO_2 concentrations in urban air in London. Abdul-Wahab and Al-Alawi (2002) use a NN model to predict the tropospheric (surface or ground) hourly basis (24 hours) ozone concentrations as a function of meteorological conditions and various air quality parameters (CH_4, $NMHC$, CO, CO_2, NO, NO_2, SO_2 and O_3) in Kuwait. Chelani, Chalapati Rao, Phadke, and Hasan (2002) conduct a SO_2 prediction by NN models and indicates that NN models are able to give better predictions with less residual MSE than those given by MA models.

NN models combine multiple training algorithms to detect complex nonlinear relationships between dependent and independent variables and all possible interactions between predictor variables. However, NN models have multiple solutions associated with local minima and for this reason may not be robust over different samples.

The Support Vector Machine Regression (SVR) is a time series prediction method developed by Vapnik (Vapnik, 1999; Drucker, Burges, Kaufman, Smola, & Vapnik, 1997; Scholkopf, Burges, & Smola, 1999). SVR departs from more traditional time series prediction methodologies in the strict sense where there is no "model" to make the prediction only depend on the data.

Given a set of time series data $x(t)$, where t is a series of N discrete samples. In time series prediction, training data is obtained at $t = 0, 1, 2, ..., N-1$, and a predict is computed at $t \geq N$. Equation (4.15)) and (4.16) give the gives the linear and non-linear regression, respectively.

$$f(x) = (\boldsymbol{w} \cdot \boldsymbol{x}) + b, \tag{4.15}$$

$$f(x) = (\boldsymbol{w} \cdot \phi(\boldsymbol{x})) + b. \tag{4.16}$$

where "$\cdot$" means a dot product (Cortes & Vapnik, 1995). $\phi(\boldsymbol{x})$ is the kernel function which makes SVM working for non-linear regression.

The goal of SVR is to find a near-optimal solution, which includes a criteria for finding an "optimal" set of weights, weights $\boldsymbol{w}$ and threshold b. Firstly, weights is flatted by the euclidean norm ($\|\boldsymbol{w}\|^2$). Secondly, an empirical risk (error) is generated by the estimation process of the value, which is to be minimized. Then, the minimized regularized risk $R_{reg}(f)$ as defined as:

$$R_{reg}(f) = R_{emp}(f) + \frac{\lambda}{2}\|\boldsymbol{w}\|^2, \tag{4.17}$$

where f is a function of $\boldsymbol{x}(t)$ and the capacity control factor λ is a scale factor regard as regularization constant which reduces "over-fitting" of data and minimizes negative effects of generation. The empirical risk is defined as

$$R_{emp}(f) = \frac{1}{N}\sum_{i=0}^{N-1} L(\boldsymbol{x}(i), y(i), f(\boldsymbol{x}(i), \boldsymbol{w})), \tag{4.18}$$

where i is an index to discrete time series t = {0,1,2,...N-1} and $y(i)$ is the training data of predicted value being sought. $L(.)$ is a "lose function" to be defined.

ϵ-insensitive loss function is the most common loss function to solve for the optimal weights and minimize the regularized risk, which is formulated: minimize $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} L(y(i), f(\boldsymbol{x}(i), \boldsymbol{w}))$, where

$$L(y(i), f(\boldsymbol{x}(i), \boldsymbol{w})) = \begin{cases} 0 & \text{if } |y(i) - f(\boldsymbol{x}(i), \boldsymbol{w})| \leq \epsilon \\ |y(i) - f(\boldsymbol{x}(i), \boldsymbol{w})| - \epsilon & \text{otherwise.} \end{cases} \tag{4.19}$$

C is a positive constant which includes the $(1/N)$ summation normalization factor and ϵ refers to the precision by which the function is to be approximated. They are both user defined constants and are typically computed empirically.

Dual optimization problems can be formed as Equation (4.20) by using Lagrange multipliers

$$\begin{array}{ll} \text{Maximize:} & -\frac{1}{2}\sum_{i,j=1}^{N}(\boldsymbol{a}_i - \boldsymbol{a}_i^*)(\boldsymbol{a}_j - \boldsymbol{a}_j^*)\left\langle \boldsymbol{x}(i), \boldsymbol{x}(j)\right\rangle \\ & -\epsilon\sum_{i=1}^{N}(\boldsymbol{a}_i - \boldsymbol{a}_i^*) + \sum_{i=1}^{N} y(i)(\boldsymbol{a}_i - \boldsymbol{a}_i^*) \quad . \\ \text{Subject to:} & \sum_{i-1}^{N}(\boldsymbol{a}_i - \boldsymbol{a}_i^*) = 0 : \boldsymbol{a}_i, \boldsymbol{a}_i^* \in [0, C] \end{array} \tag{4.20}$$

The solution for the weights is based on the Karush-Kuhn-Tucker conditions, which state at the point of the optimal solution, the product of the variables and constraints equal zero. $f(\boldsymbol{x})$ as the sum of the optimal weights times the dot products between the data points as:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N}(\boldsymbol{a}_i - \boldsymbol{a}_i^*)\left\langle \boldsymbol{x}, \boldsymbol{x}(i)\right\rangle + b, \tag{4.21}$$

Those data points on or outside the ϵ tube with non-zero Lagrange multipliers $\boldsymbol{a}$ are defined as support vectors.

To figure out the non-linear SVR regression, it is necessary to map the input space $\boldsymbol{x}(i)$ into a (possibly) higher dimension feature space $\Phi(\boldsymbol{x}(i))$. The solution of the SVR does not rely on the input data, a kernel function that satisfies Mercer's conditions can be written as:

$$k(\boldsymbol{x}, \boldsymbol{x}^{'}) = \left\langle \Phi(\boldsymbol{x}), \Phi(\boldsymbol{x}^{'}) \right\rangle, \tag{4.22}$$

which can be put back into equation (4.21) and the optimal weights w can be calculated in feature space in exactly the same fashion.

There are some well known kernel functions listed below (Table 4.1).

Table 4.1: SVM Kernels

Polynomial (homogeneous):	$k(\boldsymbol{x}, \hat{\boldsymbol{x}}) = (\boldsymbol{x} \cdot \hat{\boldsymbol{x}})^d$
Polynomial (inhomogeneous):	$k(\boldsymbol{x}, \hat{\boldsymbol{x}}) = (\boldsymbol{x} \cdot \hat{\boldsymbol{x}} + 1)^d$
Radial Basis Function (RBF):	$k(\boldsymbol{x}, \hat{\boldsymbol{x}}) = exp(-\gamma \|\boldsymbol{x} - \hat{\boldsymbol{x}})\|^2)$, for $\gamma > 0$
Gaussian Radial basis function:	$k(\boldsymbol{x}, \hat{\boldsymbol{x}}) = exp(-\frac{\|\boldsymbol{x} - \hat{x})\|^2}{2\sigma^2})$
Sigmoid:	$k(\boldsymbol{x}, \hat{\boldsymbol{x}}) = tanh(\kappa \boldsymbol{x} \cdot \hat{\boldsymbol{x}} + c)$ for some (not every) $\kappa > 0$ and $c < 0$

In literature, SVR has been considered as the best prediction model for wind speed and ambient air pollutants. W. Lu et al. (2002) found that SVR with RBF kernel outperforms neural networks (NN) and hack propagation (BP) when forecasts air quality parameters including sulphur dioxide (SO_2), nitrogen oxides (NO_2), nitric oxide (NO), nitrogen dioxide (NO_I), carbon monoxide (CO) and respirable suspend particles (RSP) in Hong Kong. Sotomayor-Olmedo et al. (2011) predicted airborne contaminants such as PM_{10} and $PM_{2.5}$ in London-Bloomsbury at south England, and compared with neural networks and various fuzzy clustering algorithms, SVR was demonstrated producing the best prediction result. In Mohandes, Halawani, Rehman, and Hussain (2004)'s study, SVR delivered the best prediction over a 12 years (between 1970 and 1982) wind data, as compared to traditional autoregression modeling with moving averages, and artificial neural networks.

Owing to the kernel in SVM can be linear and non-linear, SVM offers the benefit of solving non-linear environmental problems. On the other hand, SVM solution is associated with the problem of lack of operation transparency, and that its accuracy is heavily influenced by quality of training data.

4.3 Performance Measurement

This section presents in-depth evaluation of existing prediction models, where environmental event prediction is measured by prediction accuracy, dynamic capability, spatio-temporal capability.

4.3.1 Accuracy

There are many approaches in measuring the accuracy of environmental prediction algorithms. We now describe some of these measures, which will be used for the performance evaluation of environmental prediction described in Chapter 9.

Mean absolute error (MAE) estimates the average difference between the predicted value and the true value. Compared to other measurement, MAE is unambiguously a more natural measure of average error. To represent the prediction performance in the term of dimensioned evaluations and inter-comparisons of average model-performance error, MAE is the best choice (Willmott & Matsuura, 2005). The MAE is formulated as,

$$MAE = \frac{\sum_{i=1}^{N} abs(\boldsymbol{P}_i - \boldsymbol{O}_i)}{N}, \tag{4.23}$$

where N is number of the observations, $\boldsymbol{P}_i$ and $\boldsymbol{O}_i$ are the i^th observations of predicted value and observation value.

Mean bias error (MBE) represents a measure of overall bias error or systematic error via calculating the average difference between the two datasets. As an average value, the positive and negative differences between observations are eliminated (L. Ji & Gallo, 2006). MBE is defined as,

$$MBE = N^{-1} \sum_{i=1}^{N} (\boldsymbol{P}_i - \boldsymbol{O}_i). \tag{4.24}$$

Root mean square error (RMSE) is a frequently used measure of the differences between predicted values from a model or an estimator and the observed values. The difference is called residual. RMSE measures average deviation, similar to standard deviation, but it is concerned with deviations from the observation value whereas S is concerned with deviations from the mean (Schmidt & Wulf, 1997). RMSE is defined as,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\boldsymbol{P}_i - \boldsymbol{O}_i)^2}{N}}. \tag{4.25}$$

Note that RMSE is a non-linear function biased towards large errors when large differences occur. To illuminate the sources or types of errors, (Kolehmainen, Martikainen, & Ruuskanen, 2001) introduced the systematic RMSE (RMSE_S) and unsystematic RMSE (RMSE_U),

$$RMSE_S = \sqrt{\frac{\sum_{i=1}^{N}(\hat{\boldsymbol{P}}_i - \boldsymbol{O}_i)^2}{N}} \tag{4.26}$$

$$RMSE_U = \sqrt{\frac{\sum_{i=1}^{N}(\boldsymbol{P}_i - \hat{\boldsymbol{P}}_i)^2}{N}}, \tag{4.27}$$

where $\hat{\boldsymbol{P}}_i$ is the solution of a least-squares linear regression of $\boldsymbol{P}$ on $\boldsymbol{O}$. RMSE_S is related to an inability of the model to capture the general trends in the observations, which hypothetically could be improved; and RMSE_U is used to represent random errors which could be caused by small scale processes, which are not expected to be resolved by the prediction model. In practice, RMSE_S is mostly composed of zero while RMSE_U is substantially close to RMSE, the model is probably as good as it can be, and vice versa.

Index of agreement (IOA) measures the ability of the model to capture the variability and the mean of the observations (Willmott, 1982). IOA is ranged from 0 (no agreement) to 1 (prefect agreement). This evaluation is best applied to compare two sets of model results. For example, if $IOA_1 = 0.8$ and $IOA_2 = 0.85$, it means that model 2 is 5% outperform than model 1. The IOA is defined by:

$$\begin{aligned} IOA &= 1 - \frac{\sum_{i=1}^{N}(\hat{\boldsymbol{P}}_i - \boldsymbol{O}_i)^2}{\sum_{i=1}^{N}(\boldsymbol{P'}_i + \boldsymbol{O'}_i)^2} \\ \boldsymbol{P'}_i &= abs(\boldsymbol{P}_i - \bar{\boldsymbol{O}}) \\ \boldsymbol{O'}_i &= abs(\boldsymbol{O}_i - \bar{\boldsymbol{O}}), \end{aligned} \tag{4.28}$$

where $\bar{\boldsymbol{O}}$ corresponds to the mean observation.

4.3.2 Dynamic Prediction

Environmental conditions are varying continuously over time with many unexpected environmental phenomena, it is expected that predictions of environmental events should be uncertain. Environmental prediction models is pressed to consider this uncertainty in the magnitude of an event for a given return period. Thus, dynamic prediction is required for continuous environmental change estimation and prediction model reconstruction.

Dynamic prediction is often used on data with continuous time variables. In order to better understand behavior of prices in financial markets, Kohara, Ishikawa,

Fukuhara, and Nakamura (1997) gained a smaller prediction error from the continuous updated event-knowledge and online neural networks than other multiple regression analysis without real-time learning. Dyrløv Bendtsen, Nielsen, von Heijne, and Brunak (2004) used neural network and hidden Markov model algorithms to improve prediction of signal peptides, where both components have been updated at the latest cleavage site position. Castro-Neto, Jeong, Jeong, and Han (2009) applied online SVR to trace the incident and atypical conditions in traffic flow forecasting, which significantly improved prediction accuracy of short-term traffic flow under atypical conditions.

Consider online updating prediction models, two general approaches has been discussed in the literature. One approach is to update the training data on which the prediction model is built. (Kohara et al., 1997; Dyrløv Bendtsen et al., 2004) extracted current knowledge from the whole set of captured data. They simply re-trained the prediction models, whenever the latest data becomes available. Apparently, this is cost expensive both in time and computation, moreover, this often fails to adequately perform real-time prediction. The other approach is computationally update the prediction model by learning the latest data only. The benefit is fast updating, as only new knowledge is stored into an incremental learning model (Castro-Neto et al., 2009; Van Lint, 2008). However, it is hard to guarantee the quality of stored knowledge (in the model itself), which is predicated on the quality and reliability of incoming data.

4.3.3 Spatio-temporal Prediction

When data is collected or analyzed across time as well as space, we often consider it as a spatio-temporal problem. However, according to Hernández-Campos, Karaliopoulos, Papadopouli, and Shen (2006), until recently, there has not been a theory of spatio-temporal processes separate from the already well established theories of spatial statistics and time series analysis.

It is widely recognized that the construction of spatio-temporal models would be a conceivable approach. Cressie and Huang (1999) proposed a generic approach to develop parametric models for spatio-temporal processes. These models heavily rely on spectral representation for the theoretical space-time covariance structure, but which only generalize results for pure spatial processes (Matern, 1986). Hence, they have disregarded another important factor, which is time. Moreover, Ilin and Luttinen (2010) presented a spatio-temporal model based on variational Gaussian-process factor analysis. Their research also has such a limitation of neglecting the posterior temporal correlations to factorize the approximation of a completed spatio-temporal model.

Existing prediction models perform inadequately when tasked with spatio-temporal environmental event prediction. Firstly, accurate environmental forecast is often difficult due to the lack of prior knowledge of the geometric topology of the monitoring equipment in relation to other stations (Vlahogianni, Karlaftis, & Golias, 2007; T. Cheng & Wang, 2008). Secondly, sharing spatial knowledge derived from among a collection of distributed data capture systems, facilitates the creation of an inefficient machine learning model (Ryoo & Aggarwal, 2009; Zinzen, Girardot, Gagneur, Braun, & Furlong, 2009). This is because that environmental data are sometimes short and may suffer from the missing data problem. Moreover, problems associated with asynchronous data capture are inherent in distributed environmental data collection.

4.4 Applications to Air Quality Control

Environmental prediction recently have been applied successfully in many air pollution control systems.

Ozone protects harmful ultraviolet (UV) rays for the Earth. It is very difficult for anything to survive on the surface without the layer of ozone in the atmosphere. Salazar-Ruiz, Ordieres, Vergara, and Capuz-Rizo (2008) and Coman, Ionescu, and Candau (2008) proposed to use prediction models on ozone concentrations to supporting decisions in the matter of ozone pollution in Mexicali and France. Another approach by Sousa, Martins, Alvim-Ferraz, and Pereira (2007) proposed to use ANN to predict the tropospheric ozone concentrations for evaluating the relative influence of precursor concentrations and meteorological variables on ozone formation from anthropogenic emissions (from industry and traffic). Another ozone concentration application by W.-Z. Lu and Wang (2008) proposed to use SVM model to predict the short-term ground-level ozone (O_3) for reducing related injuries caused by elevated level of tropospheric O_3. Ozone evaluation is only one imporant aspect of environmental protection, another problem, quite often associated with ozone depletion is from industrial and other emissoion sources.

Predicting concentrations of exhaust emissions is to assist decision making of environmental protection, one method proposed by Alonso et al. (2007) used artificial neural network (ANN)-based prediction models to predict NO_x, PM, CO, HC released from diesel engines for evaluating the negative environmental impact related to their emissions for air quality control in European. Another application described by Karacan (2008) used ANN models to predict the ventilation methane emission from U.S. longwall mines for protecting the safety of underground coal miners. Another application in Athens, Greece was to control urban air quality as-

sisted by ANN models to predict CO, NO_2, SO_2, and O_3 (Moustris, Ziomas, & Paliatsos, 2010). More interestingly, from the perspective of this research is the impact on personal health.

Health studies have shown premature death is correlated with exposure to particulate matter $< 2.5\ \mu m$ in diameter ($PM_{2.5}$). Baker and Scheff (2007) proposed to use prediction models to forecast concentration of $PM_{2.5}$ in the central and eastern United States for supporting emission control plans that likely focus on reducing emissions of SO_2 and NO_2. An investigation of traffic-related air pollutants for health impact warnings was carried out by Hochadel et al. (2006), who proposed a geographical information systems (GIS) based regression models to predict long-term average concentrations of PM_x from domestic heating and industries in North-Rhine Westphalia, Germany. A correlation based approach to predict $PM_{2.5}$ in Santiago for designing an air quality attainment plan, was documented by Saide et al. (2011). Outdoor air pollution was an interesting and important topic for this research, however, research was conducted on pollutants in the home.

Some studies of indoor air quality are motivated by a desire to understand the origins of risks to the health of householders, because people spent most of time indoors. Spengler and Chen (2000) applied indoor air quality prediction system to design a healthier building. After that, Xing, Hatton, and Awbi (2001) studied different indoor arrangements with air quality prediction based on knowledge from outdoor air flow model for testing ventilation performance. An extensive review was undertaken by B. Yu et al. (2009) into the research in the area of air-conditioning systems and indoor air quality control. The purpose of that review was to evaluate different air conditioning systems based on pollutant simulation (i.e. negative air ions, non-thermal plasma, number of mass transfer units and particulate matter).

4.5 Conclusion

Pollution prediction estimates the future trend of pollution, which is helpful for environmental events monitoring. This chapter discusses existing computational methods for environmental event prediction, summarizes actual requirements of environmental prediction models, and identifies future challenges of research for computational environmental prediction methods.

Existing environmental prediction models include statistical models and artificial intelligence models. Statistical models are conducted on historical data only without considering any other variables. These methods address merely one single specific pollutant (e.g., NO_x). Artificial intelligence models extract knowledge from massive amounts of multi-dimensional data. These methods are capable of

analyzing the pollution types, amount, and even pollutants interactions.

Existing prediction models perform inadequately when tasked with real-time and spatio-temporal environmental analysis. An inadequacy is the failure of the predication model when dealing with unqualified and unreliable streaming input data. To overcome this problem retraining procedures are used on the whole dataset, which are computationally expensive in terms of time and resources. Another problem, is that the spatio-temporal models suffers from the spatially sparse environmental data, that is the geometric topology of the monitoring equipment in relation to other stations is often independent and geometrically non-uniform. In addition, the asynchronous data capture, short collection period and number of missing values already have caused the difficulty of sharing knowledge in global prediction models.

Chapter 5

Feature Extraction for Air Quality and Land Use Analysis

An important component of computational environmental study concerns the analysis of data recorded over time and space. Yet environmental data is often multidimensional, noisy, and with huge size. The quality of data is not guaranteed. Redundancy and missing data, both may involve in constructing of real environmental dataset due to equipment failures, system maintenance, and operation errors. Feature extraction thus is required for reducing the dimensionality without decreasing recognition accuracy and/or discovering a subset of features that increase the recognition performance beyond that of the original data set. Upto now, many feature extraction researches have been conducted on diverse environmental problems. This chapter discusses the environmental data feature extraction with a focus on the environmental problems addressed by the thesis.

5.1 Introduction

Environmental data is often a collection of measurements across multiple variables, exhibiting relationships in many dimensions. Air quality data include the following well known chemical compounds: nitrogen (N_2), oxygen (O_2), argon (Ar), carbon dioxide (CO_2), neon (Ne), helium (He), methane (CH_4), krypton (Kr), hydrogen (H_2), nitrous oxide (N_O), xenon (Xe), ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide NO_2, ammonia (NH_3) and fine particulate matter (PM_{10} and $PM_{2.5}$) in the atmosphere. Water quality data is constructed by the levels of pH, biochemical oxygen demand (BOD), chemical oxygen demand (COD), dissolved oxygen (DO), total hardness (TH), heavy metals, nitrate, orthophosphates, pesticides and surfactants in the observed water samples. Soil quality data consists of the levels of phosphorus, nitrogen, potassium, sulfur, calcium, magnesium, hydrogen*, carbon*, oxygen*, copper, molybdenum, chlorine, zinc, boron, cobalt and manganese in the soil samples. These variables can provide multidimensional environmental data for computational analysis.

Often sensor data are noisy because environmental measurements are highly volatile, which results in the capture of highly noisy data. In practice, the quality of environmental data is influenced by many natural phenomena such as ambient lighting and wind turbulence, equipment failure, insufficient sampling or measurement errors. This causes many unwanted spurious values appear in the captured environmental data. Other phenomena interfere with the measuring equipment itself, once again resulting in the introduction of noise errors into the measured data. Moreover, the capabilities and performance characteristics of the measuring equipment itself may introduce additional noise. For example, speckle noise (Celik, 2009) can be produced as a consequence of the inherent characteristics of the measuring equipment itself.

Computational analysis of environmental data aims to determine pathological values of parameters that produce dangerous living conditions. Three major challenges of computational analysis is that there are only a limited number of variables associated with the environmental problem, however these data for a specific problem, often results in a big collection of dependent and independent variables. This is further complicated as the analysis model has to attempt to process more information than is necessary (information overload (Fodor, 2002)) for making an appropriate decision. Finally, the third challenge is the inclusion of irrelevant or redundant variables that could distort the resulting investigation of the environmental problem (McCune, 1997).

In response to the aforementioned challenges, it is required for computational environmental analysis to extract, from this data, only the parameters useful for the subsequent decision process (Acciani, Chiarantoni, Fornarelli, & Vergura, 2003). Feature extraction of environmental data simplifies the problem of multidimensional search and analysis, which is the key of fast problem recognition by extracting robust information and reducing redundant samples (S. Zhang et al., 2007). Also, through feature extraction, the input data are transformed efficiently into a reduced representation set of features, which improves the speed of most environment analytical models without losing authentication accuracy. Moreover, feature extraction enables the discovery of a subset of features and interrelationships between environmental variables. After carefully choose features from the input data, strong information is possible to be extracted and used for pattern recognition for environmental analysis.

We restrict our focus to air quality and land use problems specifically. In our research, we have found many common types of feature extraction methods to magnify the main contributions and interrelationship constraints that exist across domains in environmental data; we have studied the term frequency feature selection in terms of verifying the valid/efficient samples to filter out noise from input data;

furthermore, feature extraction can be used to magnify the texture information for edge detection when dealing with land use problems.

This chapter is organized as follows: Section 5.2 introduces the common feature extraction methods which magnify the contributions of air parameters for emission source detection. Section 5.3 presents the brief concepts of existing feature extraction methods which magnifies the texture information for edge detection for land use observation. Section 5.4 describes the validation of feature extraction methods each of which selects important time instance(s) for environmental analysis. Finally, a summary is provided in Section 5.5.

5.2 Air Quality Authentication

Principal component analysis (PCA) and Linear discriminant analysis (LDA) are two common feature extraction methods for air quality authentication.

5.2.1 Principal Component Analysis (PCA)

PCA as a very popular feature extraction method for environmental data analysis, is used to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables by an orthogonal transformation (Jolliffe, 1986). Given a high-dimensional environmental data $\boldsymbol{X}^{n\times d}$, PCA seeks a linear transformation $\boldsymbol{W}$ that maps the original d-dimensional space onto a l-dimensional feature subspace where $l < d$. The reduced feature vectors are defined as $\boldsymbol{x}_i' = \boldsymbol{W}^T\boldsymbol{x}_i$, $i = 1, \ldots, n$, $\boldsymbol{x}_i' \in \boldsymbol{X}^{n\times l}$. Here, each column of $\boldsymbol{W}$ represents an eigenvector $\boldsymbol{v}$, which can be obtained by solving the following eigen-decomposition,

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \tag{5.1}$$

where $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{X}^T$ represents the covariance matrix and λ_i represents the eigenvalue associated with the eigenvector $\boldsymbol{v}_i$.

One strength of PCA is its ability to extract features from raw data without prior knowledge, this means that PCA overcomes some of the problems associated with information overload when performing air quality analysis. One such application, described by Henry and Hidy (1979), applied PCA to air quality data in order to analyze particulate sulfate, meteorological and air quality data for the areas of Los Angeles and New York City. Another application of this technique was deployed in the City of London, Derwent et al. (1995), which used PCA to extract information from airborne chemical components for the purpose of identifying of emission sources. Other researchers, Harkat, Mourot, and Ragot (2006), have found that the

PCA model can be used to improve the performance of automatic detection of NO_2, SO_2 and VOCs when applied to air quality monitoring networks.

5.2.2 Linear Discriminant Analysis (LDA)

LDA is used to convert a set of observations into a linear combination of features which characterizes or separates two or more classes of objects or events by an orthogonal transformation (Fisher, 1936). Given a high-dimensional environmental data $\boldsymbol{X}^{n\times d}$, LDA seeks an orientation $\boldsymbol{W}$ with size $d \times l$, and $\boldsymbol{W}$ has l column vectors known as the eigenvectors.

To define the LDA explicitly, suppose that $\boldsymbol{X}$ has c classes, $\boldsymbol{C} = \{\boldsymbol{c}_i | i = 1, 2, \ldots, c\}$ is the finite set of c class labels, c_i denotes the ith class label. As a supervised learning method, $\boldsymbol{W}$ is calculated by r training samples $\boldsymbol{Z}$ ($r < n$), $\boldsymbol{Z} = \{\boldsymbol{z}_j | j = 1, \ldots, r\}$. The training set can be subdivided into c subsets $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, …, $\boldsymbol{Z}_c$, where each subtraining set belongs to c_ith class and consists of r_i instances such that $r = \sum_{i=1}^{c} r_i$.

The between-class scatter matrix $\boldsymbol{S}_B$ is calculated with centroid $\boldsymbol{\mu}$ of $\boldsymbol{Z}$, as given in (5.2)

$$\boldsymbol{S}_B = \sum_{i=1}^{c} n_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \tag{5.2}$$

where $\boldsymbol{\mu}_i$ is the centroid of $\boldsymbol{Z}_i$.

The within-class scatter matrix $\boldsymbol{S}_W$ is the sum of c scatter matrices, is calculated as

$$\boldsymbol{S}_W = \sum_{i=1}^{c} \sum_{j=1}^{r} (\boldsymbol{z}_{ij} - \boldsymbol{\mu}_i)(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_i)^T \tag{5.3}$$

Then, according to (Duda, Hart, et al., 1973), Fisher's criterion as a function of $J(\boldsymbol{W})$ is given as

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T \boldsymbol{S}_B \boldsymbol{W}}{\boldsymbol{W}^T \boldsymbol{S}_W \boldsymbol{W}}. \tag{5.4}$$

Fisher LDA attempts to seek the linear transformation matrix $\boldsymbol{W}$ to maximize the above criterion. $\boldsymbol{W}$ can be obtained by solving the eigendecomposition as,

$$\boldsymbol{S}_W^{-1} \boldsymbol{S}_B \boldsymbol{w} = \lambda \boldsymbol{w}, \tag{5.5}$$

where $\boldsymbol{w}$ represents eigenvector corresponding to each column of the transformation matrix $\boldsymbol{W}$.

LDA in air quality analysis is used mainly for classification. Atkinson et al. (1999) applied LDA to analyse the associations between outdoor air pollution and visits to accident and emergency departments in London for respiratory complaints. Vardoulakis, Fisher, Pericleous, and Gonzalez-Flesca (2003), demonstrated that LDA

can take advantage of the pollutant concentration measures extracted from traffic emissions in urban areas. Masiol, Squizzato, Ceccato, Rampazzo, and Pavoni (2012) validated the significant advantage of using LDA to test the differences in the PM_{10} levels and chemical profiles with varying atmospheric circulation and long-distance pollutant transportation. In practice, LDA feature shows more favorable to classification than that of PCA in that LDA is trained on data with truth class labels.

5.3 Land Use Image Understanding

For land use observation, feature extraction is often conducted on a satellite image to detect changes and enhance understanding. Given a landscape image $I^{M\times N}$, extraction is used to magnify the texture information by detecting edges on image pixels in different angles. The simplest edge detection model is Gaussian blurring (Lindeberg, 1998; W. Zhang & Bergholm, 1997), which is defined as,

$$I(x) = \frac{I_r - I_l}{2}\left(\operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma}\right) + 1\right) + I_l, \tag{5.6}$$

where $x \in M \times N$ is the pixel value, $I_l = \lim_{x\to-\infty} I(x)$ is the intensity at the left side of the edge, $I_r = \lim_{x\to\infty} f(x)$ is for the right, and σ is scale parameter or the blur scale of the edge.

5.3.1 Gradient Feature

The most common image edge operator is an approximation of the intensity gradient within the image. In this approach, there are two orthogonal the gradient vectors $\boldsymbol{g}_x$ and $\boldsymbol{g}_y$, which are combined to compute the gradient magnitude $\boldsymbol{g}$ as,

$$\boldsymbol{g} = \sqrt{\boldsymbol{g}_x^2 + \boldsymbol{g}_y^2}. \tag{5.7}$$

In practice, many gradient-approximating kernels as in Table.5.1 have been developed to increase efficiency of gradient estimation.

In environmental research, gradient edge operators are often used for the detection of urbanisation, forest fragmentation and rainfall problems. For example, Hoffhine Wilson, Hurd, Civco, Prisloe, and Arnold (2003) applied gradient-based edge detection for discovering urban sprawl in New Jersey between 1989 and 1999. An application to uncover forest fragmentation problems in the USA, Fuller (2001) utilized gradient features to enhance texture information. Another application, once again in the USA, Shepherd, Pierce, and Negri (2002) proposed the use of the gradient feature for the purpose of calculating accurate area of rainfall in the Atlanta

Table 5.1: Kernels of Well-known Image Edge Operators

Name	$\boldsymbol{g}_x$	$\boldsymbol{g}_y$
Frei-Chen (Freeman, 1961)	$\begin{bmatrix} -1 & 0 & 1 \\ -\sqrt{2} & 0 & \sqrt{2} \\ -1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & \sqrt{2} & 1 \\ 0 & 0 & 0 \\ -1 & -\sqrt{2} & -1 \end{bmatrix}$
Roberts (Roberts, 1963)	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Prewitt (Prewitt, 1970)	$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$
Sobel (Gonzalez, 1987)	$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$
Laplacian (Gonzalez, 1987)	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	

area. However, gradient edge operators are sensitive to noise when detecting edges and their orientations. For coastline and desertification monitoring, gradient edge operators often suffer from speckle noise introduced in satellite image capture. Often a small error introduced in this way can degrade the performance of the edge detection method. Moreover, it can degrade the accuracy of image segmentation, if the noise appears in a sinusoidal path along the boundary between two regions.

5.3.2 Wavelet Feature

The wavelet transform can be interpreted as a multi-scale edge detector that represents the singularity content of an image at multiple scales and different orientations, thus, the noise can be removed more efficiently than using gradient feature. The basic idea of wavelet based edge detection is to calculate the Fourier transform of a function in the "sliding window" (Gabor, 1946), which refers to a small area within the image. The Gabor filter is the most popular wavelet feature for edge detection can be formulated as,

$$\boldsymbol{G}(x, y; \lambda, \theta, \psi, \sigma) = \exp\left(-\frac{x' + y'^2}{2\sigma^2}\right) \exp\left(x\left(2\pi\frac{x'}{\lambda} + \psi\right)\right), \tag{5.8}$$

where

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \tag{5.9}$$

and λ represents the sinusoid's wavelength, θ represents the orientation, ψ is the phase offset, and σ denotes the spread of the Gaussian window (Daugman, 1985).

Given a 2-D image I with a size of $M \times N$, we set Gabor filter with the set of orientations θ^t, where $t = 1, 2, ..., T$. According to (5.8) and (5.9), we obtain T different images,

$$\boldsymbol{G}^I_{\theta^t} = \left\{ g^I_{\theta^t}(m,n) | 1 \leq m \leq M, 1 \leq n \leq N, \text{and } t = [1, 2, ..., T] \right\}, \tag{5.10}$$

where $\boldsymbol{G}^I_{\theta^t}$ is a set of real-valued 2-D subbands signals, representing the magnitude of complex-valued wavelet coefficients obtained at the direction θ^t.

Wavelet has been used for satellite imagery analysis in field of area surveillance, sea level and urban sprawl monitoring. Y. Ji, Chang, and Hung (2004) applied wavelet to satellite image segmentation for the area of Adana, Turkey. Niedermeier, Romaneessen, and Lehner (2000) used the Gabor filter based edge detection to monitor coastlines in the area of Elbe estuary (located in the intertidal zone of the German Bight). Corbane, Faure, Baghdadi, Villeneuve, and Petit (2008) monitored the urban sprawling of Greater Toronto Area (GTA) Canada by employing wavelet feature for distinguishing different image objects. In principle, wavelet maps a high-resolution image to different scales of sub-band, which enhances the texture information of satellite images.

5.4 Data Selection for Emission Source Categorization

Data from environmental monitoring is often considered as a time series, $\boldsymbol{X^{T \times d}} = \{\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(t), \ldots, \boldsymbol{x}(T)\}$ where T is elapsed time and $\boldsymbol{x}(t)$ represents an data sample consisting of d chemical contaminants. A time series could be very long, sometimes containing millions of observations. For feature extraction, we apply a sliding window w to $\boldsymbol{X}$ that will produce a sequence of sub time series, and we store them as $\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2 \ldots, \boldsymbol{c}_n\}$, where $\boldsymbol{c_i} = \{\boldsymbol{x}[(t) : (t + w - 1)]\}$ represents a subsequence, and $n = T - w + 1$ is the number of subsequences.

5.4.1 Term Frequency Sample Validation

Unlike LDA and PCA, the term frequency based approach processes data samples as signals in different sliding windows. Term frequency measures how often a number is found in a collection time, and it is often applied to sample validation for noise filtering.

A term frequency based feature extraction selects useful time points within $\boldsymbol{C}$. For example, fast Fourier transform (FFT) is a simple and well-known frequency-

based feature selection method, that has been widely used in environmental analysis, and that is defined as,

$$\begin{aligned} X'_k = & \sum_{n=0}^{N-1} a_n e^{-2\pi i n k/N} \\ = & \sum_{n=0}^{N/2-1} a_{2n} e^{-2\pi i (2n) k/N} + \sum_{n=0}^{N/2-1} a_{2n+1} e^{-2\pi i (2n+1) k/N} \\ = & \sum_{n=0}^{N/2-1} a_n^{even} e^{-2\pi i n k/(N/2)} + e^{-2\pi i k/N} \sum_{n=0}^{N/2-1} a_n^{odd} e^{-2\pi i n k/(N/2)}, \end{aligned} \tag{5.11}$$

where N represents size of sliding window, $k = 0, 1, \ldots, N$, $e^{-2\pi i/N}$ is an N-th primitive root of unity, and thus can be applied to analogous transforms over any finite field. In many cases, N as the size of $\boldsymbol{X}'$ in the first non-singleton dimension can be shortened or enlarged with more understanding and experience on environmental data monitoring. To reduce the cross track bias that appears in environmental data, N as a wave number can be shortened for the smoothing purpose (Fishman, Wozniak, & Creilson, 2003).

Term frequency techniques validate data samples in a sub-sequence, which is useful for filtering noise of environmental data. T. B. Hansen and Johansen (2000) applied term frequency technique to filter out noise from radar pictures for improving the performance of air-soil pollution monitoring. Menzel and Goschnick (2000) used term frequency techniques to smooth signal of emission gas and improved the performance of detecting non-stationary gas mixtures. Calder (2007) proposed a term frequency based dynamic factor process convolution model for multivariate spatio-temporal processes and illustrated the utility of this approach in modeling large air quality monitoring data. In practice, term frequency techniques improve the performance of air quality analysis by concentrating noise in the orthogonal low frequency terms.

5.4.2 Pearson's Correlation based Variable Selection

In time series analysis, correlation is a popular feature selection method. Given extracted subsequences data $\boldsymbol{C}$, for each subsequence $\boldsymbol{c}_i$, the Pearson's correlation (ρ) between the first attribute and second attribute can be calculated as:

$$\rho_{\boldsymbol{X},\boldsymbol{Y}} = \frac{\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y})}{\sigma_{\boldsymbol{X}}\sigma_{\boldsymbol{Y}}} = \frac{E[(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}})]}{\sigma_{\boldsymbol{X}}\sigma_{\boldsymbol{Y}}}, \tag{5.12}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two time series variables, present the first attribute and second attribute of $\boldsymbol{c}_i$, *cov* is the covariance, $\sigma_{\boldsymbol{X}}$ and $\sigma_{\boldsymbol{Y}}$ are standard deviations, $\boldsymbol{\mu}_{\boldsymbol{X}}$ and $\boldsymbol{\mu}_{\boldsymbol{Y}}$ are the expected value, and E is the expected value operator. $\rho_{\boldsymbol{X},\boldsymbol{Y}}$ is ranged from +1 to -1. A positive correlation ($\rho_{\boldsymbol{X},\boldsymbol{Y}} \to 1$) means that, as one variable/time series $\boldsymbol{X}$ becomes large, the other $\boldsymbol{Y}$ also becomes large, and vice versa. $\rho_{\boldsymbol{X},\boldsymbol{Y}} \to +1$ means a perfect positive linear relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$. In case of negative

correlation ($\rho_{\boldsymbol{X},\boldsymbol{Y}} \rightarrow$ -1), as one variable ($\boldsymbol{X}$) increases the other ($\boldsymbol{Y}$) decreases, and vice versa.

Pearson's correlation coefficient condenses the comparison of two variables down to a single scalar, which indicates the strength of the relationship between the two variables. In the practice of air quality analysis, Katsouyanni et al. (1996) had selected NO_2 and SO_2 to trace O_3 based on the Pearson's correlation values within 24 hours time windows for the APHEA project (Air Pollution and Health: A European Approach). Through Pearson's correlation analysis, Statheropoulos, Vassiliadis, and Pappa (1998) validated that humidity, temperature, sunshine duration, wind velocity and wind direction are the most relative meteorological factors to analyze pollution from CO, NO, NO_2, O_3, smoke and SO_2 in the city of Athens by the Pearson's correlation values. Also, Schwartz (1999) calculated the Pearson's correlation coefficients to determine the most harmful pollutant when trace air pollution and hospital admissions for heart disease in eight U.S. counties. Unlike the above feature extraction methods which do dimension reduction or data transformation, Pearson's correlation analysis discovers the relationship between environmental problems and their main sources by calculating one assessable value.

5.5 Conclusion

We have provided a description of various feature extraction methods for addressing environmental problems. For air quality authentication, through dimension reduction air quality data are transformed into a reduced representation feature-set. This improves the speed of most analytical models without losing authentication accuracy. For land use observation, features such as gradient and wavelet feature are often used to enhance understanding of satellite images by representing the singularity content of image in multiple scales and different orientations. Feature extraction is also used for emission source categorization in the approach of selecting valid samples. Methods like, term frequency based sample validation and Pearson's correlation based variable selection, measure how often a number is found in a collection time, and they are often applied to sample validation for noise filtering.

The evaluation of above feature extraction methods in terms of computational environmental analysis allow us to discover more important knowledge from environmental data in the following chapters.

Chapter 6

One-step-more Incremental Learning for Image Series Change Detection

This chapter proposes a novel incremental learning based image change detection method capable of detecting sequences of changes over image series. Given two images for change detection, an intelligent agent is trained by incremental learning on the source image. When detecting change, the agent conducts one-step-more incremental learning on the target image to detect the difference against what has been just learned from the source image. For continuously detecting changes against the third image, the agent can upgrade its knowledge on the target image by performing incremental learning on top of its current knowledge. For performance evaluation, we perform extensive change detection experiments on both static images and image series. The results show that the proposed approach not only provides consistently accurate image detection, but also demonstrates substantial memory efficiency improvements when compared to existing methods.

6.1 Background

Image change detection through machine learning methods is gaining increasing interests in the research community (R. J. Radke, Andra, Al-Kofahi, & Roysam, 2005; Aach et al., 1993; Bazi et al., 2005a; Celik & Ma, 2010; Coppin et al., 2004). Existing image change detection methods focus on seeking differences between a pair of images (Im et al., 2008a; Ma et al., 2010; Verbesselt et al., 2010). Simple differencing methods use an uncomplicated algorithm to detect changes at pixel level or block level between two images (R. Jain, 1984; Yakimovsky, 1976). It is simple and fast, whereas impulse noise often reduces efficiency of simple differencing method, because automatically adjusting detection threshold is always a difficult task. Therefore unsupervised learning change detection and supervised learning change detection have been developed to deal with this problem (R. J. Radke et al., 2005). To identify the differences between a pair of images, a probabilistic model is often utilized in unsupervised learning methods to determine the probability of changes for every pair of corresponding pixels in the two images (Aach et al., 1993; Bazi et

al., 2005a; Celik & Ma, 2010). On the other hand, supervised learning techniques typically rely on classification models to distinguish different elements (e.g. objects or pixels) in an incoming image. For example, in (Coppin et al., 2004), objects in an image can be identified through a classification process based on an object model trained by manually selected training images. In other research on supervised learning methods (Dengkui et al., 2008), changes are detected by classifying pixel pairs into changed and unchanged classes.

When applying change detection techniques to a series of images taken at successive time intervals, the objective is often to seek differences among those images based upon the index of similarity or dissimilarity (Salmon et al., 2011). To achieve this goal, a majority of existing unsupervised learning methods require the probabilistic model to be completely rebuilt each time (Bazi et al., 2005a; Celik & Ma, 2010). Two images, namely the *source image* and the *target image* need to be stored in memory in order to build the probabilistic model. This is a computationally costly approach. Similarly, for most supervised learning methods, it is necessary, but computationally expensive, to recollect training samples and to rebuild the classification model, especially when new image elements (i.e. objects or pixels) appear in incoming images (Coppin et al., 2004; Dengkui et al., 2008).

In order to process a huge amount of data and effectively reduce the data storage requirement for image series change detection, in this chapter, incremental learning technology will be explored to maintain and update the classification model rather than rebuilding it. Incremental learning updates its model over new information from additional data that becomes available after a model has already been created based on a previously available dataset (Pang, Ozawa, & Kasabov, 2005). In the literature, incremental learning has been approved as an efficient machine learning method for processing large data streams, as reported in various real world applications (Fei-Fei, Fergus, & Perona, 2007; Opelt, Pinz, & Zisserman, 2006; K. Wang & Stolfo, 2004; Maloof & Michalski, 2004; Chen, Pang, Kasabov, Ban, & Kadobayashi, 2009; Polikar, Upda, Upda, & Honavar, 2001; Boiral, 2002). In addition, the memory requirement for storing a classification model is much less than storing two large-sized images.

The proposed image series detection is derived from a brain-like pattern recognition mechanism. As we know, a person learns a pattern incrementally. For example it takes some time to get familiar with a campus structure but, later, he/she will easily detect the difference if a new building appears on the campus. For image series change detection we implemented, in this work, a novel brain-like one-step more incremental learning system, whose principle is illustrated in Figure 6.1. An intelligent agent discovers the knowledge of image X_1 via a great number of steps of incremental learning. When image X_2 comes up, one-step-more incremental learn-

ing on X_2 is conducted by the agent to detect any change against what is learned. The agent is able to easily detect change against the second image, because after extensive multi-rounds incremental learning, the agent is familiar with X_1. For further detecting changes of X_2 to X_3, owing to incremental learning, the agent does not need to learn X_2 from scratch, but grow X_2 knowledge from its current memory on X_1 (we assume, without loss of generality, image series $\{X_i\}_{i=1}^{t}$ is about the same background, and image changes are mainly from foreground objects variation.) In summary, the proposed method has three notable characteristics: 1) able to continuously update itself for sequential detection tasks without re-training the model from scratch, 2) capable of detecting any unknown objects in the target image, 3) has high memory efficiency as it conducts one-pass incremental learning, which only stores the current knowledge base in memory.

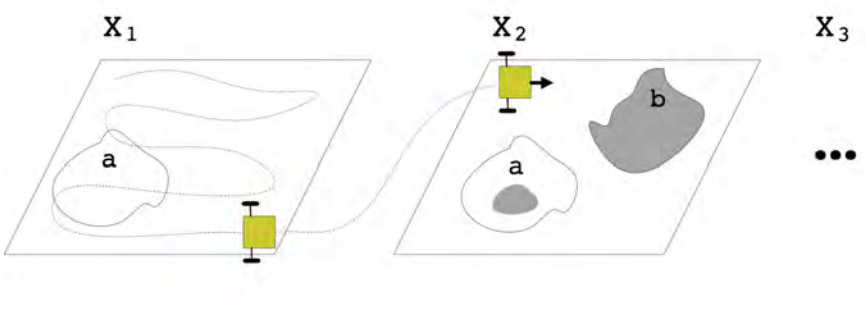


Figure 6.1: Illustration of one-step-more incremental learning for image series change detection.

6.2 Related Works and Motivations

Image change detection methods are generally categorized as simple differencing (R. Jain, 1984; Yakimovsky, 1976), unsupervised learning detection (Celik & Ma, 2010; Duskunovic et al., 2000; Aach & Kaup, 1995), and supervised learning detection (Coppin et al., 2004; Dengkui et al., 2008). Simple differencing authenticates the change by simple calculation via the gray-scale image sequence of two images. Unsupervised learning change detection utilizes an automatic analysis on the differences of images based on clustering methods. Supervised change detection utilizes classification methods to identify where change happens on images based on the learning process of classifiers.

6.2.1 Existing Methods

Simple differencing methods locate changes by computing a predefined variation measurement and authenticate the change by a straightforward calculation of mean difference (R. Jain, 1984) or mean ratio (Yakimovsky, 1976). The accuracy of these methods are often affected by impulse noise (Vasicek & Bidlo, 2011), the single-pixel bright or dark spots caused by channel decoder damages, dyeing down of signal in communication links, communication subscriber's moving, video sensor's noise and others (Z. Wang & Zhang, 1999; Hwang & Haddad, 1995; A. K. Jain, 1989). For effective change detection, simple differencing methods are often incorporated with a threshold (R. J. Radke et al., 2005), which optimizes the change detection by a great amount of derivations and calculations based on experts' experience (Schowengerdt, 1983; Jensen, 1981; Mukai et al., 1987; Zhan et al., 2000).

Unsupervised learning methods estimate image differences (Black, Fleet, & Yacoob, 2000) by feature extraction. In this category, given an $M \times N$ image pair $\{X_1, X_2\}$, unsupervised learning approaches such as wavelet decomposition, principle component analysis, and clustering are employed for feature extraction. Afterwards, simple differencing is applied to compute the difference image $D(r)$ where $1 \leq r \leq M \times N$ denotes the index of pixel. The optimization can be achieved through constructing, 1) a mixture distribution on $D(r)$s from solving change probability density functions on hypotheses $P(D(r)|H_0)$ and $P(D(r)|H_1)$, where H_0 and H_1 denote no-change and change respectively, or 2) a Gaussian distribution (Aach et al., 1993; Bazi et al., 2005a; Celik & Ma, 2010), or 3) a Markov random field (Duskunovic et al., 2000; Aach & Kaup, 1995).

Supervised learning methods address change detection by a classification model built on a given **training set**. This type of methods generally can be categorized as, (1) conceptual maps comparison approach. Concepts (i.e., class labels) in the approach are manually segmented from the given image set. Accordingly, a classifier is trained to segment image and build the conceptual map for each input image. Then, image changes can be masked by comparing a pair of conceptual maps (Coppin et al., 2004). (2) pixel pairs change identification approach. The approach labels a pair of pixels as changed and unchanged, trains accordingly a classifier using the given image set, then identifies image changes by classifying a pixel pair into changed and unchanged classes (Dengkui et al., 2008).

6.2.2 Image Series Change Detection

Time series image change detection measures differences on images presented at successive time intervals. Previous methods generally rely on tracking a foreground object to detect the changes across different images. These methods assume that

only one foreground object exists in the image series (Tian et al., 2011; Borges & Izquierdo, 2010), which enables fast segmentation and trace of the object, and which gives often reliable change detection results. Figure 6.2 shows the working process of typical image series change detection. After the first change detection model M_1 being built for X_2 to X_1, for continuous detecting the change of X_3 to X_2, we have to drop M_1 and build a mew model M_2 for X_3 to X_2. It's worth noting that these methods in principle are working merely for a pair of images; and for image series change detection, re-building detection model requires expensive memory and computational costs.

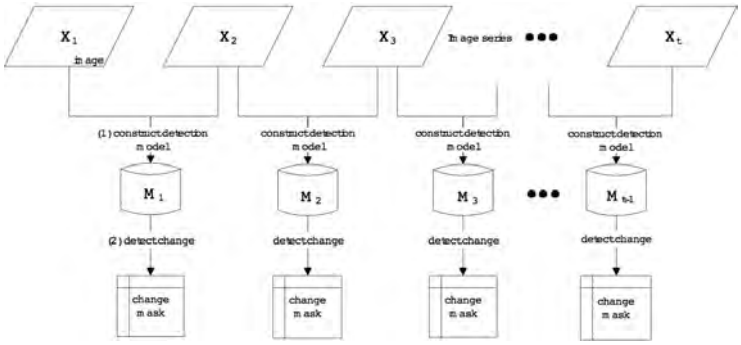


Figure 6.2: Graphical representation of existing image change detection methods for image series change detection.

6.2.3 Incremental Learning for Continuous Image Change Detection

Incremental learning is capable of processing large amount data streams with low re-calculation cost, thus can be considered as one optimal solution for image series change detection. Figure 6.3 shows the mechanism of image series change detection by incremental learning. Given an image series $X_1, X_2, \ldots, X_t$, through incremental learning the knowledge of source image X_1 can be transferred into M_1, a model that describes all objects of the image X_1. M_1 then can be used as a reference to detect the changes on the target image X_2 (via the proposed one-step-more incremental learning). For continuous change detection of X_3 against X_2, M_1 also can be updated to M_2 appending new knowledge of X_2 to M_1 through incremental learning. Different from previous works, that re-construct detection model to detect changes for each time instance, in the proposed method, we use a one-pass updating model M_t. The

implication is that instead of building change detection models from scratch, which have expensive memory and re-calculation costs, we consider first constructing a detection model, then updating the model for successive images change detection.

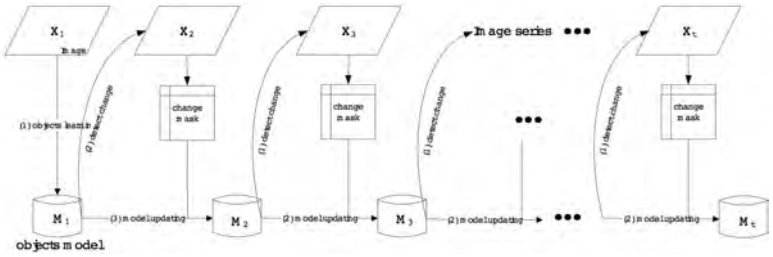


Figure 6.3: Graphical representation of incremental learning based image series change detection.

6.3 Proposed Image Change Detection Method

The idea of the proposed image change detection method is based on agent-based pattern recognition to set an agent to learn source image X_1 by continuous incremental learning and to perform one-step-more incremental learning on X_2 to detect the difference of X_2 to X_1.

Given $M \times N$ a series of images, $X_1 = \{x_1(m,n)|1 \leq m \leq M, 1 \leq n \leq N\}$ and $X_2 = \{x_2(m,n)|1 \leq m \leq M, 1 \leq n \leq N\}$, acquired at the same scene but at two different time instances with different objects, respectively. Assuming that such images are registered with respect to each other, and to avoid maximally the possible differences caused by photographic condition changes, the images are taken with minimum lighting change at two time instances. This section introduces a novel method for detecting change of X_2 against X_1 through agent-based incremental learning.

The proposed method includes three core stages, (1) image feature encoding; (2) incremental learning on X_1, and (3) agent-based change detection on X_2. Each stage consists of one or two sub-stages, which are depicted in the following subsections 6.3.1 - 6.3.3. Figure 6.4 gives the diagram of proposed agent-based incremental learning image change detection method.

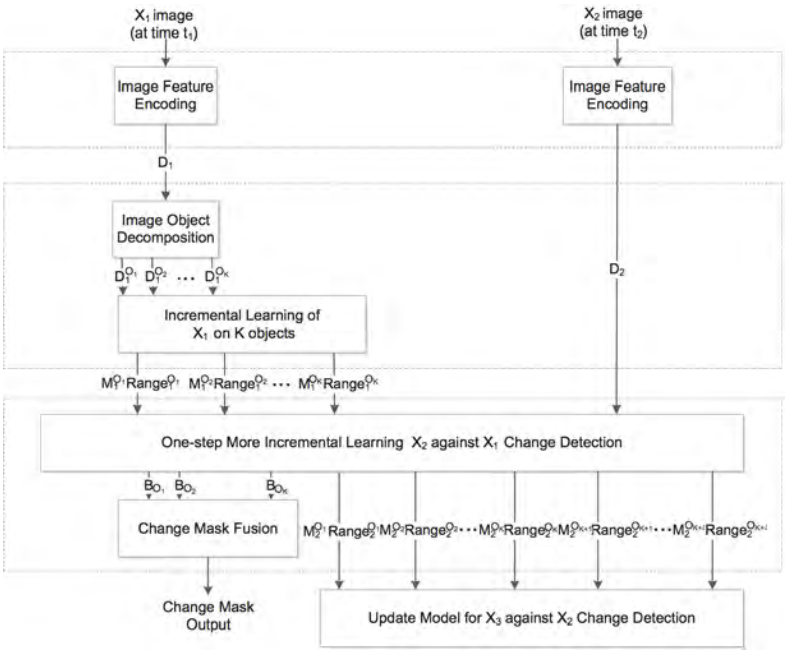


Figure 6.4: The diagram of proposed agent-based incremental learning image change detection system.

6.3.1 Image Feature Encoding

In order to provide useful image features for the proposed method, we select texture and color features to form image data $\boldsymbol{D}$. Texture feature has been widely used in object recognition, image content analysis and many others and is popularly emerged by the Gabor filter (W. Li, Mao, Zhang, & Chai, 2010). Because the Gabor filter can be interpreted as a nonlinear function that maps images from original space to feature space. Color features can be used to considerably improve object recognition performance (Choi, Ro, & Plataniotis, 2011; W. Li, Wang, Wang, & Chen, 2008). Due to the proposed approach is based on object recognition, $\boldsymbol{D}$ is formed by texture features $\boldsymbol{G}$ provided by the Gabor filter and color features $\boldsymbol{Color}$ from RGB of an image as

$$\boldsymbol{D} = [\boldsymbol{G}, \boldsymbol{Color}]. \tag{6.1}$$

According to previous research (Farmer & Jain, 2005; Tang, 2010; Zhuang & Wang, 2010), there are some typical features for objects detection in image series or video, such as color, area, roundness, contour, etc. Therefore (6.1) can be extended as (6.2)

for future development,

$$\boldsymbol{D} = [\boldsymbol{G}, \boldsymbol{Color}, \boldsymbol{roundness}, \boldsymbol{contour}, etc]. \tag{6.2}$$

Without lose of generality, we use d to denote the length of a feature vector. In our practice, we have $d = 19$.

For characterizing image contents, we apply first an image processing method - Gabor filter - to enhance images characteristics, such as texture and contour information. A general Gabor filter function is expressed as,

$$\boldsymbol{G}(x, y; \lambda, \theta, \psi, \sigma) = \exp\left(-\frac{x' + y'^2}{2\sigma^2}\right) \exp\left(x\left(2\pi\frac{x'}{\lambda} + \psi\right)\right), \tag{6.3}$$

where

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \tag{6.4}$$

and λ represents the sinusoid's wavelength, θ represents the orientation, ψ is the phase offset, and σ denotes the spread of the Gaussian window (Daugman, 1985).

Given a set of orientations θ^t,where $t = 1, 2, ..., T$, and according to (6.3) and (6.4), using an image I as input, we can obtain T different images $G^I_{\theta^t}$, which are the data representatives of these Gabor-filtered images in T orientations.

Let the 2-D image I with a size of $M \times N$ be processed by the Gabor filter, and the magnitude of complex-valued subbands is denoted as:

$$\boldsymbol{G}^X_{\theta^t} = \left\{ g^X_{\theta^t}(m, n) | 1 \leq m \leq M, 1 \leq n \leq N, \text{and } t = [1, 2, ..., T] \right\}, \tag{6.5}$$

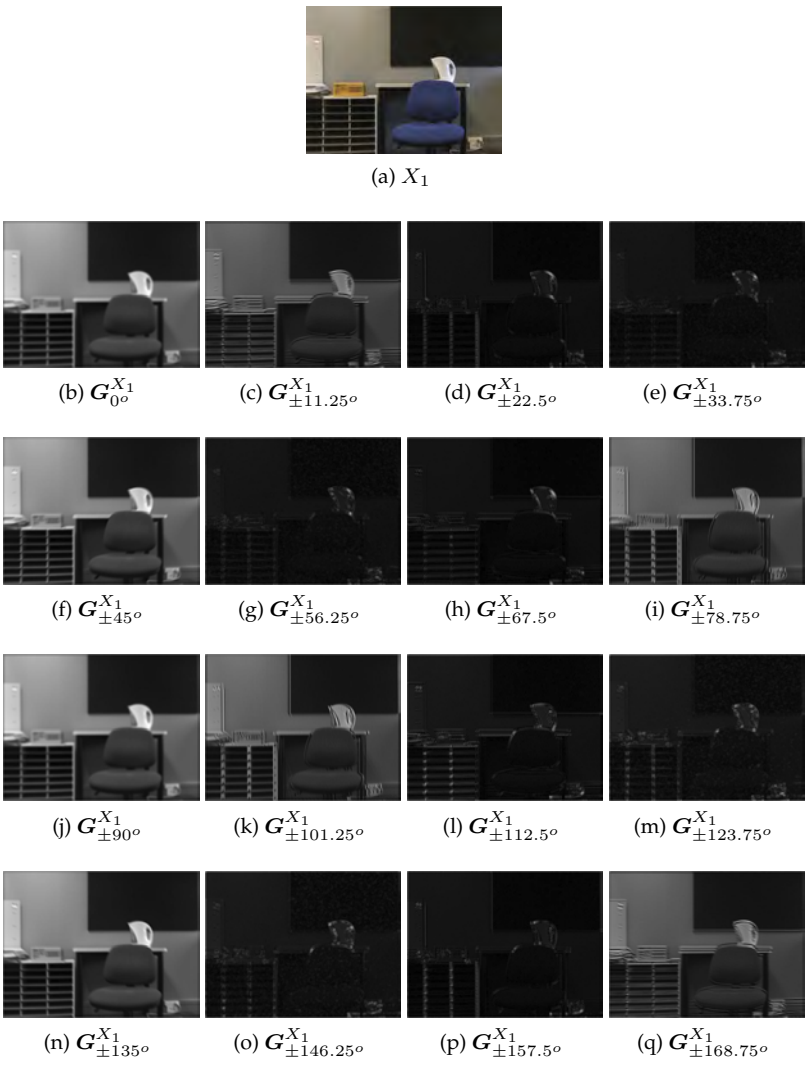
where $\boldsymbol{G}^I_{\theta^t}$ is a set of real-valued 2-D subbands signals, representing the magnitude of complex-valued wavelet coefficients obtained at the direction θ^t. The magnitudes of each directional subband's coefficients are computed and collectively denoted by $\boldsymbol{G}^{X_1}_{\theta^t}$ and $\boldsymbol{G}^{X_2}_{\theta^t}$ for the multiple temporal images X_1 and X_2. Figure 6.5 shows an original image X_1 and its corresponding magnitudes of the complex-valued directional subbands.

To identify the objects of X_1, we extract both texture set and color set of each pixel to form the source data $\boldsymbol{D}_1$.

6.3.2 Incremental Learning of X_1 Objects

Image objects decomposition

To discover content composition of X_1, we conduct clustering on $\boldsymbol{D}_1$ to discriminate different image objects from each other and group similar objects together. Given



(a) X_1

(b) $\boldsymbol{G}_{0^o}^{X_1}$ (c) $\boldsymbol{G}_{\pm 11.25^o}^{X_1}$ (d) $\boldsymbol{G}_{\pm 22.5^o}^{X_1}$ (e) $\boldsymbol{G}_{\pm 33.75^o}^{X_1}$

(f) $\boldsymbol{G}_{\pm 45^o}^{X_1}$ (g) $\boldsymbol{G}_{\pm 56.25^o}^{X_1}$ (h) $\boldsymbol{G}_{\pm 67.5^o}^{X_1}$ (i) $\boldsymbol{G}_{\pm 78.75^o}^{X_1}$

(j) $\boldsymbol{G}_{\pm 90^o}^{X_1}$ (k) $\boldsymbol{G}_{\pm 101.25^o}^{X_1}$ (l) $\boldsymbol{G}_{\pm 112.5^o}^{X_1}$ (m) $\boldsymbol{G}_{\pm 123.75^o}^{X_1}$

(n) $\boldsymbol{G}_{\pm 135^o}^{X_1}$ (o) $\boldsymbol{G}_{\pm 146.25^o}^{X_1}$ (p) $\boldsymbol{G}_{\pm 157.5^o}^{X_1}$ (q) $\boldsymbol{G}_{\pm 168.75^o}^{X_1}$

Figure 6.5: An original image X_1 and the corresponding magnitudes of the complex-valued directional subbands.

dataset $\boldsymbol{D}_1 = \{\boldsymbol{D}_1(r)|1 \leq r \leq M \times N\}$ representing source image X_1 in feature space, where $\boldsymbol{D}_1(r)$ is the feature vector of pixel $X_1(m,n)$, the goal of clustering is

to discover in source image K objects, $O_1, O_2, \ldots, O_K$ such that

$$\begin{array}{ll} \boldsymbol{D}_1^{O_k} \neq \phi, & \text{for } k = 1, \ldots K \\ \boldsymbol{D}_1^{O_k} \cap \boldsymbol{D}_1^{O_\eta} = \phi, & \text{for } k, \eta = 1, \ldots K; k \neq \eta \\ \bigcup_{k=1}^{K} \boldsymbol{D}_1^{O_k} = \boldsymbol{D}_1 & \end{array} , \tag{6.6}$$

and the image objects belonging to same cluster are similar in the sense of the given metric, while the objects belonging to different clusters are dissimilar in the same sense.

Incremental learning of objects

To acquire knowledge from source image X_1, we set an intelligent agent powered by a one-pass incremental learning algorithm, to perform incremental learning of objects obtained above.

For each object, we define object model $\boldsymbol{M}_1^{O_k}$ as a covariance on feature space,

$$\boldsymbol{M}_1^{O_k} = \begin{bmatrix} \boldsymbol{D}_1^{O_k} & -\boldsymbol{e} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{D}_1^{O_k} & -\boldsymbol{e} \end{bmatrix}, \tag{6.7}$$

where $\boldsymbol{e} = [1, \ldots, 1]^T$.

It is worth noting that the learning of $\boldsymbol{M}_1^{O_k}$ in practice is an incremental learning procedure. When $\boldsymbol{D}_1^{O_k}$ is a huge size dataset, the calculation of (6.7) confronts often the difficulty of out of memory. Thus, we derive the rule of incremental learning as,

1. LEMMA. *Given object model M on $\boldsymbol{D}$, and $\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D_1} \\ \boldsymbol{D_2} \end{bmatrix}$, then M can be learned incrementally as*

$$\begin{aligned} \boldsymbol{M} &= \begin{bmatrix} \boldsymbol{D} & -\boldsymbol{e} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{D} & -\boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}_1 & -\boldsymbol{e} \\ \boldsymbol{D}_2 & -\boldsymbol{e} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{D}_1 & -\boldsymbol{e} \\ \boldsymbol{D}_2 & -\boldsymbol{e} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} \boldsymbol{D}_1 & -\boldsymbol{e} \end{bmatrix}^T & \begin{bmatrix} \boldsymbol{D}_2 & -\boldsymbol{e} \end{bmatrix}^T \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \boldsymbol{D}_1 & -\boldsymbol{e} \end{bmatrix} \\ \begin{bmatrix} \boldsymbol{D}_2 & -\boldsymbol{e} \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{D}_1 & -\boldsymbol{e} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{D}_1 & -\boldsymbol{e} \end{bmatrix} + \begin{bmatrix} \boldsymbol{D}_2 & -\boldsymbol{e} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{D}_2 & -\boldsymbol{e} \end{bmatrix} \end{aligned} \tag{6.8}$$

Though Lemma 1 we can slice a large $\boldsymbol{D}_1^{O_k}$ into pieces with suitable size and calculate $\boldsymbol{M}_1^{O_k}$ incrementally.

To enhance the understanding of the objects, we review the learning of each object by a second round of incremental learning on each pixel,

$$\boldsymbol{M'}_1^{O_k}(r) = \boldsymbol{M}_1^{O_k} + \begin{bmatrix} \boldsymbol{D_1^{O_k}}(r) & -\boldsymbol{e} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{D_1^{O_k}}(r) & -\boldsymbol{e} \end{bmatrix}. \tag{6.9}$$

To measure the influence of an individual pixel to its object model, we calculate the eigenvalue of the updated $\boldsymbol{M'}_1^{O_k}(r)$. Repeating the above operation on all pixels that belong to the object, we obtain an influence range as,

$$Range_1^{O_k} = \left[\min(\text{eig}(\boldsymbol{M'}_1^{O_k}(r))), \max(\text{eig}(\boldsymbol{M'}_1^{O_k}(r)))\right], \tag{6.10}$$

where $eig(.)$ denotes the eigenvalue calculation implemented in two steps: (1) perform eigen decomposition on $\boldsymbol{M'}_1^{O_k}(r)$ as $\boldsymbol{M'}_1^{O_k}(r)\boldsymbol{V} = \boldsymbol{\lambda V}$, and (2) let $\text{eig}(\boldsymbol{M'}_1^{O_k}(r)) = max(diag(\boldsymbol{\lambda}))$.

Through the above two rounds of incremental learning, source image X_1 is learned with its learning knowledge represented as K object models $\{< \boldsymbol{M}_1^{O_k}, Range_1^{O_k} > \}_{k=1}^{K}$. The memory cost is $K \times ((d+1)^2 + 2)$ since $\boldsymbol{M_1}^{O_k} \in \mathbb{R}^{(d+1)\times(d+1)}$ and $Range_1^{O_k} \in \mathbb{R}^{1\times 2}$. Note that this is significantly smaller than the original image $M \times N$.

6.3.3 One-step-more Incremental Learning on X_2 for Change Detection

The idea of change detection against target image X_2 is to measure the influence of every target image pixel to the object model that is associated with its corresponding source image pixel. Here, we assume, without loss of generality, every target image pixel is about the same object as its corresponding pixel in the source image.

Detect change against X_1

For the rth pixel in target image, on the top of the object model $M_1^{O_k}$ that the rth source image pixel associates, one-step-more incremental learning is conducted on the rth target image pixel,

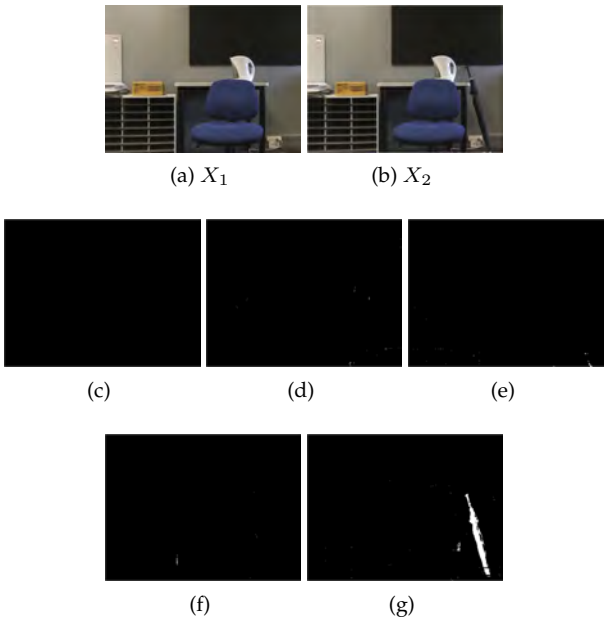
$$\boldsymbol{M''}^{O_k}(r) = \boldsymbol{M}_1^{O_k} + \left[\boldsymbol{D_2^{O_k}}(r) \quad -\boldsymbol{e}\right]^T \left[\boldsymbol{D_2^{O_k}}(r) \quad -\boldsymbol{e}\right], \tag{6.11}$$

where $\boldsymbol{D}_2^{O_k}(r)$ represents the feature vector of the rth target image pixel. We assume the pixel corresponds to the object O_k of the source image.

To evaluate the influence of the rth pixel in target image to the source image object models, again we calculate the eigenvalue on the object model after one step more incremental learning as $\text{eig}(\boldsymbol{M''}^{O_k}(r))$. Then, the status of the rth pixel as compared to the source image objects is determined by,

$$b_{O_k}(r) = \begin{cases} 0, & \text{eig}(\boldsymbol{M''}^{O_k}(r)) \in Range_1^{O_k} \\ 1, & \text{eig}(\boldsymbol{M''}^{O_k}(r)) \notin Range_1^{O_k}. \end{cases} \tag{6.12}$$

Consequently, a binary image (or mask) $B_{O_k} = \{b_{O_k}(m,n)|1 \leq m \leq M, 1 \leq n \leq N\}$ are formed for each object defined in X_1. Figure 6.6 gives an example of the proposed change detection, where the change of X_2 against X_1 is detected by (6.12) on 5 X_1 objects, respectively.



(a) X_1 (b) X_2

(c) (d) (e)

(f) (g)

Figure 6.6: Change detection conducted on different objects (c) B_{O_1}, (d) B_{O_2}, (e) B_{O_3}, (f) B_{O_4}, (g) B_{O_5}.

Mask fusion

Based on object mask B_{O_k} obtained above, the final binary change mask (BCM) is generated by performing object mask fusion as,

$$BCM = \bigcup_{\forall O_K} B_{O_k}, \tag{6.13}$$

where $\bigcup$ performs a binary union operation. Figure 6.7 shows the final merged binary change detection mask BCM based on (6.13).

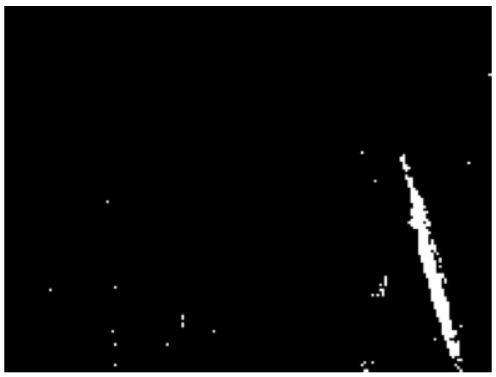


Figure 6.7: The change detection mask, BCM, resulted by merging all 5 binary masks according to (6.13).

Knowledge updating for continuous image change detection

As the result of incremental learning on X_1, we obtain object models $\{< \boldsymbol{M}_1^{O_k}, Range_1^{O_k} > \}_{k=1}^{K}$. When continuous change detection is required after the X_2 to X_1 change detection (i.e. detecting change of a X_3 to X_2), straightforwardly we repeat the above process to conduct the same incremental learning on X_2 to obtain X_2 object models, then perform one-step-more incremental learning to detect the changes of X_3 against X_2. We calculate X_2 object models by object models as:

$$\{< \boldsymbol{M}_2^{O_k}, Range_2^{O_k} >\}_{k=1}^{\eta} = \mathcal{F}(\{< \boldsymbol{M}_1^{O_k}, Range_1^{O_k} >\}_{k=1}^{K}, X_2, B_{O_k}), \qquad (6.14)$$

where $\eta \geq K$ is determined by the similarity of X_2 to X_1, and $\mathcal{F}$ is the updating function, which can be implemented by the following steps,

1. If pixel $X_2(r)$ is marked as "changed" in B_{O_k}, the following operations are efficiently performed:
   (a) Conduct one-step-more incremental learning via (6.9) on top of all exiting object models except the kth object (i.e., $\boldsymbol{M}_1^{O_j} | j \in [1, K], j \neq k$);
   (b) Evaluate the influence of the $X_2(r)$ on object models of source image $\boldsymbol{M}_1^{O_j} | j \in [1, K], j \neq k$ by (6.10);
   (c) Assign $X_2(r)$ to jth object if $\text{eig}(\boldsymbol{M}''^{O_j}(r)) \in Range_1^{O_j}$; otherwise, create new objects through (6.6) and generate new object models by (6.7), (6.9), (6.10);
2. For all unchanged pixels, they keep the same object identity.

For image series change detection, the above updating procedure can be repeated to obtain object models on subsequential images, so that change detection can be conducted continuously.

6.4 Experimental Results and Discussion

6.4.1 Performance Evaluation Criteria and Robustness Test

For quantitative performance evaluation, following (Celik & Ma, 2010; Bazi et al., 2005a) we use changes in region of interest (ROI) to generate the ground truth. Given the first image X_1, we modify the pixel intensity values of X_1 in an arbitrarily chosen ROI to create the second test image X_2 via,

$$x_2(m,n) = \begin{cases} x_1(m,n) - A, & \text{if } (m,n) \in \textbf{ROI} \\ x_1(m,n), & \text{otherwise} \end{cases}, \tag{6.15}$$

where $(m,n) \in$ ROI and A represents the absolute-valued difference between X_1 and X_2 in the selected ROI area, which is assigned as 40 in our experiments.

Let L be the total number of pixels, we calculate the following quantities of computed binary change mask for comparing the computed change detection mask against the ground truth change detection,

$$P_{TP} = TP/L,\ P_{FP} = FP/L,\ P_{TN} = TN/L, \text{ and } P_{FN} = FN/L, \tag{6.16}$$

where TP is the number of "changed" pixels that were correctly detected (True positives), FP is the number of "unchanged" pixels that were incorrectly detected as "changed" (False positives), TN is the number of "unchanged" pixels that were correctly detected (True negatives), and FN is the number of "changed" pixels that were incorrectly detected as "unchanged" (False negatives) (R. J. Radke et al., 2005). Then we measure change detection by the correct classification accuracy P_{CC} and the false classification accuracy P_{FC}, respectively,

$$\begin{aligned} P_{CC} &= P_{TP} + P_{TN} \\ P_{FC} &= P_{FP} + P_{FN} \end{aligned}. \tag{6.17}$$

Except for above measurements, for image series change detection, we are particularly interested in computational efficiency in terms of the average number of calculation over each pixel.

6.4.2 Region of Interest Detection

Following (R. J. Radke et al., 2005; Celik & Ma, 2010; Coppin et al., 2004; Bruzzone & Prieto, 2002), we conduct first region of interest detection experiments to investigate the ability of the proposed algorithm on combating different scenarios of an indoor environment. In doing that, we generate according to (6.15) total 250 pictures based on 5 background scenarios that we select on purpose in our research lab to cover the case study of different light conditions, backgrounds, and time frames. Figure 6.8 gives the example images of these 5 background scenarios. We conduct experiments for region of interest detection on those 250 pictures. The experimental results are summarized in Table 6.1, where the performance is evaluated on the 6 measurements discussed in previous section, and the results are formated in the form of 'average accuracy $\pm$ standard division'. As seen, the accuracy varies over different scenarios. However, the robustness of the proposed method is apparent in that both the average correct classification accuracy PCC and the false classification accuracy PFC are shown consistently above 95%.

Scenario 1 Scenario 2 Scenario 3 Scenario 4 Scenario 5

Figure 6.8: 5 scenarios in a computer lab to cover case study of light condition, background.

Table 6.1: Change Detection Performance Results of Incremental Learning Image Change Detection on Tasks with 5 Scenarios

Image pair						
Scenarios	P_{TP} (%)	P_{FP} (%)	P_{TN} (%)	P_{FN}(%)	P_{FC} (%)	P_{CC} (%)
Scenario 1	6.1999 $\pm$ 2.9023	0.0011 $\pm$ 0.0013	89.4803 $\pm$ 4.4816	4.3188 $\pm$ 2.7924	4.3198 $\pm$ 2.7925	95.68 $\pm$ 2.7925
Scenario 2	4.0390 $\pm$ 2.2047	9.4105e-004 $\pm$ 0.0016	94.3023 $\pm$ 2.6746	1.6577 $\pm$ 0.9751	1.6587 $\pm$ 0.9749	98.34 $\pm$ 0.9749
Scenario 3	8.0554 $\pm$ 5.0419	1.9531e-004 $\pm$ 3.7379e-004	90.2724 $\pm$ 5.9366	1.6720 $\pm$ 1.2531	1.6722 $\pm$ 1.2531	98.33 $\pm$ 1.2531
Scenario 4	3.1127 $\pm$ 2.2234	2.0715e-004 $\pm$ 8.6662e-004	94.9211 $\pm$ 3.2231	1.9660 $\pm$ 1.5007	1.9662 $\pm$ 1.5006	98.03 $\pm$ 1.5006
Scenario 5	1.0248 $\pm$ 0.8763	4.3205e-004 $\pm$ 0.0016	98.4273 $\pm$ 1.2518	0.5475 $\pm$ 0.6382	0.5480 $\pm$ 0.6386	99.45 $\pm$ 0.6386

6.4.3 Static Image Change Detection

Different sets of images (i.e., two images in one set), as shown in Figure 6.9(a) and Figure 6.9(b), were taken in our research lab. The final change detection masks resulted in each experiment are presented in Figure 6.9(c), correspondingly. In the experiments, we assume, with out loss of generality, maximum 5 objects exist in each image, thus the initial number of clusters for k-mean clustering is set as 5 (i.e., $K = 5$). As seen, Figure 6.9 gives at the first two rows two examples for single object change detection, where the first image change is a chair, which is a artificial change manually added to the image. The second is from two real-world images, taken at different time under the same photographic condition with a differentiation of a standing person in-between. The remaining two rows of Figure 6.9 presents two case studies for multi-object change detection. The image change for the first case is also artificial. The differences are a chair plus a bottle. For the second real-world change, a box and a billboard are used as observing objects.

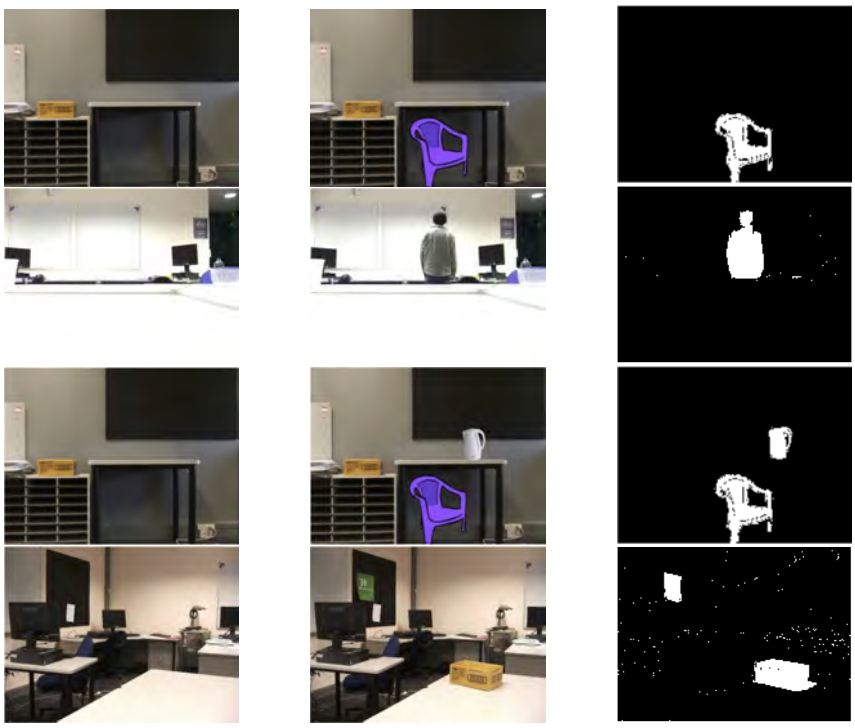
6.4.4 Image Series Change Detection

We applied the proposed method to detect a sequence of changes occurred in image series. Figure 6.10 presents the experimental results. Figure 6.10 (a) to (e) are the image series for continuous image change detection. The difference between image (a) and (b) is a cup. A stapler was added in (c). A small inflater is the difference between (c) and (d). (e) has an additional wallet as compared to (d). As seen, Figure 6.10 (f) to (i) give the change masks resulted from each step of change detection.

To compare the proposed incremental learning based change detection with traditional unsupervised learning change detection, we investigate image series change detection in terms of the computational efficiency and memory usages of algorithms

Computational efficiency

Unsupervised learning change detection exploits probability density function to establish difference image in order to achieve change detection automatically. However, for each pair of pixel, unsupervised learning change detection requires an iterative process to conduct the maximum likelihood estimation for tackling incomplete-data problems (Bazi et al., 2005a; Celik & Ma, 2010). Thus, for each image pair, the computational cost of unsupervised learning change detection can be counted by *pixel re-calculation times* (i.e., the number of iterations calculation on each pair of pixels). However in the proposed method, re-calculation happens only when a pixel is



(a) (b) (c)

Figure 6.9: Change detection results using different sets of images: (a) Input images X_1, (b) Input images X_2, and (c) final change detection masks **BCM.**

identified changed on its belonging in target image (i.e,. the pixel in source image belongs one object, but in target image, it belongs to a different object or contributes to a new object). This reduces substantially the computational cost for image series change detection, because the number of changed objects is relatively small, thus most pixels do not change their object belonging. Figure 6.11 records the computational costs in terms of *pixel re-calculation times* for an image series change detection by an unsupervised learning change detection and proposed incremental learning method.

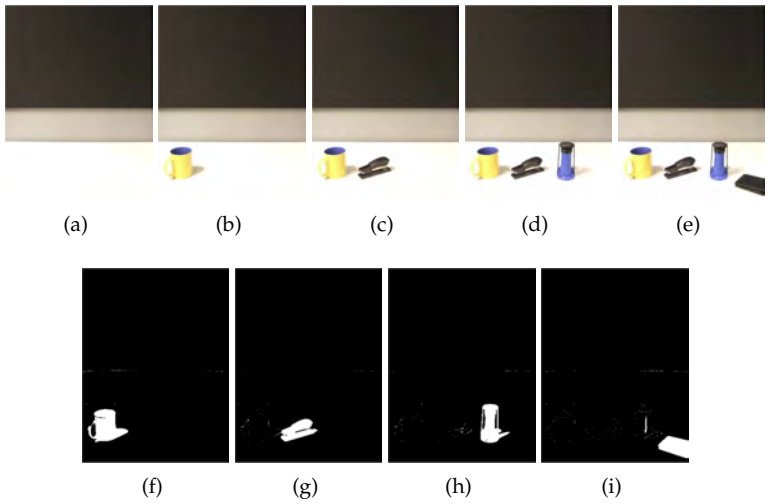


Figure 6.10: Change detection results using a sets of images: (a) is the source image X_1 and (b) to (e) are target images X_2 to X_5. (f) to (i) are BCMs for each step of change detection.

Memory usage

Existing computational methods require basically two images for change detection to be stored in memory throughout the computing process. Thus, the memory usage increases with the size of image. In contrast, the proposed method stores just the source image object models $\{< \boldsymbol{M}_1^{O_k}, Range_1^{O_k} >\}_{k=1}^{K}$ and its label matrix. Thus its memory usage is determined by the number of objects. Given a pair of images containing 3 objects, Figure 6.12 compares the simple differencing method and proposed method for change detection, when the two images are enlarged equally from 205×272 to 682×908. As seen, during the procedure the proposed method has a very slight memory increase against the image size. However, the simple differencing method expands its memory usage eight times, when the image size increases to 682×908.

6.5 Conclusions and Future Works

Image change detection technology plays an important role in diverse practical applications and has therefore attracted increasing research interests (R. J. Radke et al.,

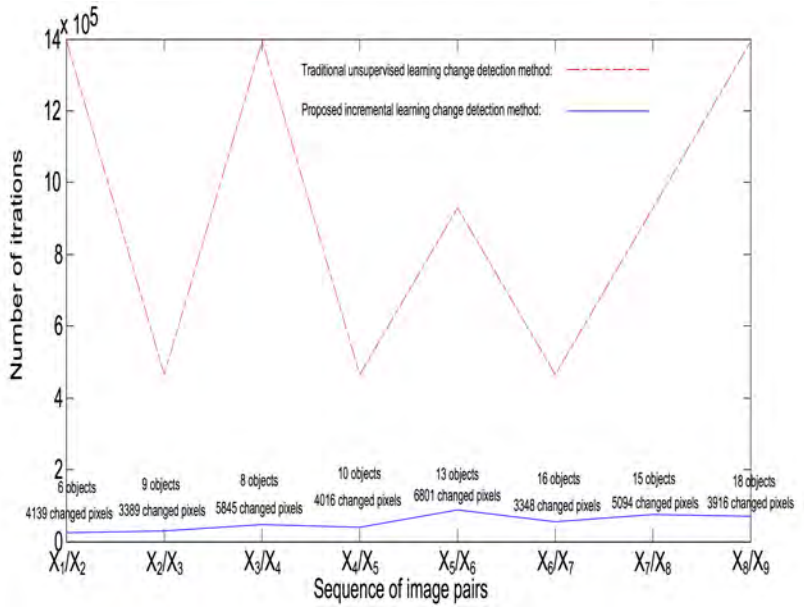


Figure 6.11: Computational costs comparison, proposed method vs. unsupervised learning change detection.

2005; Aach et al., 1993; Bazi et al., 2005a; Celik & Ma, 2010; Coppin et al., 2004). Due to the computational complexity of rebuilding the change detection model and the accompanying high memory usage, as far as we know, most of the existing change detection methods have difficulties of detecting changes in a series of images (Tian et al., 2011; Borges & Izquierdo, 2010). To tackle this problem, an innovative agent-based one-step-more change detection approach has been developed in this chapter. Our new approach utilized an agent-based learning model for efficient detection of continuous changes in image series. With the help of incremental learning technology, the computational cost of initializing and updating the change detection model can be also significantly reduced.

The core technologies employed in the proposed image change detection method include (1) agent-based incremental learning for efficient knowledge extraction from image series, and (2) one-step-more incremental learning for flexible and accurate image change detection. Given two consecutive images in the image series, the source image is first utilized by the incremental learning agent to build up the knowledge base of the change detection model. Driven by the model, one-step-more incremental learning is further performed on the target image to determine any changes in comparison with the source image. After that, incremental learn-

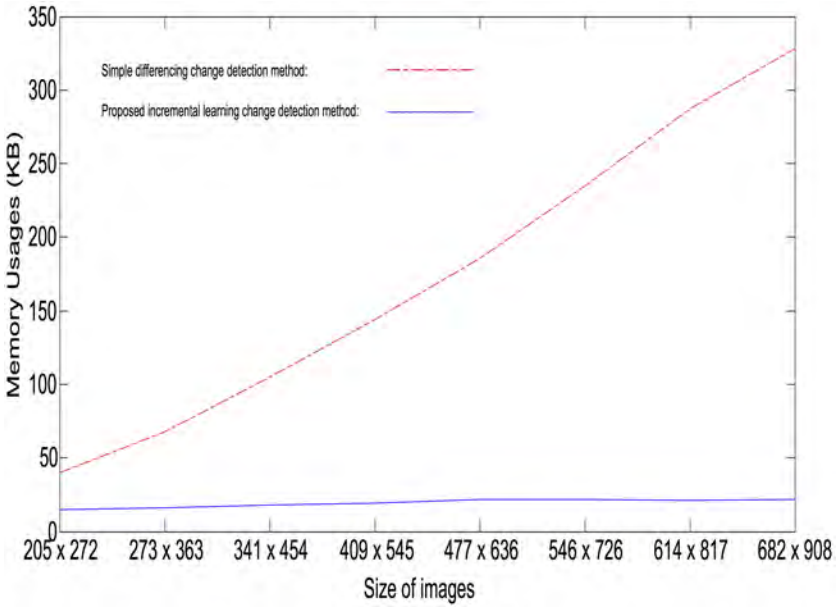


Figure 6.12: Memory usage comparison, proposed method vs. simple differencing method.

ing will be performed again to update the change detection model derived from the source image according to the target image. The updated model enables us to detect any continuous changes in upcoming images.

Lighting often causes unexpected frequency distortion on indoor images. As a result of this distortion and the impulse noise, as presented in Chapter 3, the distribution of image features extracted is complicated and often not linearly separable. Therefore, the linear SVM model utilized in this chapter may not be flexible enough in order to satisfactorily manage the lighting effects. Meanwhile, shadows in indoor images can sometimes cause incorrect objects decomposition. In our experiments shadows always appear at the intersection of two objects. Due to the influence of time-varying shadows, the learning model may not be able to correctly interpolate the textures.

In the future, efforts will be made to enhance the robustness of the proposed change detection method in the presence of the two major physical variations (i.e. lighting and shadows). Extensive comparisons with other image change detection methods reported in the literature will also be conducted to further improve our understanding of the proposed approach and its technical strengths.

Chapter 7

Agent based Land Encroachment Detection

Land management and planning is essential to assist the economic growth, sustainable resource use and environmental protection of a city. This chapter describes a novel approach to automatic encroachment detection to assist in land management decision making. The approach begins with training the agent to identify and understand the land cover/use features (such as buildings, parks, trees and roads) that are predominant in the region of interest, and carries out segmentation on the park data using the intelligent agent developed from the training samples. Experiments on park images from Auckland New Zealand show the effectiveness of the proposed approach.

Experiments have been performed on data collected in the form of digital images of parks from the Manukau and North shore areas of Auckland that include some of the 800 public parks in Auckland, New Zealand. The results of the experiments carried out successfully on four examples of encroachment show the effectiveness of the approach. The impact/cost/scale of encroachment is evaluated by multiplying the size of the encroachment by the rateable value of the land that has been encroached upon.

7.1 Introduction

We define land encroachment as the change in the perceived or actual use of land from either (a) human caused encroachment: the use of public land for private purposes, or (b) environmental caused encroachment: the change in the ability of land for its original purpose from an environmental change.

The use of public land for private purposes has been identified as a problem affecting public parks in the Auckland region. Dacey, Barbour, and Fernando (2011) proposed a participatory approach to the problem of encroachment. This approach involves visiting each park with GPS equipment to obtain data on the actual physical boundary. This data collection proved to be very time-consuming and no significant encroachment was detected. The data collection was carried out on 20 parks and took three weeks - there are over 800 parks in the Auckland region. So, a quicker

way of identifying parks with likely encroachment is needed.

This chapter examines the use of digital image analysis in automatically detecting encroachment on public parks in two specific areas in Auckland, New Zealand. Since the launch of the first Earth Resources Technology Satellite in 1972 (ERTS-1, later renamed Landsat 1), there has been significant activity related to mapping and monitoring environmental change as a function of anthropogenic pressures and natural processes according to Treitz (Treitz & Rogan, 2004). National park units and protected areas face critical management challenges because of changing land-cover types and variability of landscape contexts within and adjacent the park boundaries according to Wang (Y. Wang et al., 2009).

In some studies, land cover change is synonymous with land use change. However, when dealing with encroachment, it is sometimes necessary to differentiate between land cover change and land use change. An example of this type of encroachment is where an area of grassland belonging to a public park has been fenced off by the resident of a property that is adjacent to the park. The land cover of the majority of the grassland has not changed but its land use has changed from public to private. Detection of this type of change requires the use of high resolution data as features such as fences are hard to detect and/or difficult to distinguish from features, such as paths, that are not indicative of encroachment.

The strengths of the proposed method is that it can detect the types of land encroachment identified above. Additionally, it is suited to the type of land cover (urban and rural mixed) found in Auckland, as we design, build and test a solution for our specific purpose. Although trees naturally form part of a public park, they need to be differentiated because they can blur the boundary of a park and can obscure encroachment.

7.2 Related Work

In this section we review current research in the domain of land encroachment detection for the purpose of informing our research study. We look at previous researches from two perspectives: the first perspective is to examine the types of encroachment covered by the research, and secondly we examine the type of methods used in the research for land cover analysis. Our new approach is justified because the existing methods are unsuitable for our specific needs, and the applications examined are not reproducible in the New Zealand context.

7.2.1 Applications of Land Encroachment Detection

There is global interest in the area of land encroachment and its importance in science, the environment and socioeconomic assessments (Xian, Homer, & Fry, 2009). It is essential that the exact nature of the land characteristics of an area being studied is understood because it affects the decision of which land feature detection method to use. A variety of applications of land encroachment detection, with a focus on the discovery of the nature of encroachment, are discussed here.

There are examples of research that are in the rural context (Ghosh, Mishra, & Ghosh, 2011) and examples in the urban-rural context(He, Weib, P., Zhang, & Zhao, 2011) . Several investigations of land encroachment detection examine ecosystems around national parks and concentrate on suburban growth, deforestation and fragmentation. These investigations predominantly study human-induced land cover/use changes and an example can be seen in (Y. Wang et al., 2009). There are investigations that concentrate on the assessment of vegetation changes (Waser et al., 2008) and on burned area identification (Bruzzone & Preito, 2000). Other research examines various methods of detecting encroachment on different types of terrain and one research paper analysed the extent of agricultural encroachment in mountain ranges (Belaid, 2003).

A remote sensing approach using land cover mapping, (Malik & Husain, 2006), uses satellite sensor data evaluation for mapping different land-covers in the suburb of Rawalpindi (Pakistan) to assess the impact of urbanization on the scrub forest. Similarly, the case studies presented in (Ghosh et al., 2011) look at image change detection in rural areas in Mexico and Italy, for the purpose of monitoring urban growth as opposed to deforestation, using LANDSAT data.

In a recent survey of land encroachment management applications (Kennedy et al., 2009), many of the studies were initiated by government and civil authorities for the purpose of understanding, decision making and policy making. The purpose of the research in these areas is to increase understanding of change and how to model environmental and ecosystem impacts, and for land use planning (Xian et al., 2009; Y. Wang et al., 2009; Le Hegarat-Mascle, Ottle, & Guerin, 2005; He et al., 2011).

We now discuss a collection of methods for land encroachment detection.

7.2.2 Encroachment Detection Methods

This section discusses methods for encroachment detection using image change detection. In (Kennedy et al., 2009), discusses the broader context of applications of remote sensing and the design of monitoring projects. The study discusses guidelines for collaboration between scientists and land managers to develop remote sensing solutions for more effective landscape monitoring, as well as providing comprehen-

sive resources and attributes for remote sensing projects and high level algorithmic design considerations. Much research work has been carried out in the area and applications of land encroachment using image change detection. The change detection techniques we identify here include correlation image analysis, wavelet-fuzzy analysis, Fuzzy clustering, Fuzzy segmentation, change-vector analysis, image segmentation and object-based image analysis

In Object-based change detection using correlation image analysis and image segmentation (Im, Jensen, & Tullis, 2008b) used correlation image analysis and image segmentation for object-based change detection. The study introduced change detection based on object/neighbourhood correlation image analysis and image segmentation techniques. Five different change detection methods were investigated to determine how new contextual features could improve change classification results. The techniques could be used for detecting encroachment but they rely on the availability of multi-temporal images - not always possible for the Auckland area. The techniques operate on multi-spectral satellite images which are expensive and time-consuming to obtain.

In Wavelet-fuzzy hybridization: Feature-extraction and land-cover classification of remote sensing images (Shankar, Meher, & Ghosh, 2011) a wavelet feature based supervised scheme for fuzzy classification of land covers in multi-spectral remote sensing images is proposed. The proposed scheme is developed in the framework of wavelet-fuzzy hybridization, a soft computing approach. The approach could be utilized to aid automatic encroachment detection but, again, it uses multi-spectral satellite images which are expensive and time-consuming to obtain for Auckland.

The technique used in Fuzzy clustering algorithms for unsupervised change detection in remote sensing images (Ghosh et al., 2011) is based on fuzzy clustering approach and takes care of spatial correlation between neighboring pixels of the difference image produced by comparing two images acquired on the same geographical area at different times. The case studies presented in (Ghosh et al., 2011) look at change detection in images in rural areas. So the method may not be appropriate for the Auckland region which is a mixture of urban and rural. Also, the case studies discussed by (Ghosh et al., 2011) deal with multi-spectral image data which is not easily available for the Auckland region.

Fuzzy image segmentation was proposed recently as an alternative GEOBIA method for conducting discrete land cover classification. In quantitative land cover change analysis using fuzzy segmentation (Lizarazo, 2011) a variant of fuzzy segmentation is applied for continuous land cover change analysis. The approach could be adapted to examine encroachment detection but it does need ground-proofing using GPS data. Also, the case studies are looking at urban areas which is suited to Auckland as it is also a region that has urban elements. The case study presented in

the paper uses data from LANDSAT so the method would need to be adapted to be used on the aerial photograph data that is available for Auckland.

Although traditional spectral-based change-vector analysis (CVA) can effectively detect LULC change in many cases, it encounters difficulties in rural urban fringe areas (RUFAs) because of deficiencies in the spectral information of satellite images. To detect LULC changes in RUFAs effectively, this paper proposes an extended CVA approach that incorporates textural change information into the traditional spectral-based CVA (He et al., 2011). The method described would need to be adapted to work on the available aerial photographic data. The method is specifically developed for use in RUFAs of which there are several in Auckland. So this method could be used but needs to be adapted as LandSat data for Auckland is not easily available.

Assuming the linear mixing model for CR pixels, the problem is that both the multi-temporal class features and the pixel composition in terms of classes are unknown. The proposed algorithm is then based on the iterative alternate estimation of each unknown variable. At each iteration, the class features are estimated, thanks to the knowledge of the composition of some pixels, and then the pixel composition is re-estimated knowing the class features. The subset of known composition pixels is the subset of pixels where no change has occurred, i.e. the previous land cover map is still valid. It is derived automatically by removing at each iteration the pixels where the new composition estimation disagrees with the former one. Finally, for the final estimation of the pixel composition, a Markovian chain model is used to guide the solution. The approach has been first validated using simulated data with different spatial resolution ratios (Le Hegarat-Mascle et al., 2005)

Object-based image analysis has proven its potentials for remote sensing applications, especially when using high-spatial resolution data according to (Z. Wang, Jensen, & Im, 2010). An automatic region-based image segmentation algorithm for remote sensing applications is described by (Z. Wang et al., 2010) introduces a new automatic algorithm based on k means clustering (RISA), specifically designed for remote sensing applications. Although the technique described uses a segmentation algorithm, the technique is not ideally suitable for automatic encroachment detection in the Auckland area as it requires digital image data obtained from satellites.

To summarize, none of the techniques described above are particularly suited to automatic detection of encroachment in the Auckland area because either the characteristics of the land features are dissimilar to those found in Auckland or the type of encroachment detected by the technique is not included in the four types of encroachment that we are interested in.

7.3 Agent-based Land Encroachment Detection

Agent-based image analysis has been used in a variety of applications such as range-image segmentation (Mazouzi, Guessoum, Michel, & Batouche, 2008), pattern recognition (Gaceanu, 2012), and off-road vehicle guidance, (Broggi & Cattani, 2006). The advantage offered by such a solution includes: (1) image processing parallelization; and (2) the flexibility to concentrate on either spatial or temporal changes. A comprehensive review on agent-based image segmentation can be found in (Mishra, Srivastava, Shukla, & Singhal, 2011).

Addressing land-encroachment problems of public parks in the area of Auckland New Zealand, we conduct land encroachment detection by implementing the one-step-more incremental learning method developed in Chapter 6. The aim of the approach is to detect any changes between two multi-temporal images of a park, where we identify the two images as source image X_1 and target image X_2, the same as in Chapter 6.

We set an agent to learn objects from source image, then carry knowledge to detect changes occurs to target image. To learn as much as possible from X_1, in terms of the type of land cover, additional park land images are collected for training the agent to know 5 land cover patterns such as Grass, Buildings, Trees, Fences and Roads, which represent the land cover/use characteristics that are most likely to be found in or around public parks.

The flowchart in Figure 7.1 shows an outline of the proposed land encroachment detection, where we present only the main path of feature extraction, model training, and change detection. We use Gabor filters for the feature extraction of land

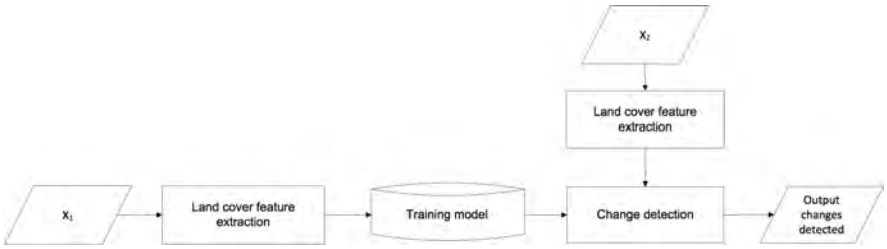


Figure 7.1: The process of applying one-step-more incremental learning method for land encroachment detection.

images X_1 and X_2. To extract objects from X_1, we conduct k-mean clustering on $\boldsymbol{D}_1$ (X_1 feature set) to identify the similar components. We set agent to learn knowledge from the extracted objects for training the detection model. As detecting change on

X_2, we use the trained agent to compare the components of X_2 against X_1.

7.3.1 Land Cover Feature Extraction

The land cover of a public park is usually grassland. However, in some cases, the types of land cover in a public park can include buildings, play grounds, paths and roads. The automatic detection of encroachment has to be able to distinguish five types land cover.

The employed method involves the use of the Gabor filter. In image processing, a Gabor filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination. As seen from (6.3), a 2-D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The Gabor filters are self-similar: all filters can be generated from one mother wavelet by dilation and rotation.

Let X_1 and X_2 with a size of $M \times N$ be processed by Gabor filter, we obtain by (6.5) the magnitudes of complex-valued subbands as,

$$\boldsymbol{G}_{\theta^t}^{X_1} = \left\{ m_{\theta^t}^{X_1}(m,n) | 1 \leq m \leq M, 1 \leq n \leq N, \text{and } t = [1,2,...,T] \right\}, \tag{7.1}$$

$$\boldsymbol{G}_{\theta^t}^{X_2} = \left\{ m_{\theta^t}^{X_2}(m,n) | 1 \leq m \leq M, 1 \leq n \leq N, \text{and } t = [1,2,...,T] \right\}. \tag{7.2}$$

The selection of appropriate parameters for Gabor filtering in literature has been evaluated in many ways. Turner (1986) evaluated the smoothing parameter approximately 2/3 improves the texture discrimination. Bianconi and Fernández (2007) tested the orientations of Gabor filter on 1280 pictures with textures on roads, trees and blocks, and explored that the best texture classification performance happened when divided the pictures into 16 non-overlapping sub-images. The same as in Turner (1986); Bianconi and Fernández (2007), we set the number of orientation T as 16 and the smoothing parameter S_x/S_y as 2/3 in our Gabor filter to magnify the texture information of X_1 and X_2.

7.3.2 Detection Model

As the result of above Gabor feature extraction, we obtain the feature set of X_1 and X_2 as $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$. To extract objects from X_1, we conduct k-mean clustering on $\boldsymbol{D}_1$ to discriminate components, group similar contents, and form a set of X_1 objects classified into five classes such as grass, trees, fences, houses and roads, respectively.

Given $\boldsymbol{D}_1 = \{\boldsymbol{D}_1(r) | 1 \leq r \leq R\}$, $\boldsymbol{C} = \{\boldsymbol{c_k} | 1 \leq k \leq K\}$ denotes the set of core vectors, initially, $k = 1$ and $\boldsymbol{c_k} = \boldsymbol{D}_1(1)$. Normally, clustering is to assign data that

are more "alike" into one cluster (Z. Huang, 1998). To determine the word "alike", a function for measuring similarity among data can be formulated,

$$\text{Similarity}(\boldsymbol{a}, \boldsymbol{b}) = \begin{cases} 0 & \text{if two vectors } \boldsymbol{a} \text{ and } \boldsymbol{b} \text{ are similar} \\ 1 & \text{otherwise.} \end{cases} \tag{7.3}$$

According to (7.3), we compare the similarity between each $\boldsymbol{D}_1(r)$ and $\boldsymbol{c}_k$, where $r = 1, 2, \ldots, R$ and $k = 1, 2, \ldots, K$. if $Similarity(\boldsymbol{D}_1(r), \boldsymbol{c}_k) = 0$, then assign $\boldsymbol{D}_1(r)$ to kth cluster, and we move $\boldsymbol{c}_k$ to $\frac{R_k \boldsymbol{c}_k + \boldsymbol{D}_1(r)}{R_k + 1}$, otherwise $Similarity(\boldsymbol{D}_1(r), \boldsymbol{c}_k) = 1$, assign $\boldsymbol{D}_1(r)$ to $(k+1)$th cluster, and let $\boldsymbol{c}_{k+1} = \boldsymbol{D}_1(r)$. Hereafter, let $r = 1$ we repeat this process until the last input vector $\boldsymbol{D}_1(R)$ is assigned into one cluster. At the end of clustering, the cluster labels $Y_1 = \{1, 2, ..., K\}$ and $\acute{\boldsymbol{D}}_1 = [\boldsymbol{D}_1, Y_1]$ are obtained.

Thus, the source land image is transformed into five object models $\boldsymbol{M}_1^{O_k}$, each stored as a 16×16 matrix, as shown in (6.7). To enhance the understanding of the objects, we conduct a second round incremental learning on each pixel as (6.9) to obtain an influence range $Range_1^{O_k}$ by (6.10). As detecting change against X_2, we conduct the one-step-more incremental learning on the target feature set $\boldsymbol{D}_2(r), r = 1, \ldots M \times N$ and obtain a tentative object model $\boldsymbol{M}_1''^{O_k}$. If $\text{eig}(\boldsymbol{M}_1''^{O_k}(r)) \in Range_1^{O_k}$, then there is no change, and vice versa.

7.4 Experimental Results and Discussion

In our experiments, four types land encroachments are investigated. They are sharing boundaries with commercial and/or industrial properties, sharing boundaries with residential properties, sharing boundaries with residential properties where the boundary is not clearly delineated physically on the ground, and sharing boundaries with residential properties where the boundary is obscured physically on the ground and/or from the air. (Dacey, Song, & Pang, 2013)

7.4.1 Data

The data used in the experiments is digital image data of New Zealand specifically the Auckland area. Auckland City Council covers the entire Auckland region in which there are over 800 designated public parks and reserves to examine for possible encroachment. The range in the parks' size is substantial, for example in the Manukau area the parks range in size from 38 square meters to 1995880 square meters. From a land management viewpoint the parks are divided into four categories:

1 Parks that share boundaries with commercial and/or industrial properties. The surrounding features of a park and the length of shared may be important

factors in the likelihood of encroachment occurring at a park. Commercial and industrial properties tend to be larger then residential so the shared boundary is longer. Teviot Reserve, on the North Shore, is an example of this type of park.

2 Parks that share boundaries with residential properties and are completely or almost completely isolated from features such as roads. The access to a park may be an important factor in the likelihood of encroachment occurring at a park. Isolated parks may not be visited frequently and this may encourage relatively small scale encroachment. Wyllie Park, in Manukau, is an example of this type of park.

3 Parks that share boundaries with residential properties where the boundary is not clearly delineated physically on the ground. There are parks where there are no natural (e.g. hedges) or man-made (e.g. fences) physical boundaries between the park and the residential properties. These parks may be more prone to encroachment. Diana reserve, on the North Shore, is an example of this type of park.

4 Parks that share boundaries with residential properties where the boundary is obscured physically on the ground and/or from the air. There are parks where the boundary is difficult to access physically from the park side and where the boundary is obscured visually in aerial photographs by dense vegetation. These parks may be more prone to encroachment due to difficulty in being able to confirm the physical location of the boundary. Also, dense vegetation may obscure temporary encroachment such as garbage dumping and camping. Auburn Reserve, on the North Shore, is an example of this type of park.

The data sets used for the detection of encroachment are digital images of areas that include public parks.

1 Aerial Photographs. The aerial photographs are captured on-line. The scale is set manually to 7.5cm/234m at time of capture and the ratio is one pixel to 0.5 meters. The images are in JPEG format and are collected from Google Earth.

2 Geo-Referenced Aerial Photographs. The aerial photographs have a set scale of 7cm/20m and are in the New Zealand Transverse Mercator 2000 projection. The aerial photographs are downloaded as raster data in JPEG format and are collected from the Auckland city Council via the GIS viewer (Council, 2010).

3 Land Boundary Data. The boundary data is used to verify that encroachment has been detected and to estimate the size of encroachment. When this research was started, there were seven Auckland regional councils and of those councils approached 2 of them, Manukau and North Shore supplied boundary data in GIS shape file format.

To demonstrate the automatic encroachment detection technique, for this chapter, four parks: Wyllie Park, Teviot Reserve, Auburn Reserve, and Diana Reserve are selected as representative examples of the four categories of park discussed above. Information about each park is given below:

1. Wyllie Park, Manukau, Auckland. Lat Long Location: $36^o59'17.95$" S, $174^o50'46.34$" E. The park is classified as a public open space and consists of general grassland with a path running across it from West to East. It is surrounded by residential and industrial properties. It has an area of 3956.00 square meters.

2. Auburn Reserve, North Shore, Auckland. Lat Long Location: $36^o47'23.08$" S $174^o46'03.42$" E. The park is classified as a community recreational area. It is surrounded by residential and commercial properties. Play centre and sports areas restricted access to Northern and eastern parts of the park. Lack of fences and accessibility problems made it difficult to determine South-Western boundary of the park. It has an area of 17577.102922 square meters.

3. Teviot Reserve, North Shore, Auckland. Lat Long Location: $36^o46'01.24$" S $1174^o43'26.01$" E. The park is classified as a neigbourhood reserve and consists of general grassland. It is surrounded by residential properties on all four sides. It has an area of 6252.332693 square meters.

4. Diana Reserve, North Shore, Auckland. Lat Long Location: $36^o46'39.92$" S $174^o43'48.05$" E. The park is classified as a neigbourhood reserve and consists of general grassland. It is surrounded by residential properties on three sides and a road on one side. It has an area of 6565.99622 square meters.

7.4.2 Experimental Setup

The experiments were carried in three phases. The first phase involves teaching the agent prior knowledge about the land cover patterns such as grass, buildings, trees, fences and roads, as they are most commonly found in and around public parks in Auckland. Note that, for this purpose, we collect data only from non-park areas for the training of the agent.

After the first phase learning, the agent now has the prior knowledge (i.e., the number of clusters k and initial mean vector of each cluster). In the second phase,

we set the equipped agent to perform spatial incremental learning on X_1 in order to obtain knowledge from the source image.

The third phase is change detection on X_2. We set the X_1 learned agent to repeat the incremental learning process on X_2 to detect change of X_2 against X_1.

7.4.3 Region of Interest Detection

Scenario 1 Scenario 2 Scenario 3 Scenario 4 Scenario 5

Figure 7.2: 5 scenarios to cover case study of changes on fences, house, trees, roads and park areas.

Following (Celik & Ma, 2010; Coppin et al., 2004; R. J. Radke et al., 2005; Bruzzone & Prieto, 2002), we conduct region of interest detection experiments to investigate the ability of the proposed algorithm on combating different scenarios. In doing that, we generate according to (6.15) total 200 pictures based on 4 scenarios that we select on purpose to cover the case study of different changes on fences, house, trees and park areas. Figure 7.2 gives the example images of these 4 scenarios. We conduct experiments for region of interest detection on those 200 pictures. The experimental results are summarized in Table 7.1, where the performance is evaluated on the 6 measurements discussed in previous section, and the results are formatted in the form of 'average accuracy $\pm$ standard division'. As seen, the accuracy varies over different scenarios. However, the robustness of the proposed algorithm is apparent in that both the average correct classification accuracy PCC and the false classification accuracy PFC are shown consistently above 95%.

The problem areas highlighted by the experiments are those were there are high concentrations of trees. The problem is that areas with trees in them have many

Table 7.1: ROI Detection of Agent-based Image Change Detection on Tasks with 5 Scenarios

Scenarios	P_{TP} (%)	P_{FP} (%)	P_{TN} (%)	P_{FN}(%)	P_{FC} (%)	P_{CC} (%)
Fences	1.5556 ± 1.0808	1.9444e-004 ± 7.1704e-004	98.2774 ± 1.1827	0.1667 ± 0.2393	0.1669 ± 0.2393	99.8331 ± 0.2393
House	3.0298 ± 2.5475	1.3333e-004 ± 2.8747e-004	96.9642 ± 2.5512	0.0058 ± 0.0176	0.0060 ± 0.0176	99.9940 ± 0.0176
Parks	3.9692 ± 2.6202	2.0556e-004 ± 5.5217e-004	96.0126 ± 2.6180	0.0179 ± 0.1226	0.0181 ± 0.1226	99.9819 ± 0.1226
Trees	4.1400 ± 2.7879	4.3889e-004 ± 7.1888e-004	95.6379 ± 2.9261	0.2216 ± 0.3586	0.2221 ± 0.3586	99.7779 ± 0.3586
Road	2.6118 ± 1.6043	1.6667e-005 ± 6.6638e-005	97.3857 ± 1.6067	0.0025 ± 0.0128	0.0025 ± 0.0128	99.9975 ± 0.0128

different types of colors and densities. True Positive - the land feature was correctly identified as belonging to a particular class (e.g. a road pixel was identified as a road). False Positive - the land feature was incorrectly identified as belonging to a particular class(e.g. a roof pixel was identified as a road). True Negative - the land feature was correctly identified as not belonging to a particular class False Negative - the land feature was incorrectly identified as not belonging to a particular class(e.g. a road pixel was not identified as a road class)

7.4.4 Land Cover Detection Experimental Results

Four specific experiments are carried out:

1 To detect permanent encroachment such as buildings (Permanent land cover encroachment).

2 To detect encroachment in the form of areas of a park fenced off (Permanent land use encroachment).

3 To detect temporary encroachment such as vegetable-growing (Temporary encroachment).

4 To detect possible areas of encroachment that may obscured by natural objects such as trees (Physical boundary encroachment).

Scenario I: Permanent land cover encroachment

Permanent land cover encroachment is where a building is constructed illegally on a public park. In land management terms this is the most serious form of encroachment. Although it is relatively simple to detect the presence of a building on a park, one difficulty in detecting permanent encroachment is the misclassification of legitimate constructions. One solution to this problem is to capture multiple images over an extended time-frame and make continuous comparisons. The type of structure built can range from a movable private structure (e.g. a shed), an immovable private

structure (e.g. a deck) through to commercial buildings and government building. We apply the above stated method to the land encroachment detection on the Wyl-

(a) park area (b) park with land encroachment (c) detected land encroachment

Figure 7.3: The scenario of land encroachment detection on Wyllie Park.

lie Park shown in Figure 7.3a, where a house encroachment is introduced near the top-left corner in Figure 7.3b. Afterwards, the resulted change image Figure 7.3c is produced through detection.

Scenario II: Permanent land use encroachment

We provide a permanent land use encroachment example where a fence is built illegally on a park, see Figure 7.3. In land management terms this is the another serious form of encroachment. This type of encroachment can be one of the most difficult to detect because the land cover change is usually not as noticeable as that of a building. This is especially true when the fence or enclosure structure is only a few pixels thick in the observed image. Contrariwise, when these enclosure boundaries increase in pixel width they become much easier to detect. Figure 7.4a present the

(a) park area (b) park with land encroachment (c) detected land encroachment

Figure 7.4: The scenario of land encroachment detection on Teviot Reserve.

scenario of land encroachment detection on the Teviot Reserve where fence element is introduced on the council boundary of this park, which is shown in Figure 7.4b. The resulted change image Figure 7.4c is produced through detection by proposed method.

Scenario III: Temporary encroachment

Temporary Use Encroachment, for example vegetable-growing, is not as serious as permanent encroachment as it is normally easier to resolve by the authorities. This type of encroachment can be difficult to detect especially when the colour of the vegetation is similar to the naturally occurring colours in the park. Another example of this type of encroachment is the detection of illegal garbage disposal on public land, this problem is usually easier to detect as the garbage colours usually differ in contrast and hue from the background. Temporary encroachment could become permanent if it is not dealt with by the authorities in a timely manner. Other types of temporary encroachment include organic dumping (e.g. grass cuttings) and inorganic dumping (e.g. furniture), grazing (e.g. horses) and intrinsic value (e.g. a fruit tree planted). In this scenario, data input, Figure 7.5a and Figure 7.5b show

(a) park area (b) park with land encroachment (c) detected land encroachment

Figure 7.5: The scenario of land encroachment detection on Auburn Reserve.

the Auburn Reserve with the council boundary and the ground truth image with an introduced vegetation element, respectively. Figure 7.5c is the resulted change image for this type of encroachment detection by the proposed method.

Scenario IV: Physical boundary encroachment

Physical boundary encroachment is a difficult problem to solve, especially when the boundary encroachment is concealed under a canopy of vegetation. However, not all vegetation canopies will hide illegal encroachment. For the fourth scenario, we

(a) park area (b) park with land encroachment (c) detected land encroachment

Figure 7.6: The scenario of land encroachment detection on Diana Reserve.

introduce land encroachment by plant elements in the Diana Reserve as seen in Figure 7.6a and Figure 7.6b. The difference extraction from proposed method is shown in Figure 7.6c. In many instances, this plant life will be introduced by the authorities, especially when it is in the confines of the park. However, the elements detected on park boundaries should indicate a possible encroachment, as these elements may have been illegally placed or provide canopies for illegal encroachment. These type of boundary triggers should warrant further investigation by the authorities.

7.5 Conclusions and Future Work

The investigation was conducted for land encroachment on 26 public parks in the area of Auckland, New Zealand. We restrict our focus on five objects including fences, house, parks, trees and road. The average correct classification accuracy of ROI detection produced by the proposed method is 99.91%. Then, the land encroachment is carried out successfully on four categories encroachment show the effectiveness of the method.

The main disadvantage of the method is that it cannot work as a stand-alone solution - ground survey data collection needs to take place to confirm the occurrence of encroachment. The advantages of the method are that permanent encroachment (such as buildings and fences) can be detected and temporary encroachment (such as vegetable-growing and garbage dumping) can also be detected. Another disadvantage of the approach is that it is not as reliable in detecting encroachment and/or changes in areas where there is a high concentration of tree as these area have a high variation in colors and concentration.

Chapter 8

Online Indoor Emission Sources Detection

Feature extraction is a key factor for indoor air quality (IAQ) studies. Unlike existing methods analyzing merely primary pollution, we consider alternatively the secondary pollution (i.e., chemical reactions between pollutants in addition to pollutant level), calculate in-between pollutants correlation coefficients for characterizing and distinguishing emission events. Extensive experiments show that our method works well on emission source data and improves the detection accuracy by a support vector machine compared to typical feature extraction methods such as PCA and LDA.

8.1 Introduction

Some studies of indoor air quality are motivated by a desire to understand the origins of risks to the health of householders, and the contribution of indoor emission sources relative to outdoor sources as both imply quite different intervention strategies. Common indoor sources of airborne particles include combustion sources (primarily heating and cooking) and tobacco smoking. Other sources include 'ornamental/amenity' combustion (candles, incense, etc), hygiene products (e.g. solvents, pesticides) and activities (e.g. dusting). Identifying the contribution of each source, and exposure to it, is central to the effort to understand health effects and manage risks. The magnitude, frequency and prevalence of these sources are strongly related to individual lifestyles and behaviours and their impacts are related to home volume, ventilation and use, i.e. also related to behaviour. Thus, there is huge potential for large variation in indoor emissions, air quality and exposures between homes, as well as between occupants. For this reason a technique was sought to identify and quantify indoor emission sources in a form that could be deployed rapidly with ease in multiple homes at low-cost.

There are a number of chemical substances of air that affects human health. Level of chemical substances such as carbon monoxide (CO), nitrogen oxides (NO and NO_2), sulphur dioxide (SO_2), ozone (O_3) and particulate matter (PM_{10} and $PM_{2.5}$) is increasingly important due to the harmful effects on human health. Euro-

pean Union and national environmental agencies have set standards and air quality guidelines for allowable levels of theses substances in the air. When the pollutant concentration levels exceed air quality guidelines, short-term and chronic human health problems may occur. Therefore, monitoring IAQ and detecting emission sources become an increasingly popular topic.

In general, there are two major pollution in indoor air quality analysis (Gurjar, Molina, & Ojha, 2010). Primary pollution is emitted directly into the atmosphere, such as carbon monoxide (CO) and carbon oxide (CO_2) gas from burning, or particulate matter (PM) released from household products. Level of pollutants can be easily detected by sensors, existing measurement studies focus on analyzing relationship between levels of pollutants and human health for people who suffer from chronic conditions (Amann et al., 2011; Heinrich, 2011; Bönisch et al., 2012). Secondary pollution results from chemical reactions between pollutants in the atmosphere. However sensors cannot detect the reactions between pollutants in atmosphere, because the continuous reactions are invisible knowledge such as the carbonation reaction in different temperatures (continuously change between CO_2 and CO) (W. Wang et al., 2010). To capture the reaction between pollutants is needed for emission sources detection.

So far, many pattern recognition models have been used to detect emission sources. Linear discriminant analysis (LDA) (Watson et al., 2011; Joly & Peuch, 2012), principle component analysis (PCA) (Wannaz, Carreras, Rodriguez, & Pignata, 2012; Pacyna et al., 2010), auto associative networks (ANN) (Ates, Erdinc, Uzunoglu, & Vural, 2010), genetic algorithms(GA) (Cuadros, Melo, Maciel Filho, & Wolf, 2012) as feature extraction methods are used to magnify the main orthogonal contributions which explain most of the pollutants for an emission source. We argue that relationship between pollutants may offer a useful knowledge for an important component of emission source detection. The ability to capture complex patterns that occur as interaction between pollutants can support emission source detection system.

Starting from observation of emission sources, we have developed a correlation coefficient based approach to support emission sources detection. This approach considers the relationship among pollutants comparisons with other existing data filtering method. We will show in this chapter that this approach outperforms the up-to-data feature extraction methods applied on emission source detection.

8.2 Related works

LDA, a supervised method, finds a linear combination of features which characterizes or separates two or more classes of objects or events (Fisher, 1936), which

benefits data classification. Joly et al.(Joly & Peuch, 2012) used LDA to analyze eight indicators (pollutants) to benefit rural and urban sites separation. A preliminary study, conducted by (Marié, Chaparro, Gogorza, Navas, & Sinito, 2010) focuses on analyzing magnetic carriers to detect emission sources from primary sources (vehicles), and on roads (paved area), road borders and surroundings areas. They found LDA is helpful to magnify magnetic carriers in pollution sources. LDA has also been used to support distinguishing two mountain valleys in Central Pyrenees as pollutant behavior analyzing method for evaluating the data (Blasco, Domeño, López, & Nerín, 2011). To determine local and regional sources of PM_10 and its geochemical composition in a coastal area, LDA is employed to test the extent differences in the PM_{10} levels and chemical profiles with varying atmospheric circulation and long-range transports were significant (Masiol et al., 2012).

PCA is a unsupervised method, which is used to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables by an orthogonal transformation (Jolliffe, 1986). Different to LDA, PCA extracts important features to support classifiers. PCA as a necessary and significant practical alternative methods is used to split a data set into a number of groups of observations which are distinct in terms of typical group values of corresponding pollutants (W.-Z. Lu, He, & yun Dong, 2011). To indicating the influence of sea salt in Tin-Jin, China, Kong et al.(Kong et al., 2010) applied PCA to analyze particulate matter emitted from coal combustion, marine aerosol, vehicular emission and soil dust. A research has been conducted to identify six major factors for observation of melting snow, PCA performs a significant benefit to their monitoring methods (J. Huang, Choi, Hopke, & Holsen, 2010).

Correlation coefficient is a measure of the relationship between two variables (W. Pearson, Davidson, & Britten, 1977). In computational chemistry study, Pearson correlation analysis can be used as tool to help the immunochemistry better understand the processing of antibody recognition of hapten molecules in competitive immunoassay (Z. Wang, Luo, Cheng, Zhang, & Shen, 2011). Pearson correlation is also used to analyze influence of streams on nearshore water chemistry (Makarewicz, Lewis, Boyer, & Edwards, 2012). Rasmussen's study (Rasmussen et al., 2012) proposed a global chemistry-climate models based on correlation coefficient to characterize the surface O_3 response to year-to-year fluctuations in weather. For food Chemistry, correlation coefficient is also benefit to analyze relationship between ORAC and Maillard reaction-like products (Brudzynski & Miotto, 2011).

The main defect of existing feature extraction methods for emission source detection is that they rely on internal data to capture the unique attributes of each entity, thus, they are not effective in discovering knowledge of interaction among pollutants. For a general emission source, the consistency ratio could be lower, but

correlation coefficient that captures the relationship among various pollutants levels as discussed above.

8.3 Methodology

The proposed method to define representative indoor events is based on the processing of air quality time series and consists in three steps: (i) selecting an appropriate continuously sliding window and fitting range of continuous sequence of air quality network data, (ii) removing of short term fluctuations associated with the influence of local emission sources from the original measurements taking into account the correlation determined from the correlation coefficient analysis, and (iii) mapping correlation factors into an non-linear space for emission sources recognitions.

Table 8.1: A Summarization of The Notation Used in This Chapter

$\boldsymbol{X} \in \mathcal{R}^{T \times M}$	The basic data structure which can dover a majority of the indoor conditions
t	The index of time point
T	Total elapse time
$\boldsymbol{x}(t) \in \mathcal{R}^{1 \times M}$	samples at time t
w	The size of sliding window
M	The total number of observed pollutants
m	The index of pollutant
$N = T - w + 1$	The total number of subsequence
i	The index of sliding window
$\boldsymbol{C}$	A set of matrices extracted by slicing time series with sliding window w
$\boldsymbol{c}_i \in \mathcal{R}^{w \times M}$	An extracted subsequence from the ith sliding window
$r_{i,j,k}$	The correlation coefficient of two pollutants (i.e., jth and kth) for the ith sliding window

8.3.1 Time Series Analysis

Given a time series data $\boldsymbol{X} = \{\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(t), \ldots, \boldsymbol{x}(T)\}$ involving emission L events, in which $\boldsymbol{x(t)} = \{a_1(t), a_2(t), \ldots, a_M(t)\}$ denotes a data sample consisting of M chemical contaminants and T represents elapsed time. A single point is no sense for emission sources detection, but a time series could be very long, sometimes containing millions of observations. To analyze relationship between pollutants, it is desirable to apply a sliding window w to $\boldsymbol{X}$ that will produce a sequence of time series. Figure 8.1 shows the procedure of sliding windows subsequence extraction with any of the real-valued representations. As a result, we store all extracted subsequences as $\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2 \ldots, \boldsymbol{c}_n\}$, where $n = T - w + 1$ is the number of subsequences

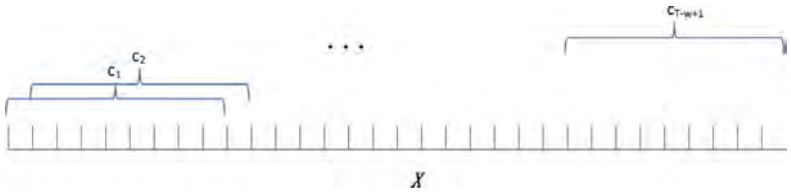


Figure 8.1: The procedure of subsequences extraction, where X is time series and c is slide window.

and $\boldsymbol{c_i} = \{\boldsymbol{x}[(t) : (t + w - 1)]\}$ represents a subsequence. Note that the corresponding label y_i is included in the calibration data according to the type of the emission event that subsequences involves.

8.3.2 Data Filtering

To remove short term fluctuations associated with the influence of local emission sources, correlation coefficient is included in this work to establish the relationships between pollutants for a fixed length time window c. Correlation is a measure of the strength of relationship between pollutants in a subsequence $\boldsymbol{c}_i$.

As $\boldsymbol{c}_i$ approximates the original time series with a combination of M pollutants in the i^{th} sliding window, we represent $\boldsymbol{c}_i$ as a combination of M column vectors ,

$$\boldsymbol{c_i} = \{\boldsymbol{a_{i,1}}, \boldsymbol{a_{i,2}}, \ldots, \boldsymbol{a_{i,M}}\}, \tag{8.1}$$

where $\boldsymbol{a_{i,m}} = \{\boldsymbol{a}_{i,m}(1), \boldsymbol{a}_{i,m}(2) \ldots, \boldsymbol{a}_{i,m}(w)\}$ represents the mth pollutant vector for the i^{th} sliding window.

For each sliding window i , we define correlation model as a covariance of every two pollutants,

$$r_{i,j,k} = r_{(\boldsymbol{a}_{i,j}, \boldsymbol{a}_{i,k})}, j \neq k, j, k = 1, 2 \ldots M, \tag{8.2}$$

in which the population correlation between two pollutants is calculated as:

$$\begin{aligned} r_{(\boldsymbol{a}_{i,j}, \boldsymbol{a}_{i,k})} &= \frac{Cov(\boldsymbol{a}_{i,j}, \boldsymbol{a}_{i,k})}{\sqrt{Var(\boldsymbol{a}_{i,j}) * Var(\boldsymbol{a}_{i,j})}} \\ &= \frac{E((\boldsymbol{a}_{i,j} - E(\boldsymbol{a}_{i,j}))(\boldsymbol{a}_{i,k} - E(\boldsymbol{a}_{i,k})))}{\sqrt{E(\boldsymbol{a}_{i,j} - E(\boldsymbol{a}_{i,j}))^2 E(\boldsymbol{a}_{i,k} - E(\boldsymbol{a}_{i,k}))^2}} \\ &= \frac{\sum_{t=1}^{w} (\boldsymbol{a}_{j,i}(t) - \bar{\boldsymbol{a}}_{j,i})(\boldsymbol{a}_{i,k}(t) - \bar{\boldsymbol{a}}_{i,k})}{\sqrt{\sum_{t=1}^{w} (\boldsymbol{a}_{i,j}(t) - \bar{\boldsymbol{a}}_{i,j})^2 \sum_{t=1}^{w} (\boldsymbol{a}_{i,k}(t) - \bar{\boldsymbol{a}}_{i,k})^2}}, \end{aligned} \tag{8.3}$$

where E is expected value. The correlation coefficient $r_{(\boldsymbol{a}_{i,j}, \boldsymbol{a}_{i,k})}$ is a number between -1 and 1. In general, the correlation expresses the degree that, on an average,

two pollutants change correspondingly. If one increases when the second one increases, then there is a positive correlation. In this case the correlation coefficient will be closer to 1. If one decreases when the other variable increases, then there is a negative correlation and the correlation coefficient will be closer to -1. As a result, we obtain a correlation efficient matrix $\boldsymbol{D} \in \mathcal{R}^{n \times \frac{M(M-1)}{2}}$ as,

$$\begin{aligned} \boldsymbol{D} &= \begin{Bmatrix} r_{1,1,2} & r_{1,1,3} & \cdots & r_{1,M,M-1} \\ r_{2,1,2} & r_{2,1,3} & \cdots & r_{2,M,M-1} \\ \vdots & \cdots & & \vdots \\ r_{n,1,2} & r_{n,1,3} & \cdots & r_{n,M,M-1} \end{Bmatrix}, \\ &= \{\boldsymbol{r}_i\}_{i=1}^{n} \end{aligned} \tag{8.4}$$

in which $\boldsymbol{r}_i \in \mathcal{R}^{1 \times \frac{M(M-1)}{2}}$ represents one data instance. The class label of the instance y_i is given according to events of emission sources for each slide window. Now, the problem of emission event detection is to seek an optimal solution to $f^* : \boldsymbol{r}_i \rightarrow y_i, i = 1, \ldots, n$

8.3.3 Incremental SVM Classification

Given training set $\boldsymbol{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^{N}$, with the label $y_i \in (-1, +1)$ indicating the class to which the vector $\boldsymbol{x}_i \in \mathcal{R}^d$ belongs. SVM seeks a linear separating hyperplane with a maximum-margin in the higher feature space. For linearly inseparable case, a non-linear kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j)|i \neq j, i, j = 1, 2, \ldots N$ is applied to transform the input space to a higher dimensional feature space so that the classes become linearly separable.

The normal form of SVM classifier is defined as follows:

$$f(x_i) = sgn(\boldsymbol{w} \cdot \phi(\boldsymbol{x}_i)) + b, \tag{8.5}$$

where "$\cdot$" means a dot product and $\phi(\boldsymbol{x}_i)$ refers to the kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j) \rangle$, which enables performing a linear classification in higher dimensional feature space. The corresponding decision making is then obtained by considering the sign of $(f(\boldsymbol{x}))$.

When new samples in the feature space violates the KKT conditions, incremental SVM updates the support vector set of a trained SVM classifier according to $\boldsymbol{w}$, and ensures that the decision is the same accuracy as the original SVM (Downs, Gates, & Masters, 2002).

Given a newly presented dataset $\boldsymbol{D}' = \{\boldsymbol{x}_i, y_i\}_{i=1}^{N'}$. By summarizing (8.5) over new coming samples and re-defining $\boldsymbol{w}$, we then update the SVM decision function

(8.5)

$$f(x_i) = sgn(\sum_{i=1}^{N+N'} (\hat{\boldsymbol{w}}_i y_i \cdot \phi(\boldsymbol{x}_i)) + b), \tag{8.6}$$

by maintaining the KKT conditions constant (Parrella, 2007) that defines the latest optimal solution with latest coming sample.

8.4 Experimental Results and Discussion

In this section, we conduct experiments where we apply the proposed correlation coefficient method on a number of available data and compare it with the conventional PCA and LDA methods in terms of classification performance. We use here a support vector machine for both feature extraction methods to find the class label of a test vector.

8.4.1 Data Set Description

In this study we use the NIWA developed PACMAN (Particles and Context Measurement Autonomous Node) device. The PACMAN instrument is able to record air quality as well as context information at 1Hz resolution. More detailed information can be found in (Olivares, Longley, & Coulson, n.d.). Table 8.2 shows the details of the parameters and sensors used in the PACMAN units.

Table 8.2: Summary of Measured Parameters and Sensors in The PACMAN Units

Parameter	Sensor	Notes
Carbon monoxide (CO)	Hanwei MQ-7	72hr warm up period 90s time resolution
Carbon dioxide (CO_2) Temperature (T)	Hanwei MG-811 LM335A	2hr warm up period Measuring temperature outside the enclosure
Movement (M) Distance (d)	PIR Maxbotix range finder	60^o field of view 6.5m range

Data were collected in a set of semi-controlled tests over several days in October 2012 where a single PACMAN unit was placed in the lounge of an unoccupied

house, as shown in Figure 8.2. Known particle emission activities were conducted and logged manually. The emission activities include:

1. Frying canola oil on electric hob.
2. Frying olive oil on an electric hob.
3. Frying olive oil on a gas hob.
4. Spray of household pesticide.
5. Lighting a cigarette and allowing it to smoulder.

The experimental protocol involved in general four stages: pre-activity sampling (baseline measurements), emission activity, emission activity halted and pollution allowed to mix in the indoor air, and venting of the house by opening external doors and windows and using a fan to aid indoor-outdoor air exchange.

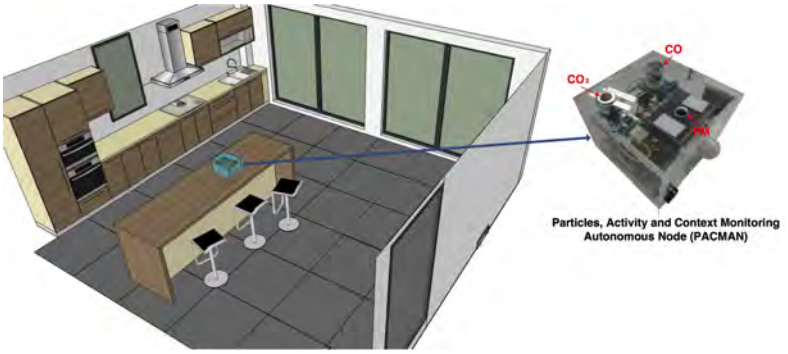


Figure 8.2: The experiment room.

As labeling data, an event is counted as samples in between the times when the first flag is set (i.e., pre-activity sampling stage) until the the emitting activity ended (i.e., emission activity halted). For the first case study, except the five events mentioned above, venting session and normal session (10 mins before each event and 10 mins after each event) are applied and labelled as class 6 and 7. For the second case study, all data is utilized and the training data is random selected from case study 1.

8.4.2 Data Quality Control

PACMAN was operated continuously during the tests logging data at 1 Hz. The data was checked for invalid instrument readings but was not calibrated, there-

fore the data presented here does not correspond to actual pollutant concentrations. As it was mentioned before, the data was analysed using a moving window approach. This poses a problem to the definition of the labels associated with the events/emission sources. Using the manual logbook as a base, the different emission activities were labelled on the records but for the sliding windows analysis, only windows that had more than 50% of their datapoints with a given label were considered as part of that event.

8.4.3 Extracted Knowledge from the Proposed Method

Figure 8.3 shows the differences between categorization by considering only pollutants level and that by involving the correlations between pollutants in this case. As seen from Figure 8.3a and 8.3b, before data filtering it is very hard to distinguish the two events "Frying canola oil, electric hob" and "Smoking". In contrast, the differences of these two events are magnified in Figure 8.3d and 8.3e in which we measure the events instead by the correlation between pollutants. Further, as observing the two events in three dimensional space as Figure 8.3c and 8.3f, we find that the capability of categorization based on interaction among pollutants are much better than that on the pollutants level.

8.4.4 Accuracy

To conduct the experimentation, original feature vectors are reformatted from a $M - dimension$ space to an $\frac{M(M-1)}{2} - dimensional$ space by correlation coefficient and $M - dimension$ space by PCA. The "classification accuracy in percentage" ($accuracy$) is reported for all the $\frac{M(M-1)}{2}$ dimensions from correlation coefficient and M dimensions from PCA. The larger is the is the classification accuracy, the better is the performance of the algorithm. The results are provided in Table 8.3 for both correlation coefficient and PCA method. We also report in this table results when no feature extraction method is used and SVM is applied in the original $M - dimensional$ space to get the classification performance. This result is reported in the last row of this table. We present the generalization performance averaged over 25 trials.

In Table 8.3, it is evident that the proposed correlation coefficient method is outperforming LDA and PCA technique in terms of the averaged classification accuracy. We also observe from this table that the proposed method is performing much better than the case when no feature extraction is done. Note that, the result of no feature extraction on original data is classified by the best optimized SVM.

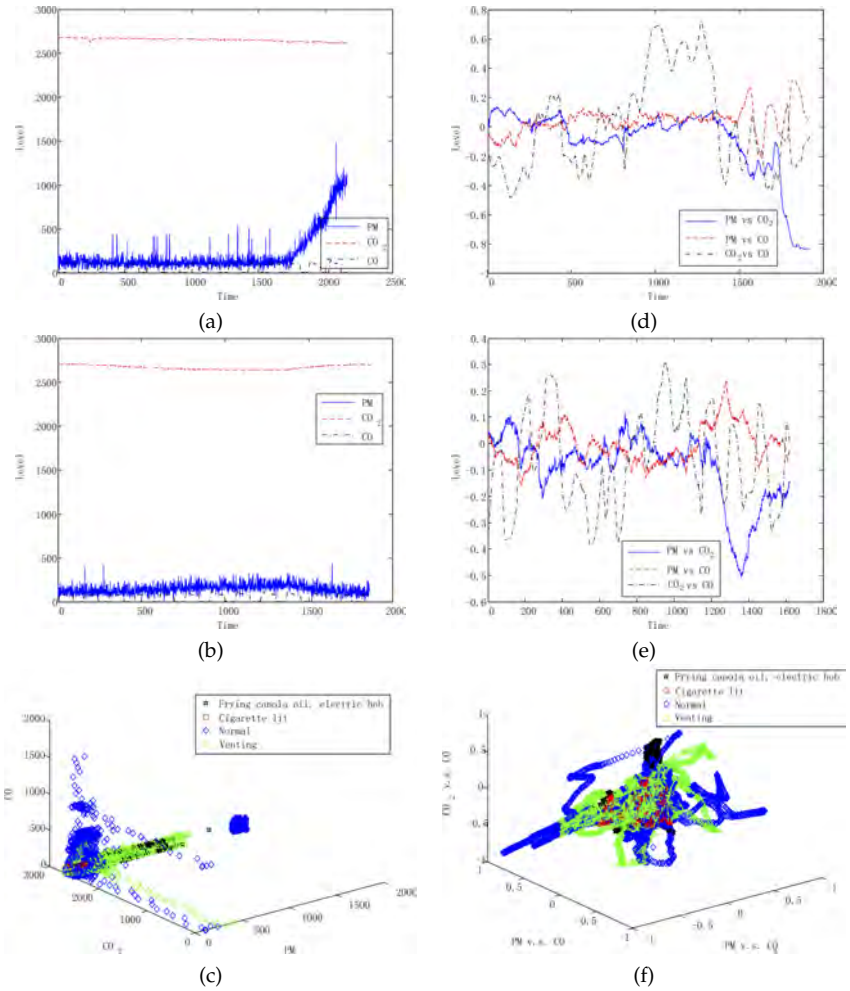


Figure 8.3: (a) Levels of organic contaminants obtained from PACMAN for event "Frying canola oil, electric hob"; (b) levels of organic contaminants obtained from PACMAN for event "Smoking"; (c) the native organization of organic contaminants for the two events with session "Venting" and "Normal" distributes on 3D space; (d) deterministic components obtained after data filtering for event "Frying canola oil, electric hob"; (e) deterministic components obtained after data filtering "Smoking"; (f) the filtered data for the two events with session "Venting" and "Normal" distributes on 3D space.

Table 8.3: A Comparison of Algorithms Using Classification Accuracy in Percentage as a Prototype

Correlation coefficient								
Accuracy	Confusion matrix							
80.14%	class	1	2	3	4	5	6	7
	1	**305.44**	9.72	0.20	0.44	3.84	11.48	37.52
	2	16.76	484.28	0.92	0.16	4.16	21.44	151.96
	3	0.20	1.32	387.12	1.36	0.24	2.60	10.36
	4	0.32	0	0.52	**125.16**	0.20	0.16	0.36
	5	8.36	14.08	0.84	0.52	188.08	11.12	88.04
	6	16.16	38.96	2.68	0.40	13.32	**773.24**	170.92
	7	25.36	57.92	6.24	0.36	26.60	62.44	**1882.20**

LDA								
Accuracy	Confusion matrix							
49.54%	class	1	2	3	4	5	6	7
	1	3.76	1.20	0	0	2.88	24.92	335.88
	2	2.96	304.44	0	12.20	1.40	13.16	345.52
	3	0.28	0	343.60	0	0	0.04	59.28
	4	0	0.76	0	17.24	0	0	108.72
	5	8.96	0.00	0	0	5.24	39.40	257.44
	6	27.96	16.88	25.08	0	23.52	273.44	648.80
	7	37.20	61.44	38.76	32.60	48.92	227.28	1614.92

PCA								
Accuracy	Confusion matrix							
29.49%	class	1	2	3	4	5	6	7
	1	155.44	19.68	0.24	2.52	65.76	29.36	95.64
	2	18.80	344.60	57.08	0.12	27.40	64.64	167.04
	3	1.36	150.72	91.52	0.12	26.48	46.68	86.32
	4	30.60	21.00	15.16	8.40	0	13.92	37.64
	5	26.00	19.52	4.36	0.04	**210.00**	7.00	44.12
	6	134.92	348.64	84.40	6.36	59.52	126.28	255.56
	7	388.96	295.96	282.04	11.68	233.52	259.52	589.44

Without data filtering								
Accuracy	Confusion matrix							
57.95%	class	1	2	3	4	5	6	7
	1	57.00	5.48	0	0	0.68	0.92	304.56
	2	1.08	**580.64**	0	0	0.84	36.24	60.88
	3	0	0	**403.20**	0	0	0	0
	4	5.08	0	0	100.16	0	0	21.48
	5	0	0.04	0	0	7.44	1.64	301.92
	6	31.24	75.88	2.52	0	37.84	381.40	486.80
	7	62.72	139.08	110.60	100.08	92.24	88.48	1467.92

For class 1 (Frying canola oil on electric hob), given that 25 times cross-validation showed averaged true positives is 305.44 produced by the proposed method, which is significantly higher than the 3.76, 155.44 and 57.00 produced by LDA, PCA and

the case without data filtering respectively. Similarly for class 4, 6 and 7, the proposed method has a clear superiority to the remaining three methods. A detail comparison on class truth positive accuracy is given in Figure 8.4. Although the proposed method is not giving the best performance for class 2, 3, and 5, its performance is seen very close to the best performed method.

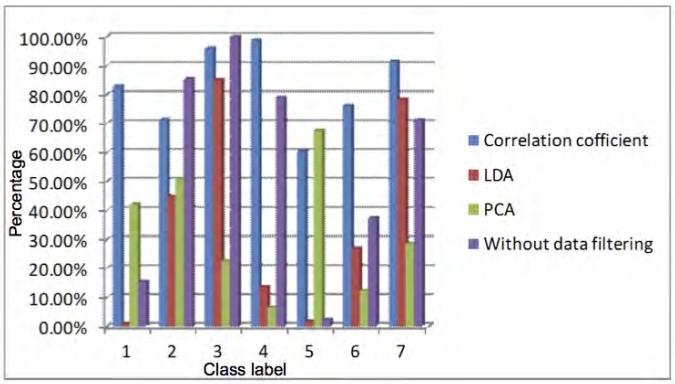


Figure 8.4: The comparison of truth positive accuracy for 7 emission events classification.

8.4.5 Sensitivity and Robustness

To verify the sensitivity and performance of the three feature extraction method, the accuracy and standard deviations under different sliding window sizes were analyzed. In the following, the results of each performance index will be demonstrated and discussed in detail.

Table 8.4 reveals the average accuracy, growth, and standard deviation of the proposed method in comparison to PCA, LDA and without data filtering under the condition of different sliding window sizes. The trend of these methods performances is shown in Figure 8.5. The lowest performance (19.32%) of the proposed method appears in the first sliding window (with $size = 30$). As larger sliding window sizes were imported into the system, the performances of the proposed method rises and reaches the highest accuracy (80.35%). Also the standard deviations of the proposed method are stead for different sliding window sizes. In addition, the experimental results of the proposed method has positive growth rate throughout the eight observed sliding windows. As illustrated in the experimental results, the pro-

Table 8.4: Performance of Target Feature Extraction Method in Different Size of Sliding Windows

Target Feature Extraction Method		The sliding window size							
		30	60	90	120	150	180	210	240
Correlation coefficient	average accuracy	19.32%	49.06%	61.50 %	68.36%	72.59%	78.19%	80.35%	80.14%
	growth	-	29.74%	12.44 %	6.86 %	4.23 %	5.60%	2.17%	-0.21 %
	stdev	0.0001	0.6185	0.7064	0.6170	0.4945	0.4358	0.4906	0.4958
LDA	average accuracy	46.06%	46.32 %	47.77%	46.81%	48.70%	49.32%	50.01%	49.54%
	growth	-	0.26%	1.45%	-0.96%	1.89%	0.63%	0.69%	-0.47%
	stdev	4.6216	2.9167	3.4096	4.3471	4.1156	3.6775	3.9270	4.7798
PCA	average accuracy	21.61%	26.44%	25.91%	29.82%	26.68%	27.63%	28.07%	29.49%
	growth	-	4.83%	-0.53%	3.92%	-3.15%	0.96%	0.44%	1.42%
	stdev	8.1703	5.4214	4.1153	5.1970	6.6487	4.3547	5.1603	5.8455
Raw data	average accuracy	51.74%	47.27%	47.17%	52.45%	54.69%	56.01%	58.47%	57.95%
	growth	-	-4.47%	-0.10 %	5.28%	2.24%	1.33%	2.45%	-0.52%
	stdev	4.5254	5.1515	4.8287	4.95	4.5594	4.3840	1.4620	1.0551

posed method consistently out performs the conventional PCA in all sliding windows. Among the small sliding window sizes, LDA and without data filtering provide better performances than proposed method. However, the proposed method surpasses LDA and without data filtering when the sliding window size reaches 60 seconds. Moreover, the proposed method shows good robustness properties, as measured by standard deviations.

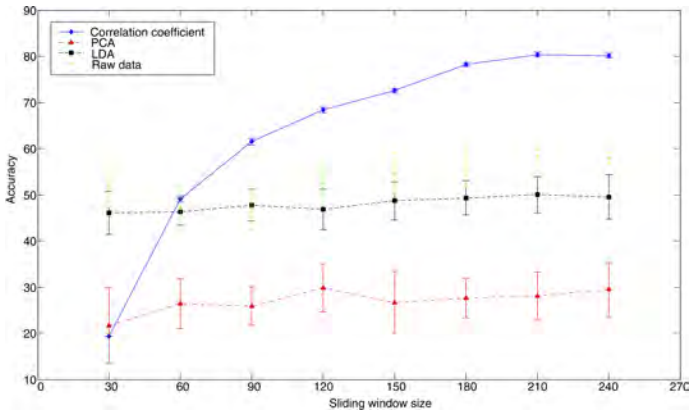


Figure 8.5: The trend of the performance of target feature extraction methods.

8.5 Conclusions

In this chapter, we address the problem of feature extraction based emission source detection, where the interactions among pollutants are not considered. Currently,

most approaches only focus on pollutant level analysis, which cannot capture the interaction among pollutants in emission events. Thus, we propose a novel approach correlation coefficient based emission source detection to address this problem. In particular, correlation coefficient assigns data attributes from levels of pollutants to relationships among pollutants so that the emission source detection model holds with knowledge of interaction among pollutants. Extensive experiments have confirmed the effectiveness and efficiency of our correlation coefficient based emission source detection under various experimental settings.

Chapter 9

Online Spatio-temporal $PM_{2.5}$ Prediction

Machine learning requires sufficient and reliable data to enhance its performance. Yet environmental data sometimes is short and/or contains missing data. Thus existing prediction models built on machine learning often fail to predict environmental problems accurately. Towards online spatio-temporal $PM_{2.5}$ prediction, this chapter investigates (1) how to use spatial domain data for the training of a temporal prediction model, (2) how to perform $PM_{2.5}$ spatial prediction for a city over its limited number of monitoring stations.

9.1 Introduction

Recent studies have shown that exposure to $PM_{2.5}$ can be more dangerous to human health than coarse particles, i.e. PM_{10}, because smaller particles can travel more deeply into our lungs and cause more harmful effects. Thus, the growing of long-term exposure to $PM_{2.5}$ is closely related to the increased mortality as well as diseases, such as lung cancer and cardiopulmonary disease (Pope et al., 2004).

In the literature, $PM_{2.5}$ prediction methods generally can be summarized into two categories: ratio prediction and tracer model. Ratio prediction calculates average ratios between concentrations of $PM_{2.5}$ and PM_{10} in a specific area. Tracer model applies mathematical models to compute the trend of $PM_{2.5}$ based on a set of inputs such as concentrations of SO_2, $VOCs$ and PM_{10}. In practice, a ratio prediction is not sustainable for long-term $PM_{2.5}$ forecasting, as $PM_{2.5}/PM_{10}$ ratios are not stable even for the same area and the same period of a year (Yuanxun, Yuanmao, Yingsong, & Delu, 2006). However, a tracer model gives often more accurate long term $PM_{2.5}$ prediction as the model uses multiple inputs rather than just the PM_{10} value for predictive analytics.

In New Zealand, PM_{10} is measured extensively but $PM_{2.5}$ data is very limited, however there is strong demand from the Ministry for the Environment and from Regional Councils for assessing levels of $PM_{2.5}$ around the country, but without incurring the considerable cost of duplicating the PM_{10} monitoring network with $PM_{2.5}$ monitors. Thus, an interesting research is to exploit the existing PM_{10} net-

work and Auckland's $PM_{2.5}$ database to determine if robust estimates of $PM_{2.5}$ can be derived from the PM_{10} data. More specifically, we intend to model daily average $PM_{2.5}$ prediction based on an hourly basis PM_{10} that has been collected by 13 monitoring stations in Auckland, New Zealand.

The challenge on $PM_{2.5}$ prediction modelling is, data from most monitoring stations consists of a number of short periods of time series, which are insufficient for training a model for long term $PM_{2.5}$ prediction; This happens because that even if urban areas have monitoring stations, data is not always available due to system maintenance and other incidental events. Therefore, an accurate estimate of $PM_{10}/PM_{2.5}$ during these periods is extremely difficult.

In a region that does not have homogeneous land use, the spatial position of data collection becomes important. This is because different areas in the region have characteristics that distinguish them from other areas. The data we use for $PM_{2.5}$ prediction has the geographical characteristics of urban residential, urban commercial, urban industrial, urban commercial/residential, semi-urban commercial, semi-urban residential, semi-urban commercial/residential and rural Park. Unlike traditional prediction models that focus on forecasting for a specific location, we develop in this chapter an new incremental support vector regression (SVR) considering geographical characteristics for spatio-temporal $PM_{2.5}$ prediction.

This chapter is organized as follows: In section 9.2, we present the brief concepts of existing prediction models and compare them from the literature. We propose spatial data aided support vector regression (SaIncSVR) in section 9.3.1 to achieve spatio-temporal and real-time $PM_{2.5}$ prediction, and we propose a spatial prediction in section 9.3.2. In section 9.4, we represent the experiments results with comparison to batch SVR learning. Finally, a conclusion is provided in Section 9.5.

9.2 Related Works

$PM_{2.5}$ prediction can be generally categorized as ratio prediction and pollutants tracer model. Ratio prediction analyses ratios between PM_{10} and $PM_{2.5}$ concentrations in a specific area (Hwa-Lung & Chih-Hsin, 2010; Hueglin et al., 2005; Chang & Lee, 2008). Pollutants tracer model allies mathematical models on fundamental principles to simulate the atmospheric chemistry and physics (Prybutok, Yi, & Mitchell, 2000; Lin & Cobourn, 2007; Appel, Gilliland, Sarwar, & Gilliam, 2007).

9.2.1 Ratio Prediction

The ratios of $PM_{2.5}/PM_{10}$ have been applied in many areas. C.-S. Li and Lin (2002) observes $PM_{2.5}/PM_{10}$ ratio in urban areas of Taipei, the ratio is about 0.68 in high

traffic area and 0.57 in downtown area. Cohen et al. (2002) studies $PM_{2.5}/PM_{10}$ ratios in six Asian regions (Australia, Hong Kong, Korea, Philippines, Vietnam and Japan), the ratio is about 0.45. A 2001 study shows strong correlation between characterizations of PM_{10} and $PM_{2.5}$ is studied in Italy and Swiss (Marcazzan, Vaccaro, Valli, & Vecchi, 2001; Gehrig & Buchmann, 2003). However, ratio prediction is limited in difference geographic areas (Hwa-Lung & Chih-Hsin, 2010). In addition, in practice, ratio prediction depend on data integrity. Missing data, which always occurs in real environmental data, affects the performance of ratio prediction.

9.2.2 Tracer Model

Mathematical models such as Backward-trajectory, Bayesian Maximum Entropy method (BME), Hidden Markov Models (HMM) and Neural Network predictions have been used for $PM_{2.5}$ prediction by many researches (S. Cheng et al., 2011; Nazelle, Arunachalam, & Serre, 2010; Dong et al., 2009; Ordieres, Vergara, Capuz, & Salazar, 2005).

Back trajectory analysis uses interpolated measured or modeled meteorological fields to estimate the most likely central path over geographical areas that provided air to a receptor at a given time (Salathé Jr & Hartmann, 1997). This method essentially traces a parcel of air backward in hourly stages during a specified period. Y. Song et al. (2006) investigates $PM_{2.5}$ concentrations in Beijing associated with two dust storm events in 2000, Back trajectory shows excellent agreement with a previous calculation using organic tracers in a chemical mass balance (CMB). model. A Canadian study in 2003 (Lee, Brook, Dabek-Zlotorzynska, & Mabury, 2003) measures $PM_{2.5}$ concentrations daily from Feb 2000 to Feb 2001 using back trajectory. Eight main source categories contributing to the $PM_{2.5}$ during the year-long period were evidently recognized. An ensemble air parcel back trajectories has been applied to analyze fine particulars in U.S. (Antony Chen, Doddridge, Dickerson, Chow, & Henry, 2002). The research represents the use of back trajectory model to forecast fine PM separately in winter and summer. However, backward trajectory is limited on analyzing short-term data as well as processing missing data (Thoss, Wang, & Miller, 2001).

BME is a method that computes generalized Bayesian probability values for expert systems by applying maximum entropy analysis (Christakos, 1990). In the field of $PM_{2.5}$ prediction, BME is used to optimize trend of $PM_{2.5}$ as a non-linear regression problem (Georgopoulos et al., 2005). As a probability time series prediction of $PM_{2.5}$ concentrations, a 2010 study (Nazelle et al., 2010) applies BME to form a prediction model for attainment demonstration in North Carolina. Other studies such as Cheon, Kim, Lee, and Lee (2009) and Georgopoulos et al. (2005) forecast $PM_{2.5}$ concentrations in Philadelphia using BME. Similar to backward trajectory model,

BME model is not effective when dealing with short-term data and missing data.

HMM is considered as a simple dynamic Bayesian network, which is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states (Baum & Petrie, 1966). HMM is modeled by a set of variables such as CO and NO_3 in the term of $PM_{2.5}$ prediction by calculating probability values generalized by HMM. Currently, HMM is the most popular forecasting model in many researches. Dong et al. (2009) represents the $PM_{2.5}$ concentration levels by factors temperature, pressure, cloudiness, wind speed, solar radiation, dewpoint and humidity in Cook County, Chicago. Schmidler et al. (2000) applies HMM to observe O_3, CO, $PM_{2.5}$ in different seasons for conducting prediction models based on same seasons concentrations data in previous years. McMillan, Holland, Morara, and Feng (2010) conducts many Markov chains to simulate trend of $PM_{2.5}$ by learning leveraging GSM such as total traveling distance, average speed differences, average speed, average traveling distance, number of unique cell IDs, number of cell ID changes and freeway annotation. Therefore, the performance of HMM forecasting depends on the information provided for prediction.

Neural Networks, often referred to as artificial neural networks, are an interconnected group of natural or artificial neurons that uses a mathematical or computational model for information processing based on a connectionistic approach to computation (McCulloch & Pitts, 1943). Neural Networks conduct prediction model of $PM_{2.5}$ from previous knowledge. An example of that, Ordieres et al. (2005) form a air quality research on analyzing and benchmarking a neural-network approach to the prediction of average $PM_{2.5}$ concentrations in El Paso (Texas) and Ciudad Juarez (Chihuahua). This research compares Multilayer Perceptron (MLP), Radial Basis Function (RBF) and Square Multilayer Perceptron (SMLP) based on input data including average levels of $PM_{2.5}$ during the first 8 h of the day, maximum level of PM2.5 during the first 8 h of the day, average temperature during the first 8 h of the day, average relative humidity during the first 8 h of the day, average wind speed during the first 8 h of the day, average wind bearing during the first 8 h of the day and wind direction index. The RBF shows up the best performance with the shortest training times. A Chilean study in 2000 (Pérez, Trier, & Reyes, 2000), applies multilayer neural networks, linear regression and persistence to predict $PM_{2.5}$ concentrations based on data obtained at an official monitor station in the city of Santiago, Chile. W. Lu, Fan, and Lo (2003) conduct a research for $PM_{2.5}$ prediction based on neural network method in downtown area of Hong Kong. Most of neural networks is time series prediction but spatial independent.

9.2.3 Methods for Spatio-temporal $PM_{2.5}$ Prediction

For spatio-temporal $PM_{2.5}$ prediction, Backward-trajectory uses the $PM_{2.5}/PM_{10}$ ratio as an input and provides a jump function to store the ratios change from one place to another (Hwa-Lung & Chih-Hsin, 2010). BME traces directly the $PM_{2.5}/PM_{10}$ ratios by Bayesian probability values for different places (Chang & Lee, 2008). Both Backward-trajectory and BME give only prediction on season trend, not the expected daily $PM_{2.5}$ value, because they use the $PM_{2.5}/PM_{10}$ ratio as the input to prediction modelling, while the $PM_{2.5}/PM_{10}$ ratios for different locations are required to be averaged over different seasons, since the PM concentrations vary seasonally depending on the nature of predominant emission sources and meteorological factors. HMM uses probabilities to derive the prediction, and simulate the regression function by Markov chains, which provides more accurate forecasting (Dong et al., 2009). In HMM models, the spatial characteristic information is merged into the input time series data to support calculating initial probability and transition probability (Dong et al., 2009).

Research on spatio-temporal prediction on a specific environmental problem is limited so far. Existing prediction models are often built on batch learning. A Backward-trajectory or BME model must be regularly re-trained to incorporate seasonal changes, because $PM_{2.5}/PM_{10}$ ratios can be calculated only after a season. The HMM models compute the probabilities of increase or decrease based on the temporal data (i.e., the probability of increase or decrease at certain time point) and spatial data (i.e., the probability of increase or decrease at certain location). However these approaches require to fix firstly the location before performing any temporal prediction, and as such, they are not truly spatio-temporal prediction model. Support vector regression (SVR) has been proposed as a good alternative to artificial neural networks in prediction, having a high generalization performance in time series modeling (de Brito & Oliveira, 2012). SVR has stronger generalization ability because the support vector machine theory is based on the minimization principle to structure risk (Wu, Xie, Song, & Hu, 2008).

9.3 Methods

The spatial element in this work comes from examining data collected from 13 different monitoring set. This work focuses on analysing the $PM_{2.5}$ concentrations according to the distance from each monitoring set to the city center. We assume that, the highest $PM_{2.5}$ concentration should show up in the city centre and then gradually decreases as the distances from the city centre to the monitoring set increases, because the city centre has more people, traffic and activities. The concentration is

expected to be at its lowest in the rural areas. Based on this assumption, the concentration of $PM_{2.5}$, in a region such as Auckland, should be a step distribution.

The research then applies incremental learning to consider the latest $PM_{2.5}$ concentration as incremental learning strategy has been proposed in this chapter to provide accurate and efficient $PM_{2.5}$ prediction. The proposed incremental learning model based environmental analysis has the following advantages on spatio-temporal environmental data:

1. significant benefits of spatial and temporal knowledge discovery,

2. mining a large amount of data efficiently with an updated model only without keeping all data,

3. capability of learning the latest knowledge from incoming data in both spatial and temporal domain.

In our work, we propose the support vector machine regression as such a learning model to predict daily $PM_{2.5}$ average based on a 24 hours basis PM_{10} concentrations in Auckland county with an incremental learning procedure. We focus on learning from both spatial and temporal domain, maintaining knowledge in support vectors and processing newest samples in time.

9.3.1 SaIncSVR: Spatial Data Aided $PM_{2.5}$ Prediction

Let $\boldsymbol{x}_t$ be a dynamics and multi-dimensional data. Suppose the present time point is t, then a prediction y_t is computed over the training data $\boldsymbol{X}_{t-1} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{t-1}\}$. Thus, the goal is to find a function $\hat{y}_t = f(\boldsymbol{x})$ that matches the actually obtained targets y_t of time instance t for all the training data.

As training $f(\boldsymbol{x})$, Support Vector Regression (SVR) according to (Parrella, 2007) solves the following optimization problem,

$$\begin{aligned} &\min_{w} \tfrac{1}{2}\boldsymbol{w}^T \cdot \boldsymbol{w} \\ s.t. &\begin{cases} y_i - (\boldsymbol{w}^T \cdot \phi(\boldsymbol{x}) + b) & \leq \epsilon \\ (\boldsymbol{w}^T \cdot \phi(\boldsymbol{x}) + b) - y_i & \leq \epsilon \end{cases}, \end{aligned} \tag{9.1}$$

where $\phi(\boldsymbol{x})$ is the kernel function, $\boldsymbol{w}$ is the margin and $\boldsymbol{x}_i$ is training sample and y_i is the training label.

As (9.1) is for classification, a boundary is set to test the tolerance on errors num-

ber, which is committed:

$$\begin{aligned} &\underset{\boldsymbol{w},b}{min}\ \tfrac{1}{2}\boldsymbol{w}^T\cdot\boldsymbol{w}C\textstyle\sum_{i=1}^{l}(\xi_i+\xi_i^*) \\ &s.t.\begin{cases} y_i-(\boldsymbol{w}^T\cdot\phi(\boldsymbol{x})+b) & \leq\epsilon\xi_i \\ (\boldsymbol{w}^T\cdot\phi(\boldsymbol{x})+b)-y_i & \leq\epsilon+\xi_i^* \\ \xi_i,\xi_i^*, i=1\ldots l \end{cases}, \end{aligned} \tag{9.2}$$

where ϵ is the lose function, ξ and ξ^* are slack variables which quantify the estimation errors greater than ϵ, penalty parameter C controls the norm of the weights $\boldsymbol{w}$.

As the result of (9.2), the regression model is formulated as a kernelled classification boundary,

$$f(\boldsymbol{x})\equiv\sum_{i=1}^{l}\theta\phi(\boldsymbol{x},\boldsymbol{x_i})+b. \tag{9.3}$$

The main drawback of batch SVR is that the model assumes the complete set of data is available for training, which is not suitable for updating itself when processing huge data streams. In addition, the model relies on reliable and sufficient data. Environmental data however sometimes is short and/or contains missing samples. For spatio-temporal $PM_{2.5}$ prediction, it requires the abilities of considering always the latest change in the air, and dealing with the short-term data and missing samples. Thus, we consider here the incremental learning of SVR.

Incremental Support Vector Regression

The incremental support vector regression (IncSVR) considers minimizing the optimization problem (9.2) and dual problem by maintaining Karush-Kuhn-Tucker (KKT) condition when a new sample is presented.

Let $h_i(\boldsymbol{x}_i)=f(\boldsymbol{x}_i)-y_i$ for all time instances $i=1,...,t$, according to (Shevade, Keerthi, Bhattacharyya, & Murthy, 2000), the KKT conditions define that each sample falls into one of the following three sets,

$$\begin{aligned} S= &\ \{i|(\theta_i\in[0,+C]\wedge h(\boldsymbol{x}_i)=-\epsilon)\vee \\ &\ (\theta_i\in[-C,0]\wedge h(\boldsymbol{x}_i)=+\epsilon)\}; \\ E= &\ \{i|(\theta_i=-C\wedge h(\boldsymbol{x}_i)\geq+\epsilon)\vee \\ &\ (\theta_i=+C\wedge h(\boldsymbol{x}_i)\leq-\epsilon)\}; \\ R= &\ \{i|\theta_i=0\wedge|h(\boldsymbol{x}_i)|\leq\epsilon\}, \end{aligned} \tag{9.4}$$

where S is the support set which contains the support vectors, i.e. the samples who contributes to the current solution; R represents remaining set, which contains the samples that do not contribute; and E denotes the error set which stores the remaining samples.

The training of IncSVR firstly classify all training points into three sets according to the above KKT conditions that define the latest optimal solution (Parrella, 2007). When a new sample $\boldsymbol{x}_t$ arrives, the IncSVR algorithm re-labels previous training set $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{t-1}$ in an iterative way till optimality is reached again. Similarly, the data used to train previous training models can be discarded, if they are not useful in machine updating. As a result, the updated regression function of form (9.3) is obtained in terms of $\boldsymbol{x}_t$, ϵ,C:

$$f'(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}_t, \epsilon, C). \tag{9.5}$$

Spatial Data Aided IncSVR (SaIncSVR)

Let $\{\boldsymbol{x}_{tj}\}$ be a spatio-temporal dynamics, where $\boldsymbol{x}$ is an hourly based PM_{10} concentration, $t = 1, \ldots T$ represents each day, and j is the number of monitoring station. In our application, input variables include a 28-dimensional state variable including 24 hours PM_{10} concentration, and 4-dimensional station spatial variables such as distance from this location to the closest beach in north, east and west, and altitude of this location,

$$[\boldsymbol{x}_t, \boldsymbol{s}_j] = [\boldsymbol{x}_{t_1}, \ldots, \boldsymbol{x}_{t_{24}}, \boldsymbol{s}_j], \tag{9.6}$$

where $\boldsymbol{s}_j = [\text{north}, \text{east}, \text{west}, \text{altitude}]$ represents the station spatial data, and $j = 1 \ldots M$ denotes the number of the station. We denote in the rest of chapter $\boldsymbol{x}_{tj}$ as the jth station PM_{10}, $\boldsymbol{X}_{(t-1)j}$ as the jth station historical PM_{10} data.

Let us suppose that we were able to select the parameter triple $(\boldsymbol{x}_{tj}, \epsilon, C)$ yielding a regression function $f_j(\boldsymbol{x}_t)$ of form (9.3) that truly describes the relationship between PM_{10}, station spatial data, and the target $PM_{2.5}$.

For spatio-temporal $PM_{2.5}$ prediction, a straightforward method is to perform local temporal prediction (LTP), which is to apply IncSVR to each local station. As a result, we obtain M local IncSVR models $f_j(\boldsymbol{x}_t), j = 1 \ldots M$ working individually for each station $PM_{2.5}$ prediction. Then, the spatio-temporal prediction model $\mathcal{F}_{st}$ is written as,

$$\mathcal{F}_{st} = \{\mathcal{F}_1(\boldsymbol{x}_t), \mathcal{F}_2(\boldsymbol{x}_t), \ldots, \mathcal{F}_M(\boldsymbol{x}_t)\}, \tag{9.7}$$

where $\mathcal{F}_j$ represents an IncSVR built on $\boldsymbol{X}_{(t-1)j}$. Because no noise from other stations is included in the modelling, is likely to be accurate under the condition of reliable and sufficient data provided for training.

However, environmental data contains lots of missing samples. To mitigate the loss, an alternative solution is to perform global temporal prediction (GTP), which is to build one IncSVR model for the prediction of all stations. The spatio-temporal prediction model $\mathcal{F}_{st}$ is given as,

$$\mathcal{F}_{st} = \mathcal{F}(\boldsymbol{x}_{tj}), j = 1, \ldots, M, \tag{9.8}$$

where $\mathcal{F}$ represents the IncSVR built on all stations historical data $\boldsymbol{X}_{t-1}$. Note that GTP assumes the contribution of experience from one station $PM_{2.5}$ prediction is positive to all other stations.

Further towards knowledge fusion across spatial and temporal dimensions, we consider utilizing spatial data to facilitate the learning of IncSVR, and propose spatial aided global temporal prediction (SaGTP) solution, which can be formulated as,

$$\mathcal{F}_{st} = \mathcal{F}([\boldsymbol{x}_{tj}, \boldsymbol{s}_j]), j = 1, \ldots, M, \tag{9.9}$$

where $\mathcal{F}$ represents the IncSVR built on all station historical data plus corresponding station spatial data $[\boldsymbol{X}_{(t-1)}, \boldsymbol{S}_j]$. Note that, SaGTP assumes that the geographical characteristic of a monitoring station has a positive impact on $PM_{2.5}$ temporal prediction.

9.3.2 A Solution to $PM_{2.5}$ Spatial Prediction

In this section, we study $PM_{2.5}$ spatial prediction for a city over its limited number of monitoring stations.

Without considering the impact of geographical features to pollution distribution, we assume that the pollution for a city follows a Gaussian distribution; the highest pollution appears in the center area of the city since more people, traffic and events means absolute more pollutant emission; the pollution decreases in outward area of the city.

In the example of Auckland city, the level of $PM_{2.5}$ appears the highest in the central area of the city, and decreases gradually over the distance to city center. Figure 9.1 shows the map of Auckland region, with the location of total 13 monitoring stations identified. We set station Newmarket (-36.8698, 174.7776) as the center, categorizing the whole Auckland area according to the distance of rest of monitoring station to the center station.

Given N stations, their averaged $PM_{2.5}$ concentrations, distance to city center $r(n)$, orientation $\theta(n)$. Here, $0 \leq n \leq N$ denotes the index of monitoring stations. According to the above pollution distribution assumption, we use a bessel function

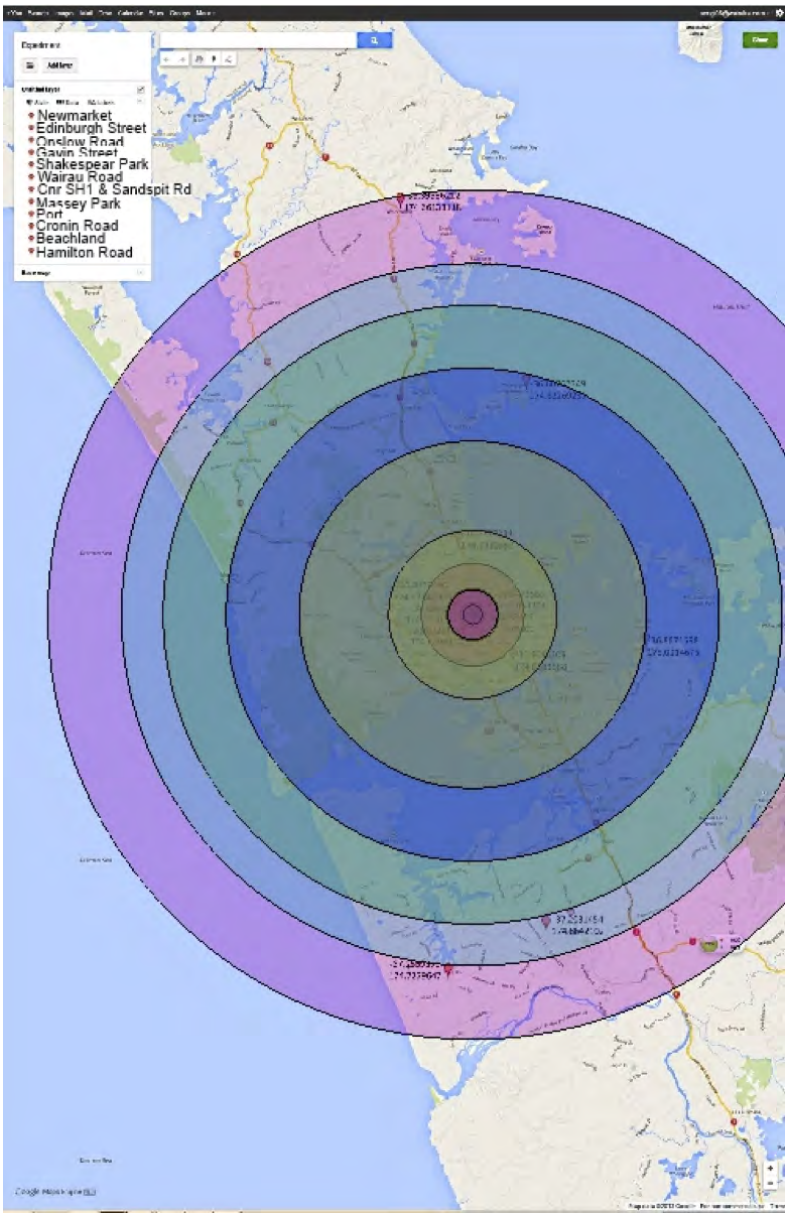


Figure 9.1: Spatial prediction assumption.

to approximate the trend variation across all stations as,

$$J_n(r(n), \theta(n)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n\tau - x\sin(\tau))}\, d\tau, \tag{9.10}$$

where τ is the gamma function, a shifted generalization of the factorial function to non-integer values.

Consider the decrease of $PM_{2.5}$ concentrations from city center to rural area, we manipulate the above bessel curvel to follow a typical Gaussian distribution by,

$$F(r,\theta) = F(r) = \xi\mathbb{G}(r) \cdot J_n(r). \tag{9.11}$$

where ξ the scale of the prediction, and $G(r)$ is the Gaussian distribution function,

$$G(r) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r-\mu)^2}{2\sigma^2}}, \tag{9.12}$$

in which $\mu = 0$. σ is the decreasing amplitude of the pollution, which is determined by training data.

To train the above prediction model, we define loss function when the difference from the actual value arises for each available monitoring stations

$$L = \frac{1}{T}\sum_{t}\sum_{r,\theta} \|F(r,\theta) - y(t,r,\theta)\|. \tag{9.13}$$

By using geometry, we find the distance between the gradient descent curve and the training values, so we want optimize σ, and ξ. Our optimization problem is:

$$F^* = \underset{\sigma,\xi}{\operatorname{argmin}} \frac{1}{T}\sum_{t}\sum_{n} \|F(r,\theta) - y(t,r,\theta)\|, \tag{9.14}$$

where F^* is the optimal spatial prediction function.

9.4 Experimental Results and Discussion

In this section, we experiment the proposed SVR model for daily $PM_{2.5}$ prediction using hourly PM_{10} data collected from 13 monitoring stations in Auckland. Specifically we conduct the following three case studies: 1) applying spatial data to facilitate $PM_{2.5}$ time series prediction, 2) applying the proposed SaIncSVR to spatio-temporal $PM_{2.5}$ prediction for 13 monitoring stations in Auckland, and 3) applying spatial prediction to spatio-temporal $PM_{2.5}$ prediction for 13 monitoring stations in Auckland.

9.4.1 Data

The data set was obtained from 13 sites in Auckland Council's air quality monitoring network by National Institute Of Water & Atmospheric Research (NIWA), New

Zealand. These stations are constantly monitoring air quality and provide daily averaged data in terms of PM_{10} and $PM_{2.5}$. The PM_{10} data were recorded hourly (24 values per day), for the period from 9^{th} August 2006 to 31^{st} December 2012. The $PM_{2.5}$ data were recorded as one value per day from 1^{st} January 1996 to 31^{st} December 2012 with the exception of 2010 (i.e., 2010 data is not available).

Table 9.1 gives the description of available PM_x datasets from 13 monitoring stations in Auckland region. As seen in the table, the size of dataset on each station has a big diversity. The smallest dataset is from Onslow Road (Partisol) that has only 117 samples; whereas the biggest dataset from Gavin Street Penrose has 1919 samples, which is almost 20 times of the smallest dataset. Note that among the 13 stations, 6 stations including Newmarket, Onslow Road (partisol), Onslow Road (BAM), Hamilton Road, Massey Park and Beachland, each has total less than one year historical data. In other words, those stations have less than four seasons data, which is insufficient for training a temporal prediction model without input from spatial domain. In addition, the missing data is another problem of these datasets. As seen, Newmarket has the biggest number of missing samples which is more than 3 times of that station data; and Hamilton Road has the smallest number of missing samples which is about 1% of the station data.

9.4.2 Experimental Setup

As predicting $PM_{2.5}$ for each day, we apply the trained model to all stations that have both PM_{10} and $PM_{2.5}$ data available. For each station, we compute the daily predicted $PM_{2.5}$ based on known PM_{10} data, evaluate the prediction performance by comparing the predicted $PM_{2.5}$ against the real value, and update the prediction model by conducting an incremental learning of the geographical SVR on the known $PM_{2.5}$ and PM_{10} data.

9.4.3 Performance Evaluation

The evaluation methods selected in this chapter include mean absolute error (MAE), mean bias error (MBE), root mean square error (RMSE), and index of agreement (IOA). In addition, the observation mean (obsMean) and model mean (modelMean) are also used to evaluate the similarity between the prediction given by model and real values. Note that not all stations have both PM_{10} and $PM_{2.5}$ data available for each day after 31^{st} December 2006, thus the prediction performance in our experiments is evaluated on the whole set of PM_{10} and $PM_{2.5}$ data collected from different stations and different days.

Table 9.1: Description of PM_x datasets from 13 Monitoring Stations in Auckland, New Zealand

Station	Samples	Time	Missing samples (%)	Location	To west (km)	To east (km)	To north (km)	Altitude
Newmarket	396	08/10/06 - 12/31/09	1239 (75.78%)	-36^o 52′ 10.36”, $+174^o$ 46′ 39.55”	7.53	9.75	2.98	58
Edinburgh Street	1239	07/03/07 - 12/17/12	421 (25.36%)	-36^o 55′ 35.43”, $+174^o$ 48′ 30.09”	22.90	41.47	16.29	68
Onslow Road (Partisol)	243	08/10/06 - 09/01/07	387 (61.43%)	-36^o 52′ 21.24”, $+174^o$ 44′ 58.12”	5.28	2.62	3.41	47
Gavin Street	1919	08/10/06 - 12/31/12	88 (4.38%)	-36^o 54′ 15.23”, $+174^o$ 49′ 5.70”	33.22	11.93	7.06	21
Shakespear Park	1323	04/12/08 - 12/31/12	67 (4.82%)	-36^o 36′ 28.00”, $+174^o$ 49′ 22.00”	2.32	0	0.88	13
Wairau Road	1627	06/16/07 - 12/31/12	54 (3.21%)	-36^o 46′ 16.83”, $+174^o$ 44′ 34.95”	30.77	3.12	12.95	8
Cnr SH1 & Sandspit Rd	556	04/20/07 - 11/20/08	40 (6.71%)	-36^o 54′ 34.19”, $+174^o$ 56′ 50.58”	22.29	6.62	24.01	48
Massey Park	290	02/21/09 - 12/31/09	31 (9.66%)	-37^o 3′ 49.18”, $+174^o$ 56′ 51.51”	3.91	32.05	19.65	12
Port	668	02/21/11 - 12/31/12	15 (2.20%)	-36^o 50′ 47.66”, $+174^o$ 47′ 7.13”	3.74	2.02	0	4
Cronin Road	615	04/19/08 - 12/31/09	13 (2.07%)	-37^o 12′ 10.57”, $+174^o$ 51′ 51.09”	23.14	41.70	16.40	68
Beachland	337	05/25/11 - 04/30/12	9 (2.60%)	-36^o 53′ 0.47”, $+175^o$ 0′ 14.28”	1.46	5.00	1.07	33
Onslow Road (BAM)	241	01/01/07 - 09/04/07	8 (3.21%)	-36^o 52′ 21.24”, $+174^o$ 44′ 58.12”	5.28	2.62	3.41	47
Hamilton Road	315	02/18/09 - 12/31/09	3 (0.94%)	-36^o 47′ 32.84”, $+175^o$ 1′ 36.01”	0.89	11.86	1.35	6

9.4.4 Applying Spatial Data in Temporal Prediction

In the first case study, we test how a temporal prediction by SVR can be facilitated by spatial data using the proposed model. For the purpose of comparison, we perform daily $PM_{2.5}$ prediction by the following three methods,

(a) LTP. The prediction is conducted locally on each station.

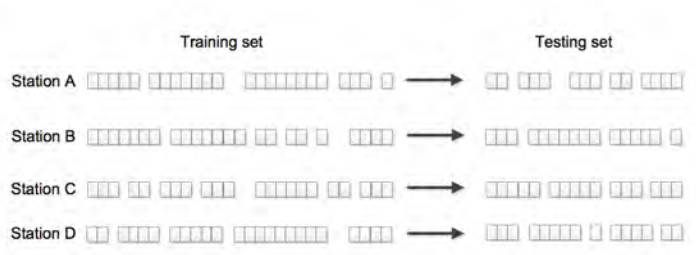
(b) GTP. The prediction is conducted using a collection of data from all observed stations for training.

(c) SaGTP. The prediction is conducted using a collection of data from all observed stations plus corresponding station spatial data for training.

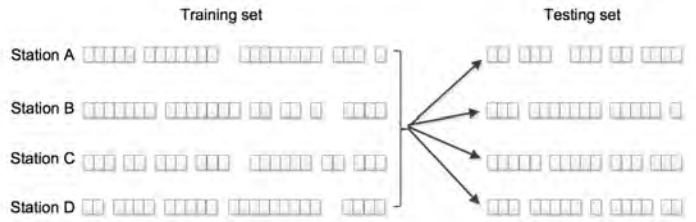
Figure 9.2 illustrates the three daily $PM_{2.5}$ prediction methods, where streams of blocks represent the time series data collected from different monitoring stations, and where those streams are shown all discontinued, indicating that each station has more or less missing samples.

For each prediction experiment, we use data before 31^{st} December 2009 for training, and data after 1^{st} January 2011 for testing. We observe specifically the 4 stations including Gavin Street, Wairau Road, Edinbourgh Street and Shakespear Park in that the data collected at those stations are insufficient to build a standard temporal prediction model.

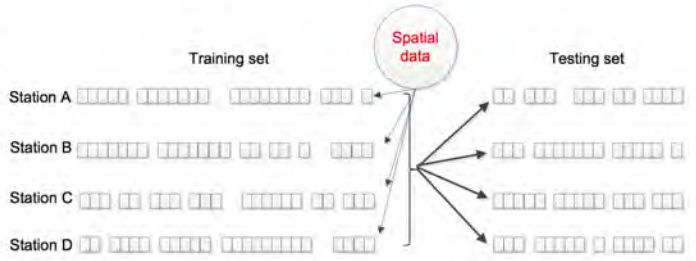
Table 9.2 records the station daily $PM_{2.5}$ forecasting accuracies produced respectively by LTP, GTP and SaGTP for the period from January 2011 to December 2012. As seen from the table, LTP clearly outperforms GTP and SaGTP for total 20 performance observations (on 4 stations and 5 measurements). This is understandable because no noise information from other stations is included in the prediction modelling, LTP gives normally the best accuracy. Second to LTP, SaGTP wins for 6 observations (4 on Wairau Road, and 2 on Shakespear Park), this is surprising because SaGTP uses not only all 13 stations historical data, but also spatial data for the training of the prediction model. Obviously, the inclusion of spatial data does not destroy the prediction model, but facilitates the learning of SVR. To further investigate the behavior of SaGTP, for each station we drill into some short periods that have missing samples, and we calculate the differences of LTP to SaGTP error and GTP to SaGTP error, respectively.



(a)



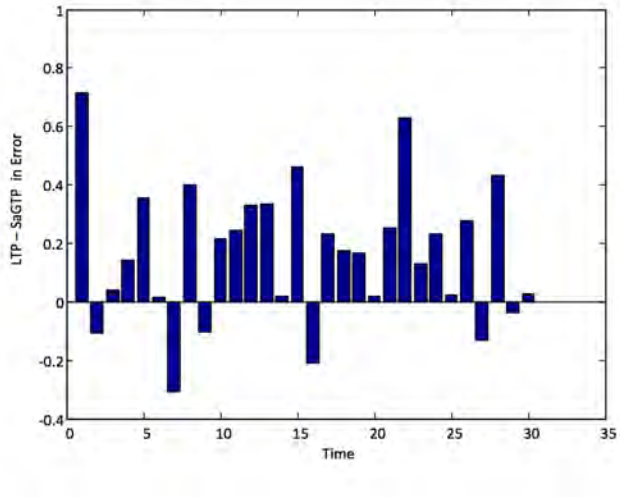
(b)



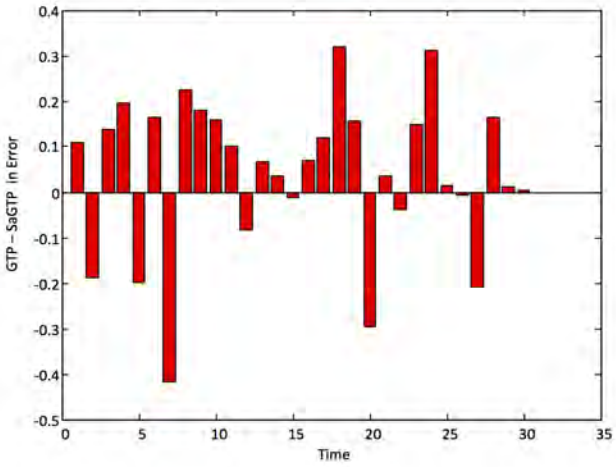
(c)

Figure 9.2: Three daily $PM_{2.5}$ prediction methods (a) local temporal prediction (LTP), (b) global temporal prediction (GTP), and (c) spatial data aided global prediction (SaGTP).

Figure 9.3 to 9.6 presents the comparison results for the four stations. For Gavin Street, within the period of 30 days from 2^{nd} January to 1^{st} February 2012, there is one day (9^{th} January) data missing, SaGTP is shown in Figure 9.3 outperforming



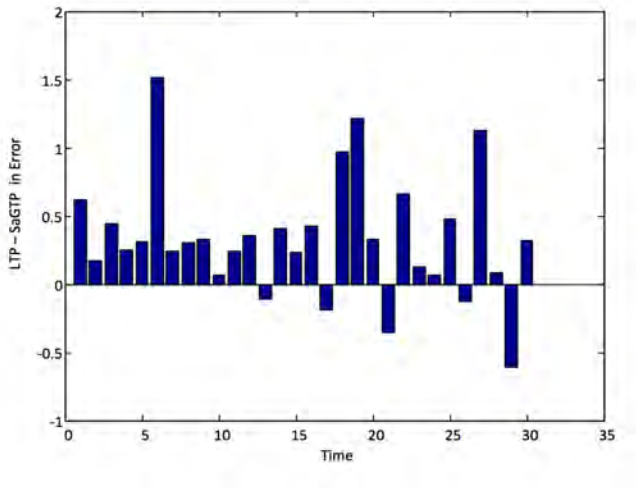
(a) LTP versus SaGTP



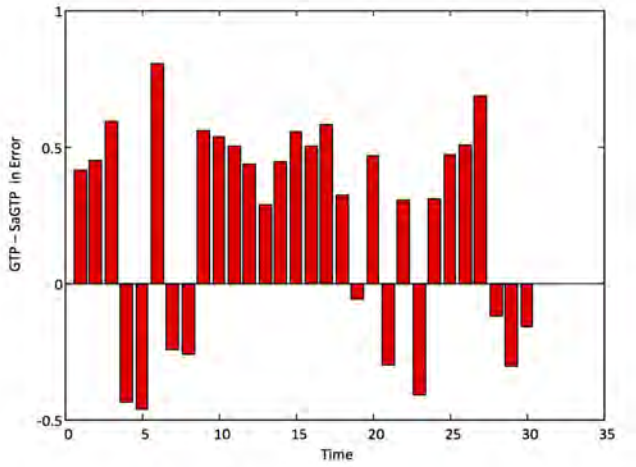
(b) GTP versus SaGTP

Figure 9.3: The difference between SaGTP and other two methods in prediction error for the period from 2^{nd} January, 2012 to 1^{st} February, 2012 on station Gavin Street.

LTP for 25 days and GTP for 22days. For Wairau Road station, there is also one day (6^{th} December) data missing within the period of 30 days from 1^{st} December to 31^{st}



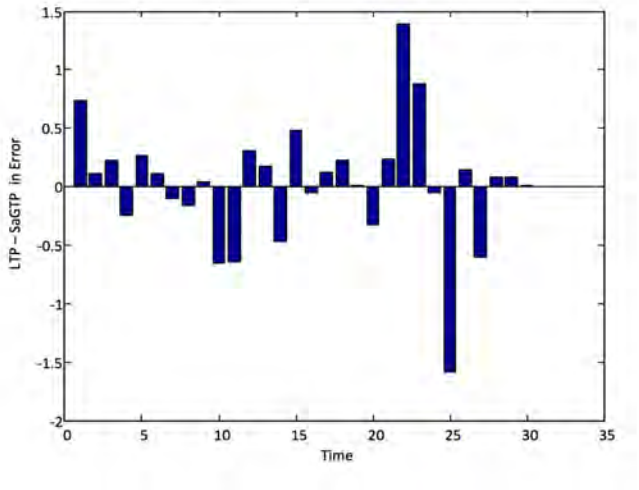
(a) LTP versus SaGTP



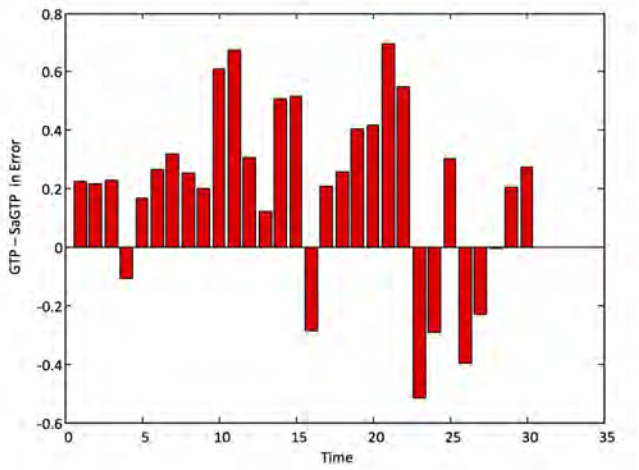
(b) GTP versus SaGTP

Figure 9.4: The difference between SaGTP and other two methods in prediction error for the period from 2^{nd} January, 2011 to 31^{st} January, 2012 on station Wairau Road.

December 2012. SaGTP however as reported in Figure 9.4 wins LTP for 23 days and GTP for 22 days of the 30 days. The same for Edinburgh Road station, although data



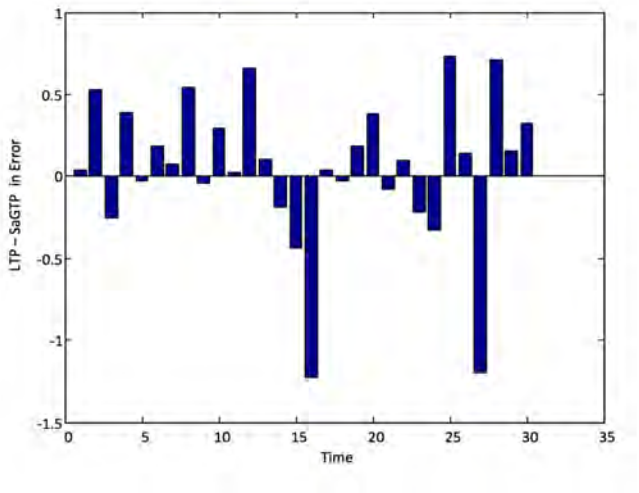
(a) LTP versus SaGTP



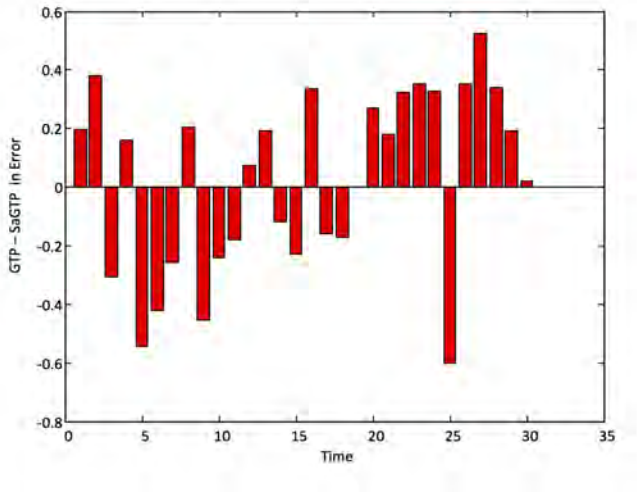
(b) GTP versus SaGTP

Figure 9.5: The difference between SaGTP and other two methods in prediction error for the period from 18^{th} March, 2011 to 17^{th} April, 2011 on station Edinburgh Street.

of 24^{th} March is absent. SaGTP by Figure 9.5 is better than LTP for 20 days and GTP for 24 days within the period of 30 days between 18^{th} March and 17^{th} April 2011.



(a) LTP versus SaGTP



(b) GTP versus SaGTP

Figure 9.6: The difference between SaGTP and other two methods in prediction error for the period from 5^{th} April, 2011 to 6^{th} May, 2011 on station Shakespear Park.

Table 9.2: Results of LTP, GTP and SaGTP Daily $PM_{2.5}$ Prediction for Station Gavin Street, Wairau Road, Edinbourgh Street and Shakespear Park. The Period of Prediction is from January, 2011 to December, 2012

Station	Number of missing samples	Prediction Method	Evaluation					
			$obsMean$	$modelMean$	$RMSE$	MAE	IOA	MBE
Gavin Street	54	LTP	6.5543	**6.1807**	**1.5904**	**1.2310**	**0.9206**	**-0.3736**
		GTP	6.5543	6.0651	1.6290	1.2497	0.9145	-0.4892
		SaGTP	6.5543	6.1528	1.6234	1.2311	0.9086	-0.4015
Wairau Road	20	LTP	6.8552	6.3713	2.1770	1.4401	**0.9007**	-0.4838
		GTP	6.8552	6.0164	2.0978	1.4680	0.8956	-0.8387
		SaGTP	6.8552	**6.3990**	**2.0167**	**1.3781**	0.8951	**-0.4562**
Edinburgh Street	20	LTP	4.2153	3.8831	**1.2315**	**0.8609**	0.8482	-0.3322
		GTP	4.2153	**4.2386**	1.3049	0.9418	**0.8698**	**0.0233**
		SaGTP	4.2153	3.9420	1.3022	0.9174	0.8608	-0.2733
Shakespear Park	16	LTP	4.1773	4.0499	**0.9674**	**0.6934**	**0.8745**	-0.1273
		GTP	4.1773	4.3330	1.1766	0.8706	0.8546	0.1557
		SaGTP	4.1773	**4.0539**	1.1192	0.8250	0.8567	**-0.1234**

LTP for 21 days and GTP 18 days of the 31 days, as recorded by Figure 9.6.

In summary, the geographical characteristics have been applied in this experiment and compared to results obtained without applying spatial information. The results from using spatial information to aid time series prediction are close to prediction results on the time series prediction only and are better than those without spatial information in the historical data. It is apparent that in static situations, the SaGTP efficiently predict environmental problems when missing samples appear, which indicate that spatial data is useful in facilitating the learning of SVR for $PM_{2.5}$ prediction.

9.4.5 Spatio-temporal $PM_{2.5}$ Prediction

In the second case study, we conduct spatio-temporal $PM_{2.5}$ prediction experiments, applying the proposed SaIncSVR to predict concentrations of $PM_{2.5}$ for all 13 monitoring stations in Auckland. The experiment is conducted in the setting of GTP.

In this experiments, we select first the stations that have data available from 9^{th} August 2006 to 31^{st} December 2006, and use the selected 3 stations (including Newmarket, Onslow Road (partisol) and Gavin Street) data to initially train the IncSVR model.

Table 9.3 gives the spatio-temporal $PM_{2.5}$ prediction of 13 monitoring stations for the period from 1^{st} January 2007 to 31^{st} December 2009 and 1^{st} January 2011 to 31^{st} December 2012. As seen from the table, the proposed SaIncSVR presents better $PM_{2.5}$ prediction than the IncSVR for 8 of total 13 stations, which includes Newmarket, Onslow Road (partisol), Onslow Road (BAM), Cnr SH1 & Sandspit Rd, Wairau Road, Edinburgh Street, Shakespear Park, and Cronin Road. For Gavin Street, the SaIncSVR wins still on 4 of total 5 performance observations. These 9 stations covers top 7 stations with the biggest number of missing samples, which indicates that the proposed SaIncSVR is capable of mitigating loss due to a big number of missing samples incurred. However for station Hamilton Road, Massey Park, Port of Auckland and Beachland, which incurs much less missing samples than the top 7 stations. SaIncSVR prediction is surprisingly not as good as that of the IncSVR, which implies that spatial dimension data for those stations cannot be exploited by IncSVR, thus is not facilitating the learning of IncSVR.

9.4.6 Spatial $PM_{2.5}$ Prediction

In the third case study, we test how $PM_{2.5}$ concentrations can be traced in a city by the proposed spatial prediction model. For the same reason of missing data, we use in this experiment data before 31^{st} December 2011 for training, and data after

Table 9.3: SaIncSVR Versus IncSVR for Spatio-temporal $PM_{2.5}$ prediction on 13 monitoring stations in Auckland New Zealand.

Station	Number of missing data	SaIncSVR						IncSVR					
		$obsMean$	$modelMean$	$RMSE$	MAE	IOA	MBE	$obsMean$	$modelMean$	$RMSE$	MAE	IOA	MBE
Newmarket	1095	9.1367	**8.8068**	**3.7580**	**2.5658**	**0.6536**	**-0.3299**	9.1367	8.5672	3.8725	2.6898	0.6541	-0.5695
Edinburgh Street	421	5.0353	**5.1943**	**1.3477**	**0.9913**	**0.9333**	**0.1590**	5.0353	5.4596	1.4496	1.0909	0.9231	0.4243
Onslow Road (Partisol)	74	7.1320	**6.2985**	**2.6738**	**1.5856**	**0.8184**	**-0.8335**	7.1320	6.0996	2.9055	1.7250	0.7761	-1.0324
Gavin Street	85	7.0132	6.5993	**1.7302**	**1.2620**	**0.9151**	-0.4139	7.0132	**6.7698**	1.7488	1.2778	0.9127	**-0.2434**
Shakespear Park	67	4.1056	**4.2886**	**1.0775**	**0.8106**	**0.8668**	**0.1830**	4.1056	4.4880	1.1564	0.8814	0.8515	0.3824
Wairau Road	54	7.0771	**6.8255**	**1.9112**	**1.2977**	**0.9147**	**-0.2516**	7.0771	6.6182	2.0123	1.3586	0.9073	-0.4589
Cnr SH1 & Sandspit Rd	40	8.1439	**8.2783**	**1.6998**	**1.2411**	**0.8629**	**0.1344**	8.1439	6.8175	2.2154	1.6830	0.7879	-1.3264
Massey Park	31	6.9207	5.6543	2.7257	1.8792	0.8763	-1.2664	6.9207	**6.3212**	**2.4119**	**1.5730**	**0.8985**	**-0.5995**
Port	15	6.7365	5.7897	1.9070	1.3232	0.8139	-0.9469	6.7365	**6.0676**	**1.7775**	**1.2135**	**0.8377**	**-0.6690**
Cronin Road	13	4.2898	**4.3380**	**1.6925**	**0.9827**	**0.8363**	**0.0482**	4.2898	4.7284	1.7946	1.1331	0.8191	0.4386
Beachland	9	4.7418	4.6981	1.3298	1.0164	0.8648	-0.0437	4.7418	**4.7457**	**1.3416**	**1.0327**	**0.8662**	**0.0039**
Onslow Road (BAM)	8	6.6091	**6.2043**	**1.3991**	**0.9799**	**0.9328**	**-0.4049**	6.6091	6.0198	1.6413	1.0948	0.9036	-0.5894
Hamilton Road	3	4.4530	3.8427	1.4808	1.0871	0.8271	-0.6103	4.4530	**4.7256**	**1.3488**	**0.9946**	**0.8439**	**0.2726**

1^{st} January 2012 for testing. We observe specifically the 6 stations including Gavin Street, Wairau Road, Edinbourgh Street and Shakespear Park, Port and Beachland.

Different to other prediction models, the spatial prediction model is trained on averaged concentrations of $PM_{2.5}$ for all stations,

We apply the spatial prediction model described in Section 9.3.2. As the result of training, we obtain a daily $PM_{2.5}$ spatial prediction model for Auckland city. Figure 9.7 discloses the prediction model in 2D and 3D spaces, respectively.

Table 9.4 records the station daily $PM_{2.5}$ forecasting accuracies produced by the proposed spatial prediction for the period from January 2012 to December 2012. For each station in observation, we we evaluate the difference between observation mean and model mean by ($|obsMean - modelMean|$ / $obsMean$).

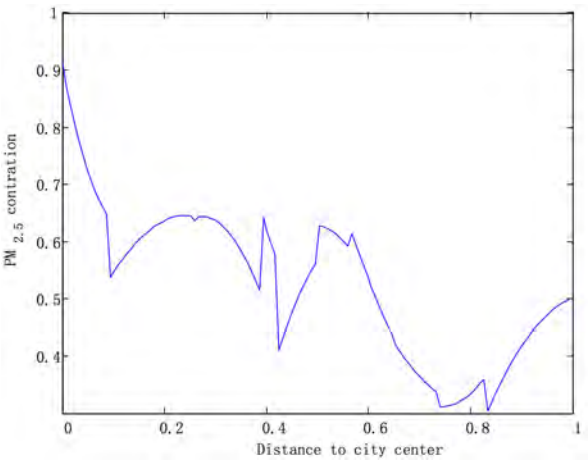
As we can see from the table, the difference rate for station Gavin Street is 19.74% and Wairau Road is 21.77%, which shows the $PM_{2.5}$ concentrations for these two stations are successfully predicted. However for the other 4 stations, the rate all goes above 55.71%. This indicates that the proposed method is working merely for a certain region, however unable to maintain the integrity of the distribution of $PM_{2.5}$ for the whole city.

This reasons that we can explore on the explanation for the fair performance of the proposed spatial prediction model include, (1) training data is not sufficient, as data collection for most stations are not continuous, which causes data is missing for a big number of days. The status of missing data for each station can be found in Table 9.1; (2) In addition, notably, there are anomalies in the data that appear to be dependent of the geographical feature of monitoring station. For example in Station Auckland Port and Beachland, monitoring takes place inside a park very close to the center monitoring station, the reading from these two stations should be as high as that of center station, however often shows unexpectedly low; and in station Cnr SH1 & Sandspit Rd, monitoring takes place next to a busy highway outside the city centre, the concentration of $PM_{2.5}$ is highly impacted by the motorway traffics and turns to be regionally very high.

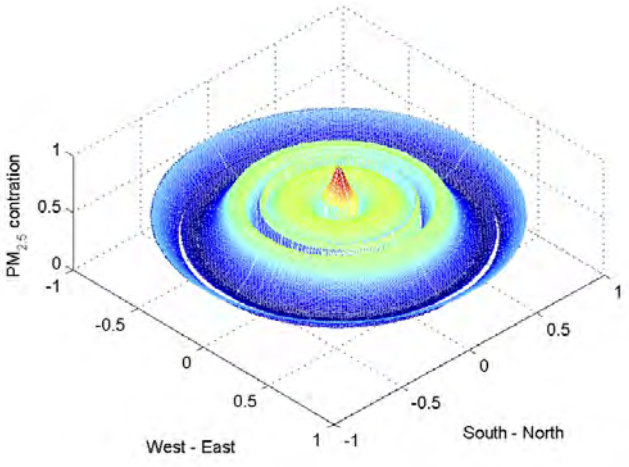
9.5 Conclusion

This chapter addresses the problem of spatio-temporal prediction of $PM_{2.5}$ by PM_{10}, under the condition of serious missing data with the historical data from the monitoring stations.

We derive computational models for (1) global temporal prediction, conduct spa-



(a)



(b)

Figure 9.7: $PM_{2.5}$ spatial prediction model illustration in (a) 2D and (b) 3D spaces.

tial domain knowledge fusion and build one global prediction model over all stations data; (2) spatial data aided prediction, conduct spatio-temporal cross dimensional knowledge fusion and use alternative spatial data to facilitate the training of a prediction model; and (3) $PM_{2.5}$ spatial prediction, build $PM_{2.5}$ prediction model over a limited number of monitoring stations.

Table 9.4: Spatial Prediction on 6 Monitoring Stations in Auckland New Zealand from 1^{st} January 2012 to 31^{st} December 2012

Station	$obsMean$	$modelMean$	$Difference$
Gavin Street	6.5634	7.8589	1.2954
Wairau Road	6.7556	5.2852	1.4704
Edinburgh Street	4.2514	7.5434	3.2920
Shakespear Park	4.1841	6.5150	2.3309
Port	6.6055	3.0216	-3.5839
Beachland	4.0000	6.5544	2.5544

As a result, a spatial data aided incremental SVR has been proposed, and it is able to more appropriately handle the short-term and missing data problem that arises in many existing forecasting models. We implemented the proposed SaIncSVR for the spatio-temporal $PM_{2.5}$ prediction on 13 monitoring stations in Auckland, New Zealand. The results show that the proposed SaIncSVR has strong capability on predicting a spatio-temporal environmental problem, in spite of various conditions of missing data. However, SaInSVR is found not working as good as a typical IncSVR for 4 of 13 stations, although these 4 stations have much less missing data than the remaining stations have. This could be explained that the geographical characteristic of these stations might be very different places, which makes the truth PM_{10} to $PM_{2.5}$ prediction model of these stations has very little or no correlation at all, and additional alternative spatial data is not able to mitigate such big gap. On the other hand, this also reflects that the applied spatio-temporal cross dimensional knowledge fusion in SaIncSVR is under the condition of the geographical correlation/similarity of the monitoring stations.

The proposed spatial prediction method is shown working just for a particular sub-region, and unable to maintain the integrity of the prediction for the entire region. Our interpretation regarding this shortcoming is that spatial prediction is influenced by complex interrelationships among the geographic characteristics of the city, the position of the monitoring stations, and the location of natural and man-made pollution emission sources, etc. Thus, a future direction for spatio-temporal environmental prediction is to discover the impacts of the geographical characteristics of monitoring stations to pollution variation.

Chapter 10

Conclusion and Future Works

A variety of environmental problems now seriously affects our world. These environmental problems have causes, numerous effects, and most importantly, potential solutions. The knowledge gained from environmental change facilitates the discovery of pollution events or the examination of the effectiveness of any adopted protection strategies. A thorough understanding of emission sources enhances our ability to detect changes in the environment. With greater understanding of environmental changes and their potential impacts, it is possible to predict pollution events and allow experts to propose possible protection strategies. These strategies could keep our ecosystems free from contamination or minimise the impact of those events. This thesis presented a systematic research on computational environmental analysis, which covers the following topics land use change detection, indoor emission source detection, and outdoor air quality prediction. One highly successful strategy for this research was the analysis of raw data to gain greater understanding of the problem. This initial analysis allowed us to test various environmental change detection algorithms for various applications including land use. We then analysed several indoor pollution source detection methods and built pollution detection models for indoor applications. Finally, the model was updated to model outdoor pollution and enable the development of an air quality prediction method.

10.1 Conclusion

We have found that environmental change detection has been widely applied to detect air pollution, water pollution and soil/land pollution. Additionally, emission, vegetation, global temperature, and global sea level are all under observation of environmental change detection. While in this thesis we reviewed existing change detection methods and proposed four criteria including accuracy, dynamic threshold, sensitivity, and continuity to evaluate change detection methods. These evaluations support simulation and model checking, in addition to some of the change detection techniques presented in the thesis, hence making it possible to apply different detection techniques to the same environmental change detection systems specification.

For the purpose of studying environmental event prediction models, this thesis validated previous works of environmental event forecasting. The prediction models have been applied to monitor many environmental problems such as Ozone holes, global warming and soil acidity and nutrient status. This thesis proposed three criteria to evaluate existing prediction models, namely accuracy, real time adaptation and spatio-temporal ability. This lets us qualify how close a prediction model is to an ideal specification for an environmental event forecasting system.

Feature extraction for environmental data allow us to improve performance of an environmental analysis. We have discussed the standard feature extraction methods for both sensor data (data stream) and remote sensor data (satellites images). We found that feature extraction methods such as PCA and LDA have been widely applied on reducing dimensions of environmental data or discovery important information from the input data in order to speed up the analysis without losing detection accuracy or even improve the detection accuracy. Gradient feature and wavelet feature are often utilized to magnify the texture information for remote sensor data to improve the edge detection in monitoring land use problem. In addition, noise can be filtered out from environmental data by term frequency based feature extraction such as wavelet-based and Pearson's correlation based sample validation. These types of methods can help us decide knowledge discovery methods according to specific environmental problems.

In order to detect continuous change over image series we have to provide consistently accurate change detection and improve the memory efficiency. This motivated us to develop the one-step-more incremental learning for image series change detection. The proposed method can learn knowledge on the source image for the purpose of detecting the difference against the target image. In developing the one-step-more incremental learning detection model we had to first understand the objects in source image and then find a way for storing those objects for future detection. Given two consecutive images in the image series, the source image is first utilized by the incremental learning agent to build up the knowledge base of the change detection model. Driven by the model, one-step-more incremental learning is further performed on the target image to determine any changes in comparison with the source image. After that, incremental learning will be performed again to update the change detection model derived from the source image according to the target image. The updated model enables us to detect any continuous changes in upcoming images.

After that, we applied one-step-more incremental learning detection model on land management problem in Auckland. The experiment was conducted on data collected in the form of digital images of parks from the Manukau and North shore areas of Auckland that include some of the 800 public parks in Auckland, New

Zealand. The model carried out successfully on 4 examples of encroachment show the effectiveness of the approach. The impact/cost/scale of encroachment is evaluated by multiplying the size of the encroachment by the rateable value of the land that has been encroached upon.

There are three primary pollutions considered in our indoor air quality analysis which includes carbon monoxide (CO), carbon oxide (CO_2), and particulate matter (PM). Secondary pollution results from chemical reactions between pollutants in the atmosphere. However, sensors cannot detect the reactions between pollutants in atmosphere, because the continuous reactions are invisible knowledge. As a result, most existing analyzing methods stop at discovering primary pollution. For better execution efficiency, we proposed a method, which considers alternatively the secondary pollution and detects emission sources by calculating in-between pollutants correlation coefficients. As validated in experimental results on data retrieved from a PACMAN unit in October 2012, our performance study shows that the proposed method is highly efficient and outperforms the typical feature extraction methods such as PCA and LDA.

Particulate matter ($PM_{2.5}$) are tiny pieces of solid or liquid matter in the air, which has been associated with adverse health effects.

We investigated $PM_{2.5}$ prediction over 13 monitoring stations in Auckland, New Zealand. Addressing the problems of outdoor data collection, we studied how spatial domain data can be used for the training of a temporal prediction model, and derive mathematically a spatial data aided incremental support vector regression (SaIncSVR) for spatio-temporal $PM_{2.5}$ prediction. The experimental results indicate that the proposed SaIncSVR presents better $PM_{2.5}$ prediction than the ordinary IncSVR for 8 of total 13 stations.

We decided to implement an alternative algorithm to gain more understanding of the geographical characteristics associated with $PM_{2.5}$ distribution. With this in mind, we proposed a Gaussian geometry model to trace the average levels of pollution at 13 monitoring stations in Auckland. The proposed model shows promise for a particular sub-region, however, it performed poorly compared with SaIncSVR. However, the work warrants further investigation for problems with limited samples captured from unique but specific spatial distributions.

10.2 Future Works

This work has been challenging and fulfilling and has a number of future directions.

1. One-step-more incremental learning has often failed when lighting changes. Lighting often causes unexpected frequency distortion on indoor images. As

a result of this distortion and the impulse noise, the distribution of image features extracted is complicated and often not linearly separable. Therefore, the linear SVM model utilized in this thesis may not be flexible enough in order to satisfactorily manage the lighting effects.

2. To confirm the effectiveness and efficiency of our correlation coefficient based emission source detection under various experimental settings, we need more experiment data consisting of multiple events repeated over a period of time.

3. To provide efficiently training data for $PM_{2.5}$ prediction, data collected synchronously from different stations is required. This helps us understand the influence to $PM_{2.5}$ from the condition of the geographical correlation/similarity of the monitoring stations.

4. To discover the geographical characteristics of monitor stations and their impacts to $PM_{2.5}$ prediction.

List of Publications

Song, L., Pang, S., Longley, I., Olivares, G., & Sarrafzadeh. A (in press). Spatial-temporal $PM_{2.5}$ Prediction by Spatial Data Aided Incremental Support Vector Regression. International Joint Conference on Neural Networks.

Song L., Pang L., Sarrafzadeh A, Ban T., and Inoue D. (2013). An Incremental Learning Approach to Continuous Images Change Detection, Proceeding of the 10th IEEE International Conference on Fuzzy Systems and Knowledge Discovery.

Dacey, S., Song, L., & Pang, S. (2013, January). An Intelligent Agent Based Land Encroachment Detection Approach. In Neural Information Processing (pp. 585-592). Springer Berlin Heidelberg.

Lai, A., Song, L., Peng, Y., Zhang, P., Wang, Q., & Pang, S. (2012, January). Exploring crude oil impacts to oil stocks through graphical computational correlation analysis. In Neural Information Processing (pp. 309-317). Springer Berlin Heidelberg.

Pang, S., Song, L., & Kasabov, N. (2011). Correlation-aided support vector regression for forex time series prediction. Neural Computing and Applications, 20(8), 1193-1203.

References

Aach, T., & Kaup, A. (1995). Bayesian algorithms for adaptive change detection in image sequences using markov random fields. *Signal Processing: Image Communication*, *7*(2), 147-160.

Aach, T., Kaup, A., & Mester, R. (1993). Statistical model-based change detection in moving video. *Signal Processing*, *31*(2), 165-180.

Abdul-Wahab, S. A., & Al-Alawi, S. M. (2002). Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. *Environmental Modelling & Software*, *17*(3), 219–228.

Abella, S. R., & Covington, W. W. (2004). Monitoring an arizona ponderosa pine restoration: Sampling efficiency and multivariate analysis of understory vegetation. *Restoration Ecology*, *12*(3), 359–367.

Acciani, G., Chiarantoni, E., Fornarelli, G., & Vergura, S. (2003). A feature extraction unsupervised neural network for an environmental data set. *Neural Networks*, *16*(3), 427–436.

Alonso, J. M., Alvarruiz, F., Desantes, J. M., Hernández, L., Hernández, V., & Moltó, G. (2007). Combining neural networks and genetic algorithms to predict and reduce diesel engine emissions. *Evolutionary Computation, IEEE Transactions on*, *11*(1), 46–55.

Amann, M., Bertok, I., Borken-Kleefeld, J., Cofala, J., Heyes, C., Hglund-Isaksson, L., ... Winiwarter, W. (2011). Cost-effective control of air quality and greenhouse gases in europe: Modeling and policy applications. *Environmental Modelling & Software*, *26*(12), 1489 - 1501. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1364815211001733` doi: 10.1016/j.envsoft.2011.07.012

Anderson, H. R., Limb, E. S., Bland, J. M., Ponce de Leon, A., Strachan, D. P., & Bower, J. S. (1995). Health effects of an air pollution episode in london,

december 1991. *Thorax*, *50*(11), 1188-1193.

Antony Chen, L.-W., Doddridge, B. G., Dickerson, R. R., Chow, J. C., & Henry, R. C. (2002). Origins of fine aerosol mass in the baltimore–washington corridor: implications from observation, factor analysis, and ensemble air parcel back trajectories. *Atmospheric Environment*, *36*(28), 4541–4554.

Appel, K. W., Gilliland, A. B., Sarwar, G., & Gilliam, R. C. (2007). Evaluation of the community multiscale air quality (cmaq) model version 4.5: sensitivities impacting model performance: part i ozone. *Atmospheric Environment*, *41*(40), 9603–9615.

Ardli, E., & Wolff, M. (2009). Land use and land cover change affecting habitat distribution in the segara anakan lagoon, java, indonesia. *Regional Environmental Change*, *9*, 235-243. (10.1007/s10113-008-0072-6)

Armitage, P., Moss, D., Wright, J., & Furse, M. (1983). The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*, *17*(3), 333 - 347.

Ates, Y., Erdinc, O., Uzunoglu, M., & Vural, B. (2010). Energy management of an fc/uc hybrid vehicular power system using a combined neural network-wavelet transform based strategy. *International journal of hydrogen energy*, *35*(2), 774–783.

Atkinson, R., Anderson, H., Strachan, D., Bland, J., Bremmer, S., & Ponce de Leon, A. (1999). Short-term associations between outdoor air pollution and visits to accident and emergency departments in london for respiratory complaints. *European Respiratory Journal*, *13*(2), 257–265.

Babcock, B., Babu, S., Datar, M., Motwani, R., & Widom, J. (2002). Models and issues in data stream systems. In *Proceedings of the twenty-first acm sigmod-sigact-sigart symposium on principles of database systems* (pp. 1–16).

Baker, K., & Scheff, P. (2007). Photochemical model performance for pm2. 5 sulfate, nitrate, ammonium, and precursor species so2, hno3, and nh3 at background monitor locations in the central and eastern united states. *Atmospheric Environment*, *41*(29), 6185–6195.

Barlow, W. E., White, E., Ballard-Barbash, R., Vacek, P. M., Titus-Ernstoff, L., Carney, P. A., ... Kerlikowske, K. (6 September 2006). Prospective breast cancer risk prediction model for women undergoing screening mammography. *Journal of the National Cancer Institute*, *98*(17), 1204-1214.

Baum, L. E., & Eagon, J. (1967). An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bull. Amer. Math. Soc*, *73*(3), 360–363.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of

finite state markov chains. *The annals of mathematical statistics*, *37*(6), 1554–1563.

Baum, L. E., & Sell, G. R. (1968). Growth transformations for functions on manifolds. *Pacific J. Math*, *27*(2), 211–227.

Bazi, Y., Bruzzone, L., & Melgani, F. (2005a). An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal sar images. *Geoscience and Remote Sensing, IEEE Transactions on*, *43*(4), 874-887.

Bazi, Y., Bruzzone, L., & Melgani, F. (2005b). An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal sar images. *Geoscience and Remote Sensing, IEEE Transactions on*, *43*(4), 874 - 887.

Belaid, M. A. (2003). Urban-rural land use change detection and analysis using gis and rs technologies. In *Proceedings of 2nd fig regional.*

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. *Journal of hydrology*, *249*(1), 11–29.

Bianconi, F., & Fernández, A. (2007). Evaluation of the effects of gabor filter parameters on texture classification. *Pattern Recognition*, *40*(12), 3325–3335.

Bishop, C. (1994). Novelty detection and neural network validation. *Vision, Image and Signal Processing, IEE Proceedings -*, *141*(4), 217 -222.

Black, M. J., Fleet, D. J., & Yacoob, Y. (2000). Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, *78*(1), 8 - 31.

Blasco, M., Domeño, C., López, P., & Nerín, C. (2011). Behaviour of different lichen species as biomonitors of air pollution by pahs in natural ecosystems. *Journal of Environmental Monitoring*, *13*(9), 2588–2596.

Boiral, O. (2002). Tacit knowledge and environmental management. *Long Range Planning*, *35*.

Bönisch, U., Böhme, A., Kohajda, T., Mögel, I., Schütze, N., von Bergen, M., ... Polte, T. (2012). Volatile organic compounds enhance allergic airway inflammation in an experimental mouse model. *PloS one*, *7*(7), e39817.

Boor, V., Overmars, M., & van der Stappen, A. (1999). The gaussian sampling strategy for probabilistic roadmap planners. In *Robotics and automation, 1999. proceedings. 1999 ieee international conference on* (Vol. 2, p. 1018 -1023 vol.2). doi: 10.1109/ROBOT.1999.772447

Borges, P., & Izquierdo, E. (2010, may). A probabilistic approach for vision-based fire detection in videos. *Circuits and Systems for Video Technology, IEEE Transactions on*, *20*(5), 721 -731. doi: 10.1109/TCSVT.2010.2045813

Bovolo, F., Bruzzone, L., & Marconcini, M. (2008). A novel approach to

unsupervised change detection based on a semisupervised svm and a similarity measure. *Geoscience and Remote Sensing, IEEE Transactions on*, *46*(7), 2070 -2082. doi: 10.1109/TGRS.2008.916643

Boznar, M. (1997). Pattern selection strategies for a neural network-based short term air pollution prediction model. *Intelligent Information Systems, IASTED International Conference on*, *0*, 340. doi: http://doi.ieeecomputersociety.org/10.1109/IIS.1997.645285

Broggi, A., & Cattani, S. (2006, August). An agent based evolutionary approach to path detection for off-road vehicle guidance. *Pattern Recogn. Lett.*, *27*(11), 1164–1173.

Brown, T. J., Hall, B. L., & Westerling, A. L. (2004). The impact of twenty-first century climate change on wildland fire danger in the western united states: An applications perspective. *Climatic Change*, *62*, 365-388. (10.1023/B:CLIM.0000013680.07783.de)

Brudzynski, K., & Miotto, D. (2011). The relationship between the content of maillard reaction-like products and bioactivity of canadian honeys. *Food Chemistry*, *124*(3), 869 - 874. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0308814610008320` doi: 10.1016/j.foodchem.2010.07.009

Bruzzone, L., & Preito, D. (2000). An adaptive parcel-based technique for unsupervised change detection. *International Journal of Remote Sensing*, *21*(4), 817–822.

Bruzzone, L., & Prieto, D. (2000). Automatic analysis of the difference image for unsupervised change detection. *Geoscience and Remote Sensing, IEEE Transactions on*, *38*(3), 1171 -1182.

Bruzzone, L., & Prieto, D. (2002). An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *Image Processing, IEEE Transactions on*, *11*(4), 452 -466.

Calder, C. A. (2007). Dynamic factor process convolution models for multivariate space–time data with application to air quality assessment. *Environmental and Ecological Statistics*, *14*(3), 229–247.

Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Rojo-Alvarez, J., & Martinez-Ramon, M. (2008). Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *Geoscience and Remote Sensing, IEEE Transactions on*, *46*(6), 1822 -1835. doi: 10.1109/TGRS.2008.916201

Carlton, G. N., & Smith, L. B. (2000). Exposures to jet fuel and benzene during aircraft fuel tank repair in the u.s. air force. *Applied Occupational and Environmental Hygiene*, *15*(6), 485 - 491.

Castro-Neto, M., Jeong, Y.-S., Jeong, M.-K., & Han, L. D. (2009). Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert systems with applications*, *36*(3), 6164–6173.

Celik, T. (2009). Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *Geoscience and Remote Sensing Letters, IEEE*, *6*(4), 772–776.

Celik, T., & Ma, K.-K. (2010). Unsupervised change detection for satellite images using dual-tree complex wavelet transform. *Geoscience and Remote Sensing, IEEE Transactions on*, *48*(3), 1199-1210.

Chang, S.-C., & Lee, C.-T. (2008). Evaluation of the temporal variations of air quality in taipei city, taiwan, from 1994 to 2003. *Journal of environmental management*, *86*(4), 627–635.

Chavez, P. (1996). Image-based atmospheric corrections-revisited and improved. *Photogrammetric engineering and remote sensing*, *62*(9), 1025.

Chelani, A. B., Chalapati Rao, C., Phadke, K., & Hasan, M. (2002). Prediction of sulphur dioxide concentration using artificial neural networks. *Environmental Modelling & Software*, *17*(2), 159–166.

Chelliah, M., & Ropelewski, C. F. (2000). Reanalyses-based tropospheric temperature estimates: Uncertainties in the context of global climate change detection. *Journal of Climate*, *13*(17), 3187-3205. doi: 10.1175/1520-0442(2000)013¡3187:RBTTEU¿2.0.CO;2

Chen, Y., Pang, S., Kasabov, N., Ban, T., & Kadobayashi, Y. (2009). Hierarchical core vector machines for network intrusion detection. In C. Leung, M. Lee, & J. Chan (Eds.), *Neural information processing* (Vol. 5864, p. 520-529). Springer Berlin / Heidelberg.

Cheng, S., Yang, L., Zhou, X., Wang, Z., Zhou, Y., Gao, X., ... Wang, W. (2011). Evaluating $pm_{2.5}$ ionic components and source apportionment in jinan, china from 2004 to 2008 using trajectory statistical methods. *Journal of Environmental Monitoring*, *13*(6), 1662–1671.

Cheng, T., & Wang, J. (2008). Integrated spatio-temporal data mining for forest fire prediction. *Transactions in GIS*, *12*(5), 591–611.

Cheon, S.-P., Kim, S., Lee, S.-Y., & Lee, C.-B. (2009). Bayesian networks based rare event prediction with sensor data. *Knowledge-Based Systems*, *22*(5), 336–343.

Choi, J. Y., Ro, Y. M., & Plataniotis, K. (2011, may). Boosting color feature selection for color face recognition. *Image Processing, IEEE Transactions on*, *20*(5), 1425 -1434. doi: 10.1109/TIP.2010.2093906

Christakos, G. (1990). A bayesian/maximum-entropy view to the spatial estimation problem. *Mathematical Geology*, *22*(7), 763–777.

Christakos, G., & Serre, M. L. (2000). Bme analysis of spatiotemporal particulate

matter distributions in north carolina. *Atmospheric Environment*, *34*(20), 3393–3406.

Chu, F., & Zaniolo, C. (2004). Fast and light boosting for adaptive mining of data streams. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in knowledge discovery and data mining* (Vol. 3056, p. 282-292). Springer Berlin / Heidelberg.

Cohen, D. D., Garton, D., Stelcer, E., Wang, T., Poon, S., Kim, J., ... others (2002). Characterisation of pm2. 5 and pm10 fine particle pollution in several asian regions. In *Proceedings of 16th international clean air and environment conference* (pp. 153–158).

Coiner, J. C. (1980). Using landsat to monitor changes in vegetation cover induced by desertification processes. In *14th international symposium on remote sensing of environment* (p. 1341-1351). San Jose, Costa Rica.

Coman, A., Ionescu, A., & Candau, Y. (2008). Hourly ozone prediction for a 24-h horizon using neural networks. *Environmental Modelling & Software*, *23*(12), 1407–1421.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International journal of remote sensing*, *25*(9), 1565-1596.

Corbane, C., Faure, J.-F., Baghdadi, N., Villeneuve, N., & Petit, M. (2008). Rapid urban mapping using sar/optical imagery synergy. *Sensors*, *8*(11), 7125–7143.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273–297.

Council, A. C. (2010). *Auckland gis viewer.* Retrieved from `http://maps.aucklandcouncil.govt.nz/aucklandcouncilviewer/`

Couvreur, C., Fontaine, V., Gaunard, P., & Mubikangiey, C. G. (1998). Automatic classification of environmental noise events by hidden markov models. *Applied Acoustics*, *54*(3), 187–206.

Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., & Slattery, S. (2000). Learning to construct knowledge bases from the world wide web. *Artificial Intelligence*, *118*(1-2), 69 - 113.

Cressie, N., & Huang, H. C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, *94*(448), 1330-1340.

Cuadros, J. F., Melo, D. C., Maciel Filho, R., & Wolf, M. R. (2012). Fluid catalytic cracking environmental impact: Factorial design coupled with genetic algorithms to minimize carbon monoxide pollution. *CHEMICAL ENGINEERING*, *26*.

Dacey, S., Barbour, R., & Fernando, A. (2011). Participatory land use management

of public spaces in new zealand. In *Proceedings of the jurse conference on urban remote sensing and planning* (pp. 449–452). Munich, Germany.

Dacey, S., Song, L., & Pang, S. (2013). An intelligent agent based land encroachment detection approach. In *Neural information processing* (pp. 585–592).

Dagan, S. (2000). Comparison of gas chromatography-pulsed flame photometric detection-mass spectrometry, automated mass spectral deconvolution and identification system and gas chromatography-tandem mass spectrometry as tools for trace level detection and identification. *Journal of Chromatography A*, *868*(2), 229 - 247.

Dasgupta, D., & Forrest, S. (1995). Novelty detection in time series data using ideas from immunology. In *In proceedings of the international conference on intelligent systems.*

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A Optics and image science*, *2*(7), 1160–1169.

de Brito, R. F., & Oliveira, A. L. (2012). A foreign exchange market trading system by combining ghsom and svr. In *Neural networks (ijcnn), the 2012 international joint conference on* (pp. 1–7).

de Gruijl, F. R., Longstreth, J., Norval, M., Cullen, A. P., Slaper, H., Kripke, M. L., ... van der Leun, J. C. (2003). Health effects from stratospheric ozone depletion and interactions with climate change. *Photochem. Photobiol. Sci.*, *2*, 16-28.

Dengkui, M., Hui, L., Jiping, L., Hua, S., Zhuo, Z., & Yujiu, X. (2008). A svm-based change detection method from bi-temporal remote sensing images in forest area. In *Knowledge discovery and data mining, 2008. wkdd 2008. first international workshop on* (p. 209-212).

Derwent, R., Middleton, D., Field, R., Goldstone, M., Lester, J., & Perry, R. (1995). Analysis and interpretation of air quality data from an urban roadside location in central london over the period from july 1991 to july 1992. *Atmospheric Environment*, *29*(8), 923–946.

Desrosieres, A., Naish, C., & Hunter, P. (2006). The politics of large numbers. a history of statistical reasoning. *The Mathematical Intelligencer*, *28*, 69-70. (10.1007/BF02987012)

Dong, M., Yang, D., Kuang, Y., He, D., Erdal, S., & Kenski, D. (2009). Pm¡ sub¿ 2.5¡/sub¿ concentration prediction using hidden semi-markov model-based times series data mining. *Expert Systems with Applications*, *36*(5), 9046–9055.

Downs, T., Gates, K. E., & Masters, A. (2002). Exact simplification of support vector solutions. *The Journal of Machine Learning Research*, *2*, 293–297.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support

vector regression machines. *Advances in neural information processing systems*(9), 115-161.

Duda, R. O., Hart, P. E., et al. (1973). *Pattern classification and scene analysis* (Vol. 3). Wiley New York.

Duskunovic, I., Heene, G., Philips, W., & Bruyland, I. (2000). Urban area detection in sar imagery using a new speckle reduction technique and markov random field texture classification. In *Geoscience and remote sensing symposium, 2000. proceedings. igarss 2000. ieee 2000 international* (Vol. 2, p. 636-638 vol.2).

Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G., & Brunak, S. (2004). Improved prediction of signal peptides: Signalp 3.0. *Journal of molecular biology*, *340*(4), 783–795.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677–697.

Ennaceur, A., & Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural Brain Research*, *31*(1), 47 - 59.

Farmer, M., & Jain, A. (2005, dec.). A wrapper-based approach to image segmentation and classification. *Image Processing, IEEE Transactions on*, *14*(12), 2060 -2072. doi: 10.1109/TIP.2005.859374

Fei-Fei, L., Fergus, R., & Perona, P. (2007). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, *106*(1), 59 - 70. (Special issue on Generative Model Based Vision)

Felipe-Sotelo, M., Andrade, J., Carlosena, A., & Tauler, R. (2007). Temporal characterisation of river waters in urban and semi-urban areas using physico-chemical parameters and chemometric methods. *Analytica chimica acta*, *583*(1), 128–137.

Fischer, S. L., Dryer, F. L., & Curran, H. J. (2000). The reaction kinetics of dimethyl ether. i: High-temperature pyrolysis and oxidation in flow reactors. *INTERNATIONAL JOURNAL OF CHEMICAL KINETICS*, *32*(12), 713-740.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, *7*(2), 179–188.

Fishman, J., Wozniak, A., & Creilson, J. (2003). Global distribution of tropospheric ozone from satellite measurements using the empirically corrected tropospheric ozone residual technique: Identification of the regional aspects of air pollution. *Atmospheric Chemistry and Physics*, *3*(4), 893–907.

Fodor, I. K. (2002). *A survey of dimension reduction techniques.* Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory.

Fortin, M.-J., Olson, R., Ferson, S., Iverson, L., Hunsaker, C., Edwards, G., ... Klemas, V. (2000). Issues related to the detection of boundaries. *Landscape Ecology*, *15*, 453-466. (10.1023/A:1008194205292)

Francke, K.-P., Miessner, H., & Rudolph, R. (2000). Cleaning of air streams from organic pollutants by plasmacatalytic oxidation. *Plasma Chemistry and Plasma Processing*, *20*, 393-403. (10.1023/A:1007048428975)

Fransson, N., & Gärling, T. (1999). Environmental concern: Conceptual definitions, measurement methods, and research findings. *Journal of environmental psychology*, *19*(4), 369–382.

Freeman, H. (1961). On the encoding of arbitrary geometric configurations. *Electronic Computers, IRE Transactions on*(2), 260–268.

Friedman, M. S., Powell, K. E., Hutwagner, L., Graham, L. M., & Teague, W. G. (2001). Impact of changes in transportation and commuting behaviors during the 1996 summer olympic games in atlanta on air quality and childhood asthma. *JAMA: the journal of the American Medical Association*, *285*(7), 897–905.

Fries, A. B. W., Ziegler, T. E., Kurian, J. R., Jacoris, S., & Pollak, S. D. (2005). Early experience in humans is associated with changes in neuropeptides critical for regulating social behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(47), 17237-17240. doi: 10.1073/pnas.0504767102

Fuller, D. O. (2001). Forest fragmentation in loudoun county, virginia, usa evaluated with multitemporal landsat imagery. *Landscape Ecology*, *16*(7), 627–642.

Gabor, D. (1946). Theory of communication. part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, *93*(26), 429–441.

Gamba, P., Dell'Acqua, F., & Lisini, G. (2006). Change detection of multitemporal sar data in urban areas combining feature-based and pixel-based techniques. *Geoscience and Remote Sensing, IEEE Transactions on*, *44*(10), 2820 -2827. doi: 10.1109/TGRS.2006.879498

Gardner, M., & Dorling, S. (1999). Neural network modelling and prediction of hourly no¡ sub¿ x¡/sub¿ and no¡ sub¿ 2¡/sub¿ concentrations in urban air in london. *Atmospheric Environment*, *33*(5), 709–719.

Gece, G., & Bilgiç, S. (2009). Quantum chemical study of some cyclic nitrogen compounds as corrosion inhibitors of steel in nacl media. *Corrosion Science*, *51*(8), 1876–1878.

Gehrig, R., & Buchmann, B. (2003). Characterising seasonal variations and spatial distribution of ambient pm10 and pm2. 5 concentrations based on long-term swiss monitoring data. *Atmospheric Environment*, *37*(19), 2571–2580.

Georgopoulos, P. G., Wang, S.-W., Vyas, V. M., Sun, Q., Burke, J., Vedantham, R., ...

Özkaynak, H. (2005). A source-to-dose assessment of population exposures to fine pm and ozone in philadelphia, pa, during a summer 1999 episode. *Journal of Exposure Science and Environmental Epidemiology*, *15*(5), 439–457.

Ghosh, A., Mishra, N., & Ghosh, S. (2011). Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. *Information Sciences*, *181*, 699–715.

Giorgi, F., & Mearns, L. O. (2002). Calculation of average, uncertainty range, and reliability of regional climate changes from aogcm simulations via the reliability ensemble averaging (rea) method. *Journal of Climate*, *15*(10), 1141-1158.

Glass, G. E., Cheek, J. E., Patz, J. A., Shields, T. M., Doyle, T. J., Thoroughman, D. A., ... Bryan, R. (2000). Using remotely sensed data to identify areas at risk for hantavirus pulmonary syndrome.

Gonzalez, R. C. (1987). P. wintz digital image processing. *Addision-Wesley Publishing Company*, 275–281.

Goyal, P., Chan, A. T., & Jaiswal, N. (2006). Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmospheric Environment*, *40*(11), 2068–2077.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, *60*, 549–576.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646-675. doi: 10.1093/poq/nfl033

Guardian, S. (2008). *Study links traffic pollution to thousands of deaths.* Author. Retrieved from `http://www.guardian.co.uk/society/2008/apr/15/health`

Guo, Y., Jia, Y., Pan, X., Liu, L., & Wichmann, H.-E. (2009). The association between fine particulate air pollution and hospital emergency room visits for cardiovascular diseases in beijing, china. *Science of The Total Environment*, *407*(17), 4826 - 4830.

Gurjar, B. R., Molina, L. T., & Ojha, C. S. P. (2010). *Air pollution: health and environmental impacts*. CRC Press.

Gaceanu, R. D. (2012). *Agent based pattern recognition*. Unpublished doctoral dissertation, Babes-Bolyai University and Eotvos Lorland University.

Hackett, B., Comerma, E., Daniel, P., & Ichikawa, H. (2009). Marine oil pollution prediction. *Oceanography*, *22*, 168-175.

Hansen, E. A., & Harris, A. R. (1975). Validity of soil-water samples collected with porous ceramic cups1. *Soil Sci. Soc. Am. J.*, *39*(3), 528-536.

Hansen, T. B., & Johansen, P. M. (2000). Inversion scheme for ground penetrating radar that takes into account the planar air-soil interface. *Geoscience and*

Remote Sensing, IEEE Transactions on, *38*(1), 496–506.

Harkat, M.-F., Mourot, G., & Ragot, J. (2006). An improved pca scheme for sensor fdi: Application to an air quality monitoring network. *Journal of Process Control*, *16*(6), 625–634.

Haverkamp, D., & Poulsen, R. (2003). Change detection using ikonos imagery. In *Asprs 2003 annual conference proceedings.* Anchorage, Alaska.

Hawthorne, G., & Elliott, P. (2005). Imputing cross-sectional missing data: comparison of common techniques. *Australian and New Zealand Journal of Psychiatry*, *39*(7), 583–590.

He, C., Weib, A., P., S., Zhang, Q., & Zhao, Y. (2011). Detecting land-use/land-cover change in ruralurban fringe areas using extended change-vector analysis. *International Journal of Applied Earth Observation and Geoinformation*, *13*, 572–585.

Heinrich, J. (2011). Influence of indoor factors in dwellings on the development of childhood asthma. *International journal of hygiene and environmental health*, *214*(1), 1–25.

Henry, R., & Hidy, G. (1979). Multivariate analysis of particulate sulfate and other air quality variables by principal components-part i: Annual data from los angeles and new york. *Atmospheric Environment (1967)*, *13*(11), 1581–1596.

Hernández-Campos, F., Karaliopoulos, M., Papadopouli, M., & Shen, H. (2006). Spatio-temporal modeling of traffic workload in a campus wlan. In *Proceedings of the 2nd annual international workshop on wireless internet.* New York, NY, USA: ACM.

Hill, J., Stellmes, M., Udelhoven, T., Roder, A., & Sommer, S. (2008). Mediterranean desertification and land degradation: Mapping related land use change syndromes based on satellite observations. *Global and Planetary Change*, *64*(3-4), 146 - 157. (Climate Change and Desertification)

Hirsch, R., Ternes, T. A., Haberer, K., Mehlich, A., Ballwanz, F., & Kratz, K.-L. (1998). Determination of antibiotics in different water compartments via liquid chromatography-electrospray tandem mass spectrometry. *Journal of Chromatography A*, *815*(2), 213 - 223.

Hochadel, M., Heinrich, J., Gehring, U., Morgenstern, V., Kuhlbusch, T., Link, E., ... others (2006). Predicting long-term average concentrations of traffic-related air pollutants using gis-based information. *Atmospheric Environment*, *40*(3), 542–553.

Hoffhine Wilson, E., Hurd, J. D., Civco, D. L., Prisloe, M. P., & Arnold, C. (2003). Development of a geospatial model to quantify, describe and map urban growth. *Remote sensing of environment*, *86*(3), 275–285.

Huang, J., Choi, H.-D., Hopke, P. K., & Holsen, T. M. (2010). Ambient mercury

sources in rochester, ny: results from principle components analysis (pca) of mercury monitoring network data. *Environmental science & technology*, *44*(22), 8441–8445.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, *2*, 283-304. (10.1023/A:1009769707641)

Hueglin, C., Gehrig, R., Baltensperger, U., Gysel, M., Monn, C., & Vonmont, H. (2005). Chemical characterisation of $pm_{2.5}$, pm_{10} and coarse particles at urban, near-city and rural sites in switzerland. *Atmospheric Environment*, *39*(4), 637–651.

Hwa-Lung, Y., & Chih-Hsin, W. (2010). Retrospective prediction of intraurban spatiotemporal distribution of $pm_{2.5}$ in taipei. *Atmospheric Environment*, *44*(25), 3053–3065.

Hwang, H., & Haddad, R. (1995, apr). Adaptive median filters: new algorithms and results. *Image Processing, IEEE Transactions on*, *4*(4), 499 -502. doi: 10.1109/83.370679

Iimura, Y., Saito, K., Fujii, Y., Kumagai, J., Kawakami, S., Komai, Y., ... Kihara, K. (2009). Development and external validation of a new outcome prediction model for patients with clear cell renal cell carcinoma treated with nephrectomy based on preoperative serum c-reactive protein and tnm classification: The tnm-c score. *The Journal of Urology*, *181*(3), 1004 - 1012. doi: 10.1016/j.juro.2008.10.156

Ilin, A., & Luttinen, J. (2010). Variational gaussian-process factor analysis for modeling spatio-temporal. *EGU General Assembly 2010*, *12*.

Im, J., Jensen, J., & Tullis, J. (2008b). Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing*, *29*(2), 399–423.

Im, J., Jensen, J. R., & Tullis, J. A. (2008a). Object - based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing*, *29*(2), 399-423. doi: 10.1080/01431160601075582

Jain, A. K. (1989). *Fundamentals of digital image processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Jain, R. (1984). Difference and accumulative difference pictures in dynamic scene analysis. *Image and Vision Computing*, *2*(2), 99 - 108.

Jensen, J. R. (1981). Urban change detection mapping using landsat digital data. *Cartography and Geographic Information Science*, *8*, 127-147(21).

Ji, L., & Gallo, K. (2006). An agreement coefficient for image comparison. *Photogrammetric engineering and remote sensing*, *72*(7), 823.

Ji, Y., Chang, K. H., & Hung, C.-C. (2004). Efficient edge detection and object

segmentation using gabor filters. In *Proceedings of the 42nd annual southeast regional conference* (pp. 454–459).

Jolliffe, I. T. (1986). *Principal component analysis* (Vol. 487). Springer-Verlag New York.

Joly, M., & Peuch, V.-H. (2012). Objective classification of air quality monitoring sites over europe. *Atmospheric Environment*, *47*, 111–123.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., & Kolehmainen, M. (2004). Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, *38*(18), 2895–2907.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., ... Yamanishi, Y. (2008). Kegg for linking genomes to life and the environment. *Nucleic Acids Research*, *36*(suppl 1), D480-D484.

Karacan, C. Ö. (2008). Modeling and prediction of ventilation methane emissions of us longwall mines using supervised artificial neural networks. *International Journal of Coal Geology*, *73*(3), 371–387.

Katsouyanni, K., Schwartz, J., Spix, C., Touloumi, G., Zmirou, D., Zanobetti, A., ... others (1996). Short term effects of air pollution on health: a european approach using epidemiologic time series data: the aphea protocol. *Journal of epidemiology and community health*, *50*(Suppl 1), S12–S18.

Keith, L. H. (1988). *Principles of environmental sampling*.

Keith, L. H. (1991). *Environmental sampling and analysis: a practical guide*. Chelsea, Mich: Lewis Publishers.

Kennedy, R. E., Townsend, P. A., Gross, J. E., Cohen, W. B., Bolstad, P., Wang, Y. Q., & Adams, P. (2009). Remote sensing change detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects. *Remote Sensing of Environment*, *113*, 1382–1396.

Kirchmer, C. J. (1987). Estimation of detection limits for environmental analytical procedures. In *Detection in analytical chemistry* (p. 78-93). Washington, DC: American Chemical Society. doi: 10.1021/bk-1988-0361.ch004

Klanova, J., Kohoutek, J., Hamplova, L., Urbanova, P., & Holoubek, I. (2006). Passive air sampler as a tool for long-term air pollution monitoring: Part 1. performance assessment for seasonal and spatial variations. *Environmental Pollution*, *144*(2), 393 - 405. (Passive Air Sampling of Persistent Organic Pollutants - Passive Air Sampling of Persistent Organic Pollutants)

Kohara, K., Ishikawa, T., Fukuhara, Y., & Nakamura, Y. (1997). Stock price prediction using prior knowledge and neural networks. *Intelligent systems in accounting, finance and management*, *6*(1), 11–22.

Kolehmainen, M., Martikainen, H., & Ruuskanen, J. (2001). Neural networks and

periodic components used in air quality forecasting. *Atmospheric Environment*, *35*(5), 815–825.

Kong, S., Han, B., Bai, Z., Chen, L., Shi, J., & Xu, Z. (2010). Receptor modeling of pm2. 5, pm10 and tsp in different seasons and long-range transport analysis at a coastal site of tianjin, china. *Science of the Total Environment*, *408*(20), 4681–4694.

Kreinovich, V., Longpr, L., Starks, S. A., Xiang, G., Beck, J., Kandathi, R., … Hajagos, J. (2007). Interval versions of statistical techniques with applications to environmental analysis, bioinformatics, and privacy in statistical databases. *Journal of Computational and Applied Mathematics*, *199*(2), 418 - 423. (¡ce:title¿Special Issue on Scientific Computing, Computer Arithmetic, and Validated Numerics (SCAN 2004)¡/ce:title¿ ¡xocs:full-name¿Special Issue on Scientific Computing, Computer Arithmetic, and Validated Numerics (SCAN 2004)¡/xocs:full-name¿) doi: 10.1016/j.cam.2005.07.041

Kumar, U., & Jain, V. (2010). Arima forecasting of ambient air pollutants (o3, no, no2 and co). *Stochastic Environmental Research and Risk Assessment*, *24*(5), 751–760.

Larry, W. (2006). *World water day: A billion people worldwide lack safe drinking water.* About.com.

Lazic, V., Colao, F., Fantoni, R., & Spizzicchino, V. (2005). Laser-induced breakdown spectroscopy in water: Improvement of the detection threshold by signal processing. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *60*(7-8), 1002 - 1013. (Laser Induced Plasma Spectroscopy and Applications (LIBS 2004) Third International Conference - LIBS Malaga)

Lee, P. K., Brook, J. R., Dabek-Zlotorzynska, E., & Mabury, S. A. (2003). Identification of the major sources contributing to pm2. 5 observed in toronto. *Environmental science & technology*, *37*(21), 4831–4840.

Le Hegarat-Mascle, S., Ottle, C., & Guerin, C. (2005). Land cover change detection at coarse spatial scales based on iterative estimation and previous state information. *Remote Sensing of Environment*, *95*, 464–479.

Li, C.-S., & Lin, C.-H. (2002). $pm_1/pm_{2.5}/pm_{10}$ characteristics in the urban atmosphere of taipei. *Aerosol Science & Technology*, *36*(4), 469–473.

Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pka values. *Proteins: Structure, Function, and Bioinformatics*, *61*(4), 704–721.

Li, W., Mao, K., Zhang, H., & Chai, T. (2010, sept.). Selection of gabor filters for improved texture feature extraction. In *Image processing (icip), 2010 17th ieee international conference on* (p. 361 -364). doi: 10.1109/ICIP.2010.5653278

Li, W., Wang, C., Wang, Q., & Chen, G. (2008, dec.). Design of comprehensive color

feature vector for tile color online inspection. In *Information science and engineering, 2008. isise '08. international symposium on* (Vol. 2, p. 24 -27). doi: 10.1109/ISISE.2008.312

Li, Y., Alvarez, O. A., Gutteling, E. W., Tijsterman, M., Fu, J., Riksen, J. A. G., ... Kammenga, J. E. (2006). Mapping determinants of gene expression plasticity by genetical genomics in c. elegans. *PLoS Genet*, *2*(12), e222. doi: 10.1371/journal.pgen.0020222

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological monographs*, *72*(3), 445–463.

Lin, Y., & Cobourn, W. G. (2007). Fuzzy system models combined with nonlinear regression for daily ground-level ozone predictions. *Atmospheric Environment*, *41*(16), 3502–3513.

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, *30*(2), 117–156.

Liu, D., Song, K., Townshend, J. R., & Gong, P. (2008). Using local transition probability models in markov random fields for forest change detection. *Remote Sensing of Environment*, *112*(5), 2222 - 2231. (Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue)

Liu, H., Zhou, C., Cheng, W., Long, E., & Li, R. (2008). Monitoring sandy desertification of otindag sandy land based on multi-date remote sensing images. *Acta Ecologica Sinica*, *28*(2), 627 - 635.

Lizarazo, I. (2011). Quantitative land cover change analysis using fuzzy segmentation. *International Journal of Applied Earth Observation and Geoinformation*.

Llewellyn-Jones, D., Edwards, M. C., Mutlow, C. T., Birks, A. R., Barton, I. J., & Tait, H. (2001). Aatsr: global-change and surface-temperature measurements from envisat. *ESA Bulletin*, *105*, 11-21.

Lo, C. (2000). Relative radiometric normalization performance for change detection from multi-date satellite images. *Photogrammetric engineering and remote sensing*, *66*(8), 967.

Longsworth, L. G. (1964). Determination of ph. theory and practice. *Journal of the American Chemical Society*, *86*(18), 3912-3912.

Loomis, D., Castillejos, M., Gold, D. R., McDonnell, W., & Borja-Aburto, V. H. (1999). Air pollution and infant mortality in mexico city. *Epidemiology*, *10*(2), 118-123.

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing*, *25*(12), 2365 - 2401.

Lu, W., Fan, H., & Lo, S. (2003). Application of evolutionary neural network

method in predicting pollutant levels in downtown area of hong kong. *Neurocomputing*, *51*, 387–400.

Lu, W., Wang, W., Leung, A., Lo, S.-M., Yuen, R. K., Xu, Z., & Fan, H. (2002). Air pollutant parameter forecasting using support vector machines. In *Neural networks, 2002. ijcnn'02. proceedings of the 2002 international joint conference on* (Vol. 1, pp. 630–635).

Lu, W.-Z., He, H.-D., & yun Dong, L. (2011). Performance assessment of air quality monitoring networks using principal component analysis and cluster analysis. *Building and Environment*, *46*(3), 577 - 583. Retrieved from http://www.sciencedirect.com/science/article/pii/S036013231000274X doi: 10.1016/j.buildenv.2010.09.004

Lu, W.-Z., & Wang, D. (2008). Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme. *Science of the total environment*, *395*(2), 109–116.

Luhar, A. K., & Patil, R. (1989). A general finite line source model for vehicular pollution prediction. *Atmospheric Environment (1967)*, *23*(3), 555 - 562. doi: 10.1016/0004-6981(89)90004-8

Ma, L., Zheng, N., Yuan, Z., & Zhang, X. (2010). A novel dual-probe adaptive model for image change detection. *Signal Processing Letters, IEEE*, *17*(10), 863 -866. doi: 10.1109/LSP.2010.2063023

Makarewicz, J. C., Lewis, T. W., Boyer, G. L., & Edwards, W. J. (2012). The influence of streams on nearshore water chemistry, lake ontario. *Journal of Great Lakes Research*.

Malik, R. N., & Husain, S. Z. (2006). Land-cover mapping: A remote sensing approach. *Pakistan Journal of Botany*, *38*(3), 559–570.

Maloof, M. A., & Michalski, R. S. (2004). Incremental learning with partial instance memory. *Artificial Intelligence*, *154*(1-2), 95 - 126.

Mandalia, H. M. (2005). Using support vector machines for lane-change detection. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *49*, 1965-1969(5).

Marcazzan, G. M., Vaccaro, S., Valli, G., & Vecchi, R. (2001). Characterisation of pm10 and pm2. 5 particulate matter in the ambient air of milan (italy). *Atmospheric Environment*, *35*(27), 4639–4650.

Marié, D. C., Chaparro, M. A., Gogorza, C. S., Navas, A., & Sinito, A. M. (2010). Vehicle-derived emissions and pollution on the road autovia 2 investigated by rock-magnetic parameters: A case study from argentina. *Studia Geophysica et Geodaetica*, *54*(1), 135–152.

Martinez, A. W., Phillips, S. T., & Whitesides, G. M. (2008). Three-dimensional microfluidic devices fabricated in layered paper and tape. *Proceedings of the*

National Academy of Sciences, *105*(50), 19606-19611.

Mas, J. F. (1999). Monitoring land-cover changes: A comparison of change detection techniques. *International Journal of Remote Sensing*, *20*(1), 139 - 152.

Masiol, M., Squizzato, S., Ceccato, D., Rampazzo, G., & Pavoni, B. (2012). A chemometric approach to determine local and regional sources of {PM10} and its geochemical composition in a coastal area. *Atmospheric Environment*, *54*(0), 127 - 133. Retrieved from http://www.sciencedirect.com/science/article/pii/S1352231012002270 doi: 10.1016/j.atmosenv.2012.02.089

Matern, B. (1986). *Spatial variation*. New York: Springer-Verlag.

Matsuzawa, K., Sekiguchi, N., Takagi, K., Dezawa, S., Nishikawa, H., & Sakuma, T. (1984). Development of metal-enclosed switchgear equipped with automatic dehumidifier. *Power Apparatus and Systems, IEEE Transactions on*, *PAS-103*(10), 3014 -3020. doi: 10.1109/TPAS.1984.318305

Mazouzi, S., Guessoum, Z., Michel, F., & Batouche, M. (2008). An agent-based approach for range image segmentation. In N. Jamali, P. Scerri, & T. Sugawara (Eds.), *Massively multi-agent technology* (pp. 146–161). Berlin, Heidelberg: Springer-Verlag.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133.

McCune, B. (1997). Influence of noisy environmental data on canonical correspondence analysis. *Ecology*, *78*(8), 2617–2623.

McMillan, N. J., Holland, D. M., Morara, M., & Feng, J. (2010). Combining numerical model output and particulate data using bayesian space–time modeling. *Environmetrics*, *21*(1), 48–65.

Melgani, F., & Bazi, Y. (2006). Markovian fusion approach to robust unsupervised change detection in remotely sensed imagery. *Geoscience and Remote Sensing Letters, IEEE*, *3*(4), 457-461.

Menzel, R., & Goschnick, J. (2000). Gradient gas sensor microarrays for on-line process controla new dynamic classification model for fast and reliable air quality assessment. *Sensors and Actuators B: Chemical*, *68*(1), 115–122.

Mishra, P., Srivastava, N., Shukla, K., & Singhal, A. (2011). Agent based image segmentation methods: A review. *Interational Journal Comp. Tech. Appl.*, *2*(3).

Mitchell, T. M. (1999). Machine learning and data mining. *Commun. ACM*, *42*(11), 30–36.

Mohandes, M., Halawani, T., Rehman, S., & Hussain, A. A. (2004). Support vector machines for wind speed prediction. *Renewable Energy*, *29*(6), 939–947.

Moser, G., & Serpico, S. (2009). Unsupervised change detection from multichannel sar data by markovian data fusion. *Geoscience and Remote Sensing, IEEE*

Transactions on, *47*(7), 2114 -2128.

Moser, G., Serpico, S., & Vernazza, G. (2007). Unsupervised change detection from multichannel sar images. *Geoscience and Remote Sensing Letters, IEEE*, *4*(2), 278-282.

Moustris, K. P., Ziomas, I. C., & Paliatsos, A. G. (2010). 3-day-ahead forecasting of regional pollution index for the pollutants no2, co, so2, and o3 using artificial neural networks in athens, greece. *Water, Air, & Soil Pollution*, *209*(1-4), 29–43.

Mukai, Y., Sugimura, T., Watanabe, H., & Wakamori, K. (1987). Extraction of areas infested by pine bark beetle using landsat mss data. *Photogrammetric engineering and remote sensing*, *53*(1), 77-81.

Nazelle, A. d., Arunachalam, S., & Serre, M. L. (2010). Bayesian maximum entropy integration of ozone observations and model predictions: an application for attainment demonstration in north carolina. *Environmental science & technology*, *44*(15), 5707–5713.

Nelson, D., Shorter, J., McManus, J., & Zahniser, M. (2002). Sub-part-per-billion detection of nitric oxide in air using a thermoelectrically cooled mid-infrared quantum cascade laser spectrometer. *Applied Physics B: Lasers and Optics*, *75*, 343-350. (10.1007/s00340-002-0979-4)

Nemmour, H., & Chibani, Y. (2006). Multiple support vector machines for land cover change detection: An application for mapping urban extensions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *61*(2), 125 - 133.

Niedermeier, A., Romaneessen, E., & Lehner, S. (2000). Detection of coastlines in sar images using wavelet methods. *Geoscience and Remote Sensing, IEEE Transactions on*, *38*(5), 2270–2281.

Nollet, L. M., & De Gelder, L. S. (2000). *Handbook of water analysis* (Vol. 102). CRC press.

Norman, Geoffrey, R., Sloan, Jeff, A., Wyrwich, & Kathleen, W. (2003). *Interpretation of changes in health-related quality of life the remarkable universality of half a standard deviation* (Vol. 41). Hagerstown, MD, ETATS-UNIS: Lippincott. (5 Medical care)

Olivares, G., Longley, I., & Coulson, G. (n.d.). Development of a low-cost device for observing indoor particle levels associated with source activities in the home.

Opelt, A., Pinz, A., & Zisserman, A. (2006, june). Incremental learning of object detectors using a visual shape alphabet. In *Computer vision and pattern recognition, 2006 ieee computer society conference on* (Vol. 1, p. 3 - 10). doi: 10.1109/CVPR.2006.153

Ordieres, J., Vergara, E., Capuz, R., & Salazar, R. (2005). Neural network prediction model for fine particulate matter (pm¡2.5) on the us–mexico border in el paso (texas) and ciudad juárez (chihuahua). *Environmental Modelling & Software*,

20(5), 547–559.

Paavola, M., Ruusunen, M., & Pirttimaa, M. (2005). *Some change detection and time-series forecasting algorithms for an electronics manufacturing process.* University of Oulu. (Oulu University Library OAI repository Test [http://herkules.oulu.fi/OAI/OAI-script.xsql] (Finland) ER)

Pacyna, E. G., Pacyna, J., Sundseth, K., Munthe, J., Kindbom, K., Wilson, S., ... Maxson, P. (2010). Global emission of mercury to the atmosphere from anthropogenic sources in 2005 and projections to 2020. *Atmospheric Environment*, *44*(20), 2487–2499.

Pang, S., Ozawa, S., & Kasabov, N. (2005, oct.). Incremental linear discriminant analysis for classification of data streams. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *35*(5), 905 -914. doi: 10.1109/TSMCB.2005.847744

Pang, S., Song, L., & Kasabov, N. (2011). Correlation-aided support vector regression for forex time series prediction. *Neural Computing & Applications*, *20*, 1193-1203. (10.1007/s00521-010-0482-5)

Parrella, F. (2007). Online support vector regression. *Master's Thesis, Department of Information Science, University of Genoa, Italy*.

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, *36*(4), 859–866.

Patten, D. M. (1992). Intra-industry environmental disclosures in response to the alaskan oil spill: a note on legitimacy theory. *Accounting, Organizations and Society*, *17*(5), 471–475.

Pearson, K. (1897). Mathematical contributions to the theory of evolution.–on a form of spurious correlation which may arise when indices are used in the measurement of organs. *The Royal Society*, 489-498.

Pearson, W., Davidson, E., & Britten, R. (1977). A program for least squares analysis of reassociation and hybridization data. *Nucleic acids research*, *4*(6), 1727–1738.

Percy, K. E., & Ferretti, M. (2004). Air pollution and forest health: toward new monitoring concepts. *Environmental Pollution*, *130*(1), 113 - 126. (Effects of Air Pollution on the Central and Eastern European Mountain Forests)

Perez, P., & Reyes, J. (2002). Prediction of maximum of 24-h average of pm10 concentrations 30h in advance in santiago, chile. *Atmospheric Environment*, *36*(28), 4555–4561.

Pérez, P., Trier, A., & Reyes, J. (2000). Prediction of pm¡ sub¿ 2.5¡/sub¿ concentrations several hours in advance using neural networks in santiago, chile. *Atmospheric Environment*, *34*(8), 1189–1196.

Phua, M.-H., Tsuyuki, S., Furuya, N., & Lee, J. S. (2008). Detecting deforestation with a spectral change detection approach using multitemporal landsat data: A case study of kinabalu park, sabah, malaysia. *Journal of Environmental Management*, *88*(4), 784 - 795.

Polhemus, N., & Cakmak, A. (1981). Simulation of earthquake ground motions using autoregressive moving average (arma) models. *Earthquake Engineering & Structural Dynamics*, *9*(4), 343–354.

Polikar, R., Upda, L., Upda, S., & Honavar, V. (2001, November). Learn++: an incremental learning algorithm for supervised neural networks. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *31*(4), 497 -508. doi: 10.1109/5326.983933

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA: The Journal of the American Medical Association*, *287*(9), 1132-1141.

Pope, C. A., et al. (2004). Ambient particulate air pollution, heart rate variability, and blood markers of inflammation in a panel of elderly subjects. *Environmental health perspectives*, *112*(3), 339.

Prewitt, J. M. (1970). *Object enhancement and extraction* (Vol. 75). Academic Press, New York.

Prybutok, V. R., Yi, J., & Mitchell, D. (2000). Comparison of neural network models with arima and regression models for prediction of houston's daily maximum ozone concentrations. *European Journal of Operational Research*, *122*(1), 31–40.

QI, P. (2003). Toward large arrays of multiplex functionalized carbon nanotube sensors for highly sensitive and selective molecular detection. *Nano Lett.*, *3*, 347-351. Retrieved from `http://ci.nii.ac.jp/naid/80015901366/en/`

Radke, R., Andra, S., Al-Kofahi, O., & Roysam, B. (2005). Image change detection algorithms: a systematic survey. *Image Processing, IEEE Transactions on*, *14*(3), 294 -307. doi: 10.1109/TIP.2004.838698

Radke, R. J., Andra, S., Al-Kofahi, O., & Roysam, B. (2005). Image change detection algorithms: a systematic survey. *Image Processing, IEEE Transactions on*, *14*(3), 294-307.

Raes, J., Foerstner, K. U., & Bork, P. (2007). Get the most out of your metagenome: computational analysis of environmental sequence data. *Current Opinion in Microbiology*, *10*(5), 490 - 498. (¡ce:title¿Antimicrobials/Genomics¡/ce:title¿) doi: 10.1016/j.mib.2007.09.001

Rangel, T. F. L. V. B., Diniz-Filho, J. A. F., & Bini, L. M. (2006). Towards an integrated computational tool for spatial analysis in macroecology and

biogeography. *Global Ecology and Biogeography*, *15*(4), 321–327. doi: 10.1111/j.1466-822X.2006.00237.x

Rasinmaki, J. (2003). Modelling spatio-temporal environmental data. *Environmental Modelling & Software*, *18*(10), 877 - 886. (Integrating Environmental Modelling and GI-Technology)

Rasmussen, D., Fiore, A., Naik, V., Horowitz, L., McGinnis, S., & Schultz, M. (2012). Surface ozone-temperature relationships in the eastern us: A monthly climatology for evaluating chemistry-climate models. *Atmospheric Environment*, *47*(0), 142 - 153. Retrieved from http://www.sciencedirect.com/science/article/pii/S1352231011012040 doi: 10.1016/j.atmosenv.2011.11.021

Ratle, F., Kanevski, M., Terrettaz-Zufferey, A.-L., Esseiva, P., & Ribaux, O. (2007). A comparison of one-class classifiers for novelty detection in forensic case data. In H. Yin, P. Tino, E. Corchado, W. Byrne, & X. Yao (Eds.), *Intelligent data engineering and automated learning - ideal 2007* (Vol. 4881, p. 67-76). Springer Berlin / Heidelberg.

Rebenitsch, L., Owen, C. B., Ferrydiansyah, R., Bohil, C., & Biocca, F. (2010). An exploration of real-time environmental interventions for care of dementia patients in assistive living. In *Proceedings of the 3rd international conference on pervasive technologies related to assistive environments* (pp. 34:1–34:8). New York, NY, USA: ACM.

Risk assessment guidance for superfund (rags). (2009). Waste and Cleanup Risk Assessment. Retrieved from http://www.epa.gov/oswer/riskassessment/ragsa/index.htm

Roberts, L. G. (1963). *Machine perception of three-dimensional solids* (Tech. Rep.). DTIC Document.

Robertson, A. (1959). The sampling variance of the genetic correlation coefficient. *Biometrics*, *15*(3), 469-485. (ArticleType: research-article / Full publication date: Sep., 1959 / Copyright ? 1959 International Biometric Society)

Robin, J. P., & Denis, V. (1999). Squid stock fluctuations and water temperature: temporal analysis of english channel loliginidae. *Journal of Applied Ecology*, *36*(1), 101–110.

Rosin, P. L., & Ioannidis, E. (2003). Evaluation of global image thresholding for change detection. *Pattern Recognition Letters*, *24*(14), 2345-2356.

Rossi, T. M., & Braun, J. E. (1997). A statistical, rule-based fault detection and diagnostic method for vapor compression air conditioners. *HVAC&R Research*, *3*(1), 19 - 37.

Russell, H., Sampson, J. S., Schmid, G. P., Wilkinson, H. W., & Plikaytis, B. (1984). Enzyme-linked immunosorbent assay and indirect immunofluorescence

assay for lyme disease. *Journal of Infectious Diseases*, *149*(3), 465-470.

Ryoo, M. S., & Aggarwal, J. K. (2009). Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *Computer vision, 2009 ieee 12th international conference on* (pp. 1593–1600).

Sahali, D., Pawlak, A., Valanciute, A., Grimbert, P., Lang, P., Remy, P., ... Guellen, G. (2002). A novel approach to investigation of the pathogenesis of active minimal-change nephrotic syndrome using subtracted cdna library screening. *Journal of the American Society of Nephrology*, *13*(5), 1238-1247.

Saide, P. E., Carmichael, G. R., Spak, S. N., Gallardo, L., Osses, A. E., Mena-Carrasco, M. A., & Pagowski, M. (2011). Forecasting urban pm10 and pm2. 5 pollution episodes in very stable nocturnal conditions and complex terrain using wrf–chem co tracer model. *Atmospheric Environment*, *45*(16), 2769–2780.

Salathé Jr, E. P., & Hartmann, D. L. (1997). A trajectory analysis of tropical upper-tropospheric moisture and convection. *Journal of Climate*, *10*(10), 2533–2547.

Salazar-Ruiz, E., Ordieres, J., Vergara, E., & Capuz-Rizo, S. (2008). Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence-based models in mexicali, baja california (mexico) and calexico, california (us). *Environmental Modelling & Software*, *23*(8), 1056–1069.

Salmon, B., Olivier, J., Wessels, K., Kleynhans, W., van den Bergh, F., & Steenkamp, K. (2011, june). Unsupervised land cover change detection: Meaningful sequential time series analysis. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *4*(2), 327 -335. doi: 10.1109/JSTARS.2010.2053918

Schmalensee, R., Joskow, P. L., Ellerman, A. D., Montero, J. P., & Bailey, E. M. (1998). An interim evaluation of sulfur dioxide emissions trading. *The Journal of Economic Perspectives*, *12*(3), 53–68.

Schmidler, S. C., Liu, J. S., & Brutlag, D. L. (2000). Bayesian segmentation of protein secondary structure. *Journal of computational biology*, *7*(1-2), 233–248.

Schmidt, R. A., & Wulf, G. (1997). Continuous concurrent feedback degrades skill learning: Implications for training and simulation. *Human Factors*, *39*(4), 509–525.

Scholkopf, B., Burges, C. J. C., & Smola, A. J. (1999). *Advances in kernel methods - support vector learning*. The MIT Press.

Schowengerdt, R. A. (1983). Techniques for image processing and classification in remote sensing. *New York: Academic Press*, 249.

Schwartz, J. (1999). Air pollution and hospital admissions for heart disease in eight

us counties. *Epidemiology*, *10*(1), 17–22.

Schwartz, J., & Marcus, A. (1990). Mortality and air pollution j london: a time series analysis. *American journal of epidemiology*, *131*(1), 185–194.

Schwender, H., Zucknick, M., Ickstadt, K., & Bolt, H. M. (2004). A pilot study on the application of statistical classification procedures to molecular epidemiological data. *Toxicology Letters*, *151*(1), 291 - 299. (Festschrift dedicated to Christian Hodel)

Scrimger, J. A. (1985). Underwater noise caused by precipitation. *The Journal of the Acoustical Society of America*, *78*, S2.

Senawongse, P., Dalby, A. R., & Yang, Z. R. (2005). Predicting the phosphorylation sites using hidden markov models and machine learning methods. *Journal of chemical information and modeling*, *45*(4), 1147–1152.

Sesnie, S. E., Gessler, P. E., Finegan, B., & Thessler, S. (2008). Integrating landsat tm and srtm-dem derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment*, *112*(5), 2145 - 2159. (Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue)

Shankar, B., Meher, S., & Ghosh, A. (2011). Wavelet-fuzzy hybridization: Feature-extraction and land-cover classification of remote sensing images. *Applied Soft Computing*, *11*, 2999–3011.

Shepherd, J. M., Pierce, H., & Negri, A. J. (2002). Rainfall modification by major urban areas: Observations from spaceborne rain radar on the trmm satellite. *Journal of Applied Meteorology*, *41*(7), 689–701.

Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., & Murthy, K. R. K. (2000). Improvements to the smo algorithm for svm regression. *Neural Networks, IEEE Transactions on*, *11*(5), 1188–1193.

Singh, A. (1989). Review article digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, *10*(6), 989 - 1003.

Soil pollution. (2011). Pollution Issues. Retrieved from `http://www.pollutionissues.com/Re-Sy/Soil-Pollution.html`

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using landsat tm data: When and how to correct atmospheric effects? *Remote Sensing of Environment*, *75*(2), 230 - 244.

Song, Y., Zhang, Y., Xie, S., Zeng, L., Zheng, M., Salmon, L. G., ... Slanina, S. (2006). Source apportionment of pm2. 5 in beijing by positive matrix factorization. *Atmospheric Environment*, *40*(8), 1526–1537.

Sotomayor-Olmedo, A., Aceves-Fernandez, M. A., Gorrostieta-Hurtado, E., Pedraza-Ortega, J. C., Vargas-Soto, J. E., Ramos-Arreguin, J. M., & Villaseñor-Carillo, U. (2011). Evaluating trends of airborne contaminants by

using support vector regression techniques. In *Electrical communications and computers (conielecomp), 2011 21st international conference on* (pp. 137–141).

Sousa, S., Martins, F., Alvim-Ferraz, M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, *22*(1), 97–103.

Spengler, J. D., & Chen, Q. (2000). Indoor air quality factors in designing a healthy building. *Annual Review of Energy and the Environment*, *25*(1), 567–600.

Statheropoulos, M., Vassiliadis, N., & Pappa, A. (1998). Principal component and canonical correlation analysis for examining air pollution and meteorological data. *Atmospheric Environment*, *32*(6), 1087–1095.

Tang, J. (2010, april). A color image segmentation algorithm based on region growing. In *Computer engineering and technology (iccet), 2010 2nd international conference on* (Vol. 6, p. V6-634 -V6-637). doi: 10.1109/ICCET.2010.5486012

Taylor, J. W., McSharry, P. E., & Buizza, R. (2009). Wind power density forecasting using ensemble predictions and time series models. *Energy Conversion, IEEE Transactions on*, *24*(3), 775–782.

Terry, D., Goldberg, D., Nichols, D., & Oki, B. (1992). *Continuous queries over append-only databases* (Vol. 21) (No. 2). ACM.

Thoss, M., Wang, H., & Miller, W. H. (2001). Generalized forward–backward initial value representation for the calculation of correlation functions in complex systems. *The Journal of Chemical Physics*, *114*, 9220.

Tian, Y., Feris, R., Liu, H., Hampapur, A., & Sun, M.-T. (2011, sept.). Robust detection of abandoned and removed objects in complex surveillance videos. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *41*(5), 565 -576. doi: 10.1109/TSMCC.2010.2065803

Titov, V. V., Gonzalez, F. I., Bernard, E., Eble, M. C., Mofjeld, H. O., Newman, J. C., & Venturato, A. J. (2005). Real-time tsunami forecasting: Challenges and solutions. In *Developing tsunami-resilient communities* (pp. 41–58). Springer.

Touloumi, G., Pocock, S., Katsouyanni, K., & Trichopoulos, D. (1994). Short-term effects of air pollution on daily mortality in athens: a time-series analysis. *International journal of epidemiology*, *23*(5), 957–967.

Treitz, P., & Rogan, J. (2004). Remote sensing for mapping and monitoring land-cover and land-use changean introduction. *Progress in Planning*, *61*, 269–279.

Turner, M. R. (1986). Texture discrimination by gabor functions. *Biological Cybernetics*, *55*(2-3), 71–82.

Van Lint, J. (2008). Online learning solutions for freeway travel time prediction. *Intelligent Transportation Systems, IEEE Transactions on*, *9*(1), 38–47.

Vapnik, V. N. (1999). An overview of statistical learning theory. *Neural Networks, IEEE Transactions on*, *10*(5), 988-999.

Vardoulakis, S., Fisher, B. E., Pericleous, K., & Gonzalez-Flesca, N. (2003). Modelling air quality in street canyons: a review. *Atmospheric environment*, *37*(2), 155–182.

Vasicek, Z., & Bidlo, M. (2011, june). Evolutionary design of robust noise-specific image filters. In *Evolutionary computation (cec), 2011 ieee congress on* (p. 269 -276). doi: 10.1109/CEC.2011.5949628

Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010). Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, *114*(12), 2970 - 2980. doi: 10.1016/j.rse.2010.08.003

Vitousek, P., D'Antonio, C., Loope, L., & Westbrooks, R. (1996, September). Biological Invasions as Global Environmental Change. *American Scientist*, *84*, 468-478.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2007). Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. *Computer-Aided Civil and Infrastructure Engineering*, *22*(5), 317–325.

Vu, T. T., Matsuoka, M., & Yamazaki, F. (2004, sept.). Lidar-based change detection of buildings in dense urban areas. In *Geoscience and remote sensing symposium, 2004. igarss '04. proceedings. 2004 ieee international* (Vol. 5, p. 3413 - 3416 vol.5).

Wang, K., & Stolfo, S. (2004). Anomalous payload-based network intrusion detection. In E. Jonsson, A. Valdes, & M. Almgren (Eds.), *Recent advances in intrusion detection* (Vol. 3224, p. 203-222). Springer Berlin / Heidelberg.

Wang, W., Ramkumar, S., Li, S., Wong, D., Iyer, M., Sakadjian, B. B., ... Fan, L.-S. (2010). Subpilot demonstration of the carbonation- calcination reaction (ccr) process: High-temperature co2 and sulfur capture from coal-fired power plants. *Industrial & Engineering Chemistry Research*, *49*(11), 5094–5101.

Wang, Y., Mitchell, B. R., Nugranad-Marzilli, J., Bonynge, G., Zhou, Y., & Shriver, G. (2009). Remote sensing of land-cover change and landscape context of the national parks: A case study of the northeast temperate network. *Remote Sensing of Environment*, *113*, 1453–1461.

Wang, Z., Jensen, J., & Im, J. (2010). An automatic region-based image segmentation algorithm for remote sensing applications. *Environmental Modelling and Software*, *25*, 1149–1165.

Wang, Z., Luo, P., Cheng, L., Zhang, S., & Shen, J. (2011). Hapten–antibody recognition studies in competitive immunoassay of α-zearalanol analogs by computational chemistry and pearson correlation analysis. *Journal of*

Molecular Recognition, *24*(5), 815–823.

Wang, Z., & Zhang, D. (1999, jan). Progressive switching median filter for the removal of impulse noise from highly corrupted images. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, *46*(1), 78 -80. doi: 10.1109/82.749102

Wannaz, E. D., Carreras, H. A., Rodriguez, J. H., & Pignata, M. L. (2012). Use of biomonitors for the identification of heavy metals emission sources. *Ecological Indicators*, *20*, 163–169.

Waser, L., Baltsavias, E., Ecker, K., Eisenbeiss, H., Feldmeyer-Christe, E., Ginzler, C., ... Zhang, L. (2008). Assessing changes of forest area and shrub encroachment in a mire ecosystem using digital surface models and cir aerial images. *Remote Sensing of Environment*, *112*, 1956–1968.

Watson, J. G., Chow, J. C., Chen, L.-W. A., Lowenthal, D. H., Fujita, E. M., Kuhns, H. D., ... others (2011). Particulate emission factors for mobile fossil fuel and biomass combustion sources. *Science of the Total Environment*, *409*(12), 2384–2396.

Wei, C., Rogers, W. J., & Mannan, M. (2006). Detection of autocatalytic decomposition behavior of energetic materials using aptac. *Journal of Thermal Analysis and Calorimetry*, *83*(1), 125-130. (10.1007/s10973-005-6982-3)

West, L. (2005). *World water day: A billion people worldwide lack safe drinking water.* About.com. Retrieved from `http://environment.about.com/od/environmentalevents/a/waterdayqa.htm`

WHO. (2002). *Estimated deaths & dalys attributable to selected environmental risk factors.* World Health Organization. Retrieved from `http://www.who.int/entity/quantifying_ehimpacts/countryprofilesebd.xls`

WHO. (2005). *Indoor air pollution and health.* World Health Organization. Retrieved from `http://www.who.int/mediacentre/factsheets/fs292/en/`

Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, *63*(11), 1309–1313.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research*, *30*(1), 79.

Wu, T., Xie, K., Song, G., & Hu, C. (2008). A multiple svr approach with time lags for traffic flow prediction. In *Intelligent transportation systems, 2008. itsc 2008. 11th international ieee conference on* (pp. 228–233).

Xian, G., Homer, C., & Fry, J. (2009). Updating the 2001 national land cover database land cover classifcation to 2006 by using landsat imagery change

detection methods. *Remote Sensing of Environment*, *113*, 1133–1147.

Xing, H., Hatton, A., & Awbi, H. (2001). A study of the air quality in the breathing zone in a room with displacement ventilation. *Building and environment*, *36*(7), 809–820.

Xue, G.-R., Lin, C., Yang, Q., Xi, W., Zeng, H.-J., Yu, Y., & Chen, Z. (2005). Scalable collaborative filtering using cluster-based smoothing. In *Proceedings of the 28th annual international acm sigir conference on research and development in information retrieval* (pp. 114–121).

Yakimovsky, Y. (1976). Boundary and object detection in real world images. *J. ACM*, *23*, 599–618.

Yu, B., Hu, Z., Liu, M., Yang, H., Kong, Q., & Liu, Y. (2009). Review of research on air-conditioning systems and indoor air quality control for human health. *International journal of refrigeration*, *32*(1), 3–20.

Yu, T.-C., Lin, C.-C., Lee, R.-G., Tseng, C.-H., & Liu, S.-P. (2012). Wireless sensing system for prediction indoor air quality. In *High speed intelligent communication forum (hsic), 2012 4th international* (pp. 1–3).

Yuanxun, Z., Yuanmao, Z., Yingsong, W., & Delu, L. (2006). Pixe characterization of pm10 and pm2. 5 particulate matter collected during the winter season in shanghai city. *Journal of radioanalytical and nuclear chemistry*, *267*(2), 497–499.

Zhan, X., Defries, R., Townshend, J. R. G., Dimiceli, C., Hansen, M., Huang, C., & Sohlberg, R. (2000). The 250 m global land cover change product from the moderate resolution imaging spectroradiometer of nasa's earth observing system. *International Journal of Remote Sensing*, *21*(6), 1433 - 1460.

Zhan, X., Sohlberg, R. A., Townshend, J. R. G., DiMiceli, C., Carroll, M. L., Eastman, J. C., … DeFries, R. S. (2002). Detection of land cover changes using modis 250 m data. *Remote Sensing of Environment*, *83*(1-2), 336 - 350.

Zhang, S., Xie, C., Zeng, D., Zhang, Q., Li, H., & Bi, Z. (2007). A feature extraction method and a sampling system for fast recognition of flammable liquids with a portable e-nose. *Sensors and Actuators B: Chemical*, *124*(2), 437–443.

Zhang, W., & Bergholm, F. (1997). Multi-scale blur estimation and edge type classification for scene analysis. *International Journal of Computer Vision*, *24*(3), 219–250.

Zhang, X., Zwiers, F. W., Hegerl, G. C., Lambert, F. H., Gillett, N. P., Solomon, S., … Nozawa, T. (2007). Detection of human influence on twentieth-century precipitation trends. *Nature*, *448*(7152), 461-465. (10.1038/nature06025)

Zhou, C., Luo, J., Yang, C., Li, B., & Wang, S. (2000). Flood monitoring using multi-temporal avhrr and radarsat imagery. *PE&RS - Photogrammetric Engineering & Remote Sensing*, *66*(5), 633-638.

Zhuang, D., & Wang, S. (2010, oct.). Content-based image retrieval based on

integrating region segmentation and relevance feedback. In *Multimedia technology (icmt), 2010 international conference on* (p. 1 -3). doi: 10.1109/ICMULT.2010.5630973

Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M., & Furlong, E. E. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature*, *462*(7269), 65–70.