Utility of Predictive Policing: Evaluating a Spatial-Temporal Forecasting Model in an Operational Deployment

Ryan G. Prox, O.O.M.

B.A., Simon Fraser University, 1996
M.A., Simon Fraser University, 2007

Doctor of Law Enforcement and Security

Australian Graduate School of Policing and Security
Charles Sturt University

September 13th, 2019
STATEMENT OF AUTHORSHIP

This dissertation is submitted as partial fulfilment of the requirements for the degree of Doctor of Policing and Security by course work. I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma at Charles Sturt University or any other educational institution, except where due acknowledgment is made in the dissertation. Any contribution made to the research by colleagues with whom I have worked at Charles Sturt University or elsewhere during my candidature is fully acknowledged.

I agree that this dissertation be accessible for the purpose of study and research in accordance with the normal conditions established by the Executive Director, Library Services or nominee, for the care, loan and reproduction of theses.

............................................................

Ryan G. Prox
ACKNOWLEDGEMENTS

A sincere thank you to my senior supervisor Dr. Patrick Walsh for encouraging me when I needed it and for helping me to navigate through the challenges of the thesis process. My deepest gratitude to Chief Constable Adam Palmer for supporting my research, inspiring creative solutions, and sponsoring my critical examinations of policing, which were only possible through his kind encouragement. I am also grateful to Betty Ling, a champion of innovative geographic information system solutions, whose expertise proved invaluable during our thoughtful deliberations on the best approaches for processing the data. A special thank you to Dr. Lauren Scott Griffin, the leading Spatial Analysis Engineer from Esri Inc., for her suggestions on how to analyze and interpret the results and for her patience in providing assistance on how to best apply geospatial statistical tools to a very complex data sample. My heartfelt appreciation to Dr. Chisen Goto Prox, my loving wife, for her enduring support, tireless work reviewing countless drafts, and for helping make this research possible.
ABSTRACT

The advent of big data has fuelled a machine-learning revolution, which in turn has included the emergence of predictive policing by law enforcement agencies, although little is known about the efficacy of this strategy and the conditions under which it can be effective in reducing levels of crime. The study involved an evaluation of a predictive policing pilot project implemented by the Vancouver Police Department in British Columbia, Canada, and focused on the use of a machine-learning system designed to conduct spatial-temporal crime forecasting on residential break and enters. It was also designed to contribute to the published literature on predictive policing and to offer guidance for police services considering adoption of this technology, by providing a template that could be used to assess the strategy’s effectiveness. The evaluation was conducted through the use of a pilot project that extended over 6 months from April to September 2016, during which time patrol resources were deployed to specific locations, at particular times of the day, as determined by a predictive policing model.

The effectiveness of the deployments in reducing residential break and enter rates throughout this time period was compared to a control period during which police patrols were not directed by the predictive model. A multimodal approach that included the use of geo-temporal data analysis techniques to evaluate the distribution, intensity of patterns, and volume of residential break and enters during the 6-month pilot project was compared to data obtained during the previous 4 years, to determine whether directed police resources based on the predictions, had an effect. The evaluation indicated the pilot project had a quantifiable effect in 4 out of the 6 months, with inconclusive findings for the remaining. The role of big data, machine learning, and artificial intelligence in forecasting, as well as its potential benefits and concerns were also discussed, with a focus on data integrity, ethics, and the human components of predictive policing.
# CONTENTS

STATEMENT OF AUTHORSHIP ................................................................. 2  
ACKNOWLEDGEMENTS ............................................................................... 3  
ABSTRACT ............................................................................................................. 4  
CONTENTS ................................................................................................................... 5  
LIST OF FIGURES ....................................................................................................... 8  
LIST OF TABLES .......................................................................................................... 10  
LIST OF ACRONYMS ................................................................................................. 11  
CHAPTER 1 — INTRODUCTION .............................................................................. 12  
  BACKGROUND ........................................................................................................... 12  
  RATIONALE ............................................................................................................... 13  
  THEORETICAL BASE ................................................................................................. 14  
  RESEARCH OBJECTIVES ......................................................................................... 15  
  APPROACH OF THE STUDY ..................................................................................... 15  
  OUTLINE OF THE STUDY ......................................................................................... 16  
CHAPTER 2 — LITERATURE REVIEW ...................................................................... 17  
  Evaluation Research ................................................................................................. 17  
  Crime Analysis ......................................................................................................... 20  
  Geospatial Analysis ................................................................................................. 22  
  Geospatial-Temporal Data Analysis Techniques ..................................................... 25  
  Artificial Intelligence, Big Data, and Data Quality ................................................... 30  
  Predictive Policing Trends ....................................................................................... 38  
    Defining Predictive Policing .................................................................................. 38  
    Criminological Theories on Predictive Policing ...................................................... 44  
  Predictive Policing Studies ...................................................................................... 49  
  Ethical Concerns in Predictive Policing ................................................................... 55  
  Data Validity Concerns ......................................................................................... 59  
  Human Rights Concerns ......................................................................................... 64  
  Residential Burglary Definition ............................................................................... 68  
  Environmental Criminology Theories on Property Crimes .................................... 69  
  Policing and Technological Adoption ...................................................................... 74  
CHAPTER 3 — METHODOLOGY ............................................................................. 79  
  Methodological Approach ...................................................................................... 79  
  Data Ethics Consideration ...................................................................................... 79
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Example of 100-m Forecast Area</td>
<td>84</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Example of 500-m Forecast Area</td>
<td>85</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Illustration of Typical 100-m Patrol Area</td>
<td>88</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Illustration of Typical 500-m Patrol Area</td>
<td>90</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Fishnet of COV with 100-m grid</td>
<td>97</td>
</tr>
<tr>
<td>Figure 6</td>
<td>3D Space-Time Data Cube</td>
<td>99</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Residential Burglaries by Lower Mainland Region Jurisdictions for 2015</td>
<td>108</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Residential Burglaries by Lower Mainland Region Jurisdictions for 2016</td>
<td>109</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Typical Monthly Residential Break and Enter Temporal Distribution</td>
<td>110</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Break and Enter Residential Comparison between 2016 and 4-Year Average</td>
<td>114</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Break and Enter Residential Change Detection April 2016 to April 4-Year Average</td>
<td>116</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Break and Enter Residential Change Detection May 2016 to May 4-Year Average</td>
<td>117</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Break and Enter Residential Change Detection June 2016 to June 4-Year Average</td>
<td>118</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Break and Enter Residential Change Detection July 2016 to July 4-Year Average</td>
<td>119</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Break and Enter Residential Change Detection August 2016 to August 4-Year Average</td>
<td>120</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Break and Enter Residential Change Detection September 2016 to September 4-Year Average</td>
<td>121</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Spatial Autocorrelation Report for Change Detection April to September</td>
<td>123</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Break and Enter Distribution Comparison 2016 to 4-Year Average</td>
<td>125</td>
</tr>
<tr>
<td>Figure 19</td>
<td>Residential Break and Enter High-Intensity Level Comparison 2016 to 4-Year Average by Month</td>
<td>127</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Residential Break and Enter Medium-Intensity Level Comparison 2016 to 4-Year Average by Month</td>
<td>128</td>
</tr>
<tr>
<td>Figure 21</td>
<td>Residential Break and Enter Low-Intensity Level Comparison 2016 to 4-Year Average by Month</td>
<td>128</td>
</tr>
<tr>
<td>Figure 22</td>
<td>April 2016 Residential Break and Enters between 1601 to 0759 hours with No Resource Allocation to Treatment Area</td>
<td>130</td>
</tr>
<tr>
<td>Figure 23</td>
<td>May 2016 Residential Break and Enters between 1601 and 0759 hours with No Resource Allocation to Treatment Area</td>
<td>131</td>
</tr>
</tbody>
</table>
Figure 24 June 2016 Residential Break and Enters between 1601 to 0759 hours with No Resource Allocation to Treatment Area .......................................................... 132
Figure 25 July 2016 Residential Break and Enters between 1601 to 0759 hours with No Resource Allocation to Treatment Area .......................................................... 133
Figure 26 August 2016 Residential Break and Enters between 1601 to 0759 hours with No Resource Allocation to Treatment Area .................................................. 134
Figure 27 September 2016 Residential Break and Enters between 1601 to 0759 hours with No Resource Allocation to Treatment Area ........................................ 135
Figure 28 Emerging Hot Spot Analysis 2016 April to September – 2-Week Interval Aggregate Cube ..................................................................................................... 139
Figure 29 Visualized 3D Space Time Data Cube – April to September 2016......................... 140
Figure 30 Residential Burglaries by Lower Mainland Region Jurisdictions for 2015 ..........144
Figure 31 Residential Burglaries by Lower Mainland Region Jurisdictions for 2016 ..........145
Figure 32 Break and Enter Residential Comparison between 2016 and 4-Year Average .................................................................................................................. 147
Figure 33 Expanded Police Activity Zone .............................................................................. 152
LIST OF TABLES

Table 1  Methodology – Data Collection Processes ............................................................ 103
Table 2  Methodology – Data Analysis Techniques ............................................................ 103
Table 3  Percentage change in Residential Break and Enters by month compared to the previous 4-year average per month. .............................................................. 106
Table 4  Comparison of Statistical Measures and Tests for RBNE Counts ...................... 111
Table 5  Break and Enter Residential Counts ...................................................................... 112
Table 6  Monthly Residential Break and Enter Hot Spot Percentage Coverage ............... 122
Table 7  RBNE Surface Area Distribution Change ............................................................... 124
Table 8  Intensity Level Comparison 2016 to 4-Year Average by Month ......................... 126
Table 9  Geographic Accuracy of Prediction Boxes with No Resource Allocation ............ 136
Table 10 Emerging Hot Spot Categories .............................................................................. 137
Table 11 Comparison of Statistical Measures and Tests ..................................................... 146
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>Three-dimensional representation</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial intelligence</td>
</tr>
<tr>
<td>ArcGIS</td>
<td>An Esri tool for compiling geographic data and performing spatial analysis</td>
</tr>
<tr>
<td>BC</td>
<td>Province of British Columbia</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer-aided dispatch system</td>
</tr>
<tr>
<td>CompStat</td>
<td>CompStat computer comparison statistics</td>
</tr>
<tr>
<td>COV</td>
<td>City of Vancouver</td>
</tr>
<tr>
<td>CSP</td>
<td>Community safety personnel</td>
</tr>
<tr>
<td>ESRI</td>
<td>Environmental Systems Research Institute</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>FOIPPA</td>
<td>Freedom of Information and Protection of Privacy Act</td>
</tr>
<tr>
<td>GDPR</td>
<td>General Data Protection Regulation</td>
</tr>
<tr>
<td>GWR</td>
<td>Geographical weighted regression</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic information systems</td>
</tr>
<tr>
<td>GO</td>
<td>General occurrence report</td>
</tr>
<tr>
<td>ILP</td>
<td>Intelligence-led policing</td>
</tr>
<tr>
<td>KDE</td>
<td>Kernel density estimation</td>
</tr>
<tr>
<td>LAPD</td>
<td>Los Angeles Police Department</td>
</tr>
<tr>
<td>NASEM</td>
<td>National Academies of Sciences, Engineering, and Medicine</td>
</tr>
<tr>
<td>NIJ</td>
<td>National Justice Institute</td>
</tr>
<tr>
<td>OSL</td>
<td>Ordinary least squares</td>
</tr>
<tr>
<td>PCSO</td>
<td>Police community support officer</td>
</tr>
<tr>
<td>PreCobs</td>
<td>Pre-crime observation system</td>
</tr>
<tr>
<td>PRIME-BC</td>
<td>Police Records Information Management Environment system used in BC</td>
</tr>
<tr>
<td>RBNE</td>
<td>Residential break and enter (residential burglary)</td>
</tr>
<tr>
<td>RMS</td>
<td>Records management system</td>
</tr>
<tr>
<td>RTM</td>
<td>Risk terrain modelling</td>
</tr>
<tr>
<td>SSL</td>
<td>Strategic subjects list</td>
</tr>
<tr>
<td>SLSC</td>
<td>Stop LAPD Spying Coalition</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>VPD</td>
<td>Vancouver Police Department</td>
</tr>
</tbody>
</table>
CHAPTER 1 — INTRODUCTION

This study examines the utility of predictive policing through an efficacy evaluation of a predictive policing pilot project implemented by the Vancouver Police Department in British Columbia (BC), Canada. The inquiry focused on the use of a machine-learning system designed to conduct spatial-temporal crime forecasting on residential break and enters (RBNE). This evaluation was conducted via the use of a 6-month pilot project that was carried out during the period from April to September 2016. During this time, patrol resources were deployed to specific locations and at particular times of the day, as determined by a predictive policing model. The effectiveness of police deployments in reducing or influencing the distribution and intensity of RBNEs was analyzed and the results were compared with expected norms with no intervention.

BACKGROUND

Artificial intelligence and machine learning are at the forefront of a technological revolution that affects virtually every facet of society. Artificial Intelligence (AI), is a broad concept of computers carrying out human cognitive functions, while machine learning is a computer science domain within AI in which the learning takes place by the machines through the provided data (Vempati, 2016). Therefore, AI is a broad overarching term related to technology emulating human tasks intelligently and machine learning is a subset discipline within AI that focuses on machines that can learn on its own without being explicitly programmed (Vempati, 2016). From the safety of the food supply chain to autonomous self-driving cars as well as police enforcement strategies, machine learning is permeating people’s daily lives with little oversight or accountability, and with limited understanding and awareness among the general public. The advent of big data has fuelled the machine-learning technical revolution, and this has impacted policing as well. The emergence of predictive policing as a crime prevention and crime interdiction strategy is illustrative of this trend.

Although many police agencies around the world have implemented, or plan to adopt, predictive policing, this trend has not been informed by sound empirical research, and the numerous challenges surrounding the analytics required to populate a predictive...
policing model have remained largely unaddressed. These include biases in the data as well as organizational predispositions in the use of the technology. This has contributed to various perspectives which suggest that predictive policing is a source of both optimism and concern within academia, law enforcement, and the public at large. These challenges are reflected in the literature and in recent government responses, including the creation by the European Union (EU) of the General Data Protection Regulation (GDPR, 2016).

Predictive policing, however, “is not fundamentally about making crime-related predictions. It is about implementing business processes … to take action to pre-empt the predicted crime” (Perry, McInnis, Price, Smith, & Hollywood, 2013, p. 128). As such, this study evaluated not only the application of predictive policing during a pilot project in one jurisdiction, but also the actions taken by the police in response to the predictions.

RATIONALE

In the policing community, predictive policing has been elevated as a means of achieving greater efficiencies and creating an environment in which police services can operate more effectively with limited resources (Prox & Griffiths, 2015). Police services have typically applied intelligence-led policing (ILP) practices, including the recent introduction of predictive policing technology, to support and enhance tactical and strategic operations (Peterson, 2005; Ratcliffe, 2008, 2016). However, despite the ubiquitous application of ILP within the policing environment, “this practice has traditionally been undertaken without adequate evaluation or the formation of predefined metrics to assess whether the outcomes have matched the intended objectives” (Azen, 1991; Perry et al., 2013; Peterson, 2005; Sargsyan & Prox, 2019, p. 31; Weisburd & Braga, 2006b). There is a limited number of publications on predictive policing applications, and there are even fewer that are methodologically-sound evaluations of the use of crime forecasting.

The primary rationale of this research was to address the gaps in the literature in three ways: (a) to provide a realistic assessment of the predictive policing landscape and the motivators driving this technological revolution; (b) to inform police agencies about realistic outcomes when adopting predictive policing, dependent on their data volume and technical capacity; and (c) to provide an actionable template for conducting an empirical evaluation of any predictive policing implementation.
The outcomes of this study may have broader implications for policing, in that future research on predictive policing might be expected to utilize the predefined metrics for assessment and criteria for objectivity and data analysis that are used here. By establishing a distinct methodology to research predictive policing outcomes, this study has addressed the current absence of established criteria and added to existing knowledge on predictive policing research. This information will assist police services by providing a replicable and evidence-based evaluative methodology. Currently, police services are generally unable to conduct evaluations of crime forecasting technology, due to the absence of an evaluative framework and an understanding of the metrics that are required to assess the efficacy of this strategy (Hunt, Saunders, & Hollywood, 2014; Short, Brantingham, Bertozzi, & Tita, 2010; Weisburd & Braga, 2006b).

**THEORETICAL BASE**

The research approaches applied in the study were all grounded in criminological theories and past research practices that sought to explain the relationships between offenders, victims, environment and other influencing factors, with geospatial analysis having a heavy influence on how these features were evaluated. An underlying premise of predictive policing is that the commission of a crime is the result of choice, rather than a consequence of other factors such as psychology or social learning. From this perspective, it is possible to predict crime premised on systematic, logic-based probabilities (Bachner, 2013; Bennett-Moses & Chan, 2018). This assumption is grounded in theories that speak to the predictability of crimes based on the actions of rational, logical individuals.

The classical school of criminological theories views individuals as rational beings, who examine the potential consequences when considering their actions, including committing crime (Kennedy, 2008). This theoretical perspective supports predictive policing principles, in that the actions of the potential offenders can be estimated and anticipated, rather than those explanations that view crime as not grounded in logic. Routine activities theory is also a key component of predictive policing, since police resources are deployed to forecasted locations with the intention that their presence as capable guardians acts as a deterrent (Cohen & Felson, 1979; Kennedy, 2008; Pollock, Joo, & Lawton, 2010).

The concept of predictive policing is also supported by criminological theories that assume the past can be used to make inferences about the future. Past criminal activities
can provide future locations of possible crime based on assumptions of routine activity, rational choice, crime pattern, and repeat victimization theories (Akers, 2013; Azen, 1991; Braga & Weisburd, 2010a; Brantingham & Brantingham, 1981; Clarke & Felson, 1993; Cohen & Felson, 1979; Kennedy, 2008; Perry et al., 2013). As there are various paradigms that can contribute to the explanation of predictive policing principles and applications, the “blended theory” (Perry et al., 2013, p. 3) of existing criminological theories has been used to examine how crimes can be predicted based on historical data and assumptions about individual motivations.

RESEARCH OBJECTIVES

The purpose of the research was to evaluate the efficacy of a machine-learning-based spatial-temporal crime forecasting model applied to RBNEs. More specifically, the quantitative study examined whether a reduction and change occurred in the distribution and intensity of RBNEs in the study area as a result of police interventions guided by the forecasting model. The hypothesis was that if the forecasting model was effective, it would predict the locations of RBNEs in the study area. In addition, based on the predictions, there would be a statistically significant change in the intensity and distribution of RBNEs, as well as an overall reduction in the number of offences committed within the study area. An ineffective or flawed forecasting model would be expected to have little influence on the distribution and levels of crime incidents when compared to historical data, which would be consistent with the theories that support crime prediction (Guerette & Bowers, 2009; Santos, 2016).

APPROACH OF THE STUDY

This quantitative study examined the outcomes of the predictive policing pilot project conducted by the Vancouver Police Department in BC, Canada. The distribution, intensity, and number of RBNEs during the pilot study that ran from April to September 2016 were compared to those from the previous 4 years, in order to evaluate the efficacy of a machine-learning system designed to conduct spatial-temporal crime forecasting on RBNEs based on trend analysis. A multimodal approach that included the use of geo-temporal data analysis techniques as well as summary and inferential statistics, was used to analyze the data. Two innovative geographic information systems (GIS) analysis techniques were applied to analyze the data quantitatively. Changes in crime patterns were calculated on kernel density.
The application of emerging trend analysis was calculated using a Mann-Kendall time-series test to detect any previously undetermined temporal relationships that were statistically significant. The approach to this quantitative evaluation study was also influenced by criminological theories developed by notable criminologists, such as Short et al. (2010) and Mohler, Short, Brantingham, Schoenberg, and Tita (2011), using geospatial analysis. Geospatial analysis is based upon criminological theory which asserts that crime can be predicted, and on the whole these methodologies have been quantitative. This is partly due to the nature of the statistical algorithms requiring big data sources, but is also because the theories originated in geography and seismology, which have been dominated by quantitative rather than qualitative approaches (Mohler, 2011; Short et al., 2010).

OUTLINE OF THE STUDY

This introductory chapter is followed by a literature review in Chapter 2, which sets out the framework for the key concepts used in the evaluation. Literature on evaluation research, crime analysis, and spatial-temporal data analysis is included to provide context for the methodology used in this study. This is followed by an examination of the key concepts of this study, including predictive policing and residential burglary. Once predictive policing has been defined, the criminological theories that support its application in operational policing are examined, followed by an exploration of big data, AI, and the adoption of new technologies with specific reference to policing. Previous studies and evaluations of predictive policing are presented in order to examine both the opportunities for the use of this policing strategy and the outcomes that have been achieved. Chapter 3 provides details of the data collected, and how the information was processed using various GIS analysis techniques. The results of the evaluation are provided in Chapter 4 using summary statistics, figures, tables, and charts that show distributions, changes, and trends in the evaluation outcome when compared to historical data and neighbouring jurisdictions. This is followed by a discussion in Chapter 5 on the implications of the findings of the study. To close the research, Chapter 6 highlights the features of the study that comprise a contribution to the literature on predictive policing, and sets out a number of recommendations for future research and for practitioners. In the next chapter, the literature review addresses the research question and objective, as well as the key terms of this study.
CHAPTER 2 — LITERATURE REVIEW

This chapter provides an overview and analysis of the literature relevant to the current research. It also locates and contextualizes the research questions within the existing literature and demonstrates how this inquiry has made a significant contribution to it. As multiple quantitative methods originating in various disciplines were used for this study, a review of the literature providing the methodological framework contextualizing the research question will be provided in the first part of this chapter. This is followed by the expansion of key terms and notable applications of predictive policing within the industry, with a focus on the critical issues and challenges being faced by law enforcement. The first part of the chapter focuses particularly on the following key knowledge areas as they relate to the methodology utilized in this study: evaluation research, crime analysis, geospatial analysis, and geospatial-temporal data analysis techniques. In the second part of the chapter, which relates to the context of the study, the literature review focuses on the following topics: AI, big data, and data quality; predictive policing definition and trends; predictive policing studies; ethical concerns in predictive policing; data validity concerns; human rights concerns; residential burglary definition; environmental criminology theories on property crimes; and policing and adoption of technology.

The purpose of this chapter is to provide context for the present study as well as to identify how it fills the gaps that currently exist in this field. The following section focuses on the literature that addresses the methodological aspects of the current research, which itself is an evaluation research that utilizes multiple quantitative methods originating in crime analysis as well as geospatial analysis.

Evaluation Research

Evaluation research, conducted for the purposes of assessing a specific program, is “essentially an information gathering and interpreting endeavor that attempts to answer a specified set of questions about a program’s performance and effectiveness” (Rossi, Freeman, & Lipsey, 2004, p. 62). Evaluation research has various purposes, including influencing decisions regarding the future of the program being assessed by gauging its effectiveness and accountability (Rossi et al., 2004). In addition, evaluation research can
contribute to substantive and methodological knowledge (Rossi et al., 2004). The current research is classified as an outcome evaluation, which can be referred to as summative, impact, or effectiveness evaluation (Royse, 2008). The primary goal of the outcome evaluation, as the name suggests, is to summarize and assess the effectiveness of a program. This process enables researchers to examine the relationship between the independent and dependent variables and to determine whether there was a cause and effect while also considering threats to validity (Lösel, 2008). While an accurate description of the program being evaluated is essential, an understanding of the relevant standards and criteria to conduct the assessment is equally important (Rossi et al., 2004). Understanding the program and the intended program outcomes contributes to the asking of relevant questions, thereby leading to appropriate data collection (Rossi et al., 2004). The types of data collected also influence statistics and analysis to support the evaluation (Newcomer & Conger, 2015).

In the current research, the relationship between the implementation of the predictive policing pilot by the Vancouver Police Department (VPD) and actual real-life RBNEs over the period of the study was examined. The evaluation method employed in the research answered questions concerning the outcome and impact of the pilot project implementation, using quantitative methodology. Quantitative methodology was chosen for this study to address the requirement for outcome and impact evaluation to “plan for collecting data that will permit a persuasive demonstration that observed changes are a function of the intervention and cannot readily be accounted for in other ways” (Rossi et al., 2004, p. 70). Although qualitative and quantitative methods can be used in combination for evaluation studies (i.e., mixed methods), the chosen approach needs to reflect both the question and the audience which the researcher wishes to address (Newcomer, Hatry, & Wholey, 2015). Since the VPD pilot project assessed in the current research was geographically based and since the predicted locations were generated by a forecasting model, a randomized controlled trial, which is standard in the natural sciences (Losel, 2008), was not a viable option for this study. Given that the jurisdiction and pilot study boundaries were already determined by the VPD, there was no randomness to the application of the forecasted locations. However, due to the fact that areas outside of the pilot and all other neighbouring jurisdictions were not part of the project implementation, the data from these
areas could be compared. This classified the current research as a quasi-experimental design, defined as “assessment design that tests the existence of a causal relationship where random assignment is not possible. Typical quasi-experimental designs include pre-post designs, comparison group designs and interrupted time-series designs” (Newcomer et al., 2015, p. 56).

Within evaluation research, randomized controlled trials have the advantage over other types of outcome evaluation in that most threats to internal validity can be ruled out and applied instead to measures of policing and situational crime prevention (Lösel, 2008, p. 147). In a review of empirical evidence on the effects of hot spot policing, Braga (2001) identified evaluation studies that fit the criteria, as established by his systematic review of the methods literature. In order for an evaluation to be assessed as having empirical evidence, the only studies that were selected were those with either randomized experimental or non-randomized quasi-experimental design that had before and after measures focusing on police enforcement in crime hot spots (Braga, 2001). The research design of the nine evaluation studies reviewed included an analysis that contrasted the number of calls for service or other indicators from the baseline time period with those of the experimental time period, a comparison between the control and experimental groups, and a trend analysis of pre- and post-time periods (Braga, 2001). This review of Braga’s (2001) evaluation studies was relevant to the current research project for several reasons. The review defines the criteria for studies that could be considered to have empirical evidence. This inquiry went beyond establishing a baseline to compare to the pilot period in order to incorporate neighbouring jurisdictions for the purposes of trend analysis. In addition, the studies reviewed were directly related to the intensity of police resources in a specific location. Broken down into the simplest components, the current research focused on measuring the effectiveness of police resource allocation to a specific location.

Lösel (2008) argued that sound experiments, including those incorporating pre- and post-comparisons of areas, should also consider issues such as crime displacement. The current research did use pre- and post-comparison of crime statistics while also addressing the possibility of displacement by examining spatial-temporal changes between the years, as well as conducting analyses on crime density, intensity, and distribution to strengthen the evaluation. One of the ways in which the current research contributes to the literature and
methodology of evaluation studies is that randomized experiments and quasi-random experiments are still rare in criminology (Farrington, 2003; Welsh, 2006). Indeed Lösel (2008) recommended that future evaluation research in criminology should include sound experiments on clearly circumscribed measures while “address[ing] broader framing conditions” (p. 152) and examining what happens, not only in the program being evaluated but also in the control conditions. In the current study, this was achieved by analyzing data from the surrounding jurisdictions and their trends, compared with the jurisdiction under evaluation.

This level of attention towards conducting an empirically sound evaluation study was relevant to this research, since a review of findings on police culture has concluded that general core values are shared within the organization: these include authoritarianism, suspicion, conservatism, and cynicism (Murphy & McKenna, 2007, p. 6). Individuals who hold these values are likely to be more receptive to information that can be independently validated and tested; therefore, a research methodology with these qualities has the greatest chance of being accepted. Given that the purpose of the current research was to assess the efficiency and effectiveness of the predictive model applied in an operational police setting, it was important for the research to resonate within the framework of the policymakers, senior managers, finance departments, and other levels of governance to ensure that the research findings would contribute to an evidence-based decision process and would then be utilized. Furthermore, the current research was conducted in direct response to the lack of evaluative processes within the literature. The key rationale for this research was to establish methodologies and predefined metrics for assessing predictive applications. It is hoped that future studies will utilize the methods used for analysis of predictive policing evaluations and will also gain insights from, and improve upon, some of the limitations of this study. The following section defines crime analysis and how it was used to support evidence-based outcomes in the current research.

**Crime Analysis**

The International Association of Crime Analysts (2014) defined crime analysis as follows:

A profession and process in which a set of quantitative and qualitative techniques are used to analyze data valuable to police agencies and their communities. It includes the analysis of crime and criminals, crime victims, disorder, quality of life
issues, traffic issues, and internal police operations, and its results support criminal investigation and prosecution, patrol activities, crime prevention and reduction strategies, problem solving and the evaluation of police efforts. (p. 2)

In order to adopt this definition of crime analysis within the current research, quantitative techniques were used in the study to analyze crime data for evaluating police effort. Decision making, efficiency, and accountability are supported by crime analysis through the use of statistics and crime mapping (Canter, 2000; Santos, 2016). The following descriptive statistics were used specifically in the current research: univariate statistics, which summarize a variable, and bivariate statistics, which describe the relationship between two variables (Newcomer & Conger, 2015). Percentages and averages of RBNEs during the 6 months of the pilot study provided the summary statistics and context of the pilot study compared to the previous 4-year average of the same months. Regional crime statistics were also examined. The use of valid and reliable statistics for evaluation supports the need for criminological evaluation research to include sound experiments on clearly circumscribed measures (Lösel, 2008).

Crime analysis supports operational policing and decision-making processes, and includes systems such as comparison statistics (CompStat; Santos, 2016; Willis, Mastrofski, & Weisburd, 2007). In 1994, the New York Police Department developed CompStat, which deals specifically with crime data to help improve communication within a police service by raising awareness and accountability (Willis et al., 2007). “CompStat is a crime management system that combines crime analysis and geographic information systems with modern management principles” (Willis et al., 2007, p. 148), and as it has developed it has been adopted by a number of major metropolitan police agencies throughout the world (Willis et al., 2007). Police departments have found CompStat to be an effective way to track crime accurately and have utilized the information to deploy officers to address community concerns and incidents in a timely manner by using up-to-date information to facilitate the thorough awareness, analysis, and follow-up of criminal patterns (Jang, Hoover, & Joo, 2010). The information that is tracked and used to direct police actions is similar to the descriptive and inferential statistics that were collected for the current research study, and which were then used systematically to examine crime statistics.
Hot spot policing, which is also guided by crime analysis, strives to address emerging issues and to implement proactive strategies before they become problematic and difficult to control (Bowers, Johnson, & Pease, 2004; Braga, 2005). Although the term ‘hot spot’ is used frequently by researchers and by police, there is some variation in the way in which it is applied. For example, researchers have referenced a hot spot as a small area, such as a specific address (Eck & Weisburd, 1995), while others have referred to hot spots as city blocks (Weisburd & Green, 1994) and larger areas within the municipality (Block & Block, 1995) as well as neighbourhoods (Shaw & McKay, 1969).

Although there is no formally agreed-upon definition of a crime hot spot, the “common understanding is that a hot spot is an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization” (Eck, Chainey, Cameron, Leitner, & Wilson, 2005, p. 2). Cool or cold spots are the inverse of hot spots, with less crime than the average. With GIS, police departments have the ability to conduct spatial analyses, including hot spot analysis, on complex data, and to identify more patterns and associations (Anselin, Cohen, Cook, Gorr, & Tita, 2000; Eck et al., 2005). As such, “hot spot policing is a strategy that involves the targeting of resources and activities towards those places where crime or police incidents are most concentrated. The strategy is based on the premise that crime and disorder are not evenly dispersed within neighbourhoods or regions, but rather are clustered in small locations” (Braga & Weisburd, 2010a, p. 132), with common influencers within those locations (Braga & Weisburd, 2011; Braga et al., 1999). Critical to the successful implementation of hot spot policing is the internal GIS capacity of the organization. The GIS framework and associated geospatial analyses are key to understanding crime (Eck & Weisburd, 1995), and were critical in conducting this research. The GIS framework is examined in greater detail in the following section.

Geospatial Analysis

Place, as the subject of interest for crime, can be traced to early social ecologists in 19th century France; however, the capabilities introduced by computerized mapping and spatial statistical analysis have contributed to the increased advancement and growth of place in crime prevention (Anselin et al., 2000; Chainey & Ratcliffe, 2005; Wortley & Mazerolle, 2008). Key criminological theories on place were developed in the 1970s and 1980s, such as
the routine activities theory (Cohen & Felson, 1979), situational crime prevention (Clarke, 1980), rational choice theory (Cornish & Clarke, 1987), and environmental criminology (Brantingham & Brantingham, 1981), has contributed to the growth of geospatial analysis by law enforcement (Wang, 2012). This section focuses on the methods, technologies, and applications of spatial analysis for the purposes of examining data. Criminological theories based on place, or environmental criminology, including routine activities theory developed by Cohen and Felson (1979), crime pattern theory (Brantingham & Brantingham, 1981), and rational choice theory (Cornish & Clarke, 1987), which were all relevant to the current research study, are discussed in the second part of this chapter, within the sections pertaining to predictive policing or break-and-enter theories.

This evaluation study examined RBNEs from a macro approach of place, “looking at aggregates of places such as regions, states, cities, communities and neighborhoods” (Eck & Weisburd, 1995, p. 23). Multiple studies have investigated the actual place at an individual level (Brantingham & Brantingham, 1981; Hunter, 1998; Rengert, 1980), focusing on facilities defined as special-purpose structures (e.g., schools and housing projects) and examining the relationship between them and crime (Eck & Weisburd, 1995; Kelling & Wilson, 1982). The current research focused on place as a larger unit of analysis, since aggregate RBNE data presented no ethical concerns and the purpose of the study was to evaluate whether or not the pilot project effectively predicted crime in an area, not at a specific residence. At a basic level, a location on a map can represent crime spatially. The computer mapping and GIS data that accompany this technology have enabled spatial representation of measurement, analysis, and relationships (Anselin et al., 2000; Manning, 2008; Weisburd & Lum, 2005). Esri (n.d.) defined GIS as follows:

A ... framework for gathering, managing, and analyzing data. Rooted in the science of geography, GIS integrates many types of data. It analyzes spatial location and organizes layers of information into visualizations using maps and 3D scenes. With this unique capability GIS reveals deeper insights into data, such as patterns, relationships and situations – helping users make smarter decisions. (para. 3)

The GIS framework has been useful in various fields, including medicine (Nykiforuk & Flaman, 2011), marketing (Hess, Rubin, & West, 2004), oceanography (Manley & Tallet, 1990), and disaster management (Montoya, 2003). Specifically, the use of GIS has also been growing within the fields of criminology, crime, and policing. Given the extensive use of the
growing literature on GIS in the policing fields, a review of the existing literature was carried out by Zhang, Hoover, and Zhao (2014). Their study was important for the current research since the purpose of their review was to establish whether the GIS studies being conducted within the policing field were empirically based and replicable. Of the 1,121 abstracts reviewed by Zhang et al.’s research team, only two GIS studies were deemed to have met the criteria of rigorous empirical testing and evaluation of a crime intervention in policing. Zhang et al.’s review demonstrated the lack of empirically based, replicable GIS studies in policing, and this current inquiry has added relevant research to address the gap in the literature. The two rigorous studies examined by Zhang et al. evaluated the effects of crime mapping interventions and could be replicated, but in focusing on police perceptions of the interventions, were qualitative in nature. As such, the current study, which was quantitative in nature, served to address what the review of relevant and foundational academic works indicated as lacking in the literature.

Wang (2012) conducted a study examining why GIS was being applied in policing, establishing in his survey that there were six distinct applications for GIS within policing. His findings were relevant to this research, as they were used to categorize the way in which the GIS was applied in the present study. The six applications as identified by Wang are as follows: as a “partner” to field officers; in crime investigation; as a prevention tool for policy implementation, as well as for resource planning; as an instrument for testing crime theories; and as a communication tool (Wang, 2012, p. 23). The current research study used the aggregated GIS data generated to track locations and times of resource deployment. Further, the geo-coordinates to which police resources were deployed and the length of time the resources were active at a location were also recorded and analyzed. The GIS data used for this current study were categorized under the “partner” application. Wang also found that out of all the possible applications for GIS as a crime investigation and prevention tool, the predictive policing applications of GIS were considered to be one of the more advanced spatial analysis tools. As an instrument, GIS assists in identifying future events based on environmental factors that are associated with past incidents. The use of GIS as a communication tool was also utilized throughout the predictive pilot process at the VPD, including the production of the forecasted outputs through to the field resources and onward to the communication of the end results for the purposes of analysis and evaluation.
**Geospatial-Temporal Data Analysis Techniques**

The research design and methodologies used in this current research study were influenced by geospatial analysis and involved in-depth statistical measures using time and space as variables for analysis. The relationship between spatial-temporal data is well studied by academics and geographers (Anselin et al., 2000; Rummens, Hardyns, & Pauwels, 2017). Methods such as risk terrain modelling (RTM) employ GIS to assess crime risk by using map layers that represent spatial influence and crime intensity, and risk heterogeneity can also be measured by presenting underlying factors in the environment that contribute to crime, including all risk factors within a given location and assessing why a crime might occur in a particular location (Caplan & Kennedy, 2010; Moreto, Piza, & Caplan, 2014).

Research into the relationship between objects that include the variables of time and space have introduced a number of statistical measures that can determine whether there is a correlation between the features being examined. Spatial temporal analysis is a form of hot spot evaluation, in that it takes into consideration the fact that crime patterns can change over time and can be affected by environmental factors such as seasons, weather, dates, time intervals, repeat victimization, locations, demographics, and economic factors (Perry et al., 2013, p. 44). A multitude of factors can influence crime patterns, and the data related to these factors can contribute to the forecasting of crimes. A heat map is an example of a spatio-temporal analysis (Perry et al., 2013), and was one of the types of evaluations applied in the current research study.

Spatial autocorrelation focuses on identifying relationships and measures an association between geospatial data that are typically associated with the same spatial area (Cliff & Ord, 1979; Griffith, 2011). Therefore, spatial autocorrelation analyzes the strength of the relationship between objects with other neighbouring objects. Research relies on evaluating whether there is a relationship or correlation between two variables, and on determining a level of confidence that the relationship is either random or statistically significant. One such test is Moran’s I, which looks at whether or not the object is clustered, and at the likelihood that the clustering is a result of random chance, demonstrated by a positive value close to 1 (Anselin, 1995; Moran, 1950). Conversely, a negative value near -1 suggests that the objects are dispersed, but in a manner that indicates patterned autocorrelation and not simply a result of random chance (Leslie & Kronenfeld, 2011). The
Moran’s I statistical test also returns a Z-score and p-value as measures of statistical significance to assist in determining whether to reject the null hypothesis. The Z-score and p-value are associated with a standard normal distribution, in which the researcher must determine either his or her degree of confidence or the risk that he or she is willing to accept for rejecting a null hypothesis (Leslie & Kronenfeld, 2011; Mitchell, 2005). For example, a Z-score greater than 1.96 and a p-value of less than 0.05 would equate to a 90% confidence level that the results were statistically significant; therefore, the researcher would reject the null hypothesis (Law & Collins, 2017).

It is important to note that there are two subsets of the measures of spatial autocorrelation; namely, global measures whereby a spatial pattern occurs consistently across the entire geographic area, and local measures which look at each object individually to help identify unique patterns occurring in different zones of the spatial area (Wang, Hu, Wang, & Li, 2017). Local measures are important when examining regional, provincial, or national datasets, since area-specific differences may have a significant impact on the relationships amongst the variables being examined. Global measures work best when examining more homogeneous regions, or smaller-scale areas that are less likely to have regional variance (Cliff & Ord, 1973). For the purposes of this current research study, the Global Moran’s I was used for data analysis. Griffith defines the range as, “a Moran’s Index value near +1.0 indicates clustering, while an index value near -1.0 indicates dispersion” (Goodchild, 1986; Griffith, 1987, p. 23; Silverman, 1986). Goodchild notes, “without looking at statistical significance, there is no basis for knowing if the observed pattern is just one of many possible versions of random. In the case of the Spatial Autocorrelation Test, the null hypothesis is that there is no spatial clustering of the values associated with the geographic features in the study area” (1986, p. 67). Conversely, “when the p-value is small and the absolute value of the Z-score is large enough that it falls outside the desired confidence level, the null hypothesis can be rejected. If the index value is greater than 0, the set of features exhibits a clustered pattern. If the value is less than 0, the set of features exhibits a dispersed pattern” (Silverman, 1986, p. 23).

A determination of clustering was important for the data analysis of the current research study, in that crime patterns that historically were clustered in certain geo-temporal positions would be altered, with the net effect of spatial and temporal
displacement, diffusion, and dispersal. At the onset of this research, the focused allocation of police resources to burglary incidents was expected to shift future incidents to new geo-temporal locations where police resources were absent (Santos, 2016). It was anticipated that these new crime incidents would appear in a patterned distribution that was inconsistent with previously observed clusters and patterns captured in the historical data spanning the past 4 years. For the short term, stable neighbourhoods that traditionally experienced limited burglary incidents were expected to see a sharp increase until the forecasting model had identified the new patterns and had subsequently allocated resources to these new geo-temporal locations (Guerette & Bowers, 2009).

Another commonly used spatial statistic is the Gi*, which was developed by Getis and Ord (1992) to evaluate whether features are spatially clustered in relation to surrounding features. The Gi* or G statistic distinguishes between hot and cold spots by identifying spatial concentration through calculating neighbouring distance within clusters (Getis & Ord, 1992). Unlike the Moran’s I test, which can only determine whether objects are clustered or dispersed, the Gi* can determine where high or low values are clustered together and indicate spatial concentrations (Davis, 2002). In other words, the Gi* statistic tool identifies objects that may have a low or high value but may or may not be statistically significant. For example, Getis and Ord noted,

To be a statistically significant hot spot, a feature will have a high value and be surrounded by other features with high values as well. The local sum for a feature and its neighbors is compared proportionally to the sum of all features; when the local sum is very different from the expected local sum, and when that difference is too large to be the result of random chance, a statistically significant Z-score results. (p. 102)

As an inferential statistic, a Z-score is typically conducted to test whether the difference is statistically significant (Griffith, 2011). The importance of both spatial autocorrelation tests is that they help to define spatial-temporal characteristics and provide insight into how these are affecting an object, including whether there is a statistically significant relationship amongst the objects. Spatial autocorrelation analysis is gaining greater prominence across the academic spectrum and is now being used to examine correlations amongst crime data and population characteristics to economic indicators where traditional regression analysis was the mainstay (Law & Collins, 2017).
Through their research, Ord and Getis (1995) have demonstrated that their approach is best suited for geo-temporal data, where spatial association of feature data must include distance statistics. This requirement for inclusion of the distance statistics is to take account of the relationship of neighbouring feature data. The Getis-Ord Gi* statistical significance is dependent on comparing a feature with its neighbours and the sum of all the features to determine whether they fall outside the expected value (Ord & Getis, 1995, p. 44). For their study data, the Getis-Ord Gi* returned either a positive or negative Z-score, in which the more intense clustering represents high values and a statistically significant negative Z-score, indicating clustering of low values or cold spots (Ord & Getis, 1995). The Getis-Ord Gi* statistic for each bin identified statistically significant hot and cold spots by intensity, providing a hot spot analysis for each bin and indicating whether for each 2-week interval (temporal bin interval) there were identifiable clusters occurring in relation to their neighbours. When applying the Getis-Ord Gi* statistic against a space-time data cube, it is important to identify the object clusters as either hot or cold, since they are limited in scope to the time interval of the bin parameter. Given the nature of the Getis-Ord test being applied to time-slices at specific bins, and examining neighbouring relationships only within that same time-slice, the approach, while revealing of temporally specific bin patterns and neighbouring phenomena, does not capture vertical axis relationships that are cross temporal. Therefore, this approach identifies instances where a hot or cold spot appears for the 2-week intervals; however, it does not reveal time-series relationships on either side of the 2-week bin interval and how these relate. In other words, while spatial neighbour bins are used to determine statistical significance, temporal neighbour bins are not factored into the Getis-Ord Gi*. These distinctions are relevant to the current discussion, as Getis-Ord Gi is later used in the study and further expanded within the following methodology chapter.

Another spatial temporal analysis technique termed emerging hot spot analysis is used to “identify new, intensifying, persistent, diminishing, oscillating and sporadic hot spots” (Law & Collins, 2017, p. 22), which can help determine statistically significant spatial and temporal relationships. Of the trend category tests available, this modelling approach can detect clustering changes through the measurement of point densities. The more informative aspect of the approach is that it factors in conceptualization of spatial relationships by including bin values from the same neighbourhood time slice as well as bins
from previous time steps (Law & Collins, 2017). In this application, the Mann-Kendall time-series trend test is a “rank correlation analysis” (Kendall & Gibbons, 1990, p. 64) for the object counts and their time sequence, in which the step interval parameter spans the time-series. Therefore, the Mann-Kendall examines neighbourhood distance and relationships as calculated by the Getis-Ord statistic and neighbourhood relationships that span temporal steps, but factor in each temporal bin slice neighbourhood cluster (Anderson, 2009; Law & Collins, 2017). In a sense, the Mann-Kendall trend test is a space-time implementation of the Getis-Ord Gi* statistic that examines the neighbouring bin spatially and temporally and was applicable to the current study using parameter inputs obtained from the pilot study, which is discussed further in the methodology chapter (see Chapter 3).

The process of generating a space-time data cube involves calculating the Mann-Kendall statistic for each bin x/y location according to the parameters input for bin spatial area (Law & Collins, 2017). In this context, x/y refer to spatial grid references that are synonymous to longitude and latitude. Each bin’s Z-score and p-value were evaluated using the Mann-Kendall trend test. The Mann-Kendall test generated a trend Z-score and p-value for each location, with the previously generated Getis-Ord Gi*, Z-score, and p-value for each bin categorized. This correlational analysis is then rank ordered and compared to the expected results, using a benchmark of “no trend” over the time series to determine whether the results are statistically significant (Anderson, 2009; Kendall & Gibbons, 1990; Scott & Pratt, 2009). The process as described by Law and Collins (2017) was applicable to the current research study, as the Z-score and p-value generated for each bin location and for each time interval slice within the separate bins provided the capability to examine complex data trends that were temporally and spatially dependent. In the case of the current research study, this process afforded the ability to evaluate fluctuating trends that expanded beyond simple location-based crime points to geo-temporal dependent trends. Specifically, the time-series output was flattened for the entire surface area and categorized according to the eight trends the model could specify. Table 10 offers a detailed explanation of the Mann-Kendall time-series emerging hot spot categories, and for a description of the emerging hot spot categories see Law and Collins’s work (p. 84).

Up to this point, the review of the literature has focused on the research methodologies relevant to the current study. Equally important is an evaluation of the
practical context of the study, which is the efficacy of the spatial-temporal crime forecasting model and the police actions guided by the model in reducing crime. Specifically, the forecasting model in this research was applied to RBNEs and directing police actions to assess whether there was a reduction in the distribution and intensity of these crimes.

In order to define predictive policing, a broader understanding of AI and big data is required. With the advent of big data fuelling the machine-learning technical revolution, predictive policing has been similarly impacted by both controversy and questions as to the ethics, bias, and validity of the technology. This came to the forefront and formed the backdrop for this research. The manner in which police services adopt new technology is an important factor influencing the spread of predictive policing throughout law enforcement. Theoretical foundations that suggest possible explanations as to how predictive policing functions at a humanistic level include a range of models, such as rational choice and routine activity theory. These criminological theories are explored in the following section, as well as the latest theory that aims to discount all previous theories, with the explosion of big data promoting an alternative view that everything can be mathematically explained through AI. Foundations of predictive policing principles are described, including the evolution and context that frames the introduction of predictive policing.

**Artificial Intelligence, Big Data, and Data Quality**

Machine-learning and AI neural-networks are at the forefront of a technological revolution that affects almost every facet of society, including policing and analysis. From IBM’s Watson, which defeated two champions of the game show Jeopardy!, to Google and Uber’s self-driving cars (Goodell, 2018), these trends continue to evolve (Brynjolfsson & McAfee, 2017; Dunjko & Briegel, 2018; Kibria et al., 2018). The subject of AI is often discussed in the context of big data, in that the actual processing, analyzing, and making sense of the information of the volume of data is often done by AI. “AI bases its operation on accessing huge amounts of information, processing it, analyzing it and, according to its operation algorithms, executing tasks to solve certain problems” (Shabbir & Anwer, 2018, p. 9). The term AI refers to developing machines and robots that have characteristics of humans in terms of behaviours, learning, sensing, and meaning making (Shabbir & Anwer, 2018).
Within the field of criminology, computer scientists from University College London and the University of Sheffield developed a machine-learning AI for the European Court of Human Rights which can predict verdicts with 79% accuracy, including the weighing of moral considerations (Aletras, Tsarapatsanis, Preotiuc-Pietro, & Lampos, 2016). In the policing community, the use of big data to feed computer modelling as a forecasting tool for crime prevention has been a growing trend that enhances predictive policing (Chan & Bennett-Moses, 2016; Kang & Kang, 2017; Rummens et al., 2017). In many cases, discussion of predictive policing includes a review of big data, and any policing practices that use large datasets are often labelled as big data (Hardyns & Rummens, 2018; Perry et al., 2013; Uchida, 2009). Big data is a field, domain, concept, theory, and an idea that involves the storage, processing, and analysis of large pools of data for the purposes of gaining intelligent insight which, in turn, is used to help inform decisions (Gupta & Saxena, 2016, p. 2). The term big data can refer to a number of things and has many nebulous definitions (De Mauro, Greco, & Grimaldi, 2016). Such definitions can range from relatively recent technological innovations to a new field of study, new technical profession (Gupta & Saxena, 2016; De Mauro et al., 2016), or phenomena that are cultural, technological, and scholarly in scope (Boyd & Crawford, 2012, p. 663).

At its core meaning, big data relates to the idea that vast amounts of information, which cannot be processed manually or through simple computation, can now be analyzed in a useful and timely manner. In their review of existing literature for the 15 most popular meanings of the term ‘big data’, De Mauro et al. (2016) proposed a definition of big data, describing it as “the Information asset characterized by such a High Volume, Velocity and Variety [as] to require specific Technology and Analytical Methods for its transformation into Value” (De Mauro et al., 2016, p. 132). According to Gupta and Saxena (2016), big data is a five-dimensional paradigm that allows for dealing with volume, velocity, variety, veracity, and value of data (p. 2). With the increasing amount of data being generated every second, data can be overwhelming, not only to process usefully, but also for basic storage. The volume and velocity, the amount, and the speed at which the data units are being generated exceed what can be understood by the human brain (Shabbir & Anwer, 2018; Vempati, 2016). The amount of data generated on a daily basis was reported at 2.5 quintillion bytes as of 2018, and, in terms of velocity, 90% of the data in the world had been
generated in the previous 2 years (Marr, 2018). An example of the sheer volume and the velocity of potential data available for processing included the staggering amount of social media data being generated every minute, consisting of 12,986,111 texts sent, 2,083,333 photos on Snapchat, and 4,333,560 YouTube videos (Domo, 2018). Both crime analysis and geospatial analysis require data, and the evolution of big data within policing have enabled advanced data mining, which is noted as a prerequisite for predictive policing.

In addition to providing a definition and the taxonomy of predictive policing, Perry et al. (2013) sought to clarify the term ‘data mining’. Perry et al. indicated that data mining was also referred to as predictive analytics and used synonymously with big data, which was defined as using large amounts of data units for the purposes of finding useful patterns and trends (p. 13). The identification of patterns is important, especially for spatial and temporal distribution and predictive analytic software that are able to make visible and give form to knowledge about regular occurrences of crimes that lie hidden in data sets…. In the context of predictive policing, the promise of the pattern is thus to serve as a base for the extrapolation of possible criminal futures and to render those futures actionable for prevention programmes. This means that patterns pre-structure how the police act upon crime and society at large. Moreover, patterns legitimize this trend towards automated policing rationalities as they provide interventions with calculative rationalities. (Kaufman, Egbert, & Leese, 2018, p. 1)

However, there are other definitions that differentiate the terms data mining and predictive analytics. According to Friedman (1998), the definition of data mining is vague, and how it is defined tends to relate to the background and perceptions of the person defining the term. Data mining can be discussed in the realms of computer science in the context of mining data using algorithms (Skillicorn, 1999), and of statistics since it addresses similar issues (Friedman, 1998). Although Perry et al. (2013) found the terms to be synonymous, the current research study did not use the terms synonymously. This present study supports the perspective that the identification of the patterns is what allows prediction to occur and legitimizes the subsequent actions of the police to disrupt these patterns. Although some predictive policing definitions do not specifically mention patterns and their use to rationalize police actions, Schlehahn et al. (2015) have determined that there are two other uses of data and patterning in policing: descriptive and prescriptive analytics (p. 1). It is the
identification of patterns that provide specific or general factors to reduce crime risk which are used in predictive policing (Schlehahn et al., 2015).

Although predictive analytics may have data mining as a component, whereby algorithms are applied to large data repositories in order to categorize and analyze big data, this is simply a process for deconstructing data to provide meaning. This differs from generating forecasts from data repositories using advanced algorithms. For the purposes of this paper the terms predictive analytics and data mining are not used interchangeably. The term ‘AI’ is also associated with the processing of big data and in some instances is referred to when discussing predictive models. For these reasons, big data and AI are examined in the following section, where it is important to draw a distinction in the use of the terminology as it relates to predictive policing, as well as the highlighting the direct relevance to the current research study.

Specific to police data, hundreds of public open datasets related to crime statistics are available, ranging from police-involved fatalities in the United States (US) from 2000 to 2016, to Monthly Uniform Crime Reporting Statistics for Buffalo, New York, to data on crimes in Atlanta, Georgia, from 2009 to 2017 (Data.World, n.d.). To illustrate the volume of police data within these datasets, Chicago police update their crimes database on a daily basis with more than 65,000 data records going back to 2001, which is in excess of the rows viewable in Microsoft Excel (United States Government, Data Catalog, 2018). The sheer volume of the number of rows alone shows that big data is so large that it cannot be processed through simple, readily-available software such as Microsoft Excel, and that a higher level of expertise or technological assistance is required, not only to view the data but also to process the information. Variety, or specifically the availability of different types of data, is also part of big data. Traditionally, databases required structured data that could be put into tabular form; however, more than 80% of the data being generated now are unstructured (Gupta & Saxena, 2016, p. 3). Storage and the process of unstructured data, such as those from photos, videos, recordings, and narrative text can be processed, analyzed, and utilized with big data (Gupta & Saxena, 2016).

All these data sources are potentially of interest to policing and can be utilized for the purposes of public safety, if processed and analyzed in a meaningful way. Veracity speaks to the validity, correctness, and, therefore, the accuracy and usability of the data
(Gupta & Saxena, 2016). As such, not all 2.5 quintillion bytes of data generated on a daily basis can be accurate, nor can all of that information be of any use. A good example is Facebook removing 559 pages and 251 accounts in 2018 that were spreading misinformation and fake news (Carson, 2019; Frier, 2018). Beyond fake news, there are also other data veracity issues, including human errors of misspelling, misunderstanding, as well as typos, auto-correct mistakes, and key entry errors that generate invalid or inaccurate information. Specific to policing is the fact that any inconsistencies or misinformation can have unintended legal and civil implications.

The difficulty with big data is that not only is there a vast amount of information generated at great velocity that needs to be assessed for validity and value, but there is also the added complexity with the technologies capable of processing big data. An emerging industry service, which includes Apache Hadoop and MongoDB, now provides various technologies that are capable of storage and analysis in addition to containing components that act as platforms, infrastructure, and associated services. Other technology services and platforms have a combination of functions, such as business intelligence that incorporates visualization, marketing, and other aspects (Gonzalez, 2015; Gupta & Saxena, 2016). The relevance of this complexity to policing is that not only are informed decisions needed in terms of the types of data available, but also how best to extract useful information from the massive datasets that police services are now building. As PredPol (2018b) stated on their website in a post titled “So you Think you can Build your own Predictive Policing Platform”, this is not an easy undertaking for police services, and many would find the prospect daunting and unattainable. The first step involves building a model that reflects the needs of data and prediction requirements, while ensuring data and information integrity. Additional considerations are to monitor the reliability of the system to safeguard information, guarantee that the software operates within legal requirements, and ensure that the outputs are applicable and valuable to end users (PredPol, 2018a). Given that big data brings with it complexities that require technical expertise and a robust information technology infrastructure, the lure of a ready-made solution, as proposed by software companies that offer to tailor their services to increase efficiencies with fewer resources for the purposes of crime reduction, is a strong one. Unfortunately, even with the explanation of the algorithms and the theories behind the modelling, the level of information
technology, mathematical, technological, and legal expertise that is required to fully appreciate the implications of utilizing crime forecasting is considerable. Combined with the types of ethical, civil, and legal issues raised in this study, it is also critical for the police agencies that implement big-data-fuelled predictive policing methods and other forecasting models to evaluate thoroughly the data sources, applications, and intended outcomes so that they are able to meet the threshold of public scrutiny as well as legal obligations.

As the complexities, volume, and velocity of data increase, the black box problem has emerged. This problem of “higher performance but less clarity” (Ozkan, 2018, p. 9) exists for machine learning models that are highly complex and provide accurate predictions, but are difficult for researchers to understand or interpret in order to work out how the outcome was achieved. Data sources, relationships, and the patterns among them continue to expand exponentially into that which is difficult to comprehend with the human mind, thereby increasing black box problems. This also adds to the epistemological authority of patterns produced by algorithms (Kaufmann et al., 2018), since it becomes more difficult to identify how the outcomes were achieved. Big data is not a requirement, but it enhances predictive policing (Chan & Bennett-Moses, 2016). That is, having big data does not equate to the ability of a police agency to conduct predictive policing, nor does an agency require big data to employ rudimentary crime analysis practices. However, it is the case that a model or an advanced algorithm cannot be used if there is no big data (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). For example, in the case of the current research study, if VPD had not had a sufficient number of historical data points or numbers of incidents for the predictions to be calculated by the model, the application would have been rendered impractical.

To contextualize, the predictive model piloted at VPD produced a set of six predictions every 2 hours and the intended application was for police resources to be assigned to all six locations to deter or interdict crime. This predictive model and its operationalization would have been counterproductive if there had been an insufficient number of residential burglaries to support this type of deployment. For example, had there been only 10 residential burglaries per month in the jurisdiction, the number of times that police officers had attended a prediction location and nothing had happened would exceed the number of times that would have led to deterrence or interdiction. Continuing with this
example of only 10 residential burglaries per month, the chances that a burglary would occur on any given day is less than 33%; as such, on a given day, it is more likely that the police resources would be directed to locations where no burglary would have occurred. In this scenario, resources deployed to forecasted areas would have had little deterrent effect because the chance of any occurrence was low to begin with. In summary, there is no benefit in deploying resources when there is not enough data to form patterns or trends, or if the number of crimes does not warrant a big-data predictive approach (Kaufmann et al., 2018). Predictive policing may not be in the best interests of, or an efficient solution for police departments that do not have the volume of data required. Only offences that are statistically identifiable because they follow patterns with a sufficient volume of historical data to support the algorithmic computations, justify the use of the technology. If the volume of the crime type is insufficient then it becomes almost impossible for the algorithm to extrapolate a pattern (Kaufmann et al., 2018). It is necessary for police agencies to assess the information they have, determine what the data units are potentially able to predict, consider whether pattern identification is possible, and anticipate what the outcome might look like before proceeding with potential solutions.

For North Americans, AI is already pervasive in people’s lives spanning across the finance, marketing, administration, automotive, purchasing, and technical sectors, and is expected to permeate all aspects of domestic life (Shabbir & Anwer, 2018). Associated with AI is the concept of machine learning, which was demonstrated by the forecasting model utilized in the VPD pilot project. Machine learning is a computer science domain that “has developmental processes in which repeated exposures of a system to an information-rich environment gradually produce, expand, enhance, or reinforce that system’s behavioural and cognitive competence” in that environment or relatively similar ones (Vempati, 2016, p. 21). Learning produces changes in the state of the system that endure for some time, often through some mechanism of explicit or implicit memory formation (Vallor & Bekey, 2017, p. 4).

The machine learning is through producing outputs based on various relevant features from the input; the algorithm allows the network to be “trained” with each new input pattern until the network weights are adjusted in such a way that the relationship between input and output layers are optimized (Vallor & Bekey, 2017; Vempati, 2016). Thus,
The network gradually “learns” (Vallor & Bekey, 2017, p. 5) from repeated “experience” (p. 5), or multiple training runs with input datasets, how to optimize the machine’s “behavior” (p. 5), or outputs, for a given kind of task. The vast amount of information ingested at great velocity that needs to be assessed for validity and value (Gupta & Saxena, 2016) and further processed for pattern analysis requires an AI-based system, given the growth of big data that moves these processes beyond the capacity of traditional programming solutions (Shabbir & Anwer, 2018).

The assessment of validity and value becomes an issue with machine learning as well, and the supervision, or lack thereof, with which humans oversee what AI will learn (Vallor & Bekey, 2017). The current study did not focus on big data, AI, or machine learning per se, but rather utilized the outputs that they produced. The results of the evaluation showed that the first month of the pilot implementation in April 2016 did not have the same effect on forecasting as did the following months, because the predictive system had a chance to learn and train on the data. The resulting inferential statistical analysis also showed that when the input from the data source omitted the location information, this resulted in repeat locations being forecast due to the lack of updated data. This analyses reinforces the notion that learning must occur with vetted data, and that the process will require some degree of monitoring by humans to verify that which is being learned aligns with the outcomes sought, and that the quality of the data feeding the system adheres to a set standard.

Data units are key components of predictive policing, as “all predictive techniques depend on data and both volume and quality of data affect the usefulness of the approach” (Perry et al., 2013, p. 13). The amount and quality of data collected will have a direct relationship to the predictions that can be made as well as to the effectiveness of the output. The data also must accurately reflect reality (Chan & Bennett-Moses, 2016) for predictive policing to be effective. Chan and Bennett-Moses (2016) argue that the crime data units forming the basis on which technology operates are not an accurate representation and are limited to what leadership determines should be reported or observed (p. 4). Data quality is essential to the rigour of analysis and directly affects the outcomes. In a thorough review of operationalized uses of predictive policing methodologies and technologies in various agencies, Perry et al. (2013) found predictive methodologies that used conventional crime
analysis tools, applied to low to moderate volume of data with a low to moderate level of complexity required to predict, were still dependent on a significant volume and high quality of data, which was similar to that required of complex technologies. This reinforces the fact that, regardless of the level of analysis conducted, all predictive policing techniques depend on high-quality data and in sufficient volume to enable analysis. Having explored AI, big data, and data quality as it relates to predictive policing, its definitions and trends are explored in the next section.

**Predictive Policing Trends**

The concept of predictive policing is not new to policing, in that statistics and geospatial analysis for the purposes of forecasting crime have been in use for decades (Perry et al., 2013, p. 2). In practice, predictive policing borrows from, and builds upon, policing principles and models that are already well established (Pearsall, 2010). It is an extension of traditional policing practices; however, with policing having entered the “information (IT) era” (Rosenbaum, 2007, p. 15), there is a greater reliance on data to inform and direct interdiction strategies to prevent and respond to crime. With the emphasis on technologies and data, predictive policing has also become increasingly more complex. An understanding of the current and historical definition, as well as an examination of the application and use of predictive policing, is essential in order to discuss and contextualize the current research study. This section begins with the definition of predictive policing, followed by an examination of theories of crime that help explain how it works.

**Defining Predictive Policing**

When examining the literature, parallels emerged between the definition of predictive policing and ILP, both ubiquitous terms used in law enforcement. For example, when ILP was first introduced, neither the Audit Commission nor the National Criminal Intelligence Service succeeded in defining what it was, which resulted in the challenges and merits of ILP being discussed without a clear definition (Ratcliffe, 2003). Similarly, predictive policing does not have an accepted definition (Ratcliffe, 2004), although its benefits and possible drawbacks are discussed in a number of research publications (Koper, Taylor, & Kubu, 2009; Mantello, 2016; Perry et al., 2013; Ridgeway, 2018). In addition, when researchers attempt to define ILP, they often include a compilation of concepts and variables, such as a business
model as well as a managerial philosophy used to direct police operations (Ratcliffe, 2016) or an application of quantitative techniques (Perry et al., 2013). There is a lack of cohesiveness and consistency in the way in which the term ILP is applied, but regardless of this disconnect, it is fair to conclude that technology is an essential component of both ILP and predictive policing, given that the definitions of both incorporate technology as a component, as is demonstrated by the following examples.

With both ILP and predictive policing being technologically dependent, researchers noted a direct association between the success of an ILP implementation and the ability of a police service to leverage technology successfully (Weisburd & Braga, 2006a). Predictive policing is dependent on having an advanced technological capacity, especially when defined as being synonymous with the phrase data mining for mathematical patterns (Perry et al., 2013). Data are also an essential component, given that “all predictive techniques depend on data” (Perry et al., 2013, p. 32), and that the leveraging of data to support an ILP framework appears to blur the distinction between the two, with minor variation on how the data units are applied. The definitions of ILP often include reference to using data in tandem with information technology, and analysis for the purposes of crime prevention to target resource allocation (Boyd & Crawford, 2012). Given this definition of ILP, and the similarities in the attempt to contextualize and define predictive policing, there is no doubt that the two are related. Predictive policing can be considered as one of the many possible ILP model applications, in which police agencies can operationalize the broader concept of interpreting the criminal environment.

Despite the lack of a universal definition that encompasses all the activities and concepts of predictive policing that can be agreed upon by academics and practitioners, several researchers have attempted to define the term by applying an overly broad and nondescript approach. In one of the earliest attempts, Uchida (2009) stated, “Predictive policing refers to any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention” (p. 2). This was the working definition that John Morgan presented at the first predictive policing symposium held in California in 2009 and organized by the National Institute of Justice (NIJ). The working definition consisted of five elements: integrate information and operations, see the big picture, use cutting-edge analytics and technology, link to performance of organization, and
adapt to changing positions (Uchida, 2009, p. 2). This definition focused on information and technology as well as the incorporating of adaptation and change within the organization as part of the concept. Predictive policing was equated to business performance measures, and the “favorable outcomes” (Uchida, 2009, p. 3) for the police would be crime reduction and better quality of life.

The National Law Enforcement and Corrections Technology Centre of Excellence on Information and Geospatial Technology at the NIJ sponsored a study by Perry et al. (2013), in which the authors examined predictive policing in the US with the aim of conducting a comprehensive overview in order to create a reference guide for police departments interested in predictive policing. In their report, Perry et al. defined predictive policing as “the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevention of crime or solve past crimes by making statistical predictions” (p. xiii). This definition is often used in studies and reports as a starting point for discussion on predictive policing (Karppi, 2081; Lum & Isaac, 2016; Nix, 2015; Ratcliffe, 2015; Seele, 2017); however, it limits predictive policing to an analytical technique, rather than a philosophy or business model. The scope of Perry et al.’s definition is quite narrow, in that it is specific about statistical predictions for the purposes of intervention and prevention of crimes while also including the possibility of solving past crimes based on statistical predictions. Nor does this definition provide any information about where the data originated, or the types of information required to conduct the analysis. However, although the definition is narrow in scope, it does incorporate many techniques and practices that could have been considered as traditional policing, such as simple hot spot analysis and heat maps.

Despite seeming contradictory, Perry et al.’s definition is narrow in capturing only analytical techniques but does include a broad range of such techniques that can be included in the definition. The definition could also have benefited from including more than the techniques and technologies while excluding the examination of basic analytical techniques. The guide did, however, compile both historical and current literature ranging from academic literature, government publications, and vendor information and open source data available on predictive policing, as well as case studies (Perry et al., 2013). Perry et al. found that predictive policing and crime forecasting were terms that were used
 interchangeably, and with regard to the current study a similar adoption of the usage of the term was applied for consistency with the literature and research.

Specific to Perry et al.’s (2013) study, the definition of predictive policing seems to be a modification of the 2009 NIJ symposium’s working definition that predictive policing is “the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevention of crime or solve past crimes by making statistical predictions” (p. xiii). As a result of reviewing and compiling the available information, one of the main contributions of this publication by Perry et al. (2013) was the creation of a taxonomy of predictive policing methods, which helped further define predictive policing into four categories: methods for predicting (a) crimes, (b) offenders, (c) perpetrators’ identities, and (d) victims of crime (p. 8). These four categories of predictive policing are based on what it claims to predict. This taxonomy has also been proposed by other researchers (Koper et al., 2009; Mantello, 2016; Ridgeway, 2018) and is a way to narrow down the focus and specify the discussion on predictive policing. This taxonomy also conforms to the law enforcement landscape, in which the most common examples of predictive policing are forecasting crime locations and predicting offenders at risk of committing crimes, with a primary focus on geospatial information. Since this current research study evaluated the VPD crime forecasting system, which identifies places and times where crimes may occur and thus falls into the category of method of predicting crimes, this particular category is discussed in more detail in the paragraphs that follow.

The method of predicting crimes is defined as approaches that anticipate places and times with an increased crime risk (Perry et al., 2013, p. 8), and the prediction can be made using various methods with varying degrees of complexity and data requirements. Crime prediction can take the form of hot spot analysis, mathematical modelling, and spatio-temporal analysis (Perry et al., 2013), and most of the analyses will result in map-based outputs, as the crime predictions are generally based on locations, and frequently use GIS. The approaches that use larger scale data and more complex data relationships typically include advanced hot spot models, regression classification and clustering repeat modelling, spatio-temporal analysis methods, and risk terrain analyses that are considered to be predictive policing (Perry et al., 2013, p. 10). The primary focus in the current research study was on methods of predicting crimes. However, a discussion on predictive approaches
identifying individuals is relevant here since it is the most controversial of the predictive policing methods. Predictive methods identifying individuals raise many issues, including legal and ethical concerns about civil rights infringements and unintended negative consequences with individuals being falsely identified as criminals (Ridgeway, 2018; Saunders, Hunt, & Hollywood, 2016). Identifying individuals regarded as displaying a higher likelihood of committing crimes, as well as predicting the identities of offenders and victims (Perry et al., 2013) are both based on personal data associated with individuals, rather than being generalities for prediction. The main legal and ethical concerns regarding predictive policing are discussed in the ethics section of this chapter, including the issues associated with identifying individuals. However, it is important to note here that location-based crime incident prediction and methods of anticipating crime have been less controversial, since personal identities are not at risk and aggregated location-based information focuses on residences as targets of police enforcement.

Most likely as a result of Perry et al.’s (2013) report and the interest of law enforcement agencies in the topic, as well as various implementations that occurred after the first predictive policing symposium in 2009, the definition of predictive policing by NIJ (n.d.) had, by 2014, evolved from its original working definition to be broader in scope, and it now covers many possible applications of predictive policing while retaining specific concepts:

Predictive policing tries to harness the power of information, geospatial technologies and evidence-based intervention models to reduce crime and improve public safety. This two-pronged approach — applying advanced analytics to various data sets, in conjunction with intervention models — can move law enforcement from reacting to crimes into the realm of predicting what and where something is likely to happen and deploying resources accordingly. (NIJ, n.d., What is Predictive Policing section, para. 1)

Compared to the NIJ’s 2009 working definition (Perry et al., 2013), the 2014 definition incorporated outcomes, specified the use of geospatial technologies and computer models and the types of datasets that could be used, and articulated what could be used to target (NIJ, n.d.). This definition was intentionally broad and showed that the term predictive policing could be applied to any human action, software, practice, or policy that attempted to use information, geospatial technology, and evidence-based intervention models for the
purposes of reducing crime. The NIJ further incorporated comprehensive terms, such as “computer modelling” and “socio-demographic data”, to illustrate how predictive policing could be applied to a range of variables. The NIJ (n.d.) definition also draws parallels between business analytic concepts and industry trend analysis, and those of offender identification and crime incident locations, as evident in the following excerpt from the overview of the summaries in their Predictive Policing:

Predictive policing leverages computer models — such as those used in the business industry to anticipate how market conditions or industry trends will evolve over time — for law enforcement purposes, namely anticipating likely crime events and information actions to prevent crime. Predictions can focus on variables such as places, people, groups or incidents. Demographic trends, parolee populations and economic conditions may all affect crime rates in particular areas. Using models supported by prior crime and environmental data to inform different kinds of interventions can help police reduce the number of incidents. (What is Predictive Policing section, para. 3)

This excerpt shows that predictive policing is at times considered an application of ILP, in that it is a decision-making tool derived from evidence to prevent and reduce crime (Ratcliffe, 2003). Depending on the targeted output, predictive policing could also form part of an enforcement strategy. Despite the emphasis on tangible outcomes and defining a range of input variables, the case studies conducted to assist in the definition suggested that police agencies were simply adopting predictive policing techniques and strategies with few or no evaluative processes (Bond-Gram & Winston, 2013).

In another report on predictive policing, published through the IBM Centre for the Business of Government, Bachner (2013) examined the practical definition of predictive policing, placing it within the broader context of ILP, and noting specifically how it was developed out of the drive towards evidence-based decision-making (Bachner, 2013; James, 2013; Ratcliffe, 2004). Components of predictive policing, such as the data used and statistical methods applied, were included in three case examples of police services that had applied predictive policing practices. In addition, Bachner described key components of predictive policing models that provided a general overview of predictive policing, since the past, as a predictor of the future, relies on theories that
we can make probabilistic inferences about future criminal activity based on existing
data. In short, we can use data from a wide variety of sources to compute estimates
about phenomena.... Past data from both conventional and unconventional sources
combine to yield estimates, with specialized degree of certainty, about what will
happen in the future. (p. 15)

This emphasizes a salient point on which predictive policing depends, i.e., “the theory that
the past can be used to predict the future relies on the assumption that crime, like any
phenomenon that can be discussed in probabilistic terms, has both a systematic and non-
systematic component” (Bachner, 2013, p. 14). In this theory, non-systematic components
are unpredictable, whereas systematic components are predictable based on logic, which
can be explained by rational choice theory (Cornish & Clarke, 1987). Criminological theories
are important in that “every therapy, treatment program, prison regimen, police policy or
criminal justice practice is based, either explicitly or implicitly, on some explanation of
human nature in general or criminal behaviour in particular” (Akers, 2013, p. 11), and
predictive policing is no exception. This section of the literature review has summarized the
predictive policing definitions that were current and available at the time this study was
being conducted. The next section explores theories that have influenced predictive policing
in practice.

Criminological Theories on Predictive Policing

The Classical School of criminological theories assumes that individuals have a choice in
committing crimes, unlike theories that have underlying assumptions that crime is
committed based on factors, such as biology, psychology, environment, and social learning,
that are beyond the control of the individual (Akers, 2013). Given that crime is considered to
be a choice, deterrence is a key component of classical criminological theories (Akers, 2013).
Deterrence is the fear of sanctions or punishment that results in a criminal act not being
committed (Paternoster, 2010). Deterrence theory is built on the principle that individuals
are rational beings that consider consequences, and that their actions are influenced by
these potential consequences (Kennedy, 2008). To be an effective deterrent (i.e., to have an
effect on potential offenders, thereby reducing the crime rate), the offender must perceive
an increased risk of negative consequences resulting from the additional police presence
(Paternoster, 2010). In a review of deterrence studies, Paternoster (2010) concluded that,
“It is reasonable to believe that increasing the number of police officers on the street does
deter some amount of crime, and increasing the risk of incarceration does as well” (p. 818). Hot spot policing is premised on the fact that the presence of police resources will deter crimes from occurring at the location of deployment (Braga, Papachristos, & Hureau, 2012).

This also holds true for predictive policing, since the allocation of resources to the area identified by the forecast reduces the amount of criminal activity at that location. As previously noted, high crime rates occurring in a micro-place or a small geographic area compared to the rest of the city is regarded as making that place a hot spot (Braga & Weisburd, 2010b). Hot spot analysis uses historical crime data to identify higher risk locations at which specific crimes might take place, based on the premise that crime is more likely to occur where it has done in the past (Perry et al., 2013, p. 13). Therefore, hot spot policing is the implementation of strategies and tactics that involve the deployment of police resources to those areas that have higher crime concentrations (Braga & Weisburd, 2010b). This focus on location and offenders is referred to as environmental criminological theory. Environmental criminology is based on three premises: (a) criminal behaviour is influenced by the environment and some environments are criminogenic and affect behaviour, (b) distribution of crime is patterned and is not random in space and time, and (c) knowledge of items (a) and (b) allows crime to be reduced (Wortley & Mazerolle, 2008). Environmental criminology also assumes that motivated offenders are focused on particular settings, which guide crime analysis, prevention, and evaluation effectiveness (Brantingham & Brantingham, 1981; Santos, 2016), while the criminological frameworks that fall into this category of environmental criminology include rational choice theory (Cornish & Clarke, 1987), routine activities theory (Clarke & Felson, 1993), and crime pattern theory (Brantingham & Brantingham, 1981).

Rational choice theory also shares the same philosophy as deterrence theory, in that an individual's decision-making process is based on risk and rewards when a decision is made to commit a crime (Santos, 2016). Given that the decision making is based on a rational process of risks and rewards, the systematic component during this process can be used to estimate probability (Bachner, 2013); that is, a deterrent can be predicted. The rational choice theory then assumes that crimes can be prevented in the future if risks to potential offenders outweigh the rewards (Santos, 2013). This current research study examined the efficacy of prediction in application.
Taken to the next application of the theory, crime displacement can be understood as a function of rational choice theory (Cornish et al., 1987). Different types of displacement are possible as a result of hot spot policing (Weisburd et al., 2005). Five potential types of displacement have significance to the outcome of predictive police deployments: spatial, temporal, target, tactical, and type of crime (Hakim & Rengert, 1981, p. 11). These types of displacement are possible, since crime moves in response to targeted law enforcement actions (Barr & Pease, 1990). The space and temporal displacements are changes in location and time and relate to specific zones or areas, while tactical or method as well as crime-type displacements involve alterations in how the offender commits crimes, and are individual-based (Weisburd et al., 2005). The concepts of deterrence and possible displacement are also related to the routine activities theory, which states that for a crime to happen there must be a motivated offender, a suitable target, and a lack of capable guardians occurring at the same time and place (Cohen & Felson, 1979). Given that criminals operate in their comfort zones, the idea that crime is predictable can be supported (Perry et al., 2013, p. 2), because the actions of the offenders are potentially predictable based on their routine activities.

The “blended theory” (Perry et al., 2013, p. 3) of predictive policing, based on the routine activity, rational choice, and crime pattern theories, is what Perry et al. (2013) referred to as the justification for being able to predict crime. The blended theory is one of the numerous possible criminological frameworks that might be used to explain and form a basis for the way future crimes could be predicted, with many possible combinations and paradigms to consider within predictive policing contexts. These theories explain how predictive policing practices have been conceptualized and the perspectives that have contributed to how such approaches are being operationalized. The next section explains the way in which these theories were applicable to this current research study.

The routine activities theory proposes that crime decreases with the presence of capable guardians (Cohen & Felson, 1979). In the current research study, the police resources were the capable guardians, acting as the deterrent to motivated offenders for the RBNE targets. The deployment of police resources to the projected locations, as predicted by the forecasting model, provided capable guardians for these locations, since an offender would be deterred from carrying out residential burglaries in the area. As such, the
47

assumption, based on the routine activities theory, is that the incidents of RBNE would
decrease as a result of the predictive policing pilot project. These theories provide possible
discussion points regarding any changes in the number and distribution of RBNEs within the
study area. The application of criminological theories as it relates to property crime, and
more specifically to RBNE, is discussed later in this chapter.

Chris Anderson (2008) expressed his concern that big data could render theories
obsolete, since traditional social sciences and their ways of sampling were being replaced by
the use of vast amounts of data that could capture a whole population, rather than a small
set. The idea of analyzing data with no theory is that with all the data, rather than just a
sample from the population being examined, a pattern could be developed and interpreted
instead of theoretical models being applied. Predictive policing algorithms are based on this
idea and on patterns from big data which emphasize that

patterns form the epistemic core of predictive analytics in general, and of policing
software or algorithms more specifically. They give form to information hidden in
datasets, which grants them considerable epistemological authority—not only in the
context of policing. Moreover, patterns do act in their role as authorities, since they
guide police action—whether in everyday practices of dispatching officers or in
intelligence mapping activities. (Kaufman et al., 2018, p. 15)

The argument for big data replacing theory was further reiterated by Chris Anderson
(2008) who explored the prevalence of big data entering every facet of human interaction as
the focus shifted to outcomes, rather than to any explanation of the overarching drivers and
motivations influencing human behaviour. Thus the new era of theory shifts from an
explanation of the reason why to

a world where massive amounts of data and applied mathematics replace every
other tool that might be brought to bear. Out with every theory of human behaviour,
from linguistics to sociology. Forget taxonomy, ontology and psychology. Who
knows why people do what they do? The point is they do it, and we can track and
measure it with unprecedented fidelity. With enough data, the numbers speak for
themselves. (Anderson, 2008, p. 23)

Anderson (2008) illustrated his point further by using biology and allowing algorithms to find
patterns; this shows that the concept of no theory can even be applied to the non-social
sciences. Applying this premise to the current study, it is interesting to note that patterns
can be detected between historical crime data and that collected from the pilot study, given sufficient data points and the application of finely tuned technology. Various criminological theories can also be applied as possible reasons a particular outcome occurred, however, both the causal relationships and the application of criminological theory are inferred. There is no feasible research method to determine exactly why crime decreases in a particular area and why potential criminals do not commit the crimes; but an inference can be made based on theory and historical data. Big data, and the patterning of all potential data, leads to the possible by-passing of inferences and causal relationships by determining all possible data combinations and displaying what is, rather than what might be. This was of interest to the current research study as an evaluation research for a predictive policing application. Big data and patterning are an essential part of predictive policing applications, and the possibility of patterning out all possible combinations of what is, rather than providing possible inferential relationships, should be given some thought.

Chan and Bennett-Moses (2016) directed this question specifically to criminology and the impact of big data on the discipline. They concluded there were two areas in which big data was being used in criminology, namely the use of social media data for research and the use of models and algorithms as a predictive policing tool. Chan and Bennett-Moses further stated,

The use of Big Data as one type of research data is unlikely to transform the practice of criminology except new analytical techniques are required to process the very large volumes of data. Even the use of computer modelling and analytics does not by and large constitute a significant departure from past criminological practices, but there is one exception; the use of machine learning procedures in predictive analysis is one area where established ways of doing criminology may well be threatened. This relates to the claim that with Big Data, causal theory is irrelevant because correlation is sufficient. (p. 36)

Kaufman et al. (2018) have questioned the ethical implications of this approach, in which patterns hold epistemological authority and challenging this is difficult. However, Ozkan (2018) argues that in certain cases machine learning models may be the best suited to a given criminological problem, given a sample size closely resembling the population of interest, since it is possible to miss a variable of value. Without all the information and data, an influential variable may well be overlooked; as such, machine learning models may be
able to capture relationships that could be missed with less information (Ozkan, 2018). Since inductive reasoning is the rationale behind the data-driven approach of big data without hypothesis, it should not be discounted as being less relevant or less valid than other theory-based research even though it does not align with the scientific hypothesis-driven research practices of criminology (Ozkan, 2018). The idea that big data, used specifically in the way that this predictive pilot employed it, is the epistemological source and that the resulting patterns have an inherent power (i.e., it represents an analyzed whole) is an interesting one and needs to be considered, since technology and patterns are often perceived to have unbiased foresight. As a result of resources being deployed based on the crime-forecasting model, further power is given to big data. As noted by Kaufman et al. (2018), the ethics of big data is often overlooked with this approach, whereby biased data, the product of human endeavour and potentially flawed algorithmic design, could result in a erroneous output and analysis. These concepts are explored further in the Ethical Concerns in Predictive Policing section of this chapter. The next section continues with the examination of the application of predictive policing, with a focus on publications that were conducted prior to this evaluation study. To summarize, criminological theories that apply to this study are related to theories on prediction, prevention, and place. These include theories on deterrence, rational choice, displacement, and routine activities theory. The possibility that big data can replace theory was also explored in this section.

**Predictive Policing Studies**

One gap within the literature is the lack of verifiable and replicable empirical testing of predictive policing. Few research studies have critically examined the subject of predictive policing. Relating to this research, the pilot study at VPD occurred in 2016; at that time, studies on predictive policing were just emerging. One investigation into predictive policing studies found, despite the high expectations and optimism for predicting crime, that issues arose relating to the evaluation of the models used for the predictions (Berk & Bleich, 2013). In particular, Berk and Bleich (2013) found that studies involving predictive models claimed to be more accurate than older tools; however, when put through some rigorous evaluation, these models were shown to be no better at prediction than traditional regression methods. This evaluation study aimed to address the gap in the literature as well as to ensure that the evaluation process was rigorous. The outcomes of this research may assist in providing an
evaluative roadmap to bridge the gap between industry and academic research on predictive policing. In other words, predictive policing studies published by industries involved in the marketing of technology are different from publications examining criminal justice applications within the academic literature sphere. Both types of literature – the ad hoc evaluative practices reported in the industry and academic studies that adhere to defendable research practices – are examined in the sections below in order to provide a context for the current research.

Of concern is the lack of interest in empirical testing and a bias related to the evaluation of models developed for crime forecasting. The idea that technology and predictive policing or crime forecasting can conserve policing resources is pervasive, even though studies of predictive algorithms have indicated a low level of detail, such as providing general areas of predicted crime locations and using the same statistical model for a full range of crime types, as opposed to examining certain offences selectively. For example, Mohler et al. (2011) demonstrated that algorithms used for seismic data could be used on historical burglary data for predictions. Algorithms are used to predict epidemic-type aftershock sequences and other similar types of earthquake; Mohler et al. (2011) demonstrated through figures and charts that the same formulas could be extrapolated to crime data. They claimed that crimes had a pattern similar to the epidemic-type aftershock sequences, and by making some customized alteration for crime data, the same formula could be repurposed from seismic aftershocks to crime forecasting. Although Mohler et al. used Los Angeles Police Department (LAPD) burglary data in their research, there was no indication of how successful this formula was in predicting crime, and no validation through any practical application of the technology. This is a good example of a study that is academic and far removed from the practices of law enforcement. Highly technical in nature, the figures and charts in Mohler et al.’s report show distributions and value estimates in the form of formulas versus tangible outputs. The content was not useful to the current research, but Mohler et al.’s article offers a good example of the disconnect between practitioners and academics; a gap that the current study fills in the area of predictive policing (2011).

In contrast, there are studies that self-fulfil predictions by loading datasets that directly influence the outcome. Canter, Coffey, Huntley, and Missen (2000) tested Dragnet,
a geographic decision support system, using historical data of 79 solved serial killer cases in the US. For all 79 cases, all serial killer home bases were found to be located within the predicted search area, which reduced the potential search area, as directed by traditional means, by 11% (Canter et al., 2000). In discussing their inquiry’s limitations, Canter et al. acknowledged that their dataset consisted only of those serial killers who had been caught. It could have been a self-fulfilling prophesy, given that only serial offenders who fitted the criteria of committing crimes in close proximity to their home base were included in the dataset. Canter et al.’s research, with their self-reported limitations, is a reminder that many predictive policing models are based on similar data inputs and should be examined for any analogous outcomes that could be a direct result of the type of data fed into the system.

Another example of bias in predictive policing studies involves those that focus only on the positive outcomes of a forecast, while ignoring or failing to examine the prevalence of inaccurate predictions or false positive outcomes. Henry et al. (2014) used historical crime data to create an equation to predict serious violent crime in Chicago. To create the predictive equation the authors made several assumptions, including the notion that minor crimes were the precursors of serious violence (Henry et al., 2014). Henry et al. were involved in the formulating and evaluating of the equation, as well as following up with an application of the model to assess whether it did indeed predict serious violent crime in specific areas. Four of the 10 tracts identified experienced an actual increase of one or more aggravated batteries in summer 2010, but the results were inconclusive (Henry et al., 2014). The prediction equation was built on three seasons and was used to forecast the fourth season. No other data were involved in the study. The authors reported many incorrectly predicted areas, which calls into question the accuracy of the prediction. The methods they used were of interest to this research study, as this was an example of predictive analytics that was deemed to have positive results, whereas in reality it was at best inconclusive. In addition, the researchers were involved in the creation of the models as well as their evaluation, which could be viewed as a conflict of interest.

Kennedy, Caplan, and Piza (2011) conducted another study that examined forecasting, and the results were deemed to be positive. They investigated risk terrain modelling (RTM) and suggested applying place-based intervention strategies over targeting of individuals for the purposes of forecasting crimes. A key principle of their research was
that individuals who are at the greatest risk of committing crimes will perpetrate crimes in specific locations that are deemed to be higher risk, thereby forming the backdrop for RTM; as such, forecasts of crime locations are provided by operationalizing risk factors and assigning a value to the presence of these risk factors. To present the data, GIS layers in the form of a grid are statistically produced, creating a terrain map that yields locations at greatest risk.

A case study of Newark and the application of RTM illustrated the level of detail required in order to operationalize the model and apply the techniques to shooting incidents. Kennedy et al. (2011) emphasized that hot spot analysis had limited value for predicting future occurrences and generalized that place-based interventions might provide a better approach to influencing crime. This assertion fails to take into account significant progress made by European police services, such as the Netherlands National Police, who use a combination of location-based crime strategies and targeted enforcement on high-risk offenders (Willems, van der Mei, & Haasdijk, 2014). It also seems to devalue the credibility of the study, in that there are many examples of law enforcement agencies adopting intervention strategies that are not place-based, yet reporting a significant impact on crime. Kennedy et al. also assert that historical crime data have less significance when considered in the light of computations that examine social, physical, and behavioural factors; this contradicts the prevailing research on predictive policing, whereby other factors (social, environmental, physical) are typically represented to some degree within the crime occurrences (Fitterer, Nelson, & Nathoo, 2015). Issues with RTM also include the ad hoc selection of risk factors as well as the difficulty of obtaining the selected data on a consistent basis and in a timely manner. Kennedy et al.’s results were not compared to the actual outcomes observed in the high-risk areas and, with the exception of hot spot analysis, their research failed to examine RTM in relation to other approaches (Kennedy et al., 2011). The RTM approach is still unproven and is not sufficiently developed to justify some of the sweeping claims made by Kennedy et al. While the authors did use GIS layers, their study was significantly different from the current evaluation research, in that it examined social and behavioural factors which were not included in the current study’s focus on historical crime data.
Academic reviews and articles often refer to the lack of objective and impartial research on the effectiveness of predictive policing technology (Berk, 2011; Hunt et al., 2014; Perry et al., 2013; Robinson & Koepke, 2016; Rosenbaum, 2007; Shapiro, 2017). In contrast, the research by Hunt et al. (2014) – an evaluation of the Shreveport experiment for the Rand Corporation – is often quoted and cited in literature related to predictive policing, and stands out as representative of one of the few studies that adheres to academic research standards and replicability. Hunt et al. conducted a post-experiment review in Shreveport using reliable evaluative research methods and adherence to academic standards. Their report appears unbiased and demonstrates that they were well versed in the subject matter. Their review was sufficient in both scope and detail to enable them to uncover underlying issues with the initial experiment’s methodology and flawed implementation practices. The authors also identified a number of shortcomings in both the type and level of detail of the predictions generated in the Shreveport experiment, as well as the activities of the police service in attempting to respond to the predictions. The predictions generated were in fact hot spot areas with a greater risk for crime incidents compared to the surrounding area.

In the Shreveport experiment, the prediction model was run once a month and provided grid cells of an approximately 125m area, with a medium to high probability (Hunt et al., 2014). After 6 months, the experiment had not resulted in a statistically significant reduction in crime within the treatment area (null effect), and Hunt et al. noted that the predictive maps were not truly predictive, given the latency of 1 month between updates. Attempts to run the statistical model more frequently resulted in little predictive value or change, since the identified locations remained stable over time. In their post-experiment review, Hunt et al. found little difference between reactive hot spot maps showing historical crime and the statistical predictions. Their findings were significant for the current research since their work has been heralded as a reliable predictive policing evaluation study.

As seen with the Shreveport experiment (as cited in Hunt et al., 2014), academics review and evaluate predictive policing implementations from various perspectives. In addition to examining the predictive model used in the Shreveport study, researchers Hunt et al. (2014) also went beyond the technology to look at the human element and how that influenced the model being implemented. Factors such as how police personnel, both line-level officers and management, perceived this policing model had a profound effect on
whether a deployment was successful or not. Impediments to the Shreveport experiment included a lack of coordination and communication with patrol officers which resulted in a diminished interest in the experiment (Hunt et al., 2014). There appeared to be a disconnect between the preventative strategies, the predictions and the practitioner; this served to increase the lag in the preventative strategies being actioned and may have contributed to the null effect. Published in 2014, Hunt et al.’s review remains current, and provides an extensive examination of the attempt by a single police service to implement a predictive policing model, thereby illustrating the issues involved in adopting a new model with limited experience and understanding of the process and resources needed. The objective nature of Hunt et al.’s (2014) report and the level of detail covered highlights their study as one of the few authoritative, objective, and empirically robust reviews on the subject of predictive policing. The current research study also evaluated the pilot following 6 months of forecasting, examining police deployments, and the possibility of changes in the number of crimes, specifically RBNEs.

With the increasing popularity of the use of predictive policing technologies over the past few years, more recent works have been published since the completion of Hunt et al.’s (2014) study, and those with a strong relevance to the current research study are discussed in this section. Gerstner (2018) also recommended further research on predictive policing. His predictive pilot project took place in Baden-Wurttemberg, focused specifically on residential burglary, and the results were empirically evaluated. The pre-crime observation system (PreCobs; Institut für musterbasierte Prognosetechnik, n.d.; see also Jansen, 2018) was utilized to determine the likelihood that a residential burglary would occur in a specific area at a particular time; alerts were sent to local police stations; and anonymized GPS data from police cars were used to measure changes in the density of resources during the alerts (Gerstner, 2018). Gerstner acknowledged that comparing the number of crime incidents from the test period to previous years was common but could not be used as definite proof of the effectiveness of the predictive policing method employed. The differences between the current research and Gerstner’s study are that his study employed qualitative elements, including interviewing operators of the system and conducting online surveys regarding the pilot; and that the data analysis was limited to comparing residential burglary data to previous years and measuring the intensity of police resources via GPS. Thus far, Gerstner’s
study is the closest to the current research in terms of topic and methods. However, Gerstner took a different perspective on the evaluation, since he collected qualitative data from officers about their opinions of predictive policing technology, whereas the current research focused on the efficacy of predictive policing, not only by using data from previous years but also by examining the trends observed in neighbouring jurisdictions and using GIS to analyze spatial-temporal diffusion.

Meijer and Wessels (2019) reviewed predictive policing literature, demonstrating the continued interest in this topic that has developed recently. They found 24 academic articles specific to predictive policing published between 2010 and 2017 (Meijer & Wessels, 2019, p. 2). According to the authors, the first related work was an article published in 2010, while the first book and conference paper were published on the subject of predictive policing in 2013 (Meijer & Wessels, 2019). Of interest to the current research was their conclusion that empirical evidence did not support the claims of a decrease in crime, and that the adoption of these technologies was based on anecdotal evidence. Meijer and Wessels recommended that independent tests should be conducted, and that evidence-based testing should be required in the field of predictive policing. Hardyns and Rummens (2018) also reached this conclusion, providing an overview of predictive policing applications through their literature review and by conducting interviews. The recommendations of both sets of reviewers were addressed by the current research through empirical testing of the predictive policing model and application.

**Ethical Concerns in Predictive Policing**

An emerging concern within civil society is the lack of transparency in predictive policing and the technosocial consequences of algorithmic bias as it relates to civil rights. In August 2016, Robinson and Koepke reported that 20 of the 50 major law enforcement agencies in the US used some form of predictive or crime forecasting tools, with another 11 agencies seriously considering the same procedures. While their examination of the crime forecasting utilized in police agencies resulted in the identification of the differences and similarities amongst the tools, the authors also identified one constant concern, which was that the algorithms that predicted or forecast the crime locations were often a mystery to the agencies that employed them, since the technology itself was considered proprietary. This lack of transparency was also a common concern amongst many researchers (Chan & Bennett-
Moses, 2016; Shapiro, 2017), human rights activists (Lum & Isaac, 2016; Robinson & Koepke, 2016), and statisticians (Datta, Sen, & Zick, 2016). The issues around concerns for civil liberties, individual rights, and privacy matters can be attributed to the trend that most predictive policing and crime forecasting implementations have not been tested empirically for validity and reliability. Nor has there been transparency or clarity on the application of models or the use of the output (Chan & Bennett-Moses, 2016; Edwards & Urquhart, 2016; Lum & Isaac, 2016; Robinson & Koepke, 2016). Furthermore, researchers have noted a lack of empirical assessment of the effects of policing efficacy using GIS. Zhang et al. (2014) found that out of the 1,121 abstracts reviewed, only two studies met the criteria of evaluating the effects of crime mapping interventions by police, and that both of these focused on perception rather than on outcomes. The limited number of empirical assessments of predictive policing models, including the lack of transparency and ethical issues associated with their application, remains a concern. Transparency and ethical concerns are examined in the following sections, along with a conclusion as to how these issues have been addressed in the current research.

Since the introduction of predictive policing in North America, both support and concern have been voiced as to the validity of the technology and the potential for misuse by law enforcement agencies. The main issues of concern have been raised by civil rights advocacy groups, primarily in the US; these relate to the potential bias of the technology, the violation of civil liberties, and the over-policing of marginalized communities (Hirsh, 2016, Lindsey, 2018; Stop LAPD Spying Coalition [SLSC], 2018; Waldman, Chapman, & Robertson, 2018). Within the US, the primary vendors developing and implementing predictive policing are PredPol (2018b), Palantir (n.d.), and HunchLab (Azavea, n.d.; see also Mantello, 2016). High-profile deployments of the technology have been reported from Los Angeles, Chicago, and New Orleans, whose use of this technology is the most widely publicized (Ridgeway, 2018). Taken in context, the US Department of Justice indicated that many police services were exploring the use of predictive policing, and estimated that over 70% would attempt to implement the technology within the next 2 to 5 years (Koper et al., 2009).

One of the reasons for the uptake of predictive policing is the perceived savings to police services (Beck & McCue, 2009). Fiscal restraints and cutbacks affect police services,
forcing many agencies to reduce the number of sworn officers in order to save money; this also drives many police services to adopt a stance of doing more with less and to implement more efficient practices (Beck & McCue, 2009). Predictive policing has been elevated as a means of achieving greater efficiencies because it enables police services to work more effectively with limited resources. Beck and McCue (2009) provided a case example of how Wal-Mart had used predictive technology to anticipate supply chain demands and adjust accordingly before experiencing a product shortage. The authors then attempted to draw a connection between the use of the supply chain technology and predictive policing, from which it appeared that they had an inherent bias toward the merits of predictive policing and often promoted it as solution to many of the economic woes affecting police services. Furthermore, Beck and McCue (2009) accepted the merit and value of predictive policing as a given in their research; they lacked a critical perspective and failed to address some of drawbacks and inherent risks in adopting this model, such as data validity and human rights concerns. They also asserted that predictive policing could prevent crime using fewer policing resources, and their research revealed an underlying assumption that predictive policing worked as claimed by the vendors selling the products.

Contradictory research indicating that predictive policing is no more effective than hot spot policing is not particularly well documented, and a gap exists in the literature in this area of measuring effectiveness. Beck and McCue (2009) suggested that law enforcement had been open to innovation and change over the last 40 years. However, much has been written in academic publications highlighting the reluctance of law enforcement towards change and its adherence to entrenched systems of tradition-bound practices (Bond-Graham & Winston, 2013; Ratcliffe, 2016; Rosenbaum, 2007). The assertions of Beck and McCue (2009) on predictive policing are inaccurate regarding a number of points, such as their description and account of community policing and the first implementation of ILP (Ariel, Weinborn, & Sherman, 2016; Rosenbaum, 2007; Wortley & Mazerolle, 2008). More importantly, however, Beck and McCue (2009) named examples of risk-based police deployments that had no direct relationship to predictive policing. The risk-based deployments cited used historical crime data over a 5-year timeframe to identify seasonal trends and had no predictive aspect whatsoever. Even so, and despite the significant limitations of Beck and McCue’s (2009) study, their research did provide important insight
into the biases seen in law enforcement. Their work also highlighted the disconnect between the realities of predictive policing and the acceptance of marketing propaganda in the absence of credible research to support these claims. The current research includes an evaluation of predictive policing that is free from the potential biases of sales and marketing, since the forecasting model being evaluated is not a commercial product.

While some may dismiss the disparity in perceptions as a typical bias, frequently seen in law enforcement, there appear to be police services that claim significant crime reduction improvements and that endorse specific vendors and predictive policing products (Beck, 2012; Beck & McCue, 2009; Rosenbaum, 2007). Bond-Graham and Winston (2013) conducted an in-depth evaluation of the most prominent predictive policing company in the American market, PredPol (2018a), stating in their article that the future of predictive policing was good branding. Written from an investigative journalist perspective, the article looks at the advent of predictive policing into mainstream policing, with a detailed exploration of the politics behind the recent surge in predictive policing and some of the behind-the-scenes special interest groups and lobbyists that were supporting this uptake in technology. Bond-Graham and Winston exposed a negative side of predictive policing, with supporting interviews that indicated disingenuous marketing, such as contractual clauses that obliged customers to provide favourable testimonies, regardless of the results. Further, a review was attempted of the outcomes in police services where PredPol (2018b) had been implemented but proved impossible, since inconsistent reporting was uncovered, without any clear crime control gains being achieved (Bond-Graham & Winston, 2013). Given the limited research and evaluation into the merits of predictive policing, this article stands out as one of the few critical reviews of a policing trend that is being adopted by an increasing number of police services (Bond-Graham & Winston, 2013). While theirs is not an academic publication and lacks the rigour and standards of scientific research methods, Bond-Graham and Winston did delve into a “dark side” of policing and politics that is rarely exposed. Their article contextualized the current trend toward predictive policing seen in the US, especially in terms of the political and marketing forces at play, and is a commentary on how some vendors misrepresent the use and effectiveness of the technology through a lack of independent analysis or use of research methodology that can be replicated. This was important for the current research study, as Bond-Graham and Winston provide a critical
view of the corporate interests that are motivated by profit-promoting predictive policing technology, an aspect which has not been addressed in academic and/or police practitioner discussions on the topic. There is a need within the academic literature to acknowledge the influence of corporate interests in the adoption of predictive policing technologies.

Ensign, Friedler, Neville, Scheidegger, and Venkatasubramanian (2017) analyzed the code and PredPol Oakland data that was developed by Lum and Isaac (2016) to study the runaway feedback loop of the algorithm used by the PredPol (2018a) predictive policing system. The runaway feedback loop occurs when “police are repeatedly sent back to the same neighbourhoods regardless of the true crime rate” (Ensigh et al., 2017, p. 1). Ensigh et al.’s examination was of interest to the current research since it was an empirical study that investigated exactly what the algorithm does in a predictive system. By simulating the learning mechanism of the PredPol algorithm, Ensigh et al. found that runaway feedback does exist. In the context of this research, the purpose of examining the models by Ensigh et al. was mathematical, and although the results have ethical implications, the focus was only on the outcomes produced by the algorithm with varying data. The results of the modelling found that runaway feedback could be reduced, depending on how crime was defined and inputted into the system. This is an example of literature that examines models without considering ethical and human impacts were they to be applied in a real-world setting. Lum and Isaac (2016) looked at the ethical considerations of what Ensigh et al. referred to as “runaway feedback” when applied to a community by police using predictive policing methodologies, and when these become civil and human rights concerns. In conducting the current study, the researcher remained conscious of the impact that data validity would have on civil and human rights and took steps to mitigate these factors. Literature discussing data validity and human rights, and how the current study specifically addresses these issues, is examined in the next section.

Data Validity Concerns

Dirty data refers, in general, to incorrect or non-standard representations of data. Specifically in the case of policing, there is concern that this could include data units that are “derived from [or] influenced by corrupt, biased and unlawful practices, including data that has [sic] been intentionally manipulated or ‘juked’ as well as data that is [sic] distorted by individual and societal biases” (Richardson, Schultz, & Crawford, 2019, p. 4). While poor
data generated as a result of human or technical errors or mistakes, such as typos or missed information is of concern, dirty data focuses specifically on knowingly-produced misinformation or information that results not only from individual and societal biases, as Richardson et al. (2019) have stated, but also those that exist in police organizations and culture and that translate into practice and policies. In the case of policing, the possible cost of dirty data is more significant than it would be in a field such as marketing. The intention may be good, but the resulting compilation of data for further use for other purposes may translate into unintended consequences.

One such possible data source in Canada is the practice of street checks or carding, in which the police’s intended purpose of keeping records of interactions with the public has turned into a dataset, thereby revealing that there are racial disparities among those being stopped and questioned by the police. An investigation by the Ontario Human Rights Commission in 2018 found that there was “a lack of legal basis for police stopping or detaining Black civilians in the first place; inappropriate or unjustified searches during encounters; and unnecessary charges or arrests” (Executive Summary section, para. 2). The investigation revealed that even though Black civilians constituted only 8.8% of Toronto’s population, they accounted for 28.8% of ‘use of force’ cases, 36.0% of shootings, 61.5% of deadly encounters, and 70.0% of fatal shootings in the city (Ontario Human Rights Commission, 2018). Although there is no direct evidence linking the over-representation in violent crimes to the disproportionate volume of street checks, it is possible that the two are related. It is also possible that officers who interacted with the community exercised discretion and conducted enforcement action in good faith; however, despite their best intentions, their own personal biases may have skewed or dirtied the data units collected.

Once collected, the data are entered into a system and represented in the database in the same format as other data points. Thus, regardless of what the biases or the underlying reasons might have been for collecting the data, the information gathered becomes normalized and presented like all other data, with the presumption that it is valid and can be taken at face value. It is difficult, if not impossible, to determine the motivations and biases of individuals and organizations that have led to the collection of certain data, and it is even harder to assess omissions of data and the reasoning behind those decisions. It is
safe to assume that any set of data will never be value free, since its collection was based on
the idea that the data points have some value.

As potential sources available for mining and usage expand, there is growing attention
towards the inherent biases contained within historical data being fed into predictive
policing models (Hvistendahl, 2016; Lum & Isaac, 2016; Robinson & Koepke, 2016; Shapiro,
2017). In this regard, Hassani, Huang, Silva, and Ghodsi (2016) reviewed the data mining
techniques and functions of over 100 applications in crime analysis, and then classified each
into five main categories: entity extraction, cluster analysis, association rule, classification
techniques, and social network analysis (p. 141). Hassani et al.’s results showed that in the
context of crime analysis, big data can vary significantly and offers endless opportunities for
further applications within the law enforcement milieu. There is no shortage of studies that
test new algorithms for hot spots and crime forecasting using crime data in the US (Adepeju,
Rosser, & Cheng, 2016; Chamberlain & Boggess, 2016; Chastain, Qui, & Piquero, 2016), in
Canada (Tayebi, Glässer, Ester, & Brantingham, 2016), along with studies conducted in
Europe (Bertozzi, Johnson, & Ward, 2016), and in countries such as Brazil (Chainey & da Silva,
2016) and India (Datta, 2016). The algorithm applications are limitless for academic and
modelling exercises on specific offences such as gun crimes (Drawve, Moak, & Berthelot,
2016), as well as all crimes and efficient patrolling locations (Ndunge & Shibwabo, 2016).
However, the real-world applications and the use of the outcomes and outputs should be
examined carefully, since there are inherent biases that may influence the products created.

Most predictions and forecasting of crime rely on existing crime data, and the
algorithms calculate what may happen in the future based on the information that has been
ingested into predictive policing models and technologies. Depending on the potential
biases inherent in the data, the model might result in replication and possible amplification
of the information biases that already existed in the data. The result, “at best, this renders
the predictive models ineffective; at worst, it results in discriminatory policing” (Lum & Isaac,
2016, p. 2). The dataset used by the system is based on crime incident history collected by
the police, which is not representative of all crimes. Areas selected to be policed are not
random, since they depend upon both the police and the citizens of the community who
choose to report those crimes, which leads to “selection bias meets confirmation bias” (Lum
& Isaac, 2016, p. 3). The bias is perpetuated, as the predictions based on the biased data will
only increase the presence of the police in the very area that initially had been over-policed. These areas then continue to be over-represented in the dataset from which the system learns. As Crawford (2016) states, “Predictive programs are only as good as the data they are trained on and that data has a complex history” (p. 22). This contextual component of the data that feeds the predictive software and programs cannot be ignored when evaluating the efficacy of any predictive policing applications. Civil and human rights concerns are addressed in a later section in this chapter; however, it is important to note here that it is the use of historical data that contributes to the concerns raised regarding biases.

Although some of the concerns are with the predictive methods and data being used to forecast crimes, there are also concerns surrounding the actual use of the outputs by the police. According to Shapiro (2017), “Checks and balances are needed to mitigate police discretionary power. We should be wary of relying on commercial products that can have unanticipated and adverse effects on civil rights and social justice” (pp. 458–459). The validity of the predictive policing systems currently being used without independent analyses or reviews is called into question, and without the ability of the police to question how the methods work, oversight of police actions is also stymied (Hirsh, 2016). The current study addressed this directly by independently assessing the predictive system, examining factors such as the data sources used, and the application of police resources based on the outputs of the predictive system.

Unlike circumstances in which police services have implemented new technology, such as a new radio system, predictive policing requires carefully calibrated measurements and large volumes of unbiased data that have been de-duped and error-corrected in order to support machine learning. Data quality can be the enemy of predictive policing, since unforeseen systemic issues that penetrate most organizations and that are related to quality control, standardized coding, and inherent bias, can affect whether or not an implementation will be successful. These issues are difficult to control and often derail organizations seeking to obtain productivity gains with machine-learning processes. For example, in circumstances in which the data contain location bias and the information is used to generate forecasts, the subsequent deployment of police resources will further reinforce and exacerbate this issue. In this situation, the common outcome is for the
predictive system to increase police presence in those areas where the police were originally active and engaged in crime control measures. The data collected continues to corroborate the over-representation of resources, and the machine learning reinforces this bias (Lindsey, 2018; Ridgeway, 2018; SLSC, 2018). Ultimately, “predictive programs are only as good as the data they are trained on and that information frequently has a complex history” (Crawford, 2016, p. 22).

Mathematical models have been evaluated primarily in concept and are hypothetical in nature, which has meant that when the system was put into practice its efficacy was unknown. In the current study, the researcher remained conscious of data sources that were fed into the system to ensure that the datasets did not directly influence the outcome and become self-fulfilling predictions. This initial step is an important one, as data validity and data biases may be inherent and will need to be examined prior to any efficacy being tested on predictive products or applications. As was the case with the current study, it is important to assess the enforcement strategies when planning the operationalization of the model (e.g., high-density, multi-story apartments may not be conducive to certain prevention or enforcement actions).

Data collection for the current study included 4 years (2012–2015) of aggregate crime incident data for the study area and the surrounding region for the purposes of establishing a baseline measurement of crime trends and patterns. The 4 years-worth of residential crime data included data units for neighbouring municipalities and a comparison city within the region (i.e., the City of Surrey), that were used in the current research for the purposes of temporal and spatial comparison. Data were collected over the 6-month period of the pilot project for the area under study, as well as the bordering municipalities and the City of Surrey. Percent changes for the pilot study months compared to the previous 4 years were also obtained. Crime trends for Vancouver and the three bordering cities, including Surrey, were determined from monthly fluctuations in the previous year, and a histogram was created to capture the trends for the same time period during the pilot project, which provided insight into the distribution of the RBNE variable across multiple geospatial jurisdictions. In addition to the trends by jurisdiction a further two datasets were added to the trend graph, representing the study time with resource allocations and the control study time between 1601 and 0759 hours. Although the operationalization was outside the scope
of the current study, it is worth noting that the time for resource allocation and the control was determined by aggregating the previous year’s temporal evaluation of RBNE by day of week and time, while the highest concentration of crime was determined and used as the study period. The availability of the forecasting information from the control hours allowed for an evaluation of the forecasting model without police resources and the impact it might have had on the model at this time.

Human Rights Concerns

At the forefront of human rights concerns in policing are AI-based crime forecasting systems and the issues that algorithmic policing has created, such as potential legal, ethical, and civil liberties implications. Within the field of predictive policing, the act of attempting to predict persons at risk of offending is the most controversial. Police departments in Los Angeles, Chicago, and New Orleans have received the most attention and their use of predictive technology has raised the spectre of violations of individual rights, abuse of civil liberties, and infringements of privacy rights (Ridgeway, 2018; Saunders et al., 2016). The type of data used by these systems differs from that of location forecasting since it relies on person-specific details, identified associates, and nodes of activity. For the most part, the data units are subjective in nature and dependent on an individual officer making specific judgements and assumptions when reporting on individuals. In many cases, police intelligence files are used to augment the evaluation of an individual’s risk of committing an offence (SLSC, 2018; Waldman et al., 2018). The risk assessment of a person’s potential to commit an offence is based on the weighting of specific factors that are felt by the software designers or the police themselves to have predictive value. Factors such as number of associates who are linked to gang activity, prior gun violence, seriousness of criminal charges, and use of narcotics are collected and incorporated into the prediction (Ridgeway, 2018; Saunders et al., 2016). Generally speaking, many of these weighted factors are arbitrary in nature, with very little empirical evidence to support their predictive value (Mantello, 2016).

Contextual and data bias factors can be seen in the way Los Angeles has used both Palantir (n.d.) and PredPol (2018a) technology. Citizen groups have expressed concern that the LAPD is disproportionately policing neighbourhoods that are predominately Hispanic and African American (Mantello, 2016; SLSC, 2018). The same groups criticize the police for being more akin to an occupying force rather than a community service, because law-
abiding citizens of these communities are regularly subjected to indiscriminate checks that are disproportionate to checks conducted in other neighbourhoods (Lindsey, 2018; SLSC, 2018). Significantly, neither Palantir (n.d.) nor PredPol (2018a) have independently verifiable mechanisms for ensuring the ethical use of crime forecasts. The prevailing assumption regarding data ingestion is that “more is better” and that every police contact and incident should be used by their systems, despite the absence of any mechanism to ensure unbiased outputs (Ridgeway, 2018; Waldman et al., 2018).

Many sources have heralded predictive policing as a solution for a range of issues, including gun violence and suppression of gang crime, and a number of studies have highlighted both empirical and ethical concerns in the application of predictive policing models (Camacho-Collados & Liberatore, 2015; Farrell & Pease, 2014; Goode, 2011; Maguire, 2000; Malik, Maciejewski, Towers, McCullough, & Ebert, 2014; Saunders et al., 2016; Smith, 2018). The Chicago Police Department piloted what it regarded as predictive policing using the strategic subjects list (SSL), while Saunders et al. (2016) evaluated the Chicago Police Department’s pilot. They focused on the impact of the prevention strategy but did not explore whether the approach was valid or reliable in predicting the crimes. The evaluation study consisted of 43 semi-structured interviews in 22 police districts over the course of 9 months. In addition, observations were made from 48 CompStat meetings (Saunders et al., 2016, p. 361).

The SSL was a project between the Illinois Institute of Technology and the Chicago Police Department and was aimed at reducing gun violence by focusing resources on individuals identified as being at highest risk of such violence. This was done by generating a list based on predetermined scoring criteria through examining previous homicide victims and common precursors of violence. These factors were developed into scoring metrics with an automated process that ranked ordered individuals deemed to be high risk (Saunders et al., 2016). The results of the evaluation found that there was little or no attention given to the SSL. Of the 405 homicides in Chicago that had occurred during the year, only 1% were identified in the SSL. This equated to just three people involved in homicide from the list of 426 individuals identified in the SSL, further reinforcing that there are limits to the application of predictive policing and highlighting that there are challenges associated to the wider application of machine learning. In practice, the list assisted only marginally in
identifying possible individuals for arrest for gun violence rather than for prevention of gun crimes (Saunders et al., 2016). This also raised privacy and civil rights concerns, since the intention in generating the list was to reduce gun crimes and be preventative in nature. An individual in the SSL group was more likely to be arrested for a shooting, but no more likely to be the victim of a homicide or shooting than individuals in the comparison group (Saunders et al., 2016, p.34). This approach, while classified as predictive by the police service, identified a variety of issues, such as how risk assessment tools could generate potential subjects of interest, and the way they were used, which could have unintended consequences. The SSL is different from other predictive and crime forecasting models in that it focuses on individuals rather than geospatial areas (Saunders et al., 2016). This approach needs to be considered and applied with caution as it has the greatest potential implications for privacy and civil rights.

Similar to the Chicago Police Department’s SSL is the LAPD’s adoption of a program built by Palantir (n.d.), named Gotham. This was deployed as part of a project called “Operation LASER” (Lapowksy, 2018, para. 2), the acronym for Los Angeles’ strategic extraction and restoration, with the key objectives of identifying and deterring people likely to commit crimes (Lindsey, 2018; Waldman et al., 2018). The system ingested details on a range of individuals, incorporating police interviews, police contact reports, criminal history, parole reports, and gang intelligence reports. Other outsourced data, such as pizza delivery records, social media, and toll road data were also rumoured to be part of the data sources (SLSC, 2018; Waldman et al., 2018). Civil rights advocates argue that this evaluative approach and the use of technology simply reinforce police biases and racial stereotyping, while the human rights advocacy group SLSC (2018) argue that Operation LASER (Lapowksy, 2018, para. 2) has reinforced racial biases and that the system has resulted in the labelling and criminalization of individuals with no prior criminal history who are only loosely associated with a known offender. The main issue is the lack of transparency in the way the system categorizes someone as being “at risk to offend” and, in many cases, the data used by the system are influenced by the subjective views and personal biases of police officers (Lindsey, 2018; Mantello, 2016; SLSC, 2018). The technology is akin to a scoring matrix based on a set of criteria determined by Palantir (n.d.). The validity of the criteria is questionable, and the project outcomes have yet to be determined.
Shapiro (2017) examined another popular predictive system, HunchLab (Azaveam, n.d.), a machine-learning AI algorithm and its application. As an ethnographer, Shapiro concluded that the system “operat[ed] in good faith” (p. 458) but lacked oversight, while concerns around the bias of the original data that fed the system were not addressed. Chan and Bennett-Moses (2016) argued that, due to the use of user-generated data sources and the lack of transparency and causal explanations of the algorithms, the expectation of trusting the output of the process has itself been called into question. An attempt to measure the influence of any input into an algorithm shows that more transparency is required before understanding how and what has contributed to a prediction. However, this transparency needs to be balanced with privacy protection, since reporting on the inner workings of the system could lead to identifying both citizens and individual records that have contributed to a prediction.

Most of the articles related to forecasting equations and modelling do not discuss the possibility that there could be negative costs associated with crime predictions. However, Ferguson (2012, 2015) attempted to address some of the Fourth Amendment issues related to crime forecasting. Ferguson (2012) considered the legal terms “probable cause” and “reasonable suspicion” (2012, p. 1) in various applications of predictive policing, and also emphasized the importance of understanding and evaluating predictive policing methods, technologies, and products, including measuring the accuracy of the model. Ferguson (2015) also argued that transparency in the way the model operated was necessary for the technologies to be able to withstand legal challenges, noting that new sources and the volume of data had influenced the ability for law enforcement to develop “reasonable suspicion” (2015, p. 329), and that the reasonable suspicion test, as it developed historically, would not be applicable given that greater volumes of information and more individualized data would become available about the suspect. The human rights and constitutional issues raised by Ferguson (2015) are also applicable in the criminal justice system of countries other than the US, including Canada, Australia, and the United Kingdom.

Walsh and Miller (2015) argued that ethical and policy issues, including methods, context, and targets relating to intelligence collection for the purposes of national security, had to be re-examined, since the Edward Snowden intelligence leaks had raised public awareness of privacy issues surrounding the practices of security services. The same
suggestions were applicable to the practices of crime forecasting and predictive policing, in that citizens needed to be informed and to be provided with information as it related to the methods being employed, transparency regarding the technology itself, and the steps taken to ensure compliance with the legal framework, including ethical guidelines and privacy policies. The cost of predictive policing had also to consider any possibility of infringement of citizens’ civil and constitutional rights (Weisburd, Feucht, Hakimi, Mock, & Perry, 2011).

Many of the issues related to civil rights violations and the over-policing of marginalized and ethnic minority communities in the US do not apply to the VPD’s experience with predictive policing. Focusing exclusively on property crime, vetting police-generated reports for the data importation process, and implementing specific solutions to guard against crime-forecast hot spots in marginalized segments of the City of Vancouver (see Chapter 5, Data Concerns), help to ensure that the issues raised in Los Angeles and elsewhere in the US do not apply to the VPD’s use of crime forecasting technology. However, this does not imply that the use of machine-learning and AI technology is risk free.

This second part of the literature review focussed on the definition of AI, big data, and predictive policing, and included associated studies and trends discussing ethics and theories. The next section focuses specifically on the definition, studies, and theories related to RBNE, the crime that was forecast by the VPD pilot study.

Residential Burglary Definition

As the current study specifically examined the capabilities and effectiveness of a predictive system as applied to a single crime type, a brief description of the type of activities being predicted is given below. This study investigated predictions of RBNEs; under the Canadian Criminal Code (1985), this crime is defined as “the offence of breaking and entering into a dwelling-house through illegal entry for the purposes of committing the offence of theft” (s. 348, para. 1). The Canadian Centre for Justice Statistics (Statistics Canada, 2019) established the national scoring and data collection rules for police services relating to the offence of RBNEs, which follow the national Uniform Crime Reporting Survey (Statistics Canada, 2019). Statistics Canada (2019) also designed Uniform Crime Reporting to measure the prevalence of crime in Canadian society, including inherent characteristics of crime incidents, in order to facilitate crime analysis, resource planning, and policy development, while at the same
time protecting persons charged and victim information (Statistics Canada, 2019). For the purposes of the current study, the terms residential burglary and RBNE were used interchangeably, as this practice is reflected in the literature (Santos, 2016). Criminological theories that are applicable to RBNE, or more broadly to property crimes and how this might contribute to the methods, results, and discussion of this evaluative study will be discussed next.

**Environmental Criminology Theories on Property Crimes**

Environmental criminological theories, as discussed in the “Predictive Policing Definition and Trends” section of this chapter, have been tested in practice settings with varying outcomes on property crimes (Barr & Pease, 1990, Brantingham & Brantingham, 1981; Cornish & Clarke, 1987). For the current study, the researcher was interested in theories derived from environmental criminology as applied to property crime, specifically RBNEs, and related to the concepts of predictive policing. Routine activities theory (Cohen & Felson, 1979) and rational choice theory (Cornish & Clarke, 1987) were particularly relevant to this study. These theories both assume that distribution of crime is due to motivated offenders being in the same time and space as suitable targets (Cohen & Felson, 1979; Cornish & Clarke, 1987). Several studies have demonstrated that property crime, specifically RBNEs, is one of the most patterned, and therefore predictable types of crime (Farrell & Pease, 2001; Johnson et al., 2007; Short et al., 2010; Short, D’Orsogna, Brantingham, & Tita, 2009). Past crime data are the most basic used for crime predictions and are based on the well-established theory of repeat victimization (Bachner, 2013). The theory states that people and places which have been victimized have a higher likelihood of being victimized again (Farrell & Peace, 2001, p. 18). The assumption that future incidents can be predicted from past data using systematic components is foundational to predictive policing for property crimes (Bachner, 2013). An exact-repeat burglary event is a consecutive burglary at the same location and near-repeat burglary events are those within a predetermined spatial area (Short et al., 2009). Repeat victimization and near-repeat victimization theories have been used to examine armed street robberies (Haberman & Ratcliffe, 2012) and domestic burglaries, and have been found to be a useful application (Chainey & da Silva, 2016).

Risk heterogeneity is a term that also relates to victimization risk, in that some places are at a higher risk of being targeted by offenders when compared to other locations, due to
their place-based risk. Short et al. (2009) found risk heterogeneity to be applicable to time as well; these authors developed mathematical models to predict property crimes using residential burglary data from Long Beach, California. However, their results found that a repeat and near-repeat burglary model, based on risk heterogeneity alone, could not explain the patterns of victimization (Short et al., 2009, p. 82). In seven of the nine studies examined by Braga (2001), applying hot spot policing through the use of past crime data to predict RBNE incidents resulted in crime reductions. This was of interest to the current research, as Braga’s (2001) results had demonstrated that RBNE was a crime that could be successfully studied using pattern analysis techniques, although the difficulty for the author lay in assessing the effects of displacement and diffusion. Chastain et al. (2016) developed a model that combined multiple criminological theories, namely journey to crime, social disorganization, and routine activity theories, to predict residential burglary locations based on historical crime data from the Dallas Police Department. These theories were chosen as they had been applied historically in an attempt to explain residential burglary target selection in differing but not mutually exclusive ways (Chastain et al., 2016). The current study also utilized routine activities theory (Cohen & Felson, 1979) and rational choice theory (Cornish & Clarke, 1987), since they relate to the distribution of crime and have been applied to property crime, specifically residential burglaries.

Targeted resource deployment in response to areas identified as being at high risk, as is the case with hot spot policing, was examined by using a weighted displacement quotient to analyze crime reduction in Camden New Jersey (Ratcliffe & Breen, 2011). Hot spot policing had been implemented and put into practice in numerous countries with varying degrees of success, but according to the authors (Ratcliffe & Breen, 2011), more studies were needed in this area, particularly those that used a rigorous and replicable methodology (Frogner, Andershed, Lindberg, & Johansson, 2013; Gerell, 2016), because many studies were unable to assess or evaluate the efficacy of the actual interventions used in an effort to reduce crime (Paternoster, 2010). Resources were deployed in response to the hot spot areas, and the researchers found a decrease in crime in this targeted area as well as an added benefit of diffusion to other areas that were also experiencing the decrease (Ratcliffe & Breen, 2011). Such crime reduction and diffusion applied specifically to property crimes (Ratcliffe & Breen, 2011). Ariel et al. (2016) examined whether the
positioning of police community support officers (PCSOs), which are resources with limited arrest powers and no weapons, could reduce crime in a hot spot for the purpose of determining if presence alone was sufficient for crime reduction in a high-risk area.

The directed resource allocation of the PCSOs was of interest to this research study, as the VPD pilot also used community police officers, who have limited powers and no weapons. As in Ariel et al.’s study, the role of these community police officers in the predictive policing pilot was to act as a deterrent solely through their presence. The authors did not identify what activity was conducted at the location when the PCSOs attended the hot spot, as they were tasked with being visible in the location during the time specified. Outcomes measured were the number of calls for service during the 12-month period of the study, which was then contrasted with data from the 24 months preceding the study. Ariel et al. analyzed the changes in the treatment and control areas before and after their study, and found that the calls for service were reduced by approximately 20% and that victim-generated crimes were reduced by 39% by the PCSO patrols when compared to the control data (Ariel et al., 2016, p. 295). They also measured the effectiveness by diffusion of benefits of the intervention, using standardized mean differences. Although Ariel et al.’s study was located in the United Kingdom, given the similar objectives, the effectiveness of PCSO presence in their study was relevant to the current research, which utilized GPS, location-based forecasting and deployment of resources to the location to measure the effectiveness of the police resources deployed in reducing RBNEs in the City of Vancouver. One difference from the current study is that the hot spots in Ariel et al.’s inquiry were constant, and the researchers did not measure any changes the resourcing might have had on the hot spot until the conclusion of their study period; nor did their findings speak to the effectiveness of the hot spot itself for accuracy of crime forecasting.

There have been further deterrent studies using the GPS locator on police vehicles and incident records (Davies & Bowers, 2015). The use of GPS data from police was of interest to the current study, since the location information from this research was obtained through police GPS data. The authors compared crime incident data against the GPS data, which contained street segments, time, travel speed, and whether the police vehicle stopped (Davies & Bowers, 2015). The purpose of the comparison was to determine whether there was a relationship between the GPS data and the crime data that would
quantify deterrence effects. In an approach similar to that employed by Davies and Bowers (2015), the current study compared the GPS data and the RBNE crime data in order to quantify the deterrence effects.

Some researchers argue that in order for a police resource presence to be effective as a deterrent, the offender must perceive a potential increase in negative consequences (Koper, 1995; Paternoster, 2015; Weisburd et al., 2005). In other words, there is no deterrent if the offender does not associate the police presence as a negative consequence, or if there is a lack of connection between the increased police resource presence and negative consequences on the part of the potential offender. Researchers have concluded that at least some decrease in criminality can be attributed to police presence (Koper, 1995; Paternoster, 2015, p. 250; Weisburd et al., 2005); however, studies based on police presence alone found that the number of police officers and the rate of homicide, robbery, and burglary rates were inversely related (Marvell & Moody, 1996).

Deterrence theory is connected to displacement theory, in that a rational offender would consider the greater risk of the negative consequences of committing a crime in a location with increased police presence, and to commit the offence would simply relocate to another area with a perceived lower risk (Paternoster, 2010). The VPD pilot study inherently incorporated these criminological theories, given that resources were deployed to a location at a time when RBNE predictions were provided by the spatial-temporal forecasting model, with the assumption that the presence of police resources as guardians would influence the decision-making process of a potential offender. Although deterrence is a core component of the criminal justice system, the effect of deterrence has not been empirically tested in a way that can be called credible (Eck & Maguire, 2000; Horney & Marshall, 1992; Kleck, Sever, Li, & Gertz, 2005); as such, academics regard the supporting research as “flimsy” (Paternoster, 2010, p. 766). Paternoster (2010) argued that it is difficult to determine how strong an effect deterrence might have because it is hard to isolate and measure, and it is equally difficult to measure its effect through outcomes of the criminal justice system. However, previous research has also supported the theory that police presence can reduce crime in a hot spot (Koper, 1995; Weisburd et al., 2005). Researchers have found that crime reductions in the range of 10–25% are possible when hot spot policing is applied properly and used to direct patrols (National Academies of Sciences,
More specifically, a police presence led to higher rates of arrest and a decreased rate of burglary (Wilson & Boland, 1977), while Sampson and Cohen (1988) also found an increased police presence led to higher rates of arrest and decreased crime rates for burglary and robbery in 35 cities within the US. Two other studies conducted on deterrence found an inverse relationship between police presence and crimes, including burglary rates (Levitt, 2002; Marvell & Moody, 1996). This was of interest to the current study, since the predictive policing pilot would only be effective in reducing the number of residential burglaries in the study area if the deployment of police resources acted as a deterrent to the crime.

As noted earlier in the chapter, RTM, which was created to forecast shootings, has also been modified and adapted to forecast other infractions including property crime (Caplan & Kennedy, 2010). The RTM analysis uses GIS data to assess crime risk through map layers that represent spatial influence and crime intensity (Caplan & Kennedy, 2010). RTM analysis can also measure risk heterogeneity by presenting underlying factors in the environment that contribute to crime by presenting all risk factors in a given location, and by focusing not only on the crimes to identify, but also on assessing why a crime might occur in that particular area (Moreto et al., 2014). Moreto et al. (2014) created a model that incorporated both RTM analysis and a near-repeat calculator to forecast residential burglaries, and concluded that risk of burglary increased in certain locations because they had factors that made these crimes more likely to occur and attract potential offenders. Although there is a potential for the models to be able to predict future crime occurrences, the predictive studies and methods associated with mathematical modelling are hypothetical in nature, derived from data that did not occur in real time, and were not operationalized.

Perry et al. (2013) noted that regression, classification, and near-repeat modelling were some of the mathematical methods used to predict crimes. Of particular interest to the current study was Fitterer et al.’s (2014) mathematical model, which was used to examine the City of Vancouver property crime rates to predict RBNEs. Unlike hot spot policing, which focuses primarily on historical crime data, Fitterer et al.’s model, which was founded on near-repeat theory, incorporated datasets on urban environment, socio-
demographics from census data, road network information, graffiti data, and tax information. This primary framework became a proof of concept for the further development of the VPD crime forecasting model that formed the basis of the current study. Compared with the Moreto et al. (2014) model and the Fitterer et al. (2014) model, the approach in the current study was of greater depth, since it enabled the researcher to complete an examination and evaluation of a model that has been used in a real-world setting and operationalized. In contrast to Moreto and Fitterer, and to other frameworks and models, this study filled a gap in the academic literature by applying a replicable methodology for evaluating crime forecasting to direct police actions.

This section of the literature has focused on the context of the evaluation in this study, which was predictive policing and its effect on RBNEs. Specifically, the ethics, bias, and data issues and theories as they related to this study were discussed. Related to ethics and bias, the manner in which police services adopted new technology was of interest to the current study since it was an important factor influencing the adoption of crime forecasting technologies by the police. The next section examines possible theories and explanations regarding policing and adoption of technology, which directly relates to the rationale and validation being used by many police services to justify the adoption of what might be considered unproven predictive policing solutions.

**Policing and Technological Adoption**

Increasing numbers of law enforcement agencies around the world have expressed interest in predictive policing. As at August 2016, more than half of the 50 major law enforcement agencies in the US were using some form of predictive or crime forecasting tools or were seriously considering the adoption of such practices (Robinson & Koepke, 2016). In Europe, predictive policing capacities are increasing and are being developed to meet their specific operational requirements. For example, in Germany and Sweden, several police agencies are utilizing PreCobs, developed and built by the Institut für musterbasierte Prognosetechnik (n.d.; see also Jansen, 2018), while the Amsterdam Police built their own Crime Anticipation System (Willems, 2014; Willems et al., 2014). Several police agencies in the United Kingdom, including the London Metropolitan Police, have tested the adoption of American systems, such as PredPol (2018a) and Palantir (n.d.) but are also considering developing their own options (Jansen, 2018). Police services eager to look progressive and
proactive are adopting predictive policing without undertaking decision-making based on realistic outcomes, protection from biased policing, and upholding community expectations.

The progressiveness and performance of a law enforcement agency is believed to be influenced by the technology used by the department (Gottschalk & Holgersson, 2006), and, with many police agencies adopting new technology, those that lag behind are looking into keeping up with the perceived trends in policing. It is often the case that policymakers and executives who make decisions on technology adoption do not take into account the end users or the public who are affected by the choices that are made regarding policing technologies. Policing has evolved over time with the technological advancements of society, as was the case with the adoption of cars, radios, and dispatch systems (Weatheritt, 1986). These informal and subjective influencers on decision-making in police services have a direct role to play as to whether valid and defendable research will be undertaken, particularly if it contradicts a prevailing viewpoint (Griffiths, 2014; Mastrofski, 2006).

The use of new technological advancements is now influencing policing (Bowling, Iyer, Reiner, & Sheptycki, 2016; Bowling, Marks, & Murphy, 2008) in a way that previous eras have not, since all facets of administrations and operations, including resource allocation and investigations, are being reached. Diffusion, not in the criminological context, but from a marketing and communication standpoint, is “the process in which an innovation is communicated through certain channels over time among the members of a social system” (Rogers, 2003, p. 21). Diffusion can relate to a manufacturer of social change, whereby an “alteration occurs in the structure and function of a social system. When new ideas are invented and then diffused, they are either adopted or rejected” (Rogers, 2003, p. 28). This adoption or rejection results in consequences leading to social change. The main elements of diffusion then facilitate this process from innovation to social change or, according to Rogers (2003), the process follows through the continuum from “innovation, communication channels, time and social system” (p. 22).

This same phenomenon applies to predictive policing, in that the concept of predictive policing is perceived as an innovation, a new idea to the police and the law enforcement community. The term ‘innovation’ is synonymous with technology in many diffusion studies (Rogers, 2003). This process of innovation may then apply to new or existing technology that is known to have been adopted in other circumstances, although an
individual may still be assessing the application and relevance of the innovation. In a police service in which a practitioner may become aware of the use of technology by other police services, the innovation may be strictly technical and not well established in the industry; however, it may be the only one that has come to that person’s attention. Well-advertised predictive software products can fall into this category because awareness has been brought to the police organization over other unknown but potentially more appropriate solutions. A new software or type of analysis may have been introduced at a conference or by vendors, and the knowledge of the innovation may have already been present in the organization but not in the knowledge base of the decision-makers, nor at the stage of being evaluated for possible adoption.

The evaluative nature of the current study helped in providing empirical evidence to determine whether predictive policing technology should be adopted as part of the VPD operations beyond the pilot study. In addition, the perception of compatibility and relative advantage of a new technology is essential in determining if it will be adopted successfully into an organization (Rogers, 2003, p. 20). In the case of predictive policing technology, the relative advantage is usually the perceived cost savings based on strategic resource allocation. The emphasis on compatibility with existing norms is important since the resources tasked to action the results of the technological output, such as location of a crime prediction, are expected to be deployed, altering some of their routines but without making changes in their norms such as basic training and protocols. Compatibility must also be considered for existing technologies. According to the technology adoption model which examines organizational adoption of technology in economics (Davis, 1985; Chuttur, 2009; Venkatesh & Davis, 1996) and which is highly applicable to police organizations, perceived usefulness and ease of use are also key factors.

In the case of policing and historical adaptations, there appears to be a trend within law enforcement, as well as an awareness that policing has entered the information era requiring technologies to analyze and respond to crimes (Rosenbaum, 2007). In addition, the value and merit of a police service, as measured by citizens, are perceived through professionalism, response time, and arrest rate, as well as through use of technology (Chu, 2001). Employing a focus group method, Colvin and Goh (2005) applied the technology adoption model to the adaptation of technology by police officers; they found that the four-
factor model of the exploratory factor analysis, namely “ease of use, usefulness, information quality, and timeliness, was a better theory for this population” (p. 43). The authors stated that technology had an impact on police practices. In their study, the police personnel were the subject matter experts, not the policymakers or the executives who typically influenced the actual implementation or purchase decisions of a particular technology.

For the current study, the resources deployed to the predictive locations for preventative and enforcement activities did not include individuals tasked with reviewing, assessing, and making strategic decisions on whether predictive policing practices, models, and processes should be implemented. The police resources deployed to the locations were non-voluntary. Thus the current evaluative study was important, not only to assess the effectiveness of the pilot project on outcomes, but also to assist with further implementation, given that the results were positive. It is important to note that if the adoption is not accepted at all organizational levels, particularly by those who are using the system and being deployed, there is a possibility that the technology will not be used to its full potential. The purchase and implementation of predictive technologies are separate from acceptance. In addition, perceptions of being proactive, intelligence-led, and evidence-based are important for police services, in that the community in which they serve must view officers as capable and competent (Chu, 2001); to that end the current evaluation assisted in providing the empirical evidence on the predictive policing pilot project.

There is also the belief that technology in law enforcement is justified and a necessity that must be maintained and advanced, because criminals constantly make use of available technologies for their own benefit and to victimize others (Nunn, 2001, p. 12). This perception justifies the implementation of different types of policing technologies, including predictive policing (Nunn, 2001). Technology is believed to affect the operation of the criminal justice system in policies and practices, and the implementation of certain technologies has changed the ways in which law enforcement is able to operate (Holt, 2013). Technological adaptation in the area of predictive policing by law enforcement needs to be approached with caution and transparency. The current research assisted in providing transparency in the way efficacy was defined and measured in the VPD pilot study. Due to the desire for police agencies to become technologically advanced and to increase their perceived effectiveness through use of the newest and the latest technology, it is important
to examine how research is being conducted in this area, what is available for law enforcement, and whether it is possible to make informed choices on implementation.

This concludes the literature review, which explored the main characteristics of AI, machine learning, big data, and predictive policing that form the foundation of this study and are instrumental to the evaluative process used in the study. Furthermore, key policing concepts related to the research of property crime and associated studies, theories, and trends, which help to contextualized the research question within the existing literature, were also examined. The methodological approaches used are explored in Chapter 3.
CHAPTER 3 — METHODOLOGY

**Methodological Approach**

This chapter discusses the methodological approach of the study, including the ethical considerations, justification of methods, data collection, processing and analysis methods and the context of the research. To answer whether or not a machine-learning, spatial-temporal crime forecasting model used in an operational police deployment was effective, an evaluation study was conducted using quantitative crime analysis data collection and analytical methods. This is an evaluation research, as the results have implications for the VPD pilot project and the findings will have immediate use to determining the future of the initiative and implementation (True, 1989). Quantitative crime analysis data and methods encompass the use of numerical or categorical data and statistical methods, which include frequencies and percentages (Santos, 2016, p. 5). The quantitative methods employed include establishing baseline measurements of RBNE trends using frequencies, calculating means, averages and percentage changes, as well as pattern and geospatial changes. Based on the data collected and analyzed, the results answer the question as to whether the machine learning, spatial-temporal crime forecasting model met the intended outcome of a reduction in residential burglaries through directed police action, and whether it was an effective tool for forecasting residential burglaries.

The following sections of the chapter outline the type of data collected, the processes used to obtain the data, and the methods used to collate and categorize the data for detailed analysis. Each data source and its purpose is examined in this chapter, which also provides context as to the origins of the data and its relevance within the larger framework of the VPD's crime forecasting pilot.

**Data Ethics Consideration**

This study used secondary data, specifically de-identified, aggregated crime data that were publicly available. Other data analyzed included the predicted crime locations generated by the crime forecasting model and the geo-coordinate-enabled locations of resource deployments collected by the VPD. This information was only available as aggregate, non-identifiable statistics. The release of the data by the VPD was provided in a format that had
been processed to remove any information that could be used to identify any individuals. This type of information and data use complies with provincial government privacy legislation, specifically the Freedom of Information and Protection of Privacy Act (FOIPPA, 1996). Canadian law enforcement agencies are prohibited from disclosing or releasing any crime data that could be used to identify a victim of crime or violate a person's privacy rights (FOIPPA, 1996, s. 15). As a result, data cannot be released by the VPD until all identifying details have been removed prior to disclosure. Therefore, the crime data used met the same criteria and standards that apply to all publicly available crime data. Nonetheless, the study obtained a letter of release, certified by the Chief of the VPD that the data released for the study had been cleared for public use (see Appendices A and B for copies of the VPD permission letters). In addition, this study was submitted and approved by the Charles Sturt University Human Research Ethics Committee, with protocol number H17017 assigned for annual progress reporting and monitoring (see Appendix C for a copy of the approval letter).

**Research Context**

The VPD crime forecasting system was developed independently of this evaluation process and study. However, to situate the data obtained from the system, a brief introduction of the study area and a contextualization of the data is presented below.

The City of Vancouver (COV) is 114 km² with a population of 631,486; COV is the largest city in British Columbia, Canada, and the eighth largest municipality in Canada. Greater Vancouver metropolitan area includes neighbouring cities and is the third largest in Canada (COV, n.d.-a). Vancouver has 23 distinct neighbourhoods referred to as areas (COV, n.d.-a). The streets generally form a grid, with named streets running north and south and numbered avenues running east and west (COV, n.d.-a). For the purposes of the pilot, the West End, Downtown, and Strathcona areas were excluded and the remaining 20 neighbourhoods of “West Point Grey, Kitsilano, Fairview, Mount Pleasant, Grandview-Woodland, Hastings-Sunrise, Dunbar-Southlands, Arbutus Ridge, Shaughnessy, South Cambie, Riley Park, Kensington-Cedar Cottage, Renfrew-Collingwood, Musqueam, Kerrisdale, Oakridge, Marpole, Sunset, Victoria-Fraserview, and Killarney were included. Downtown is the business district of Vancouver” (COV, n.d.-a) and 93.9% of the dwellings are apartments with five or more storeys; there are no single detached, semi-detached or detached duplex dwellings in this neighbourhood (COV, n.d.-a). The West End is also high density, in that
77.8% of dwellings are apartments with five or more storeys, while apartments under five storeys represent 21.4% of the remaining dwellings (COV, n.d.-a). In Strathcona, 81.2% of all dwellings are apartments (COV, n.d.-a). Given that the above-noted neighbourhoods comprise mainly high-density, multi-storey apartments, they did not allow for the type of enforcement and prevention resource allocation planned; therefore, these areas were excluded from the pilot study.

The name Vancouver is often used to refer to Metro Vancouver, or Greater Vancouver, which includes the COV as well as 20 other municipalities, one electoral area, and one Treaty First Nation (Metro Vancouver, n.d.). The cities within the regional district relevant to this study that were used for comparison purposes included Burnaby, Richmond, and North Vancouver, all of which directly border the COV and City of Surrey, which is only slightly smaller in population than Vancouver and faces similar urban issues (Owen, 2009). The population breakdown of the surrounding municipalities includes Surrey at 518,007, Burnaby at 232,755, and Richmond at 198,309 (Metro Vancouver, n.d.). Aggregate crime incident data were collected for the study area within the COV as well as the surrounding region covering a 5-year period for the purposes of trend analysis and comparison. Specifics of the data collected and its uses are discussed in detail in the following sections of this chapter.

Justification of Methods
Based on the aggregate crime data available and the output of the forecasting model used for the pilot project, a quantitative methodology and analysis was the logical choice for this study. An establishing of trends in time and location was required for the purposes of comparing the pilot results with what would be considered to be the norm, namely had a forecasting model not been implemented and subsequent directed police action been undertaken. Given that the purpose of this study was evaluative (i.e., intended to assess the efficiency and effectiveness of the predictive model and its application in an operational police setting), it was important that the research resonated within the framework of the policymakers, senior managers, finance departments, and other levels of governance. This was important for enabling the research findings to be more easily accepted and, more importantly, utilized. At a fundamental level, this study, which examined crime statistics and defined effectiveness based on a reduction of the number of RBNEs, needed to be
quantitative since both the data and the data analysis were purely quantitative in nature. Specifically, the study was premised on measuring variances in crime trends, geo-temporal changes in crime patterns, and validity tests that measured statistically significant changes in the distribution of residential burglaries correlated with police resource allocations, anticipated geo-temporal locations, and actual crime incidents. Further detailed information and justification of the data collection, validation, processing, and analysis will follow.

**Data Collection**

Residential break and enter data, including the period covering the VPD pilot study was used in this study. This section includes eight subtopics: (a) aggregate RBNE crime data; (b) VPD forecast model description and outputs, Forecast Model 1 – geographic constrained prediction area; (c) Forecast Model 2 – moderate scale prediction area; (d) resource deployment strategy; (e) patrol plain clothes teams – targeted enforcement; (f) community safety program team – high-visibility mobile deterrence prototype; (g) data collection – resource deployment data; and (h) data validation, data processing. Each is discussed in the subsections below.

**Aggregate Residential Break and Enter Crime Data**

Aggregate crime incident data were collected to establish a baseline, using 4 years of crime data between 2012 to 2015 from within the pilot project boundaries and for the 6-month period from April 2016 to September 2016 during which the study was conducted. The RBNE crime data were extracted by the VPD from the police records information management environment (PRIME-BC; Royal Canadian Mounted Police, 2015) records management database, using the following criteria: each month (April–September) per year (2012–2016), between the occurred times of 0800 and 1600 hours, and within the predictive study area. In addition, RBNE data were extracted for the times of 1601 and 0759 hours, within the pilot study boundaries as detailed in the “Research Context” section of this chapter.

The RBNE data from the remaining areas within the COV jurisdiction but outside the pilot project were also collected, as a control with which to evaluate prevailing crime trends for residential burglaries. The extraction also covered the same time intervals, from the periods of 0800 to 1600 hours and 1601 to 0759 hours. Further, regional crime statistics
from surrounding municipal jurisdictions were collected for use as a comparison with those of the COV and the study area. Baseline historical property crime data for 4 years prior to the study were collected for the purposes of comparison with the residential burglary statistics generated during the pilot project.

Crime statistics that had been cleansed of identifying data and represented anonymized statistics were obtained for the Metro Vancouver policing jurisdictions from the provincial records management system and were verified against the published Lower Mainland District Crime Statistics Dashboard (Royal Canadian Mounted Police, 2016). This facilitated an extrapolation of general regional trends when compared to the data from the forecasting pilot study. It is considered best practice to use between 3 to 5 years of data for trend analysis, whether this is for annual, quarterly, or monthly comparisons, with 3 years representing the minimum baseline that a statistician would typically use (Bergeron & Rivard, 2017; Santos, 2016, p. 419). The use of a larger timescale allows for better controls for outlier activity and mitigates any unique influences that may have occurred in isolation in any one year (Santos, 2016, p. 419).

Two outputs were generated by the crime forecasting system. Details on how the information and outputs were generated will follow.

**VPD Forecast Model Description and Outputs**

The crime forecasting system, a focal point of this research, was developed by a consortium comprised of an academic, public, and private partnership. The current research study was independent of any validation and testing conducted by the consortium and the developers of the model. Rather, the current study aimed to assess the value of using a forecasting system in a real-world policing environment, using innovative evaluative practices and methodologies that had not been applied to evaluations of predictive policing deployments in the past. The intention was to provide a template for the evaluation of future forecasting studies and to determine whether this particular case-study met its intended objectives.

To begin, the forecasting system generated two outputs that were predetermined by the software developers and were in keeping with the intended use of the model by the police service. The reasons and rationale for these decisions on output and forecast sizes
were outside of the scope of the study but are noted here to help contextualize the structure of the data and the type of data being evaluated.

The predictive model generated a total of six forecasts that comprised a single set for each 2-hour interval. The single set of six forecasts included three 100-m and three 500-m areas (see Figure 1 & 2).

Figure 1 Example of 100-m Forecast Area

Forecasts were generated each morning by the system at 0600 hours, covering the period from 0800 to 0759 hours the following day. Each of the 100-m and 500-m prediction areas was active for a 2-hour period throughout the day, from which new forecasts would come into effect covering different geospatial areas for another 2-hour period. These forecasts were added to a GIS user interface that was accessible on police mobile terminals in the vehicles designated for the pilot study.

As noted, the development of the forecasting model was conducted by a consortium of academics, private, and public sector participants who collaborated to build a machine-learning system. While the actual development of the model is outside of the scope of the current study, it is important to note that the system was designed to highlight and render the top 10% of a potential total possible number of forecasts. The model was built to generate a probability for RBNE for each 100-meter grid for the total area of the City of Vancouver at two-hour intervals. In essence, each 100-meter grid comprising the area of the City of Vancouver has a probability ranking for a RBNE incident occurring, with the top 10% of the forecasted locations having an almost indistinguishable chance in occurring. During the model development and testing it was determined that the top 10% forecasts equated to approximately six actual forecast locations, with a 50% mixture of 100-meter and 500-
meter identified zones. Therefore, the model was designed to generate three separate 100-m forecasted locations and three separate 500-m forecasted locations at each 2-hour interval, thereby completing a set of six forecasts. The top 10% of the possible forecasts were determined by the development consortium to have the highest confidence for capturing a RBNE occurrence.

The intended application of the model was to inclusively identify the top six locations every 2 hours that had the potential for residential burglary to occur. Therefore, the model’s success was premised on any one of the six locations encountering an incident. This is a slight departure from the traditional line of understanding of predictive systems, in which one might expect the highest-ranked forecast to have the best chance of capturing an incident. This was not the case with the VPD’s forecasting model, and this departure from the traditional deployment of predictive systems confounded the evaluation of the model, given that traditional successes or correlation measures could not be applied.

It is also important to distinguish the intended application of the forecasting system in an operational environment. The model required police resources to be deployed to all six locations comprising a set of forecasts, for each 2-hour interval, in order for the model to function properly in an operational setting. This is an important distinction, as it changes how the system is utilized and subsequently evaluated. The likelihood that a residential burglary will occur at any one of the six predictions within the set is barely discernible, which puts an increased onus on the police service to maintain consistent and directed enforcement at multiple locations for a sustained period. To summarize, for the model to be effective in an operational police setting, a total of six forecasts (3 x 100 m and 3 x 500 m) are generated for each 2-hour interval, with the intended application of the model that police resources will be assigned to all six locations to deter or interdict crime.

Figure 2 Example of 500-m Forecast Area

![Example of 500-m Forecast Area](image)
**Forecast Model 1 – Geographic Constrained Prediction Area**

Within the set of six predictions, a forecast output termed Forecast Model 1 – a geographic constrained prediction area – provided three spatial temporal areas at risk of RBNEs at 2-hourly intervals from 0800 to 0759 hours the following day. However, due to limited resources, police deployments only targeted a specific timeframe within that forecast range, with police resources being deployed from 0800–1600 hours each day, 7 days a week. Forecast Model 1 generated three 100-m across buffers. For the four 2-hour intervals in which police resources were deployable to 100-m grid cells throughout the day, a total of 12 predicted locations had resources allocated daily to these geographically constrained prediction areas per day.

**Forecast Model 2 – Moderate Scale Prediction Area**

The second forecast output referred to as Forecast Model 2 – moderate scale prediction area – similarly provided spatial temporal areas at risk of RBNEs at 2-hourly intervals for the same timeframe as Forecast Model 1. However, this model differed, as the spatial detail was lower at 500 m x 500 m across, in the shape of a regular quadrilateral polygon. This prediction area covered a significantly larger area, and the VPD opted to use a modified deployment approach. The deployment approaches used by the VPD are discussed in detail below.

**Resource Deployment Strategy**

There were two resource deployment strategies employed by the VPD, with each adapted to outputs from the two forecast models:

- Forecast Model #1 (100m across prediction zones)
  *Patrol Plain Clothes Teams – Targeted Enforcement Prototype*

- Forecast Model #2 (500m across prediction zones)
  *Community Safety Program Team – High-Visibility Deterrence Prototype*

As noted, Forecast Model 1 provided a higher level of spatial detail (10,000m²) which, for the majority of neighbourhoods within the pilot study area, covered approximately half a block. The deployment of Patrol Plain Clothes Teams covered the 0800 to 1600 hour period at predicted locations at 2-hourly intervals. Forecast Model 2 covered a larger geographic
area, whereby the deployment approach made use of community safety personnel (CSP) employed in teams that mirrored the same times as the police officers.

**Patrol Plain Clothes Teams – Targeted Enforcement at Model 1**

Plain clothes police patrol teams that formed the targeted enforcement were deployed in unmarked patrol cars at 100m Model 1 predicted locations. These teams, while deployed in unmarked patrol cars and in a plain-clothes capacity, had a mandate to actively engage potential persons of interest, such as known offenders and any suspicious activity that occurred within their assigned predicted location. The VPD’s rationale and goal for a lower visibility presence was not to displace and deter crime within their area of responsibility, but rather to actively engage with and interdict offenders from committing or planning to commit a property offence. Officers in plain clothes had the ability to self-deploy to laneways and observe potential offenders in an inconspicuous manner. They were also able to conduct street checks (stop and check) of known property offenders and chronic offenders who entered the area, as well as to monitor the area for suspicious activity. Specifically, these teams were deployed in an observe-and-approach strategy, in which they were expected to engage actively in enforcement action when the situation merited it.

At the end of each 2-hour interval, targeted enforcement teams relocated to the next set of forecast locations for the subsequent interval. The teams used a proactive approach that was deemed appropriate for the location, in which the officers conducted street checks, foot patrol, and exercised preventative or pre-emptive strategies for known offenders, such as individuals in breach of bail conditions, identified in the area.

In the interest of involving as many officers possible in the pilot project, the teams were regularly rotated, with each area of the patrol operations providing one officer per 4-day deployment cycle. This served two important functions. First, it engaged a large number of officers in the pilot project for the 6-month period. This wide exposure helped to educate officers about the technology and provided a unique perspective as to how it might assist them in their duties after the pilot had ended and the system had been fully deployed. Equally importantly, it may have helped to hedge against the potential for the deployed officers to stagnate as the pilot project progressed through the spring and summer. Previous predictive studies have indicated that this is a very real concern and in several well-
documented case studies, officer apathy and lack of productivity negatively influenced the outcome and results (Hunt et al., 2014; Perry et al., 2013). While an evaluation of officer participation and their perceptions is beyond the scope of this study, it is interesting to note that the VPD took this perspective into consideration when drafting the implementation strategy.

Figure 3 illustrates the typical size of a Forecast Model #1 Geographical Constrained Prediction Area and a possible police deployment approach within this one-block radius.

**Figure 3 Illustration of Typical 100-m Patrol Area**

*Community Safety Program Team – High-Visibility Deterrence at Model 2*

Forecasting Model 2 used a modified strategy, given the larger geographic area covered by each forecast and the strategy being deterrence and diffusion as opposed to interdiction and enforcement. A forecasted area spanning 500m translated into approximately eight to 10 city blocks. The larger area excluded the potential for direct observation and foot patrols. Instead, an area covering up to 10 blocks necessitated the use of a vehicle in order to provide sufficient coverage to the area for a full 2-hour interval. While continual patrolling
of a 10-block area will provide a noticeable police presence and deterrence effect, it does have limitations in terms of actively monitoring the movement of potential offenders and also reduces the likelihood of observing an offence in progress.

The CSP members were deployed on shifts that matched the Patrol Plain-Clothes Teams – Targeted Enforcement Prototype police officers from 0800 to 1600 hours. The CSP teams formed the High-Visibility Mobile Patrol Deterrence Prototype (CSP Resources) with the aim of projecting a high-visibility presence within the forecast area through persistent patrols. These teams actively looked for suspicious activity during their patrols and engaged in proactive measures, such as monitoring laneways and identifying potential known offenders entering the area. Where appropriate, CSP High-Visibility Mobile Patrol Deterrence teams called on regular patrol officers for assistance or to engage in enforcement activity while operating within their authority and mandate.

CSP members are considered to be distinct from regular police officers, with limited powers and reduced training. While regarded as peace officers, their status is restricted, and they are authorized to perform certain duties only (VPD, 2015; Wiebe, 2012). For example, CSPs are not permitted to conduct street checks of known offenders or carry out checks of suspicious activity, as their personal safety could put at higher risk. Even so, their functions are multifaceted, ranging from supporting regular police officers with lower risk duties to providing a high-visibility presence within the community, where regular officers are not required (VPD, 2015, pp. 12–15). The formation of the CSP program was intended to create greater efficiencies within the police service by maximizing the availability of police officers to engage in proactive policing activities (Wiebe, 2012). Figure 4 provides an illustrated example of a typical Forecast Model 2 area assigned to a CSP team to patrol.
The difference between the application of the outputs, as well as the factors that determined how these outputs would interconnect with resource deployment strategies, was beyond the scope of this quantitative evaluation research, which focused on the outcomes of the pilot study as it was operationalized. However, background information was provided within the literature review to help contextualize the relevance of the outputs as influencers on deployment tactics used by the police and to indicate how operational necessity can transform the use of technology in a real-world setting.

**Data Collection – Resource Deployment Data**

The aggregate resource utilization data contained information regarding locations and times when police teams were deployed, the length of time resources attended the forecast locations, and the types of activities that were engaged in at these locations. Noteworthy aspects of the data were the x/y geo-coordinates where police teams were deployed, and the length of time the resources were active in the predictive locations. Collection of the data enabled a detailed analysis of the effectiveness of the deployment strategies applied to the treatment area, and comparisons with areas where no resources were deployed. As noted in the ethics section, the police resource deployment data was aggregated.
information and contained nothing that would enable identification of the individual resources deployed.

**Data Validation**

Prior to the release of police-generated crime incident data, referred to as general occurrence reports (GO), the VPD conducted a quality control and validity check to verify the accuracy of the police data and to validate the locations and times that residential burglaries were reported to have occurred. This data audit was conducted by the VPD, since it had been determined that for any detailed analysis to retain validity, the data used must undergo a comprehensive review. To accomplish this task, every RBNE GO that occurred within a predicted location (100m and 500m grid) covering the 6 months of the pilot project, were reviewed and the synopsis or narrative pages were read by VPD staff to determine whether the incident had occurred at a date and time that could be reliably determined, and whether it corresponded to the date and time recorded in the relevant GO fields. The VPD then expunged any sensitive information from the data that did not meet its strict criteria for the public release and dissemination of police information under the provincial FOIPPA (1996) legislation. The released data contained only generic fields containing x/y coordinates, date, time, and incident categorization. These data are publicly available on the VPD Open Data Catalogue website (COV, n.d.-b), with the only difference being the added quality control measures that were applied to the data prior to release.

A total of 418 RBNEs occurred during the predictive study period of April 1, 2016, through to September 30, 2016, from 0800 to 1600 hours. The 418 incidents were manually reviewed by the VPD quality control team to confirm the accuracy of occurred date and time and incident classification. The quality control team identified RBNE incidents in which the occurred date and/or time were incorrect, not known, or were more than five hours over the estimated occurred time. In addition, some RBNE incidents were incorrectly coded, as the incident had occurred at a non-residential structure. These RBNEs related to sheds, apartment parkades or storage areas, and detached garages. The total number of RBNE incidents that met the exclusion criteria was 220. Therefore, with the removal of 220 RBNE incidents from the total of 418 RBNE incidents, 198 RBNE incidents were left for analysis, with only 47.3% of the raw RBNE data meeting the data quality standards.
After initial analysis using only 198 RBNE incidents, it was determined that the number of RBNEs was not adequate for the application of commonly used analytic methods. For example, the total number of incidents remaining (198 RBNEs) did not meet the optimal data sample size required to conduct the specific trend and effectiveness analysis that had previously been carried out by Silverman (1986) on geospatial trend data. However, this did not adversely influence the study, as alternative methods were applied that made use of recently developed geo-temporal data analysis technology and techniques. Details on the analytic approaches adopted, and those not feasible because of data limitations, are discussed in detail in the following sections.

**Data Processing**

For the purposes of this study the geospatial and temporal data provided by the VPD were formatted and parsed into a larger geospatial relational database. The database selected for this process was an Esri (n.d.) geo-database. The rationale for this decision was that the Esri platform provides a central storage and retrieval repository that facilitates advanced analysis and the use of SQL queries and data mining techniques. Furthermore, the Esri geo-database supports multiple file formats, such as shapefiles, while offering adequate performance through optimized storage capacity. Although other GIS software that could have been used for processing the data is available, because of the volume and type of data collected, as well as the necessity for the information to be searchable and able to be queried, the Esri geo-database was the only feasible option with which to handle the geospatial and temporal data in an expeditious manner. Conversion and compatibility of data were also of concern, in addition to the extended processing time and the potential quality of the output to be affected.

Once ingested into the Esri geo-database, the data were edited and structured to ensure consistency in format and to provide an aggregate of 5 years of geo-spatial and temporal data for the COV and the surrounding region. This was necessary for the facilitating of a comprehensive analysis of crime trends and patterns, and to apply advanced geospatial statistical tools to the data. The process of ingesting the data was at times a complex manual approach, such as the requirement for street names and addresses to be converted into latitude (y-coordinate) and longitude (x-coordinate) to enable further geospatial analysis. In addition, the database schema was customized so that key
information could be managed and represented in a table format, thereby making it possible to use advanced querying and processing of large datasets. Once the data were cleansed using standardized formatting and geoprocessing, a consolidated representation of RBNE crime incident data, along with the police resource allocation, geospatial temporal data, and crime forecasts in the form of vector polygons, were run through a geo-database enabling analysis.

In summary, the data collection and processing involved establishing baseline measurements of trends and patterns of historical RBNEs. This process also involved obtaining the geospatial and geo-temporal forecast data associated with the pilot study area and period as well as gathering the same data on the deployment as a result of the forecasts.

**Data Analysis**

Analysis of the data included the use of both descriptive and inferential statistics for comparing the 4-year baseline data to that of the pilot project when the crime forecasting model was deployed by the VPD. Kernel density estimation (KDE), Global Moran’s I, and GIS analysis were conducted on the data collected, as described in the previous section. Tightly related to these analysis techniques are the use of spatial regression for quantifying spatial patterns, and examining the spatial relationships behind the observed patterns. The modelling of spatial relationships allows researchers to determine which factors are potentially influencing a particular variable, and to provide greater understanding of a phenomenon (Fotheringham, Brunsdon, & Charlton, 2015). Researchers can use this approach to test a hypothesis and to help in answering questions regarding the data being examined. Ordinary least squares (OLS) regression, which is the best known of all spatial regression techniques, was not applied in this study for a number of reasons (Mitchell, 2005). The dominant consideration rests with the difficulty of defining a global model of the variables under examination. The OLS regression requires the fitting of a regression equation to every feature in the dataset (Mitchell, 2005). However, in the case of this study, there is no direct linear relationship between forecast locations and crime incidents. Forecast locations are considered to be geospatially dispersed throughout the study area at each time interval, with each of the six forecasts representing an equal probability of a positive or negative relationship to crime incidents. However, the remaining five forecasts that do not show a relationship still produce a successful forecasting model, despite the lack
of a linear relationship. This confounds the use of OLS, as well as the more geospatial centric regression model that is being used within the GIS industry and is known as geographical weighted regression or GWR (Cliff & Ord, 1973; Scott & Pratt, 2009). Both GWR and OLS require the application of a spatial regression equation, either to each individual feature or to the global model (Scott & Pratt, 2009).

Regardless of the specific technique applied, standard regression modelling does not fit the study data sample, given the potential for five of the six forecasts failing to indicate a linear relationship, even when the forecasting model is operating at peak efficiency and meeting the intended operational benchmarks. For example, OLS and GWR are not appropriate methods for determining the effectiveness of the forecasting model because the forecasts are generated at 2-hour time windows. With OLS, there is a requirement to aggregate the data points, which essentially nullifies the specified time component (Davis, 2002). Additionally, OLS and GWR are not effective when the range of values is small, as is the case when the actual crime counts for each fishnet are either 0 or 1 (Davis, 2002; Leslie & Kronenfeld, 2011). For the 6-month pilot study there were insufficient values produced for the total spatial boundary to allow OLS to be calculated accurately. The use of OLS and GWR is further confounded by the definition of success, with five out of six forecasts having an expected null incident return, where OLS and GWR analysis would focus on aggregate data but when the actual unit of examination is object-based. This introduces spatial heterogeneity issues, whereby the explanatory measures for the entire model that are used for testing the hypothesis, do not adequately describe the outcomes at a specific location (Kendall & Gibbons, 1990; Scott & Pratt, 2009). To address this situation, statistically significant time changes across crime incidents and the geo-temporal accuracy of the forecasted locations were examined, which mitigated the spatial heterogeneity issues. The next section summarizes how and why these analytical methods were used.

**Kernel Density Estimation**

A comparison of spatial-temporal changes between the years was conducted using KDE, “a non-parametric method that estimates a smoothed probability surface, mapping the spatial intensity of crime over space” (Nie, Wang, Du, Ren, & Tian, 2015; Silverman, 1986, p. 76). This method uses a mathematical function known as a kernel, through which it averages the location of a point with respect to nearby points (Nie et al., 2015). Therefore,
the cell size and search radius input parameters are important factors when executing the kernel density tool from ArcGIS (Esri, 2016). For this study, the forecast model generated 100m and 500m predictive boxes. This necessitated an approach that matched these parameters while offering the ability to obtain the average of the location points with respect to the location of other points that were within the 500m search radius. In addition, KDE was used for this study because, as a method, it is considered the most accurate approach for analyzing and visualizing changes in geo-temporal crime data (Chainey, Tompson, & Uhlig, 2008; Nie et al., 2015).

A Gaussian statistical function was applied to control for diffusion and outlier activity. Spatial autocorrelation testing was used to determine the likelihood that the observed results were random or by chance, and not related to the forecasting and subsequent deployment of resources. It was important for the study to determine whether or not the crime forecasting model and the resulting resource strategies employed by the VPD had simply moved crime around the study area, or whether previous concentrations of residential burglaries had effectively been reduced and to identify potential displacement due to the pilot project.

The *Kernel Density* analysis was applied to determine the density of the RBNE crime points for each month (April–September) per year (2012–2016), as well as to visualize the spatial distribution of the RBNE crime events per square meter. The mathematical function that Esri (2016) ArcGIS uses in the kernel density tool is based on Silverman’s research in this field (Law & Collins, 2017; Mitchell, 2005).

**Spatial Autocorrelation Test – Global Moran’s I**

The GIS x/y coordinates representing the resources deployed in response to the forecasting were analyzed using the Esri (2016) ArcGIS geospatial analytical software. The Spatial Autocorrelation Test (Global Moran’s I) (Nie et al., 2015) was run to evaluate the confidence level that the crime reduction and resulting crime patterns were probably not a result of random chance. The Moran’s test measures spatial autocorrelation in the similarity of location and value features (Griffith, 2011; Silverman, 1986). Specifically, “given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered,
dispersed, or random” (Griffith, 2011, p. 12). Values were calculated for the Moran’s Index with both a Z-score and p-value evaluating the significance of that index.

**GIS Spatial Analysis**

The spatial analysis was performed to assess the crime density, intensity, and distribution of RBNE incidents. This required the use of Esri (2016) ArcGIS 10.4.1, with the Spatial Analyst and Geostatistical Analyst extensions, and Esri ArcGIS Pro 2.3 which provided the capacity to conduct advanced three-dimensional space-time cubes by aggregating vector-based data points. This analysis software allowed for the application of *Emerging Hot Spot Analysis* and the calculation of the Getis-Ord Gi that identified statistically significant spatial clusters of high and low values with an associated Z-score, p-value and confidence level (Cliff & Ord, 1973; Ord & Getis, 1995). Further, a Mann-Kendall test can be applied, which is a rank correlation analysis for the bin count that examines the values and counts the instances that fall into each interval and their time sequence (Getis & Ord, 1992). This approach can then examine “variance for the values in a bin time series, the number of ties, the number of periods, and the observed results” (Law & Collins, 2017, p. 34) versus the expected values, to determine if the difference is statistically significant (Esri, 2016).

There was a limitation on what statistical analysis could be applied, based on the number of values generated during the 6-month pilot study within the spatial boundary of the study area. More specifically, the limited number of incidents did not allow for the OLS to be calculated accurately. An innovative approach was used, whereby the focus of the analysis shifted towards examining statistically significant changes across crime incidents and the geo-temporal accuracy of the forecasted locations. To start, a Fishnet was created in a grid pattern that consisted of 100m x 100m grid cells; this produced the reference grid that was required for computing kernel density with the number of incidents available. The 100m x 100m grid cell size was chosen as it was the finest level of detail of the predictive box size used in the study. Creating a fishnet permits the detection of each 100m x 100m grid cell where the intensity and concentration of break and enter residential crimes are located, as well as providing an estimate of the distribution of clustered break and enter residential crimes that were displaced (Mitchell, 2005). The fishnet created for the predictive study consisted of 10,632 grid cells, each of 100m x 100m, and is illustrated in Figure 5.
The Kernel Density analysis was applied to determine the density of the RBNE crime points for each month (April–September) per year (2012–2016), as well as to visualize the spatial distribution of the RBNE crime events per square meter, as previously mentioned. The mathematical function that Esri (2016) ArcGIS uses in the Kernel Density tool is based on Silverman’s research in this field (Law & Collins, 2017; Mitchell, 2005). The standard classification method used for the resulting kernel density for each month from 2012–2016 was Natural Breaks (Jenks) with four classes (Law & Collins, 2017; Nie et al., 2015). The Natural Breaks (Jenks) classification was chosen since it is considered to be the model classification strategy for determining class breaks within the four classes to maximize the between-class differences and minimize the within-class differences (Scott & Warmerdam, 2005).

The only issue in using this classification approach is that it generates unique classifications for each kernel density dataset, which would cause difficulty when comparing disparate kernel density maps (Harries, 1999). Therefore, to solve this issue and to simplify the interpretation of the numeric density values, the numeric density values in each kernel density map were reclassified to four categories: 0 represents “No Data/Not Significant”; 1
represents “Low Intensity”; 2 represents “Medium Intensity”; and 3 represents “High Intensity”. This process was undertaken by applying a reclassify technique in Esri (2016) ArcGIS that mirrored the same approach used by Scott and Pratt (2009) when attempting to apply regression modelling techniques to spatial-temporal data with difficult-to-define qualifying spatial patterns. This allowed for the assignment of a rescaled value to a common scale, namely the four categories of density values.

Using the fishnet as a reference grid, along with identical categories for each kernel density map, the aggregate of the mean category values was calculated for each month and for each year between 2012–2015. This was then used to determine the common hot spots of the RBNE crime events that occurred in the same month of each year from 2012–2015. The results of the aggregated mean category values were used to compare with the same month in 2016 to identify any changes and the resulting impact on these hot spots, compared to the previous 4 years. This provided an empirical value that was used for detailed analysis and comparison. For visualization of the changes and impacts, change detection maps were generated, where a raster calculator tool was used to calculate the change difference between the month in the current year and the average 4-year period of 2012–2015. Furthermore, the number of grid cells for each intensity level within a cold or hot spot (n) was divided by the total number of grid cells, where all cold/hot spots occupied (N), expressed as a percentage. This was calculated by \((n/N) \times 100\). The change difference was represented on illustrative maps for each month in 2016 compared to the monthly average covering each 2012–2015 period.

The data were then formatted into a relational space-time cube to determine whether emerging hot spots or the expected pattern of RBNE was being altered in a measurable way. As with the evolution of crime forecasting and machine-learning, both of which were made possible with the introduction of the newest iteration of high-speed processors, more advanced programming capacity and an ability to consume vast amounts of raw data, research capabilities and analytic tools have also been similarly impacted. Previously restricted to less technically advanced methods for examining geo-temporal data, the newest evolution of GIS software provides the ability to create and visualize space-time data cubes by providing a more robust ability to apply space-time statistical tools using three-dimensional (3D) modelling techniques (Law & Collins, 2017). This new capacity is
premised on generating a three-dimensional data cube made up of space-time bins, where the x- and y-axes represent the geospatial location of the crime incident objects and the t-dimension represents time. The process of aggregating objects into space-time bins also allowed for better handling of outlier activity and establishing time-step intervals that correspond to the data range being examined (Mitchell, 2005). Figure 6 provides a visual representation of a space-time data cube and a cutaway location-based, time-slice bin.

**Figure 6 3D Space-Time Data Cube**

For this study, a space-time data cube was generated by aggregating the RBNE crime incident points that included geo-coordinate details and temporal reference of date and time for the purposes of identifying time-dependent hot and cold spots. Geo-temporal bins were created using a time-step interval of 2 weeks, within the range of April 1 to September 30 inclusive, capturing the number of crime incidents that occurred at each 100m grid location. Therefore, a bin, represented 2-week step intervals at a specific 100m grid geo-location ranging from April 1 to September 30 inclusive. Anderson (2009) noted in his study of accident hotspots that the spatial proximity of features within the same temporal interval were indistinguishable when the interval failed to collect the minimum number of statistically significant incidents required for the test to be applied. For this study, a one-week interval did not obtain the minimum number of points for the space-time cube to generate a statistically significant time-step required for calculating clusters. Similarly, by
increasing the time-step interval to four weeks, the values supported the required minimums, however, the geo-temporal bins were segmented over each month for 183 days, resulting in an interval that failed to capture potential changes that could have occurred due to the aggregation of the data points. In the end, a step-interval of two weeks was used as it captured the minimum number of points to support a statistically significant analysis and the interval was granular enough to still capture any potential intensity of clustering of high or low counts. The Getis-Ord Gi statistic was then applied to each bin to examine whether the features were high or low value, and clustered spatially in relation to their proximity to other features (Law & Collins, 2017). For the study data, the Getis-Ord Gi returned either a positive or negative Z-score. The more intense clustering represents high values and a statistically significant negative Z-score indicates clustering of low values or cold spots (Ord & Getis, 1995). This then allowed for emerging hot spot analysis that identified statistically significant spatial and temporal relationships using the Mann-Kendall time series trend test.

The Mann-Kendall time series trend test examined neighbourhood distance and relationships as calculated by the Getis-Ord statistic as well as neighbourhood relationships that spanned temporal steps, while factoring in each temporal bin slice neighbourhood cluster (Anderson, 2009; Law & Collins, 2017). The process of generating a space-time data cube involved calculating the Mann-Kendall statistic for each bin x/y location according to the parameters input for bin spatial area (Law & Collins, 2017). For the current study, as discussed in the data collection section, the location parameter was calculated using 100m, which corresponded to the unit of measure rendered by the forecast polygons, and each bin temporal interval was 2-weeks. Building on the Getis-Ord Gi statistic output, each bin’s Z-score and p-value was evaluated using the Mann-Kendall trend test. The Mann-Kendall test generated a trend Z-score and p-value for each location and with the previously-generated Getis-Ord Gi Z-score and p-value for each bin categorized. This correlational analysis was then rank ordered and compared to the expected results, using a benchmark of ‘no trend’ over the time series to determine whether the results were statistically significant (Anderson, 2009; Kendall & Gibbons, 1990; Scott & Pratt, 2009).
Within the context of this study, the application of the Getis-Ord Gi statistic test on each bin within the space-time data cube was premised on identifying time-dependent hot and cold spots. Determining whether there existed transitory trends that did not reflect an overall pattern, and whether persistent hot spots were repeated through multiple time slices at similarly located bins, was directly relevant to testing the research hypothesis. Transitory hot spots that developed quickly and then disappeared shortly after appearing, could indicate micro-displacement caused by a properly functioning forecasting system. In this scenario, the forecasting system could result in displacement at traditionally active and persistent hot spots. If this was the case, the displacement would then be dispersed and of a lower intensity than what had existed previously. A properly functioning forecasting model should then be able to acquire the new micro-displacement hot spot locations and reallocate police resources to again displace and disperse the activity. Were this the case, the Getis-Ord should only reveal sporadic and inconsistent displacement that was not persistent over time. Alternatively, if the analysis revealed persistent and numerous hot spots, then there would be cause to question whether the model was capturing crime patterns accurately and deploying resources in an effective way.

The emerging hot spot analysis using the Mann-Kendall test is a supplementary method for evaluating time-series trends in the data. Following on the previous premise, very few indications of Persistent, Intensifying and Consecutive hot spots should result if a forecasting model is effective and able to deploy police resources to developing crime patterns shortly after they appear. A highly effective forecasting model might have an expected outcome of “no pattern detected” with occasional indicators of “oscillating” and “sporadic hot spots” still reflective of an effective model, whereby new and emerging crime patterns are quickly identified, and resources dispatched, resulting in a disruptive effect. An abrupt change from patterned and consistent hot spots to no patterns further reinforces that the model was effectively disrupting the crime incidents, when taken in conjunction with a marked reduction in the number of burglary incidents and reductions in the intensity (fewer incidents) of micro hot spots. When taken in totality with the previous evaluative measures noted, the use of emerging hot spot analysis and the Getis-Ord Gi test on a space-time data cube is a comprehensive method for testing the hypothesis.
**Geographic Accuracy of Prediction Boxes with No Resource Allocation**

Police resources were deployed to forecast locations between 0800 and 1600 hours during the predictive study area for the 6-month period. From the hours of 1601 to 0759 hours no police resources were deployed; however, the forecasting system continued to generate predictive boxes. These outputs were then used as a validation test of the model accuracy to determine whether RBNE occurred within the forecast areas when no interdiction measures were in place. Forecast locations for each 2-hour interval were parsed in the geotemporal database and grouped by date. Each 2-hour interval from 1601 to 0759 hours in each day from April 1 to September 30, was assigned six forecast locations that were generated for the respective period. Therefore, 48 forecasts were generated and assigned per day, for each of the 183 days during the non-operational times of the study. Each geotemporal 2-hour interval was spatially joined to the corresponding RBNE that occurred within the respective interval.

A spatial query was then run for each interval to determine which crime incidents fell outside the forecasted locations for the day and 2-hour period in question. The spatial query used a search parameter of inside or within 200m of the predictive locations. The expanded search parameter was used after reviewing the deployment data of the CSP and sworn officers, where it was determined that police resources rarely stayed exactly within their assigned area, on average expanding their deployment radius by an additional 200m. Therefore, when deployed, the police resources’ actual coverage area and interdiction zones were over 200m of the polygon, resulting in an impact on a larger area than the initial forecast location. The use of the search criteria “inside or within the 200m predictive boxes” better reflected the real-life usage of the forecast technology in the field and in the larger area impacted by the police resources. The results were aggregated by month and a geotemporal map generated to illustrate the points that were not captured.

**Methodology Summary**

The specifics of the data collection portion of the methodology has been summarized in the following table (see Table 1).
Table 1  Methodology – Data Collection Processes

<table>
<thead>
<tr>
<th>Stage</th>
<th>Analysis Technique</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>Aggregate crime incident data were collected for the study area and the surrounding region, covering a 5-year period. The intended application of the data was to establish baseline measurements of crime trends and patterns.</td>
<td>Region and Study Area Aggregate RBNE Crime Incident Data 2012-2016 Apr - Sept</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Geospatial-temporal data related to officer deployments and activity were obtained to account for police intervention strategies and influencers on crime.</td>
<td>Study Area Aggregate RBNE Crime Incident Data 2012-2016 Apr-Sept</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Geospatial-temporal crime data were collected for the study area for the 4 years prior and including the 6 months of the study period.</td>
<td>Study Area - 2012-2016 RBNE Apr-Sept</td>
</tr>
<tr>
<td>Stage 4</td>
<td>Crime forecast geospatial-temporal data were collected, which was represented as either 100-meter or 500-meter square zones.</td>
<td>Study Area - 2012-2016 RBNE Apr-Sept</td>
</tr>
</tbody>
</table>

The list of analysis techniques applied are summarized in Table 2, which details the specific data used for each stage of the analysis process.

Table 2  Methodology – Data Analysis Techniques

<table>
<thead>
<tr>
<th>Stage</th>
<th>Analysis Technique</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>Trend Analysis - Summary Statistics</td>
<td>Region and Study Area Aggregate RBNE Crime Incident Data 2012-2016 Apr - Sept</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Inferential Statistical Comparison (Z-Score to Percentile Change)</td>
<td>Study Area Aggregate RBNE Crime Incident Data 2012-2016 Apr-Sept</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Kernel Density Estimation (Change Detection)</td>
<td>Study Area - 2012-2016 RBNE Apr-Sept</td>
</tr>
<tr>
<td>Stage 4</td>
<td>RBNE Hot Spot Percentage Coverage</td>
<td>Study Area - 2012-2016 RBNE Apr-Sept</td>
</tr>
<tr>
<td>Stage 5</td>
<td>RBNE Surface Area Distribution Change</td>
<td>Fishnet reference grid coverage for total of 10,632 m sq grid using 2012-2016 RBNE Apr-Sept</td>
</tr>
<tr>
<td>Stage 6</td>
<td>RBNE Intensity Level Comparison</td>
<td>Study Area - 2012-2016 RBNE Apr-Sept</td>
</tr>
<tr>
<td>Stage 7</td>
<td>Z-Scores Inferential Statistics</td>
<td>Study Area Aggregate RBNE Crime Incident Data 2012-2016 Apr-Sept</td>
</tr>
<tr>
<td>Stage 8</td>
<td>RBNE Geographic Forecasting Accuracy</td>
<td>RBNE 1601-0759 H with No Police Resources Deployed Apr-Sept 2016</td>
</tr>
<tr>
<td>Stage 9</td>
<td>Global Moran's I Spatial Autocorrelation Test</td>
<td>Study Area - 2012-2016 RBNE 0800-1600 H</td>
</tr>
<tr>
<td>Stage 10</td>
<td>Relational Space-Time Cube</td>
<td>Study area - April to Sept 2016</td>
</tr>
<tr>
<td>Stage 11</td>
<td>Getis-Ord Gi with Z-score and p-value</td>
<td>Study area - April to Sept 2016</td>
</tr>
<tr>
<td>Stage 12</td>
<td>Mann-Kendall Time Series Emerging Hot Spot Analysis</td>
<td>Study area - April to Sept 2016 (2 week step-interval)</td>
</tr>
<tr>
<td>Stage 13</td>
<td>Getis-Ord Gi with Z-score and p-value</td>
<td>Study area - April to Sept 2016</td>
</tr>
</tbody>
</table>
Limitations

The assessment of the success of the crime forecasting model was based purely on quantitative data and methods in this evaluation study. The impact, as measured in the reduction in the number and intensity of the RBNEs in the study area, determined the efficacy. This is a limitation, in that there may have been a diffusion of benefits (Santos, 2016, p. 35) based on the operational deployment of police resources on other crime types. In addition, although the study accounts for quantitative considerations including random chance, it does not account for other outcomes of success. This could have incorporated quantitative assessments of success including public perception of reduction of RBNEs in the study area. Perceptions of the police resources deployed about the success of the forecasting model could also have been another option. However, one of the key objectives of this study was to provide a framework for a standardized, replicable evaluation methodology for a predictive technology application informed by empirical data. As such, the quantitative data, methods and analysis used were the most appropriate for this study. This concludes Chapter 3. The results obtained using the methods discussed in this chapter are presented in Chapter 4.
CHAPTER 4 — RESULTS

Chapter 4 provides figures, tables and the statistics from the results of the analysis conducted for this study.

The research study set out to determine whether a machine-learning, spatial-temporal crime-forecasting model, applied in an operational deployment, was effective. To determine effectiveness, summary statistics, inferential statistics, and geospatial and image analyses were conducted on the data collected. In brief, the results for the summary statistics comparison showed a marked reduction in RBNEs for the months of May, June, and July, with inconclusive findings for the remaining 3 months of the study. Regional crime statistics showed a gradual drop in RBNEs for the study area starting in April, with the lowest number of incidents recorded in June. In comparison, the other regional municipalities experienced more variable trends, with most recording a significant increase in June. The results of the statistical tests showed May, June, and July as not falling within a normal distribution and unlikely to have occurred by random chance. However, the months of April, August, and September fell within a normal distribution and with a high probability of occurring.

When compared to the historical 4 years of monthly incident counts for the study period, only the months of April and August were above the average, with the remaining 4-month recording noticeable drops. Broadly speaking, the change detection analysis applied to each month showed a general cooling in the number of incidents and intensity, with a shift from higher intensity clustering to medium and lower intensity with fewer clusters. The geographic accuracy testing of the model showed the model performed best in the months of May thru August, with a total prediction accuracy of 83.6%. The emerging hot spot analysis identified only one diminishing hot spot, with the remaining tests indicating no identified pattern. The space-time cube analysis identified two areas with pronounced clustering that were persistent in several space-time bins for a 4-week period. Those areas aside, the remaining bins indicated “not significant”, except during the time period when the police data were corrupted because of an issue with the records management system (RMS). The following section provides a detailed examination of the study results and findings, with illustrative charts and tables of the various tests administered.
Summary Statistics

The RBNE statistics were examined for the months of April, May, June, July, August and September in 2016, which was the 6-month period during which the VPD pilot study was conducted. The aggregate number of RBNEs for each month was then compared to each corresponding month averaged over the previous 4 years. In the study area, there were decreases noted in RBNEs each month of the 6-month period that the pilot study was in operation, when compared to the same month averaged over 4 years. Table 3 presents the results for the study data. In the first month of April 2016 the data recorded a slight drop in RBNEs of -2% when compared to the 4-year average for April 2012–2015. May, June and July recorded a more significant drop in RBNEs at -21%, -27%, and -26% respectively, when compared to the 4-year average for each representative month. August recorded a minimal decrease of only -1% compared to the 4-year average and September saw a modest decrease of -15%. The percentage change in RBNEs by month compared to the previous 4-year average was to determine whether the percentage change in the number of crimes could be accounted for by chance, or might be expected to fall within the normal range based on historical data.

Table 3  Percentage change in Residential Break and Enters by month compared to the previous 4-year average per month.

<table>
<thead>
<tr>
<th>RESIDENTIAL BREAK &amp; ENTER CHANGE</th>
<th>%Change April 2016 - 4 Yr Av</th>
<th>%Change May 2016 - 4 Yr Av</th>
<th>%Change June 2016 - 4 Yr Av</th>
<th>%Change July 2016 - 4 Yr Av</th>
<th>%Change August 2016 - 4 Yr Av</th>
<th>%Change September 2016 - 4 Yr Av</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2%</td>
<td>-21%</td>
<td>-27%</td>
<td>-26%</td>
<td>-1%</td>
<td>-15%</td>
</tr>
</tbody>
</table>

To contextualize the above statistics, the -2% drop recorded in April is positioned within the normal range seen during the previous 4-years and falls within one standard deviation of the expected results without any predictive policing intervention. The probability of that count occurring is 56% of the time, based on the sample size. For May, June, and July, the drop in RBNEs fell one standard deviation outside of the normal expected range, with the probability of those rates occurring respectively in 12%, 11%, and 27% of the time. August and September both fell within the normal range, with the probability of those
same rates occurring in 99% and 64% of the time. The next section examines regional crime statistics for the purposes of further contextualizing the results obtained from the VPD pilot study.

**Regional Crime Statistics**

The pilot study was conducted in the COV; however, to better understand the trends within Metro Vancouver, RBNE crime statistics were obtained for the period of time during which the VPD pilot study was conducted, as well as for the same months in 2015. In addition, one additional month prior to the commencement of the VPD pilot study – specifically March 2016 and historically March 2015 – was obtained to determine the overall trend within Metro Vancouver. The purpose of examining the data was to compare the predictive study data with jurisdictions that bordered the COV, as well as with the Municipality of Surrey, which was deemed to be similar in its composition of crimes and issues (Owen, 2009). Of note, Surrey’s RBNE trends were similar to those of the VPD study area in 2015, in the number of incidents as well as trends, and with a steady increase in April 2015 that peaked in July 2015. The RBNE data were extracted for March to September 2015 and 2016 inclusive. In relation to the VPD study area, a trend comparison was run for 2015 for the municipalities of North Vancouver, Burnaby, Richmond, and Surrey, covering three separate time intervals. The VPD study area was represented in several ways, including a complete 24-hour period, an 0800–1600 hours period, and 1601–0759 hours period. In addition, the VPD control area and city-wide rates for RBNEs were included for comparison purposes.

As illustrated in Figure 7, the monthly fluctuations from March to September 2015 are, for the most part, mirrored across the other jurisdictions. The most prominent mirroring occurred when compared with the jurisdiction of Surrey, a city only slightly smaller than Vancouver, which encounters many of the same urban issues (Owen, 2009). In general, the crime trend observed in VPD RBNEs for the 2015 period mirrored the other municipalities, with the exceptions as noted below.

The City of Richmond recorded a minor increase from April to June 2015 and then a gradual drop in RBNEs. However, North Vancouver and Surrey matched the changes in the crime pattern month to month in 2015. Specifically, April to July saw a steady increase in RBNEs, with a peak occurring in July before a gradual decline through to September.
Burnaby mirrored most of this trend from June to August, but with variations in April and May and an unexplained increase in September that differed from the regional trend.

**Figure 7 Residential Burglaries by Lower Mainland Region Jurisdictions for 2015**

In the following chart, RBNEs were captured for the months of March to September 2016 inclusive, for North Vancouver, Burnaby, Richmond, and Surrey in relation to the VPD study area, covering three separate time intervals. These were a complete 24-hour period, 0800–1600 hours and 1601–0759 hours. In addition, VPD control area and city-wide rates for RBNEs were included. Figure 5 captures the period in which the predictive policing pilot study was active.

The trend recorded for the VPD for all the noted time intervals (0800-1600 hours, 1601–0759 hours and daily totals) of the study area was a steady decline from April to June and then an upward shift from July to August, with a slight drop in September. In comparison, the municipality of Surrey recorded a crime trend that traversed in the opposite direction to that of the VPD, with a sharp increase from April to June (55% rise) and then a drop in August, with a final spike in September. North Vancouver and Richmond mirrored each municipality’s trend. In April, an increase occurred in the number of residential burglaries. This was followed by a drop in the number of incidents in May, with a
subsequent increase in June. In July, the number of incidents dropped, but increased again in August. Finally, the incidence of RBNEs tailed off with a slight drop in the numbers in September. Burnaby followed a similar trend, with the exception of a drop in June where the other municipalities recorded an increase.

However, the divergent aspect of this comparison shows that the RBNE trend seen for the VPD study area was not replicated in the other municipalities. When the VPD study area recorded a sharp drop from March to April, North Vancouver, Burnaby and Richmond all recorded increases. Similarly, in June, the VPD study area recorded a significant drop from the previous months, whereas Surrey saw a sharp spike in RBNEs. This illustrates that the RBNE crime trend recorded for the VPD during the pilot study was not mirrored by the other municipalities; in fact the opposite trend was recorded. Figure 8 illustrates this incongruity, i.e. the inverse in the trend between VPD and the other jurisdictions. Having examined the results focusing on jurisdictional similarities and differences, the next section examines the temporal distribution of the results.

**Figure 8** Residential Burglaries by Lower Mainland Region Jurisdictions for 2016
Resource Deployment Model – Residential Break and Enters

The distribution of RBNE incidents was determined using monthly aggregate crime data from 2015, which was categorized by day of the week and hour of day. The shift model where VPD resources were deployed to forecast locations covered 0800 to 1600 hours inclusive. Based on a monthly 2015 aggregate temporal evaluation of RBNEs that typically occurred within the study area, it was determined that the majority of these crime incidents took place between 0700 and 1600 hours, with some secondary concentrations from 1700 to 2000 hours. Figure 9 presents a temporal topology chart which offers a visual depiction of a representative RBNE monthly cycle. Based on the RBNE temporal crime patterns, the evaluation can only speak to the 0800 to 1600 hours’ shift, since that was the timeframe in which police resources were specifically allocated; it therefore represents the focus of the research study. Given the temporal analysis conducted by the VPD prior to this research study, and reinforced by the researcher’s follow-up aggregate temporal analysis of RBNEs, the timeframe of 0800 to 1600 hours appears to have captured the greatest concentration of events and was therefore the most efficient use of a single unit deployment throughout the 24-hour period available.

Further evaluation and study would need to be conducted in order to assess alternative deployment times and potential shift scheduling that might have offered an improved coverage to interdict RBNE. However, given that the deployment scheduling was set by the VPD, this option was not available to the researcher. The following section will continue to focus on inferential statistics and the likelihood of the outcomes being a result of chance.

Figure 9 Typical Monthly Residential Break and Enter Temporal Distribution
Inferential Statistics

From 2012 to 2015, for each year prior to the study, and segmented by each month, the number of RBNEs was tabulated and compared with the respective month in 2016. The standard deviation and Z-scores were calculated for the range and were then used to determine whether the 2016 study results for each month were consistent or varied from the expected norm (i.e., fitted within the expected normal distribution). The Z-score was then calculated to determine percentile that the RBNE crimes rates for 2016 that fell within or outside of the normal distribution and whether it occurred by random chance. The findings are listed in Table 4.

Table 4  Comparison of Statistical Measures and Tests for RBNE Counts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr</td>
<td>79</td>
<td>71</td>
<td>74</td>
<td>57</td>
<td>79</td>
<td>72</td>
<td>9.05538514</td>
<td>0.773020683</td>
<td>Yes - April 2016 count within normal distribution with a 56% chance of occurring.</td>
</tr>
<tr>
<td>May</td>
<td>85</td>
<td>89</td>
<td>76</td>
<td>88</td>
<td>65</td>
<td>80.6</td>
<td>10.1143462</td>
<td>-1.542363651</td>
<td>No - May 2016 count not within normal distribution with a 12% of occurring by chance.</td>
</tr>
<tr>
<td>Jun</td>
<td>87</td>
<td>75</td>
<td>77</td>
<td>85</td>
<td>58</td>
<td>76.4</td>
<td>11.4804181</td>
<td>-1.602729084</td>
<td>No - June 2016 count not within normal distribution with a 11% of occurring by chance.</td>
</tr>
<tr>
<td>Jul</td>
<td>94</td>
<td>110</td>
<td>60</td>
<td>96</td>
<td>58</td>
<td>83.6</td>
<td>23.2980686</td>
<td>-1.09880353</td>
<td>No - July 2016 count not within normal distribution with a 27% of occurring by chance.</td>
</tr>
<tr>
<td>Aug</td>
<td>119</td>
<td>79</td>
<td>63</td>
<td>64</td>
<td>81</td>
<td>81.2</td>
<td>22.6980175</td>
<td>-0.008811342</td>
<td>Yes - August 2016 count within normal distribution with a 99% chance of occurring.</td>
</tr>
<tr>
<td>Sep</td>
<td>142</td>
<td>94</td>
<td>55</td>
<td>76</td>
<td>72</td>
<td>87.8</td>
<td>33.3196639</td>
<td>-0.47419446</td>
<td>Yes - September 2016 count within normal distribution with a 64% chance of occurring.</td>
</tr>
</tbody>
</table>

The crime count for the month of May 2016 did not fall within a standard normal distribution based on the Z-score. The likelihood of this distribution occurring by chance was only 12%. Nor were the months of June and July 2016 within the standard normal
distribution based on Z-scores. The probability of each of the distributions occurring by chance was 11% for June and 27% for July. Based on a normal distribution, the results show that in May, June, and July of 2016, the RBNE counts fell outside the expected norm and furthermore, based on their distance from the mean, had a low probability that the crime counts were a result of random chance. Given that the results obtained from the pilot were likely not by random chance, further examination of the RBNE incident numbers was conducted, relative to the previous 4 years.

GIS Analysis

Although the number of RBNEs was lower in the months of May, June, July, and September 2016, when compared to the average between 2012 and 2015, for the months of May, June, and July 2016 the number of RBNEs was not only lower than the average, but also lower than any count in the previous 4-year period. For trend comparison purposes of the geographic area of the study, 4 years of RBNE data prior to the study launch were tallied for each month and year. These monthly totals per year were averaged in order to compare them against the 2016 RBNE counts by month. The results indicated a significant reduction in RBNE crimes from May, June, July, and September in 2016, compared to the average 4-year period of 2012–2015. This is shown in Table 5.

Table 5 Break and Enter Residential Counts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>79</td>
<td>71</td>
<td>74</td>
<td>57</td>
<td>70.3</td>
<td>79</td>
</tr>
<tr>
<td>May</td>
<td>85</td>
<td>89</td>
<td>76</td>
<td>88</td>
<td>84.5</td>
<td>65</td>
</tr>
<tr>
<td>June</td>
<td>87</td>
<td>75</td>
<td>77</td>
<td>85</td>
<td>81.0</td>
<td>58</td>
</tr>
<tr>
<td>July</td>
<td>94</td>
<td>110</td>
<td>60</td>
<td>96</td>
<td>90.0</td>
<td>58</td>
</tr>
<tr>
<td>August</td>
<td>119</td>
<td>79</td>
<td>63</td>
<td>64</td>
<td>81.3</td>
<td>81</td>
</tr>
<tr>
<td>September</td>
<td>142</td>
<td>94</td>
<td>55</td>
<td>76</td>
<td>91.8</td>
<td>72</td>
</tr>
</tbody>
</table>

In detail, Table 5 shows the range of RBNEs per month, over the 4 years preceding the study, and the average and the number of RBNEs counted during the study. For the month of April, the distribution of RBNEs ranged from 57 to 79 within the previous 4-year period, and during the pilot in 2016 the count was 79 incidents, which is higher than the average and is at the high end of the range. For the month of August, the range was between 63 and 119 incidents, and during the pilot study it was 81 incidents, which was
equal to the 4-year average. For the month of September, the number of RBNEs during the study was lower than the previous 4-year average at 72 incidents, but was within the range of 55 to 142 incidents for the previous 4-year period. For the months of May, the distribution of RBNEs ranged from 76 to 88 incidents within the previous 4-year period, and during the pilot in 2016 the crime count was 65 incidents, lower than the range or the average of the previous 4 years. This was also the case for the months of June and July, in which the number of RBNEs was lower than both the range and average of the previous 4 years. For June, the RBNEs numbered 58 incidents during the study, while the range was 75 to 87 incidents with an average of 81. For July, the range was 60 to 110 incidents and the average was 90 incidents for the previous 4 years, while it was 58 incidents during the predictive study. For the month of September, the range was 55 to 142 incidents, and the average was 91.8 incidents, with the value for the study at 72.

Figure 10 presents a line graph representing the columns in Table 5, containing the RBNE counts for the months of April to September for the years 2012 to 2015 and 2016. This graphic representation of the columns of interest represents an overlay of the 4-year average monthly RBNE counts compared to the 2016 RBNE month counts. The months of May, June, July, and September followed a downward trend relative to the 4-year average, with significant drops respectively of 21, 23, 32 and 20 RBNEs for each month. This downward trend in 2016 was a significant deviation from the 4-year average. So far in this chapter the focus has been on the number of incidents. The next section examines in detail the locations in which these RBNEs incidents occurred.
For each month during the VPD predictive study, from April to September 2016, hot spot changes were measured using Kernel Density in 100m x 100m grid cell size. Change detection was calculated using the Esri (2016) ArcGIS raster calculator tool to generate image differences as a way of identifying hot spot changes over a period of time. This method allowed for the measure of hot spot changes between each month (April – September) in 2016, with each month (April–September) overlapping the annual average for 2012–2015, where red indicates an increase in incidents and intensity and blue indicates a cooling of the number of incidents and intensity, as illustrated in Figure 11 below. Furthermore, the difference in image values is represented in the grid cell values, which are used to determine the RBNE hot spot intensity, and, coupled with the fishnet, hot spot distribution and coverage were also determined. It is useful to note that the following figures are not hot spots, but are change detection maps that show change when comparing a single month in 2016 to the aggregate average of 2012 to 2015 for the same month.

Analysis of the change detection values for each 100m grid cell determines whether crime incidents were increasing in intensity, frequency, and numbers after the use of the predictive model, or conversely, whether there was an overall shift to fewer incidents, with a reduction in frequency of occurrences and decreased intensity. The use of 4-year
aggregate data for calculating the change detection values helped to control for anomalous months, where crime patterns might have been inconsistent with previous years for the same area under examination. This analysis cannot be taken in isolation since it does not take into account resource allocation and deployment decisions: these might influence certain patterns that could have been more reflective of crime displacement or staffing than the effectiveness of the model. Determining which factor played a role in the outcomes, whether police operational decisions or the underlying technology, is critically important when assessing the overall VPD predictive study. Subsequent analysis will delve into the effectiveness of the technology itself to help distinguish which might have had a greater influence on overall performance. The following figures present six maps, each representing a month during the 6-month pilot period between April and September (see Figures 11 – 16). The RBNE incidents in each month are represented geospatially with a brief explanation of the change detection comparing the month in 2016 during the pilot study to the aggregate average of the same month from 2012 to 2015.
For April 2016 (see Figure 11), there were several identified changes, where previously cool areas located in Kitsilano, Killarney on the border of Oakridge, Kerrisdale, and Marpole transitioned to high-intensity hot spots. This represents a shift in crime intensity whereby the area changed from lower intensity or not significant to become a higher intensity hot spot. As shown in Table 6, the percentage coverage for the month of April during the pilot study was 50.5%, higher than the average of the previous 4 years. There were several new cool spots that appeared in Kitsilano in close proximity to the new hot spots, with a general cooling effect throughout Fairview, a traditionally high-intensity area. Other areas undergoing a cooling affect were Arbutus Ridge, the northern portion of Kerrisdale, and western Killarney. These changes represent a geo-temporal shift in crime intensity and concentrations, where the new evolving heating or cooling effects were seen in close proximity to either previously-existing cool or hot spots, depending on the change recorded. Generally, high- and medium-intensity hot spots appeared in nearby and previously stable areas that did not have significant concentrations of RBNE. The same was
true for the cooling effect seen in previous hot spots. However, when overall change patterns were assessed, more medium and high hot spots than cooling spots were generated.

**Figure 12** Break and Enter Residential Change Detection May 2016 to May 4-Year Average

For May 2016 (see Figure 12), several newly-created high-intensity hot spots appeared, located in Mount Pleasant, Killarney, and at the border of Fairview and Shaughnessy. However, at 45.2% the total hot spot coverage percentage was lower than half during the research study, and was less than the average 4-year percentage coverage of 54.8% (see Table 6). Newly emerging cold spots developed in Mount Pleasant, Grandview-Woodland, Kensington-Cedar Cottage, Victoria-Fraserview, and Renfrew-Collingwood. Overall, the change detection indicates a reduction in overall hot spot intensity and frequency when compared to the previous 4 years in May, and a preponderant cooling effect influenced the change detection trend.
For June 2016 (Figure 13), there is transition from cold spots or non-significant concentrations to high-intensity hot spots located at the borders of Kensington-Cedar Cottage and Renfrew-Collingwood, as well as at the borders of Oakridge, Kerrisdale and Marpole, as previously seen in April 2016. When compared to Table 6, the total hot spot coverage percentage, however, was lower than half at 44.3% during the study and was also less than the average 4-year percentage coverage of 55.7%. Significant cooling was taking place in Mount Pleasant, Grandview-Woodland, Fairview, and Victoria-Fraserview, which was reflected in the decrease observed in total incidents for the month and a predominant cooling affect that continued from May.
For July 2016 (see Figure 14), there were newly-created high-intensity hot spots at Kerrisdale and Marpole and at the north end of the Kensington-Cedar Cottage areas. The total hot spot coverage percentage, however, was 39.7% during the pilot project and was also less than the average 4-year percentage coverage of 60.3% (see Table 6). Significant cooling was observed at previously identified hot spots, with overall reductions in both frequency and intensity of clusters taking place. The most notable changes to cold spots occurred in Fairview, Mount Pleasant, and Grandview-Woodland. The overall number of hot spots was lower than in previous months, with a general cooling trend dominating the majority of areas. This is consistent with the monthly reduction in total RBNEs recorded at -26% when compared to the previous 4-year average incidents.
For August 2016 (see Figure 15), there were notable shifts in a newly-developing high-intensity hot spot at Renfrew-Collingwood and medium-intensity hot spots at Hastings-Sunrise, Marpole, and South Cambie. The total hot spot coverage percentage was over half at 51.9%, more than the 4-year average at 48.1% (see Table 6). There were notable high-intensity hot spots, coupled with an increase in both the frequency and intensity of crime incident clusters. The transition to cool spots was limited, in area and intensity, to Fairview, Mount Pleasant, Kensington-Cedar Cottage and Grandview-Woodland. This outcome is consistent with the total number of incidents recorded in August, with only a 1% reduction noted when compared to the 4-year average for the month.
For September 2016 (see Figure 16), there were newly-evolving high-intensity hot spot located in Kitsilano, Renfrew-Collingwood, and South Cambie. The total hot spot coverage percentage was over half, at 52.4% – the highest during the pilot project – and was higher than the 4-year average of 47.6% (see Table 6). While September experienced a reduction in overall RBNE incidents compared to the previous 4-year average, the number of high-intensity hot spots that occurred was significant. Despite the increase in high-intensity clusters, the reduction in the total number of incidents and the increase in overall cold spots generated still resulted an overall improved month, compared to the previous 4 years.

Overall, the above figures illustrate the creation of low RBNEs (blue = cold spot) in areas that were previously considered high RBNE crime locations (red = hot spot) that existed from 2012 to 2015 for each subsequent month. The red areas show new high RBNE crime concentrations for the respective months in 2016, which could be attributed to new
emerging RBNE clusters or to an increase in existing RBNE crimes for an area that was not fully controlled by police deployments.

Table 6 represents the hot spot percentage coverage as depicted in Figures 11 through to 16. When analyzing the monthly hot and cold spot coverage between the average 4-year period of 2012–2015 and 2016, the results show that May, June, and July had reductions in the total hot spot coverage. For those circumstances where displacement occurred, it was recorded at a lower intensity than the recurring hot spots that were recorded each year for the same months from 2012-2015.

<table>
<thead>
<tr>
<th>Month</th>
<th>Avg. Year 2012-2015 (%)</th>
<th>Year 2016 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>49.4</td>
<td>50.5</td>
</tr>
<tr>
<td>May</td>
<td>54.8</td>
<td>45.2</td>
</tr>
<tr>
<td>June</td>
<td>55.7</td>
<td>44.3</td>
</tr>
<tr>
<td>July</td>
<td>60.3</td>
<td>39.7</td>
</tr>
<tr>
<td>August</td>
<td>48.1</td>
<td>51.9</td>
</tr>
<tr>
<td>September</td>
<td>47.6</td>
<td>52.4</td>
</tr>
</tbody>
</table>

Generally speaking, there were recurring clusters of common hot spots in the neighbourhoods of Kitsilano, Fairview, Mount Pleasant, Kensington-Cedar Cottage, and Grandview-Woodland for the same month from 2012-2015. When these recurring hot spots were compared with the same month during the predictive study period, there was a prominent shift in the crime pattern distribution recorded, with reductions in intensity from high and medium, and fewer hot spot clusters.

In addition to the analysis of the coverage and intensity levels identified in the change detection comparisons, spatial auto-correlation tests using Global Moran’s I were conducted for each of the change detection months to determine whether the pattern (hot spots) could have occurred by random chance (see Figure 14). The measurement of spatial auto-correlation helps to describe the overall patterns within the predictive area and indicates any detection of unusual deviations from the norm (Griffith, 2011).
The analysis represented in Figure 17 indicated that the change detection results were statistically significant (p-value < 0.01); however, there were minor deviations in the month of September. The calculated Global Moran I's Z-score for September was quite low compared to the months of April to August, indicating less clustering of incidents in hot spots compared to April to August months. In addition, the variance of the intensity level for September showed a more homogeneous pattern compared to the April to August months (see Figure 17).

Figure 17 Spatial Autocorrelation Report for Change Detection April to September
**Break and Enter Residential Distribution – Table and Chart**

The reference grid (i.e., fishnet) consists of a total of 10,632 100m grid cells for the predictive study area coverage. This equates to a coverage area of 106,320 km². Based on this reference grid, it is possible to estimate the distribution changes of the total hot spot coverage area by comparing the average 4-year period of 2012–2015 with 2016. The results demonstrated an overall reduction of hot spot clusters and a reduction in the total surface area for existing hot spots (shrinking in size) for the months of May, June, and July in 2016 (see Table 7 and Figure 18 below).

**Table 7 RBNE Surface Area Distribution Change**

<table>
<thead>
<tr>
<th>Month</th>
<th>Avg. Year 2012-2015 (km²)</th>
<th>Year 2016 (km²)</th>
<th>Difference (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>11810</td>
<td>12090</td>
<td>280</td>
</tr>
<tr>
<td>May</td>
<td>11660</td>
<td>9620</td>
<td>-2040</td>
</tr>
<tr>
<td>June</td>
<td>12490</td>
<td>9920</td>
<td>-2570</td>
</tr>
<tr>
<td>July</td>
<td>14760</td>
<td>9720</td>
<td>-5040</td>
</tr>
<tr>
<td>August</td>
<td>11410</td>
<td>12320</td>
<td>910</td>
</tr>
<tr>
<td>September</td>
<td>12000</td>
<td>13220</td>
<td>1220</td>
</tr>
</tbody>
</table>

Table 7 presents a summary of the surface area distribution of the RBNE comparing the 4-year average and the data collected for the research study in 2016. The table is a numerical representation of the distribution changes presented in the previous figures and tables showing the change in hot spot clusters per month. Table 7 further corroborates the observations noted in the monthly hot and cold spot detection figures noted above (see Figures 8 through to 13). Specifically, it was observed that April, August, and September experienced the overall creation of a greater number of hot spot clusters in traditionally cool areas, and that the subsequent cooling of previous hot spots was not to the same degree as the increase. Conversely, for the months of May, June, and July, there was a significant shift towards a general cooling of previous hot spot areas, as the overall hot spot coverage area in these 3 months shrank to between 2,040 and 5,040 km² (see Figure 18).
The reduction of hot spot coverage within the predictive study area was most evident from May to July (see Figure 18). In addition to the surface area analysis, it was also important to examine whether the reductions observed in hot spot coverage was similarly noted in intensity levels to determine whether the hot spots in certain areas may have dispersed and/or been displaced. The following section examines crime intensity levels for existing hot spots within the study area.

**Break and Enter Residential Intensity Level – Table and Charts**

When analyzing the intensity levels, a 4-year monthly average was used when comparing them with a specific month in 2016 during the pilot project period. From the total hot spot coverage area, it was then possible to further examine each intensity level (low, medium, and high) in detail. The intensity level classification was determined by applying the Natural Breaks (Jenks) classification method because this classification is based on grouping data by identifying similarities in distribution while minimizing variation in each class by comparing the sum of squared deviations to the mean (Mitchell, 2005, p. 54). The resulting values were reclassified to four categories (0 to 3) to facilitate analysis of the intensity levels for the predictive study. Category 0 represents “No Data/Not Significant”; 1 represents “Low Intensity”; 2 represents “Medium Intensity”; and 3 represents “High Intensity”. By combining these categorized values with the reference grid (e.g., fishnet), it was possible to determine the number of grid cells for each intensity level (n). The results in the change
difference for each month in 2016 were compared to the average months from 2012–2015. Table 8 and Figure 19 summarize the change in intensity levels for new and emerging hot spots between the average 4-year of 2012–2015 and 2016 for the same month.

**Table 8**  
**Intensity Level Comparison 2016 to 4-Year Average by Month**

<table>
<thead>
<tr>
<th>Month</th>
<th>Avg. Year 2012–2015</th>
<th>Year 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>April</td>
<td>35.4%</td>
<td>10.6%</td>
</tr>
<tr>
<td>May</td>
<td>40.5%</td>
<td>12.0%</td>
</tr>
<tr>
<td>June</td>
<td>44.0%</td>
<td>10.4%</td>
</tr>
<tr>
<td>July</td>
<td>51.5%</td>
<td>7.2%</td>
</tr>
<tr>
<td>August</td>
<td>41.2%</td>
<td>5.5%</td>
</tr>
<tr>
<td>September</td>
<td>38.8%</td>
<td>7.7%</td>
</tr>
</tbody>
</table>

The intensity level per month, when compared to each month’s previous 4-year average, recorded reductions in high-intensity hot spots for all 6 months of the study period. For the emergence of medium-intensity hot spots, there were reductions for every month except for September. For the emergence of new, low-intensity hot spots, there were reductions in June, July, and August, but increases for April, May, and August. Interestingly, for the 3 months that saw increases in new lower intensity hot spots (April, May, August), these same months recorded the greatest reductions in high-intensity hot spots. For example, April dropped from 3.4% high-intensity 4-year average to 1.1% in 2016. The month of May shifted from a historical high-intensity average of 2.3% to 1.5% in 2016 and August improved from 1.4% to 0.6%. The months of June and July recorded reductions in all three levels of intensity, which corresponds to the sharpest drops in RBNE incidents for those same months. During the study, every month saw a reduction in at least two of the intensity levels when compared to the average of the previous 4-years, and the overall trend was a reduction from high- and medium-intensity hot spots to newly-emerging low-intensity hot spots.
The result of the intensity level analysis showed that there was a reduction in the high and medium-intensity levels where hot spots were known to recur from the same months of 2012–2015. Furthermore, low-intensity levels decreased between May and July. May, June, and July showed significant reductions in the high- and medium-intensity crime levels.

An anomaly was recorded for medium-intensity level clusters in September. This was most likely to have been caused by the inaccurate locational information from PRIME-BC during a data interruption in mid-September 2016 that resulted in an inaccurate predictive model generation. There were no significant changes to the low-intensity level in the month of April, probably as a result of ongoing modifications and updates to the predictive model. While August recorded an increase in low-intensity hot spots, there was an overall reduction in the intensity of medium and high hot spots. The shift from clustered medium- and high-intensity hot spots to an increase in low-intensity clusters still represents an overall improvement, since the shift to more dispersed and less intense clusters indicates that a diffusion effect was still occurring (see Figures 20 and 21).
The following illustrates those RBNEs that occurred from 1601 to 0759 hours for each month. While no predictive policing resources were deployed to the predicted locations during this time interval, the predictive system continued to generate 100m and 500m forecasts. This interval, with no police resources allocated during the timeframe, was used as a control to compare with the period from 0800 to 1600 hours, which had allocated police resources. This forecast interval had no intervention strategies in place; therefore the
model should accurately capture (within the definition of success) those locations where RBNEs actually occurred. When reporting on model performance, the current study was complicated and limited by the design of the algorithm, where false positives were still considered true positives and which formed part of a successful prediction outcome. Under a typical assessment regime, precision and recall are both evaluated, whereby statistical tests such as a F-measure, regression coefficients, and weighted accuracy are conducted. However, these tests are premised on a clear definition of success precision. In the case of this study, there is no direct linear relationship between forecast locations and crime incidents. The model’s success is premised on any one of the six forecasted locations encountering an incident. Therefore, false positives are actually considered true positives, given the design of the model. Forecast locations are considered to be geospatially dispersed throughout the study area at each time interval, with each of the six forecasts representing an equal probability of a positive or negative relationship to crime incidents. However, the remaining five forecasts that do not show a relationship still produce a successful forecasting model, despite the lack of a linear relationship. The use of standard measures is confounded by the definition of success, with five out of six forecasts having an expected null incident return, where analysis would focus on aggregate data, but when the actual unit of examination is object-based. This introduces spatial heterogeneity issues, whereby the explanatory measures for the entire model that are used for testing the hypothesis, do not adequately describe the outcomes at a specific location. Based on the above complexities of defining success and limitations resulting from the design of the algorithm, it was not possible to apply standard measures of precision. In an attempt to report on model performance, the methodology employed accepted the model’s definition of success, whereby a RBNE incident occurring within a two-hour interval, in which one of the six forecasts captured this incident was a true positive, despite the five false positives occurring. These results are detailed below.
For the month of April (see Figure 22), there were 71 predictions in total during the
time period with no resource allocation to the treatment area with 47 falling inside, or
within the 200m of the 100m and 500m prediction boxes, resulting in 66.1% of predictions
being accurate. Forecasting locations were concentrated in Kitsilano, Oakridge, Renfrew-
Collingwood, Riley Park, and Kensington-Cedar Cottage neighbourhoods, but the accuracy
was the lowest out of the study period, with many of the actual RBNEs occurring in
neighbourhoods with very few predictions.
For the month of May (see Figure 23), there were 59 predictions in total during the time period with no resource allocation to the treatment area, with 52 falling inside, or within the 200m of the 100m and 500m prediction boxes, resulting in 88.1% of predictions being accurate. The forecasted locations, compared with the patterns observed in April, became more concentrated in new neighbourhoods and localities. Neighbourhoods with a new, higher concentration of forecasts included Mount Pleasant, Grandview Woodlands, Marpole, and Killarney, which indicated that the forecasting model was adjusting to the redistribution of crime patterns. The concentration of forecasting within Kitsilano remained constant and persistent, as it did in Renfrew-Collingwood, Oakridge, and Riley Park. This pattern may have reflected the lack of action taken to address the forecast crime issues, since areas experiencing hot spots continued with little variation.
For the month of June (see Figure 24), there were 56 predictions during this time period, with 48 falling inside, or within the 200m of the 100m and 500m prediction boxes, resulting in 85.7% of predictions being accurate. The forecast locations, compared to the patterns observed in May, shifted in intensity to new neighbourhoods and locations, although some remained the same. Neighbourhoods with a new and higher concentration of forecasts included Arbutus Ridge, Grandview-Woodlands, and West Point Grey, while clustering in Riley Park and Renfrew-Collingwood decreased. Marpole and Killarney neighbourhoods which had shifted to higher concentrations in forecasts from April to May also saw a decrease in intensity in June. Generally speaking, the model was accurately capturing the areas where RBNEs were occurring, and in the absence of police interdiction strategies, the RBNE crime patterns continued unaltered, with repeated forecasts in those areas experiencing multiple crime incidents.
For the month of July (see Figure 25), there were 73 predictions during this time period, with 63 falling inside, or within the 200m of the 100m and 500m prediction boxes, resulting in 86.3% of predictions being accurate. The forecasting locations, compared to the patterns observed in June, shifted in intensity to new neighbourhoods and locations, although some remained the same. A very high concentration of predictions had occurred in the South Cambie neighbourhood in July, although the intensity had been low in June and there were only a few in both April and May. Kitsilano on the eastern side bordering Fairview also saw an increase, while the intensity in the southern area of the neighbourhood remained constant from previous months. This was also the case for the Oakridge neighbourhood, while Victoria-Fraserview also recorded an increase in the intensity of predictions which had not been seen in the previous months.
For the month of August (see Figure 26), there were 94 predictions during this time period, with 88 falling inside, or within the 200m of the 100m and 500m prediction boxes, resulting in 93.6% of predictions being accurate. The forecast locations, compared to the patterns observed in July, remained constant for the most part, with a few minor shifts into new neighbourhoods. Kitsilano remained high in both predictions and occurrences of RBNEs. South Cambie remained high in the clustering of predictions, and forecast clustering was also experienced in the neighbouring area of Riley Park, which saw an increase in intensity of both predictions and crimes from July to August. New areas of concentration emerged in Grandview-Woodlands, similar to a pattern seen in May.
For the month of September (see Figure 27), there were 80 predictions during this time period, with 64 falling inside, or within the 200m of the 100m and 500m prediction boxes, resulting in 80.0% of predictions being accurate. South Cambie, Riley Park, and Kitsilano neighbourhoods remained at high intensity in the number of predictions as well as the number of RBNEs. Most of the crimes were also captured by the predictions, with eight of the nine in Riley Park, five of six in Kitsilano, and three of three in South Cambie. Renfrew-Collingwood saw an increase in the number of predictions, but the number of RBNEs increased by only one incident, going from 9 in August to 10 in September, although in both months the predictions generated captured all of the crimes that occurred within this neighbourhood.

With resources allocated between 0800 to 1600 hours during the predictive study area, RBNE crimes generally did not occur within the predictive boxes due to the presence of police. However, no police resources were deployed to predictive boxes between 1601 to
0759 hours, despite the system continuing to generate these forecasts. When this time interval was examined in detail, it was observed that RBNE crimes occurred inside or within the 200m predictive boxes. Based on the results represented in Table 9, the predictive model generated an average geographic accuracy of 83.6% for RBNE crimes.

<table>
<thead>
<tr>
<th>Month</th>
<th>TIME (1601-0759 hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inside or Within 200 m</td>
</tr>
<tr>
<td>April</td>
<td>47</td>
</tr>
<tr>
<td>May</td>
<td>52</td>
</tr>
<tr>
<td>June</td>
<td>48</td>
</tr>
<tr>
<td>July</td>
<td>63</td>
</tr>
<tr>
<td>August</td>
<td>88</td>
</tr>
<tr>
<td>September</td>
<td>64</td>
</tr>
<tr>
<td>TOTAL</td>
<td>362</td>
</tr>
<tr>
<td>Prediction Box Accuracy (%)</td>
<td>(362 / 433) *100 = 83.6%</td>
</tr>
</tbody>
</table>

The geographic accuracy analysis showed the months of May, June, July, and August as having the best prediction accuracy in forecasting the likelihood of a RBNE crime occurring at a location, compared to the months of April and September. Of all the RBNE incidents that did occur, there were 10 or fewer RBNE crimes that fell outside of a prediction box buffered area of 200m. Conversely, April and September prediction locations were less accurate, with April having 24 RBNE crimes that were not forecast by the prediction boxes and September having 16 RBNE crimes not forecast by the prediction boxes. To recap, this variance in April was probably the result of ongoing modification, training and updating of the machine-learning algorithm to improve accuracy. The discrepancy in forecast accuracy recorded in September was likely to have resulted from data issues with PRIME-BC that prevented the accurate locational data that was needed for generating forecast locations from being imported into the system.
Emerging Hot spot Analysis

Emerging hot spot analysis was applied from April to September using a 2-week step interval against a space-time data cube for the entire study range. When reviewing the following results, the researcher considered how the graphic representation of the Mann-Kendall statistic test generated a trend Z-score and p-value for each bin and for each time interval slice within separate bins. The fluctuation between dependent geo-temporal trends were flattened for the entire surface area, and were categorized according to the following eight trends of either hot or cold, as illustrated in Table 10 for evaluative purposes.

<table>
<thead>
<tr>
<th>Image</th>
<th>Short Description</th>
<th>Detailed Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="No Pattern Detected" /></td>
<td>No Pattern Detected</td>
<td>Does not fall into any of the hot or cold spot patterns defined below.</td>
</tr>
<tr>
<td><img src="image" alt="New Hot Spot" /></td>
<td>New Hot Spot</td>
<td>A location that is a statistically significant hot spot for the final time step and has never been a statistically significant hot spot before.</td>
</tr>
<tr>
<td><img src="image" alt="Consecutive Hot Spot" /></td>
<td>Consecutive Hot Spot</td>
<td>A location with a single uninterrupted run of statistically significant hot spot bins in the final time-step intervals. The location has never been a statistically significant hot spot prior to the final hot spot run and less than ninety percent of all bins are statistically significant hot spots.</td>
</tr>
<tr>
<td><img src="image" alt="Intensifying Hot Spot" /></td>
<td>Intensifying Hot Spot</td>
<td>A location that has been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of high counts in each time step is increasing overall and that increase is statistically significant.</td>
</tr>
<tr>
<td><img src="image" alt="Persistent Hot Spot" /></td>
<td>Persistent Hot Spot</td>
<td>A location that has been a statistically significant hot spot for ninety percent of the time-step intervals with no discernible trend indicating an increase or decrease in the intensity of clustering over time.</td>
</tr>
<tr>
<td><img src="image" alt="Diminishing Hot Spot" /></td>
<td>Diminishing Hot Spot</td>
<td>A location that has been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering in each time step is decreasing overall and that decrease is statistically significant.</td>
</tr>
<tr>
<td><img src="image" alt="Sporadic Hot Spot" /></td>
<td>Sporadic Hot Spot</td>
<td>A location that is an on-again then off-again hot spot. Less than ninety percent of the time-step intervals have been statistically significant hot spots and none of the time-step intervals have been statistically significant cold spots.</td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Oscillating Hot Spot</td>
<td>A statistically significant hot spot for the final time-step interval that has a history of also being a statistically significant cold spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant hot spots.</td>
<td></td>
</tr>
<tr>
<td>Historical Hot Spot</td>
<td>The most recent time period is not hot, but at least ninety percent of the time-step intervals have been statistically significant hot spots.</td>
<td></td>
</tr>
<tr>
<td>New Cold Spot</td>
<td>A location that is a statistically significant cold spot for the final time step and has never been a statistically significant cold spot before.</td>
<td></td>
</tr>
<tr>
<td>Consecutive Cold Spot</td>
<td>A location with a single uninterrupted run of statistically significant cold spot bins in the final time-step intervals. The location has never been a statistically significant cold spot prior to the final cold spot run and less than ninety percent of all bins are statistically significant cold spots.</td>
<td></td>
</tr>
<tr>
<td>Intensifying Cold Spot</td>
<td>A location that has been a statistically significant cold spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of low counts in each time step is increasing overall and that increase is statistically significant.</td>
<td></td>
</tr>
<tr>
<td>Persistent Cold Spot</td>
<td>A location that has been a statistically significant cold spot for ninety percent of the time-step intervals with no discernible trend, indicating an increase or decrease in the intensity of clustering of counts over time.</td>
<td></td>
</tr>
<tr>
<td>Diminishing Cold Spot</td>
<td>A location that has been a statistically significant cold spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of low counts in each time step is decreasing overall and that decrease is statistically significant.</td>
<td></td>
</tr>
<tr>
<td>Sporadic Cold Spot</td>
<td>A location that is an on-again then off-again cold spot. Less than ninety percent of the time-step intervals have been statistically significant cold spots and none of the time-step intervals have been statistically significant hot spots.</td>
<td></td>
</tr>
<tr>
<td>Oscillating Cold Spot</td>
<td>A statistically significant cold spot for the final time-step interval that has a history of also being a statistically significant hot spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant cold spots.</td>
<td></td>
</tr>
<tr>
<td>Historical Cold Spot</td>
<td>The most recent time period is not cold, but at least ninety percent of the time-step intervals have been statistically significant cold spots.</td>
<td></td>
</tr>
</tbody>
</table>

Looking for spatial and temporal relationships that were statistically significant, the Mann-Kendall time-series test identified only one new cold spot and one diminishing hot spot for each of the 100m fishnet grids across the study area covering the full 183 days. The remainder of the 100m grids indicated “No Pattern Detected” (see Figure 28).

**Figure 28** Emerging Hot Spot Analysis 2016 April to September – 2-Week Interval Aggregate Cube

![Image of Emerging Hot Spot Analysis](image)

**Space Time Cube Analysis**

A Visualized Space Time Cube in 3D was generated and a Getis-Ord Gi statistic test was run on each geo-temporal 100m fishnet grid. For the most part, the visual reading of the pilot study hot spots showed “not significant” for the majority of the areas and time periods. This is consistent with the emerging hot spot analysis that also showed “No Pattern Detected” with only a few exceptions (see Figure 26). The notable exceptions were the last 2 weeks of the pilot study in the Kitsilano neighbourhood of the city and a few isolated spots scattered in Mount Pleasant and Renfrew-Collingwood, where the Getis-Ord Gi identified several hot spots within time-bins that captured the final two weeks of the study. These results can be
attributed to the data corruption issue that occurred with the PRIME-BC data feed to the VPD, which resulted in erroneous data being ingested by the forecasting system. Data corruption issues aside, there were several clusters of hot spots identified in geo-temporal bins representing the borders of the Kitsilano and Fairview neighbourhoods during the month of August, and in Killarney for the same period (see Figure 29). These temporal dependent hot spots were consistent with the other data tests that showed a marginal drop of only 1% in RBNEs in August and the increase in hot spot percentage coverage for the same month.

Figure 29  Visualized 3D Space Time Data Cube – April to September 2016

Summary of Findings
This study approached the research question using a unique thematic, multimodal methodology, grounded in criminological theory and relying on the field of geography to determine geo-temporal data analysis techniques to assess the efficacy of a forecasting system. The findings, represented across a variety of metrics, indicate that the crime forecasting system was likely to have influenced changes in the distribution, intensity and number of RBNEs in May, June, and July 2016. For the remaining months of April, August, and September, and while testing did record changes in several metrics, the findings were inconclusive, and the changes could possibly have occurred by random chance. However, factors such as system outages and machine-learning model training are possible
explanations for the anomalous months. During the control time period (1601 to 0759 hours) when no police resources were allocated although predictions of locations for RBNEs were still being made, results showed that 83.6% of the predictions were accurate. Thus the inference is that the forecasting model was effective in predicting RBNEs. Furthermore, the resulting police actions based on these forecasts contributed to the statistically significant decrease in residential burglaries within the pilot period and indicate that it influenced the distribution and arrangement of crime incidents when compared to historical data.

This concludes the results chapter. The significance of the results presented in Chapter 4 relative to the study’s hypothesis will be discussed in Chapter 5. The study hypothesized that, if the forecasting model and the police actions directed by these forecasts were effective, there would be a statistically significant change in the intensity and distribution, as well as a reduction in the total number, of RBNEs in the study area.
CHAPTER 5 — DISCUSSION

This chapter discusses the significance of the results as presented in Chapter 4. The following sections will address the meaning of the results and discuss the implications of the methodologies used in the study.

Evaluation Outcomes

From April 2016 to September 2016, the VPD predictive policing pilot recorded decreases in RBNEs for specific months when compared to the same month comparison, using an average of the previous 4 years of monthly data. During the previous 4 years, traditional policing principles were in practice to address crime issues in the COV, including RBNE. During this time, crime prevention strategies in a simpler form had been in practice at the VPD – these were informed by and focused on conventional crime analysis techniques. However, with a shift in emphasis to big data and complex analysis of the data using machine learning, the study demonstrated an operationalized application of predictive policing. As a business practice, the use of the technology-guided police resource deployments resulted in both a disruptive effect on crime patterns and concentrations of property crime, but also an overall reduction in the total number of incidents. The following section explores the outcomes recorded in the findings and helps to contextualize the significance of the various statistical tests applied to the data comparisons. Subsequent sections will examine the methodologies used, and will determine whether the application of the model and the associated police resource deployment had a measurable and statistically significant outcome on residential burglaries.

As with many police jurisdictions in North America, the VPD adheres to a CompStat accountability and analysis process. A brief summary of CompStat will be provided to help contextualize the results presented in the previous chapter and to frame the relevance of the findings, based on existing trend and data analysis practices conducted by the VPD. Under CompStat, a summary of crime activity is presented at monthly inspections consisting of aggregated data that has been compiled by analysts to provide trend analysis and percentages changes (Santos, 2016). To contextualize, CompStat products are not utilized for predictive purposes, nor are the outputs produced with the same velocity and volume of
data that could be provided using technological solutions. For the most part, CompStat presents summary statistics in a table format with comparisons of the current 28-day cycle with that of the previous 28 days covering a multitude of crime types (Santos, 2016). In the year prior to the study, crime patterns were predominantly mirrored across the Metro Vancouver Region, with some minor variations observed. The characteristic exception was Richmond that often broke from the regional trend for property crime, with a reduction recorded for the month of July when the remaining municipalities typically saw increases in this month. This exception aside, the municipalities of the Metro Vancouver Region typically followed a similar pattern of increases and reductions in accordance with the expected seasonal variation for each month. Month-over-month comparisons do not provide meaningful information when compared in isolation. Factors such as seasonal variations, weather, vacation patterns, festivals and tourism influence crime trends from month to month, rendering a comparison from one month to the next of little value. Recurring patterns can be observed when looked at over year-to-year comparisons, by quarters or monthly. For example, summer months typically record an increase in RBNEs, possibly as a consequence of homes being left insecure, large number of vacationers resulting in unattended houses, and generally more opportunities (Cohn & Rotton, 1999; Santos, 2016). This trend is seen year after year for the summer months with little variation (Owen, 2009). To address these known seasonal variations, an examination of 4-year trends for the same month was conducted, as well as a comparison of neighbouring jurisdictions and the regional as a whole, for the same evaluative period.

The regional RBNE crime statistics for March to September 2015 were compared to the same period in 2016. The 2015 chart indicated that property trends tended to mirror across the region with a few exceptions in the more stable “bedroom” communities of Metro Vancouver, specifically Richmond (see Figure 30).
Any spikes or downward trends in one municipality are usually seen across the other municipalities. When compared to the surrounding municipal jurisdictions, in 2016 the research study area represented an inverse relationship, where RBNEs dropped from April to July, while this trend was not mirrored by the other municipalities. The research study area recorded significant reductions in June, compared to Surrey which experienced a notable spike. The study area deviated significantly from the observed trend patterns of the other surrounding jurisdictions by dropping when the others either remained static or increased.

The 2016 comparison deviated from this behaviour, with the study area showing a marked reduction from May to July, and the remaining municipalities experiencing a spike in June, with Surrey increasing from April to June and again in September.
The most interesting aspect of the comparison is the month of June, in which most municipalities experienced a significant spike in occurrences at a time when the study area was recording a substantial drop in RBNE. Also, to contextualize the graph (see Figure 31), the scale used was necessary to accommodate the crime trends of the larger municipalities of Surrey and Vancouver, which experienced the greatest number of incidents in the Metro Vancouver area. However, smaller municipalities such as North Vancouver and Richmond, with roughly 1/6 of the population of Vancouver, still experienced a modest increase for their size (Owen, 2009). For example, from May to June 2016, North Vancouver saw an increase from 20 to 31 RBNEs. For a population of approximately 137,000, this represents a significant increase in property crime from one month to the next.

Table 11 presents a comparison of statistical measures and tests. Based on a normal distribution, the results show that in May, June, and July of 2016, the RBNE counts fell outside of the expected norm and further, based on their distance from the mean, had a low probability that the crime rates were a result of random chance. Therefore, there is some indication that an external factor occurred during this period that is likely to have influenced the low RBNE counts and spatial pattern. The fall in RBNE counts can also be
seen on the associated maps that will be discussed in detail in the following sections, where there were a number of distinct clustering of RBNE crimes in May, June, and July not a result of chance. Compared to April, August, and September, there were more scattered small clusters of RBNE crimes, showing distributions that were probably not a result of chance. Table 11 and Figure 32 both examined monthly 4-year average RBNE counts from 2012 to 2015, compared to the same months in 2016.

### Table 11  
**Comparison of Statistical Measures and Tests**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr</td>
<td>79</td>
<td>71</td>
<td>74</td>
<td>57</td>
<td>79</td>
<td>72</td>
<td>9.05538514</td>
<td>0.773020683</td>
<td>Yes - April 2016 count within normal distribution with a 56% chance of occurring.</td>
</tr>
<tr>
<td>May</td>
<td>85</td>
<td>89</td>
<td>76</td>
<td>88</td>
<td>65</td>
<td>80.6</td>
<td>10.1143462</td>
<td>-1.542363651</td>
<td>No - May 2016 count not within normal distribution with a 12% of occurring by chance.</td>
</tr>
<tr>
<td>Jun</td>
<td>87</td>
<td>75</td>
<td>77</td>
<td>85</td>
<td>58</td>
<td>76.4</td>
<td>11.4804181</td>
<td>-1.602729084</td>
<td>No - June 2016 count not within normal distribution with a 11% of occurring by chance.</td>
</tr>
<tr>
<td>Jul</td>
<td>94</td>
<td>110</td>
<td>60</td>
<td>96</td>
<td>58</td>
<td>83.6</td>
<td>23.2980686</td>
<td>-1.09880353</td>
<td>No - July 2016 count not within normal distribution with a 27% of occurring by chance.</td>
</tr>
<tr>
<td>Aug</td>
<td>119</td>
<td>79</td>
<td>63</td>
<td>64</td>
<td>81</td>
<td>81.2</td>
<td>22.6980175</td>
<td>-0.008811342</td>
<td>Yes - August 2016 count within normal distribution with a 99% chance of occurring.</td>
</tr>
<tr>
<td>Sep</td>
<td>142</td>
<td>94</td>
<td>55</td>
<td>76</td>
<td>72</td>
<td>87.8</td>
<td>33.3196639</td>
<td>-0.47419446</td>
<td>Yes - September 2016 count within normal distribution with a 64% chance of occurring.</td>
</tr>
</tbody>
</table>
Similar to the regional comparison, the months of May, June, July, and September recorded significant drops when compared to the 4-year average, with reductions ranging from 20 to 32 RBNEs per month. Taken in isolation, these figures are insufficient to reject the null hypothesis. However, when taken in combination with the other metrics as a whole, the argument to assess the influence that the forecasting model may have had on the crime trends within the study area becomes considerably more compelling.

The number of RBNEs that occurred in the COV during the 2016 pilot study period was lower than the monthly averages in May, June, and July of each of the four years between 2012 and 2015. A statistical analysis of the data supported the findings of a decrease in the number of RBNEs. Inferential statistical analysis returned very low values for Z-scores that were represented as a percentile value. Furthermore, the analysis revealed that the May, June, and July RBNE counts had a very low probability of occurring by chance, while the counts for April, August, and September were within the expected value range.

Specifically, the month of April 2016 recorded a 2% reduction in RBNE when compared to the monthly 4-year average for April. This value fits within the normal range for that month, recorded from 2012 to 2015. The probability of that count occurring is 56% of the time calculated on the baseline metrics. For May, June, and July the reduction of RBNEs was 21%, 27%, and 26% respectively (see Table 3). The drop in RBNEs fell outside of the normal expected range (one standard deviation) with the probability of those rates
occurring respectively in 12%, 11%, and 27% of the time (see Table 11). August and September both fell within the normal range, with the probability of those same rates occurring in 99% and 64% of the time. Based on these figures taken in isolation from Table 11, it would indicate that April, August, and September reductions could not be attributed to having resulted directly from external influences, such as the predictive project, and that the drops recorded could possibly have been a result of yearly variation. Conversely, May, June, and July offence reductions were outside of the expected norm and were unlikely to have occurred by random change, indicating they were influenced by unanticipated factors, possibly the effects of the pilot project.

Hot spot changes using Kernel Density and calculated using the Esri (2016) ArcGIS raster calculator generated images to represent change detection in intensity over a period of time (see Figures 8 to 13). Geospatial and geo-temporal analysis techniques were used. Specifically, a reference grid combined with a change detection algorithm was used to compare monthly hot spots. The comparison was made between the RBNE hot spots that were common from 2012–2015 and the same areas in 2016. There were two notable findings. First, when comparing the monthly distribution, common hot spots that occurred in 2012–2015 were displaced in 2016. Stated another way, for several of the months, hot spots transformed to cold spots during the pilot study. When 2012–2015’s common hot spots were compared to new RBNE crime clusters, there was an overall reduction in total cluster surface area for the months of May, June, and July. In other words, change detection analysis for May, June, and July showed a significant reduction in the overall hot spot area coverage, as well as notable reductions in high- and medium-crime intensity levels during the predictive study period, when compared to the 4-year average of 2012–2015. Spatial auto-correlation tests determined that the RBNE crime clusters (new hot spots) for April to September 2016 did not occur by random chance, which may be inferred to have resulted from targeted enforcement directed by the forecasting system.

For the month of April 2016, during the start of the pilot study, a higher than average percentage coverage of high-intensity hot spots was visible, representing a temporal shift in crime intensity that had existed previously in a lower intensity and degraded to a higher intensity hot spot; alternatively, it had previously been a stable area with no significant concentrations but had shifted into a high-intensity hot spot (see Figure 8). On the borders
of the Oakridge, Kerrisdale, and Marpole neighbourhoods, a high-intensity hot spot can be seen in April, June, July, with a medium intensity in May and August.

The theory of repeat victimization states that “people and places that have been victimized have a higher likelihood of being victimized again” (Farrell & Pease, 2001, p. 67), and it is possible that this has been demonstrated through this area showing repeatedly as a hot spot. Although a granular, detailed examination of the addresses was not conducted to pinpoint whether these were exact repeat residential burglary occurrences or near-repeat events within a set spatial neighbourhood (Short et al., 2010), it is likely that there was a risk heterogeneity in this neighbourhood that needed to be examined. This raises an important issue: being able to predict crime and allocate resources for prevention or enforcement is one potential use of predictive policing; but understanding the underlying reasoning and why a particular neighbourhood might be repeatedly victimized and finding methods other than allocating police patrol resources for prevention, is of equal importance. Specifically for the study, a closer look at the neighbourhood characteristics shows that Kerrisdale is 60.9% detached housing with 50.6% of the housing being single detached; Oakridge is 50.7% detached housing with 38.3% being single detached; while Marpole has 37% of housing as detached (COV, n.d.-a). There is a public transit hub, a community college, and a shopping centre in the area, and easy access to neighbouring cities outside of the jurisdiction of the VPD. This would be consistent with the crime pattern theory that for areas that are more likely to be frequently visited by potential offenders, there is a higher likelihood of crime occurring (Santos, 2016).

This, coupled with routine activity theory, can provide additional insight into crime patterns, where it is possible that certain activities are associated with a transport hub, with easy access to transiting the city limits. Furthermore, close proximity to shopping centres and a nearby college could facilitate offenders blending in on foot but, being difficult to access by cars, including police vehicles, could also provide a theoretical explanation as to why the area is prone to repeated victimization. In contrast, rational choice theory would suggest that the presence of a police officer has a deterrent effect on individuals who might consider committing a RBNE. The decision-making process, based on a risks and rewards assessment made by a potential offender, may have been influenced by the allocation of police resources (Santos, 2016) and would thereby have affected the aggregate property
crime numbers in the study area. The assumption is that the police presence added another element of risk and the potential for negative consequences for criminally-motivated individuals (Santos, 2016).

In addition, as crime moves in response to targeted law enforcement activity (Barr & Pease, 1990, p. 33), it is possible that RBNEs occurring in the study area had moved into a jurisdiction other than the COV. Interestingly, the City of Surrey had a directly inverse relationship compared to the COV study area in August and September, at a time when the pilot study area saw significant decreases in the number of RBNEs and there was a sharp increase in the same crime type for Surrey. Although this evaluative study takes into consideration the areas within the COV that were not included in the pilot and assesses spatial and temporal alterations in crime, from assessing the change in the distribution and intensity of residential burglaries it is possible that other neighbouring police jurisdictions also experienced displacement. In addition, it is possible that another type of displacement could have occurred, specifically, crime-type displacement (Weisburd et al., 2005). While this was an interesting possibility that came to light during the research, the answer is somewhat elusive, given the limitations of the data collection, and therefore is outside the scope of this study. Although residential burglary statistics were generally shown to have decreased, an examination of the aggregate of all crime types during the pilot period was not conducted. Therefore, it is possible that individuals who traditionally committed RBNEs may have switched to shoplifting or theft from autos, based on the targeted enforcement. This potential adaptation by offenders to a new offence category could have reduced the number of the aggregate RBNE crime statistics, but not reduced the aggregate property crime statistics as a whole, when associated crimes, such as theft from autos, are considered.

The typical monthly temporal distribution of RBNEs was examined in order to determine a shift model for the purposes of devising the most effective resource development to cover the times that were most likely to experience the targeted crime (see Figure 6). It was established that the majority of RBNEs were taking place between 0700 and 1600 hours, which resulted in police resource deployments being concentrated around this time. While this chart examined the temporal distribution from 2015’s monthly aggregate data, Figures 27 and 28 recorded the temporal patterns of RBNEs within the study area for
the hours of 0800 to 1600 hours and 1601 to 0759 hours for the years 2015 and 2016. In 2015, the majority of RBNEs for the geographic area in which the study was eventually conducted occurred between 0800 and 1600 hours for each subsequent month, with the exception of April and August. For those 2 months in 2015, the number of RBNEs was greater between the hours of 1601 and 0759 hours, but lower for the other months (see Figure 4). In 2016, the greater number of RBNEs for the months of April, May, and June occurred between 0800 and 1600 hours. However, in the later 3 months of the study, July, August and September experienced more RBNEs from 1601 to 0759 hours (see Figure 5). On initial review, it appears that there was a temporal shift in the distribution of RBNEs within the study area, following the first 3 months of the pilot project. This outcome deviated from what might have been expected, since past crime data tended to indicate that this would not be the case, with a greater number of RBNEs occurring from 0800 to 1600 hours. Although this is of interest and unexpected, the statistical validity of the changes in temporal distribution was not meaningful because there were too few incidents for an accurate calculation.

In context, there are studies indicating that although displacement has been researched there is limited evidence for a complete movement of crimes to another location or time, and when it is present, the increase has been modest (Ratcliffe, 2002). Cause and effect are therefore extremely difficult to determine, given the many variables that could potentially affect the number, location, time, and types of crime. It is possible that temporal displacement occurred after the first 3 months of the study, possibly as a result of offenders modifying their behaviour patterns and shifting their activity times to periods when no police resources were dedicated to predictive policing deployments. However, given the minimal variance between both the two time intervals and the inherent limitations of the comparison data collected, it is impracticable to draw any conclusions.

The research study had the advantage of being able to assess the time interval that had the benefit of police resource allocation versus the time interval without resource interventions, in order to assess the effectiveness of the forecasting model. The evaluation of predictive policing, as it was applied at VPD, was mostly limited to 0800 to 1600 hours, since that was the timeframe to which project resources were allocated and represented the focus of the data collection for the pilot study. For the time range of 1601 to 0759 hours,
police resources were not deployed to forecast locations that continued to be generated by
the system. This absence of police resources created an opportunity to assess how the
model performed without any interventions disrupting the predicted crime incidents from
occurring.

**Crime Forecasting Model Effectiveness**

In assessing model effectiveness, a series of assumptions was made based on an analysis of
officer deployment activity during the day. On review, it was determined that dedicated
police resources, either CSP or regular sworn officers, rarely stayed within the confines of
the forecast location. For both 100m and 500m forecast locations, police resources typically
ventured approximately 200m outside of their zone. Therefore, officer activity zones were
actually 300m and 700m respectively. Given the tight constraints of a 100m radius, which
equates to approximately four to six houses, officers were consistently recorded as
deviating at a minimum of 200m outside of their forecast zone. The same also held true for
500m forecast zones which were normally attended by mobile patrol CSP units. Driving a
police vehicle within a 500m radius for 2-hours appeared to be too limiting, and these units
also tended to cover a far larger area, with an additional 200m, or approximately one block,
being the norm. Given these operational realities, the evaluation of the effectiveness of the
directed patrols in preventing RBNEs also reflected these practices by expanding the
forecasting zone to include the actual areas being patrolled. While the timeframe used to
evaluate the model’s accuracy was outside that of regularly scheduled police resources, a
200m expanded buffer was used as part of the evaluation criteria, since it captured the
activity zone in which officers would have a deterrent effect more realistically.

**Figure 33   Expanded Police Activity Zone**
The key assumption is that if officers had been deployed to the evening and early morning shift (1601 to 0759 hours), they would have continued with the regular practices of increasing their patrol areas by at least 200m and any potential RBNEs within that area would have been either detected or deterred (see Figure 33).

In the case of the 1601 to 0759 hours timeframe evaluation covering the study period, it was found that the predictions were 83.6% accurate. In other words, 362 of the 433 RBNEs fell inside or within a 200m buffer surrounding either the 100m or 500m forecast zones (see Table 9). During the months of May, June, July, and August, when there were no data issues related to machine learning or omissions, the number of predictions that fell outside of the actual crime locations were in the range of 7% to 15% (see Table 9). Given that this is the rate of accuracy without police intervention, confidence can be placed on the predictive accuracy of the output produced by the model. Furthermore, it was significant to record that the aggregate number of RBNEs was higher in July and August for the non-resource allocated shift from 1601 to 0759 hours than it was for the 0800 to 1600 hours shift that had the resource allocation, when historically the RBNEs are concentrated between 0800 and 1600 hours. Although cause and effect are difficult to determine for any crimes, the decrease in the number of RBNEs during the time that resources were allocated vis-à-vis that of the other shift, and comparing the two results in the same study area, may indicate that the presence of police resources was possibly a factor in changing the pattern of RBNEs during the pilot. In summary, the results obtained from this study are significant when examined in the context of the prevailing regional trends, the reductions in intensity of clusters recorded, and the high forecasting accuracy that fell within the definition of model success.

**Methodological Implications**

In this section, further analysis of the results is conducted based on the implications for the methodology used.

The VPD pilot was a predictive policing programme, in that in practice it did what the NIJ (n.d.) defined as predictive policing, by using advanced analytics in conjunction with intervention. This study played a key part in establishing that the predictive policing technology and the resource allocation was effective and had the intended effect of
reducing crime, specifically RBNEs. Evidence-based decision making, which is an underlying drive towards ILP (James, 2013; Ratcliffe, 2004), holds that for predictive policing technologies to be considered evidence-based, the models must be tested in practice. True to the predictive policing explanation in Perry et al. (2013), the pilot was about implementing a business process to prevent the crime that was being predicted (p. 128).

Specific to property crime, RBNEs have been documented as being predicable and one of the most patterned crime types (Farrell & Pease, 2001; Johnson et al., 2007; Short et al., 2009, 2010). Therefore, with a predictive model premised on generating forecasts using the most patterned types of crime, as such, if the model is shown to perform as intended, it does so using a data type with the best chance of assessing its true efficacy. However, regardless of the outcome of the study on the effectiveness of the model itself, the larger issue rests with the methodological approach taken to evaluate the model and determine its effectiveness. As noted previously, a review of the literature found that there was little interest in empirically testing the benefits of predictive policing practices; there were also issues with the evaluation of the models being used (Berk & Bleich, 2013). In addition, studies that specifically examined GIS applications were also found to be lacking in rigour and measurable outcomes (Zhang et al., 2014). Although one purpose of this study was to evaluate the efficacy of a machine-learning, spatial-temporal crime forecasting model on RBNEs, another rationale for conducting the study was to establish a methodology for assessing evidence-based and replicable predictive policing applications. Given the lack of technically and scientifically sound research, publications, and literature, this study is important in that whatever the outcome may have been for the primary research question, the methodologies used for this study can be replicated by other police services.

With regard to the neighbourhoods, during the pilot there were specific locations that were identified across multiple months as high-intensity hot spots. It is possible that the intensity moved from high to medium in April to May due to targeted resource allocation, intensified again in June and July but then eased to medium intensity before eventually disappearing in September. It is important to note on this topic that the system, by design, identifies emerging property crime patterns and police resources are then deployed to disrupt this activity. This resource allocation is intended to displace and disrupt an emerging crime pattern before it can continue. The process is repeated, which at an
aggregate level has a reduction effect on property crime. This process is very similar to hot spot policing, whereby problem crime areas are targeted by police in an effort to disrupt and prevent the continuation of the activity. The operant difference is that crime forecasting aims to respond to incidents before they become issues and is more proactive as opposed to reactive by nature. As a result, if a crime forecasting system is operating at peak efficiency, there should be very few crime incident hot spots, as the interdiction strategies should prevent them from continuing in the same manner and pattern.

However, for a number of reasons there is a possibility of crime forecasts being repeated in certain areas. One reason could be the recurring victimization of the neighbourhood by certain offenders, who could be drawn to areas that lack physical deterrents or the target-hardening of the homes themselves. While the crime forecasting system should prevent the occurrence of crime-incident hot spots, the potential for an area to encounter crime-forecast hot spots still exists or, in other terms, there is a concentration of prediction boxes in a specific neighbourhood that is repeated over time. Although the reason behind the intensity is beyond the scope of the thesis since this is a quantitative study focusing on outcomes, this may be a subject of interest for a follow-up study. However, it should be noted that there were intensity changes in the hot spots, resulting in their eventual disappearance by the end of the pilot. It is also important to mention that although the pilot study outcomes appeared to be positive, and that the focus of this study was on the effectiveness of how police actions might have affected RBNEs in the area, the overall goal for public safety and crime reduction in a jurisdiction should not be limited to police resource allocation and should include solutions that address the underlying causes of crime.

The numerical values of the summary statistics in themselves are significant; however, further discussion of what the statistics meant, beyond the possibility that the changes during the pilot months exceeded what might randomly have been expected, is needed to determine whether the predictive policing pilot project was effective. It is important to note that there were two variables of interest in the study, those that were geographically defined and those temporally defined. These factors were considered to assess the accuracy of the forecasting model, and the efficiency of the resource deployments based on the forecasts. Geographically, the study area excluded West End and
Downtown Vancouver, as both areas in total had less than 0.2% of detached homes (COV, n.d.). More than 99% of residential housing in the two areas was high density, and consisted mostly of high-rise apartments. Given the fact that previous studies had found single family dwellings to be 15% more likely to be victimized by RBNEs than apartments (Bernasco & Nieuwbeerta, 2004), it was determined that the two neighbourhoods would be excluded. Furthermore, since the forecasting system could not predict floors or units within a building, making enforcement and strategic resource allocation for prevention was not possible. In the event that a building address was identified, patrolling all floors or all possible entrance and exit points in a building for “suspicious” individuals was not an efficient use of resources, especially when the locations were high-rise buildings where a motivated offender could find various entry and exit points and/or hiding places for themselves or stolen goods, had they wished to do so.

Because there was such a vast difference in the types of dwellings between the area in which the pilot was conducted and the area that was excluded, it was not possible to compare them against each other; therefore they could not be defined as a control. In addition, based on the typical monthly RBNE temporal distribution on historical data, resources were not allocated between 1601 to 0759 hours to the pilot area. This meant that the prediction boxes generated by the forecasting system were examined with both the presence and absence of officer activity according to the time interval. When examining prediction boxes when no police resources were deployed, a geographic analysis of the model indicated that 83.6% of the RBNE incidents were accurately captured within the prediction boxes during the time period from 1601 to 0759 hours. It is important to note that these results took into account the total number of prediction boxes for the monthly time period that was generated by the model, where a success was represented by one out of the total of six forecasted locations per 2-hour interval that recorded a hit.

In terms of an operational police shift model, the net influence of the model on property crime can only speak to the shift that ran from 0800 to 1600 hours, since that was the timeframe in which project resources were allocated and thus represents the focus of the evaluation. Significant improvements were recorded in reducing RBNE during this period, with historical data confirming that the majority of RBNEs occurred within that window of time. Based on the findings from the running of the pilot, it is reasonable to infer that 0800
to 1600 hours was an effective shift schedule. Further evaluation and study would need to be conducted in order to assess alternate deployment times as well as shift scheduling and what effect it had on the RBNEs. However, as a machine learning model the predictive system did have an opportunity to learn from the data it ingested, by factoring in the outcomes when interventions by the police were carried out and during the times without police intervention, even though the actual preventative allocation was applied only to the one shift interval.

Predictive policing is more than analysis; rather it is a systemic process that includes actions in response to the data processed (Perry et al., 2013). When emerging hot spot analysis on the space-time data cube was evaluated for this study, an interesting outcome was recorded. For the emerging hot spot analysis, very few patterns were revealed over the course of the full 183 days of the study. The exceptions were one new cold spot and a diminishing hot spot. For the remaining 100m fishnet grid covering the city, no other patterns were detected. This outcome is consistent with the intended functioning of the forecasting model, whereby the system is premised on the early identification of RBNE patterns, and the subsequent deployment of police resources to interdict or deter continuation of the events. When this occurs at peak efficiency, several results transpire. One possible scenario is that the offender is displaced to another neighbourhood or an area in which police resources are absent. While this may not at first seem an optimal outcome, research has shown that repeated and prolonged targeting of crime intensity locations and the ensuing displacement has a known reduction effect (NASEM, 2018; Santos, 2016). A critical aspect of this approach depends on sustained resource allocation that adjusts to new and emerging hot spots which develop as a result of the displacement – in other words, a crime prevention strategy that adapts to new and evolving hot spots. If these emerging hot spots are subsequently targeted with police resources and the offenders are repeatedly displaced, then there is a net reduction in occurrences and a diffusion of benefit effect (NASEM, 2018; Weisburd & Braga, 2006a; Weisburd et al., 2005). Research on hot spot policing is plentiful, and generally speaking, most agree that when applied properly, reductions in the range of 10–25% are attainable (NASEM, 2018, p. 119; Santos, 2016, p. 546; Weisburd et al., 2005).
A second possible outcome from the deployment of officers to forecast locations is the identification and apprehension of offenders responsible for RBNEs. This was especially the case when looked at in the context of the VPD’s deployment strategy, which included using unmarked patrol cars with plain-clothes officers. In this scenario, a tangible reduction in RBNE could possibly be attributed to the arrest and conviction of known offenders. The lack of identifiable crime patterns was probably attributable to the forecasting system operating efficiently, where police resources had disrupted and interdicted RBNE locations that the system had pre-emptively targeted. By deploying resources to areas that contained the potential for hot spots to develop, the system had effectively prevented any new hot spots from occurring, with no discernible pattern emerging over the 183 days of the study as a result. The results showed that crime prediction was only one aspect of the predictive policing process; once resource allocation was established, it had to be sustained and the forecasts had to remain dynamic for the benefits to continue.

It is important to consider that resource allocation can take many forms and a strategy must be in place for forecasts to be used effectively. The types of action taken will depend on the type of strategy being utilized, which can be a uniformed officer in a marked police car for deterrence, versus plain clothes officers in unmarked police cars for apprehension purposes. Whichever strategy is chosen as a result of the resource allocation, the outcomes will contribute further to the machine learning process that will have a direct effect on future predictions. As in the case of data selection, resource allocation strategies must be thought through at the implementation phase, and a definition of effectiveness and success for the predictive system should be defined.

Following on the emerging hot spot analysis, the Getis-Ord Gi test results were similarly expected, given the functioning of the system and resource allocation. The Getis Ord Gi tests specifically for the existence of hot spots in relation to the neighbours of each 100-meter geospatial bin. This test then examines hot spots at each specific time slice interval, in relation to neighbouring bins in the same time interval. For the month of August, several hot spot clusters were identified in the Kitsilano, Fairview and Killarney neighbourhoods. This indicated statistically significant hot spots in these areas where identifiable clusters that spanned several two-week intervals (temporal bin interval) occurred in relation to their neighbours. On closer examination, the hot spots did not occur
for more than three time-bin slices and were generally transitioning from high to medium to low intensity over the intervals in question. While the existence of any hot spot clusters is not optimal, care should be taken to contextualize the hot spots in the larger study area. The existence of hot spots was identified for a limited period of time and was confined to a small geographic area within the three neighbourhoods where the patterns were eventually extinguished, as against being persistent over time and space. The remaining geo-temporal bins within the study area did not experience any hot spot activity, with the exception of the last weeks of September which was attributable to the PRIME-BC RMS crash and subsequent data corruption.

**Big Data, Artificial Intelligence and Predictive Policing**

The current study did not focus on big data, AI, or machine learning per se; however, it did evaluate the operationalization of the output of big data-fed machine learning AI as used for the purposes of forecasting RBNE. The results of the evaluation showed that the first month of the pilot implementation in April did not have the same effect on forecasting as did the following months, after the predictive system had had a chance to learn and train on the data. The idea that the machine network learns from experience (Vallor & Bekey, 2017) occurred with this particular system in the first month of April. Based on change detection analysis, the predictive study recorded a significant impact in the months of May, June, and July. This can be seen in the reduction of RBNE counts, and a reduction in all intensity levels, as well as in the distribution of the hot spot coverage that shifted from medium and high to low intensity. April and August showed little change in the RBNE counts, which was also corroborated by the standard deviation and Z-score calculations. Taken in context, the predictive model was still undergoing a testing and development phase during April, and was modified three times for predictive accuracy.

At the onset of the pilot project in April, the model was updated and modified several times to take account of the rapidly shifting crime patterns that had quite possibly been caused by police intervention strategies at the predicted locations. During this training and adjustment period, the predictive system locked, and specific locations were repeated until the next update was processed. Although this was part of the learning process for both the system and the end users, and the issue was resolved by the first week of May, it nevertheless had implications for the month’s forecasting. This adjustment helped the
machine to learn the relationship between input data and the output being sought. Despite only minor changes in the RBNE counts for April, there was still a reduction in the high- and medium-intensity levels, which may have pushed out and increased the distribution at the low-intensity level. A similar effect was recorded in August, with an increase in low-intensity hot spots, while medium- and high-intensity hot spots reduced. It still represented an overall improvement, since the shift to more dispersed and less intense clusters indicated that a diffusion effect was still occurring.

The vulnerability of the AI and machine learning system being only as good as the data it was provided with became evident when technical issues with the provincial data provider (BC-PRIME Corp.) responsible for pushing the data, failed to send complete information. The September results can be considered an anomaly, due to the fact that from mid-September 2016 onward, PRIME-BC data did not contain locational information. Following a PRIME BC province-wide RMS crash in September, the predictive system had begun to receive erroneous and missing data fields, and the failure of provincial RMS and the resulting data loss seriously affected the predictive system’s ability to generate accurate forecasts. The month of September was therefore problematic when used for comparison and analysis purposes, and the erroneous or missing location data resulted in an absence of updates for the prediction zones; therefore, prediction boxes were repeatedly displayed in the same areas. Although, it might have been prudent to remove the September data from this analysis, it was included because the first 2 weeks (before the PRIME-BC RMS crash), did provide good results. This is shown in the RBNE counts and the reduction in the high- and low-intensity levels.

This issue of no location information experienced in September was not discovered until the conclusion of the pilot study period. Given these circumstances, greater weight should be placed on the months from May to August, when the system was running at peak efficiency. The importance of the veracity, or the validity, accuracy and usability of data crucial to big data (Gupta & Saxena, 2016) was demonstrated with the occurrence of the data push anomaly. This occurrence is a good reminder that humans are still responsible for supervising the learning process of AI and for ensuring that the data is accurate, which AI is not able to determine on its own. The reliance on AI for predictive forecasting is not equated to error-free or completely automated intelligence that can monitor itself for
validity and value. A human being needs to supervise and predetermine the types of data and determine the value of the data that will be used, as well as establishing whether there are any inherent biases, errors or issues with the data. The volume and velocity, once the data sources are determined, does not have to be controlled by humans; however, the other aspects require human beings with expertise, knowledge, and an understanding of policing and resource allocation, as well as the overall purpose and the capabilities of the technological component of the predictive policing.

In summary, with the emphasis on modelling, machine learning and algorithms, it is often forgotten that in order for the technology to perform the tasks that it was designed to do, it needs to be supplied by humans. Data is created by humans, or automated by human-generated processes into an output that is designated by humans. In no way is this process free of human decisions and processes, although the patterns developed may seem to be removed from humans and therefore have epistemological authority (Kaufman et al., 2018). The temporary technical issue from the data provider was a useful reminder for this study, and also for future studies, that the data itself does not have inherent authority and should not be taken without scrutiny.

Data Concerns
As most predictive policing technologies operate on the premise that past crime data can be used for prediction (Bachner, 2013; Kaufman et al., 2018; NIJ, n.d.), it is possible that any issues inherent with past crime data would be carried over into the future, if it is used as a forecasting source. Crime or location forecasting has been raised as an area of concern when the data being consumed by the machine-learning technology includes offences and police reports that are police-initiated and generated. The rationale behind this concern is the possibility that personal biases, prejudices, and perceptions of the police officers will artificially skew who they check, and the areas in which they conduct active enforcement (Lindsey, 2018; SLSC, 2018). These decisions may be based on personal perceptions of what they believe are higher crime areas, such as neighbourhoods that are socio-economically disadvantaged or composed of visible minorities. As a result, crime incidents recorded and subsequently used by the predictive technology can represent inherent biases and are more reflective of police preconceptions than objective measures of community issues (Mantello, 2016). The troubling aspect of this becomes accentuated when the technology blurs the
front-end collection and ingestion that ultimately forms the nucleus of the ‘big-data’ system. The outputs are often represented as objective and scientifically valid by virtue of the use of advanced technology that is touted as being racial and socio-economically neutral (SLSC, 2018; Waldman et al., 2018). Furthermore, the black box problem of outputs that are difficult to understand or interpret without technical or mathematical knowledge (Ozkan, 2018) contributes to the perceived authority of patterns and algorithms (Kaufman et al., 2018).

The black box problem, defined as “higher performance but less clarity” (Ozkan, 2018, p. 217) exists for machine learning models that are highly complex, and provides accurate predictions that are difficult for researchers to understand or interpret when determining how the outcome was achieved. However, a more detailed examination of policing practices and the inherent culture of the police service in question may reveal a different reality. In this situation, the criticism by civil rights and civil liberties advocates can be valid, especially when taken in the context of the spate of high-profile police shootings and the development of the Black Lives Matter movement (Day, 2015; Luibrand, 2015). There is concern that by removing the human element from the decision-making process on how police are deployed will result in the propagation of the belief that the outcomes are evidence-based and unbiased, because of the false perception that machines cannot be biased. However, as discussed above, the processes that allow for outputs within the predictive policing system require human input every step of the way, including the materials that comprise the foundation of learning, what is learned, and ultimately what, if anything, will be done with the forecasted information.

Although concerns were raised about the use of data that resulted in the prediction of crime locations, major concerns were directed towards predictions aimed at identifying individuals (Lum & Isaac, 2016). An individual who was visited by the Chicago Police, based on a predictive methodology that identified him, stands as an example. He was identified as being of interest, even though he had not had any recent negative contacts with the police, did not have a violent criminal record, and had not committed a crime (Lum & Isaac, 2016). Studies examining individuals such as serial burglars (Fox & Farrington, 2016) also have the same inherent bias, in that the data is only available for those individuals that have been apprehended and for whom the information is available. Specific to the VPD crime
forecasting system that was empirically tested for this study from methodology to analysis were ethical concerns, including data integrity and bias which were of central importance. In April 2016 the VPD was the first in Canada to pilot the operational deployment of a machine learning, crime-forecasting system which this study evaluated. It is worth noting that the VPD took great care to ensure that adequate measures were taken to help safeguard the data and ensure it was as bias-free as possible.

Initially the VPD sought to address potential issues, such as an unintended outcome involving targeting of citizens when predictive technology resulted in the over-policing of marginalized or ethnically-diverse neighbourhoods. Closely tied to its preventative strategy was the overall ethical application of the technology and ensuring that privacy rights were respected and there was transparency in the deployment of police resources. Following on from the actions of the EU which has recently implemented specific regulations and disclosure requirements concerning the collection and use of data amassed from the general public by both corporate and government entities (the GDPR), governments and law enforcement agencies in particular have been observing this legislation with interest (EU, 2016). Closer to home in Canada, The Montreal Declaration for Responsible Development of Artificial Intelligence (Beliveau, Bengio, & Hebert, 2018) was a proactive undertaking by an academic and government consortium to develop a set of general guidelines in the use of AI within Canada. The declaration focused on several key areas, with concerns raised over the types of data used and broad access to private data sources at the forefront, followed closely by the potential harm the use of automated decision-making systems could have on the privacy rights of individuals (Beliveau, Bengio, & Hebert, 2018). Specific technology was considered immaterial to the discussion, as the greater risk was placed on the outcomes being sought and for what purposes the outcomes might be applied.

In tandem with establishing protocols for the ingestion of data, the crime forecasting system itself was designed to focus only on locations; it does not target people, nor does it import information on individuals. Being restricted solely to location-based processing eliminates the potential for any ethical issues around the targeting of individuals, and as a final step, the crime forecasting system has embedded programming logic to ensure that specific communities or neighbourhoods do not have an over-representation of crime prediction boxes and a disproportionate allocation of police resources. Particular attention
is paid to sensitive policing zones within the COV, specifically an area known for socio-economic challenges and high rates of drug addiction called the Downtown Eastside. The system was built with ongoing tracking and audit processes to monitor crime forecasts at an aggregate level, highlighting any areas that are over-represented by the number of forecasts. It is important to note that the addition of the programming logic to exclude over-representation of prediction boxes was not the decision of the AI, but a conscious predetermination made by key individuals involved in the project. Ultimately it is the responsibility of the individuals engaged in predictive policing to take steps to address the potential consequences of a system meant to contribute to public safety.

To this end, the VPD intentionally excluded the use of police-generated data, and instead only processed citizen-generated property crime incidents. While community-generated data can still contain levels of bias and prejudices, removing police-generated data from the system inputs helps to control any underlying organizational issues, should they exist. While the VPD has a long history of positive community engagement and proactive participation with marginalized members of the community, the added step of removing any perceptions of police bias helps to strengthen public trust in the use of the technology. As the data source is derived from the community, there is a better chance for the outcomes generated to be perceived as less biased and police-centric than if the data were to be generated from police data only. However, any errors, omissions and mistakes in the data, intentional or otherwise, have consequences if used as the source of analysis. This was observed during the VPD pilot study, which saw a dramatic decrease in prediction accuracy when the data fed into the system contained erroneous or non-existent location data during the last month of the pilot testing. The resulting inferential statistical analysis showed that when the input from the data source omitted the location information, it resulted in repeat locations being forecast, due to the lack of updated data. Any resource allocation to the repeated locations only resulted in further erroneous learning, contributing to the cycle of errors. This shows that despite the care and attention provided to the initial data selection process, analysis, training, and the use of the output, along with continued monitoring by humans to validate and monitor the forecasts, are important aspects of predictive policing and big data solutions.
One of the dimensions of the big data paradigm is value, arguably the most important in the opinion of some, since it is the primary motivation and reason for processing datasets that require technology and systems to analyze and use (Gupta & Saxena, 2016). The return for the initial investment for the technology, training and change must make economic sense in the longer term for it to be considered by many police agencies, as the costs associated with start-up would normally not be included in a budget. Predictive policing has been elevated as a means of achieving greater efficiencies and creating an environment where police services can operate more effectively with limited resources (Prox & Griffiths, 2015). Police services have typically applied ILP practices, including the more recent introduction of predictive policing technology, to support and enhance tactical and strategic operations (Peterson, 2005; Ratcliffe, 2016). A dichotomy develops between the need to adopt advanced and often expensive technology, and the requirement to be fiscally prudent. This evaluation study has provided the evidence required to verify that the predictive technology piloted in VPD was effective and has the potential to provide direction for targeted resource allocation to increase efficiencies. Since financial and economic considerations weigh heavily on the drive towards policing based on technological solutions that rely heavily on data, an evaluation should be an essential component of implementation. Despite the ubiquitous application of ILP within the policing environment, it was traditionally undertaken “without adequate evaluation or the establishment of predefined metrics to assess whether the outcome matched the intended purpose” (Perry et al., 2013; Peterson, 2005; Sargsyan & Prox, 2019, p. 31; Weisburd & Braga, 2006a). This drive for efficiency and the need to use technology to be perceived by the community as being successful (Chu, 2001), influences the type and timing of its adoption by agencies. This study served the purpose of providing evidence but has also been a foundation for providing information to the public for transparency.

The VPD pilot was not based on vendor-provided software, nor was the evaluating researcher the developer of the model affiliated with the AI production. Neither the evaluation study nor the researcher received any benefit from the production of the system, and there was no product endorsement requirement influencing the outcome of the evaluation, although such potential biases and influences have been of concern in some predictive policing applications and studies (Beck, 2012; Beck & McCue, 2009; Rosenbaum,
2007). The evaluation was conducted to assess empirically not only the system’s ability to predict RBNEs reliably, but also to assess the implications of the operationalization by the VPD: there have been other modelling studies that addressed only the system predictions, but not in conjunction with resource allocation and in real time, without the influences of stakeholder bias.

This chapter examined the results and outcomes of this study as well as its broader implications for policing. Future research on predictive policing can be expected to meet predefined metrics for assessment as well as the criteria for objectivity and data analysis used in this research. This study was conducted to fill gaps in the literature and research, since police services have typically applied ILP practices, including predictive policing, without adequate evaluation or the establishment of predefined metrics for assessment (Perry et al., 2013; Peterson, 2005; Sargsyan & Prox, 2019, p. 31; Weisburd & Braga, 2006b). The study also put into practice and operationalized an evaluative methodology that could be applied to a predictive model to determine whether an algorithm can be used for the purposes of police resource allocation. The results of the efficacy of the forecasting system obtained through summary statistics, inferential statistics and geo-temporal analysis, showed that the crime forecasting system was likely to have influenced changes in the distribution intensity and number of RBNEs for the months in which the system was piloted in the VPD. More important than the positive results of the evaluation, however, is that the methods utilized in this study are compatible with the policing paradigm, namely the need for what is perceived to be evidence. This concludes the discussion chapter. The sixth and final chapter of the thesis summarizes the key research findings and recommendations based on them.
CHAPTER 6 — CONCLUSIONS

This study evaluated the utility of a 6-month pilot that applied a crime-forecasting model specifically developed to target RBNEs within a police operational setting. Within the academic literature, there are few independent evaluations and a general lack of replicable studies on the topic of predictive policing, with the many reports having been influenced or conducted by the industry that developed the technology and as a result lacking any objective or verifiable outcomes. In contrast, this study applied quantitative data, methods, and analytical measures to assess the efficacy of the use of the technology, by which means empirical evidence provided insight to these questions.

Key Research Findings

Four key research findings were produced based on the results of this study:

1. The police resource allocation, based on the crime forecasting system, likely influenced changes in the distribution, intensity, and number of RBNEs during the months of May, June, and July 2016.

2. The forecasting system was effective in that it was successful in predicting RBNEs. The predictive system forecasted locations of RBNE incidents with 83.6% accuracy during the pilot period when no resources were deployed.

3. The method used for this study can be replicated and built upon as a standardized method of evaluating technological applications related to predictive policing solutions and processes. The study produced an evaluation methodology that is valid and reliable, and informed by empirical data that provides a framework for future evaluations of geospatial predictive policing applications in operational policing environments.

4. Both data and human components are important in the implementation of predictive policing technology. Although there is significant attention provided to the use of algorithms, models and patterns, a data error that went undetected by the technology was identified by human resources, and even though recommended locations were provided by the technology, the actions taken, and the decisions on what actions would be taken, were dependent on a human
making them. The implications of these key research findings are discussed further in the following sections.

**Research Implications**

The findings from this study appear to corroborate the hypothesis that there would be a reduction and change in the distribution and intensity of RBNE in the study area, as a result of police interventions guided by the forecasting model for the months of May, June, and July 2016. The study was able to support these findings by comparing the results of the pilot project to the baseline data using a multimodal quantitative approach. The significance of this finding is that the outcome of the operationalization of the crime forecasting technology was evaluated independently and objectively, using empirical methods that can be replicated, verified and supported by other researchers.

The second facet of the hypothesis was that a high percentage of RBNEs will have been predicted by the model during the time that there was no police action taken, as determined by comparing the predicted locations with actual RBNEs. With 83.6% accuracy of the model recorded during the pilot period, this aspect of the hypothesis was supported, and the geospatial methodology used for this assessment also contributes to the third finding.

Both Findings 1 and 2 required a well-developed multimodal quantitative methodology. This study is unique in that it addresses a gap in the criminology research, publications, and literature that is concerned with predictive policing applications, and that it is empirically sound and based on operational policing practices rather than on models and theories. The intention is to publish the findings from this study, specifically the methodology utilized. The method used in this evaluation will add to the existing body of knowledge in GIS studies, as the geospatial analysis techniques used have not previously been applied to operational policing. The specifics of the data collection portion of the methodology can be summarized in the following four steps:

- First, aggregate crime incident data were collected for the study area and the surrounding region, covering a 5-year period. The intended application of the data was to establish baseline measurements of crime trends and patterns.
• Second, geospatial-temporal data related to officer deployments and activity were obtained to account for police intervention strategies and influencers on crime.

• Third, geospatial-temporal crime data were collected for the study area for the 4 years prior to and including the 6 months of the study period.

• Fourth, crime forecast geospatial-temporal data were collected, which was represented as either 100 or 500 square meter areas.

The list of analysis techniques used were as follows:

• Z-Scores on inferential statistics

• Kernel density estimation

• Mann-Kendall time series

• Emerging hot spot using relational space-time cube

• Getis-Ord Gi

The analysis portion of the methodology required image analysis that involved measurement for hot spot changes using the Kernel Density in the 100m grid cell size. The change detection was calculated using Esri (2016) ArcGIS raster calculator tool to generate image differences to identify hot spot changes over each month of the pilot study and the previous 4-year average of that month detecting change in intensity and incidents. The image analysis methodology produced a hot spot percentage coverage allowing for a comparison to be made between the previous 4-year average to the pilot study results. Further, the distribution of the total 100m grid cells for the predictive study area was calculated and the distribution changes estimated for the total surface area. The intensity level of all the concentrated areas within the pilot study were also examined and categorized into a standardized baseline of low, medium and high, and in addition compared to the previous 4-year average. A spatial auto-correction test using Global Moran’s I was conducted for each of the change detection outcomes, to determine whether the pattern of change detection could have occurred by chance.

For the control time interval, break and enter incidents that actually occurred were mapped out with the predictive boxes. A distinction was made as to whether the prediction fell inside or within 200m of the prediction boxes and counted with the percentage
calculated for those crimes that landed inside or within 200m. This provided the prediction accuracy percentage for the month and the total for the pilot period. Emerging hot spot analysis was applied to all the months of the pilot study, using a 2-week step interval against the space-time data cube for the study range. A Mann-Kendall time-series test identified cold spots and hot spots to determine if a pattern existed. A space time cube analysis was generated, and a Getis-Ord Gi statistics test was run on each of the geo-temporal grids to identify hot or cold spots.

The fourth finding is significant in that predictive policing is founded on large volumes of unbiased and error-corrected quality data; this then becomes the focus and human contributions are often overlooked. Although this study focused on the quantitative, measurable outcomes of the pilot project, it also concluded that decisions, overall project management and resource allocations are human based and need to be addressed in the implementation of any predictive policing technologies.

**Recommendations for Future Research**

There may have been a diffusion of benefits, based on the operational deployment of police resources, which was not measured in this study. There may also have been an effect on other types of crime at other locations or other times. It is recommended that future research measures the aggregate number of all property crimes, including RBNE, during and after the pilot. This approach would assess whether there was an overall reduction in property crime, which would be demonstrated by a reduction in the aggregate number of another type of crime. An examination of the aggregate number of crimes in both the study area and the surrounding jurisdictions would determine whether there were any peripheral benefits of crime reduction. There could have been a diffusion of benefit beyond the RBNE crime type into the neighbouring jurisdictions and it would be beneficial to include this examination in future studies.

By virtue of this study being quantitative by design, it did not incorporate any qualitative elements, although inclusion of a qualitative approach into the examination of predictive policing has been suggested recently (Goerzig, 2016). To this end, although the study was quantitative in focus, future research may consider collecting qualitative data about measures of success, based on interviewing or surveying deployed police personnel.
and their assessment of the efficacy of the intervention strategies. The success or the assessment of efficacy based on supervisors, managers, and project development resources may also be of benefit, in order to understand how best to integrate future technologies into operations to achieve reductions in crime. Efficiency may be defined in ways other than crime reduction and may not be the only definition of success. Other factors contributing to a successful project may be managerial and administrative in nature, such as a reduction in expenses, or could include measures such as streamlining of processes and procedures. In addition, a comparison between predictive policing technologies and hot spot policing using crime analysis should be considered, to determine whether there are any differences in outcomes. In this regard, there are no comparative studies that examine hot spot analysis programs, where the analysis is consistently generated and actioned on a timely basis and how that policing model compares to crime forecasting.

Specific to the outcomes of this study and the research findings, it would be of benefit to follow up on possible reasons as to why certain neighbourhoods were repeatedly victimized. For these selected neighbourhoods, there existed a pattern where a hot spot was detected and correctly identified by the forecasting model, resulting in police resources being deployed. The area then underwent a cooling effect, only to return to being a hot spot, and the cycle repeated, despite the ongoing resource allocation to the area. For these locations there are possible root causes and environmental factors that, although beyond the scope of this study, may have contributed to the repeat victimization. Other possible avenues for future research to enhance the current research are plentiful in this field, as more police agencies adopt predictive policing technologies and methodologies.

**Recommendations for Practitioners**

Given the ethical and civil rights implications in the use of this technology, there is in fact considerable risk and liability exposure for public sector agencies. As such, the development of an Industry Code of Conduct that has undergone an independent and objective review process, would help to ensure public confidence and protect against any technical engagements that might bring controversy to an organization. While the Canadian-developed Montreal Declaration imparts a broad ethical framework for industry and public bodies, including guiding principles and human-centric values in the use of technology, it does not provide specifics as to how to achieve this in a real and pragmatic way (Beliveau et
For companies and public bodies that are engaged in the use of machine learning algorithms, a detailed checklist of mandatory practices is needed, ranging from independent ethical review mechanisms to integrity checks for data bias, and including a minimum standard for reviewing algorithms for inherent programming bias. Tied to this are specifics as to what level of transparency is required concerning the functioning of the algorithms and the parameters for verifying and testing AI prior to usage. These are very real issues within the Canadian context, and there are currently no standard practices that speak to these necessities.

For Western countries, an industry code of conduct for automated decision making is a foregone eventuality, especially when taken in the context of developments in the EU and its recent GDPR (EU, 2016). Moving forward, there is a need to develop and draft specific guidelines and a certified industry checklist, as a proactive engagement process versus a reactive response to government intervention. This is an important consideration, as technology advances rapidly and AI becomes more and more assimilated into everyday policing tools. Beyond the current predictive study, AI is being integrated into facial recognition software, social media data mining technology, and the Internet of Everything (IoT; Koper et al., 2009; Lussenhop, 2018). It is foreseeable that law enforcement will start to branch into these areas and begin utilizing this technology as part of their operations and investigations. AI-driven facial recognition software is the likely next step, as an efficiency multiplier in forensic video. Given the controversy on the use of AI, it would be prudent for law enforcement and the public sector to engage actively in an independently-certified review and implementation process that outlines very specific guidelines and practices. Being party to these developments will help ensure that law enforcement will still be able to use advanced AI technology, and equally important, do so in a manner that maintains public trust and avoids adverse consequences.

While the study of an AI-based crime forecasting system raised many questions concerning ethics and current practices of police services, it also brought to the forefront the real-world application of algorithmic-based policing and the techno-social implications on how this technology could potentially shape and influence future policing practices. In the end, the spatial crime-forecasting model, as applied by VPD through the operational deployment of resources, effectively reduced the number of property crimes during the
pilot period, indicating that there was utility in using the technology within this specific context. How or whether this technology can be applied to a broader mandate is still undetermined, indicating there still exists a void in evaluative studies in this area and significant opportunity for future explorative research.
REFERENCES


To Whom It May Concern,

The Vancouver Police Department grants Mr. Ryan Prox permission to reproduce and use in whole or in part, any data or information collected as part of the Predictive Policing Pilot Project, including but not limited to text, tables, figures, photos, charts, graphs, drawings, images, logos and substantive portions of the evaluation, analysis and internal report. Mr. Ryan Prox may use this information and data for publication and inclusion as part of his requirements for a Doctor of Law Enforcement and Security with the Australian Graduate School of Policing and Security, Charles Sturt University, without seeking further permission.

It is understood that the data, findings and analysis will be used as part of a published dissertation and includes granting Charles Sturt University Library licence for circulating, archiving copies and permitting photocopying. This also includes a non-exclusive licence to Library and Archives Australia and its agents for microfilm and PDF use.

[Signature]

Adam Palmer
Chief Constable 1393
Vancouver Police Department
APPENDIX B: VPD Data Access Approval and Research Permission

Charles Sturt University

Chief Constable Adam Palmer
Vancouver Police Department
2120 Cambie St.
Vancouver, B.C.
Canada V5Z 4N6
September 26, 2016

Dear Chief Constable Palmer,

I am currently enrolled in the degree of Doctor of Policing and Security at the Australian Graduate School of Policing and Security, Charles Sturt University, Australia. My research relates to crime forecasting, with a focus on the Vancouver Police Department’s (VPD) pilot study on the use of this technology. Using crime statistics collected by the VPD as part of their records management system, this study will examine the number of residential burglary incidents in the study area. Descriptive statistics will also be used to summarize the baseline historical property crime data covering the last five years. Forecasted crime locations generated by the system in the form of 100 meter and 500 meter buffers will also be used to evaluate accuracy. The general areas in which resources were deployed in the form of geographic information system (GIS) summary data, namely x/y coordinates, will also be used.

As such, I seek your approval to access these data. The data I am requesting does not contain any identifiable information, such as personal details of police officers or the public. Further, no specifics related to burglaries is sought that could identify a victim of crime or a police officer. The data used will be summary in nature; that is, similarly available online through the VPD public accessible open-data catalogue and public GeoDash system. The open-data catalogue (public internet) property crime data contains the GIS x/y coordinate of the incident and the date/time of the incident. The significant difference between the open data and requested data is that it is specific to the study area versus city-wide.

If you approve my use of the VPD’s data, I would be grateful if you would acknowledge this by noting "Approved" at the bottom of this letter, including the date that you granted permission.

A copy of the research results will be included in a final dissertation that will be lodged with the Charles Sturt University library and will be available to the general public. If you require verification of any of these details please feel free to contact my research supervisor, Associate Professor Patrick Walsh at Tel: +61 993215210 or by email at pawash@csu.edu.au.

Best Regards,

Ryan Prox
Graduate Student:
Intelligence and Security Studies, Australian Graduate School of Policing and Security
Charles Sturt University, Collins Beach Road, Many NSW 2005
Tel: +1-604-313-4346 Email: ryan.prox@shaw.ca

[Signature]

[Approved]

[2016-CA-26]
APPENDIX C: Charles Sturt University Research Ethics Committee Approval

25 February 2017

Mr Ryan Prox
By email: ryanp@shaw.ca

Dear Mr Prox,

Thank you for submitting your Research Proposal to the Charles Sturt University Human Research Ethics Committee. Your research proposal was considered at the 13 February 2017 meeting.

The Charles Sturt University Human Research Ethics Committee is constituted and operates in accordance with the National Health and Medical Council’s National Statement on Ethical Conduct in Human Research (National Statement).

Based on the guidelines in the National Statement the Committee has approved your Research Proposal. Please see below details of your Research Project:

Project Title: The Utility of Predictive Policing: Evaluating a Spatial-Temporal Forecasting Model in an Operational Deployment

Approved until: 1 February 2019 (subject to annual progress reports)

Protocol Number: H17017 (to be included in all correspondence to the Committee)

Progress Report due by: 10 January 2019

You must report to the Committee at least annually, and as soon as possible in relation to the following, by completing the ‘Report on Research Project’ form:

- any serious and/or unexpected adverse events or outcomes which occur associated with the research project that might affect participants, therefore, the ethical acceptability of the project;
- any variation to the research project and/or research design (Committee approval required);
- if an extension to the research project period is required (Committee approval required); and
- once the research project is complete.

This approval constitutes ethical approval in relation to humans only. If your research involves the use of radiation, biochemical materials, chemicals or animals, separate approval is required by the appropriate University Committee.

Please contact the Governance Office on (02) 6338 4628 or ethics@csu.edu.au if you have any queries.
The Committee wishes you well with your research.

Sincerely

[Signature]

Ms Regan McIntosh
Governance Officer

Cc. Associate Professor Patrick Walsh and Associate Professor Henry Prunckun