Manufacturing & Service Operations Management

Prescriptive Analytics in Urban Policing Operations

Tobias Brandt, Oliver Dlugosch, Ayman Abdelwahed, Pieter L. van den BergDirk Neumann

Prescriptive Analytics in Urban Policing Operations

Tobias Brandt,[a,b] Oliver Dlugosch,[c] Ayman Abdelwahed,[b] Pieter L. van den Berg,[b] Dirk Neumann[c]

[a] European Research Center for Information Systems, University of Münster, 48149 Münster, Germany; [b] Rotterdam School of Management, Erasmus University, 3062PA Rotterdam, Netherlands; [c] Chair for Information Systems Research, University of Freiburg, 79098 Freiburg, Germany

Contact: tobias.brandt@wwu.de, https://orcid.org/0000-0002-2127-1921 (TB); oliver.dlugosch@is.uni-freiburg.de, https://orcid.org/0000-0001-7296-2290 (OD); abdelwahed@rsm.nl, https://orcid.org/0000-0003-2473-102X (AA); vandenberg@rsm.nl, https://orcid.org/0000-0002-8550-6769 (PLvdB); dirk.neumann@is.uni-freiburg.de, https://orcid.org/0000-0003-2178-3705 (DN)

Abstract. *Problem definition:* We consider the case of prescriptive policing, that is, the data-driven assignment of police cars to different areas of a city. We analyze key problems with respect to prediction, optimization, and evaluation as well as trade-offs between different quality measures and crime types. *Academic/practical relevance:* Data-driven prescriptive analytics is gaining substantial attention in operations management research, and effective policing is at the core of the operations of almost every city in the world. Given the vast amounts of data increasingly collected within smart city initiatives and the growing safety challenges faced by urban centers worldwide, our work provides novel insights on the development and evaluation of prescriptive analytics applications in an urban context. *Methodology:* We conduct a computational study using crime and auxiliary data on the city of San Francisco. We analyze both strong and weak prediction methods along with two optimization formulations representing the deterrence and response time impact of police vehicle allocations. We analyze trade-offs between these effects and between different crime types. *Results:* We find that even weaker prediction methods can produce Pareto-efficient outcomes with respect to deterrence and response time. We identify three different archetypes of combinations of prediction methods and optimization objectives that constitute the Pareto frontier among the configurations we analyze. Furthermore, optimizing for multiple crime types biases allocations in a way that generally decreases single-type performance along one outcome metric but can improve it along the other. *Managerial implications:* Although optimization integrating all relevant crime types is theoretically possible, it is practically challenging because each crime type requires a collectively consistent weight. We present a framework combining prediction and optimization for a subset of key crime types with exploring the impact on the remaining types to support implementation of operations-focused smart city solutions in practice.

1. Introduction

In today's data-driven world, prescriptive analytics has been gaining traction in both research and practice (Bertsimas and Kallus 2019). Although definitions for prescriptive analytics abound, we understand it as the culmination of the analytics value chain, transforming data into concrete recommendations for decision makers. Depending on the context, this may include understanding patterns in historical data; predicting future developments; and employing heuristics, learning algorithms, optimization techniques, or scenario analysis to determine the recommended way forward. Recent examples can be found in a variety of business domains, such as optimizing customer segmentation and targeting based on customer data (Nair et al. 2017) or creating decision support tools that optimize maintenance assignments (Angalakudati et al. 2014). Such leveraging of data-driven algorithms to improve

both strategic decision making and day-to-day service operations has not only become a key objective for private-sector companies, but is also becoming increasingly relevant in the public sphere. The emerging *smart city paradigm* reflects this development, with municipalities using sensors, data, crowdsourced citizen input, and analytics to improve public services and tackle complex urban challenges.

In this context, a smart city can be perceived as a system of systems (Qi and Shen 2019) with adjustments to one system potentially affecting a variety of connected systems and phenomena. For example, changes to urban mobility through the introduction of electric vehicles and smart charging strategies may impact the city's power grid (Valogianni et al. 2020). The complexity of navigating such potential externalities a given smart city solution may exert is added to the complexity of improving the focal system or service. Particularly in the context of prescriptive analytics, the latter form of complexity derives, for instance, from operationalizing data, identifying the appropriate prediction and optimization methods, and integrating these methods into a cohesive approach as explored in current research by, for example, Elmachtoub and Grigas (2020) and Wilder et al. (2019). The management of smart city operations that utilize prescriptive approaches, therefore, requires a thorough understanding of how these components are linked within the problem at hand and how a potential solution affects adjacent systems in the smart city ecosystem.

In this paper, we contribute to this understanding by exploring operations management (OM) challenges associated with the application of prescriptive analytics to urban policing operations, a critical service in communities around the globe. *Prescriptive policing* links the prediction of urban crime with the assignment of patrol cars to specific areas in the city in order to improve both crime deterrence and the patrols' response time. Much of the extant research on this topic has exclusively focused on the prediction angle, seeking to identify data sources and methods that improve crime forecasts. As we illustrate in this paper, combining those forecasts with optimization methods that assign patrol cars in an efficient manner is not a trivial task, reflecting the problem-specific complexity previously discussed. The likelihood of a given crime occurring at a specific place and time is exceedingly low; different crime types are driven by different, possibly contradictory dynamics; a constrained resource needs to be allocated such that the actual crimes are best addressed; and there are competing measures of the quality of the resulting allocations. We first consider each of these problems separately—prediction, assignment, and evaluation—before analyzing how the solution approaches work together and identifying key trade-offs.

Building upon a comprehensive data set from the city of San Francisco, we calibrate established prediction methods to produce varying levels of prediction performance and combine these with two different optimization approaches that target different objective functions. We find that, although combined approaches that leverage the prediction method with the best prediction quality (PQ) generally lead to the best decision quality (DQ) according to the optimization objective, this is not necessarily always the case. Particularly when considering the trade-off between the two target objectives, approaches that use weaker prediction methods may still provide Pareto-efficient outcomes. Effectively, this implies that decisions regarding the choice of prediction and optimization methods should not be made sequentially; instead, the choice of prediction method can be considered a key parameter of the optimization method depending on the desired trade-off between the target objectives. Furthermore, we find that the dynamics underlying different critical crime types lead to different optimal vehicle allocations. To overcome the complex challenge of making the relative weight of each crime type explicit within the objective function, we propose an approach that combines prediction and optimization with exploration for practical implementations of prescriptive policing. This *predict–optimize–explore* approach also provides a framework for decision makers to navigate between complexity that is made explicit within the focal problem and potential externalities on other systems and phenomena in smart city application beyond prescriptive policing.

In the next section, we first provide detailed background information on the policing context, including a discussion of target objectives and relevant work on crime predictions. In Section 3, we introduce the case context and outline the distinct subproblems constituting the prescriptive policing challenge as well as the solution approaches we explore in our analysis. In Section 4, we present the results of our analysis. In Section 5, we discuss key implications and the predict–optimize–explore approach. Section 6 concludes.

2. Background and Related Work

Algorithmic, data-driven support for operations in both police forces and the judicial system has substantially grown in recent years, ranging from software supporting probation decisions and sentencing to improved targeting of police resources (Metz and Satariano 2020, Sassaman 2020). The impact of this development is heavily debated with concerns over, for example, racial and sociodemographic biases, policing intensity, and accountability shaping that discussion (Ferguson 2017, Shapiro 2017). In this paper, we focus on one problem from the policing environment, namely, the dynamic assignment of a constrained number of patrol cars to different parts of the city based on crime

predictions. Although not without controversy—Sassaman (2020) argues that police forces in the United States continue to grow despite the purported efficiency gain from algorithmic decision support—the fundamental idea that patrol forces should be used in the most effective manner receives broad support. Field patrols constitute the largest share in police departments' budgets (Mak 2020), and efficiency gains have the potential to provide relief to municipal budgets that are often stretched thin. Consequently, preventive and proactive measures that strategically leverage police presence in hot spot areas to quickly react to and possibly prevent crime have become an important goal for research and practice (Braga et al. 2014).

Response time and *prevention/deterrence* are also the key objectives that are generally associated with the allocation of patrol forces. Response time represents a fundamental measure of police service quality (Dunnett et al. 2019, Leigh et al. 2019) as a shorter delay before arriving at the scene of the crime may increase chances that suspects are immediately apprehended and that victims' recollection of events are still fresh. However, research also shows that the presence of police officers has a deterring effect on crime (Sherman and Weisburd 1995, Di Tella and Schargrodsky 2004) within their immediate surroundings. As this effect is highly localized, there is likely to be a trade-off between these key objectives. Patrols that are allocated to maximize deterrence are likely to be concentrated in hot spot areas where a single patrol can cover a high number of potential incidents. In contrast, minimizing response time may induce a more spread out allocation.

For decades, improvements in police operations have focused primarily on reactive approaches to improving these key performance criteria. For instance, Chaiken and Dormont (1978) develop solutions for allocating vehicles to police precincts, allowing departments to optimize for a selection of performance criteria, including response times. Other approaches use historical crime patterns to model future crime incidence in addition to purely analytical models. For example, Taylor and Huxley (1989) describe a program for allocating police officers to precincts based on forecasted service calls and time to resolution, and Chelst (1978) uses a simulation of patrol cars to estimate the expected number of crimes that would have been intercepted by police cars in different regions.

Recent advances in data management and computing have led to an increased focus on prediction in crime prevention approaches. For instance, Al Boni and Gerber (2016b) augment the commonly used kernel density estimation (KDE, Chainey et al. 2008) and show prediction improvements for 17 different crime types using this *local KDE* approach. Al Boni and Gerber (2016a) find that training separate models for

different ZIP codes improves prediction for some crime types as the driving factors behind crime patterns may vary across geographical regions. Other approaches consider the use of autoregressive integrated moving average predictions (Chen et al. 2008) and the application of bio-surveillance techniques for crime pattern detection (Neill and Gorr 2007).

A related stream of work explores the use of novel sources of information in addition to historical crime data to infer crime incidents. Kang and Kang (2017) show that using data on demographics, housing, and education as well as image data from Google Street View can improve crime prediction. Other work investigates how data from social media, such as Twitter, can be utilized both indirectly (only the tweets' time stamps and locations) and directly (the tweets' content) to improve crime prediction (Blevins et al. 2016, Williams et al. 2016). Wang and Gerber (2015) use Twitter and Foursquare data to track and predict spatial trajectories of users to derive inferences regarding future crimes. Gerber (2014) uses a crime hot spot analysis with KDE for 25 different crime types in the city of Chicago as a base case. In a subsequent step, latent Dirichlet allocation (LDA) is applied to the tweets to extract higher level information. LDA groups words from text documents into clusters that can be interpreted as topics (Blei et al. 2003). The strength of each cluster is then used as input for the prediction. Based on this data representation, Gerber (2014) finds an improvement of the prediction quality for most crime types. However, the interpretability of the results is limited as the reason for specific topics relating to a crime increase is unclear. In contrast, dictionary-enabled approaches analyze tweets along different linguistic dimensions, such as use of certain grammatical structures or valence according to different emotional categories (e.g., anger, joy, sadness) and are also shown to be effective in improving prediction quality (Chen et al. 2015, Ristea et al. 2017).

The results from Gerber (2014) also illustrate a critical issue associated with most research on crime prediction techniques. Models are often calibrated to provide predictions on a daily basis, thereby omitting intraday spatial and temporal variance. This is problematic insofar as crime predictions are supposed to serve as input for patrol routes and similar operational decisions. With crime hot spots shifting throughout the day and probabilities of a given crime occurring at a specific area and time span being extremely small, the practical use of daily predictions is limited. We address this issue by splitting each day into six blocks of four hours, a duration that captures intraday dynamics (morning, noon, afternoon, evening, midnight, late night) and aligns well with common eight-hour shifts in police departments (Amendola et al. 2011).

In addition to this more realistic problem design, the main contribution of our work lies in furthering our understanding of how the pieces that constitute such a highly complex urban operational problem, such as prescriptive policing, beyond prediction—choice of data, choice and evaluation of vehicle assignment, aspects to include into the model, and those to exclude from it—fit together and affect each other. Although we rely on well-established prediction and assignment techniques, we show that the degrees of freedom these techniques offer already require decision makers to quantify potential trade-offs within the focal system and with respect to externalities caused on other systems and phenomena. Thereby, we illustrate a key challenge the practice of operations management in a smart city context faces and provide insights on how to overcome it.

3. Problem Description, Data, and Methodology

Before we describe the prediction, assignment, and evaluation problems of the prescriptive policing problem in detail, we briefly introduce the main data set used in this study. We obtained call-for-service data for the city of San Francisco from the authorities' official website (https://datasf.org/opendata/). The data spans the period between August 1, 2013, and September 30, 2013, and contains both criminal and noncriminal incidents, including a categorization, a short description, location, and time stamp. Table 1 shows the prevalence of different crime types. We performed standard data-cleaning procedures to the original data we obtained, such as removing corrupted or incomplete data points (e.g., missing time, location, or crime type). Furthermore, to determine the lagged values of crime densities during model training (see Section 4.1), we also collected crime data for the four weeks between July 4, 2013, and July 31, 2013.

We augment the crime data with a comprehensive set of geotagged Twitter messages (384,916 in total) from August and September 2013 to assess the benefit of such external data. As noted earlier, Gerber (2014) and Chen et al. (2015) find that geotagged social media activity can improve crime prediction performance. Specifically, we employ both the density of tweets in a given area as well as linguistic characteristics identified by the Linguistic Inquiry and Word Count (LIWC) software. LIWC analyzes each tweet along different linguistic dimensions, such as use of certain grammatical structures or valence according to different emotional categories (e.g., anger, joy, sadness). Additionally, given that Brandt et al. (2017) and Willing et al. (2017) demonstrate the value of points-of-interest (POIs) in explaining and predicting urban phenomena, we also consider data on 63,308 points of interest in San Francisco. The POI data contains location information as well as the type of POI according to a set of categories and was accessed through Google Maps.

3.1. Prediction Problem

The crime prediction problem is challenging along several dimensions. First, the occurrence of a crime at a given place and time is an extremely rare event. Considering common temporal and spatial resolutions, a crime is by several magnitudes more likely not to happen at a given place and time than otherwise. Second, the likelihood of a crime taking place at a given place and time is neither spatially nor temporally independent from its surroundings. In most cases, this likelihood would be identical to the likelihood of the same crime taking place at the same place, but several minutes later. Third, as shown in Table 1, the term "crime" encompasses a wide array of different offences with each following different and possibly contrary patterns. Hence, as has been suggested in prior work, it is useful to consider predictions for different crime types separately. Specifically, modeling each crime type independently allows deriving and acting upon predictive patterns that are idiosyncratic to a given crime type and thereby improves predictive accuracy.

For our main analysis, we focus on the two most frequent crime types: *larceny/theft* and *assault* (henceforth, we refer to larceny/theft simply as *theft*). There are multiple reasons for this choice. First, for both of these crime types, adequate allocation of police vehicles can have practical benefits: the presence of police forces has likely a deterring effect, or shorter response times can have a tangible benefit. Second, both are crime types that are prevalent, and thus, dealing more effectively with these crimes can be impactful.

Table 1. Selected Crime Types and Their Prevalence in the Data Set

Crime type	Count	Share, %
Larceny/theft	7,082	30.0
Assault	2,195	9.3
Vehicle theft	1,283	5.4
Vandalism	1,188	5.0
Drug/narcotic	1,118	4.7
Burglary	1,102	4.7
Robbery	785	3.3
Fraud	474	1.7
Stolen property	247	1.0
Weapon laws	239	1.0
Drunkenness	197	0.8
Forgery/counterfeiting	165	0.7
Trespass	150	0.6
Sex offenses, forcible	144	0.6
Others	7,747	30.8
Total	23,642	100

Thefts and assaults account for 40% of all reported incidents, and they represent harm to either a person's body or property. The relatively high number of incidents also reduces the impact of data sparsity as other, less frequent crimes would be even harder to predict. Third, this choice provides a theoretical reasoning on why social media data may be of value to the prediction of these crime types. High levels of social media activity may point to public events at a given place and time at which crowds may be targets for thefts; similarly, the valence of tweets in a given area may facilitate predictions of assaults. Fourth, the underlying dynamics of thefts and assaults are very different, which allows us to compare the impact of optimization on one or the other. Thefts are usually considered crimes requiring intent and/or opportunity, whereas assaults often result from sudden emotional arousal.

To address the spatiotemporal dependency, we follow the approach proposed by Gerber (2014) and transform point observations of crimes, tweets, and POIs into densities across the entire area through kernel density estimation. As a result, the occurrence of a crime is independent from the particular location and, instead, understood as reflective of the local vicinity. Figure 1 visualizes the resulting density surfaces for assaults and thefts for our data. The surface is created by transforming the observation area into a grid of 300×300 meter tiles and calculating the density at the centroid of each tile. It is worth emphasizing that the calculation does not only take occurrences within the focal tile into account, but rather all occurrences and weights them by distance. Consequently, these density values represent both the input and output of the prediction methods as we are seeking to predict future crime density distributions across the city. The choice of tile size aims to balance computational complexity and usefulness for the police car assignment task. In most cities, the chosen tile size would correspond to one or two blocks. However, our results are consistent for other tile sizes. The larger area with darker shades in Figure 1(b) reflects the difference in the number of incidents between thefts and assaults. Although the hot spot in downtown San Francisco (northeast) is common to both figures, the patterns in other parts of the city differ between the crime types. Several smaller theft hot spots are, for instance, not accompanied by an increase in local assault crimes. By transforming all variables into densities, spatial dependencies are made explicit, and the prediction models assess the relationships between the spatial patterns of the variables instead of individual occurrences (Willing et al. 2017).

From a temporal perspective, we distribute crimes into six blocks per day as previously discussed: 2 a.m. to 6 a.m. (late night), 6 a.m. to 10 a.m. (morning), 10 a.m. to 2 p.m. (noon), 2 p.m. to 6 p.m. (afternoon), 6 p.m. to 10 p.m. (evening), and 10 p.m. to 2 a.m. (midnight). This distribution achieves two objectives. First, the temporal dependency is taken into account, and crimes are perceived as representative of a more general notion of time of day. The blocks follow the natural course of the day, and it is reasonable to assume that crime patterns within a block are relatively consistent although differing across blocks. For example, there may be more criminal activity around bars within the midnight block compared with the morning block. Second, the four-hour blocks should also prove more useful in the practical vehicle assignment than full-day predictions as they cover intraday variation in hot spot locations and align well with the traditional eight-hour shifts of most police departments (Amendola et al. 2011).

When distributing discrete crime observations across these blocks, the result is an extremely imbalanced

Figure 1. (Color online) Density Surfaces of Crime Data

(a)

(b)
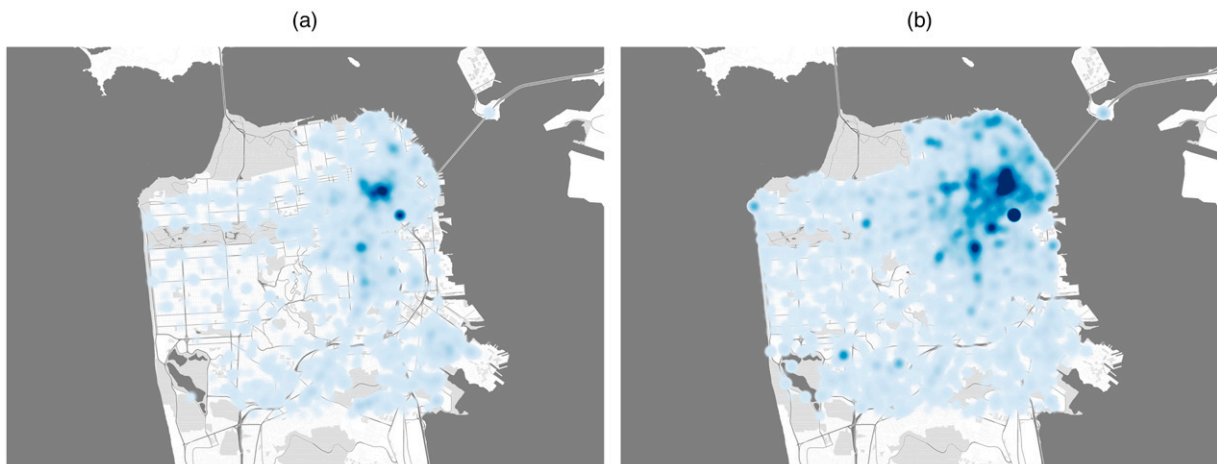
Notes. (a) Assaults. (b) Thefts.

data set with only a handful of "positive" observations taking place in a given time block. As a consequence, some measures of prediction quality, such as accuracy, are inappropriate (e.g., a naïve strategy that always predicts no crime to occur would have an accuracy of close to 100% and be hard to beat). Gerber (2014) proposes the area under the curve (AUC) of a surveillance plot as a more appropriate measure in this context. A surveillance plot is a function $s : [0,1] \to [0,1]$, mapping the share of area covered ("surveilled") to the share of crimes covered. The surveilled area is calculated by ranking the tiles according to the predicted density in a descending order as follows. Let L be the set of all tiles and let $L_y \subset L$ be the set of y tiles with the highest prediction value. Similarly, let C be the set of all crimes of the chosen category occurring during the time span considered and $C_y \subset C$ as the subset of those crimes that occur in the tiles contained in L_y. Then, $s(|L_y|/|L|) = |C_y|/|C|$. Effectively, the surveillance plot takes the share of tiles covered as input and returns the share of crimes covered. For other input values, the function s is obtained via linear interpolation. Note that the ordering of tiles is based on the prediction value, whereas $s(\cdot)$ is calculated based on the actual number of crimes per tile. Clearly, we have that $s(0) = 0$ and $s(1) = 1$.

The AUC is then obtained by calculating the area under the curve $s(\cdot)$ between zero and one, which can be calculated by taking the integral. However, as s is a piecewise linear function with $|L|$ pieces, this can also be calculated by summing over the different tiles:

$$AUC = \sum_{l=1}^{|L|} 0.5 \times \left(s\left(\frac{l-1}{|L|} \right) + s\left(\frac{l}{|L|} \right) \right) \times \frac{1}{|L|}.$$

Note that the AUC of a prediction method does only depend on the ordering of the tiles based on the prediction and not on the predicted value. Note further that the upper limit for the AUC, that is, the AUC corresponding to a perfect prediction, depends on the distribution of crimes over the tiles. For example, if each tile would have exactly one crime, the maximum AUC would be 0.5. However, because our crime data are very sparse, the maximum AUC approaches one.

Figure 2 offers some illustrative examples of surveillance plots. In Figure 2(a), we construct a simple example of three tiles (A, B, C) with one, three, and six crimes, respectively. Now, suppose that the prediction is reasonably good and that it yields density values of 0.2, 0.4, and 0.6, respectively. Arranging the tiles in descending order according to their density results in the sequence C, B, A. The corresponding surveillance plot in Figure 2(a) is, hence, defined by four points— (0.00, 0.00), (0.33, 0.60), (0.66, 0.90), and (1.00, 1.00)— and an AUC metric of 0.83. For comparison, random ordering would result in a score of 0.67 on average in this particular case. Figure 2(b) shows the surveillance

plot for a real-time slot from our data set with five assault crimes. This curve seems to be a step function, but this is not the case. Because of the small number of crimes and the large number of tiles, the curve is just very steep for tiles with a crime. Figure 2(c) aggregates the surveillance plots for all time slots, yielding a smooth curve. The AUC values in Figure 2, (b) and (c), are 0.88 and 0.85, respectively.

In summary, the prediction problem generally seeks to identify the prediction method that provides the highest AUC of the surveillance plot when predicting crime densities for the next four-hour block. However, as we are particularly focusing on the interdependencies between prediction, optimization, and evaluation, we tune models such that they produce both high-AUC predictions and low-AUC predictions. For this, we use three established methods, namely, multiple linear regression (MLR), random forest (RF), and gradient boosting (GB) (see Hastie et al. 2009 for detailed descriptions of these methods). As input data we use historical crime data, twitter data, and data on points of interest. For training the models, we use a rolling window of four weeks, implying that a model trained on data from August 1 to August 28 is used to predict assault and crime densities during the six time slots on August 29. AUC values are calculated using the actual crimes that took place during the time slot that was predicted.

3.2. Optimization Problem

The output of the prediction method is a density surface of predicted crimes during the next four-hour interval. This density surface is used to determine the optimal distribution of vehicles over the city. As discussed in Section 2, prescriptive policing can affect both the police force's ability to deter crime and its ability to quickly respond to a call for service. Regarding the first effect, people are unlikely to commit a crime within the immediate vicinity of a police officer or patrol; however, this influence diminishes relatively swiftly with increasing distance. Hence, if deterrence is the main objective, the goal is to find the distribution of vehicles that maximizes coverage, that is, the probability that the closest vehicle is within a (relatively short) predefined maximum distance from a crime. If the objective is based on the average response time, we find the distribution of vehicles such that the average distance between the closest vehicle and a crime is minimized, taking distance as a proxy for the time the police need to traverse it. Because of the overall rarity of crimes (on average, there are 0.013 thefts and 0.004 assaults per tile/time slot combination) the location of any given incident is always highly stochastic. Hence, in the optimization stage, the decision maker is bound to the density surfaces provided.

Figure 2. Illustrative Examples for Surveillance Plots
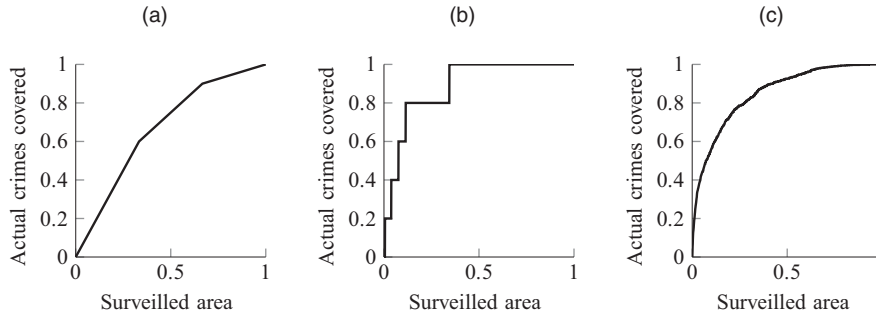
Notes. (a) Constructed example with three tiles and 10 crimes. (b) Single time slot and all tiles. (c) Aggregation over all time slots and tiles. Plots (b) and (c) are based on the multiple linear regression prediction method.

As we have previously mentioned, we transform the city area into a grid of 43×46 square tiles with an edge length of 300 meters each. With $L = \{(x, y) \mid x \in \{1, \ldots, 43\}, y \in \{1, \ldots, 46\}\}$ as the set of tiles, the symmetric distance between tile l and tile l' is denoted by $d(l, l')$. For each tile l, the crime density is predicted at its centroid and is denoted by δ_l. Note that some prediction methods could give a negative prediction for some tiles. To avoid having a negative reward for a short response time to these tiles, we set δ_l equal to zero for tiles with a negative prediction value. Regardless of the objective considered, the main decision is the assignment of police vehicles to tiles. We denote this decision by the binary decision variable v_l, which takes value one if a police vehicle is assigned to tile l and zero otherwise.

Based on the maximal covering location problem (MCLP) model introduced by Church and ReVelle (1974), we maximize the probability that the closest police vehicle is less than τ meters from a crime for the coverage objective. For this purpose, we introduce the binary decision variable w_l that takes value one if at least one police vehicle is within a distance τ from tile l. The resulting optimization problem can be formulated as the following integer linear programming problem:

$$\text{Maximize} \quad \sum_{l \in L} \delta_l \times w_l \tag{1}$$

$$\text{Subject to} \quad \sum_{l':d(l,l')\leq\tau} v_{l'} \geq w_l \qquad \forall l \in L \tag{2}$$

$$\sum_{l \in L} v_l \leq \bar{v}. \tag{3}$$

Although the average distance objective can be formulated based on the *p*-median problem introduced by ReVelle and Swain (1970), it is crucial that the optimization models run very quickly, both for the purpose of practical implementation and given the large number of instances that need to be solved in our numerical experiments. An initial exploration of the runtime shows that such a model is very slow because the full distance matrix needs to be incorporated. In contrast, for the coverage model, it is sufficient to know which pairs of tiles have a distance less than or equal to τ between them. As a result, only building the average distance model already takes more than an hour, whereas the coverage model can be built and solved in less than a second. Alternative solution methods do exist that could speed up the optimization compared with using commercial solvers (see, for example, Fischetti et al. 2016), but these methods do not give the speedup required for this study.

To overcome the excessive computation time requirements, we implement an alternative model as a proxy for the average distance model. Instead of evaluating every distance, we only evaluate the model at a prespecified number of values. This allows us to use a formulation that is similar to the MCLP formulation. Let $R = \{r_1, r_2, \ldots, r_{|R|}\}$ be the set of distances that is considered in the optimization and let a_r be the weight of distance target r in the objective. The model then maximizes the weighted average over the different distance targets. For this, we replace the decision variables w_l by $w_{l,r}$, indicating whether tile l is covered within a distance r:

$$\text{Maximize} \quad \sum_{r \in R} a_r \sum_{l \in L} \delta_l \times w_{l,r} \tag{4}$$

$$\text{Subject to} \quad \sum_{l':d(l,l')\leq r} v_{l'} \geq w_{l,r} \qquad \forall l \in L, r \in R \tag{5}$$

$$\sum_{l \in L} v_l \leq \bar{v}. \tag{6}$$

By choosing the distance targets appropriately, this model provides solutions with a small average distance. We set $a_{r_i} = r_{i+1} - r_i$ with $r_{|R|+1}$ being equal to the maximum distance between any two tiles, $\bar{d}$. This approach corresponds with rounding the distance for each tile up to the first distance in the set R. Thus, by increasing the number of distance values considered, the proxy model resembles the average distance model more closely. Naturally, however, increasing the

number of distances considered increases the computation time of the model.

3.3. Evaluation Problem

As improving both deterrence and response time are valid goals of prescriptive policing systems, we evaluate decision quality with respect to both, corresponding to the two optimization models introduced in Section 3.2. Effectively, we assume that vehicles are allocated to certain locations within the city and patrol within a certain radius around that location. Therefore, we calculate any response time values using the central location as this is where the vehicle is at a given time on average. Similarly, we assume that the vehicle has a deterring effect within a certain radius of this central location as a proxy of the deterrence effect along the patrol routes.

To assess the deterrence effect within a short distance surrounding a patrol car, we construct a decision quality measure Q_ω^1 that corresponds to optimization Problems (1)–(3). This measure reflects the share of crimes in the evaluation set (i.e., the time slots starting with August 29, 2013) that occur within a ω-vicinity of a police car. Formally, in a given time slot t, the binary variable v_l^t describes the allocation of vehicles to tiles $l \in L$. The set of assaults (respectively, thefts) is $C = \{c_1, c_2, \dots\}$, with each crime $c \in C$ being described by a location l_c and a time slot t_c. We define the coverage measure Q_ω^1 representing the deterrence effect as

$$Q_\omega^1 \equiv \frac{\sum_{c \in C} \gamma_c}{|C|}$$

with

$$\gamma_c = \begin{cases} 1, & \text{if } \min_{l \in L}\{d(l_c, l) \mid v_l^{t_c} = 1\} \leq \omega \\ 0, & \text{otherwise.} \end{cases}$$

In the numerator, we check for each crime c whether there is a police car at a distance less than ω from it given allocation V^t. Here, $d(\cdot)$ gives the distance between any two locations. Note that we use the center of the tile for the location of the police car, but the actual coordinates of the crime for the crime location. We divide the number of positive cases by the total number of crimes analyzed. In this study, we set ω to 500 meters.

The second measure for decision quality calculates the average distance between a crime and the closest police car. This measure corresponds to the objective of optimization Problems (4)–(6). We define the average minimum distance measure Q^2 as

$$Q^2 \equiv \frac{\sum_{c \in C} \min_{l \in L}\{d(l_c, l) \mid v_l^{t_c} = 1\}}{|C|}.$$

In the numerator, we calculate for each crime $c \in C$ the minimum distance from a tile with a police car at the given time to the crime location. We sum the resulting

distance for each crime and divide the result by the total number of crimes to get the average distance between a crime and the closest police vehicle.

4. Results

To gain a deeper understanding of the interdependencies between the problems we describe in the previous section, we first consider the determinants of prediction quality before proceeding to the alignment between prediction and decision quality for the different optimization objectives. Afterward, we present results on the trade-off between the two objectives and between the two crime types considered in this paper.

4.1. Prediction Quality

As a first step, we analyze the drivers of crime prediction quality. In Section 2, we outline that several papers find evidence for spatial patterns of social media activity improving full-day crime predictions. Similarly, points of interests are shown to be relevant predictors for other urban phenomena. In Table 2, we investigate whether these data sources affect prediction quality when increasing the temporal resolution of crime forecasts to four-hour time slots.

For this purpose, we focus on an MLR prediction of $\Delta_t = \{\delta_{l,t} : l \in L\}$, that is, the crime density (in this case, assaults) in all tiles in time slot t. The initial configuration exclusively relies on autoregressive features of crime density: the density values for the preceding time slot ($\delta_{l,t-1}$), the same time slot one day before ($\delta_{l,t-6}$), and the same time slot one week before ($\delta_{l,t-42}$) as well as a four-week moving average (*AvgCrime*). We then successively add various additional variables, namely, the tweet density in $t-1$ (*Tweets*), the POI density (assumed to be fixed over the observation period), and the density of tweets weighted by the share of negative (*NegEm*) and positive (*PosEm*)

Table 2. AUC Values for Varying Variable Input to Multiple Linear Regression Predicting Assaults

AvgCrime	$\delta_{l,t-42}$	$\delta_{l,t-6}$	$\delta_{l,t-1}$	Tweets	POIs	*NegEm*	*PosEm*	AUC
✓	✓	✓	✓					0.848
✓	✓	✓	✓	✓				0.848
✓	✓	✓	✓		✓			0.851
✓	✓	✓	✓	✓	✓			0.851
✓	✓	✓	✓			✓		0.849
✓	✓	✓	✓				✓	0.848
✓	✓	✓	✓	✓	✓	✓	✓	0.850

Notes. The predicted variable is the assault density in time slot t and tile l, $\delta_{l,t}$ for all tiles. Predictors are the average crime density during the preceding four weeks in that tile; the crime density from one week ago ($\delta_{l,t-42}$), one day ago ($\delta_{l,t-6}$), and one time slot ago ($\delta_{l,t-1}$); the tweet density from one time slot ago; the POI density (fixed); the negative sentiment in the tweets one time slot ago; and the positive sentiment in the tweets one time slot ago. AUC values are averaged over the observation period.

words, respectively. These shares are derived by the LIWC algorithm. The regression is trained using data from the first four weeks of the observation period as previously described. The resulting model is then used to predict the subsequent time slots using a rolling four-week window, and the AUC is calculated based on the surveillance plot.

Table 2 summarizes the average AUC, and it is clearly evident that the autoregressive components are the dominant drivers of prediction quality, at least for the MLR setting. Although adding tweets and POIs does result in some improvement, it does not exceed 0.2 percentage points for AUC values that are already quite high. Overall, the evidence suggests that social media and POI data does not substantially affect prediction quality when the temporal resolution is increased to four-hour blocks.

To gain a better understanding of the relationship between prediction quality and decision quality, we apply the random forest and gradient boosting predictors, tuning them such that they produce comparatively high and low AUC levels, respectively. By choosing an extremely low number of trees, the RF predictor performs substantially worse than the other predictors. The fact that it still outperforms the random predictor by a wide margin emphasizes that it still captures some of the predictive information—just not as well as the other configurations. The GB predictors are tuned to produce high-AUC results but fail to outperform the MLR prediction. As summarized in Table 3, this results in a set of weak predictors (the RF configurations) and a set of strong predictors (the MLR and GB configurations) that we use to analyze the relationship between prediction and decision quality going forward.

4.2. Comparing Prediction and Decision Quality

We first investigate whether the differences in the predictive quality between strong and weak predictors

translate into similar differences in the quality of decisions when those are based on predictions resulting from these models. In Figure 3, we summarize key results on the alignment between DQ and PQ. Specifically, we consider the decision quality metric—coverage or average distance—that was used in the optimization stage and present results for three different fleet sizes, namely, 5, 25, and 50 vehicles. For the coverage model, we set $\tau = \omega = 0.5$ km, and in the average distance model, we set $R = \{0.0, 0.5, 1.0, 1.5, 2.5, 5.0, 10.0$ km$\}$. We focus on assault cases; however, results for thefts are largely similar and presented in the appendix.

Across both metrics, we see that prediction and decision quality are generally well aligned although not exclusively so. When optimizing for coverage (plots a, c, and e), alignment implies that a high AUC is associated with high coverage values so that the best combinations would be found in the top right corner. Although this is the case for the scenarios with 25 and 50 vehicles, we observe that there is a slight misalignment for five vehicles because a combination that uses GB predictions with slightly lower AUC values results in higher coverage than those with the highest AUC values. The reason is likely that this lower AUC GB predictor excels at identifying the areas with the highest likelihood of assaults but performs worse than the other strong predictors in correctly ranking the remaining areas. Because the low number of available vehicles limits the share of crimes that can be covered at all, only a small share of the AUC actually matters, pushing the performance of predictive models that are otherwise inferior when considering the entire city. As vehicle numbers and the share of crimes that can be covered increase, this advantage is outstripped by the high-AUC models that predict well across the entire area.

It is noteworthy that the same low-AUC GB model that excels with respect to coverage falls behind when optimizing for average distance as shown in plots b, d, and f of Figure 3. We can also observe that PQ and

Table 3. AUC Values for Varying Prediction Methods and Configurations

Model ID	Method	AUC assaults	AUC thefts
1	Random	0.515	0.504
2	MLR	0.850	0.851
3	Random forest (5, 5)	0.753	0.785
4	Random forest (10, 5)	0.770	0.804
5	Random forest (5, 10)	0.752	0.787
6	Random forest (10, 10)	0.768	0.799
7	Gradient boosting (10,000, 0.01)	0.847	0.850
8	Gradient boosting (1,000, 0.01)	0.849	0.850
9	Gradient boosting (10,000, 0.001)	0.849	0.850
10	Gradient boosting (1,000, 0.001)	0.804	0.812

Notes. Random predictor assigns values from the interval [0, 1] randomly to tiles. The full set of variables from Table 2 is provided to Models 2–10 as input. Parameter configurations are listed as (*ntree, mtry*) for random forest models and (*ntree, shrinkage*) for gradient boosting models. Predictions were executed in *R* using the *randomForest* and *gbm* packages.

Figure 3. Alignment Between Prediction Quality and Decision Quality After Optimizing for the Respective DQ Metric (Assault)
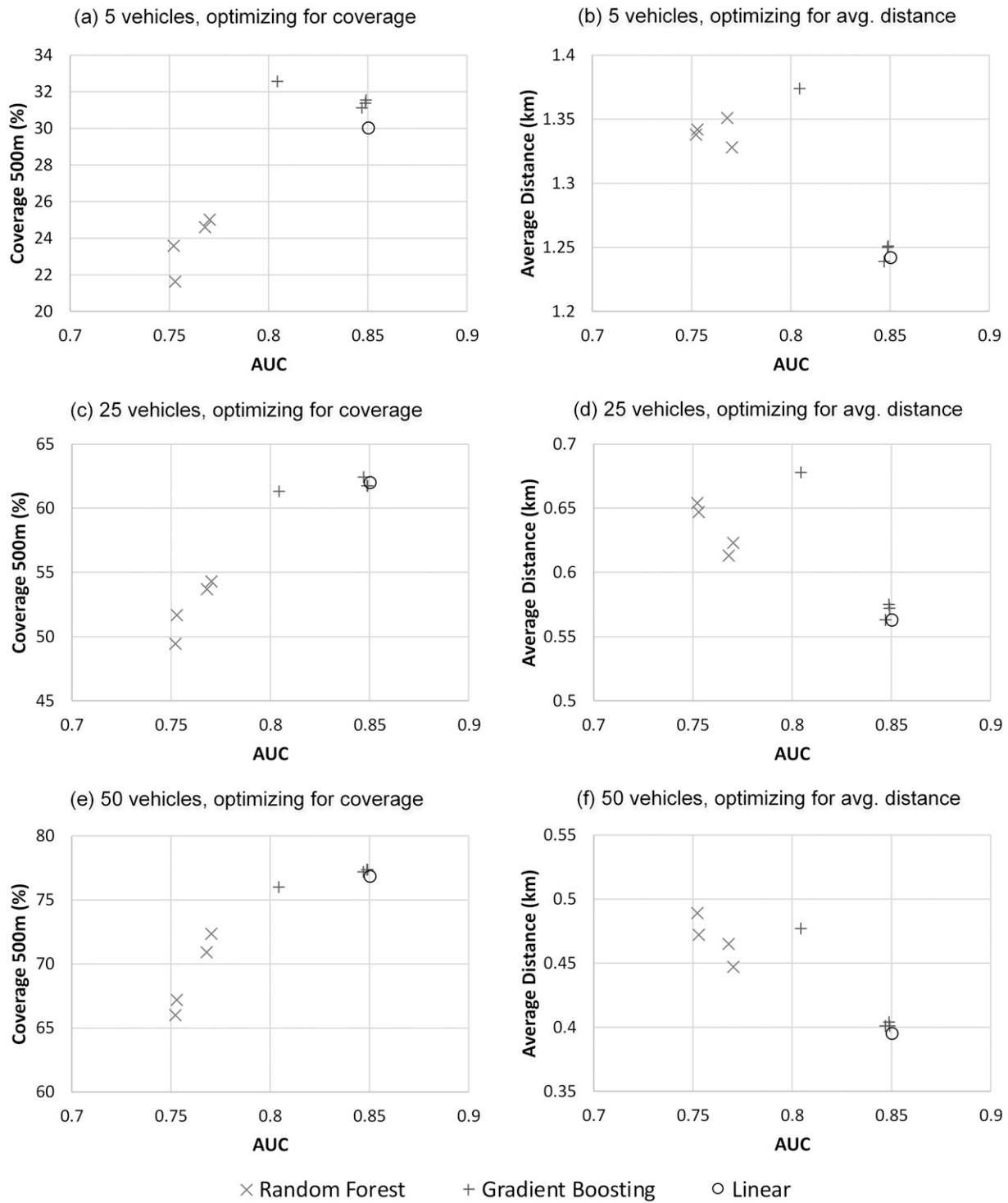
the Q^2 measure of average distance are consistently well aligned. We would expect a high AUC to be associated with low average distance values, placing the best combinations in the bottom right corner. This is indeed the case with the high-AUC GB and linear models also producing the lowest average distance for all vehicle fleet sizes.

4.3. Alignment of Decision Quality Metrics

Overall, these results illustrate that the alignment between prediction and decision quality in prescriptive policing approaches is not unambiguous even when focusing on a single crime category. The optimal combination of methods depends on various factors, including vehicle fleet size and the optimization objective

with a combination that is optimal for one DQ metric potentially being inferior for the other. In practice, it is unlikely that decision makers would focus exclusively on one single DQ metric while completely neglecting the other. Instead, police departments might seek to maximize the deterrence effect of patrol visibility in hot spot areas while also being aware of the need to keep a low response time on average (or the other way around).

When analyzing the trade-offs between these quality measures, the impact of the choice of prediction method becomes even more nuanced. Figure 4 illustrates the respective performance of each combination of prediction model and optimization objective with respect to both quality measures. Given that decision makers would aim for low average distance and high coverage, better outcomes would be located toward the top and left of the plots. At a first glance, the results in Figure 4 confirm expectations. Combinations that optimize for distance perform well along the distance-related DQ metric, and those that optimize for coverage perform well along the coverage metric. In

contrast, the performance of a given combination is generally lower when considering the opposite metric.

However, when considering the Pareto frontier among these solutions, that is, the combinations that do not allow for improvement along one metric without decreasing performance along the other, the weaker RF-based prediction models offer additional options in the trade-off between coverage and average distance when combined with coverage as the optimization objective. For low numbers of vehicles, the fact that the RF models are less tuned toward the hot spot density peaks leads to a more spread out distribution of vehicles, sacrificing coverage for a reduction in average distance. This effect is even more pronounced for thefts as illustrated in the appendix with the RF models providing a reasonable decrease in average distance for a slight decrease in coverage for fleet sizes of both 5 and 25 vehicles.

4.4. Alignment Across Crime Types
As we explain in Section 3.1, assault and theft are crimes that are driven by different factors, resulting in

Figure 4. (Color online) Trade-off Between Decision Quality Metrics for Different Combinations of Prediction Techniques and Optimization Objectives (Assault)
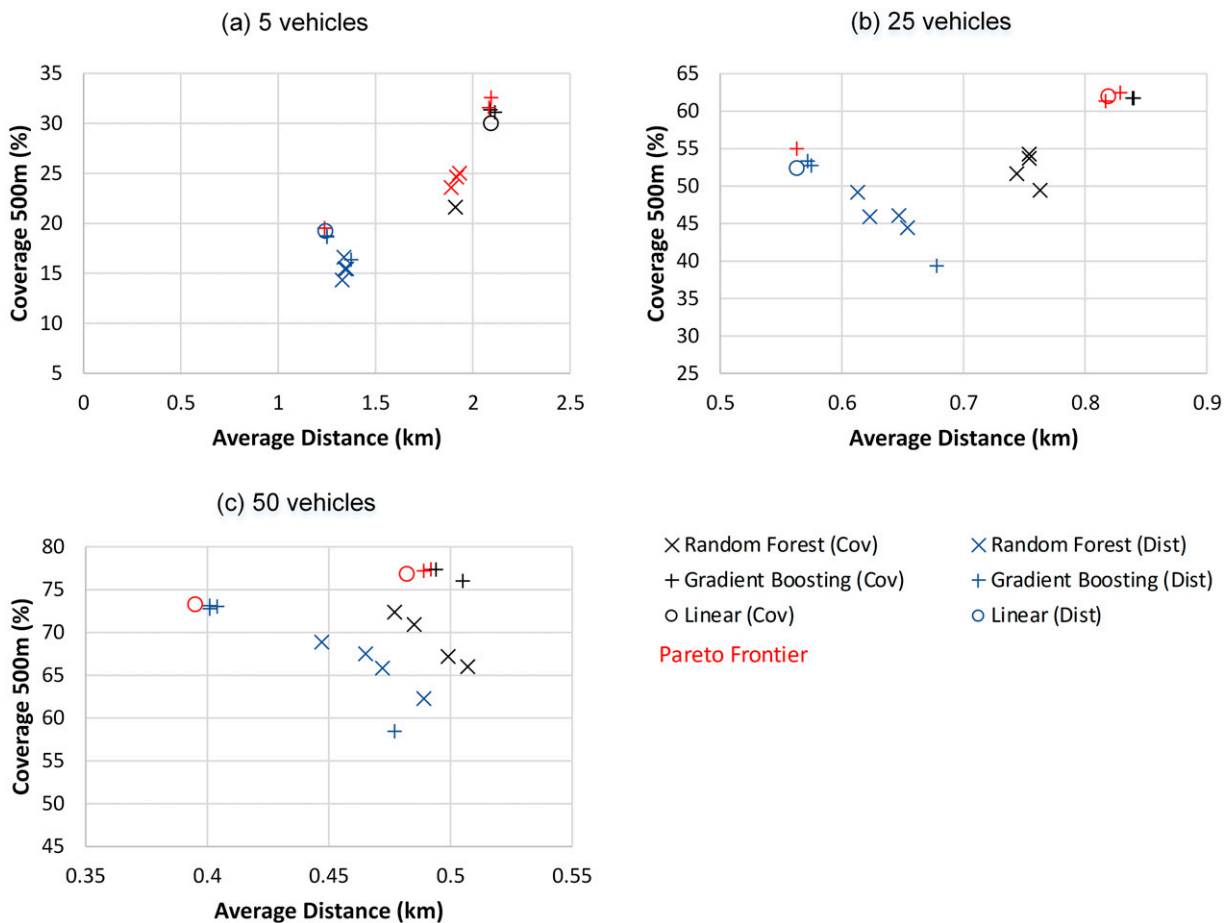
the different spatial patterns illustrated in Figure 1. As a consequence, although police patrols usually aim at handling a broad spectrum of crimes, optimizing for a particular crime type does not necessarily lead to optimal—or even good—allocations for other types. This problem is illustrated in Figure 5, in which we visualize the performance of the different configurations with respect to coverage of and distance to theft crimes. Specifically, we compare performance when vehicles are allocated based on theft predictions (see the appendix for details) to the performance when vehicles are allocated based on assault predictions. As expected, for configurations minimizing average distance, the assault-based performance is consistently worse than the theft-based performance across both DQ metrics. The effect for configurations maximizing coverage is a bit more nuanced. Although coverage of thefts decreases if vehicles are allocated based on assault predictions, the average distance actually decreases as well for the best performing configurations in the top right of the plot. The reason for this phenomenon is that, as evident in Figure 1, the spatial extent of the assault hot spot in San Francisco's downtown area is much smaller than the extent of the theft hot spot. As a consequence, fewer vehicles are required to cover this hot spot, leaving more available to be allocated to other parts of the city. Therefore, fewer of the thefts in the downtown area are covered, but the distance to thefts in the rest of the city is decreased.

In this example, optimization is biased extremely toward one crime type as vehicles are allocated solely based on assault predictions. We further explore the trade-off between different crime types by considering configurations that optimize based on a mixture of the predicted densities of both crimes. Figure 6 presents

the results for such a mixture in which assault and theft are given equal weight. Because the total number of theft cases is more than three times as high as the number of assaults, this implies that coverage of and distance to an assault are given a weight approximately three times as high as for a theft.

Compared with Figure 5, the performance decrease for thefts is now substantially reduced, reflecting the weight the predicted density of thefts now receives in the optimization stage. However, the fundamental patterns observed in Figure 5 persist with the configurations optimizing for distance performing slightly worse along both DQ metrics. For the configurations optimizing for coverage, we see again that coverage decreases, but so does average distance for the best performing models. When considering the effect on assaults in Figure 6(b), we can observe the opposite effect as the average distance for models optimizing coverage increases, and coverage also decreases. Similar to thefts, the effect on models optimizing average distance is only slight with marginal decreases in coverage and average distance for most configurations.

5. Discussion
The results presented in the previous section provide several insights regarding the nature, complexity, and potential solution approaches of the prescriptive policing problem. With respect to *prediction quality*, we can observe that, when considering four-hour prediction intervals, data on historical crime distributions represent the most important input by far. These autoregressive and moving average components dominate any potential influence of additional predictors, such as social media activity and points of interest, that have been used in other works (e.g., Gerber 2014, Chen et al. 2015). Although our results do not invalidate the

Figure 5. (Color online) Trade-off Between Decision Quality Metrics for Theft Crimes When Applying Assault-Based Vehicle Allocations (25 Vehicles)

13
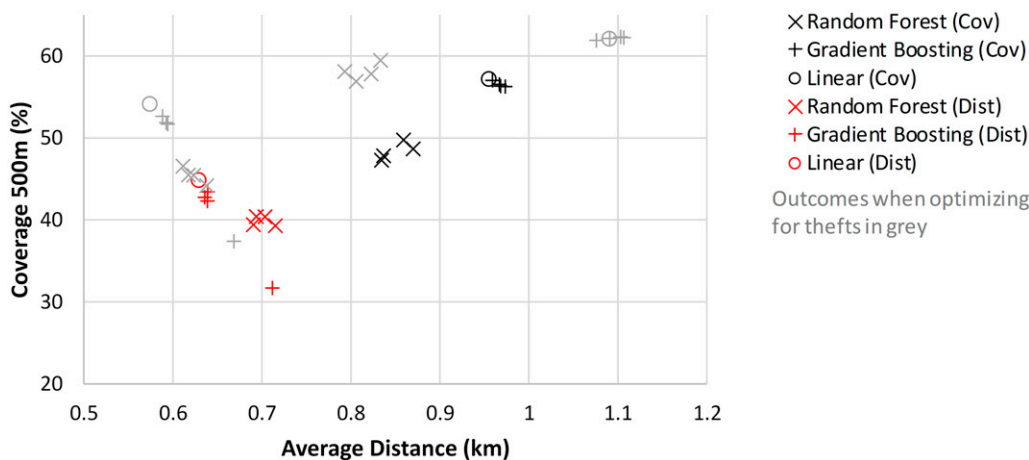
Figure 6. (Color online) Trade-off Between Decision Quality Metrics for Assault and Theft Crimes When Optimizing for a Mixture of Both Crime Types (25 Vehicles)
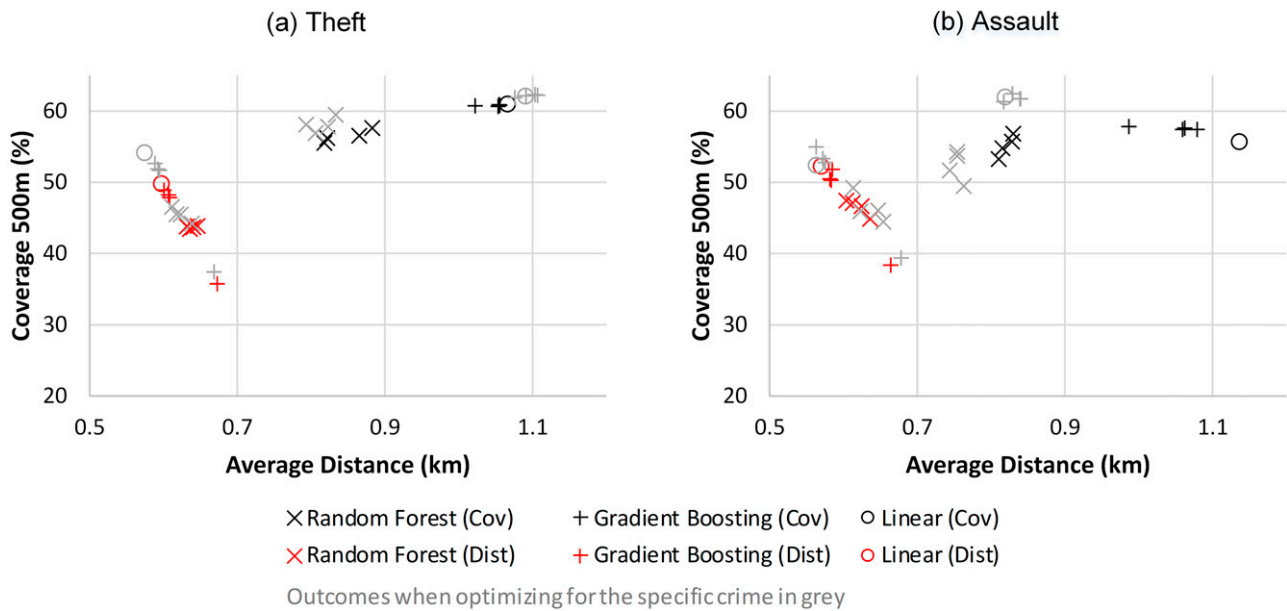
findings from these papers—the focus in them is generally on temporal resolutions exceeding four hours—they emphasize the importance of practical considerations in the context of prescriptive policing. As we have previously outlined four-hour blocks correspond well with common shift durations, whereas the actionable insights from full-day predictions are limited. The substantial impact of historical crime data on predictive quality also implies that police departments can rely on a generally readily available data source without needing to depend on external data providers. Furthermore, as we have previously discussed prescriptive policing is susceptible to biases induced by the data utilized. Focusing exclusively on crime data limits the introduction of latent biases as, for instance, socioeconomic and social media data would do. However, as Shapiro (2017) argues, using reported crimes as the data foundation may introduce a different bias as certain communities may be less likely to report a specific incident. These and similar effects need to be kept in mind when training and applying models for crime prediction in practice.

Our results further show that prediction and decision quality are generally aligned but not always. A key reason for the occasional divergence between PQ and DQ are the constraints of the optimization problem in addition to the stochastic nature of criminal incidents as a crime at a specific location and time is an extremely unlikely event. The AUC metric used to assess prediction quality evaluates the fit between the predicted crime density surface and the actual crimes that happened across the entire city. However, if only

a small number of vehicles is available, the share of the city that can be covered at all is limited. How a prediction model performs beyond this limit is irrelevant for the decision quality if the metric used is coverage. Integrating the constraint from the optimization problem into the prediction problem (Elmachtoub and Grigas 2020) by limiting the AUC calculation only to the share of the city that can theoretically be covered addresses this divergence, bringing PQ and DQ in line.

However, when we consider the alignment between multiple decision quality metrics, we can observe that the choice of prediction method becomes an important parameter irrespective of prediction quality outcomes. Although the strong predictors we consider achieve the best DQ when combined with the corresponding optimization objective, optimizing for coverage with weak predictor methods can produce allocations that also represent a Pareto-efficient trade-off between the DQ metrics (coverage and average distance) among the configurations we analyze. With low-PQ methods being less attuned to the city's hot spots, allocations based on them result in a more spread out distribution of vehicles, reducing average distance at the cost of coverage. Relaxing the resource constraint by increasing the available number of vehicles diminishes this phenomenon as a larger fleet size automatically results in a more spread out distribution once hot spots are covered.

Similar to the trade-off between multiple DQ metrics, we also find a clear trade-off when seeking to address multiple crimes. As different crime types are

driven by different factors, their spatial and temporal patterns of occurrence differ as well. Naturally, optimizing for one is likely to decrease performance metrics for the other as shown in our analysis of theft and assault crimes. However, even this effect is not unambiguous when both DQ metric are considered. Although optimizing for coverage decreases the coverage of thefts when assaults are used as (part of) the optimization input, average distance also decreases. The underlying dynamic is similar to the one for low-PQ methods as fewer vehicles are needed to cover assault hot spots than those of theft, leading to the remaining fleet being distributed in a more spread out fashion.

5.1. Implications for Prescriptive Policing and Smart City Operations

Returning to our initial considerations regarding the complexity of the focal problem and potential external effects on other systems and phenomena in the smart city context, we can perceive the variety of different crime types either as part of the problem or, as such, an external phenomenon. From a theoretical perspective, it is possible to assign weights to all crime types and integrate them into the optimization problem. As different crimes might be more or less susceptible to coverage or responsive to a short reaction time, weights might be further differentiated by target metric. However, for practical purposes, this approach is infeasible for several reasons. First, although such weights may be formulated for two crime types as in our study, explicitly articulating them for all relevant types is a nontrivial task. Although thefts and assaults account for 40% of observed crimes, there are several additional categories that both occur frequently and matter to vehicle allocation, such as vehicle theft, vandalism, burglary, and robbery. Together, these types account for another 20% of crimes. Deriving weights for them that are both transitive and reflect the potentially varying importance of the

deterrence and response time criteria for each type would be a time-consuming and subjective process. Second, any weights that are determined would likely need to vary between different time blocks as certain crimes are particularly prevalent during specific times of day. Prescriptive policing algorithms may be required to give them additional weight beyond the increased density. Similarly, seasonal effects, such as waves of certain crimes during particular times of the year, may require adjustments of the weights. Finally, in addition to being able to respond better to crimes and prevent some from happening at all, the application of prescriptive policing is likely to displace certain crimes to less-covered areas. This dynamic response to the prescriptive policing solution requires both regular updates using current data and, potentially, a weight shift as well.

Although these challenges may prevent the inclusion of all relevant crime types in the optimization process, it is still important to consider how these are affected by the prescriptive policing mechanism. For this purpose, we propose an approach that combines prediction, optimization, and exploration not just in the context of prescriptive policing, but as a general tool to navigate the complexity of smart city OM problems. Figure 7 summarizes this approach using our policing case as an example. Instead of attempting to optimize across all crimes, a small set of focal crime types is selected. Similar to the choice of theft and assault in our study, these types should matter (frequency and impact) and be affected by intelligent patrol placement. For these crime types, prediction and optimization is conducted for so-called *Pareto archetypes*, that is, the main combinations of prediction methods and optimization objectives defining the trade-off between coverage and average distance. In our examples, these would be the strong predictors combined with either objective function and the weak predictors combined with coverage maximization. The

Figure 7. (Color online) Predict–Optimize–Explore Approach for Prescriptive Policing
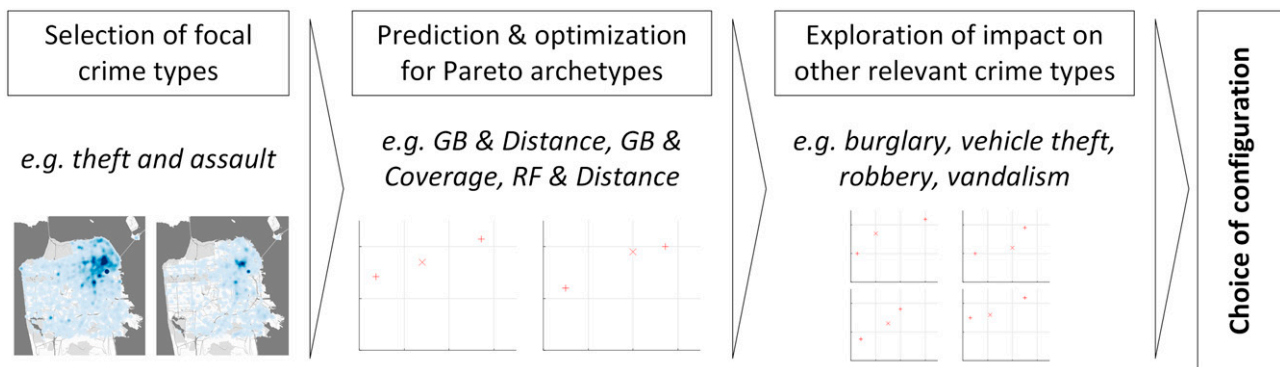
impact of the resulting allocations is then explored for other relevant crime types. Based upon this exploration, the choice regarding the configuration to be used is made.

Through this approach, the practical complexity of the prescriptive policing problem is substantially reduced. Instead of explicitly formulating the relative weights for each relevant crime type, decision makers only need to explicate relative importance for a small set of key types. For these types, a reduced set of prediction–optimization combinations is determined that offers different trade-offs between coverage and average distance for the key decision quality metrics used. The exploration stage then allows the inclusion of other crime types in a less formalized way to allow for a holistic perspective on the impact of the chosen method. However, exploration is not limited to analyzing the impact on other crime types and can similarly investigate the performance of the different configurations on other metrics and systems in the urban ecosystem. For instance, in light of the debate surrounding policing and racism in many countries, decision makers may want to analyze how certain configurations may produce outcomes that are biased against certain sociodemographic groups during the exploration stage. Thereby, the predict–optimize–explore approach offers a perspective on *fair AI* that not only sees human input as a potential source of bias (Feuerriegel et al. 2020), but also as a potential remedy to support the identification of discriminating algorithmic decisions.

5.2. Limitations and Future Work

These considerations are particularly relevant as all models are trained on historical data—data that may already reflect certain biased operational policies. Even absent explicit bias, patrols are likely to have exerted deterrence and displacement effects in the areas they were patrolling. Although difficult to access, data on patrol allocations can potentially be used to address these issues.

Our study focuses on analyzing the interactions between the separate operational challenges that constitute a prescriptive policing approach and how the practical implementation of such smart city solutions affects the urban ecosystem. We use established prediction and optimization methods to address the individual challenges, and there are certainly opportunities for future work to further refine these algorithms or explore the value of adding additional data sources. For instance, in the calculation of the average response time metric, we assume that the vehicle closest to a crime is actually available and not busy with handling another incident. Given that, on average, there are about 25 thefts or assaults in total per four-hour time slot in the city, the impact of this assumption is likely small for vehicle fleets of 25 or more cars. Nevertheless,

delving deeper into this availability problem is a promising path for future research.

6. Conclusion

Modern cities represent confluences of multiple complex subsystems, and data-driven approaches provide a new set of tools to understand and steer these systems. Focusing on the case of urban policing, we show that such a single application still consists of multiple interlinked problems with various trade-offs that affect the application's effectiveness.

Fundamentally, we derive four key lessons for practical applications of prescriptive policing. First, when considering actionable prediction intervals (e.g., four hours), historical crime data are—from the data sources we analyzed—by far, the most dominant and important. Second, when seeking to maximize coverage, it is important to consider the police department's resource constraint, that is, the number of available vehicles, in the assessment of prediction quality. Third, weaker prediction methods may still provide a viable trade-off between the deterrence effect of policing and response time. Fourth, it is important to consider how optimizing for one or a few types of crimes affects quality measures for other relevant crime types.

For the practical implementation of smart city solutions that affect multiple urban systems or phenomena, such as the variety of different crime types in the policing context, we propose an approach that combines prediction and optimization with an exploration stage. By focusing on a small subset of important crime types, problem complexity is reduced, and a set of candidate solutions offering varying trade-offs between deterrence and response time are derived. During exploration, the impact of these solutions on other relevant crime types is analyzed. As a result, a holistic impact assessment can be made without having to explicitly weight all relevant crime types against each other. Such an exploration stage can also prove useful in the context of other smart city operational challenges affecting multiple systems, such as the transition to electric mobility.
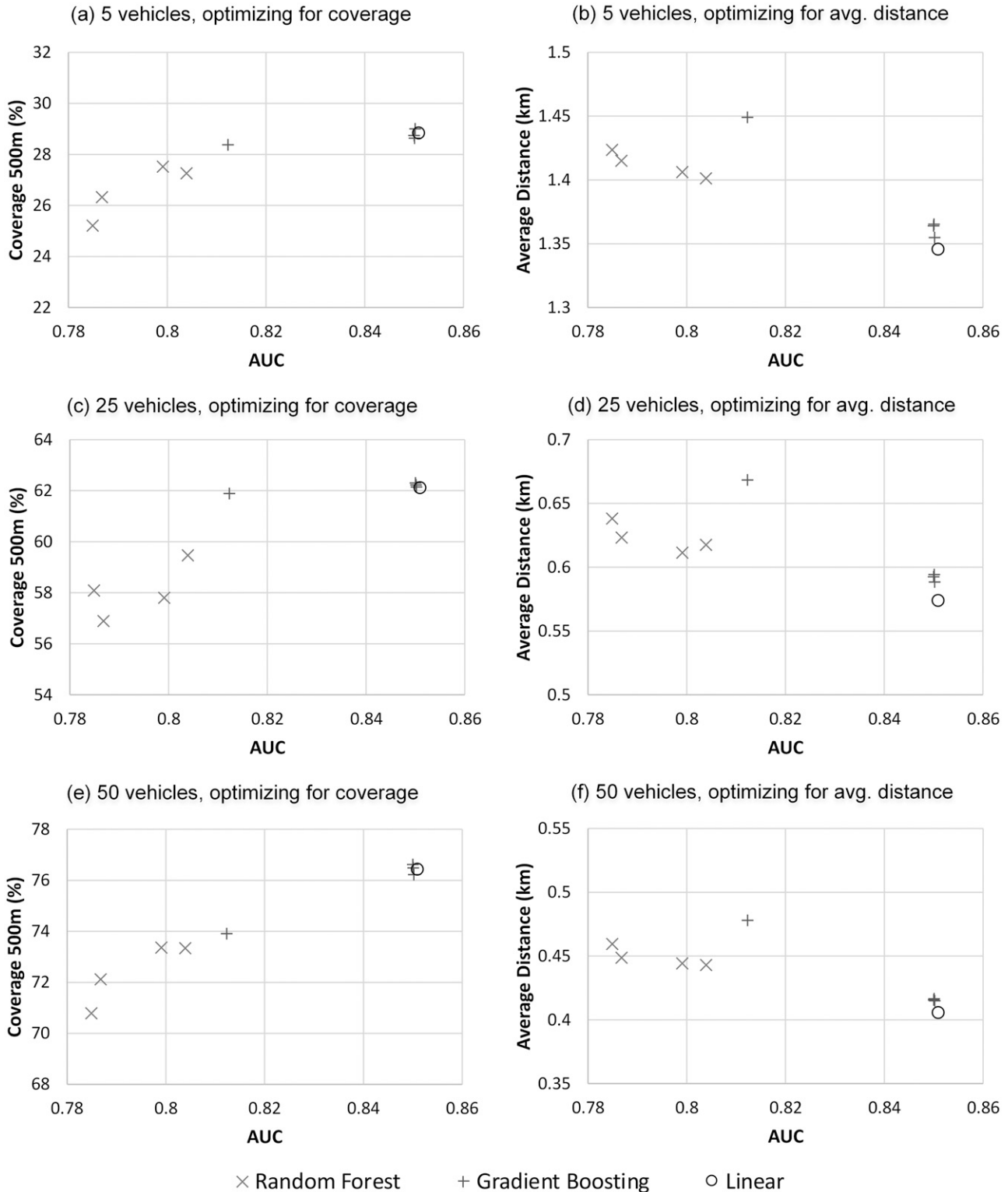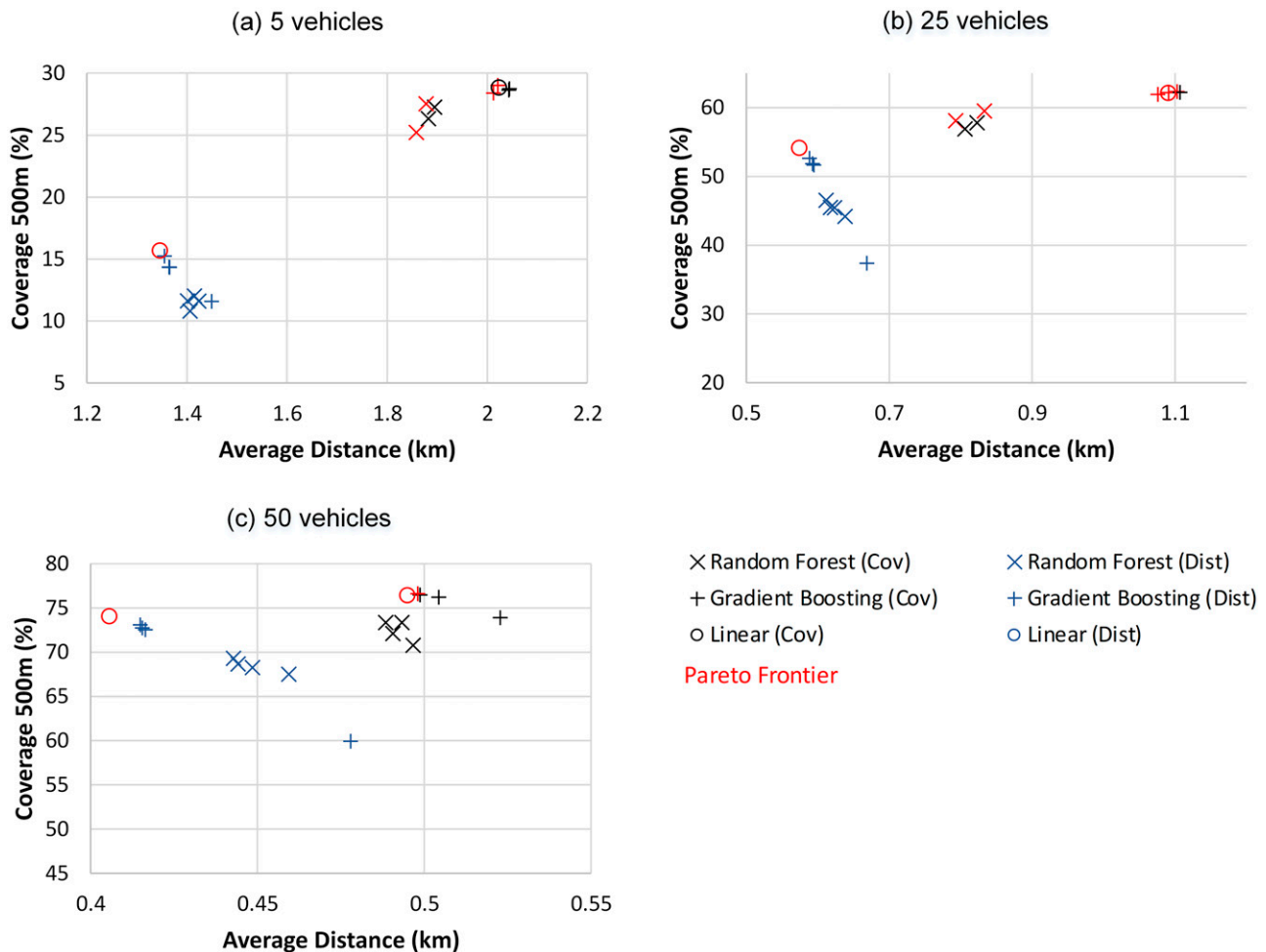
Appendix. Prediction and Decision Quality Analysis for Thefts

Figure A.1. Alignment Between Prediction Quality and Decision Quality After Optimizing for the Respective DQ Metric (Theft)

Notes. Figure A.1 illustrates the alignment between PQ and the two DQ metrics for thefts. Alignment is strong across all combinations of fleet size and optimization objective with the highest AUC values leading to the highest coverage values when optimizing for coverage (plots a, c, and e) and to the lowest average distances when optimizing for average distance (plots b, d, and f).

Figure A.2. (Color online) Trade-off Between Decision Quality Metrics for Different Combinations of Prediction Techniques and Optimization Objectives (Theft)

(a) 5 vehicles

(b) 25 vehicles

(c) 50 vehicles

✕ Random Forest (Cov)   ✕ Random Forest (Dist)
+ Gradient Boosting (Cov)   + Gradient Boosting (Dist)
○ Linear (Cov)   ○ Linear (Dist)
Pareto Frontier

Notes. Figure A.2 illustrates the trade-off between the two DQ metrics for thefts. Similar to the results for assaults, we observe that configurations with weaker prediction methods (random forest) still constitute part of the Pareto frontier for fleet sizes of 5 and 25 vehicles.

References

Al Boni M, Gerber MS (2016a) Area-specific crime prediction models. *15th IEEE Internat. Conf. Machine Learn. Appl.* (IEEE, Piscataway, NJ), 671–676.

Al Boni M, Gerber MS (2016b) Automatic optimization of localized kernel density estimation for hotspot policing. *15th IEEE Internat. Conf. Machine Learn. Appl.* (IEEE, Piscataway, NJ), 32–38.

Amendola KL, Weisburd D, Hamilton EE, Jones G, Slipka M (2011) An experimental study of compressed work schedules in policing: Advantages and disadvantages of various shift lengths. *J. Experiment. Criminology* 7(4):407–442.

Angalakudati M, Balwani S, Calzada J, Chatterjee B, Perakis G, Raad N, Uichanco J (2014) Business analytics for flexible resource allocation under random emergencies. *Management Sci.* 60(6):1552–1573.

Bertsimas D, Kallus N (2019) From predictive to prescriptive analytics. *Management Sci.* 66(3):1025–1044.

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J. Machine Learn. Res.* 3:993–1022.

Blevins T, Kwiatkowski R, MacBeth J, McKeown K, Patton D, Rambow O (2016) Automatically processing tweets from gang-involved youth: Toward detecting loss and aggression.

Matsumoto Y, Prasad R, eds. *26th Internat. Conf. Comput. Linguistics: Technical Papers* (The COLING 2016 Organizing Committee, Osaka, Japan), 2196–2206.

Braga AA, Papachristos AV, Hureau DM (2014) The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quart.* 31(4):633–663.

Brandt T, Bendler J, Neumann D (2017) Social media analytics and value creation in urban smart tourism ecosystems. *Inform. Management* 54(6):703–713.

Chaiken JM, Dormont P (1978) A patrol car allocation model: Capabilities and algorithms. *Management Sci.* 24(12):1291–1300.

Chainey S, Tompson L, Uhlig S (2008) The utility of hotspot mapping for predicting spatial patterns of crime. *Security J.* 21(1–2):4–28.

Chelst K (1978) An algorithm for deploying a crime directed (tactical) patrol force. *Management Sci.* 24(12):1314–1327.

Chen X, Cho Y, Jang SY (2015) Crime prediction using Twitter sentiment and weather. *Systems Inform. Engrg. Design Sympos.* (IEEE, Piscataway, NJ), 63–68.

Chen P, Yuan H, Shu X (2008) Forecasting crime using the ARIMA model. Ma J, Yin Y, Yu J, Zhou S, eds. *Fifth Internat. Conf. Fuzzy Systems Knowledge Discovery* (IEEE, Piscataway, NJ), 627–630.

18

Church R, ReVelle CR (1974) The maximal covering location problem. *Papers Regional Sci.* 32(1):101–118.

Di Tella R, Schargrodsky E (2004) Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *Amer. Econom. Rev.* 94(1):115–133.

Dunnett S, Leigh J, Jackson L (2019) Optimising police dispatch for incident response in real time. *J. Oper. Res. Soc.* 70(2):269–279.

Elmachtoub AN, Grigas P (2020) Smart "predict, then optimize". *Management Sci.*, ePub ahead of print, https://doi.org/10.1287/mnsc.2020.3922.

Ferguson AG (2017) Policing predictive policing. *Washington Univ. Law Rev.* 94(5):1109–1190.

Feuerriegel S, Dolata M, Schwabe G (2020) Fair AI. *Bus. Inform. Systems Engrg.* 62(4):379–384.

Fischetti M, Ljubić I, Sinnl M (2016) Benders decomposition without separability: A computational study for capacitated facility location problems. *Eur. J. Oper. Res.* 253(3):557–569.

Gerber MS (2014) Predicting crime using Twitter and kernel density estimation. *Decision Support Systems* 61:115–125.

Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning* (Springer, New York).

Kang HW, Kang HB (2017) Prediction of crime occurrence from multi-modal data using deep learning. *PLOS One* 12(4): e0176244.

Leigh J, Dunnett S, Jackson L (2019) Predictive police patrolling to target hotspots and cover response demand. *Ann. Oper. Res.* 283(1–2):395–410.

Mak A (2020) What three cities are spending on police compared with everything else. *Slate* (June 19), accessed September 18, 2020, https://slate.com/news-and-politics/2020/06/what-los-angeles-minneapolis-dallas-police-spend.html.

Metz C, Satariano A (2020) An algorithm that grants freedom, or takes it away. *New York Times Online* (February 6), accessed September 18, 2020, https://www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html.

Nair HS, Misra S, Hornbuckle WJ, Mishra R, Acharya A (2017) Big data and marketing analytics in gaming: Combining empirical models and field experimentation. *Marketing Sci.* 36(5):699–725.

Neill DB, Gorr WL (2007) Detecting and preventing emerging epidemics of crime. *Adv. Disease Surveillance* 4(13).

Qi W, Shen ZJM (2019) A smart-city scope of operations management. *Production Oper. Management* 28(2):393–406.

ReVelle CS, Swain RW (1970) Central facilities location. *Geographical Anal.* 2(1):30–42.

Ristea A, Langford C, Leitner M (2017) Relationships between crime and Twitter activity around stadiums. *25th Internat. Conf. Geoinformatics* (IEEE, Piscataway, NJ), 1–5.

Sassaman H (2020) Covid-19 proves it's time to abolish 'predictive' policing algorithms. *Wired Online* (August 27), accessed September 18, 2020, https://www.wired.com/story/covid-19-proves-its-time-to-abolish-predictive-policing-algorithms/.

Shapiro A (2017) Reform predictive policing. *Nature* 541(7638): 458–460.

Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice Quart.* 12(4):625–648.

Taylor PE, Huxley SJ (1989) A break from tradition for the San Francisco police: Patrol officer scheduling using an optimization-based decision support system. *Interfaces* 19(1):4–24.

Valogianni K, Ketter W, Collins J, Zhdanov D (2020) Sustainable electric vehicle charging using adaptive pricing. *Production Oper. Management* 29(6):1550–1572.

Wang M, Gerber MS (2015) Using Twitter for next-place prediction, with an application to crime prediction. *IEEE Sympos. Ser. Computat. Intelligence* (IEEE, Piscataway, NJ), 941–948.

Wilder B, Dilkina B, Tambe M (2019) Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. *Proc. AAAI Conf. Artificial Intelligence*, vol. 33 (AAAI Press, Palo Alto, CA), 1658–1665.

Williams ML, Burnap P, Sloan L (2016) Crime sensing with big data: The affordances and limitations of using open source communications to estimate crime patterns. *British J. Criminology* 57(2):320–340.

Willing C, Klemmer K, Brandt T, Neumann D (2017) Moving in time and space—Location intelligence for carsharing decision support. *Decision Support Systems* 99:75–85.