Supplementary Material

1 Robust Survival Modeling

In optimization problem 10, the attacker considers all potential pairs of incidents that could occur in each cell due to possible manipulation, and chooses incidents such that the overall likelihood of the model is minimized. The product term in the objective function ensures that likelihood is captured only for decisions made by the attacker, and not for all the possible options that the attacker has. The attacker's action space in terms of which cells to move the crimes to is thus represented by the binary decision variables y, and constraint (10b) ensures that the attacker can shift each incident to only one cell (since the same crime cannot be committed at the same time at two different locations).

2 Data Description

We provide a detailed description of the features we use in our predictive models here.

2.1 Feature set for poaching data

We use the following set of features in predictive models for poaching data.

- **Spatial features** We used features that modeled distances to the closest national park boundary, water, road, town, patrol post, mineral lick location, and the cost to reach to the closest village.
- **Geographic terrain** We used data to model habitat type, elevation, slope, topographic wetness index and NPP (net primary productivity) of cells.
- **Animal density** We used historical data to account for animal presence in each of the cells. This included animals like buffalo, elephant, hippopotamus, giraffe, kob, oribi, warthog, and waterbuck.

2.2 Feature set for burglary

We used the following covariates for predictive models using urban crime data.

- **Temporal cycles** We use a binary feature for weekdays and weekends. In order to look at the effect of time of day on incidents, we split each day into six zones of four hours each, and captured these by binary features.
- **Temporal and spatial incident correlation** For each cell, we looked at the past incident counts in the last week and month in it as well as neighboring cells as features to capture the effect of temporal and spatial correlation among incidents. We also treated the number of past incidents in each severity category as a feature while predicting incident severity, and considered the long-term effect of temporal correlation by looking at the average number of incidents in the past year.
- **Weather** We included a collection of features, such as rainfall, snowfall, and mean temperature to capture this effect. Weather data was collected from a weather station located in roughly the center of the county (we suppress an exact citation for blind review).
- **Risk-terrain features** We used several risk-terrain features to aid to our prediction model, including population density, housing density, and mean household income at a census tracts level. We also used data with 624 retail shops that sell liquor, 2494 liquor outlets, 41 homeless shelters, and 52 pawn shops.

3 Setup

We ran experiments for Poisson regression and survival analysis on a 2.4GHz 32-core Ubuntu Linux machine

with 32 GB RAM. The experiments for logistic regression (due to confidentiality agreements on data) had to be performed on centers approved by the national park, and were run on 3.2Ghz 6-core Ubuntu Linux machine with 16GB RAM. While evaulating the algorithm *RSALA*, we solved the optimization problem (5) using CVXPY. The only hyper-parameters in our model are the learning rate α for *AdGrad* and the attacker's definition of "neighboring grids" to which the criminals can move. We performed experiments using multiple learning rates, and found that a learning rate of 0.001 as the best parameter value.

4 Discussion

Before we conclude, we identify limitations and provide a few words of caution. First, it is important to note that while our optimization framework only accounts for spatial shifts, temporal shifts are also possible in response to learned models. It is natural to assume that attackers can shift their actions in time, by waiting to commit a crime at a particular location (or by committing it early). Through this paper, we establish the foundation of robustness in case of spatial-temporal predictions, and will address the issue of temporal shifts in future work. Next, we point out some insights about the experimental findings on non-adversarial data. A slight underestimation of frequency in high-density cells is not a drawback in itself; after all, robustness is gained so that the defender can account for adversarial manipulations. However, this raises an important question. First, how does one choose the extent of adversarial manipulation? This issue is important to address, but outside the current scope of our research. In adversarial models, the extent of spatial shifts should be chosen based on the specific problem and the type of incident at hand.