Reproducible research: assessing spatial predictions of crime

Matthew Daws

Leeds

LIDA Seminar, November 2017

Project

UK Home Office Police Innovation Fund: “More with Less: Authentic Implementation of Evidence-Based Predictive Patrol Plans”. With Andy Evans and Monsuru Adepeju here at Leeds.

My task:

- Take crime prediction algorithms from the literature, and implement in an open source way (`https://github.com/QuantCrimAtLeeds/PredictCode/`)
- Allow other researchers to see what benefit different crime prediction algorithms are likely to give.
- My background is in Mathematics; and Software Development.
- Runs until February 2018.

Project

UK Home Office Police Innovation Fund: "More with Less: Authentic Implementation of Evidence-Based Predictive Patrol Plans". With Andy Evans and Monsuru Adepeju here at Leeds.

My task:

- Take crime prediction algorithms from the literature, and implement in an open source way (`https://github.com/QuantCrimAtLeeds/PredictCode/`)
- Allow other researchers to see what benefit different crime prediction algorithms are likely to give.
- My background is in Mathematics; and Software Development.
- Runs until February 2018.

Project

UK Home Office Police Innovation Fund: "More with Less: Authentic Implementation of Evidence-Based Predictive Patrol Plans". With Andy Evans and Monsuru Adepeju here at Leeds.

My task:

- Take crime prediction algorithms from the literature, and implement in an open source way (`https://github.com/QuantCrimAtLeeds/PredictCode/`)
- Allow other researchers to see what benefit different crime prediction algorithms are likely to give.
- My background is in Mathematics; and Software Development.
- Runs until February 2018.

Project

UK Home Office Police Innovation Fund: “More with Less: Authentic Implementation of Evidence-Based Predictive Patrol Plans”. With Andy Evans and Monsuru Adepeju here at Leeds.

My task:

- Take crime prediction algorithms from the literature, and implement in an open source way (`https://github.com/QuantCrimAtLeeds/PredictCode/`)
- Allow other researchers to see what benefit different crime prediction algorithms are likely to give.
- My background is in Mathematics; and Software Development.
- Runs until February 2018.

The (Near-)repeat hypothesis

"The tendency of victims of crime to, in the nearby future, be repeat victims; and of near-by (say) buildings to also be future victims."

(Principally interested in Burglary.)

That is, a crime event at a spatial/temporal location tends to imply a higher risk, localised in space and time, for nearby locations.

- Classical prediction techniques tend to generate "hot spots" around previous locations.
- Part I: How do we do this? (Plea for reproducible research.)
- Part II: And what do we mean by "prediction" anyway? What makes a "good" prediciton?

The (Near-)repeat hypothesis

"The tendency of victims of crime to, in the nearby future, be repeat victims; and of near-by (say) buildings to also be future victims."
(Principally interested in Burglary.)
That is, a crime event at a spatial/temporal location tends to imply a higher risk, localised in space and time, for nearby locations.

- Classical prediction techniques tend to generate "hot spots" around previous locations.
- Part I: How do we do this? (Plea for reproducible research.)
- Part II: And what do we mean by "prediction" anyway? What makes a "good" prediciton?

The (Near-)repeat hypothesis

"The tendency of victims of crime to, in the nearby future, be repeat victims; and of near-by (say) buildings to also be future victims."
(Principally interested in Burglary.)
That is, a crime event at a spatial/temporal location tends to imply a higher risk, localised in space and time, for nearby locations.

- Classical prediction techniques tend to generate "hot spots" around previous locations.
- Part I: How do we do this? (Plea for reproducible research.)
- Part II: And what do we mean by "prediction" anyway? What makes a "good" prediciton?

The (Near-)repeat hypothesis

"The tendency of victims of crime to, in the nearby future, be repeat victims; and of near-by (say) buildings to also be future victims."
(Principally interested in Burglary.)
That is, a crime event at a spatial/temporal location tends to imply a higher risk, localised in space and time, for nearby locations.

- Classical prediction techniques tend to generate "hot spots" around previous locations.
- Part I: How do we do this? (Plea for reproducible research.)
- Part II: And what do we mean by "prediction" anyway? What makes a "good" prediciton?

The (Near-)repeat hypothesis

"The tendency of victims of crime to, in the nearby future, be repeat victims; and of near-by (say) buildings to also be future victims."
(Principally interested in Burglary.)
That is, a crime event at a spatial/temporal location tends to imply a higher risk, localised in space and time, for nearby locations.

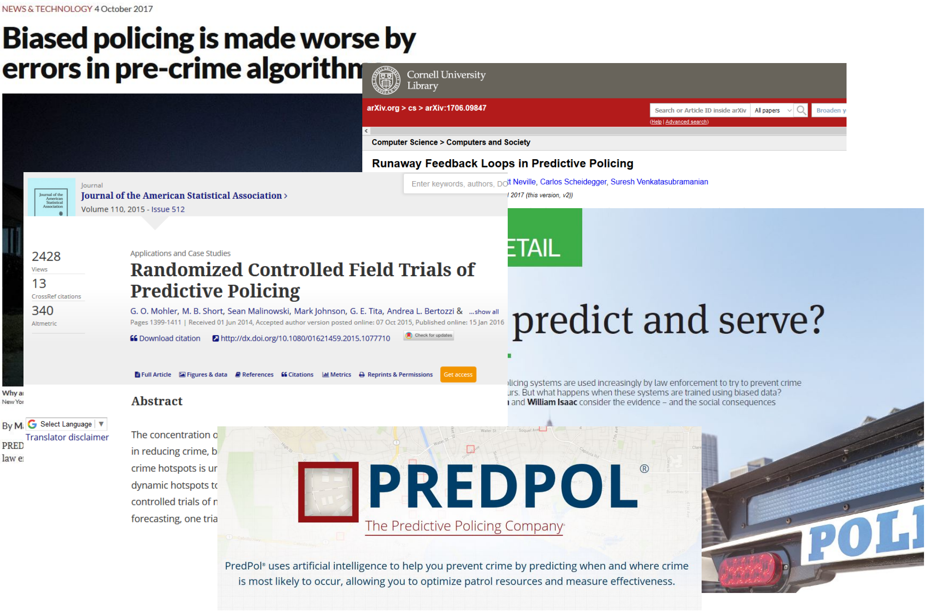
- Classical prediction techniques tend to generate "hot spots" around previous locations.
- Part I: How do we do this? (Plea for reproducible research.)
- Part II: And what do we mean by "prediction" anyway? What makes a "good" prediciton?

The (Near-)repeat hypothesis

"The tendency of victims of crime to, in the nearby future, be repeat victims; and of near-by (say) buildings to also be future victims."
(Principally interested in Burglary.)
That is, a crime event at a spatial/temporal location tends to imply a higher risk, localised in space and time, for nearby locations.

- Classical prediction techniques tend to generate "hot spots" around previous locations.
- Part I: How do we do this? (Plea for reproducible research.)
- Part II: And what do we mean by "prediction" anyway? What makes a "good" prediciton?

Publications



The algorithm

We start our investigation with an overview of PREDPOL. PREDPOL [8] assumes that crimes follow an earthquake aftershock model, so that regions that previously experienced crime are likely to experience crime again, with some decay. Mohler et al. [8] model the crime rate $\lambda_r(t)$ in region r at time t as follows: $\lambda_r(t) = \mu_r + {}_{t_r^i<t}\,\theta\omega e^{-\omega(t-t_r^i)}$ where t_n^i represents the time of an event in region r, ω quantifies the time decay of a shock, and θ captures the degree to which aftershocks are generated from an initial event. They use an expectation-maximization procedure to determine the parameters of the model.

2.2 Epidemic-Type Aftershock Sequence Model for Crime Prediction

Building on a foundation of reaction-diffusion models of crime (Short et al. 2010), we treat the dynamic occurrence of crime as a continuous time, discrete space ETAS point process (Marsan and Lengline 2008; Mohler et al. 2011; Mohler 2014). Policing areas were first discretized into 150×150 m square boxes. The conditional intensity, or probabilistic rate $\lambda_n(t)$ of events in box n at time t was determined by

$$\lambda_n(t) = \mu_n + \sum_{t_n^i<t} \theta\omega e^{-\omega(t-t_n^i)}, \tag{1}$$

where t_n^i are the times of events in box n in the history of the process. The ETAS model has two components, one modeling place-based environmental conditions that are constant in time and the other modeling dynamic changes in risk. Rather than modeling fixed environmental characteristics of a hotspot explicitly using census data or locations of crime attractors, long-term hotspots are estimated from the events themselves. In particular, the background rate μ is a nonparametric histogram estimate of a stationary Poisson process (Marsan and Lengline 2008). If over the past 365 days a grid cell has a high crime volume, the estimate of μ will be large in that grid cell. The size of the grid cells on which μ is defined can be estimated by maximum likelihood and in general the optimum size of the grid cell will decrease with increasing data. However, for a fixed area flagged for patrol, a greater number of small hotspots are more difficult to patrol than a small number of large hotspots. The 150×150 m hotspots were chosen in this study to be the size of a city block in Foothill and were then held constant across all of the experimental regions. The number of days of data used as input to the ETAS model, 365 days, was also chosen subjectively, though is consistent with other hotspot policing studies that use 1–2 years of data to select hotspots.

The second component of the ETAS model is the triggering kernel $\theta\omega e^{-\omega t}$ that models "near-repeat" or "contagion" effects in crime data. The exponential decay causes grid cells containing recent crime events to have a higher intensity than grid cells with fewer recent events and the same background rate. The main difference between the ETAS model and prospective hotspot maps (Bowers, Johnson, and Pease 2004) that model near-repeat effects is the introduction of the background rate μ. Whereas prospective hotspot maps only take into account short-term hotspot dynamics, the ETAS model estimates both long-term and short-term hotspots and systematically estimates the relative contribution to risk of each via expectation-maximization (EM)(Mohler et al. 2011; Mohler 2014). Given an initial guess for the parameters θ, μ, and ω, the EM algorithm is applied iteratively until convergence by alternating between the following two steps:

E-step

$$p_n^{ij} = \frac{\theta\omega e^{-\omega(t_n^j-t_n^i)}}{\lambda_n(t_n^j)}, \tag{2}$$

$$p_n^j = \frac{\mu_n}{\lambda_n(t_n^j)}, \tag{3}$$

M-step

$$\omega = \frac{\sum_n \sum_{i<j} p_n^{ij}}{\sum_n \sum_{i<j} p_n^{ij}(t_n^j - t_n^i)}, \tag{4}$$

$$\theta = \frac{\sum_n \sum_{i<j} p_n^{ij}}{\sum_n \sum_j 1}, \tag{5}$$

$$\mu = \frac{\sum_n \sum_j p_n^j}{T}, \tag{6}$$

where T is the length of the time window of observation.

The EM algorithm can be intuitively understood by viewing the ETAS model as a branching process (Mohler et al. 2011). First-generation events occur according to a Poisson process with constant rate μ. Events (from all generations) each give birth to N direct offspring events, where N is a Poisson random variable with parameter θ. As events occur, the rate of crime increases locally in space, leading to a contagious sequence of "aftershock" crimes (Mohler et al. 2011) that eventually dies out on its own, or is interrupted by police intervention; the former occurs naturally so long as $\theta < 1$, while the latter is unaccounted for by the model. In the E-step, the probability that event j is a direct offspring of event i is estimated, along with the probability that the event was generated by the Poisson process μ. Given the probabilistic estimate of the branching structure, the complete data log-likelihood is then maximized in the M-step, providing an estimate of the model parameters. For a detailed treatment of the EM algorithm in the context of ETAS see Veen and Schoenberg (2008) or Lewis and Mohler (2011) where the EM algorithm is shown to be equivalent to projected gradient ascent optimization on the log-likelihood.

The code

Reproducible Research

"An article about computational science in a scientific publication is *not* the scholarship itself, it is merely *advertising* of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."
— Buckheit, Donoho, "WaveLab and Reproducible Research", 1995.

"In my own experience, error is ubiquitous in scientific computing ..."
— Donoho, "An invitation to reproducible computational research", Biostatistics (2010).

Merton's norms: universalism, communalism, disinterestedness, organized scepticism.

With thanks to Victoria Stodden.

Reproducible Research

"An article about computational science in a scientific publication is *not* the scholarship itself, it is merely *advertising* of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures." — Buckheit, Donoho, "WaveLab and Reproducible Research", 1995.

"In my own experience, error is ubiquitous in scientific computing ..."

— Donoho, "An invitation to reproducible computational research", Biostatistics (2010).

Merton's norms: universalism, communalism, disinterestedness, organized scepticism.

With thanks to Victoria Stodden.

Reproducible Research

"An article about computational science in a scientific publication is *not* the scholarship itself, it is merely *advertising* of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."

— Buckheit, Donoho, "WaveLab and Reproducible Research", 1995.

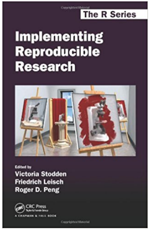
"In my own experience, error is ubiquitous in scientific computing ..."

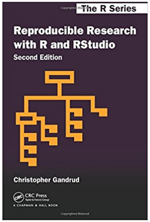
— Donoho, "An invitation to reproducible computational research", Biostatistics (2010).

Merton's norms: universalism, communalism, disinterestedness, organized scepticism.

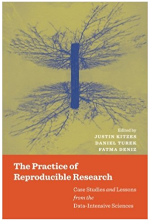
With thanks to Victoria Stodden.

Resources





- http://reproducibleresearch.net/
- https://rroxford.github.io/
- http://www.bmj.com/content/344/bmj.e4383



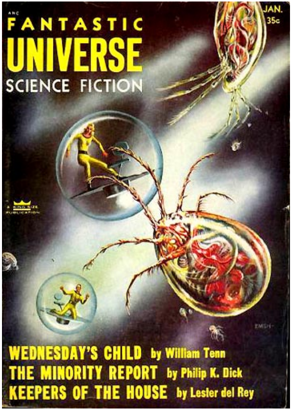
But to continue

Wikipedia entry “Hobby horse”

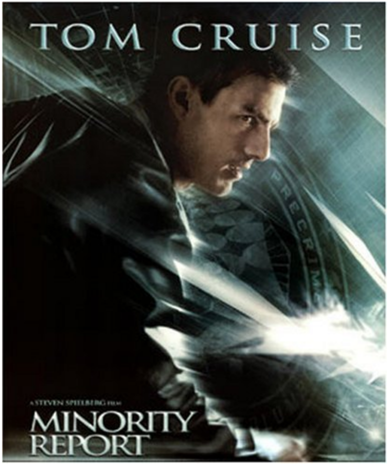
“My Uncle Toby on his Hobby-horse”, Wikipedia

What is crime prediction?

"Precrime: It Works!"



Wikipedia entry "The Minority Report"



From IMDB

What is crime prediction *actually*?

"Although much news coverage promotes the meme that predictive policing is a crystal ball, these algorithms predict the risk of future events, not the events themselves." Perry, McInnis, Price, Smith, Hollywood, "Predictive Policing", RAND report.

"Prior to each shift, Santa Cruz police officers receive information identifying 15 such squares with the highest probability of crime, and are encouraged — though not required — to provide greater attention to these areas." Joh, "Policing by numbers: Big data and the fourth amendment.

"Despite the increased emphasis on proactive policing, the core of police work remains that of responding to calls for service. . ." Groff, La Vigne, "Forecasting the future of predictive crime mapping".

What is crime prediction *actually*?

"Although much news coverage promotes the meme that predictive policing is a crystal ball, these algorithms predict the risk of future events, not the events themselves." Perry, McInnis, Price, Smith, Hollywood, "Predictive Policing", RAND report.

"Prior to each shift, Santa Cruz police officers receive information identifying 15 such squares with the highest probability of crime, and are encouraged — though not required — to provide greater attention to these areas." Joh, "Policing by numbers: Big data and the fourth amendment.

"Despite the increased emphasis on proactive policing, the core of police work remains that of responding to calls for service. . . " Groff, La Vigne, "Forecasting the future of predictive crime mapping".

What is crime prediction *actually*?

"Although much news coverage promotes the meme that predictive policing is a crystal ball, these algorithms predict the risk of future events, not the events themselves." Perry, McInnis, Price, Smith, Hollywood, "Predictive Policing", RAND report.

"Prior to each shift, Santa Cruz police officers receive information identifying 15 such squares with the highest probability of crime, and are encouraged — though not required — to provide greater attention to these areas." Joh, "Policing by numbers: Big data and the fourth amendment.

"Despite the increased emphasis on proactive policing, the core of police work remains that of responding to calls for service. . . " Groff, La Vigne, "Forecasting the future of predictive crime mapping".

Analogy with weather forecasting

I have found analogies with *probabilistic forecasting* within Meteorology to be very profitable.

"There is a 20% chance of rain in Leeds tomorrow."

- What does this *mean?*
- If we make this prediction many times, then 1 in 5 times, it should rain tomorrow. *"reliability"*.
- But maybe it rains 20% of the time in Leeds anyway (over a year, say)?
- *"resolution"* (which is hard to actually define.)

Analogy with weather forecasting

I have found analogies with *probabilistic forecasting* within Meteorology to be very profitable.

> *"There is a 20% chance of rain in Leeds tomorrow."*

- What does this *mean?*
- If we make this prediction many times, then 1 in 5 times, it should rain tomorrow. *"reliability"*.
- But maybe it rains 20% of the time in Leeds anyway (over a year, say)?
- *"resolution"* (which is hard to actually define.)

Analogy with weather forecasting

I have found analogies with *probabilistic forecasting* within Meteorology to be very profitable.

> *"There is a 20% chance of rain in Leeds tomorrow."*

- What does this *mean?*
- If we make this prediction many times, then 1 in 5 times, it should rain tomorrow. *"reliability"*.
- But maybe it rains 20% of the time in Leeds anyway (over a year, say)?
- *"resolution"* (which is hard to actually define.)

Analogy with weather forecasting

I have found analogies with *probabilistic forecasting* within Meteorology to be very profitable.

> *"There is a 20% chance of rain in Leeds tomorrow."*

- What does this *mean?*
- If we make this prediction many times, then 1 in 5 times, it should rain tomorrow. *"reliability"*.
- But maybe it rains 20% of the time in Leeds anyway (over a year, say)?
- *"resolution"* (which is hard to actually define.)

Analogy with weather forecasting

I have found analogies with *probabilistic forecasting* within Meteorology to be very profitable.

> *"There is a 20% chance of rain in Leeds tomorrow."*

- What does this *mean?*
- If we make this prediction many times, then 1 in 5 times, it should rain tomorrow. *"reliability"*.
- But maybe it rains 20% of the time in Leeds anyway (over a year, say)?
- *"resolution"* (which is hard to actually define.)

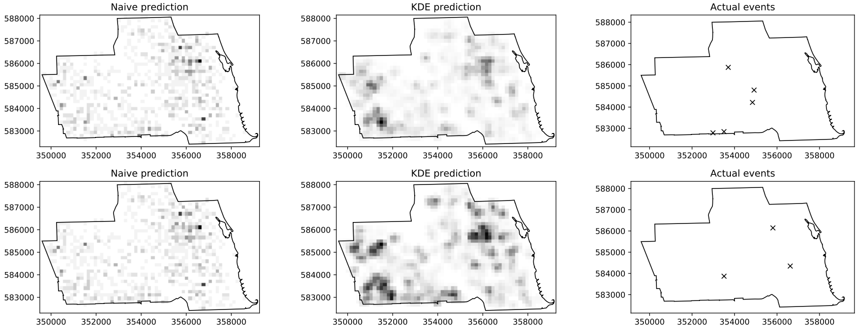
Analogy with weather forecasting

I have found analogies with *probabilistic forecasting* within Meteorology to be very profitable.

> *"There is a 20% chance of rain in Leeds tomorrow."*

- What does this *mean?*
- If we make this prediction many times, then 1 in 5 times, it should rain tomorrow. *"reliability"*.
- But maybe it rains 20% of the time in Leeds anyway (over a year, say)?
- *"resolution"* (which is hard to actually define.)

Lack of analogy



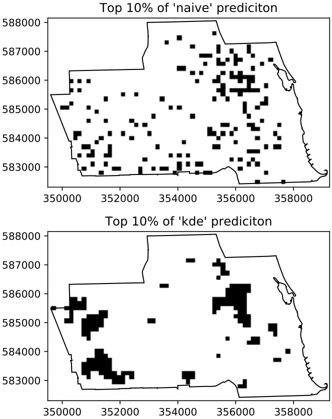
Northside of Chicago, predictions and reality for 5th Nov 2016, and 23rd October 2016.

- The probabilities involved are tiny.

Hit rate

The *de facto* standard.

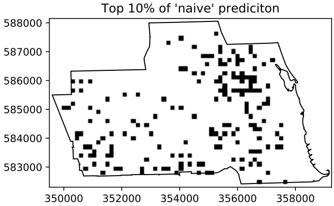
- Pick a "coverage level", say 10% of the area, which might be chosen given Policing resources.
- Pick that % of grid cells, by picking those with the highest risk first.
- Then calculate the fraction of actual events which fall in the selected grid cells.

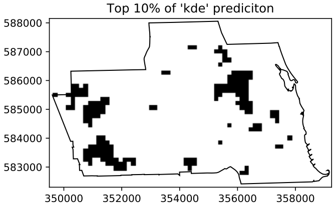


Hit rate

The *de facto* standard.

- Pick a "coverage level", say 10% of the area, which might be chosen given Policing resources.
- Pick that % of grid cells, by picking those with the highest risk first.
- Then calculate the fraction of actual events which fall in the selected grid cells.

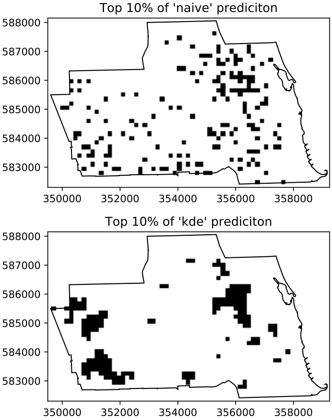




Hit rate

The *de facto* standard.

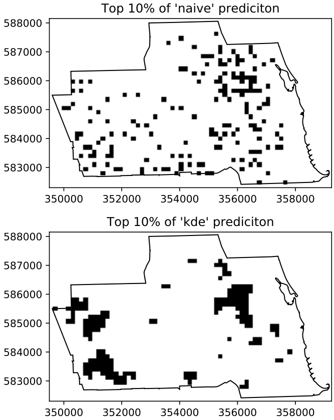
- Pick a “coverage level”, say 10% of the area, which might be chosen given Policing resources.
- Pick that % of grid cells, by picking those with the highest risk first.
- Then calculate the fraction of actual events which fall in the selected grid cells.



Hit rate

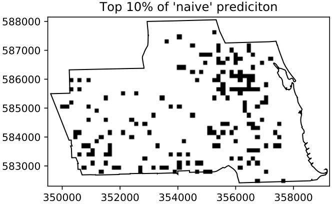
The *de facto* standard.

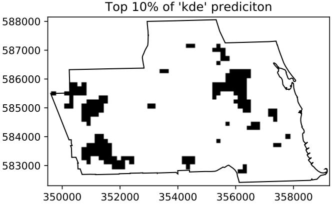
- Pick a "coverage level", say 10% of the area, which might be chosen given Policing resources.
- Pick that % of grid cells, by picking those with the highest risk first.
- Then calculate the fraction of actual events which fall in the selected grid cells.



The good, the bad, the ugly

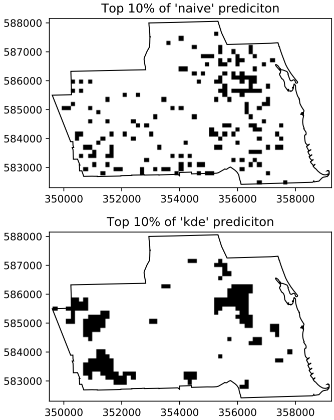
- Easy to understand, tied to *usage* of the prediction;
- But seems to me to confuse *prediction* with *hot-spot / patrol plan* creation.
- Notice the huge quantitative difference in the two examples.
- How do you deal with the selection of a coverage level?





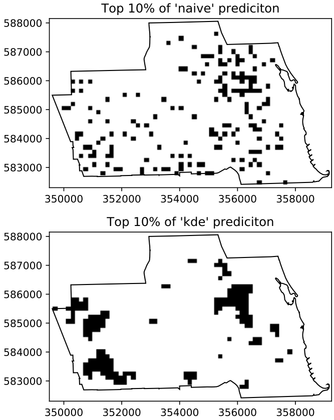
The good, the bad, the ugly

- Easy to understand, tied to *usage* of the prediction;
- But seems to me to confuse *prediction* with *hot-spot / patrol plan* creation.
- Notice the huge quantitative difference in the two examples.
- How do you deal with the selection of a coverage level?



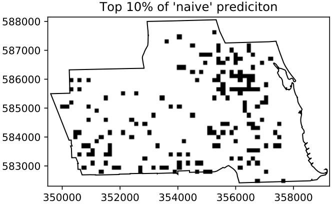
The good, the bad, the ugly

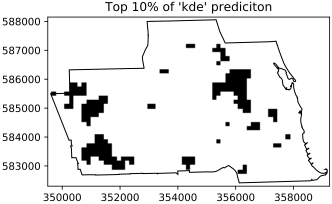
- Easy to understand, tied to *usage* of the prediction;
- But seems to me to confuse *prediction* with *hot-spot / patrol plan* creation.
- Notice the huge quantitative difference in the two examples.
- How do you deal with the selection of a coverage level?



The good, the bad, the ugly

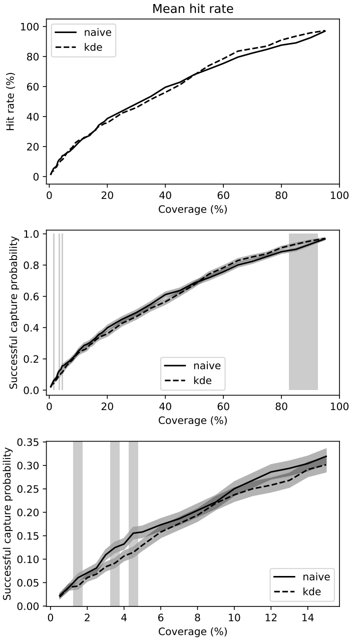
- Easy to understand, tied to *usage* of the prediction;
- But seems to me to confuse *prediction* with *hot-spot / patrol plan* creation.
- Notice the huge quantitative difference in the two examples.
- How do you deal with the selection of a coverage level?





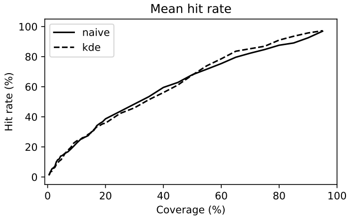
Interpret the results

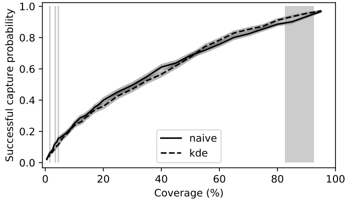
- Usual to plot *mean hitrate* against coverage. Then use some statistical test.
- But what's the model?
- Let's suppose that each trial is an independent draw from a binomial with unknown p.
- Use a flat prior. Compute the predictive posterior, plot the median and inter-quartile range.
- Gives much the same result (the number of events per day doesn't vary that much).

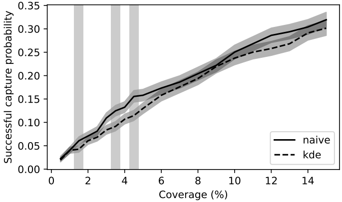


Interpret the results

- Usual to plot *mean hitrate* against coverage. Then use some statistical test.
- But what's the model?
- Let's suppose that each trial is an independent draw from a binomial with unknown p.
- Use a flat prior. Compute the predictive posterior, plot the median and inter-quartile range.
- Gives much the same result (the number of events per day doesn't vary that much).

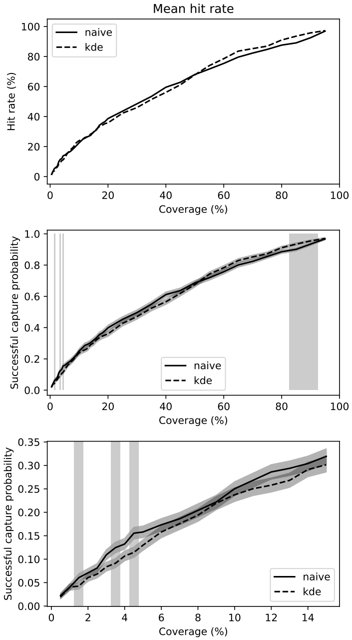






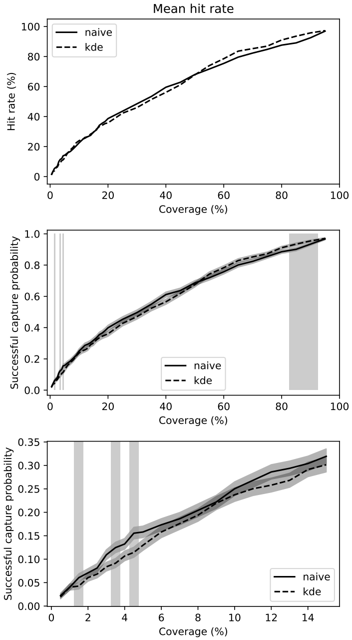
Interpret the results

- Usual to plot *mean hitrate* against coverage. Then use some statistical test.
- But what's the model?
- Let's suppose that each trial is an independent draw from a binomial with unknown p.
- Use a flat prior. Compute the predictive posterior, plot the median and inter-quartile range.
- Gives much the same result (the number of events per day doesn't vary that much).



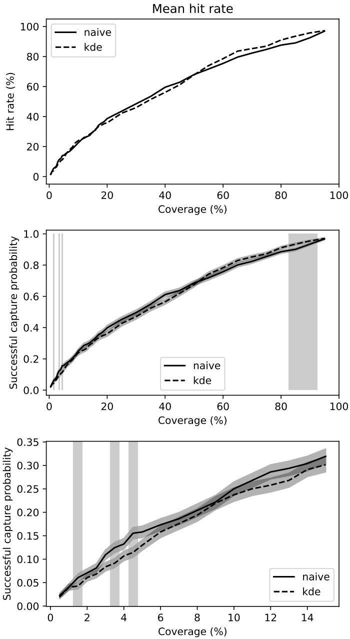
Interpret the results

- Usual to plot *mean hitrate* against coverage. Then use some statistical test.
- But what's the model?
- Let's suppose that each trial is an independent draw from a binomial with unknown p.
- Use a flat prior. Compute the predictive posterior, plot the median and inter-quartile range.
- Gives much the same result (the number of events per day doesn't vary that much).



Interpret the results

- Usual to plot *mean hitrate* against coverage. Then use some statistical test.
- But what's the model?
- Let's suppose that each trial is an independent draw from a binomial with unknown p.
- Use a flat prior. Compute the predictive posterior, plot the median and inter-quartile range.
- Gives much the same result (the number of events per day doesn't vary that much).



Brier scores

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

Return to Meteorology and probabilistic forecasting.

- Binary events: either happens (1) or not (0).
- For $t = 1, \cdots, N$ make a prediction $f_t \in [0, 1]$.
- Have actual events (o_t).
- We follow a variant from Roberts, "Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model"
  - K grid cells
  - predicted probability p_k
  - n_k actual events so n_k/N fraction.
- "Fractional Brier Score"

$$F = \frac{1}{K}\sum_{k=1}^{K}\left(p_k - \frac{n_k}{N}\right)^2$$

Brier scores

$$\boxed{BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2}$$

Return to Meteorology and probabilistic forecasting.

- Binary events: either happens (1) or not (0).
- For $t = 1, \cdots, N$ make a prediction $f_t \in [0, 1]$.
- Have actual events (o_t).
- We follow a variant from Roberts, "Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model"
  - K grid cells
  - predicted probability p_k
  - n_k actual events so n_k/N fraction.
- "Fractional Brier Score"

$$\boxed{F = \frac{1}{K}\sum_{k=1}^{K}\left(p_k - \frac{n_k}{N}\right)^2}$$

Brier scores

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

Return to Meteorology and probabilistic forecasting.

- Binary events: either happens (1) or not (0).
- For $t = 1, \cdots, N$ make a prediction $f_t \in [0, 1]$.
- Have actual events (o_t).
- We follow a variant from Roberts, "Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model"
  - K grid cells
  - predicted probability p_k
  - n_k actual events so n_k/N fraction.
- "Fractional Brier Score"

$$F = \frac{1}{K}\sum_{k=1}^{K}\left(p_k - \frac{n_k}{N}\right)^2$$

Brier scores

$$\boxed{BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2}$$

Return to Meteorology and probabilistic forecasting.

- Binary events: either happens (1) or not (0).
- For $t = 1, \cdots, N$ make a prediction $f_t \in [0, 1]$.
- Have actual events (o_t).
- We follow a variant from Roberts, "Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model"
  - K grid cells
  - predicted probability p_k
  - n_k actual events so n_k/N fraction.

$$\boxed{F = \frac{1}{K}\sum_{k=1}^{K}\left(p_k - \frac{n_k}{N}\right)^2}$$

- "Fractional Brier Score"

Brier scores

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

Return to Meteorology and probabilistic forecasting.

- Binary events: either happens (1) or not (0).
- For $t = 1, \cdots, N$ make a prediction $f_t \in [0, 1]$.
- Have actual events (o_t).
- We follow a variant from Roberts, "Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model"
  - K grid cells
  - predicted probability p_k
  - n_k actual events so n_k/N fraction.
- "Fractional Brier Score"

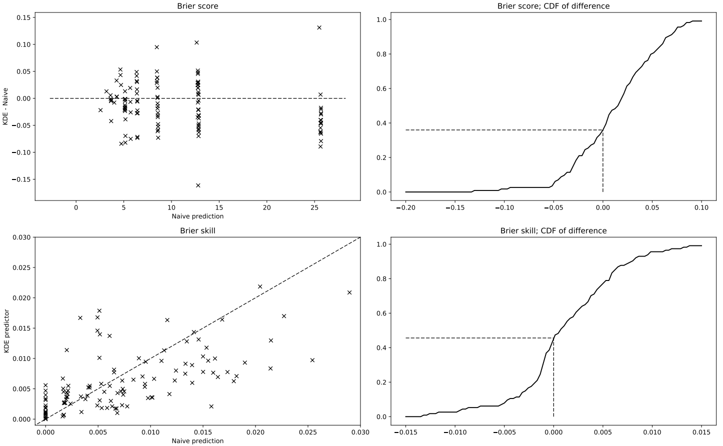
$$F = \frac{1}{K}\sum_{k=1}^{K}\left(p_k - \frac{n_k}{N}\right)^2$$

Skill score; results

$$F_{\text{worst}} = \frac{1}{K}\sum_{k=1}^{K}\left(p_k^2 + \left(\frac{n_k}{N}\right)^2\right)$$

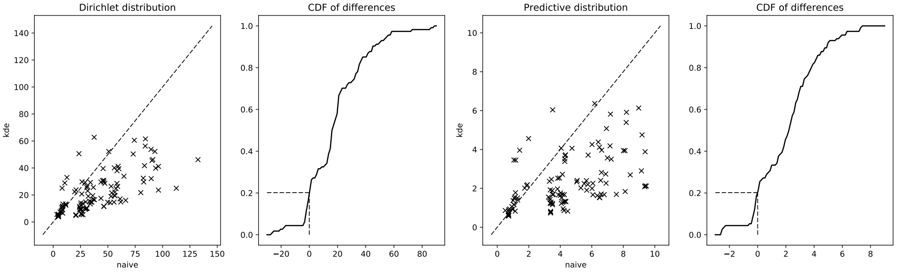
$$F_S = 1 - F/F_{\text{worst}}$$

- What are units of F?
- F_S is the "skill"; closer to 1 is better.



Bayesian information gain

- Want to capture the feeling that if we see more events on a given day, we should learn more about the quality of the prediction.
- My idea is to use the prediction to form a *prior*, the update this given the data to form a *posterior*, and then compare these with the Kullback-Leibler divergence.
- Measures the information gain from prior to posterior– a good prediction should mean less gained on learning the result.



Conclusions?

- Seems a little inconclusive.
- Hit rate, Brier scores, (other ideas we develop) show roughly a tie.
- The information gain idea is more of a clear win for the KDE method.

Original aim was to get beyond the “hit rate” as being the only game in town.

Bit of a work in progress: any ideas much appreciated!

https://github.com/QuantCrimAtLeeds/PredictCode/

Conclusions?

- Seems a little inconclusive.
- Hit rate, Brier scores, (other ideas we develop) show roughly a tie.
- The information gain idea is more of a clear win for the KDE method.

Original aim was to get beyond the "hit rate" as being the only game in town.

Bit of a work in progress: any ideas much appreciated!

https://github.com/QuantCrimAtLeeds/PredictCode/

Conclusions?

- Seems a little inconclusive.
- Hit rate, Brier scores, (other ideas we develop) show roughly a tie.
- The information gain idea is more of a clear win for the KDE method.

Original aim was to get beyond the "hit rate" as being the only game in town.

Bit of a work in progress: any ideas much appreciated!

https://github.com/QuantCrimAtLeeds/PredictCode/

Conclusions?

- Seems a little inconclusive.
- Hit rate, Brier scores, (other ideas we develop) show roughly a tie.
- The information gain idea is more of a clear win for the KDE method.

Original aim was to get beyond the "hit rate" as being the only game in town.

Bit of a work in progress: any ideas much appreciated!

https://github.com/QuantCrimAtLeeds/PredictCode/

Conclusions?

- Seems a little inconclusive.
- Hit rate, Brier scores, (other ideas we develop) show roughly a tie.
- The information gain idea is more of a clear win for the KDE method.

Original aim was to get beyond the "hit rate" as being the only game in town.

Bit of a work in progress: any ideas much appreciated!

`https://github.com/QuantCrimAtLeeds/PredictCode/`