



MIT Open Access Articles

Analytics for Power Grid Distribution Reliability in New York City

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation	Rudin, Cynthia, Seyda Ertekin, Rebecca Passonneau, Axinia Radeva, Ashish Tomar, Boyi Xie, Stanley Lewis, Mark Riddle, Debbie Pangsrivini, and Tyler McCormick. "Analytics for Power Grid Distribution Reliability in New York City." <i>Interfaces</i> 44, no. 4 (August 2014): 364–83.
As Published	http://dx.doi.org/10.1287/inte.2014.0748
Publisher	Institute for Operations Research and the Management Sciences (INFORMS)
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/99144
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/4.0/

Analytics for Power Grid Distribution Reliability in New York City

Cynthia Rudin, Şeyda Ertekin

MIT Sloan School of Management, Massachusetts Institute of Technology, rudin,seyda@mit.edu

Rebecca Passonneau, Axinia Radeva, Ashish Tomar, Boyi Xie

Center for Computational Learning Systems, Columbia University, becky,axinia,ashish,boyi@ccls.columbia.edu

Stanley Lewis, Mark Riddle, Debbie Pangsrivinij, John Shipman

Con Edison Company of New York, lewiss,riddlem,pangsrivinijd@coned.com

Tyler McCormick

Department of Statistics, Department of Sociology, Center for Statistics and the Social Sciences, University of Washington, tylermc@uw.edu

We summarize the first major effort to use analytics for preemptive maintenance and repair of an electrical distribution network. This is a large-scale multi-year effort between scientists and students at Columbia and MIT and engineers from Con Edison, which operates the world's oldest and largest underground electrical system. Con Edison's preemptive maintenance programs are less than a decade old, and are made more effective with the use of analytics developing alongside the maintenance programs themselves. Some of the data used for our projects are historical records dating as far back as the 1880's, and some of the data are free text documents typed by dispatchers. The operational goals of this work are to assist with Con Edison's preemptive inspection and repair program, and its vented cover replacement program. This has a continuing impact on public safety, operating costs, and reliability of electrical service in New York City.

Introduction

Loss of power is currently one of the most critical threats to the functioning of our society. Hospitals, universities, and almost all other businesses depend heavily on a reliable energy grid. As society moves towards electrified vehicles, high powered data centers, and ambitious new infrastructure, the demand for electricity will rise, and it is estimated that the increased demand will cause us to rapidly exceed our current ability to reliably deliver electrical power (DOE 2008, Rhodes 2013, NYBC 2010). The demand for electrical power is increasing in the U.S. at a steady rate (NYBC 2010), and with adequacy of power resources being an imminent concern in many urban centers, it is essential that all available resources for power grid reliability be spent wisely. In New York City alone, peak demand has been increasing, and a record for peak demand was set in 2006 at 13,141 megawatts, and

then another record was reached in 2011, at an even higher 13,189 megawatts (ConEdison 2011). Blackouts across the U.S. are on a dramatic rise, with the number of large outages (affecting over 50,000 consumers) more than doubling over the last 15 years, going from 41 outages between 1991-1995, to 58 outages in the second half of the 1990's, to 92 outages in the early 2000's (Amin 2011). In New York City, the estimated cost of a single large scale 29 hour blackout was approximately \$1.05 billion or \$36 million per hour (MSNBC 2003), and an estimated 90 lives (Siegel 2012).

One of the main reasons for concern about the reliability of the power grid is the age and state of the electrical infrastructure in many cities. According to the US Department of Energy's Grid 2030 report (DOE 2003), "America's electric system, 'the supreme engineering achievement of the 20th century,' is aging, inefficient, and congested, and incapable of meeting the future energy needs of the Information Economy without operational changes and substantial capital investment over the next several decades." Some parts of the energy grid are surprisingly old; in fact, some parts of the original electrical grid dating from 1880's in the time of Thomas Edison are still in operation in New York, and probably the same is true of many other older US cities. According to a 2007 survey on reliability issues from the North American Electrical Reliability Corporation (NERC 2007), of all the issues considered to affect reliability, "Aging Infrastructure and Limited New Construction" was ranked first among all technical issues, with the highest likelihood and highest impact. In Manhattan alone, we calculated that at least 5% of the low-voltage distribution cables were installed prior to 1930, and many of these old cables are still functioning reliably. However, we are currently taking our electrical grids to the limit (or beyond the limit) of what they can handle, and emergency situations are occurring more and more often.

It is only within the last decade that power companies have switched from reactive maintenance (fix it after failures occur) to preemptive maintenance (fix it before failures occur) to mitigate the number and severity of power failures on the distribution network. Specifically, these preemptive maintenance programs started in 2004 (So 2004).

The key problem for the effectiveness of these programs then becomes one of *prediction* - if it is possible to predict where failures are most likely to occur, then power companies can directly target the most vulnerable equipment for inspection and preemptive repair work. As far as we can tell, for the low-voltage distribution network, there is no other reasonable alternative to the use of prediction for targeted maintenance in order to more

effectively limit power failures. The low-voltage (secondary) distribution network in any major city is *immense*, as it comprises the full network of cables that traverses every street in the city, connecting each building and streetlight to the electrical grid. In New York City, there is enough underground cable to wrap around the world three and a half times (94,000 miles), and in Manhattan alone, there is almost enough underground cable to go around the world (21,216 miles). It is not reasonable (nor would it be a good use of resources) to suggest that all of the distribution network is replaced every few years or so, as this would be prohibitively expensive, wasteful, and disruptive. For the same reason, it is also not possible to put smart meters, or monitoring devices, on each length of cable; and even if we did this, it may not preemptively stop failures from occurring, rather it may mainly detect failures that are already occurring. In 2011, Con Edison invested \$1.2 billion to upgrade and reinforce electrical systems, in the process replacing 1,550 miles of underground electric cable, which is an enormous amount of cable, but only a small fraction of the total electrical grid. Our evidence also suggests that targeting only the oldest cables for replacement is not an optimal strategy; cable age is only one factor (and not the most important factor) in predicting failures. In order to truly target the most vulnerable equipment, we need to consider many factors, including cable age, cable and insulation material, and the history of past incidents. Predicting power failures in advance is a truly challenging problem, with the combination of the vastness of the grid, the aging of its equipment, and the sheer number of factors that one should consider in order to make accurate predictions; and the increase in power failures in the US and increase in peak demand make our need to solve this problem more urgent than ever.

In this paper, we discuss the first major effort to preemptively maintain a power grid using analytics. New York City's grid is both the oldest power system in the world, and the largest underground electrical system. If analytics can be used in New York, we hypothesize that it can be used in any major urban center. This work started at the right time: just as the new preemptive inspection programs were starting in New York City, we developed machine learning tools that target the most vulnerable equipment in order to make these programs - immediately - as effective as possible. The system-wide scale of this effort is unprecedented, as are some of the resulting analytical challenges. There are collaborative projects targeting preemptive work on all different parts of the distribution network, including the primary grid (feeders, cable sections, joints, terminators), transformers, and

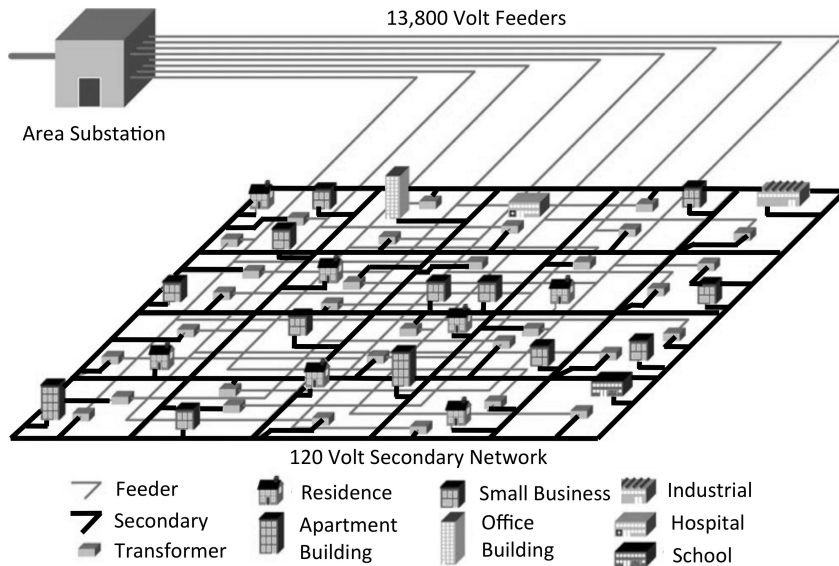


Figure 1 This schematic of Con Edison Power System for Manhattan shows the high voltage primary system, the transformers that step down the voltage, and the secondary system that traverses each street and avenue in the city. Source: Con Edison (figure was adapted).

the secondary grid. The primary grid data are mainly real-time sensor data, thus predictions of feeder failures are updated regularly, and primary (feeder) cables go along major avenues (see Gross et al. 2006). The data from the secondary grid are very diverse, including historical cable records, some of which are 120 years old, and trouble tickets typed by dispatchers while they are directing repair work in response to an incident. The secondary grid in New York City is massive, with cables under every street and avenue in the city. An illustration of a power system is in Figure 1, illustrating both the high voltage primary system, the transformers that step down the voltage from 13,800 volts to 120 volts, and the low-voltage secondary system. In this paper, we will focus mostly on secondary grid reliability, as this project has been the most challenging from an analytical perspective in many ways, and thus has given rise to the most innovations among the projects.

For the secondary grid, insulation breakdown on the cables can happen over a period of years, which can cause cable insulation material to ignite. Fire or smoke emanating from the underground system may be visible through the access points to the underground grid, which are called manholes. A picture taken inside a manhole is in Figure 2. It is possible for a fire to spread from manhole to manhole through the underground ducts. Figure 3 shows several recent manhole fires across cities in the US, each of which caused serious power failures and required substantial repair work. These are the kinds of events that the



Figure 2 This image was taken inside a manhole on lower Broadway in Manhattan in 2009. Source: (Little 2009).



February 11th, 2010 Manhattan, NY
<http://www.inquisitr.com>



January 27th, 2013 Omaha, NE
<http://www.theblaze.com>



June 5th, 2009 San Francisco, CA
<http://www.sfgate.com>

Figure 3 These are images of manhole fires caused by problems underground. The fire in Omaha in early 2013 spread through the underground ducts from manhole to manhole.

secondary grid project aims to reduce in New York, and this paper will focus mainly on mitigating the risk of these disastrous events.

The tools that have resulted from the Con Edison / Columbia projects have had, and will continue to have broad impact: the analytical tools developed during these projects can be used widely for reliability applications beyond the power grid; for instance in airline fleet maintenance and manufacturing equipment prioritization, the goals of prioritizing equipment for targeted repair are similar, and these domains sometimes even have the same types of diverse, historical and free-text data that Con Edison has (e.g., Oza et al. 2009, in the domain of airline maintenance). We have demonstrated that predictions can

be operationalized on the power grid through inspection programs, preemptive repair and upgrade programs, and vented cover replacement programs. The ideas behind the human-centered interactive tools we developed for making predictions convincing and actionable can be used broadly, given how useful these tools have been to our team of individuals with vastly different backgrounds. Finally, the innovative statistical models and machine learning tools that were developed for creating the predictions can be used for many different applications. In particular, power grid maintenance is one of the first applications of learning-to-rank technology outside of the realm of information retrieval and web search, though it is a natural fit to reliability prediction problems. The statistical tools in development for the short term prediction model described below (“Reactive Point Processes”) can be used in applications ranging from crime modeling to neuroscience, where there are self-exciting and self-regulating components contributing to overall vulnerability levels.

Besides the impact that analytics will have on New York City, an important impact of the Con Edison / Columbia projects is the knowledge that analytics can be used to maintain a power grid on a very large scale - and now that the foundation for doing this has been established, it should help us worldwide to battle the growing threats of increasing power failures, increased demand for electricity, a massive amount of aging equipment, and increased reliance on electricity in our modern world.

Objectives

Con Edison currently has several systems in place for both proactive and reactive maintenance. The reactive maintenance system for the low-voltage grid is the Emergency Control Systems (ECS), which answers calls about potential emergencies such as customer outages, manhole fires and explosions, smoking manholes, flickering lights, and low voltage. When a call comes in, Con Edison dispatchers direct Con Edison’s reaction to the event. Events can range in seriousness, where some events are mild and require minimal repairs, and some of the events can be quite serious, involving extensive cable damage in multiple manholes, and millions of dollars in repair work, and risks to public safety. These are the kinds of events that our work aims to prevent.

There are several proactive maintenance systems for the secondary grid, all of which have been in place for less than a decade, and around the same time as the collaboration with Columbia was started. The inspections program started in 2004, and at the time

of this writing in 2013, all manholes have been inspected at least once, and some have been inspected several times. The inspection program includes both scheduled inspections and ad hoc inspections. An ad hoc inspection takes place when a utility worker is already inside a manhole (for instance to connect a new service cable to a building for additional service), and fills in an inspection report while inside the manhole. Currently, scheduled inspections take place at least once every 5 years. Manhole inspectors identify possible defects with the electrical system and sort them into several categories ranging from urgent (level I repair - must be done before the utility worker leaves the manhole), semi-urgent (level II - must be done within 1 year) to non-urgent (level IV repair - can be placed on a waiting list and completed when possible). The level IV repairs are usually major infrastructural improvements like manhole enlargements or “cut and racks,” where cables are made parallel along racks on the sides of the manhole.

The first main goal of our project is to make this inspection program as effective as possible, by targeting the most vulnerable manholes for inspection and repair work.

Our work has had impact in achieving this goal, in the following ways. We have shown that:

- Prediction of manhole events from diverse historical data is possible, as measured quantitatively by blind prediction tests.
- Case studies and visualization tools can be used to communicate and evaluate particularly vulnerable manholes.
- For the most vulnerable manholes, our evidence suggests that the inspections program has reduced vulnerability levels substantially. The inspections program also has identified major repairs that could be completed, potentially leading to further substantial reductions in future vulnerability.
- Our statistical models can estimate the effect of various policy decisions for the inspection program, and project the effect of the inspections program into the future for a cost vs. benefit analysis of the program.

Con Edison has additional proactive maintenance programs for replacing solid manhole covers with vented covers. (See Figure 4.) In the past, almost all manhole covers were solid, which had considerable risks to the public: solid covers prevent gases from escaping, potentially leading to pressure buildup inside the manhole, and a manhole explosion.



Figure 4 This is an image of a solid cover (left) and a new vented cover (right). Source: Flickr, photostreams of Nick Sherman and i.follow.

The vented cover replacement program reduces the number of serious manhole events, by turning them into less serious smoking manhole events.

The second main goal of our project is to make the vented cover replacement program as effective as possible, by optimizing the placement of these vented covers. Con Edison's filing with the Public Service Commission for the vented cover replacement program is \$52.8 million for 2013-2017, and we wish to spend this to optimally reduce the risk of manhole events. We will discuss later the impact of our work, in providing a method for optimizing the vented cover replacement program to mitigate the impact of manhole events on public safety.

Challenges

In both goals discussed above, the key to success is prediction of the future that is as accurate as possible. This requires a fundamental characterization of the past. We will start by discussing the major challenge of understanding the data collected about the past by Con Edison.

The Challenge of Diverse Historical Data for A Massive and Aging Power Grid

Con Edison started collecting data from the very beginning of the power grid, and it is the world's oldest power grid. The Con Edison databases have been continually updated over the last ~ 120 years, and Con Edison now possesses very detailed data about the grid, including the dates of installation of the electrical cables, the material and size of the conductors and the insulation, the locations and other physical properties of the manholes, the manhole cover information, network information for each manhole, inspections reports, and records of past outages. It is a huge amount of data, with numerous tables, none of which were designed for the purpose of being integrated and used for predictive modeling

(it has all been “repurposed”). Integration - even of the data sources that are structured - can be problematic. For instance, there are several departments within Con Edison that each store their own cable data. Some of the departments have more reliable connectivity data than others, and some have better installation dates. Many of the databases do not store the unique identifiers for manholes, so matching the hundreds of thousands of cable records to their appropriate manholes is considerably challenging. However, our ability to predict failures accurately depends heavily on our knowledge of the state of the grid, so it is essential to handle these data as carefully as possible. In Manhattan alone, our final cable records table has over 220 thousand records, where the cables enter into 53 thousand manholes.

The unstructured data are even more of a challenge. These are the Con Edison trouble tickets, which come from the Emergency Control Systems discussed above. Our goal is to predict past events and the trouble tickets are, by far, the most complete source of event data. An example trouble ticket for a 2010 smoking manhole event is in Figure 5. The ECS trouble ticket system was started in the 1970’s, when tickets were written on carbon copy. The “B” copy was the one that was filed, and the Con Edison workers still call these tickets “B-Tickets” even though they have been electronic for over a decade. The B-Tickets are comprised of a mixture of automatically generated text and free text. The automated text is generated by the dispatchers’ actions, .e.g., 02/11/10 19:38 ACT TRBL CHNGD FROM RCUSMH TO RUWSMH BY 11511 which means the event’s designation went from a preliminary smoking manhole designation to a verified smoking manhole designation, and the operator who changed it has identifier 11511. Tickets often have contributions from several different dispatchers, each of whom is directing part of Con Edison’s response to the event. In the ticket in Figure 5, there appear to be comments from four separate dispatchers. Each dispatcher uses their own form of shorthand, that even with best efforts towards uniformity, causes the language in the tickets to be highly irregular. The ticket in Figure 5 is relatively straightforward; someone reports a smoking manhole and a smell of burning wires. The manhole (called service box 63555¹) needed to be flushed out so that repair work could be done. The manhole had a vented cover; a manhole with a vented cover is more likely to have a less serious smoking manhole event than a more serious explosion

¹ All manhole numbers, names, and addresses were changed for the purpose of anonymity.

event. The burnouts (B/O, burned pieces of cable) were cleared out by one of the utility workers. The manhole needed a cut and rack, which means that cables were not in straight racks along the sides of the manhole, but it is not clear whether the cut and rack was performed. Cut and racks, and manhole enlargements, can require considerable time and expense, and are not generally considered urgent repairs. In some sense, the event data are our most important data, because we are specifically trying to predict events. However, transforming these data into something that predicts power grid failures was extremely difficult.

A major challenge when dealing with diverse historical data is determining which data to trust. Historical data often contain conflicting or irreconcilable information, and it is important to know which source is more reliable. Sometimes it is clear that additional data processing will be needed, but it is not clear which data should next be processed to achieve the highest level of prediction accuracy. This is particularly true for pilot projects like ours, where it was not clear at the outset whether power failures could be predicted at all from these extremely raw data. The most experienced engineer who started the project with us (Steve Ierome, who has since retired) hypothesized that serious events could be predicted by past “precursor” non-serious events or past serious events. These past events indicate potential vulnerability of the underlying local network. This hypothesis was called the “hotspot” theory. However, to verify the hotspot theory required in depth handling of the trouble tickets, which the Columbia team initially could not read, and did not believe were trustworthy. (Steve turned out to be correct in his hypothesis, and the tickets turned out to be particularly useful.) Initial handling of the trouble tickets was problematic; for instance, in the trouble ticket of Figure 5, the manhole that is given as the trouble hole (source of the problem) is service box 63555. However, knowing that it is a service box and knowing its number does not always constitute a unique identifier: there can be multiple manholes with the same type and number. Further, manhole numbers can be mistyped within the tickets or the structured fields. Our mechanism for dealing with this problem is multi-faceted, where a combination of domain expertise, overlapping information between several sources, and descriptive statistics characterizing the match is used to ensure the integrity of the data within our database.

Ticket: ME1000002966
Lines: 32
Trouble Type: SMH (Smoking Manhole)
Address: 2400 SULLIVAN ST
Received Datetime: 2012-02-11 16:10:00-05
Cross St : W 3 ST
Remarks:
PAWELA KULESZ 8005556633:SMOKE. COVER/ON. LOC:FRT OF STORE.
CALL TRFD TO 911. OPERATORS ID/BADGE#:22621.SMELLS BURNING WI
RES.
02/11/10 16:10 OR/CU CNT FROM 0 TO 1 BY STAR_AUTO
02/11/10 16:10 FDR FROM TO MN/SHRSQ BY STAR_AUTO
02/11/10 16:15 OR/CU CNT FROM 1 TO 0 BY paganor
02/11/10 16:15 OR/CU CNT FROM 0 TO 0 BY S.paganor
02/11/10 16:50 ACT TRBL FROM EDSSMH TO RCUSMH BY S_
02/11/10 19:38 ACT TRBL CHNGD FROM RCUSMH TO RUWSMH BY 11511
02/11/10 19:38 REFERRED TO: M.EC1.BTUG RUWSMH FYI BY 11511
02/11/10 21:50 REMARKS BELOW WERE ADDED BY 11511
02/11/10 21:49 UG REPORTS TROUBLE STRUCTURE IS SB-63555 F/O
10211 W.3ST,ROUND,VENTED AND SMOKING LIGHTLY. NEEDS FLUSH,
02/11/10 21:51 REMARKS BELOW WERE ADDED BY 11511
102323.....FH
02/12/10 05:01 REMARKS BELOW WERE ADDED BY 11511
02/11/10 05:00 UG REPORTS ALL B/O'S CLEARED AND AN EMERGENCY
CUT & RACK NEEDED.....FH
* ELIN REPT ADDED FOR INCIDENT:SMH 02/12/10 05:02*BY 11511
**** REPORTED BY: FIRE DEPT ****
STRUC MSPLATE TYPE NUMBER COND COVTYP COVFOUND DISTANCE
(1) 014H SB 63555 MU 0 Y 00
TYPE OF CURRENT: ALTERNATING CURRENT
VOLTAGE: 120/208V
APPROPRIATE SIZE: 004
CONDUCTOR CODE: COPPER
POSSIBLE CAUSE OF THE INCIDENT: INSULATION BREAKDOWN
WEATHER CONDITIONS DURING THE INCIDENT: FREEZE
***** ELIN REPORT END *****
02/12/10 05:02 CMP DISPATCHED BY STAR_heimachfr
02/12/10 05:02 CMP ARRIVED BY STAR_heimachfr
02/12/10 05:02 CMP COMPLETE BY STAR_heimachfr

Figure 5 This is an example of a Con Edison trouble ticket for a smoking manhole event. Over 150,000 trouble tickets are used to create predictive models. On arrival, the manhole was smoking lightly, and the Con Edison workers performed necessary repairs by clearing burned portions of cable (B/O's).

The Challenge of Predicting Rare Events on a Dynamically Evolving System

The power grid is dynamic, in the sense that its state evolves continuously over time as events occur, repairs are made, equipment is replaced or is sometimes taken out of service, demand oscillates, weather conditions impact reliability, and also as the inspections and preemptive repair programs ramp up. Each of these causes the reliability of the grid to evolve over time, and requires our estimation of vulnerability to change over time as well. Changes to the power grid need to be made while it is operating: the analogy we use is that maintaining the power grid is like fixing a large aircraft while it is still flying in the air.

Very serious events are also very rare events, which leads to problems evaluating predictive performance of the system. To assess vulnerability, we need enough failure events over each time period that we can accurately estimate risks. The best mechanism we have found to do this involves the following:

- We measure risks with respect to all serious events, not just very serious events. Very serious (fires, explosions) and less serious events (smoking manholes, underground burnouts, and power outages) have the same cause, which is insulation breakdown and weakness in the local infrastructure. Using more events to learn from allows us to better characterize the failure class. We have found that using the less-serious events to build the predictive model allows it to be more accurate in predicting serious events in blind tests. In these tests, recent data are withheld from the Columbia database for out-of-sample evaluation of future event prediction models.

- We choose performance metrics designed for rare event prediction. Our data tend to be highly imbalanced, as we have much fewer instances of the failure class than the non-failure class. The most typical performance measure for machine learning, which is classification accuracy, is a completely wrong quality measure for rare event prediction. For example, with an imbalance ratio of 99 non-failures to 1 failure, a model that predicts everything as a non-failure will be 99% accurate but it will not have any practical use. Since our goal is to *rank* the manholes in order of vulnerability, we use and optimize *rank statistics* to evaluate and tune the performance of our system. Rank statistics will be discussed further below. If the predictive model yields high values of these rank statistics on blind tests, it is an excellent indicator of good future system performance.

A related challenge is that the preemptive maintenance programs themselves affect the vulnerability levels. We evaluate the treatment effect of the inspection program dynamically. Controlled experiments are not possible since the grid needs to be continually maintained, but observational studies are possible (Wu et al. 2011c,b,a). We determine the effect of the inspection program by propensity score matching (Passonneau et al. 2011). We match structures that have been inspected to similar structures that have not been inspected to see the change in vulnerability. For the vented cover programs, we match manholes over a time period before they are vented to a time period after they are vented to measure the change in vulnerability. The vulnerability levels have decreased considerably as a result of both of the programs separately, particularly for the most vulnerable class of manholes, which is the top 5,000 manholes on the ranked list.

In what follows, we will discuss the process for predicting failures, and operationalizing these predictions in New York City. This involves handling the challenges discussed above: working with incredibly raw and historical data in order to characterize the past and current state of the grid, in order to predict the future.

Solutions: Prediction by Machine Learning

Our system for predicting manhole events has several separate modules, connected together as in Figure 6. The process starts with all of the raw Con Edison tables, including the Emergency Control Systems trouble tickets, the inspection reports, several tables of cable and connectivity data, manhole location data, data from Con Edison’s electrical shock and energized equipment program (contact voltage), open mains data (data about cables that are not connected during certain intervals of time), localized network data, and vented cover data. It is a huge number of tables that all need to be fused into a single database of the highest possible quality to allow measurement and optimization of predictive performance.

The first set of modules (Geocoding, Metadata Extraction) generate structured data from unstructured data as we discuss below.

Geocoding: As our predictive models are based on local geographic information, it is essential that we know the location of each event. The trouble ticket addresses are typed by dispatchers in raw form, which cannot be immediately used to verify the location of an event. To solve this, ticket addresses are geocoded and a latitude and longitude are

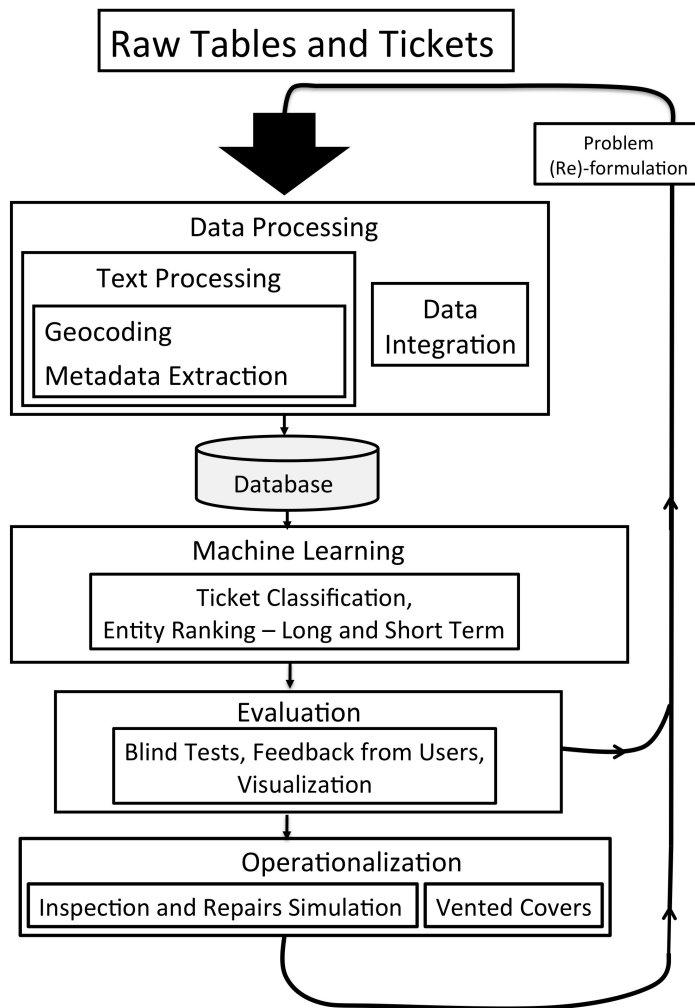


Figure 6 This is a diagram representing the knowledge discovery system for the manhole events prediction project.

assigned to the address of the ticket. Many of the ticket addresses are typed by dispatchers using irregular shorthand and with street names misspelled (e.g., C/O GREENWHICH ST rather than CORNER OF GREENWICH STREET), so the system uses a semi-automated process to convert irregular phrases to more standard ones, sends the corrected addresses to a geocoding system, and a matching process is used afterwards to match the suggestions from Google Earth to the tickets. The Geocoding module is described in depth by Dutta et al. (2008).

Metadata Extraction: The challenging free text fields of the trouble tickets have been more valuable in understanding the seriousness of an event than any of the structured fields. The free text often provides much more detail than the structured fields about the

final table are trustworthy. For instance, the manhole identifiers that we find within the ticket text are checked against the geocoded address of the ticket to ensure that they are physically close.

Machine Learning: Our ability to predict the future comes from our ability to accurately model the past. There are several machine learning models created for the secondary grid project.

- **Ticket Classification.** Using statistics of the data extracted from within the ticket texts, we classify each ticket as to whether it represents: i) a serious event, the kind we would like to prevent, ii) a less serious “potential precursor” event, iii) a non-event, where the ticket should not be used for further analysis. Our initial classification of tickets was based on a study, where several Con Edison engineers manually labeled a set of tickets that our system could learn from (Passonneau et al. 2009). Currently, the system uses a support vector machine (Vapnik 1995) to label the tickets, based on features of the tickets, including the presence of various words within the ticket (e.g., shunts or cleared), the length of the ticket, the ticket’s structured fields, etc. The number of variables optimized for the predictive model changes from year to year, and ranges from $\sim 5,000$ - $7,000$ in Manhattan to $\sim 9,000$ - $15,000$ in Brooklyn. The ticket classification defines what is meant by “serious” and “non-serious” events, and is used for fitting and evaluating both the long and short term vulnerability models for manholes, discussed below.

- **Long Term Vulnerability Modeling of Manholes by Learning-To-Rank.** Since manhole events are rare events, our data are highly imbalanced, and as we discussed earlier, our quality measures are thus *rank statistics*, or statistics calculated from a ranked list of manholes. Examples of rank statistics include: the number of misranked pairs (number of times that a manhole having an event was ranked as having lower vulnerability than one that did not have an event), and the number of manholes within the top 10% of the ranked list that had an event. If the main goal is to achieve high values of rank statistics, then it is sensible to directly optimize the model for these rank statistics; this is called “learning-to-rank” or “supervised ranking.”

Learning-to-rank developed in the information retrieval and web search community, where these methods are among the current state-of-the-art in machine learning methods for ranking webpages based on content. In web search, there are also a set of entities to be ranked - namely the webpages. Also, in web search, rank statistics such as the

discounted cumulative gain are the evaluation measures of choice. Our project is one of the first projects to use learning-to-rank methodology outside of the information retrieval community, and a number of innovative rank statistics and algorithms have been developed that can be used broadly for prioritization problems such as web search and maintenance (Rudin 2009, Ertekin and Rudin 2011).

The ranking models constructed using learning-to-rank methodology are optimized on rank statistics, usually over $\sim 10^6$ variables, taking into account $\sim 2-3$ million pairs of manholes. The variables that generally matter the most include statistics of past involvement in events (the hotspot theory), and the number and age of the electrical cables. In the next section, we discuss how the long term model is evaluated. Appendix A describes the P-Norm Push algorithm for the problem of learning-to-rank, which is used for failure prediction. The P-Norm Push optimizes a specific rank statistic that focuses on accuracy at the top of the ranked list.

- **Short Term Vulnerability Modeling of Manholes by Reactive Point Processes.** At the start of the project in 2006, before the major efforts in ticket processing, our database was not of high enough quality to allow accurate short term predictions of future events. For instance, even if we knew a serious event would occur within 60 days, we could not predict when within those 60 days it would occur (see Rudin et al. 2010). However, even in 2006 when this project started, the ultimate goal was to be able to achieve short term predictive power, based on a hypothesis of Con Edison engineer Steve Jerome, who believed it could be done. Recently, armed with much higher quality data, we revisited this goal, and designed a new statistical model of vulnerability (Reactive Point Processes, Ertekin et al. 2013) that is able to make reasonably accurate short term predictions. This is described in Appendix B.

Evaluation and Operationalization: The machine learning steps outlined just above are the key to the success of the project. Without good estimates of vulnerability, there is no chance that we would be able to assist with targeted repair and maintenance. For this reason, we evaluate the predictive model extensively, through blind test evaluations, case studies, and visualization tools. After the evaluation by the combination of quantitative blind tests, and qualitative evaluation of our data processing by our domain experts, improvements are often suggested and made, generally to the sources of data and initial data processing, in order to ensure the highest integrity of the predictive model.

The predictions are operationalized through the inspection programs and vented cover replacement programs, as discussed in the “Impact” section below. Specifically, we will discuss how: (i) The rankings are used to effectively target manholes for inspections and vented cover replacements; (ii) The RPP models are used to simulate the future impact of inspections policies on the number of events. These findings assist Con Edison in developing and supporting optimized future inspection policies and programs; (iii) For the vented cover replacement programs, a Minimum Vertex Cover problem has been solved for each network in Manhattan to determine the minimal set of manhole covers to replace in order to protect (vent) the entire network.

Impact

We discuss impacts in prediction accuracy, human-centered assessment tools, decision-making for the targeted inspections program, and decision making for the vented cover replacement program.

Impact in Ability to Target the Location of Future Failures: The extensive predictive modeling in this work is the key to effective targeting of vulnerable equipment; any increase in predictive accuracy can translate into a benefit for public safety, reliability of electrical service, and cost savings.

Con Edison has regularly performed blind tests of prediction quality, where recent data are withheld from our database, creating a true barrier to the future. Columbia/MIT produces a ranked list of manholes and this list is provided to Con Edison, who evaluates the list based on the held-out data. Figure 8 shows the evaluation results in two ways, the first one is what we call an “arrow plot” (top figure) where the manholes are listed from most vulnerable (left) to least vulnerable (right). An arrow on the line indicates that this manhole had a fire (F) or explosion (X). Many of the fires and explosions are clustered near the top of the list, indicating high quality predictive performance. In fact, for this particular blind test, which occurred in the Bronx, we were able to capture 44% of the serious manhole events that happened within 2009 within the top 10% of our ranked list, which was constructed from data before mid-2008.

The same information is conveyed in Figure 8 (bottom figure), called a “pseudo-ROC” plot. Along the x-axis is again the ranked list of manholes. Moving left to right along the

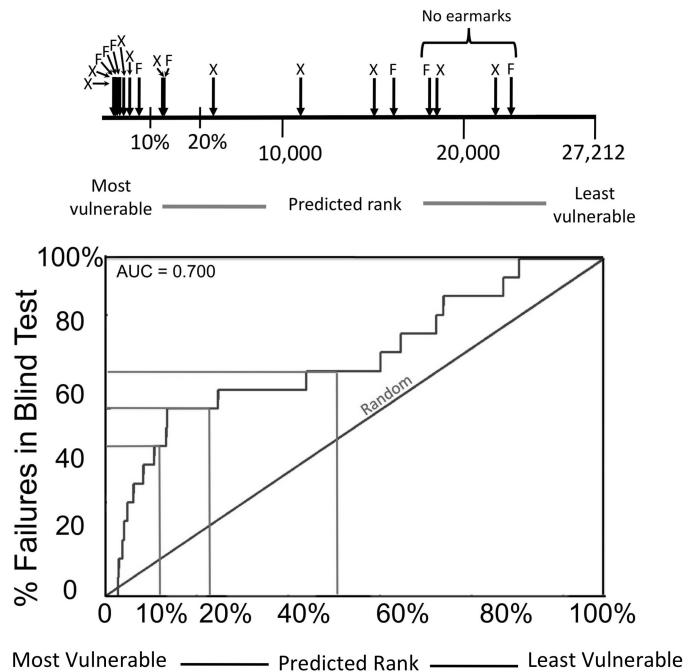


Figure 8 These are results from a blind test evaluation in the Bronx. The horizontal line in the top plot illustrates the manholes ranked according to predicted vulnerability. An arrow was placed at each manhole that experienced a serious event in 2009: 'F' is fire, 'X' is explosion. The lower plot conveys the same information in a pseudo-ROC curve. Moving from left to right along the ranked list, whenever a manhole that had an event is reached, the curve increases by one unit. Source: Rudin et al. (2011, 2012).

x-axis, each time there is an arrow on the arrow plot, the y-axis of the pseudo-ROC curve increases by one notch. Thus, the best possible pseudo-ROC curve is one that is as close as possible to the upper left corner. Pseudo-ROC curves are an excellent way to visualize a ranked list, and one who is used to seeing these curves can immediately judge the quality of the ranked list. Often we pay particular attention to the slope at the lower left of the curve, which corresponds to the quality of the ranking near the top of the ranked list. In Figure 9 we show ROC curves for blind tests from the projects on other parts of the power grid, namely the results of hammerhead rankings, primary feeder cable section rankings, transformer rankings, and joint rankings.

Impact in Human-Centered Assessment Tools for Predictive Models: With data as complex as ours, black-box (non-transparent) predictive models are not useful nor convincing. Every step of the processing needs to be able to be verified by both Columbia/MIT scientists and Con Edison engineers. We designed a suite of tools that make our full system transparent, and provide reasons for the predictions made by the

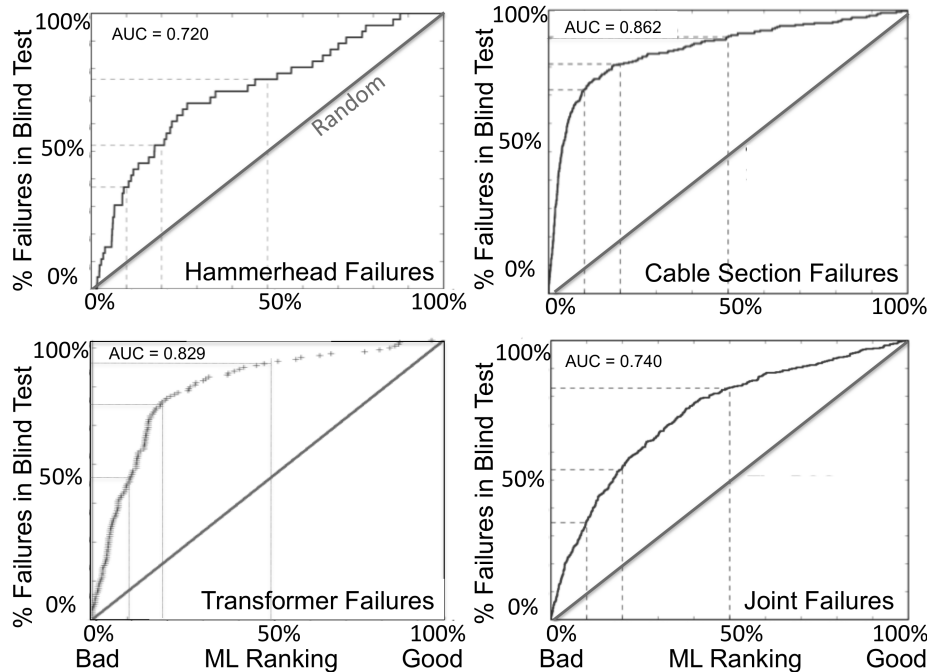


Figure 9 These are pseudo-ROC curves from other ranking projects. Source: Rudin et al. (2012).

system.

Case Study Tool: The “case study” tool is nicknamed the “report card” tool at Con Edison because it produces a full report of the processing that went into the prediction for each manhole. An example case study for a manhole in Manhattan is provided in Figure 10. This is the type of manhole that we would recommend for inspection with high priority. The first part of the report lists the set of events that the manhole has been involved in. This particular manhole, identified as service box 1024358, has a long history of involvement in past events, including 10 events for which it was the source of the problem (the “trouble hole”). For each event listed, the trouble ticket is summarized in the remaining columns.² This includes the *trouble type*, which gives some indication of the seriousness of the event, e.g., SMH for smoking manhole, SO/SOB/SOP for “side off,” “side off bridge,” or “side off power” which means a partial outage, or MHF for manhole fire. The length of each ticket is provided, measured by the number of free text lines typed by dispatchers, and an indicator for the word “shunt,” which means that repair work was performed. This summary of tickets provides an instantaneous view of how “hot” the manhole is, i.e., how

² Some additional columns were removed for the purpose of brevity.

extensive its involvement in past events was. The second part of the case study provides the manhole's inspection results for all of its inspections. This particular manhole had 5 inspections. Almost every time the manhole was inspected, it required some type of repair, either level I (urgent repair), level II (fix within 1 year), or level IV (infrastructure repairs). These inspection results also indicate potential vulnerability for the manhole. The last part of the case study gives the manhole's cable information. This manhole was built gradually, and contains cables from the 1930's, 1940's, 1960's, 1990's, and 2000's, illustrating how the power grid was built gradually over many decades. The case study tool is described in more depth by Radeva et al. (2009).

Visualization Module: Our visualization module is a human-centered tool that flexibly allows humans to navigate through the streets of New York, viewing the manholes, cables, and trouble tickets in any given area as an overlay on Google Earth. Figure 11 provides sample images from the tool, and the architecture behind it is discussed by Dutta et al. (2008). One can navigate to any desired region, and click on each of the trouble tickets in that area to see the history of events.

The visualization tools are useful for decision support, integration of domain expertise into the modeling process, data verification, and system evaluation.

Impact in Decision-Making for Targeted Inspections: The new inspections program has already had a substantial impact in reducing vulnerability levels and increasing public safety in Manhattan, particularly for the most vulnerable class of manholes (the class our ranking models specifically target). Figure 12 illustrates the effect of the inspection and repair program on different categories of manholes in 2008, during the first inspection for most of the manholes. The categories are bins within the ranked list (rank 1 to 5,000, rank 5,001 to 12,000, and so on), matched using a propensity scoring model to eliminate sources of bias. In most of the vulnerability categories, the manholes that have had level I repairs (due to the new inspection program) have had a significantly (and often substantially) reduced probability of having a (serious or non-serious) event. We would like to note in particular the following impact to energy reliability from targeting the highest vulnerability category identified by our model:

The asset id of this structure is: 4395980

strType	strNum	Address	St1	St2
SB	1024358	2311 ALLEN ST	CANAL ST	HESTER ST

List of tickets the structure was mentioned in:

Ticket	Trouble Type	Received Date	Trouble Hole	Free Text Lines	C&R	Shunt	Structure Mentioned
ME09000362	SMH	2009-09-11	*	57		*	*
ME09000303	SMH	2009-09-10	*	43			*
ME09000984	SMH	2009-03-23	*	40		*	*
ME08000122	SMH	2008-11-07	*	24			*
ME05000389	SO	2005-01-09	*	13		*	*
ME04000334	SMH	2004-05-19		72		*	*
ME04000488	SOP	2004-02-02		16			*
ME04000418	SOP	2004-02-01		24			*
ME03000701	SOB	2003-12-07	*	12			*
ME03000636	SO	2003-04-09	*	7			*
ME02000814	SO	2002-03-07		34		*	*
ME02000869	SO	2002-01-21	*	10			*
ME00000764	SO	2000-11-05	*	9		*	*
ME96000781	MHF	1996-09-09	*	30			*

Above, “C&R” means that the remarks contain a variation of “CUT and RACK”. “Shunt” means that the remarks contain a variation of “Shunts” or “Cleared.”

This structure was a trouble hole 10 times.

This structure has Vented cover.

The structure was inspected on the following dates:

Inspection Date	Level1	Level2	Level3	Level4	Reason For Visit
2005-01-25	T	F	F	F	Ad-Hoc Inspection (Incorporated into Routine Work)
2005-01-27	F	F	F	F	Ad-Hoc Inspection (Incorporated into Routine Work)
2009-03-23	F	T	F	F	Ad-Hoc Inspection (Incorporated into Routine Work)
2009-03-24	F	F	F	T	Ad-Hoc Inspection (Incorporated into Routine Work)
2009-09-12	T	T	F	T	Repairs or Follow-up Previous Inspections

Above, Level1 indicates a Level 1 repair was made. Level2 indicates either a temporary Level 2 repair was made or a Level 2 permanent repair must be done within a year, Level3 indicates either a temporary Level 3 repair was made or a Level 3 permanent repair must be done within 3 years, and Level4 is used for reporting, tracking, information and recommendation/suggestion purposes (no time frame is associated).

Here is cable information for the structure:

To ID	From ID	Num Ph Conductors	Ph Size	Conductor Material	Year	Service Type	Length
1261880	4395980	3	500	CU	2004	Main	83
1261880	4395980	3	500	CU	2004	Main	83
4395980	2644070	6	4/0	CU	1931	Main	51
	4395980	3	2	CU	1993	Service	39
	4395980	2	6	CU	missing	Service	45
	4395980	2	6	CU	1949	Service	45
	4395980	3	2	CU	1931	Service	31
	4395980	1	6	CU	1961	St Light	50

The total number of sets connected to the structure is: 8

Figure 10 This is a case study for a manhole in Manhattan’s Chinatown. Basic information for the manhole is followed by historical manhole event information in the vicinity, inspection history for the manhole, and cable contents for the manhole.



Figure 11 These are screenshots from the visualization module. In the upper plot, manholes are indicated by circles and shaded by vulnerability according to the predictive model. In the lower plot, geocoded locations of trouble tickets are indicated on the map, each labeled by trouble type and year.

- From Figure 12, our evidence suggests that for the highest vulnerability category, namely the top 5000 most vulnerable manholes identified by the model, vulnerability to serious and non-serious events was reduced 25% after level I repairs.

Figure 13 illustrates the need for the recommended repairs in the level IV category to be completed. The level IV category of repairs are infrastructure repairs, requiring considerable effort and expense. There is a waiting list of manholes needing level IV repairs,

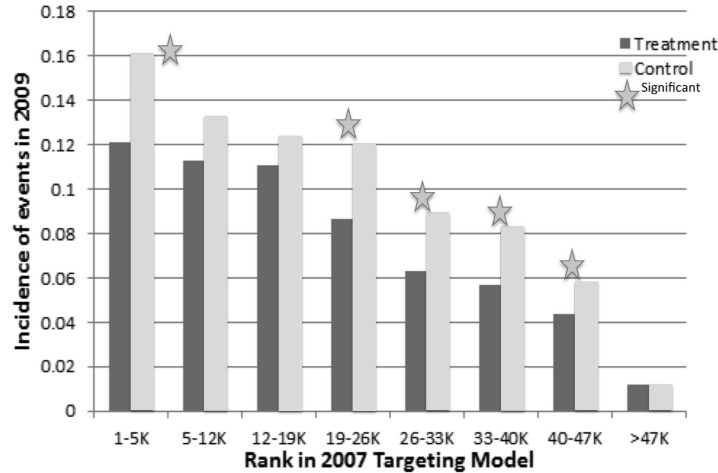


Figure 12 This figure shows rates of events (both serious and non-serious) for each vulnerability category. The manholes that have had level I repairs are generally less vulnerable than manholes that have not had these repairs.

and it is important for the repairs to be ordered according to necessity on this list. This figure shows a matched pairs comparison within each vulnerability category, where each manhole that needed a level IV repair was propensity matched with a manhole that did not need one. The height of the bar represents the number of manholes in each category that had a serious event after its inspection. The number of manholes that had serious events was higher when a level IV was needed, across all categories, and significantly higher in some categories. In particular, we know that the level IV repairs identified during the inspections in the top vulnerability category are important to reliability and safety:

- From Figure 13, we can see that for the highest vulnerability category, manholes that were identified by the inspection program as needing a level IV repair were over one and a half times more vulnerable to serious events than manholes that were not identified.

Beyond identifying specific manholes to be targeted for repair and inspection, we are able to assist with general policy decisions. In particular, we are able to quantify the effect of any inspection policy that Con Edison would potentially use in the future, which allows us to optimize the inspection policy. To quantify the effect of a given inspections policy, our long and short term prediction models are used to simulate the inspection policy over a 20 year period, showing the forecasted distribution of the number of events and number of inspections over that time. In Figure 14 we show preliminary results from a simulation incorporating “bright line” policies, or policies where each manhole is inspected every x years, where x is given on the x-axis of the plot. The projected number of serious and

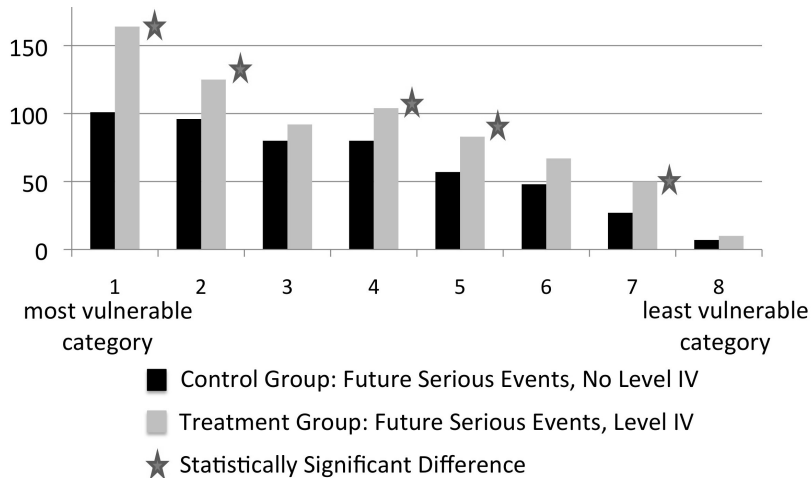


Figure 13 This figure shows the count of matched manholes having serious events after level IV repair was recommended during inspection (gray), and count of matched manholes having serious events after a clean inspection (black). Manholes needing level IV repairs are more vulnerable than those that do not.

non-serious events is given in the left plot for a range of different bright line policies, and the corresponding number of inspections is given in the right plot. There is an inverse relationship between the number of inspections and number of events: more inspections leads to fewer events. These plots permit Con Edison to choose a policy with an optimal trade off between the anticipated number of events and the cost of inspections. For instance, changing from a four year cycle to a six year cycle would lead to an average of ~ 100 fewer events per year, at the cost of an additional ~ 4000 inspections per year.

Without an accurate characterization of the past, provided by the two prediction models, and without the ability to simulate the short term effect of an inspection provided by the RPP model, it would be impossible to assist with decision support in a quantifiable and defensible way.

Impact in Decision-Making for Vented Cover Replacements: The vented cover replacement program has also made a measurable impact in public safety and electrical grid reliability. We calculated that if a manhole is vented, its likelihood of having a serious event is halved (1.58% probability of event per year before vented cover, and 0.74% probability for the same manhole after vented cover, p value 7×10^{-6}). Further, we calculated that if a neighboring manhole in the network is vented, this also reduces the probability of serious event on that manhole by approximately 20%. Thus, in order to distribute the vented

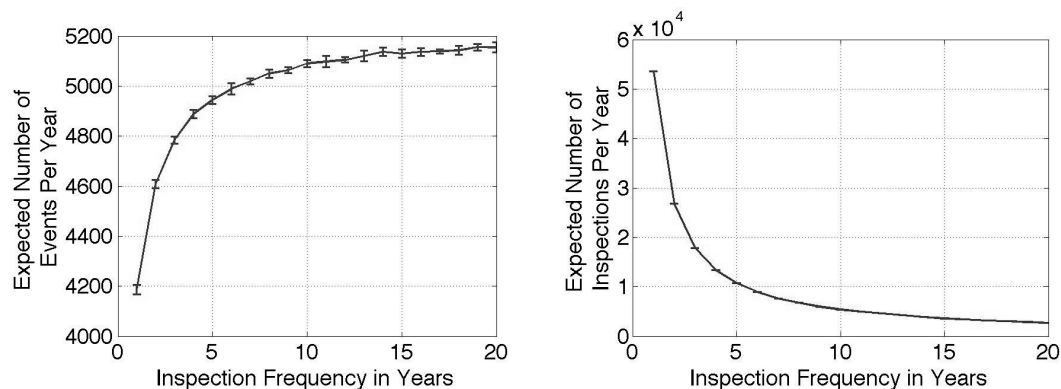


Figure 14 These plots illustrate the preliminary effect of “bright-line” policies. The left plot shows the estimated number of events per year vs. the number of years for the targeted inspection cycle. Longer inspection cycles mean fewer inspections, and more events. The right plot shows the estimated number of inspections per year vs. the number of years for the targeted inspection cycle. Ad-hoc inspections are included in the simulation.

covers, we would like to ensure that, as efficiently as possible, the covers are placed so that for each manhole, either it or one of its nearest neighbors in the underground network are vented.

This is precisely a Minimum Vertex Cover (MVC) problem, and we solved the MVC problem optimally for each of the 38 networks in Manhattan using integer linear programming. Our solution will reduce the time that the whole network is protected, and reduces Con Edison’s immediate effort in Manhattan by 15,865 vented cover replacements.

Future Outlook and Challenges

We believe one of the main practical and far-reaching impacts of this work is to establish the precedent that analytics can be used for electrical grid reliability on a massive scale. The reliability of the power grid affects millions of people every day. New York City possesses both the largest and oldest power system in the world. This system currently serves 8.2 million residents, and upwards of 50 million tourists per year. Electrical grid reliability in New York is critical for the day-to-day functioning of its residents, visitors, financial district, and businesses. We have shown in this pilot effort that it is possible to use analytics to assist in maintaining a grid of this size.

One of the major challenges of this work discussed above was the use of extremely raw historical data. We have demonstrated that with effort, historical company data banks should not be “data tombs,” they can instead be used fully to assist with operations. Con

Edison may be unusual in that it has kept so much historical data over the years, and its engineers have the expertise to interpret and find useful and reliable sources of data. It is not necessarily clear whether data salvaged from other companies can be uniformly harnessed in the same way. We encourage companies (particularly companies with older equipment) to be careful how data are stored in anticipation of its potential future use.

Efforts in preemptive power grid maintenance efforts will become critical in the near future, as we now need to battle the growing threats of increasing levels of failure, and the problem that power grids are reaching the limits of what they are able to handle. Using predictive analytics is a cost-effective and useful way to prioritize these efforts, and can be used generally to achieve a safer and more reliable energy service.

Acknowledgments

This paper is dedicated to the memory of David Waltz, director of the Center for Computational Learning Systems at Columbia University. David helped to establish the infrastructure and collaboration leading to this effort, and devoted the last decade of his life to analytics for energy grid reliability. We would like to additionally thank the MIT Energy Initiative and NSF CAREER IIS-1053407 for partial funding for this work. Finally we thank the judges, particularly Michael Gorman, for organizing the INFORMS Innovations in Analytics Competition, and inviting us to be a part of it.

References

- Amin, Massoud. 2011. Turning the tide on outages. *GreenTechmedia.com*, GridTech: Smart Grid.
- Chehrazi, Naveed, Thomas A Weber. 2011. Dynamic valuation of delinquent credit-card accounts. Submitted.
- ConEdison. 2011. Con Edison Inc. 2011 annual report. Tech. rep. Con Edison Document Center.
- Crane, Riley, Didier Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences* **105**(41) 15649–15653.
- Cunningham, Hamish, Diana Maynard, Kalina Bontcheva, Valentin Tablan. 2002. GATE: A framework and graphical development environment for robust NLP tools and applications. *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*.
- DOE. 2003. Grid 2030: A national vision for electricity's second 100 years. Tech. rep. US Department of Energy, Office of Electric Transmission and Distribution.
- DOE. 2008. The smart grid, an introduction. Tech. rep. Prepared by Litos Strategic Communication. US Department of Energy, Office of Electricity Delivery & Energy Reliability.
- Dutta, Haimonti, Cynthia Rudin, Rebecca Passonneau, Fred Seibel, Nandini Bhardwaj, Axinia Radeva, Zhi An Liu, Steve Ierome, Delfina Isaac. 2008. Visualization of manhole and precursor-type events for the Manhattan electrical distribution system. *Workshop on Geo-Visualization of Dynamics, Movement and Change, 11th AGILE International Conference on Geographic Information Science*. Girona, Spain.

- Egesdal, Mike, Chris Fathauer, Kym Louie, Jeremy Neuman, George Mohler, Erik Lewis. 2010. Statistical and stochastic modeling of gang rivalries in Los Angeles. *SIAM Undergraduate Research Online* **3** 72–94.
- Ertekin, Şeyda, Cynthia Rudin. 2011. On equivalence relationships between classification and ranking algorithms. *Journal of Machine Learning Research* **12** 2905–2929.
- Ertekin, Şeyda, Cynthia Rudin, Tyler McCormick. 2013. Predicting power failures with reactive point processes. *Proceedings of AAAI Late Breaking Track*.
- Gross, Philip, Albert Boulanger, Marta Arias, David L. Waltz, Philip M. Long, Charles Lawson, Roger Anderson, Matthew Koenig, Mark Mastrocinque, William Fairechio, John A. Johnson, Serena Lee, Frank Doherty, Arthur Kressner. 2006. Predicting electricity distribution feeder failures using machine learning susceptibility analysis. *The Eighteenth Conference on Innovative Applications of Artificial Intelligence IAAI-06*. Boston, Massachusetts.
- Krumin, Michael, Inna Reutsky, Shy Shoham. 2010. Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Frontiers in computational neuroscience* **4**.
- Little, Amanda. 2009. Our old electric grid is no match for our new green energy plans. Gridst.org.
- MSNBC. 2003. Blackout costs N.Y. City \$1 billion. MSNBC News: Business with CNBC, Stocks and Economy.
- NERC. 2007. Results of the 2007 survey of reliability issues, revision 1. Tech. rep. North American Electric Reliability Corporation Reliability Assessments & Performance Analysis Program.
- NYBC. 2010. Electricity OUTLOOK: Powering New York City’s economic future. Tech. rep. New York Building Congress Reports: Energy Outlook 2010-2025.
- Ogata, Yoshihiko. 1998. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics* **50**(2) 379–402.
- Oza, Nikunj, J. Patrick Castle, John Stutz. 2009. Classification of aeronautics system health and safety documents. *IEEE Transactions on Systems, Man and Cybernetics, Part C* **39**(6) 670–680.
- Paninski, Liam. 2004. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems* **15**(4) 243–262.
- Passonneau, Rebecca, Cynthia Rudin, Axinia Radeva, Zhi An Liu. 2009. Reducing noise in labels and features for a real world dataset: Application of NLP corpus annotation methods. *Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*.
- Passonneau, Rebecca, Cynthia Rudin, Axinia Radeva, Ashish Tomar, Boyi Xie. 2011. Treatment effect of repairs to an electrical grid: Leveraging a machine learned model of structure vulnerability. *Proceedings of the KDD Workshop on Data Mining Applications in Sustainability (SustKDD), 17th Annual ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

- Radeva, Axinia, Cynthia Rudin, Rebecca Passonneau, Delfina Isaac. 2009. Report cards for manholes: Eliciting expert feedback for a machine learning task. *Proceedings of the International Conference on Machine Learning and Applications (ICMLA'09)*.
- Rhodes, Lisa. 2013. US power grid has issues with reliability. Data Center Knowledge (www.datacenterknowledge.com): Industry Perspectives.
- Rudin, Cynthia. 2009. The P-Norm Push: A simple convex ranking algorithm that concentrates at the top of the list. *Journal of Machine Learning Research* **10** 2233–2271.
- Rudin, Cynthia, Rebecca Passonneau, Axinia Radeva, Haimonti Dutta, Steve Ierome, Delfina Isaac. 2010. A process for predicting manhole events in Manhattan. *Machine Learning* **80** 1–31.
- Rudin, Cynthia, Rebecca Passonneau, Axinia Radeva, Steve Ierome, Delfina Isaac. 2011. 21st-century data miners meet 19th-century electrical cables. *IEEE Computer* **44**(6) 103–105.
- Rudin, Cynthia, David Waltz, Roger N. Anderson, Albert Boulanger, Ansaf Salleb-Aouissi, Maggie Chow, Haimonti Dutta, Philip Gross, Bert Huang, Steve Ierome, Delfina Isaac, Arthur Kressner, Rebecca J. Passonneau, Axinia Radeva, Leon Wu. 2012. Machine learning for the New York City power grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(2) 328–345.
- Siegel, Kathryn. 2012. The true cost of power outages. Yale Environment Review.
- So, Hemmy. 2004. Council approves bill on Con Ed annual inspections. *The Villager*, Volume 74, Number 23.
- Vapnik, Vladimir N. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- Wu, Leon, Gail Kaiser, Cynthia Rudin, Roger Anderson. 2011a. Data quality assurance and performance measurement of data mining for preventive maintenance of power grid. *Proceedings of the KDD Workshop on Data Mining for Service and Maintenance (KDD4Service), 17th Annual ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Wu, Leon, Gail Kaiser, Cynthia Rudin, David Waltz, Roger Anderson, Albert Boulanger, Ansaf Salleb-Aouissi, Haimonti Dutta, , Manoj Poolery. 2011b. Evaluating machine learning for improving power grid reliability. *Proceedings of the ICML 2011 workshop on Machine Learning for Global Challenges, International Conference on Machine Learning*.
- Wu, Leon, Timothy Teräväinen, Gail Kaiser, Roger Anderson, Albert Boulanger, Cynthia Rudin. 2011c. Estimation of system reliability using a semiparametric model. *Proceedings of IEEE EnergyTech*.

Appendix A: Long Term Vulnerability Modeling of Manholes with Rank Statistics

For the problem of learning-to-rank the long-term vulnerability of manholes, we developed the P-Norm Push (Rudin 2009), which is a supervised ranking algorithm that tries specifically to optimize the fidelity of results at the top of the ranked list.

In supervised ranking methods, a training set of labeled data from the past is used to create a ranking model, and that ranking model is used to predict the future. The ranking model gives a score to each manhole, so that the more vulnerable manholes are closer to the top of the list.

Formally, we are given examples $\{(x_i, y_i)\}_{i=1\dots m}$. Here, $x_i \in \mathcal{X}$, where $\mathcal{X} \subseteq \mathcal{R}^d$ is the space of manholes during a timeframe, represented by d features characterizing its state in the past. Examples of feature values x_{ij} for manhole i are:

- the number of serious events where manhole i was the trouble hole within the timeframe.
- 1 if the manhole has a vented cover and 0 otherwise.
- the number of main phase cables within manhole i .
- the number of service cables from the 1930's within manhole i .
- 1 if manhole i had a serious event, and a clean inspection afterwards, with no other events or inspections since that time.

There are hundreds of features that have been developed for the ranking model. The labels $y_i \in \{-1, +1\}$ indicate whether the manhole had a serious event within a year after the timeframe. For instance, if the features were designed using times until 2012, then the label would be 1 if the manhole had an event in 2012 and -1 otherwise. That model, which is built using times up to and including 2012, would be then used to predict events in 2013. In particular, a test set (out-of-sample set) of $\{\tilde{x}_i\}_i$ values are derived from data up to and including 2012, and the labels for 2013 will be predicted.

The ranking model is a linear combination of the features, with coefficients $\{\lambda_j\}_{j=1}^d$ that need to be learned from the training data:

$$f_\lambda(x) = \sum_{j=1}^d \lambda_j x_{.j}.$$

To train the model, the P-Norm Push algorithm minimizes the following objective:

$$R_p(\lambda) := \sum_{k:y_k=-1} \left(\sum_{i:y_i=1} e^{-[f_\lambda(x_i) - f_\lambda(x_k)]} \right)^p,$$

This objective is a rank statistic. It is convex, smooth, and thus can be minimized efficiently. The model f_λ trained through this procedure now characterizes the training data, and can

be used to predict the labels on the unlabeled examples \tilde{x}_i . In particular, the predicted label is $\hat{y}_i = \text{sign} f_\lambda(\tilde{x}_i)$.

The user-chosen power p is the parameter p of an ℓ_p norm, which allows the model to focus on the top of a ranked list, similar to the way an ℓ_p norm focuses on the largest elements of a vector. By increasing p , one changes how much the algorithm concentrates on “pushing” high scoring negative examples (i.e. non-vulnerable manholes) down from the top of the list, leaving the vulnerable manholes higher on the list.

Appendix B: Reactive Point Processes

“Reactive Point Processes” (RPP’s) is a new statistical model that takes into account several important properties of vulnerability over time, including: i) self-excitation, where a past manhole event causes vulnerability levels to increase and then fade gradually back to the baseline level, ii) self-regulation, where an inspection and repair causes the vulnerability level to decrease and then fade gradually back to the baseline level, iii) saturation, where vulnerability levels cannot go too far from the baseline level (several inspections in a row do not continue to cause the vulnerability level to go down to zero for instance), and iv) a shift in baseline levels due to at least one past event. A key component of this model is an intensity function, which describes the instantaneous risk of an event for a given manhole, that changes through time in response to the event and inspection history of the manhole. Specifically, we assume events follow a non-homogenous Poisson process for observation j with intensity $\lambda_j(t)$ as follows (Ertekin et al. 2013):

$$\lambda_j(t) = \lambda_0 \left[1 + g_1 \left(\sum_{\forall t_e < t} g_2(t - t_e) \right) - g_3 \left(\sum_{\forall \bar{t}_i < t} g_4(t - \bar{t}_i) \right) + c_1 \mathbf{1}_{[\text{Count}(t_e)] \geq 1} \right].$$

where t_e are event times and \bar{t}_i are inspection times. The term containing g_1 and g_2 is for self-excitation, where the function g_2 increases the propensity for a manhole to have an event if an event has happened in the recent past. The term containing g_3 and g_4 encodes self-regulation, temporarily decreasing the vulnerability of manholes after inspections. Both of these terms are compound functions with saturation functions, g_1 and g_3 , imposing a maximum increase (or decrease) in vulnerability from the sum of the g_2 or g_4 terms, respectively. The saturation function captures, for example, the diminishing returns to subsequent, frequent inspections. The c_1 term adjusts the baseline vulnerability by a

constant if the number of past events, $\text{Count}(t_e)$ is at least one. That is, once a manhole has its first event, its baseline level shifts upwards permanently if $c_1 > 0$. Using this combination, we capture relevant vulnerability changes, including vulnerability spikes due to previous events, vulnerability drops due to previous repairs, and longer-term changes in baseline levels. This allows us to predict when future events are more likely to occur in the short term. Figure 15 shows an example of the vulnerability of a simulated manhole changing over time according to the model, where events are marked in red and repairs are marked in blue. Figure 16 shows functions g_2 and g_1 for Manhattan, which were traced over time for the Con Edison manholes. This figure nicely shows how well the RPP model fits vulnerability profiles of manholes in Manhattan. The baseline vulnerability level λ_0 is obtained from the long-term vulnerability model.

RPP's are a general statistical innovation, motivated by this project, that can have widespread use. RPP's are much more general than self-exciting point processes (Ogata 1998), which have been proposed for a number of applications in a wide range of fields including earthquake modeling (Ogata 1998), criminology (Egesdal et al. 2010), neuroscience (Krumin et al. 2010, Paninski 2004), credit card delinquency (Chehrazi and Weber 2011), and video viewing behavior (Crane and Sornette 2008). Self-exciting point processes capture only the correlation between an observation's current vulnerability and events that have happened in the past, or the g_2 function in the RPP model. We expect that RPP's will be useful in a range of domains where external means are available to reduce vulnerability, for example, in reliability modeling or in predicting patient outcomes in observational healthcare data.

The RPP model is used to simulate the effect of many different inspection policies, as it gives a *quantifiable estimate of the impact of each policy*, as discussed in the main text.

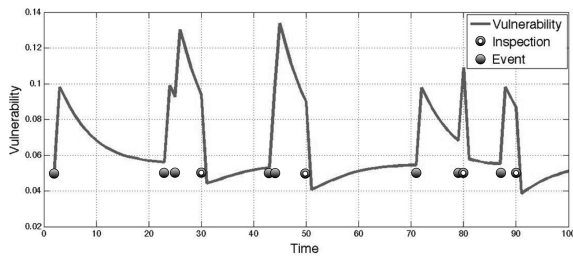


Figure 15 This curve illustrates a manhole’s vulnerability over time according to the Reactive Point Process model. Filled in circles indicate events (spikes in vulnerability), while rings indicate inspections (drops in vulnerability).

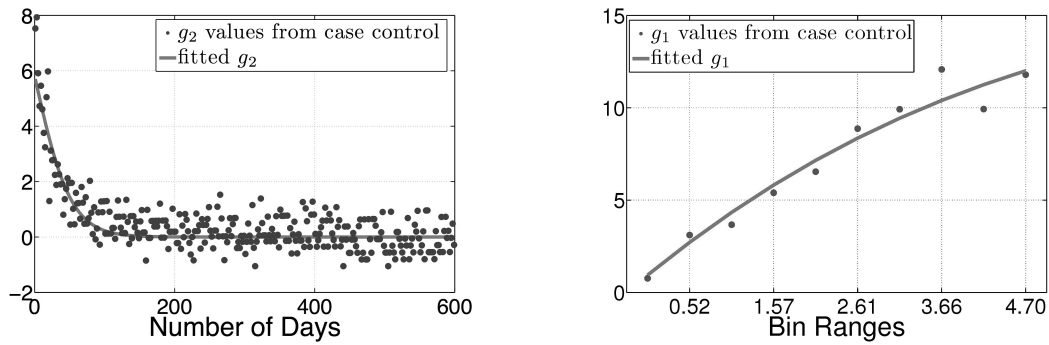


Figure 16 This figure shows nonparametric estimates for g_2 (left) and g_1 (right) for manholes in Manhattan. The RPP model fits the Manhattan data very well.