Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

COULD SIMULATION OPTIMIZATION HAVE PREVENTED 2012 CENTRAL FLORIDA ELECTION LINES?

Jingsheng Li

Digital Manufacture Research Lab
Beijing Institute of Technology
Haidian, Beijing 100081 CHINA

Theodore T. Allen
Kimiebi Akah

Integrated Systems Engineering
The Ohio State University
Columbus, OH 43210 USA

ABSTRACT

In this article, we attempt to simulate the election lines in four central Florida counties in the 2012 presidential election. To do this, we estimate the numbers of booths at all locations and the service times using data about poll closing times and numbers of ballot items at all 479 locations. Then, we investigate the relevance of an optimization formulation in which the maximum expected waiting time at all locations is minimized by reapportioning voting booth resources. We solve the formulation using a heuristic from the literature and (tentatively) conclude that, according to our estimates and assumptions, none of the locations would have been expected to close after 9:50 pm if simulation optimization had been applied to allocate voting booths. Further, our model indicates that, by applying simulation optimization compared with proportion-al allocation, the expected latest poll closing time reduces from approximately 6.8 hours to less than 2.5 hours after closing time.

1 INTRODUCTION

Allen (2013a 2013b) estimated that over 200,000 people were deterred from voting across Florida in 2012 because of the peoples' awareness of the local waiting line conditions. This estimate was based on the fact that turnout percentages were lower in locations having longer waiting lines. Since no new voters were allowed to enter the line after 7 pm, the time that the poll closed offers an estimate of the time the last voter needed to wait and vote. Figure 1 shows the percentage of eligible voters who voted at the 479 locations in the four central Florida counties plotted against the poll closing time. Apparent in the plot is the downward trend resulting in the estimate that 2% of eligible voters were lost (on average) for every hour that the polls stayed open late generating the estimated number of deterred voters in Allen (2013a).

Of course, the state has already invested millions of dollars in permitting voters to vote before Election Day, paying workers, and purchasing machines. If millions more were spent, the lines could clearly be eliminated in future elections. Yet, an interesting question for the simulation community is whether the lines could have been prevented not through purchasing additional resources but simply by reallocating the available resources. Note that the situation seems similar to the case of central Ohio in 2012. For that case, Allen and Bernshteyn (2012) show a similar plot for Figure 1 and offer remedies based on queuing approximations to address the variable ballot lengths and service times.

Operations research and simulation optimization have a long history related to allocating resources across systems in parallel. For possibly useful allocation methods see, e.g., Koopnmab (1953), Köchel (2003), Yoshimur and Fujimi (2006), Frazier and Kazachkov (2011), Ahmadbeygi and Cohn (2010), and Yang et al. (2009). Thus, we reference only the papers are most related to the voting machine allocation including Yang et al. (2013) studied several voting resource allocation formulations and generated heuristic solution method as well as rigorous bounds on solution quality.

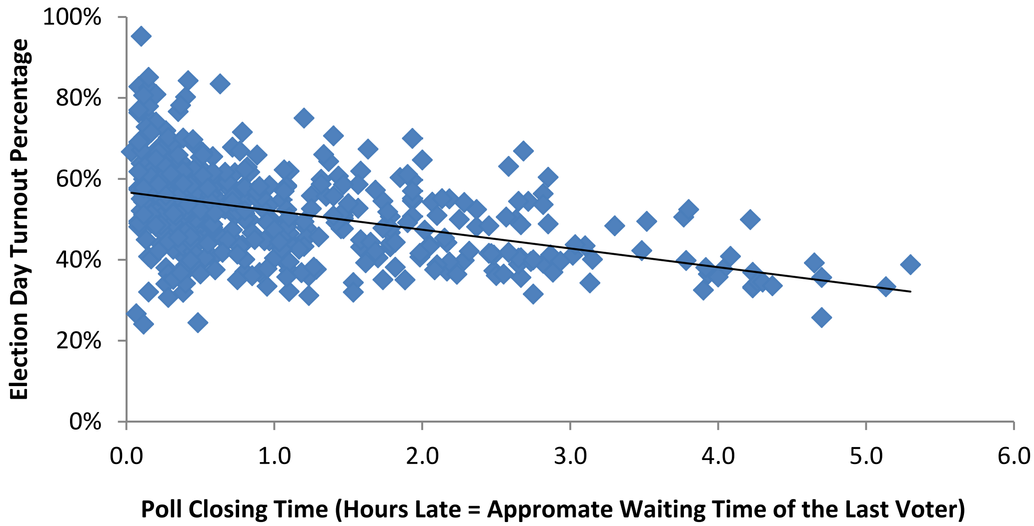

Figure 1: The hours late that the polls stayed open versus the Election Day turnout percentages

In this article, we attempt to recreate all of the inputs needed to apply the minimax formulation from Yang et al. (2013) related to the four central Florida counties. Using the developed simulation model, we seek to fill in details about the likely in the 2012 election. Also, using the optimization heuristic, we seek to explore what hypothetically might have occurred if resources had been reallocated following the recommendations from the Yang et al. (2013) had been applied.

Recreating the 2012 election lines is a considerable challenge because we do not have access currently to key inputs for the simulation. Specifically, we are missing:

- The number of key voting resources at each location (voting booths) and
- The service time distributions at each location.

Clearly, the majority of discrete event simulation applications require the availability of these data.

However, we do have data which have an indirect bearing on the resources and service times. The data that we have include:

- The poll closing times which (as mentioned previously) permit estimation of the waiting times and
- The number of races, issues, and referenda at each location.

Allen (2013a and 2013b) has emphasized the importance of ballot length in allocation because of the common practice of provisioning resources based solely on the number of eligible voters. Yet, with some ballots at certain locations in central Florida in 2012, the sum of the number of races plus issues plus referenda equaled 24 and in others it equaled 36 unique items to interpret and vote on. Possibly, the voters in the locations with the shorter ballots (24 items) had less than half the average time monopolizing the voting booths than the voters in the locations with 36 items.

In Section 2, we explore the evidence that the number of voting booths was the bottleneck in the election system. Section 3 uses a simulation model based on the assumption that voting booths were the bottleneck. Using a full factorial experiment, we attempt to estimate the needed numbers of voting booths and also the service time distribution parameters. In Section 4, we review the minimax optimization formulation from Yang et al. (2013) and the proposed heuristic. Section 5 describes the results from the hypothetical voting booth reallocation and compares the simulation model predictions for both the estimated actual allocation and the reallocation. In Section 6, we discuss the limitations of this study and opportunities for further research.

2 IDENTIFYING THE BOTTLENECK

In the United States, there are many types of procedures for voting. Generally, states permit each county to apply a different combination of equipment and process. In addition, often each location within a county often has a distinct ballot having a different number of races, issues, and referenda compared with other locations. While some states have relatively few ballot initiatives, states such as California, Florida, and Ohio often have long ballots which can require over 20 minutes of reading and processing for voters to cast their ballots once they reach the part of the process in which their inputs are recorded.

Here, we focus on the four counties in central Florida: Lake, Orange, Osceola, and Seminole. In our understanding, all four counties use the process flowcharted in Figure 2. The voters arrive and queue. Then, they register, providing identification. The registration desk gives the voter the relevant pre-printed paper ballot. When a voting booth becomes available, the voter fills out the ballot using the booth. Next, the voter brings the completed ballot to a scanning machine which scans it into the memory. Therefore, there are several candidates for the system bottleneck including: the registration counter, the voting booth, and the scanning machines. Also, different locations and counties could have different bottlenecks.

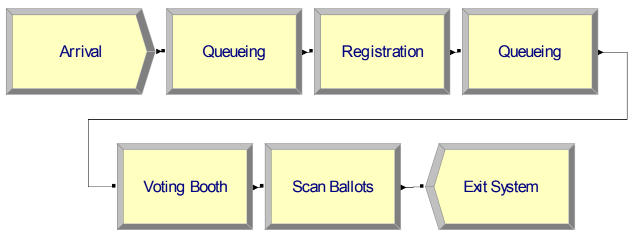

Figure 2: Steps in the central Florida voting system

Figure 3 shows the number of items voted on and the hours late that the polls closed for the 479 locations across the four central Florida counties. As mentioned previously, the hours late offers an estimate of the waiting times at that location. Note that all of the waiting times in excess of 1.7 hours occurred at locations with 29 or more items on the ballot. Since the registration service times and, to a great extent, the scanning service times are reasonably independent of the ballot length, we eliminate these as possible bottlenecks, at least for the locations having the long lines. We therefore assume that the bottlenecks are the voting booths. This follows because, the longer the ballot, the longer the time the voter monopolizes the voting booth. For this reason, our simulation models omit the registration and scanning processes and include only queuing and booth processing times. We will discuss the omission of registration and scanning together with other model limitations in Section 5.

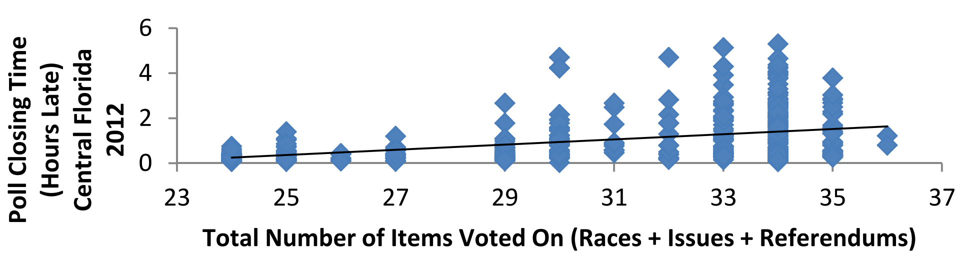

Figure 3: Poll closing times of the 479 locations versus the number of ballot items

3 NUMBERS OF BOOTHS AND ARRIVAL AND SERVICE DISTRIBUTIONS

In much of the previous election related research, the number of resources at each location and the service time distributions were treated as givens (Allen 2013a; Yang et al. 2013; Bernshteyn, 2006). This occurred because the relevant counties in Ohio and New York state used direct recording equipment (DRE) voting machines which were publically documented and/or available to us because of our working relationships with election officials. For central Florida in 2012, the number of voting booths is only known by us in aggregate from the U.S. Federal Election Assistance Commission Election Administration and Voting Survey (2012) and synthesized by Stewart (2012). In this Section, we attempt to estimate the number of booths at all locations in the four central Florida counties. Also, we attempt to estimate number of people arriving intending to vote and the service distributions for these booths, i.e., the distributions of the times required at every location by the voter while monopolizing the booth.

For the voting booths, we assume (with admittedly little information) that the number of booths in use was proportional to the number of eligible voters in each location. This assumption is supported by the pattern in Figure 2 which indicates that insufficient resources were allocated to the locations having the longest ballots. Further, many states have explicit legal provisioning of resources proportional to the number of voters. This occurs despite the fact that ballot lengths vary widely from location to location resulting in systematic effective disenfranchisement (Allen 2013a). We therefore assume that the number of booths, n_i, for location i equaled the number of eligible voters, v_i, multiplied by a factor, α, and rounded down to nearest integer, i.e.:

$$n_i = \lfloor \alpha v_i \rfloor. \tag{1}$$

To estimate α, we refer to Stewart (2012) who referenced the Election Administration and Voter Survey results. Stewart (2012) analyzed the report results and estimated a ratio of 117 actual voters per voting both. Since, the turnout in the four counties of eligible voters was 84%, we derive a ratio of 138 eligible voters per voting booth resulting in α = 1/138 = 0.00725. The resulting numbers of booths are shown in Table 1 for the first ten precincts in Seminole County.

Table 1: The simulation data based on Florida election day

Location	Election Day Eligible Voters	Number Attempting to Vote (Actual + Predicted)	Ballot Length (# Items)	Estimated Number of Booths $\lfloor \alpha v_i \rfloor$.	Mean Service Time (μ_i)	Std. Dev. ($\mu_i \beta$)
SEM001	1232	846	30	9	5.50	1.10
SEM002	1561	1142	29	11	5.30	1.06
SEM003	2185	1385	29	16	5.30	1.06
SEM004	2302	1085	31	17	5.70	1.14
SEM005	2542	1550	30	18	5.50	1.10
SEM006	2678	1159	31	19	5.70	1.14
SEM007	2368	1588	29	17	5.30	1.06
SEM008	925	606	29	7	5.30	1.06
SEM009	1073	649	29	8	5.30	1.06
SEM010	2835	1860	30	21	5.50	1.10

For the turnout, we assume that the number of people who arrived intending to vote equaled the number who actually voted plus 0.02 multiplied by the number of hours the polls closed late and the number of eligible voters. This follows the 2% rule from Allen (2013a) and Allen and Bernshteyn (2006) which is justified by Figure 1. The number attempting to vote is also indicated in Table 1. In our simulation, we distributed the arrivals of these voters over the election day uniformly so that we used a type of constrained Poisson process. Also, we assumed that none of the voters reneged for simplicity.

In addition, we assume (with further admitted arbitrariness) that the booth service times are normally distributed $N(\mu_i, \mu_i \beta)$ for location i where β is a scale factor. Further, we assume that the mean service times are given by the linear equation:

$$\mu_i = r + sq_i \tag{2}$$

Where r is the mean time required to vote on the core 12 races in 2012. Also, s is the mean time needed per item other than the 12 core races and q_i is the number of issues and referenda at location i.

To estimate the parameters β, r, and s, we performed several informal experiments. Our conclusion was that β = 0.20, r = 1.5 minutes, and s = 0.2 minutes/item offers a reasonable fit for the poll closing times. Figure 4 shows the average simulated mean or average poll closing time at the 479 locations verses the actual poll closing times. We do not expect a perfect one-to-one correspondence in part because the simulation is predicting the long run average and the actual times can be regarded as a single replicate. We use 20 replications which requires approximately 10 minutes of run time using a Dell I5-3210 2.5GHz processor for our code which is written in Visual Basic.

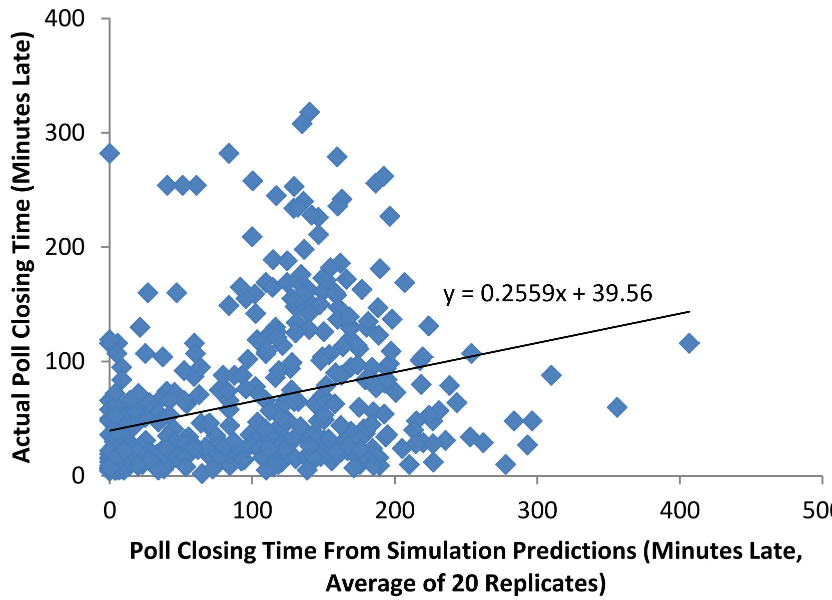

Figure 4: The predicted average poll closing time lateness versus the actual poll closing times

4 SIMULATION OPTIMIZATION REVIEW AND APPLICATION

In the previous section, we described our simulation model which constitutes an attempt to replicate what occurred during the 2012 presidential election in the four central Florida counties. In this section, we review the simulation optimization and heuristic solution method from Yang et al. (2013).

4.1 Min-Max Model

Yang et al. (2013) considered several formulations which each constitutes an attempt to measure and evaluate equity in the election systems context. All of the formulations recommended by the authors address the unequal ballot lengths and make provisions for the service time distribution variability. Perhaps the simplest of these is the so-called "minimax" formulation in which resources are allocated to minimize the maximum over locations of the expected waiting times. Formally, this can be defined:

$$min_{x_i} Z(X) \tag{3}$$

subject to

$$\sum_{i=1}^{N} x_i = M,\ x_i \geq b_i, x_i \in \{1,2,3 \dots\}$$

$$Z(X) = \max_{i \in N} W_i\,(X)$$

where N is the set of all subsystems. M is the total number of available servers. $W_i(X)$ for $i \in N$ is the waiting-time in location i. The sum $\sum_{i=1}^{N} b_i \leq M$ involves $b_i$which is the minimum number of servers required in subsystem i, and x_i is a positive integer. The decision vector is $X = (x_1, x_2, \dots, x_N)'$ so that $x_i (i \in N)$ is the number of resources (booths) allocated to subsystem (location) i. The simulation predicted outputs $w_i(x_i)$ for $i \in N$ are the waiting times at subsystem i given x_i resources (booths).

4.2 Review of Heuristic Solution Method

Yang et al. (2013) proposed a constant sample size greedy heuristic for solving the formulation in equation (3). In this approach, a small number of resources, e.g., 3 booths, is allocated to all locations. Then, an additional resource is added to the location having the highest expected waiting time based on the fixed sample Monte Carlo estimates. The procedure terminates when the number of available resources is exhausted. In our implementation, we used 20 replicates and allocated 3,260 booths across the 479 locations. The run time is approximately one hour using a Dell I5-3210 2.5GHz processor.

5 RESULTS AND LIMITATIONS

In this section, we compare two allocations. The first is based on allocating machines proportional to the number of eligible voters as described in Section 3 and equation, i.e., 138 eligible voters per booth rounded down. The second allocation was derived using the same number of booths (3,260) following the putative minimax optimal allocation derived from the Yang et al. (2013) heuristic as described in Section 4.2. We focus on two waiting time measures which are estimated by our simulation model using 20 replicates: the average or mean poll closing times and the average or mean waiting times.

Figure 5 shows the simulation predicted average poll closing times at the 479 locations for the two allocations. The dotted line is the minutes late using proportional allocation method. Overall, according to the simulation the average hours late across all locations for the proportional allocation method is 92.1 minutes with standard deviation of the mean estimate 2.5. The solid line denotes the minutes late predicted for the minimax putative optimal solution. Across the precincts, the average minutes late for minimax distribution is now 79.8 minutes with standard deviation of the mean equal to 1.7 minutes.

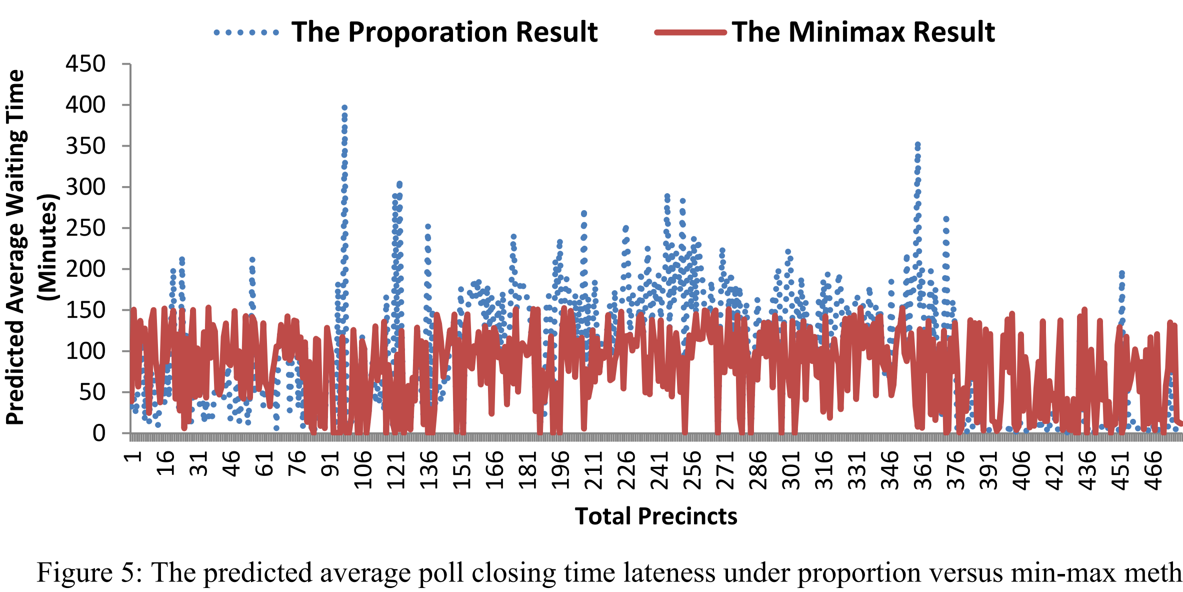

Figure 5: The predicted average poll closing time lateness under proportion versus min-max method

Similarly, Figure 6 compares the estimated average or mean waiting times for each voter at the locations, which the estimated average waiting time is 49.2 minutes and standard deviation of the mean is 0.5 under the proportion method. For the putative minimax solution, the average time and standard deviation is estimated to be 39.0 minutes with standard deviation of the mean equal to 0.5.

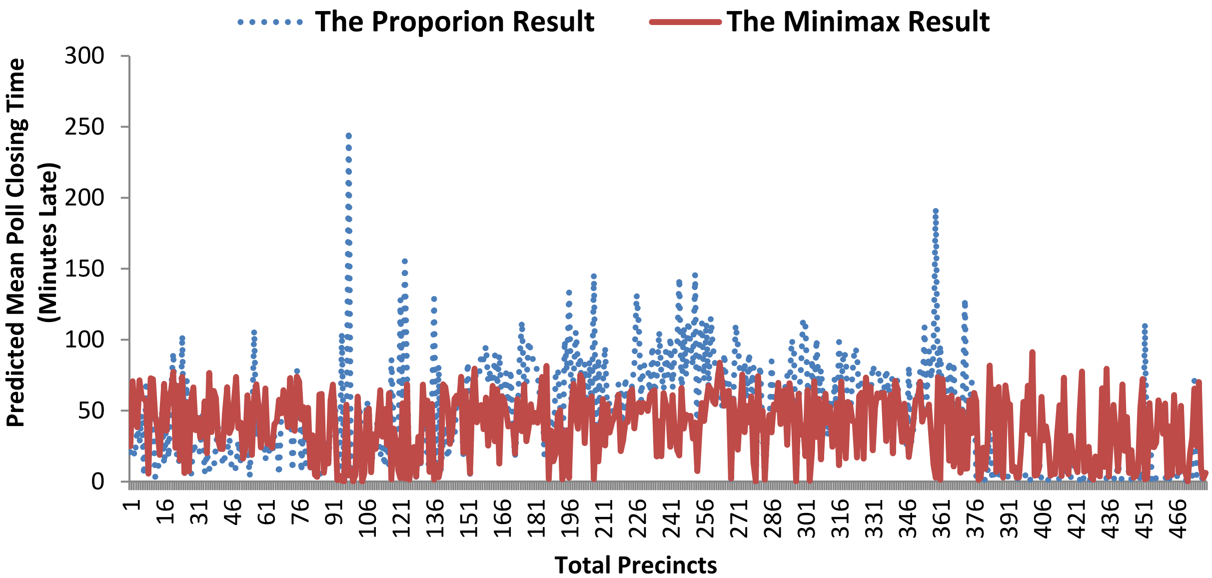

Figure 6: The predicted average waiting time under proportion versus min-max method

With such large estimated mean difference and small estimated standard deviations of the estimates, the results are statistically significant. The simulation model clearly predicts a significant reduction in the waiting times through the reallocation of resources (booths) following the minimax heuristic approach. This is clearly not surprising since the waiting times depend on both the arrival process (number of eligible voters) and service process (ballot length). The proportional method under-provisions for locations with long ballot and, in a relative sense, over provisions for locations with short ballots.

What is surprising, perhaps, is the extend of the reductions in waiting times. The model predicts that, if the minimax allocations had been applied to allocate booths, the issue of waiting times would have been largely erased. This is predicted with no net increase in the number of booths and simply derives from the more "equitable" provisioning of resources derived from solving the formulation in equation (3).

This result comes with substantial limitations deriving from our assumptions. These include all of the following:

- We assume that the bottleneck in all cases in the voting booths. In some locations, registration and/or scanning machines could create bottlenecks which we do not include in model.
- We assume that our estimates for the numbers of booths at the various locations are accurate.
- We assume that the estimated service time distributions (times required to fill out the ballots while monopolizing the booths) are accurate. In the past, we have been able to directly time voters before elections which was not easily possible in this case (Allen and Bernshteyn, 2006).
- We assume that booths can be moved between locations like voting machines without space restrictions.

The seriousness of these limitations explains why we characterize the results of this article as "tentative" pending further data collection and analysis.

Note that all of these limitations are not an unavoidable property of simulation optimization applications. All of them can and have been avoided in simulation optimization projects prior to elections in

Franklin County in 2008 and 2010. In those elections, our team members worked with officials and gathered all of the needed data for using simulation optimization prior to the election. The resulting allocations resulted in line lengths that were likely reduced from what they would have been if simulation optimization had not been applied.

6 DISCUSSION AND FUTURE WORK

In this paper, we attempted to recreate using discrete event simulation the waiting lines in the four central Florida counties during the 2012 presidential election. To do this, we needed to estimate approximately the numbers of available resources (booths) and the service time distributions. Then, we described our application of the formulation and heuristic solution method from Yang et al. (2013). The tentative conclusion is that all of the polls could have been expected to close before 9:50 pm if the booths had been allocated following the minimax heuristic recommended solutions. This conclusion is tentative because it is associated with a list of limitations and assumptions some of which we detailed. Yet, we believe that the benefits of applying simulation optimization are likely important. The common practice of allocating resources proportional to the number of eligible voters is (likely) unavoidably associated for cases in which resources are limited and ballot lengths vary. Virtually any queuing or simulation inspired reallocation that accounts for variable ballot lengths is likely to substantially improve election line performance.

There are a number of opportunities for future work. First, optimization algorithms with proven convergence and greatly improved computational efficiency are needed. The heuristic from Yang et al. (2013) is associated with rigorous bounds on solution quality. Yet, its run times generally exceed one hour and are much higher if hundreds or thousands of replications are used. Convergent methods likely based on variable numbers of replicates could offer election officials efficient software/method solutions and defensible allocations. Second, legislation and/or election procedures which involve simulation optimal allocations before elections are likely needed to avoid lines and deterred voters. Our model predicts the long lines that did occur in the real world in some locations. It also predicts the minimal lines that did occur in other locations. It is perhaps inevitable that these differences will align with demographic patterns and result (again) in systematic discrimination unless allocations that address variable ballot lengths are applied. Yet, new legislation and/or revised procedures can likely avoid waiting lines with minimal additional expenses.

ACKNOWLEDGEMENTS

We thank Muer Yang for his code and support. We also thank the People's Republic of China for supporting Jingsheng Li as a visiting scholar.

REFERENCES

Ahmadbeygi, S., Cohn A., Lapp, M. 2010. Decreasing airline delay propagation by re-allocating scheduled slack. *IIE Transactions*,42, 478-489.

Allen, T. T. 2013a. Analysis: 201,000 in Florida didn't vote because of long lines. Orlando Sentinel http://articles.orlandosentinel.com/2013-01-29/business/os-voter-lines-statewide-20130118_1_long-lines-sentinel-analysis-state-ken-detzner.

Allen, T. T. 2013b. Delving into the reasons for long lines can bring solutions. Orlando Sentinel http://articles.orlandosentinel.com/2013-01-08/news/os-ed-long-lines-voting-florida-010813-20130107_1_long-lines-ballot-length-turnout.

Allen, T. T., M. B. Bernshteyn. 2006b. Optimal voting machine allocation analysis. S. Hertzberg, Analysis of May 2006 Primary Cuyahoga County, Ohio. Election Science Institute, 71-89.

Election Administration Voting Survey 2012.
http://www.eac.gov/research/election_administration_and_voting_survey.aspx.

Frazier, P. I., A. M. Kazachkov. Guessing preferences: a new approach to multi-attribute ranking and selection. In *Proceedings of the 2011 Winter Simulation Conference*. Edited by S. Jain, R. R. Creasey, J.

Himmelspach, K. P. White, and M. Fu, 4325-4336. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
Koopman, B.O. 1953."The optimum distribution of effort." *Operation. Res.,* 1, 52-63.
Köchel, P. 2003. Optimal control of a distributed service system with moving resources: Application to the fleet sizing and allocation problem. *Int. J. Production Economics*, 81-82, 443-459.
Stewart, Stewart, C. 2012. A Back-of-the-Envelope Calculation of Florida's Capacity to Handle Election Day Turnout Without Lines. *Election Updates* http://electionupdates.caltech.edu/2012/11/14/a-back-of-the-envelope-calculation-of-floridas-capacity-to-handle-election-day-turnout-without-lines/.
Yang, M., M. J. Fry, W. D. Kelton. 2009. Are all voting queues created equal? In *Proceedings of the 2009 Winter Simulation Conference*. Edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 3140-3149. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
Yang, M., T. T. Allen, M. Fry, and D. Kelton. 2013. The Call for Equity: Simulation-Optimization Models to Minimize the Range of Waiting Times. *IIE Transactions*, 45(7), 781–795.
Yoshimura, M. and Y. Fujimi. 2006. Decision-making support system for human resource allocation in product development projects. *International Journal of Production Research*, Vol. 44, No. 5, 1 March 2006, 831–848.

AUTHOR BIOGRAPHIES

Jingsheng Li is a visiting scholar in the Department of Integrated Systems Engineering at the Ohio State University. He is a PhD candidate in the Mechanical Engineering at Beijing Institute of Technology, China. His email is li.3746@osu.edu.

Theodore T. Allen is an associate professor in the Department of Integrated Systems Engineering, College of Engineering at The Ohio State University. His interests focus on the overlap of operations research and quality improvement with emphasis on applied statistics and design of experiments (DOE). Allen has worked with numerous companies applying related techniques including as a full-time intern at The Ford Motor Company. He is a fellow of the American Society for Quality and the author of more than 40 peer-reviewed papers including two textbooks. His email address is allen.515@osu.edu.

Kimiebi Akah is an M.S. student at the Integrated System Engineering department at The Ohio State University. He received his B.S. in industrial and systems engineering from The Ohio State University. His interests include discrete event simulation and cyber security. His email address is akah.5@osu.edu.