

Work Area Optimization at a Major European Utility Company

Jinxin Yi and Emily Lada, SAS Institute Inc., Cary, NC

Anne Smith and Colin Gray, SAS Software Limited (United Kingdom)

ABSTRACT

A European utility company has several thousand service engineers who provide its customers with services that range from performing routine maintenance to handling emergency breakdowns. Each service engineer is assigned to a work area that consists of a set of postal sectors. The company wants to understand how it should configure its work areas to improve customer satisfaction, minimize travel time for its full-time service engineers, and minimize the costs of overtime and subcontractor hours. This paper describes the use of SAS/OR[®] optimization procedures to model this problem and configure optimal work areas, and the use of SAS[®] Simulation Studio to simulate how the optimal configurations might satisfy the customer service requirements. The experimental results show that the proposed solution can satisfy customer demand within the desired service-time window, with significantly less travel time for the engineers, and with lower overtime and subcontractor costs.

INTRODUCTION

The service and repair group at a major European utility company (referred to as “SRG” for confidentiality reasons) provides on-site services to customers who call in for help when their furnace, water heater, or other gas-powered appliance needs repair or maintenance. SRG employs several thousand service engineers. The engineers are assigned to specific work areas, although some of them might have to travel to other areas when the need arises. SRG divides its year into winter and summer seasons. The composition of work areas and the assignment of engineers might change from one season to the next.

SRG receives about 10 million calls from its customers each year. There are more than 10 types of calls, ranging from emergency breakdowns to annual maintenance. Each call type is given a target service level that measures how customer calls of that type are served. The service level is defined by two parameters: a time window and a percentage. For example, if a customer calls in because his water heater has broken down, and the call is made before 1 p.m., SRG must dispatch a service engineer to visit the customer the same day 90% of the time during winter. Here the time window is “same day,” and the percentage is “90%.” Because the volume and timing of these calls are fairly random, the actual service level is computed as an average over a summer or winter season.

SRG organizes its entire territory in a hierarchical structure. It splits the territory into regions; each region is made up of subregions; each subregion consists of work areas; and each work area is composed of postal sectors. SRG currently has close to 10,000 postal sectors, making up about 750 work areas. Figure 1 depicts a subregion that contains 14 current work areas. The legend gives the drivetime_max (in minutes) for each area, which is computed as the maximum driving time between any two postal sectors in each area.

PROBLEM STATEMENT

In recent years, SRG has noticed that its engineers spend a significant amount of time serving customers outside their designated work areas. For example, approximately 10% of customer calls in the subregion shown in Figure 1 were covered by engineers from outside. The increased travel time not only reduces engineers’ availability to serve calls in their own areas but also increases overtime and subcontractor hours, which can be quite costly.

There are two primary reasons for bringing outside engineers to cover customer calls. First, SRG does not have enough service engineers to handle all customer calls. Second, there is a mismatch between the skills of engineers and the skills required by customer calls. Figure 2 gives fill rates and utilization rates of the 14 work areas shown in Figure 1. The fill rate measures how calls in a certain work area are served by engineers who are assigned to the area. The utilization rate measures how engineers are used on average. Four of the areas, which are marked with red circles, have fill rates of less than 90%. This means that service engineers in these four areas were not able to handle all the customer calls—and SRG had to use outside engineers, overtime, and subcontractors to meet its contractual agreements. It is interesting to see area D, which has a 70% fill rate and an 85% utilization rate. These rates suggest that the engineers’ skills do not match the skills needed to handle the customer calls in this work area.

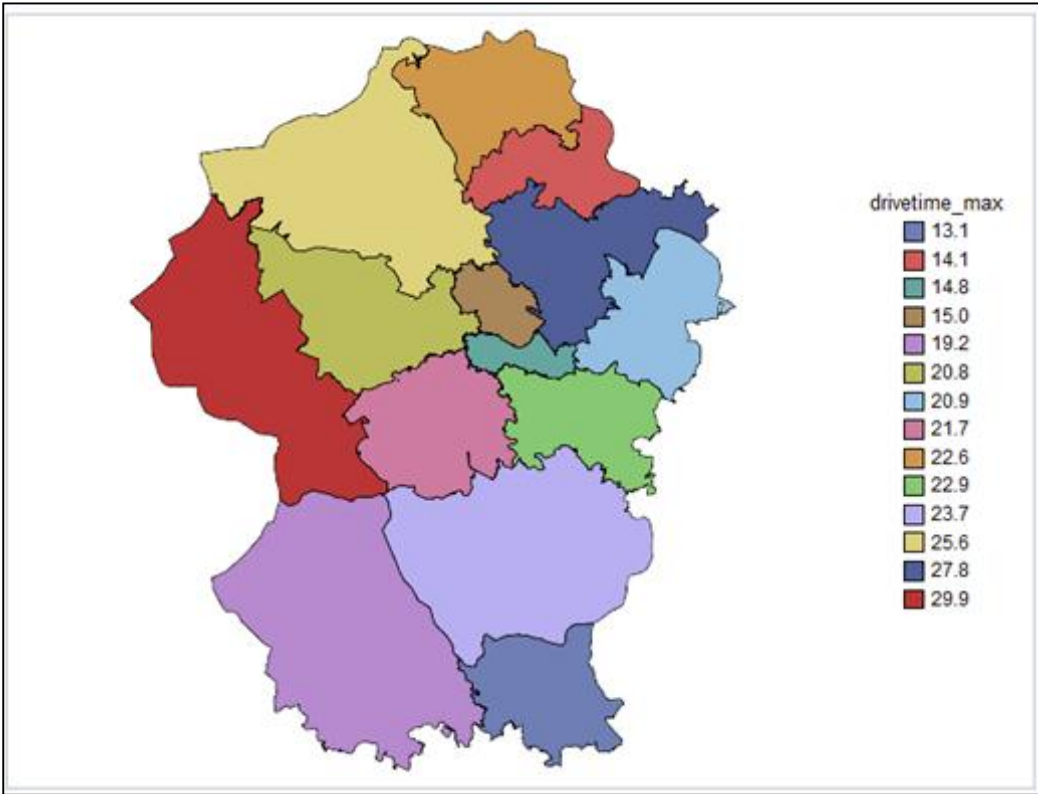


Figure 1. A Subregion of 14 Work Areas

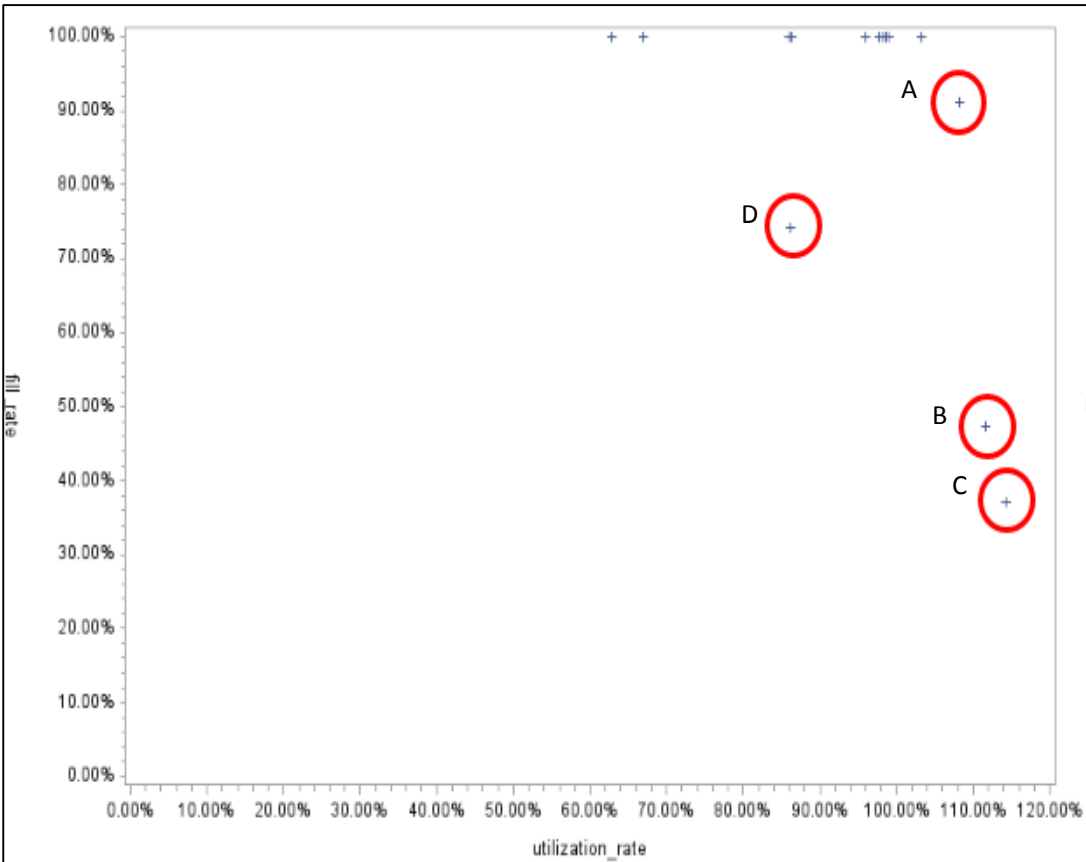


Figure 2. Fill Rate and Utilization Rate of a Subregion

THE SOLUTION

The primary goal of the project for SRG is to find a better composition of its work areas in order to minimize the total travel time of its service engineers. The secondary goal is to minimize the costs of overtime and subcontractors. SAS/OR is used to find the optimal configuration of work areas. There are four steps in finding the optimal configuration.

STEP 1: GENERATE POTENTIAL WORK AREAS

The first step is to generate potential work areas. Several factors determine the makeup of a work area:

- A work area must be a continuous geographical area without “holes.” In other words, a work area cannot fall completely inside another work area.
- A work area must be made up of postal sectors that belong to the same subregion.
- The drive time between any pair of postal sectors within a work area must not exceed 55 minutes.

To generate a potential work area, you pick a postal sector and add adjacent sectors one at a time until no more sectors can be added according to the preceding constraints. The process is repeated at each sector. Other methods of identifying potential candidate work areas have been tested; this method has been chosen because of its simplicity and its strong ability to produce good candidates. The quality of potential work areas is discussed in the next step.

STEP 2: EVALUATE POTENTIAL WORK AREAS

After potential work areas are generated, they are evaluated based on how customer calls can be served by engineers. The calls from customers in each work area can be thought of as “demand” and the work hours of service engineers in the area as “supply.” The volume of calls of each type is estimated from historical data. The volume is then transformed into engineer hours needed based on target service levels. The work hours of engineers are computed based on their work schedule.

A mismatch occurs when demand and supply do not perfectly align. It is important to point out that skill sets play a critical role here. Each call type requires a specific skill from an engineer, but not every engineer has the complete skill set. So the mismatch can also result from a lack of skills in engineers. The mismatch is the primary reason why engineers travel outside their work areas to serve calls. It is quantified as the aggregated shortage of engineer hours.

A linear programming model is developed to find the best way to match demand with supply in order to minimize aggregated shortage. For example, the aggregated shortage for the work areas in Figure 1 is about 38 hours on a daily basis. The model also minimizes overtime and subcontractor costs as a secondary objective. For example, the work areas in Figure 1 require 56 hours of overtime and 5 hours of subcontractor time.

STEP 3: CHOOSE A SUBSET OF WORK AREAS

Currently there are about 750 work areas. The number of potential work areas, however, could run into the hundreds of thousands. The third step is to select the optimal subset of potential work areas that cover SRG’s entire territory with the least amount of shortage. A mixed integer programming model is developed for this purpose. Figure 3 shows the new and optimal design of the work areas. Compared with the current number of work areas (in Figure 1), the new design has only 12 work areas.

The shortage in the new design is 9.7 hours, which is significantly less than in the current design. The overtime is about 60 hours, and no subcontractor hours are needed. You can see similar improvements across SRG’s entire service territory. This is the result of a better alignment of service engineers and customer calls, which is a more effective way to use the regular and overtime hours of the company’s full-time engineers.

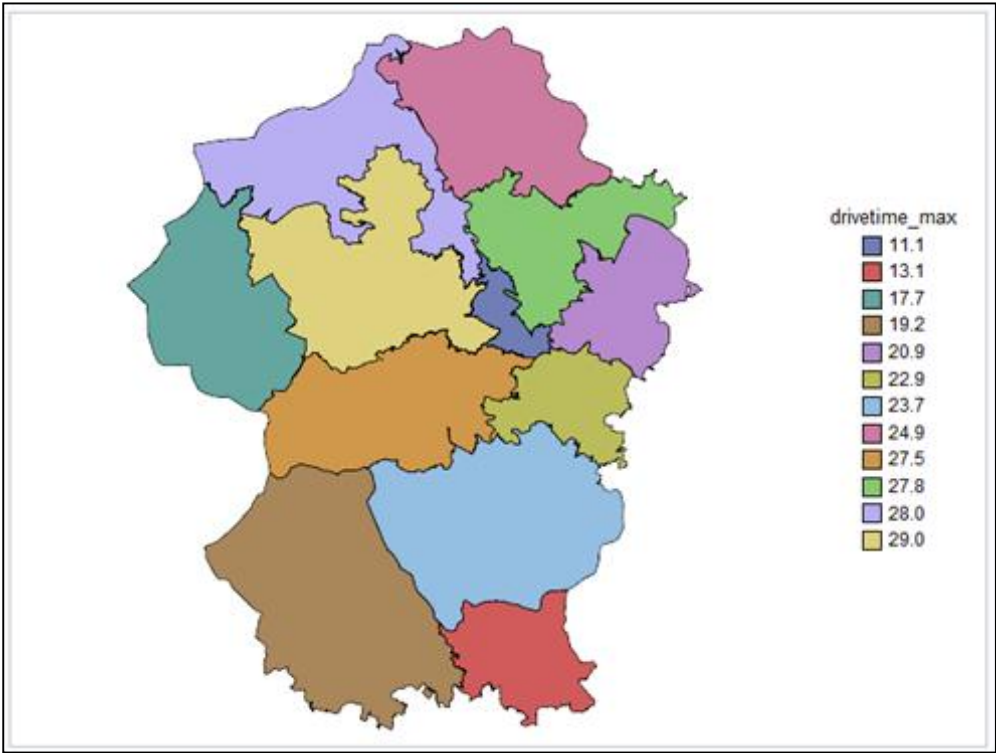


Figure 3. Optimal Composition of Work Areas

STEP 4: BUILD A SIMULATION MODEL

The fourth step is to build a discrete-event simulation model to validate the optimally configured work areas from step 3. The model simulates an arrival pattern of customer service calls over a six-month period (that is, a winter or summer season) and the assignment of those calls to a service engineer. The output data from the model can be analyzed to determine whether calls of each service level type are completed within the desired time window. The model can also be used to estimate the travel time of the engineers. Figure 4 is an overview of the modeled process flow. This process flow was translated into a discrete-event simulation model by using SAS Simulation Studio.

Figure 5 shows a high-level view of the model in SAS Simulation Studio. Each yellow square in the model is a compound block that can be opened to display the Simulation Studio logic contained within. For additional information about SAS Simulation Studio, see the SAS Simulation Studio documentation (SAS Institute Inc. 2013).

The inputs to the simulation model are in the form of SAS data sets and include drive-time data and the engineer data for an optimally configured work area. In the bottom part of Figure 5, the SAS Simulation Studio Dataset Holder block is used to hold the drive-time data set in memory so that it can be queried for each call. The entities (or objects flowing through the model) represent customer calls and service engineers. The arrival pattern of calls is modeled using a separate homogeneous Poisson process for each service type so that the time between calls of each type is exponentially distributed. The call entities have attributes, or properties, that include call type, arrival time, and priority.

The engineer entities are generated at time zero for a specific work area, and their attributes (read in from a SAS data set) include starting location, hours worked, and skill set. The engineer entities then wait in a queue (ordered by hours worked) until a call entity arrives. When a call entity is generated, it is sent to a queue where calls are prioritized according to the call type, arrival time, and desired time window. The model then attempts to find an available engineer entity with the skill required by the incoming call entity. If no such engineer entity is available, then the call entity waits. Similarly, after an engineer entity completes a call and returns to the queue of available engineer entities, it checks the call queue to see if there is a call waiting that requires any of its skills.

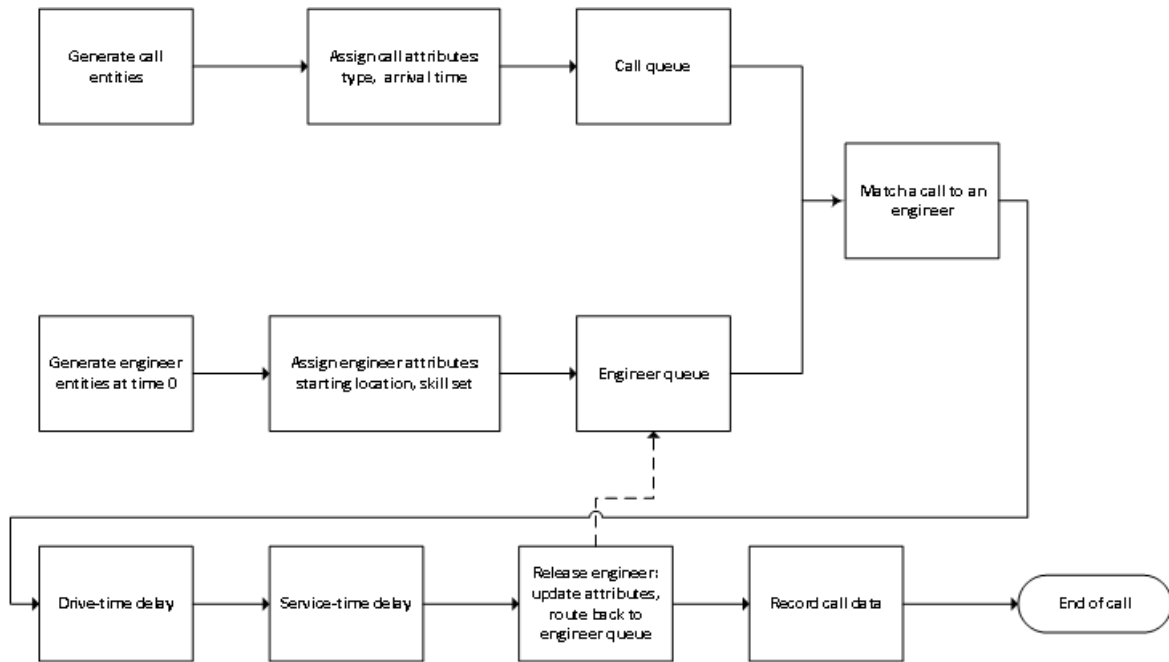


Figure 4. Simulation Model Process Flow

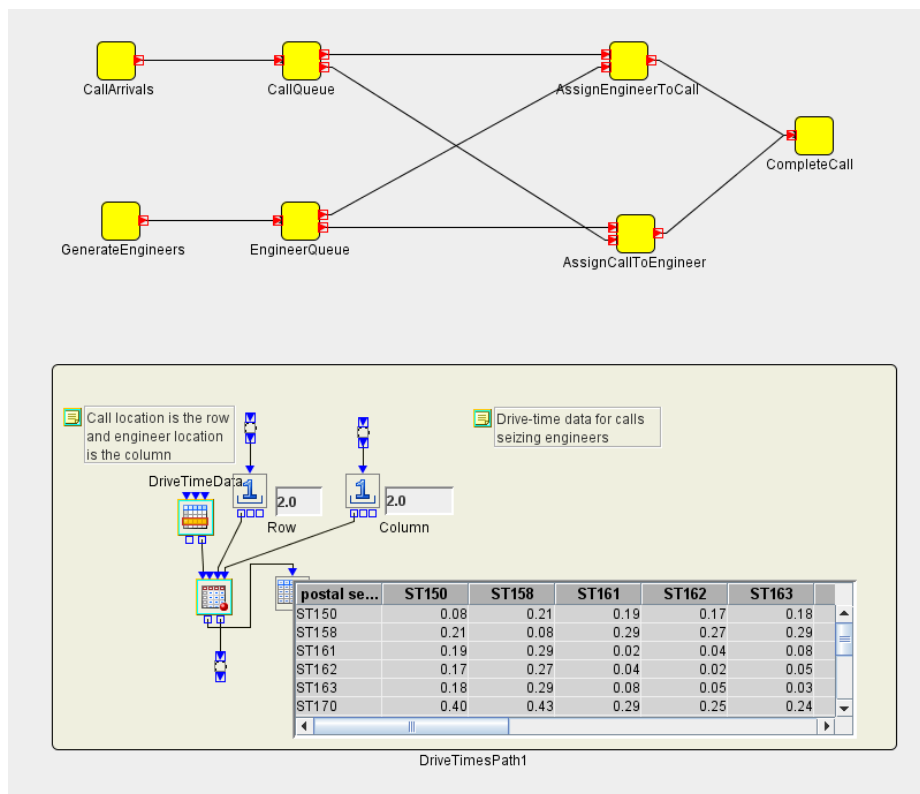


Figure 5. High-Level Overview of SAS Simulation Studio Model

When a call entity is matched to an available engineer entity, a drive-time delay is computed based on the current locations of the call and the engineer, and then a service-time delay is determined based on the type of call. When the drive-time and service-time delays elapse, the call entity releases the engineer entity, the number of hours worked for the engineer entity is updated, and the engineer entity is routed back to the queue of available engineer entities. To mark the end of the call, the completion time is recorded, all call attribute data are written to a SAS data set (including assigned engineer, drive time, service time, and call completion time), and then the call entity is discarded (deleted). Figure 6 shows the expanded CompleteCall compound block in the SAS Simulation Studio model where this logic is executed. When the simulation is finished executing, the recorded call data are analyzed to compute the percentage of calls of each type that fall within the desired completion time window.

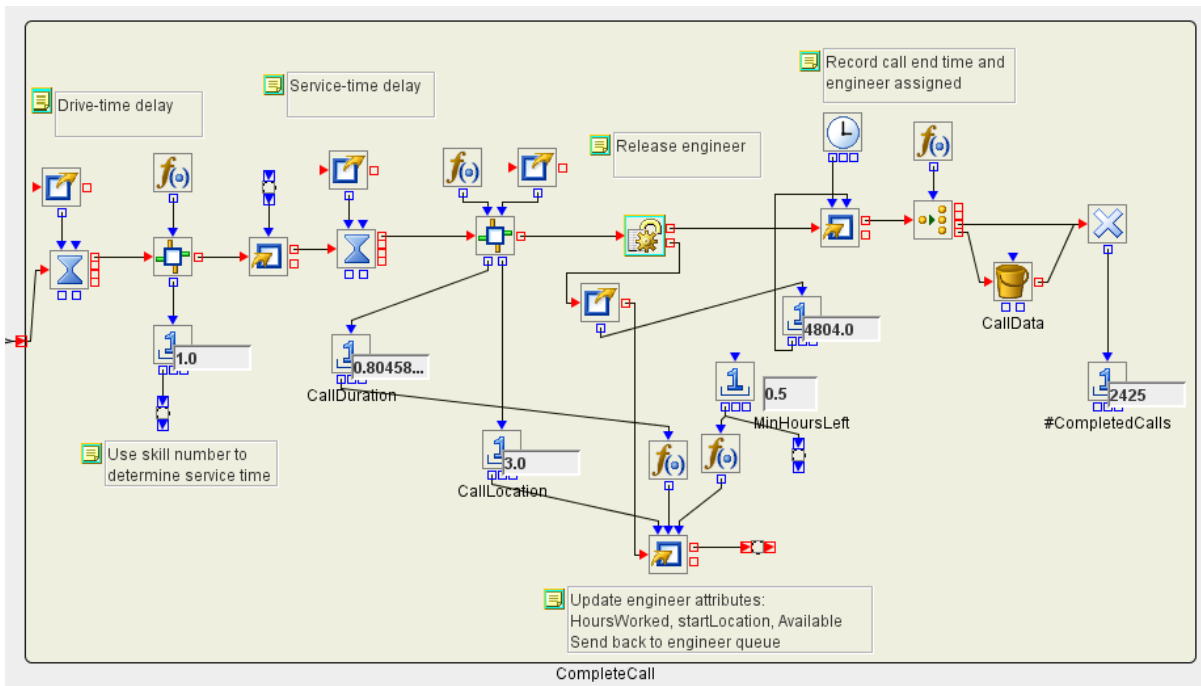


Figure 6. Expanded CompleteCall Compound Block

PROJECTED BENEFITS OF OPTIMIZATION

The primary focus of this project was to minimize drive time. Reducing drive time between calls is expected to increase the number of calls that can be completed, with resulting improvement in meeting service levels and customer satisfaction. Additional benefits are expected in the ability to meet call demand with less use of overtime and subcontractors, hence reducing costs.

The situation in the past has seen a large number of service engineers passing each other as they respond to calls outside their allocated work areas. When the call demand is more closely matched to the engineer supply within each work area, it is less likely that engineers will need to travel long distances to customers in other areas. In some cases, the optimized size of the work areas is larger than current areas, but because the engineers stay in their own area more often, the overall drive time is reduced. In addition, the modeling indicates expected reductions of up to 20% in overtime and up to 50% fewer subcontractor hours.

The optimization results suggest average drive-time reductions of up to 13%. Although additional factors not incorporated in the modeling might affect the results, implementation is expected to provide significant benefits. Initial pilot studies have confirmed that drive times have in fact been reduced.

REFERENCES

SAS Institute Inc. 2013. *SAS/OR 13.1: User's Guide*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2013. *SAS Simulation Studio 13.1: User's Guide*. Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the following author:

Jinxin Yi
 SAS Institute
 100 SAS Campus Drive
 Cary, NC 27513
 Phone: 919-531-9276
 Fax: 919-677-4444
 Jinxin.Yi@sas.com
www.sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.