

Paper 297-2009

Using the SAS/OR® OPTMODEL Procedure to Assign Students to Schools in the Wake County Public School System

Ivan Oliveira, Rob Pratt, SAS Institute Inc., Cary NC
Charles Dulaney, Wake County Public School System, Raleigh NC

ABSTRACT

With over 137,000 students and 156 schools, the Wake County Public School System (WCPSS) is the largest in North Carolina and the eighteenth largest in the United States. Each year, WCPSS faces the problem of assigning students to schools, using multiyear student population forecasts. The main goals are to reduce county-wide: socioeconomic imbalance among neighborhood schools, overcrowding, travel time, and the number of students reassigned to different schools from year to year.

These goals must be achieved while adhering to numerous constraints, which include limits on student demographics by region, limits on school-specific capacities, and the observance of predetermined special assignments. WCPSS needed to automate this previously manual assignment process in order to efficiently evaluate different policies and decisions and quickly propose new assignments.

Using the SAS/OR OPTMODEL procedure, we developed an optimization engine to solve this problem. We used SAS® Grid Manager to generate multiple solutions for analysis and SAS/GIS® to visualize the recommended assignments on county geographical maps. Throughout the analysis, we used JMP® for design of experiments and visualization of optimization results. This presentation explores the formulation and solutions of our approach.

INTRODUCTION

Every year, WCPSS faces the task of deciding how to assign students to schools. To simplify both the problem and the resulting solution, WCPSS has partitioned the county into 1,302 geographical areas called nodes. Each node, rather than each individual student, is assigned to a school. As a further simplification, we solve three separate assignment problems, one for each grade level (elementary, middle, and high). However, we are exploring ways to handle so-called “feeder patterns” that attempt to provide more continuity from one grade level to the next. A further complication is that elementary and middle schools in WCPSS have one of two calendar options: traditional (August–June) or year-round. WCPSS actually provides two assignments for each node: a “base” assignment with one calendar option, and an “alternative” assignment to a school with the other calendar option. Under certain circumstances, a student can elect to attend the alternative school instead of the base school, in which case the county no longer provides transportation for that student. The problem described here is only the base assignment, and the calendar option of a particular school affects only the school’s capacity. Although the tool provides results for all three grade levels, for the sake of brevity we show results only for high schools.

WCPSS has several metrics to gauge the quality of a particular assignment in addition to several rules that every assignment must satisfy. Until now, WCPSS has relied on intuition and time-consuming ad hoc methods to determine assignments. They wanted a more automated process to recommend good assignments quickly, while still maintaining control over the final decisions. Due to the size of the problem and the numerous interrelated constraints, even determining whether a feasible assignment exists is nontrivial, and finding one of good quality is an additional challenge. Fortunately, this constrained decision problem can be formulated as a mixed integer linear programming (MILP) problem and solved using the MILP solver available in SAS/OR software.

METHODOLOGY

We used the OPTMODEL procedure to read the data from SAS data sets, formulate the problem, call the MILP solver, implement customized heuristics to reduce computation time, and output the solution to SAS data sets. In this section, we describe the decision variables, objectives, and constraints that define the MILP problem. We also provide source code examples for the corresponding statements in PROC OPTMODEL.

DECISION VARIABLES

Although WCPSS has traditionally made assignments one year at a time, they want to devise a multiyear plan so that parents have a better sense of what to expect more than one year in advance. The main decision variables are binary variables that determine the assignment of nodes to schools for each year in the planning horizon. We use the term *period* to refer to the school year indexed from the current year, as in [Table 1](#).

Table 1 Periods

School Year	Period ID
2007–08	0
2008–09	1
2009–10	2
2010–11	3

For the current school year, say 2007–08, the period is $p = 0$ and assignments are already fixed. Our optimization model determines assignments for periods $p > 0$. Explicitly, $\text{Assign}[n, s, p] = 1$ if node n is assigned to school s in period p , 0 otherwise. The values of these variables completely determine the solution, but defining several classes of auxiliary variables turns out to be useful.

The algebraic modeling language implemented in PROC OPTMODEL makes the conversion from mathematical expressions to the corresponding SAS code transparent. For example, the declaration of $\text{Assign}[n, s, p]$ is as follows:

```
var Assign {<n,s,p> in NODE_SCHOOL_PERIOD} binary;
```

Here, NODE_SCHOOL_PERIOD is an index set of triples (n, s, p) , declared and populated using the PROC OPTMODEL SET and READ DATA statements:

```
set <str,str,num> NODE_SCHOOL_PERIOD;
read data Node_school_period into NODE_SCHOOL_PERIOD=[Node_ID School_ID Period_ID];
```

Because of upper bounds on travel time, not every node is eligible to be assigned to every school. Note that the preceding statements take advantage of this structure by using the PROC OPTMODEL sparse representation to create an efficient optimization model. An alternative approach would be to use the dense representation

```
var Assign {n in NODES, s in SCHOOLS, p in PERIODS} binary;
```

and then use the PROC OPTMODEL FOR and FIX statements to eliminate ineligible assignments:

```
for {n in NODES, s in SCHOOLS, p in PERIODS: distance[n,s,p] > max_distance}
  fix Assign[n,s,p] = 0;
```

But the sparse approach is more efficient because we never create the unnecessary variables in the first place.

Because rapid population growth has outpaced new school construction in certain areas of the county, some schools will become overcrowded. $\text{CapacityFactor}[s, p]$ is the amount by which school s is underpopulated (<1) or overpopulated (>1) in period p . $\text{MinCapacityFactor}[p]$ and $\text{MaxCapacityFactor}[p]$ are the minimum and maximum values, respectively, of $\text{CapacityFactor}[s, p]$ across all schools s .

One of the goals of WCPSS is to balance regional school populations with respect to socioeconomic attributes. Here we consider two such attributes: school population proportion of free and reduced lunch (F&R) and low English proficiency (LEP) students. The methodology can be generalized to any number of arbitrarily defined attributes. $\text{Proportion}[a, s, p]$ is the proportion of students with attribute a assigned to school s in period p . $\text{MinProportion}[a, r, p]$ and $\text{MaxProportion}[a, r, p]$ are the minimum and maximum proportions, respectively, of $\text{Proportion}[a, s, p]$ across all schools s in region r . [Table 2](#), [Table 3](#), and [Table 4](#) summarize the attributes, schools, and regions considered in the Wake County problem. The data in these tables can be stored in SAS data sets and read by PROC OPTMODEL using one READ DATA statement per data set.

Table 2 Attributes

Attribute	Attribute ID
Free and reduced lunch	F&R
Low English proficiency	LEP

Table 3 WCPSS High Schools

School Name	School ID
APEX HIGH	316
ATHENS DRIVE HIGH	318
BROUGHTON HIGH	348
CARY HIGH	368
EAST WAKE HIGH	411
ENLOE HIGH	412
FUQUAY-VARINA HIGH	428
GARNER HIGH	436
GREEN HOPE HIGH	441
HOLLY SPRINGS HIGH	455
KNIGHTDALE HIGH	466
LEESVILLE ROAD HIGH	473
MIDDLE CREEK HIGH	495
MILLBROOK HIGH	500
PANTHER CREEK HIGH	526
SANDERSON HIGH	552
SE RALEIGH HIGH	562
WAKE FOREST HIGH	588
WAKEFIELD HIGH	595
HERITAGE HIGH	H2
TBD	H6

Table 4 WCPSS Regions

Region	High Schools in Region
North	588, 595, H2, H6
Far East	411
Near East	466
South-East	436, 562
South-West	428, 455, 495
West-South	316, 368
West-North	441, 526
North-West	473
North Raleigh	500, 552
Central	318, 348, 412

Finally, reassigning a node from one school to another school is generally undesirable (for example, because it can separate students from their friends). To track such a change in assignment, define binary variable $\text{Change}[n, p] = 1$ if node n is assigned to different schools in periods $p - 1$ and p , 0 otherwise.

OBJECTIVES

The four main goals are to reduce travel time, overcrowding, socioeconomic imbalance among neighborhood schools, and the number of students reassigned to different schools from year to year. These goals are expressed mathematically by the following objective functions, where population $[n, p]$ is the number of students in node n in period p and

distance $[n, s, p]$ is the distance (or travel time) from node n to school s in period p . (Note the convention that data parameter names appear in lower case and decision variable names are capitalized.)

$$\text{AverageDistance} = \frac{\sum_{n,s,p} \text{population}[n, p] \text{distance}[n, s, p] \text{Assign}[n, s, p]}{\sum_{n,p} \text{population}[n, p]} \quad (1)$$

$$\text{CapacityDeviation} = \sum_p (\text{MaxCapacityFactor}[p] - \text{MinCapacityFactor}[p]) \quad (2)$$

$$\text{AttributeImbalance} = \sum_{a,r,p} (\text{MaxProportion}[a, r, p] - \text{MinProportion}[a, r, p]) \quad (3)$$

$$\text{Displacement} = \sum_{n,p} \text{population}[n, p] \text{Change}[n, p] \quad (4)$$

The AverageDistance objective function is straightforward. CapacityDeviation measures the difference between the largest and smallest CapacityFactor; so minimizing this objective function has the effect of spreading the overcrowding more equally across schools. AttributeImbalance plays a similar role in balancing attributes across schools within a region. Finally, Displacement counts the number of students who are reassigned from one school to another.

In PROC OPTMODEL, objectives are declared using the MIN or MAX keyword. For example, the declaration of AverageDistance is as follows:

```
min AverageDistance
  = sum{<n, s, p> in NODE_SCHOOL_PERIOD} population[n, p]*distance[n, s, p]*Assign[n, s, p]
    / (sum{<n, p> in NODE_PERIOD} population[n, p]);
```

To solve this optimization problem with multiple competing objectives, we take the classical approach of minimizing a weighted sum of all four objectives (each scaled to be between 0 and 1), with weights chosen by the user to reflect the relative importance of each objective with respect to the others. By adjusting the weights, the user can steer the solver in a different direction if a specified set of weights yields a solution that favors one objective too strongly.

CONSTRAINTS

The constraints express both the rules that must be satisfied and also the relationships among variables in mathematical terms.

$$\sum_s \text{Assign}[n, s, p] = 1 \quad \text{for } n, p \quad (5)$$

$$\text{Change}[n, p] \geq \text{Assign}[n, s, p] - \text{Assign}[n, s, p - 1] \quad \text{for } n, s, p \quad (6)$$

$$\sum_n \text{population}[n, p] \text{Assign}[n, s, p] = \text{capacity}[s, p] \text{CapacityFactor}[s, p] \quad \text{for } s, p \quad (7)$$

$$\text{MinCapacityFactor}[p] \leq \text{CapacityFactor}[s, p] \leq \text{MaxCapacityFactor}[p] \quad \text{for } s, p \quad (8)$$

$$\sum_{n,p} \text{Change}[n, p] \leq 1 \quad \text{for } n \quad (9)$$

$$\text{Proportion}[a, s, p] = \frac{\sum_{n,s,p} \text{attributePopulation}[n, a, p] \text{Assign}[n, s, p]}{\sum_{n,s,p} \text{population}[n, p] \text{Assign}[n, s, p]} \quad \text{for } a, s, p \quad (10)$$

$$\text{MinProportion}[a, r, p] \leq \text{Proportion}[a, s, p] \leq \text{MaxProportion}[a, r, p] \quad \text{for } r, a, s, p \quad (11)$$

Constraint (5) forces each node to be assigned to exactly one school in each period. Constraint (6) forces $\text{Change}[n, p] = 1$ if node n is assigned to different schools in periods $p - 1$ and p . Constraint (7) defines the $\text{CapacityFactor}[s, p]$ variable as a ratio of the school's population to its capacity. Constraint (8) defines $\text{MinCapacityFactor}[p]$ and $\text{MaxCapacityFactor}[p]$. Constraint (9) prohibits a node from being reassigned more than once within the total time horizon. Constraint (10) defines $\text{Proportion}[a, s, p]$ as the ratio of students having attribute a at school s in period p . We modeled this nonlinear constraint with a sequence of linear approximations that yield a solution that satisfies (10). Constraint (11) defines $\text{MinProportion}[a, r, p]$ and $\text{MaxProportion}[a, r, p]$.

In PROC OPTMODEL, constraints are declared using the CON keyword. For example, the declaration of constraint (5) is as follows:

```
con SumAssign_n_p {<n,p> in NODE_PERIOD: p > 0}:
  sum{s in SCHOOLS_n_p[n,p]} Assign[n,s,p] = 1;
```

Note that the index set for the constraint includes the condition $p > 0$ so that the optimization model does not include trivially redundant constraints when $p = 0$.

In addition to these constraints, the optimization model contains lower and upper bounds on decision variables. For example, no school can be more than 30% percent over or under capacity. So $0.7 \leq \text{CapacityFactor}[s, p] \leq 1.3$, although the model allows these bounds to take different values for each school-period combination. Similarly, each region's attribute proportions can be subject to hard bounds, such as $0.05 \leq \text{Proportion}[a, s, p] \leq 0.2$. That is, the proportion of students having attribute a and assigned to school s in period p must be between 5% and 20%.

SOLUTION ANALYSIS

OBJECTIVE WEIGHTS

Given that four distinct objectives must be weighted against one another, a question that arises is how to choose the corresponding coefficient values. Determining appropriate weights for the objective function is difficult because the numerical meaning of the coefficients is only vaguely defined and we start with no guidance about how to choose the correct values. In fact, we are interested only in the relative proportions of these coefficients to each other (for example, multiplying all coefficients by a scalar value has no impact on the final solution).

The answer to this question is that one should not focus on the objective value itself, but should rather explore the space of combinations of (appropriately bounded and scaled) four-coefficient vectors. In the case of two-objective optimization, efficiency frontiers are often used to visualize and explore the data. With four objectives, however, visualization and exploration require an expanded tool set. We describe how we use JMP[®] software to help us understand the space of objective weights.

Every "run" of the optimization with specified objective weights can be seen as an experiment, and our task is to design a set of such experiments so that we cover a good portion of the space in which they are defined with reasonable computational requirements. We can accomplish this by making use of the Design of Experiments (DOE) capability of JMP software. To capture only the relative proportions of the weights, we force their values to add up to 100%, which means that we are interested in a "mixture"-type DOE. For the greatest flexibility in designing our experiments, we choose a Custom Design with mixture factors for all four weights: school capacity deviation (W_{cap}), student displacement (W_{disp}), average distance (W_{dist}), and regional attribute imbalance (W_{imb}). We restrict the values to be at least 5% in order to never completely neglect any objective. Since we expect at least a three-way interaction, we choose to model our outputs with all four main effects and a Scheffé cubic model, which yields 20 distinct experimental runs. In addition to this set of experiments, we expand our exploration of the parameter space by appending 130 additional runs with weight values that are randomly generated (and scaled to sum to 100%). For the randomly generated set, we relax our minimum weight requirement on each objective to be possibly as low as 1%. This provides us with a total of 150 combinations of the four objective weights.

Having obtained our collection of weight combinations, we are free to proceed and run all 150 experiments (optimizations) sequentially. However, each of these runs is completely independent. For situations such as this, SAS/GRID provides a powerful, easy-to-use capability that enables us to run all 150 optimizations simultaneously (in parallel). Since each run is very computationally intensive (each optimization can take more than an hour to complete), the gains are significant.

The runs are performed on a grid of 150 identical 64-bit, 2.33GHz Intel Xeon processors with 16GB RAM, using the SAS Grid Manager %Distribute macros (GridDistribute.sas).

SAMPLE SOLUTIONS

The optimization tool presented here generates a three-year (period) assignment plan for the school years 2008–09, 2009–10, and 2010–11. For each of the three periods, each of the county's 1,302 nodes is assigned to exactly one of the 21 high schools. As described earlier, the aim is to minimize a weighted combination of school capacity deviation, student displacement, average travel distance, and attribute imbalances. [Table 5](#) shows the first 15 (of 150) runs, with respective weights, run times, optimality bound gap, total displacement of students over the three-year proposal horizon, and average distance of travel in the solution.

Table 5 Results: High Schools

Run #	W_cap	W_disp	W_dist	W_imb	Run Time (s)	B Gap	Tot Disp	Avg Dist
1	0.05	0.63	0.27	0.05	2121	2%	2,426	3.2
2	0.08	0.49	0.38	0.05	1713	1%	2,516	3.2
3	0.05	0.71	0.21	0.03	1788	2%	2,560	3.2
4	0.09	0.8	0.06	0.05	1912	3%	2,633	3.3
5	0.07	0.57	0.32	0.04	1875	3%	2,642	3.2
6	0.16	0.69	0.14	0.01	2515	3%	2,686	3.2
7	0.06	0.68	0.1	0.16	2290	4%	2,744	3.2
8	0.05	0.85	0.05	0.05	1949	4%	2,778	3.2
9	0.02	0.7	0.03	0.25	2842	3%	2,833	3.2
10	0.05	0.63	0.05	0.27	1553	3%	2,887	3.2
11	0.14	0.36	0.43	0.07	1799	3%	2,911	3.2
12	0.27	0.43	0.26	0.04	2149	3%	2,917	3.2
13	0.05	0.27	0.63	0.05	1568	3%	2,925	3.2
14	0.27	0.63	0.05	0.05	2338	4%	2,947	3.3
15	0.25	0.45	0.23	0.07	2217	3%	2,956	3.2

Total computational time (the time required if running sequentially) is approximately six days, whereas clock time (the actual time needed for the full computation on the grid) is approximately three hours (approximately the time required for the longest-running job). This illustrates the type of performance gains that are possible with this type of setup and SAS Grid Manager. The median run time of the optimization model is 40 minutes.

The column labeled “B Gap” in [Table 5](#) represents the optimality bound gap, defined as the percent difference between the solution objective when the model stopped its run and its best lower bound known at termination. It is best thought of as a certificate of optimality. In the full set of solutions, the model’s outputs show a median optimality gap of 5% with a standard deviation of 3%. We assume this to be acceptable for our purposes, but higher accuracy can be attained at the expense of longer run times.

Suppose that student displacement is one of the most important factors that constrain node-school reassignments. [Table 5](#) displays results in ascending order of total student displacement over the three-year planning period. Interestingly, the table shows that the run with the highest W_disp ranks eighth on the list (rather than first), since we have ordered by increasing displacement. This is attributed to the fact that we are solving mixed integer nonlinear programming problems, and local minima might trap our solution from a global minimum.

However, [Figure 1](#) shows that this effect is not severe in the sense that we successfully influence the model to act, in general, according to the weight selection inputs. Again, this is another situation where accuracy can be improved at the expense of computational time. In fact, weights greater than 0.4 seem to cause the displacement objective to dominate the solution, overriding the effect of other weights.

[Figure 2](#) is a more comprehensive illustration of the effect of weights on targeted objective terms. The diagonal plots show that the overall impact of weight variation is achieved. Off-diagonal plots show that secondary effects lead to interesting behavior in some instances. For example, as the W_disp is reduced, leading to greater displacement of students, drastic improvements in school capacity deviation and average travel distance are possible.

The question might arise of whether solutions exist that improve distance traveled and capacity deviations while remaining in the insensitive area of student displacement of weights greater than 0.4. Unfortunately, [Figure 2](#) indicates otherwise. We have used the JMP highlight feature to show that even allowing for W_disp values as low as 0.3 (so that Max_p_displacement stays within approximately 2,000 students), the highlighted points have relatively high capacity deviations. The highlighted points also have relatively low average distance, but we are unable to explore the significant gains that would enable us to reduce average distance values below 3.2 miles. It seems appropriate to conclude that following a status-quo policy (of minimal school reassignment) favors the distance objective to the detriment of school capacity deviation.

Figure 1 Total Displacement versus W_disp

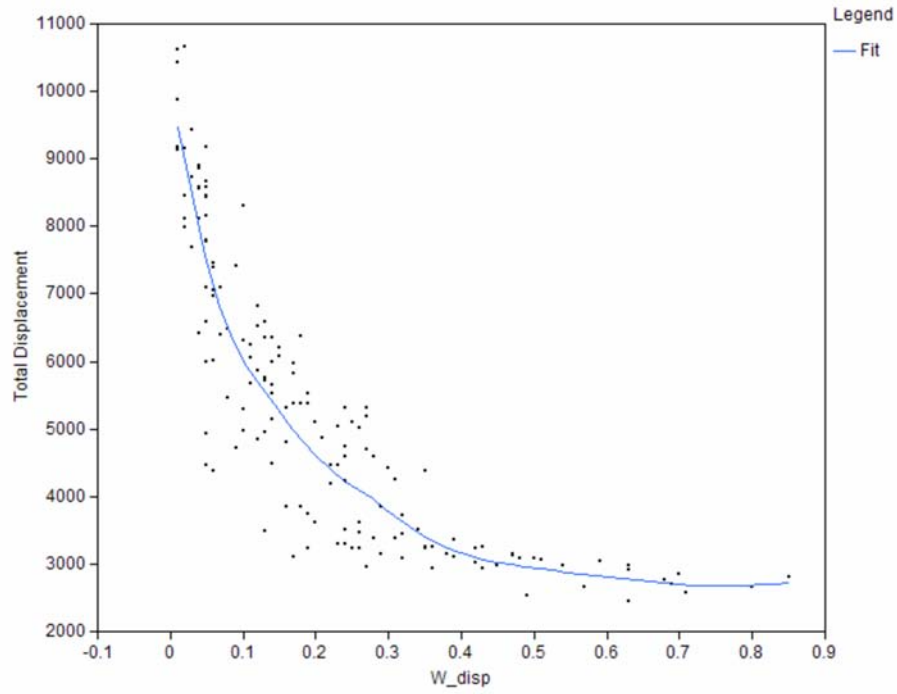
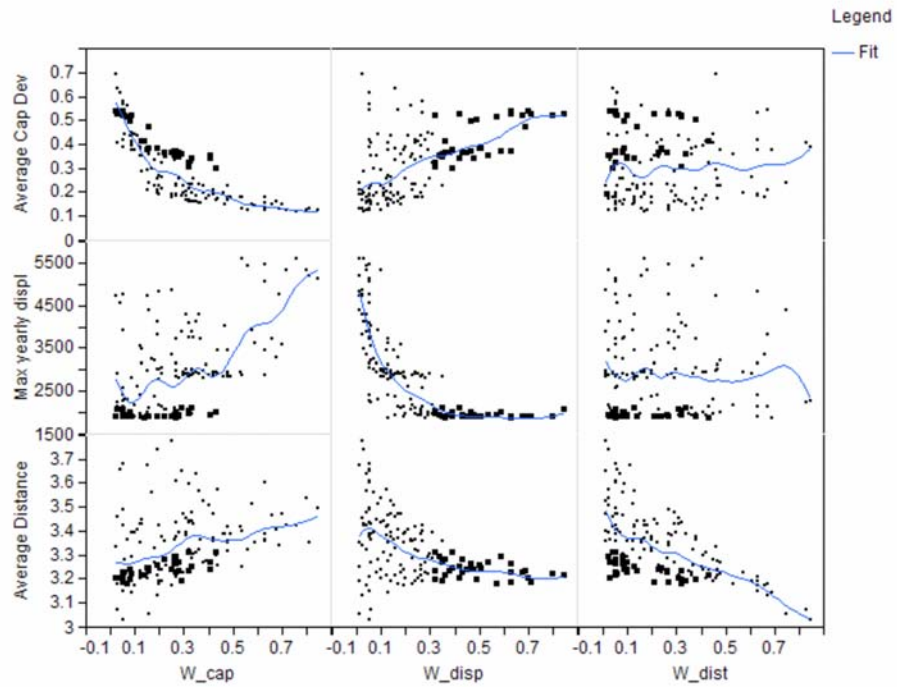


Figure 2 Objective Values versus Weights



FURTHER EXPLORATION OF PROPOSALS

For the purpose of this discussion, we assume that maintaining status-quo assignments benefits the solution. Exploration of the results enables us to select a small subset of runs for further investigation. Figure 3 shows that we can focus on the insensitive portion of the displacement objective weight W_disp (greater than 0.3) and still find points of relatively low average distance and capacity deviation that were originally highlighted in Figure 2. Four such points were found using JMP software, and they are shown in Figure 3.

Figure 3 Objective Values versus Weights, with Selected Points for Table 6

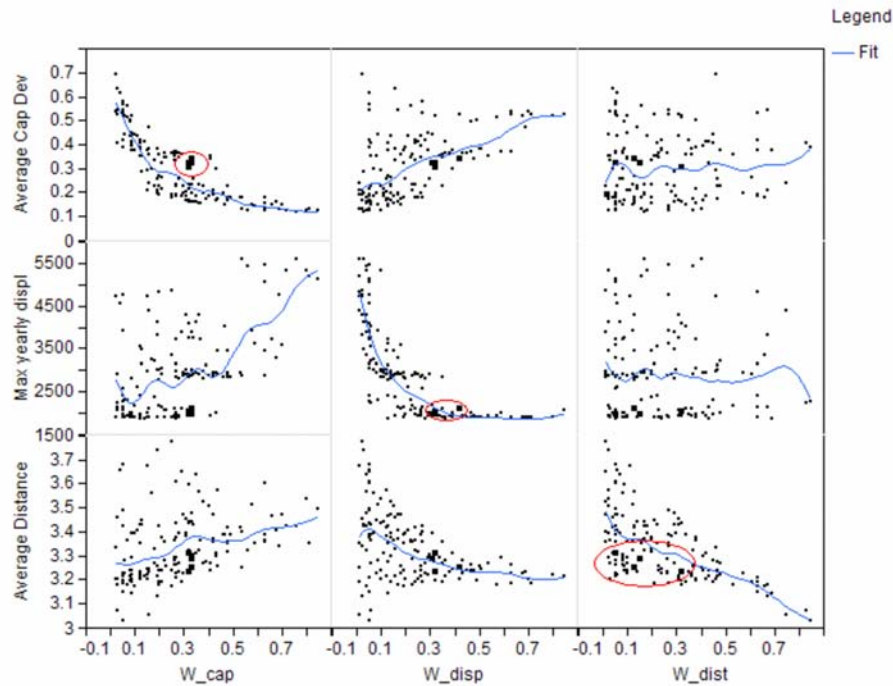


Table 6 presents additional information for the selected optimization runs. The table shows the weights used as inputs to the model and the following outputs: average distance traveled, maximum difference between capacity factors among all schools (ΔC_p , for periods $p = 0, 1, 2, 3$), except schools that are introduced in $p > 0$ (H2 and H6), and the displacement of students (D_p , for $p = 1, 2, 3$). We note that because period $p = 0$ is data read into the problem, all ΔC_0 values are equal (these values correspond to the current assignments, which are not subject to change).

Table 6 Results: High Schools, Selected Points from Figure 3

W_cap	W_disp	W_dist	W_imb	Avg Dist	ΔC_0	ΔC_1	ΔC_2	ΔC_3	D_1	D_2	D_3
0.33	0.31	0.15	0.21	3.3	39%	16%	17%	35%	1,144	289	1,930
0.32	0.32	0.05	0.32	3.3	39%	17%	16%	36%	1,487	270	1,940
0.32	0.32	0.32	0.05	3.2	39%	17%	14%	41%	786	635	2,002
0.33	0.42	0.13	0.12	3.2	39%	18%	17%	48%	933	210	2,062

Not surprisingly, Table 6 shows that solutions that aim to minimize school capacity deviation and retain the status-quo are heavily weighted in these objectives (over 64% of the weight is placed on these). However, what is perhaps surprising is the relative insensitivity to weights placed on distance, especially since all the results in Table 6 show good values of around 3.2 to 3.3 miles. This relative insensitivity to weights placed on distance gives us a certain amount of flexibility in choosing our solution. If we are interested in exploring a solution that achieves less school capacity deviation, the first row ($W_cap=33, W_disp=31, W_dist=15, W_imb=21$) seems promising. On the other hand, if we

are interested in exploring a solution that achieves less disturbance over the years by restraining displacement, the last row ($W_cap=33$, $W_disp=42$, $W_dist=13$, $W_imb=12$) seems more appropriate.

We could explore each of the solutions in [Table 6](#) (among any others that similar analysis might identify as desirable) as potentially good candidates by further investigating outputs of the optimization. For illustration, we focus on the last solution in the table, which we refer to by its weights ($W_cap=33$, $W_disp=42$, $W_dist=13$, $W_imb=12$). Bear in mind, however, that all of the outputs presented in [Table 7](#) are immediately available upon completion of our parallel run for all 150 solutions of our DOE.

By assigning 33% of the weight to the school capacity deviation objective, we expect a moderately balanced outcome from the optimization. In any case, given that all 150 solutions are feasible, we expect school total population to stay within prescribed bounds of 50% and 130% of capacity for all periods. [Table 7](#) shows the total population as a percentage of capacity of each high school over the next three periods.

Table 7 School Capacity Factors: High Schools ($W_cap=33$, $W_disp=42$, $W_dist=13$, $W_imb=12$)

School Name	$p = 0$	$p = 1$	$p = 2$	$p = 3$
APEX HIGH	100.5%	104.9%	90.2%	98.5%
ATHENS DRIVE HIGH	108.8%	96.0%	91.2%	97.3%
BROUGHTON HIGH	106.5%	104.8%	107.4%	110.6%
CARY HIGH	78.3%	86.5%	90.1%	83.1%
EAST WAKE HIGH	91.8%	94.5%	103.9%	115.1%
ENLOE HIGH	97.1%	104.8%	106.2%	108.0%
FUQUAY-VARINA HIGH	102.9%	99.9%	95.7%	95.1%
GARNER HIGH	100.2%	87.7%	90.8%	102.7%
GREEN HOPE HIGH	102.9%	101.4%	98.6%	101.0%
HOLLY SPRINGS HIGH	95.8%	99.5%	97.4%	108.0%
KNIGHTDALE HIGH	92.1%	95.6%	107.6%	97.5%
LEESVILLE ROAD HIGH	114.5%	104.7%	105.9%	108.3%
MIDDLE CREEK HIGH	76.9%	86.4%	91.9%	96.7%
MILLBROOK HIGH	98.2%	98.2%	99.0%	85.5%
PANTHER CREEK HIGH	115.5%	102.5%	104.4%	114.8%
SANDERSON HIGH	98.3%	98.0%	100.7%	104.2%
SE RALEIGH HIGH	93.6%	103.5%	106.7%	109.1%
WAKE FOREST HIGH	78.0%	86.5%	100.3%	92.0%
WAKEFIELD HIGH	93.9%	94.3%	90.2%	67.4%
HERITAGE HIGH	-	-	-	50.3%
TBD	-	-	-	51.5%
Avg (excluding last two)	97.1%	97.4%	98.9%	99.7%
Std (excluding last two)	11.0%	6.6%	6.6%	11.9%

We see that feasibility is clearly achieved since all schools stay within the imposed bounds. Furthermore, by the 2009–10 school year ($p = 2$), we achieve a much tighter distribution of school capacities than the current assignments: a mean of 98.9% and standard deviation of 6.6% versus a mean of 97.1% and standard deviation of 11.0%. This result comes despite the changing, uneven dynamics of population growth throughout the county. Note that a significant disruption in the 2010–11 school year ($p = 3$) causes significant changes in deviations; two new schools, Heritage High and one whose name is to be determined, come online. However, even then, all bounds are correctly observed, and a mean of 99.7% capacity utilization with standard deviation of 11.9% (excluding the new schools) suggests that, since it is an optimal solution, this is as good as we can expect for this selection of weights. We are free, however, to improve this objective—most likely at the detriments of others—by increasing W_cap . This type of tradeoff analysis is our recommendation of how to best make use of the proposed tool.

We recall that two socioeconomic attributes are considered here (F&R and LEP), and that one of our objectives is to minimize the imbalance of these attributes between schools that belong to the same region, as defined in [Table 4](#). Judging from our weights selection, the ($W_cap=33$, $W_disp=42$, $W_dist=13$, $W_imb=12$) solution seems to put relatively low emphasis on region imbalance (12% of the weight). It is important then to investigate this aspect of the solution, which is an output of the optimization run (and again, available for each of our 150 runs at completion).

[Table 8](#) shows the ($W_cap=33$, $W_disp=42$, $W_dist=13$, $W_imb=12$) solution's extreme values for F&R student pop-

ulation proportions within each region of the county. The lower bound and upper bound columns show the desirable hard bounds for each region; these are inputs to the problem (bounds can be determined by School Board policy, for example). The remaining columns show, for each period, the F&R proportions: “min” for the school in the region with the lowest proportion, and “max” for the school in the region with the highest proportion. Stated succinctly, one of our objectives is to minimize the difference between these numbers for each region, but constrained by the lower and upper bounds in the table. As a later table shows, the bounds can take different values for different attributes.

Table 8 Region Attribute F&R Extremes: High Schools ($W_cap=33$, $W_disp=42$, $W_dist=13$, $W_imb=12$)

Region	lower bound	upper bound	$p = 0$ min	$p = 0$ max	$p = 1$ min	$p = 1$ max	$p = 2$ min	$p = 2$ max	$p = 3$ min	$p = 3$ max
North	15.0%	40.0%	13.5%	17.5%	15.1%	16.8%	15.1%	16.7%	16.1%	16.9%
Far East	15.0%	40.0%	39.7%	39.7%	39.2%	39.2%	38.8%	38.8%	37.8%	37.8%
Near East	15.0%	40.0%	33.3%	33.3%	29.0%	29.0%	27.8%	27.8%	29.7%	29.7%
South-East	15.0%	40.0%	25.9%	33.9%	29.3%	34.7%	30.5%	34.8%	31.0%	34.3%
South-West	15.0%	40.0%	16.5%	20.8%	15.5%	19.9%	15.7%	20.4%	15.5%	20.3%
West-South	10.0%	40.0%	8.0%	20.9%	9.8%	21.7%	11.2%	22.0%	11.9%	22.2%
West-North	10.0%	40.0%	5.6%	8.4%	8.9%	9.1%	9.1%	9.2%	9.1%	9.7%
North-West	15.0%	40.0%	13.4%	13.4%	14.7%	14.7%	16.0%	16.0%	16.8%	16.8%
North Raleigh	15.0%	40.0%	24.3%	27.3%	25.8%	28.5%	27.5%	28.7%	28.6%	29.6%
Central	15.0%	40.0%	19.1%	25.3%	20.7%	23.7%	21.1%	24.8%	21.2%	24.9%

It is evident from [Table 8](#) that schools in the North, West-South, West-North, and North-West regions currently ($p = 0$) violate their lower bound requirements in F&R student proportions. The table also shows that, with one exception, the ($W_cap=33$, $W_disp=42$, $W_dist=13$, $W_imb=12$) solution adequately brings regions to compliance by 2010–11 ($p = 3$). The exception mentioned is the West-North region, which is unable to do better than 9.1% F&R (the bound is 10%). Two important points should be made here. First, we note that despite violating the constraint, the final solution is vastly improved from the current 5.6% value. Second, the tool is capable of providing a solution that is infeasible but will do so only when necessary. The implication is that the tool has told us something important about our policy decisions: the bounds imposed on the problem are too stringent. Either we can choose to accept the proposed solution as “close enough” to the desirable bounds, or we can adjust our policy to adapt to this reality (by increasing the distance that schools in the West-North region can serve, changing the lower bound on the region, incorporating other schools in this region, and so on).

As described in [Table 4](#), some regions (Far East, Near East, and North-West) contain only one school. Therefore, zero imbalance exists in these regions, and the imposed bounds are the only outcome that our tool aims to control. More than one school is present in the remaining regions, allowing for the possibility of nonzero imbalances $\Delta_p = \max_p - \min_p$ for each period p , as shown in [Table 9](#).

Table 9 Region Attribute F&R Imbalances: High Schools ($W_cap=33$, $W_disp=42$, $W_dist=13$, $W_imb=12$)

Region	Δ_0	Δ_1	Δ_2	Δ_3
North	4.0%	1.7%	1.6%	0.8%
South-East	8.0%	5.4%	4.3%	3.3%
South-West	4.3%	4.4%	4.7%	4.8%
West-South	12.9%	11.9%	10.8%	10.3%
West-North	2.8%	0.2%	0.1%	0.6%
North Raleigh	3.0%	2.7%	1.2%	1.0%
Central	6.2%	3.0%	3.7%	3.7%

It is evident from [Table 9](#) that the tool succeeds in reducing the imbalances within regions to an improved solution over current assignments. Some insight is also available from the results. The schools in region West-South (Apex High and Cary High) present the most difficulty when attempting to resolve imbalances. Schools in region West-North (Green Hope High and Panther Creek High) present the least difficulty, reflecting the relative homogeneity of the region.

[Table 10](#) shows results for the LEP attribute extremes, and [Table 11](#) shows results for the LEP imbalances.

Table 10 Region Attribute LEP Extremes: High Schools ($W_{cap}=33, W_{disp}=42, W_{dist}=13, W_{imb}=12$)

Region	lower bound	upper bound	$p = 0$ min	$p = 0$ max	$p = 1$ min	$p = 1$ max	$p = 2$ min	$p = 2$ max	$p = 3$ min	$p = 3$ max
North	3.0%	10.0%	1.8%	6.4%	3.3%	5.3%	3.4%	5.3%	3.9%	5.1%
Far East	3.0%	10.0%	5.5%	5.5%	5.4%	5.4%	5.4%	5.4%	5.2%	5.2%
Near East	3.0%	10.0%	7.7%	7.7%	6.1%	6.1%	6.1%	6.1%	7.1%	7.1%
South-East	3.0%	10.0%	1.1%	5.9%	3.8%	6.0%	4.4%	6.1%	4.8%	6.1%
South-West	3.0%	10.0%	2.2%	5.5%	2.9%	5.3%	3.1%	5.1%	3.0%	4.6%
West-South	3.0%	10.0%	2.9%	9.3%	3.2%	9.9%	4.0%	10.0%	4.3%	10.3%
West-North	3.0%	10.0%	3.1%	4.3%	3.5%	4.1%	3.7%	3.8%	3.7%	3.8%
North-West	3.0%	10.0%	5.4%	5.4%	5.2%	5.2%	5.7%	5.7%	6.2%	6.2%
North Raleigh	3.0%	10.0%	5.0%	9.5%	5.7%	9.9%	6.3%	10.1%	6.7%	11.0%
Central	3.0%	10.0%	1.0%	9.3%	3.0%	9.4%	3.2%	9.7%	3.5%	9.6%

Table 11 Region Attribute LEP Imbalances: High Schools ($W_{cap}=33, W_{disp}=42, W_{dist}=13, W_{imb}=12$)

Region	Δ_0	Δ_1	Δ_2	Δ_3
North	4.6%	2.0%	1.9%	1.2%
South-East	4.8%	2.2%	1.7%	1.3%
South-West	3.3%	2.4%	2.0%	1.6%
West-South	6.4%	6.7%	6.0%	6.0%
West-North	1.2%	0.6%	0.1%	0.1%
North Raleigh	4.5%	4.2%	3.8%	4.3%
Central	8.3%	6.4%	6.5%	6.1%

Similar conclusions can be drawn here, suggesting correlation between the LEP and F&R attributes, although results indicate that the Central region shows greater disparity between LEP students than it does for F&R. Again, many actions could be taken to deal with the results of the model, and further exploration of results could lead to solutions that place different emphasis on the outcome. For example, if we explore the ($W_{cap}=32, W_{disp}=32, W_{dist}=05, W_{imb}=32$) solution, we find much tighter imbalance distributions, shown in [Table 12](#) and [Table 13](#). (Contrast these solutions with those from [Table 9](#) and [Table 11](#), respectively.) Finally, for a geographical illustration of the ($W_{cap}=33, W_{disp}=42, W_{dist}=13, W_{imb}=12$) solution, we show SAS/GIS maps of this proposal in [Figure 4](#) (2007-08), [Figure 5](#) (2008-09), [Figure 6](#) (2009-10), and [Figure 7](#) (2010-11).

Table 12 Region Attribute F&R Imbalances: High Schools ($W_{cap}=32, W_{disp}=32, W_{dist}=05, W_{imb}=32$)

Region	Δ_0	Δ_1	Δ_2	Δ_3
North	4.0%	1.7%	0.0%	2.1%
South-East	8.0%	3.8%	3.1%	1.4%
South-West	4.3%	2.3%	2.4%	2.5%
West-South	12.9%	8.0%	4.4%	3.5%
West-North	2.8%	0.5%	0.4%	0.3%
North Raleigh	3.0%	1.3%	0.2%	0.4%
Central	6.2%	1.2%	1.7%	1.7%

Table 13 Region Attribute LEP Imbalances: High Schools ($W_{cap}=32, W_{disp}=32, W_{dist}=05, W_{imb}=32$)

Region	Δ_0	Δ_1	Δ_2	Δ_3
North	4.6%	0.9%	0.8%	0.8%
South-East	4.8%	2.0%	1.5%	0.8%
South-West	3.3%	0.3%	0.2%	0.4%
West-South	6.4%	4.3%	3.7%	3.5%
West-North	1.2%	0.7%	0.2%	0.0%
North Raleigh	4.5%	3.9%	3.2%	3.3%
Central	8.3%	6.1%	6.0%	6.1%

Figure 4 Node-School Assignment Map: Period 0 (Current)

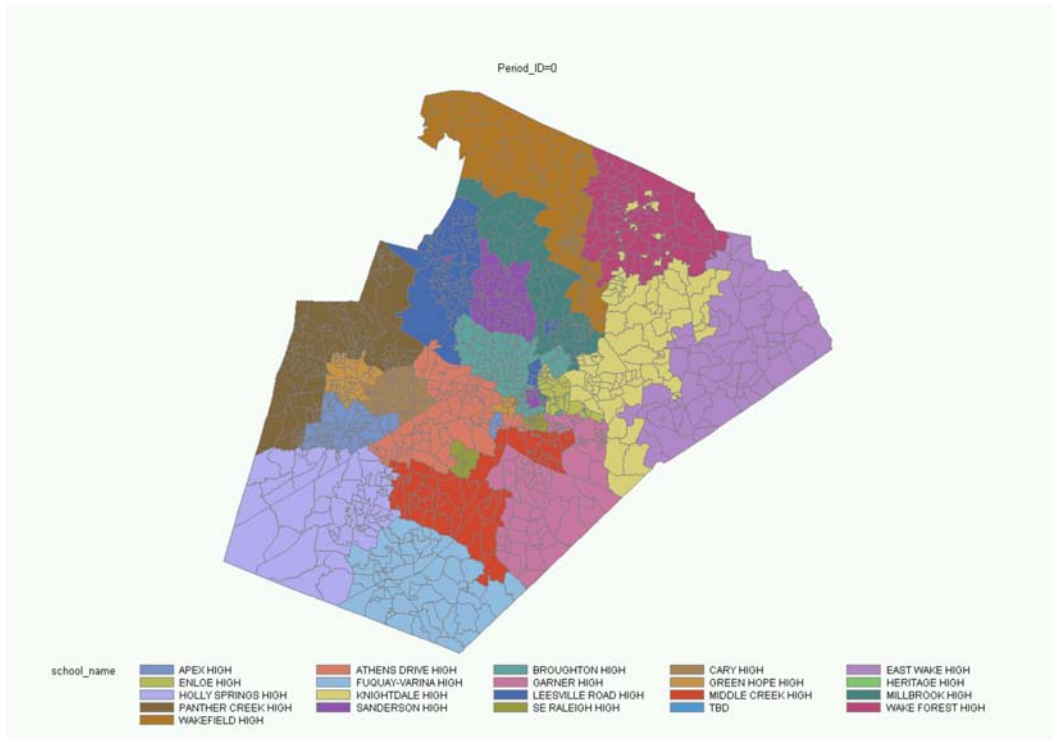


Figure 5 Node-School Assignment Map: Period 1 (W_cap=33, W_disp=42, W_dist=13, W_imb=12)

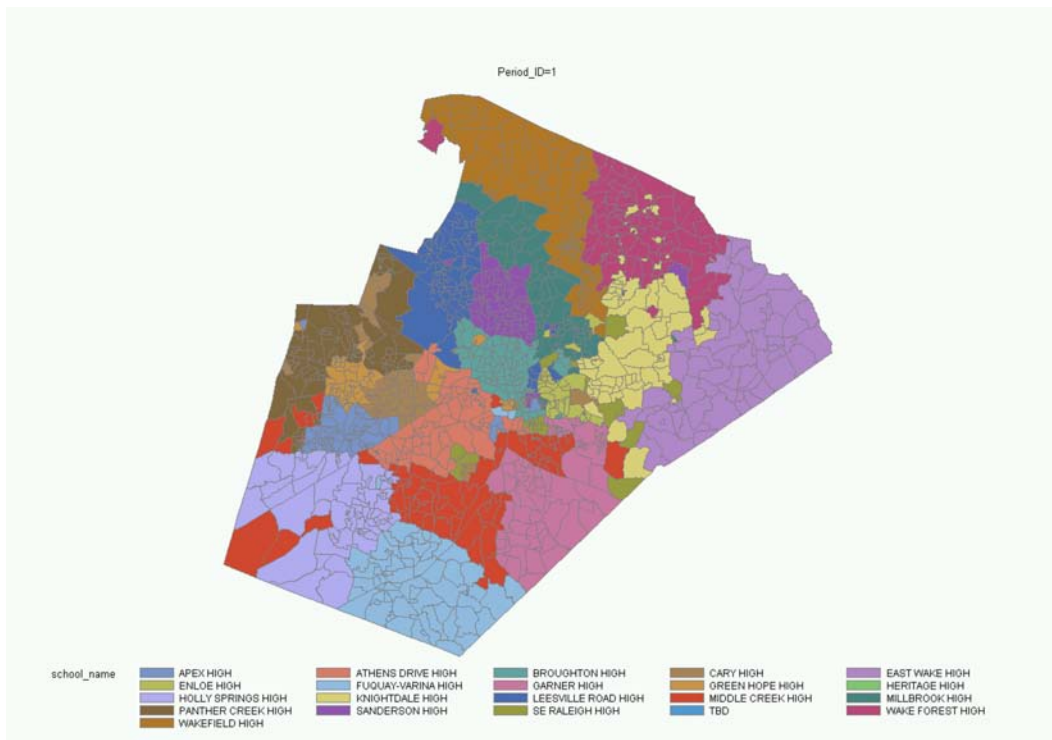


Figure 6 Node-School Assignment Map: Period 2 (W_cap=33, W_disp=42, W_dist=13, W_imb=12)

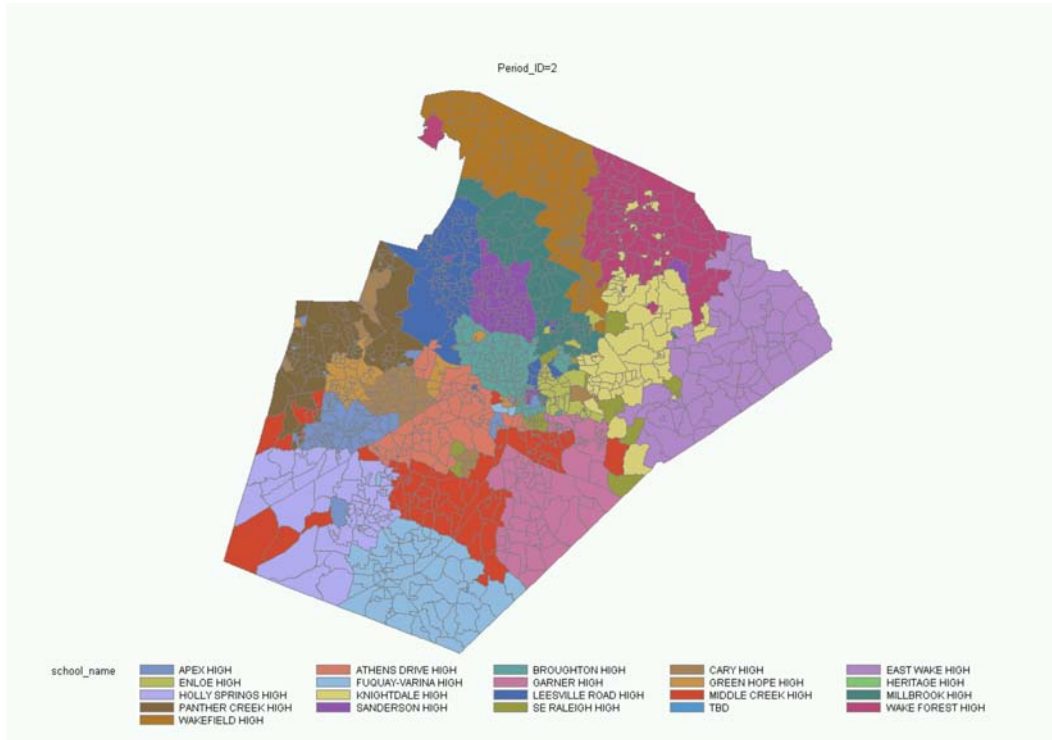
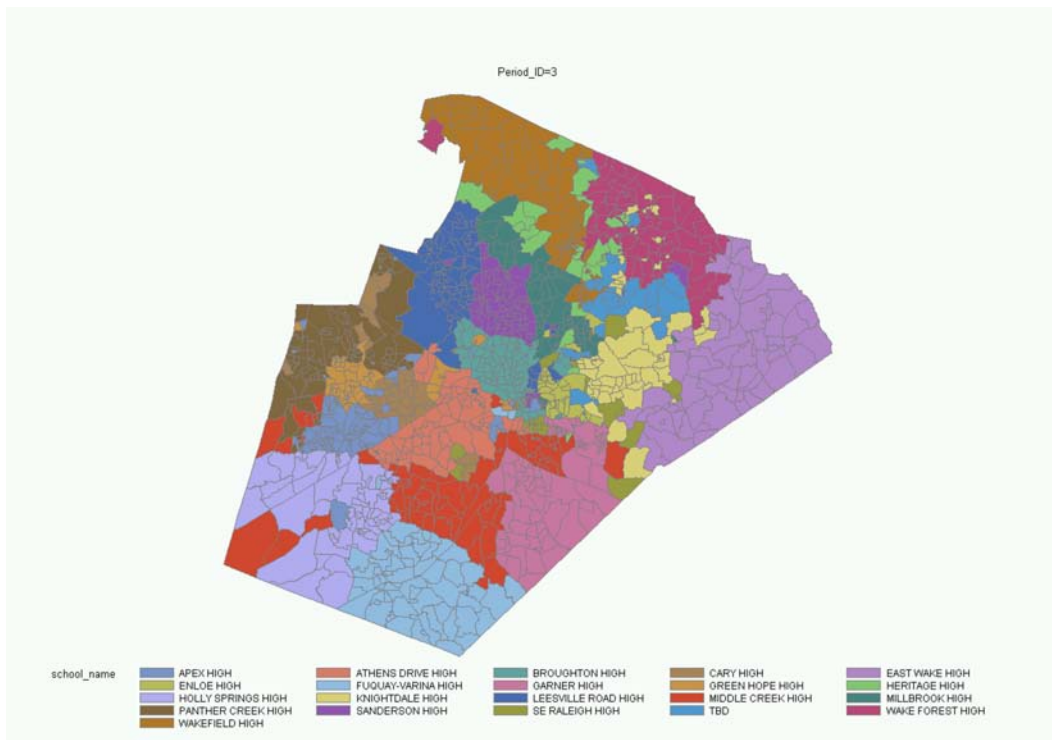


Figure 7 Node-School Assignment Map: Period 3 (W_cap=33, W_disp=42, W_dist=13, W_imb=12)



CONCLUSION

This example illustrates the power and flexibility available with PROC OPTMODEL, in addition to the ease of coding a mathematical programming model to solve a real-world problem. By leveraging SAS Grid Manager and JMP software, we also demonstrate the benefits of having such a tool embedded within the SAS system. The resulting automation and speed enable the decision makers at WCPSS to focus on higher-level concerns rather than the overwhelming low-level details of manually trying to construct an assignment that balances multiple competing objectives while satisfying numerous difficult constraints.

CONTACT INFORMATION

Ivan Oliveira, SAS/OR Research and Development
100 SAS Campus Drive
Room R5329
Cary, NC 27513
(919) 531-0097
Ivan.Oliveira@sas.com

SAS and all other SAS Institute Inc. products 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.