HAI MODEL MST TEST?

A. Yes, but it is important that one keep in mind what the MST test represents. The test is a test of a model's internal consistency, in other words, whether the respective model does what it purports to do, assuming that one accepts its particular modeling assumptions.

With respect to the HAI model, the test addresses whether the HAI model estimates the minimum amount of cable distance, via the rectangular ***main clusters***, to connect customers in the locations identified by the model, i.e., in the corresponding PNR ***main clusters***.

With respect to BCPM, the test addresses whether BCPM estimates the minimum amount of cable distance, via the road-reduced areas and connecting cable configuration, to connect customers in the locations identified by the model, i.e., in the microgrids that comprise an ultimate grid.

Hence, the conclusion one can make is that BCPM is more internally consistent than HAI 5.0a. That is, BCPM is much more likely to estimate the minimum amount of distribution distance needed to connect customers in ***its*** serving areas, i.e., ultimate grids, than is HAI 5.0a to connect customers in ***its*** serving areas i.e., main PNR polygon clusters.

Q. DO THE RELATIVE RESULTS OF THE TWO MODELS' MST TESTS CHANGE IF THE DEFINITION OF A "SERVING AREA" IN THE HAI MODEL IS EXPANDED TO INCLUDE THE ASSOCIATED OUTLIER

Q. PLEASE SUMMARIZE THE MAIN POINTS OF YOUR REBUTTAL TESTIMONY.

A. There are three points I wish to emphasize that pertain respectively to the Hatfield models' customer location, customer aggregation, and provision of distribution plant.

First, the rate of successful address-geocoding in the rural areas of Florida is very low. In fact, not a single location could be geocoded in 25 wire centers in Florida. HAI 5.0a relies on an estimation process for those locations that cannot be address-geocoded. Due to the limited ability to address-geocode customers in rural areas, HAI 5.0a's customer location methodology is reduced essentially to placing customers along the perimeter of Census Blocks.

The proponents of the HAI model have not provided any quantitative analysis of the predictive accuracy of the geocode-surrogate methodology relative to actual, real-world customer locations. In comparison, it has been demonstrated in this testimony that BCPM yields a reasonably accurate depiction of the distribution of customers across the randomly chosen Yankeetown wire center.

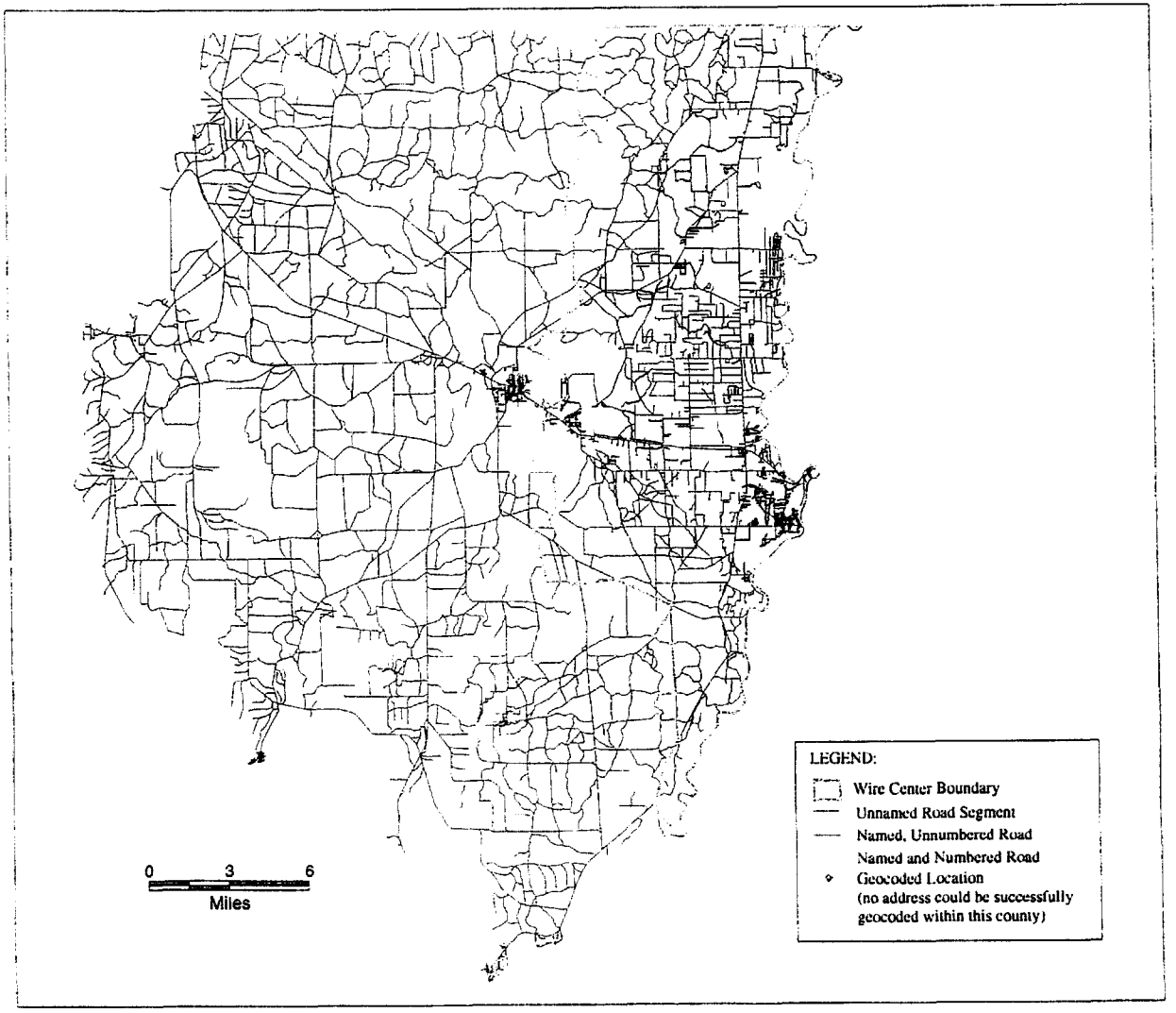
Second, the degree to which a model uses address-geocoding needs to be determined. For example the address-geocoded and surrogate locations are used only to define the perimeter of the PNR polygon clusters in the HAI preprocessing stage. Once these clusters are formed, the customer latitude and longitude information is discarded. This information never enters the Access database used by HAI 5.0a.

1,866 miles while that of BCPM totals at least 465 miles.

Q. DOES THIS CONCLUDE YOUR TESTIMONY?

A. Yes.

Dixie County, Florida
Road Network

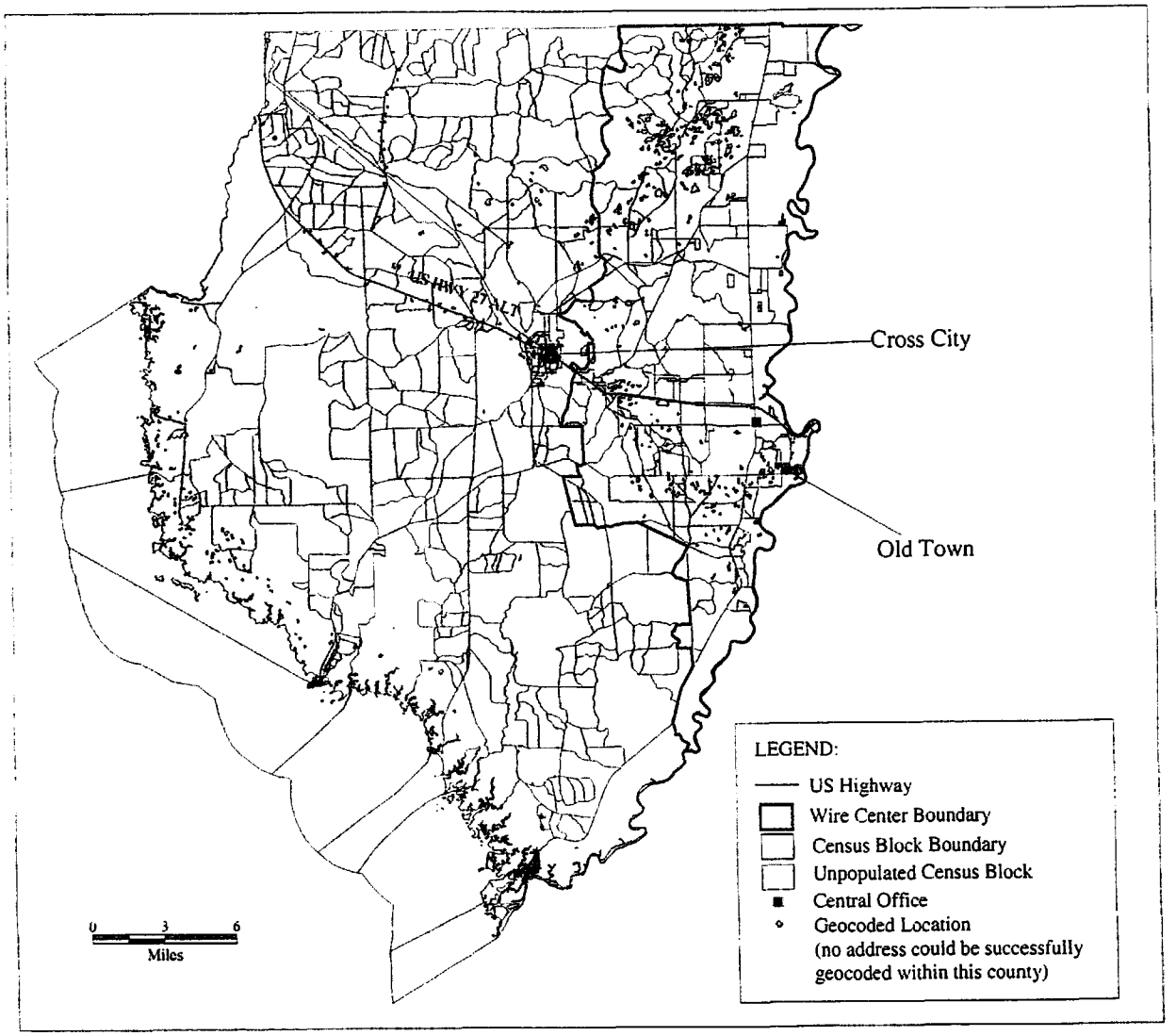


Less than 1% of the Dixie County roads (shown in red) are named and numbered and are therefore, geocodable.

Exhibit KDD-1

Dixie County, FL

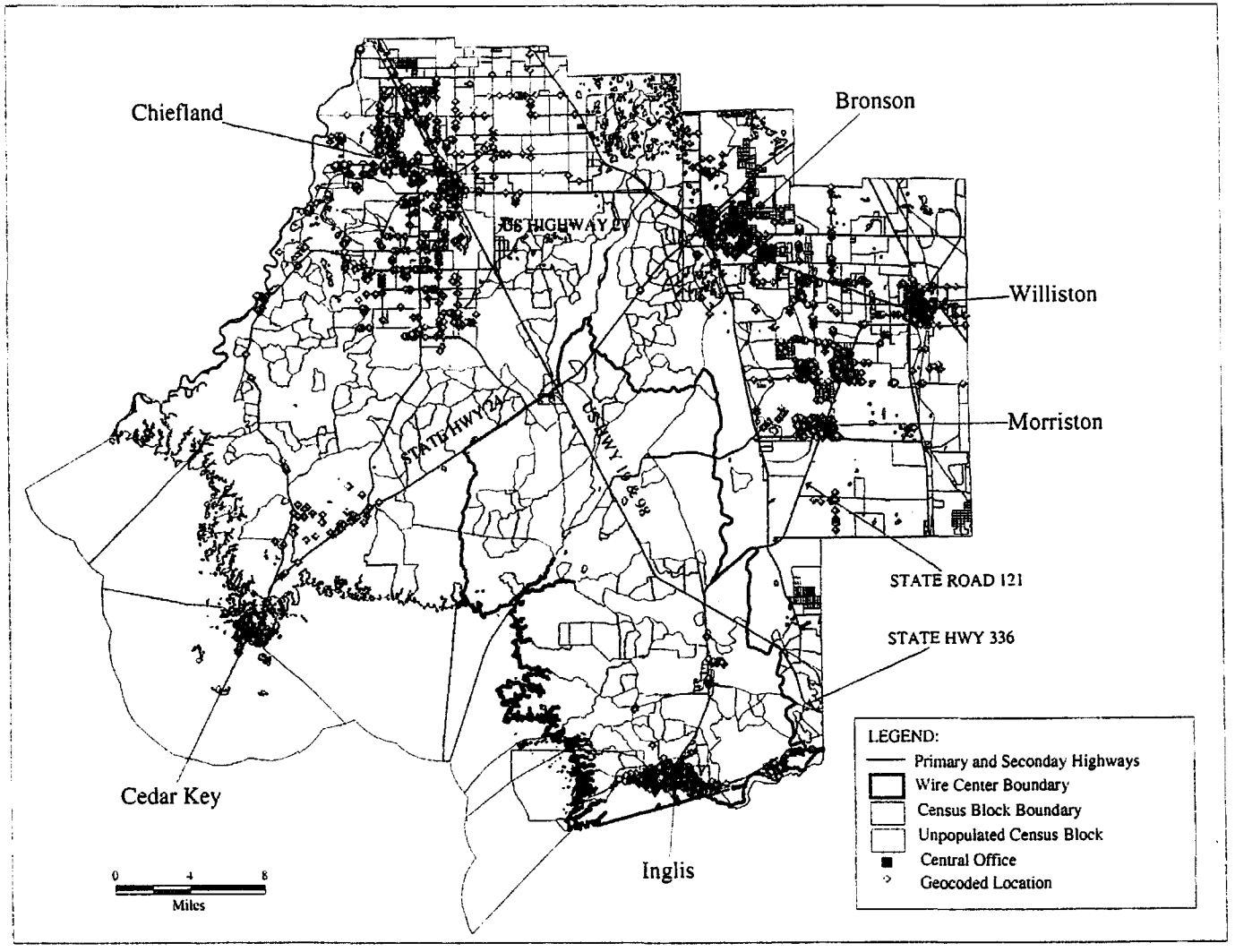
Geocoded Customer Locations



Wire Center FL 07914 01411
CLLI OLTWFLLN
Exhibit KDD-2

Levy County, FL

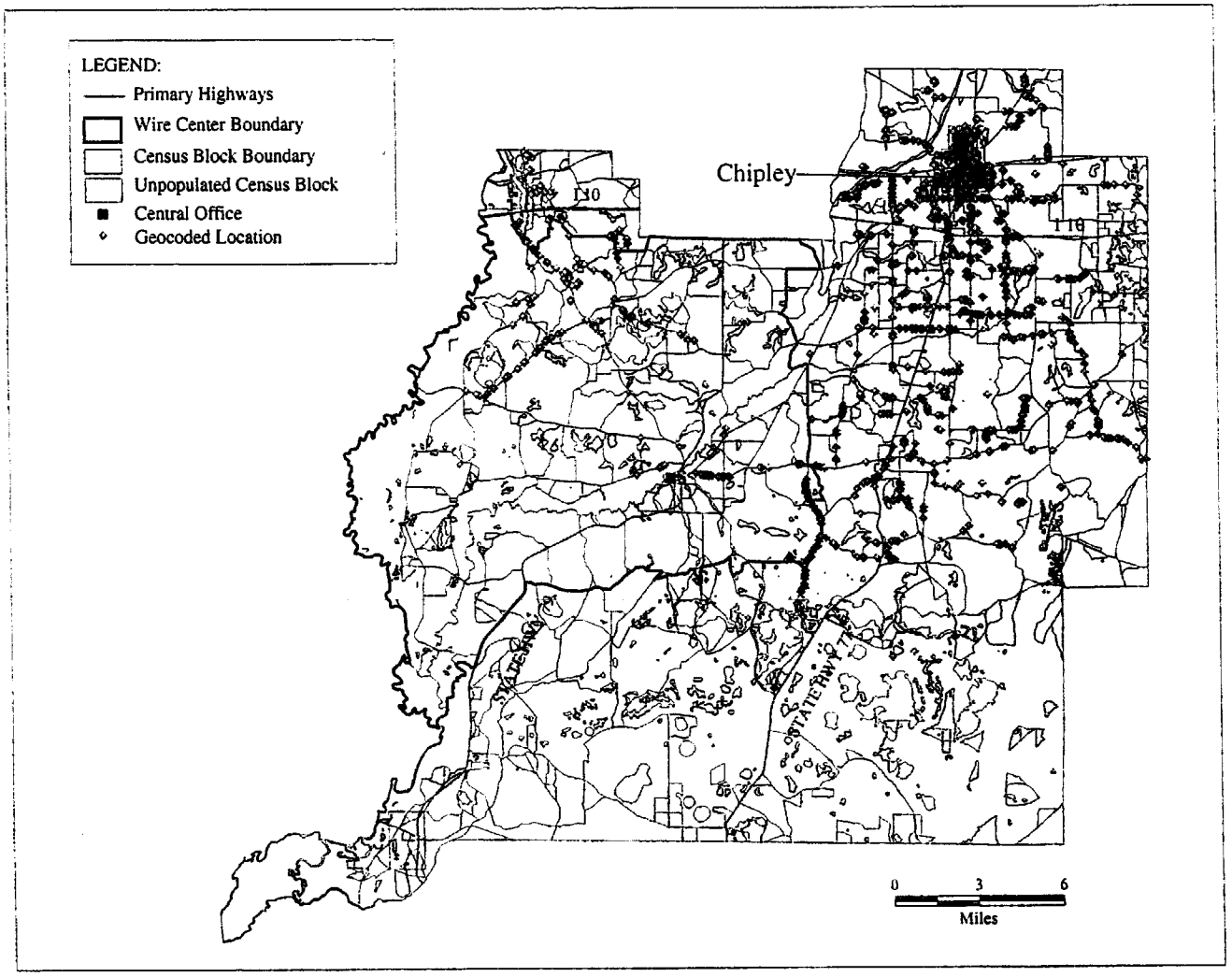
Geocoded Customer Locations



Wire Center FL 07991 01303
CLLI YNTWFLMA
Exhibit KDD-3

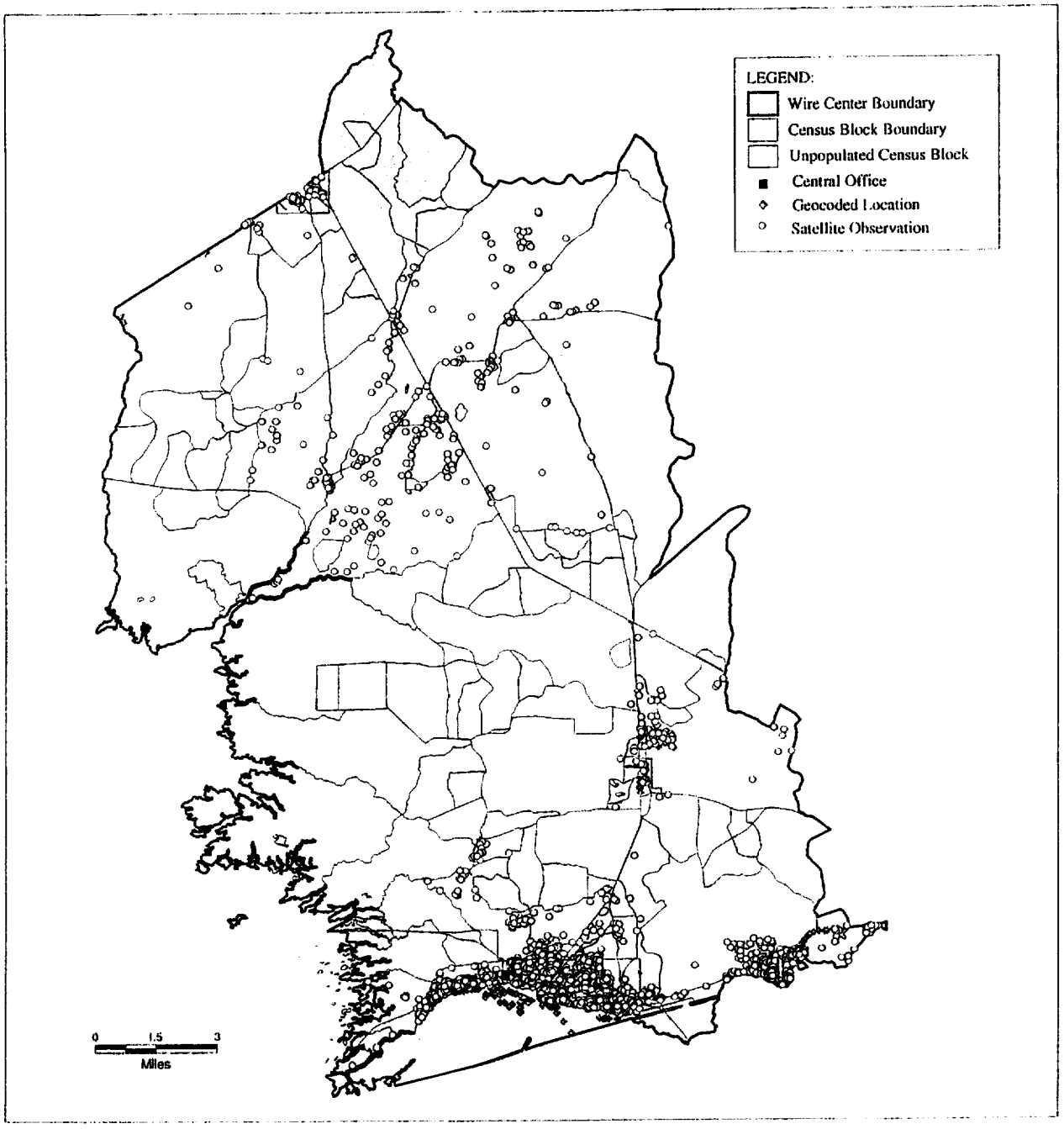
Washington County, FL

Geocoded Customer Locations



Wire Center FL 07974 01972
CLLI VERNFLMA
Exhibit KDD-4

Yankeetown Wire Center
Levy County, FL
Satellite Observations and Geocoded Customer Locations



Wire Center FL 07991 01303
CLLI YNTWFLMA
Exhibit KDD-5

Effect of Surrogate Point Placement On Minimum Spanning Tree Length

By Phil Bolian, Stopwatch Maps
For *INDETEC* International

The documentation of the HAI Model Version 5.0a claims that the placement of surrogate points uniformly about the periphery of a Census Block causes those points to be "maximally separated from one another" [Section 5.4.4, first paragraph]. The documentation claims that this placement is highly conservative ... that is, that it causes the *greatest dispersion* of points possible.

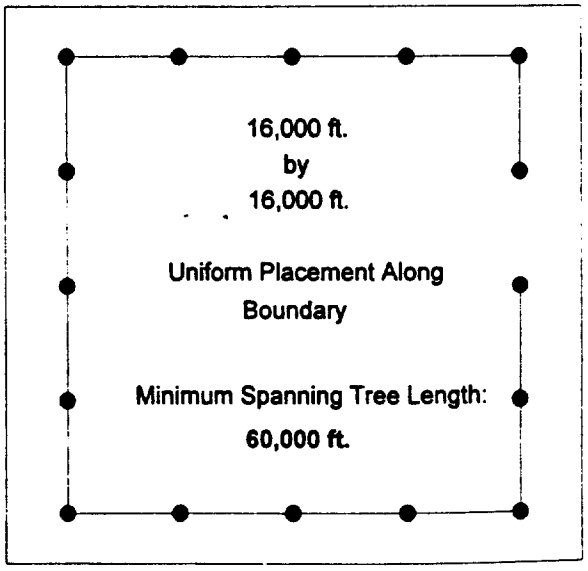
In fact, it does *not* cause the greatest dispersion of points. This paper will illustrate this by placing the same number of surrogate points in two *other* configurations:

- Uniformly *within* a Census Block
- Uniformly along *interior roads* as well as the periphery

We will then determine the dispersion (as measured by a Minimum Spanning Tree) of each of the newly placed sets of points, then compare each to the Minimum Spanning Tree for points placed about the periphery of the Census Block. We will find that the surrogate points in these new placements are either *just as dispersed as* or *more dispersed than* in the original placement about the periphery.

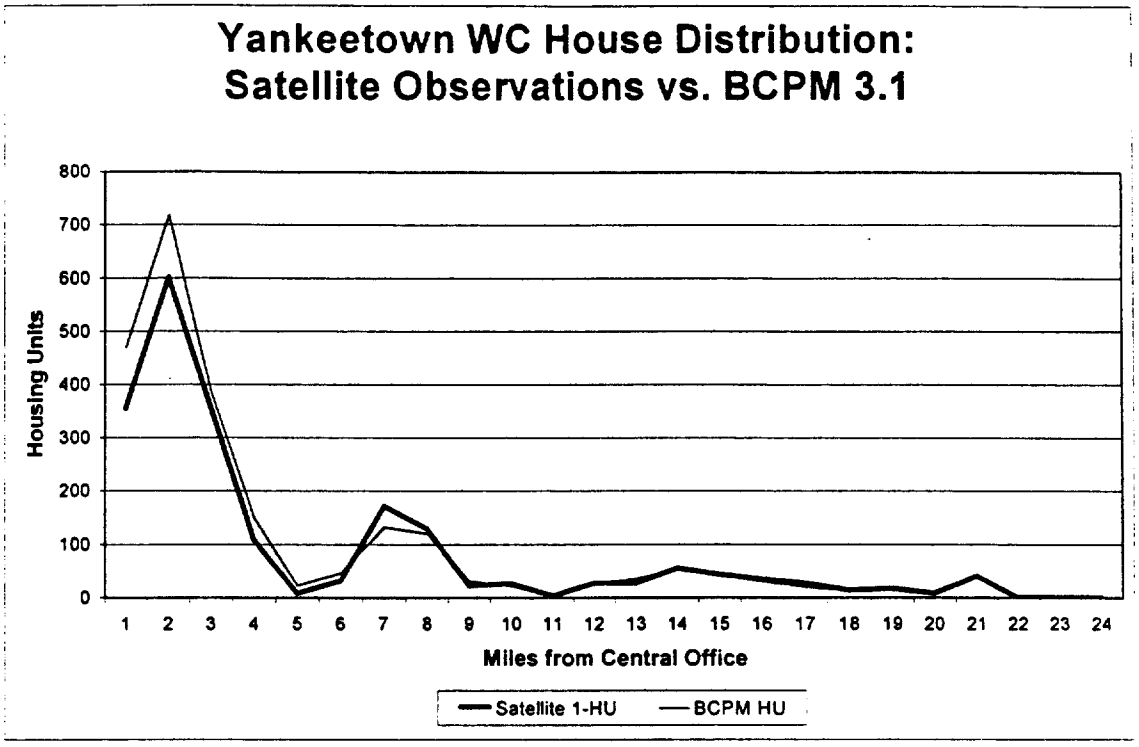
For every case, let us construct a square Census Block, conveniently (for calculation) exactly 16,000 ft. by 16,000 ft. Let us place 16 subscriber locations as surrogate points in that Census Block.

In the first case, we place these points uniformly along the periphery of the Census Block, exactly as is done for the current HAI Model. When we calculate the Minimum Spanning Tree of this set of points, we find it to be 60,000 ft., the length of the full perimeter *minus* the distance between two adjacent points.



Suppose, instead, we were to place our points uniformly distributed *within* our square Census Block. One might think that this would make them less dispersed. But then there is a set of "inner" connections to make. On the next page, as the first figure, we see one of the possible configurations of Minimum Spanning Tree for

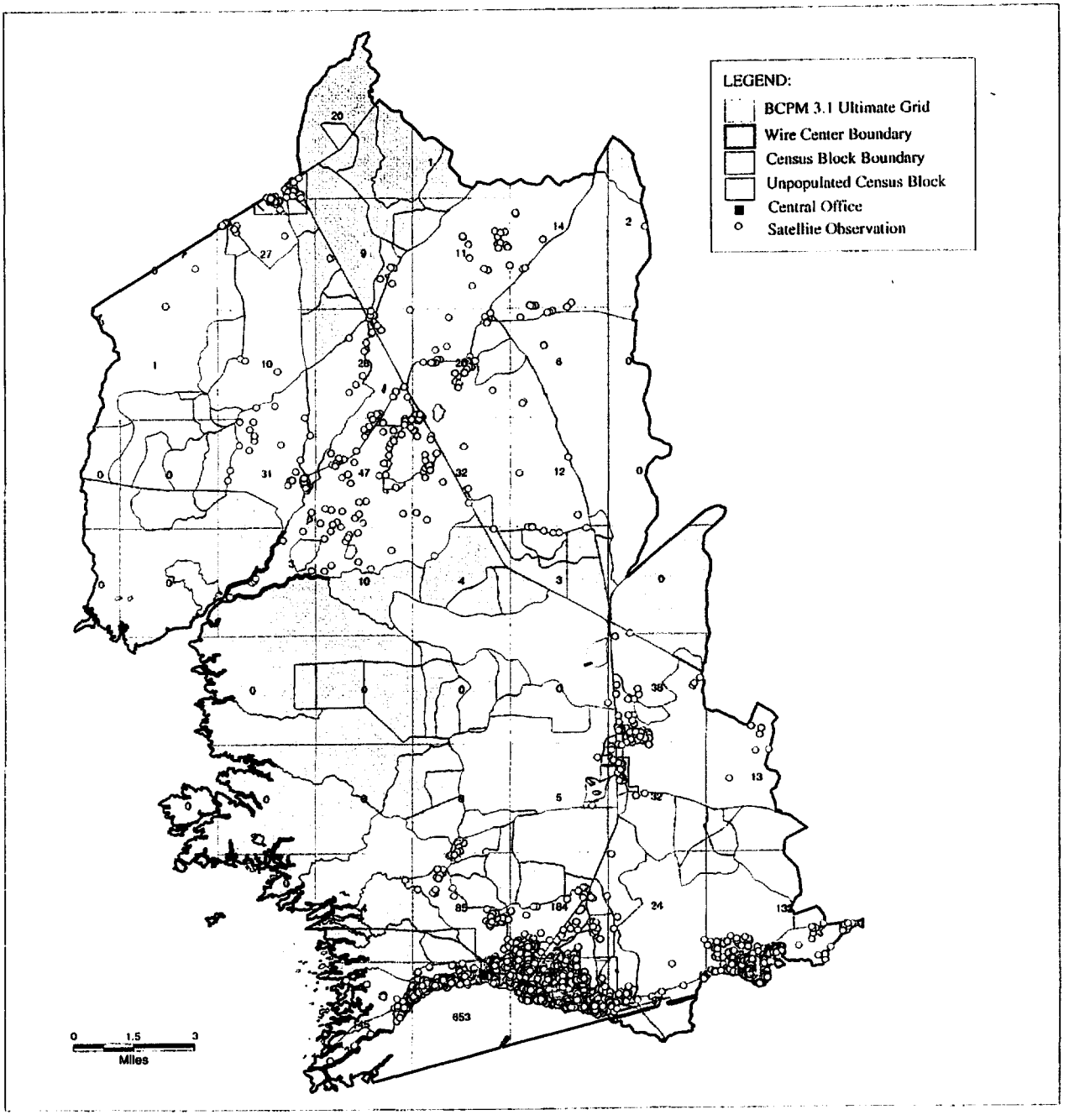
Figure 1. Yankeetown Wire Center: Distribution of Actual and BCPM Predicted Counts



Yankeetown Wire Center

Levy County, FL

BCPM 3.1 Ultimate Grids Labeled with Housing Units and Satellite Observations



Wire Center FL 07991 01303
CLLI YNTWFLMA
Exhibit KDD-10

HAI Distribution Cable Requirements

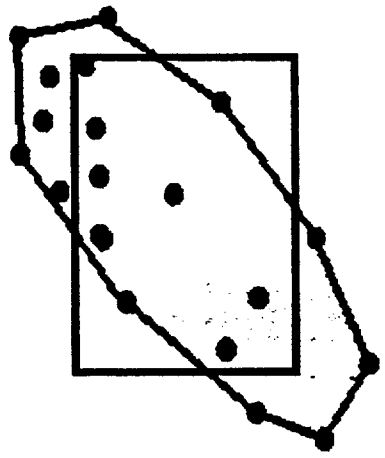
Issue: Whether the distribution plant modeled by HAI 5.0a is adequate to serve customers in their "actual" locations as identified by PNR and Associates (PNR).

Finding: The distribution route miles modeled by HAI 5.0a are too few to serve the customers in the *convex hull* clusters of geocoded and surrogate locations that underlay the rectangular clusters. The rectangular clusters are used in HAI 5.0a in the design of the network.

Hence, HAI 5.0a's estimate of the required investment in rural, low-density areas is too low.

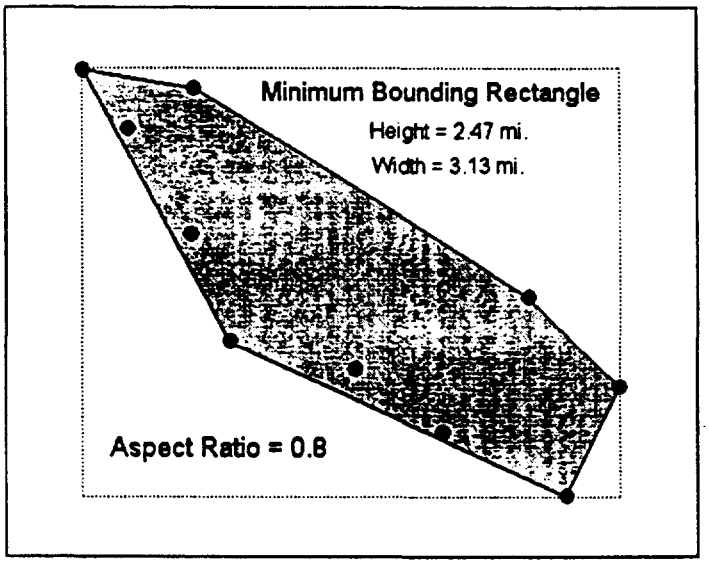
Discussion: The customer locations assumed by HAI 5.0a for the purpose of "building" plant are inconsistent with the "actual" locations in the underlying polygon (convex hull) clusters.

The figure below shows a hypothetical convex hull cluster of geocoded and surrogate locations. The rectangle shown is derived from the North-South, East-West aspect ratio and area of the convex hull cluster. Specifically, the rectangle has the aspect ratio of the rectangle that just covers the convex hull cluster (a *minimum bounding rectangle*) and the

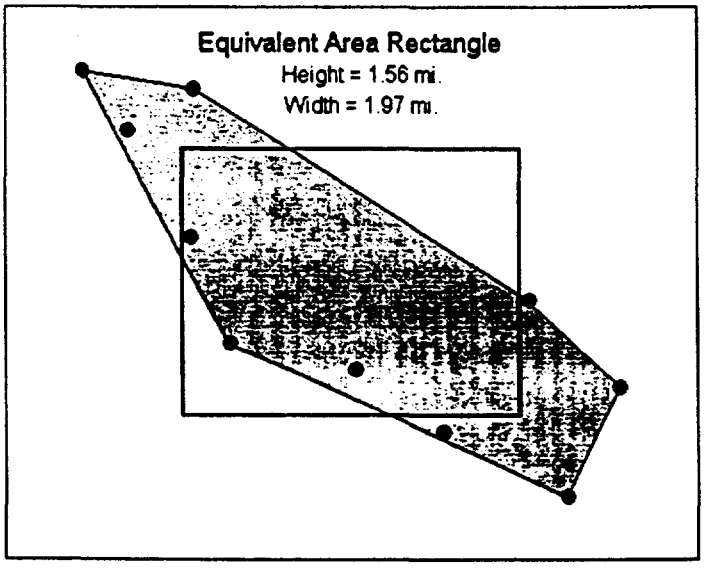


area of the convex hull cluster itself. The rectangle cluster is what is directly used by HAI 5.0a in its design of the network.

HAI 5.0a assumes that customer locations (i.e., lots) are evenly distributed within the rectangular cluster. For simplicity, assume there are 9 locations. This yields the following figure.

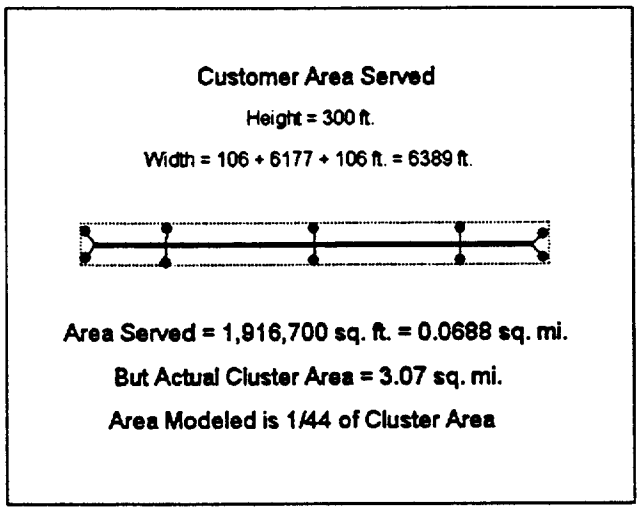


HAI then constructs a *rectangle* with the above aspect ratio; the *size* of that rectangle is determined by its *area* ... and that area is set to be the *area of the convex hull* ... in this case, 3.07 square miles.



HAI then constructs *lots* within this constructed rectangle. Each lot is twice as high as it is wide.

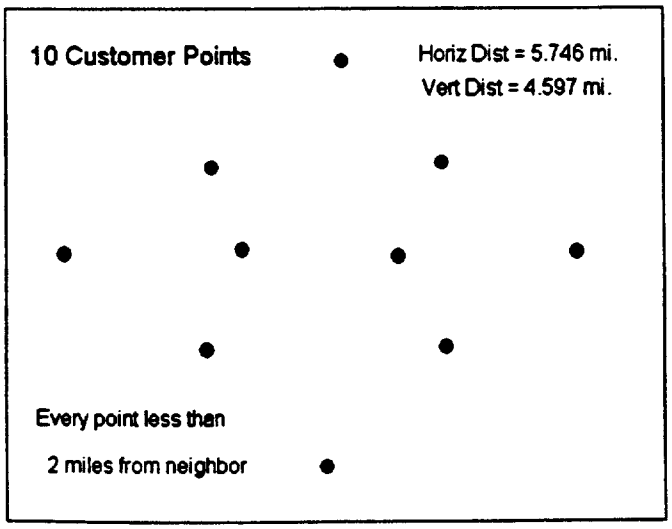
But note how closely the customers are squeezed toward the branch cable. The arrangement is unrealistic, both from the standpoint of cable length *and* from the standpoint of area served.

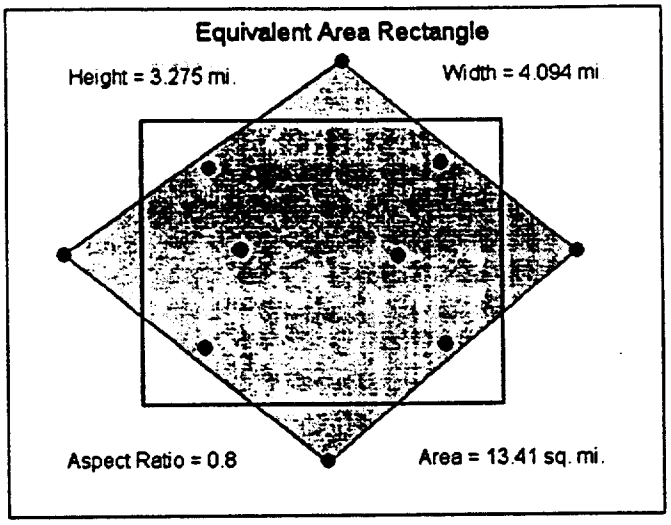


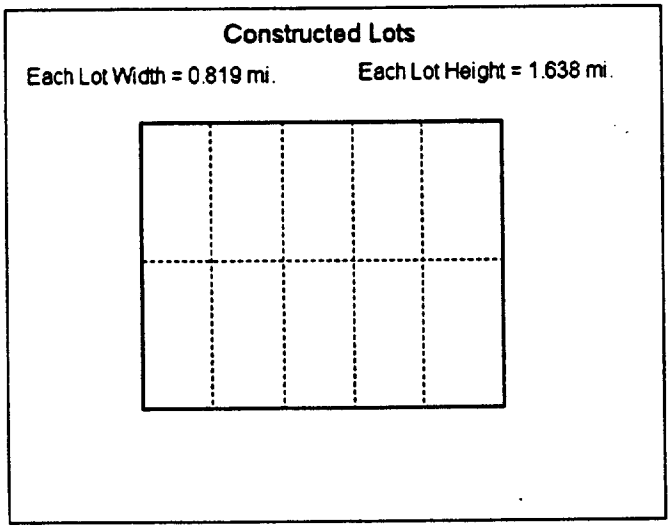
Hence, for this example, the distribution plant route miles modeled by HAI 5.0a are only 25 % of the minimum amount required to connect the 9 customers in their "actual" locations. Moreover, the area modeled as containing distribution plant is only 2 % of the area of the polygon (convex hull) cluster.

Example #2

The next example considers a much larger cluster, similar in size and density to which HAI 5.0a models in low-density areas.

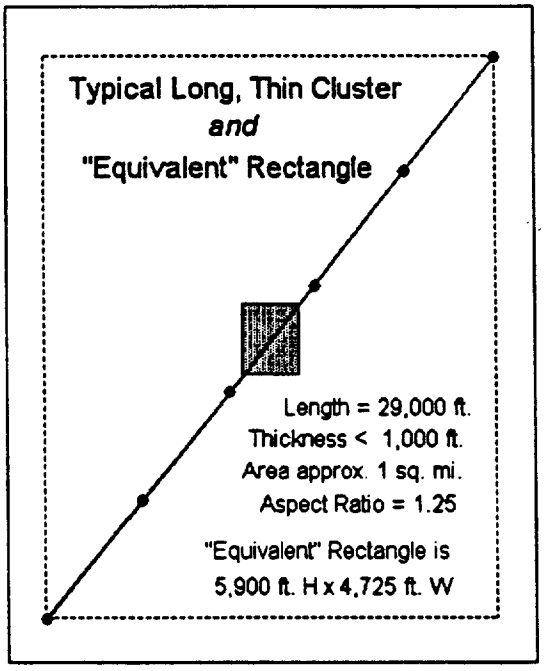



Equivalent Area Rectangle
Height = 3.275 mi.
Width = 4.094 mi
Aspect Ratio = 0.8
Area = 13.41 sq. mi.


Constructed Lots
Each Lot Width = 0.819 mi.
Each Lot Height = 1.638 mi.

sometimes it is curved, and sometimes it bends. But very typically, the convex hull of the resulting cluster is long and skinny.

The figure below shows a long and thin convex hull cluster that can occur in rural areas. The cluster consists of 6 locations strung out along a relatively straight line (road). The length of this string is 29,000' with a width of less than 1,000'. The minimum bounding rectangle for this cluster is also shown and is assumed to have an aspect ratio of 1.25. In this example, the equivalent area rectangle has an area of approximately 1 square mile.

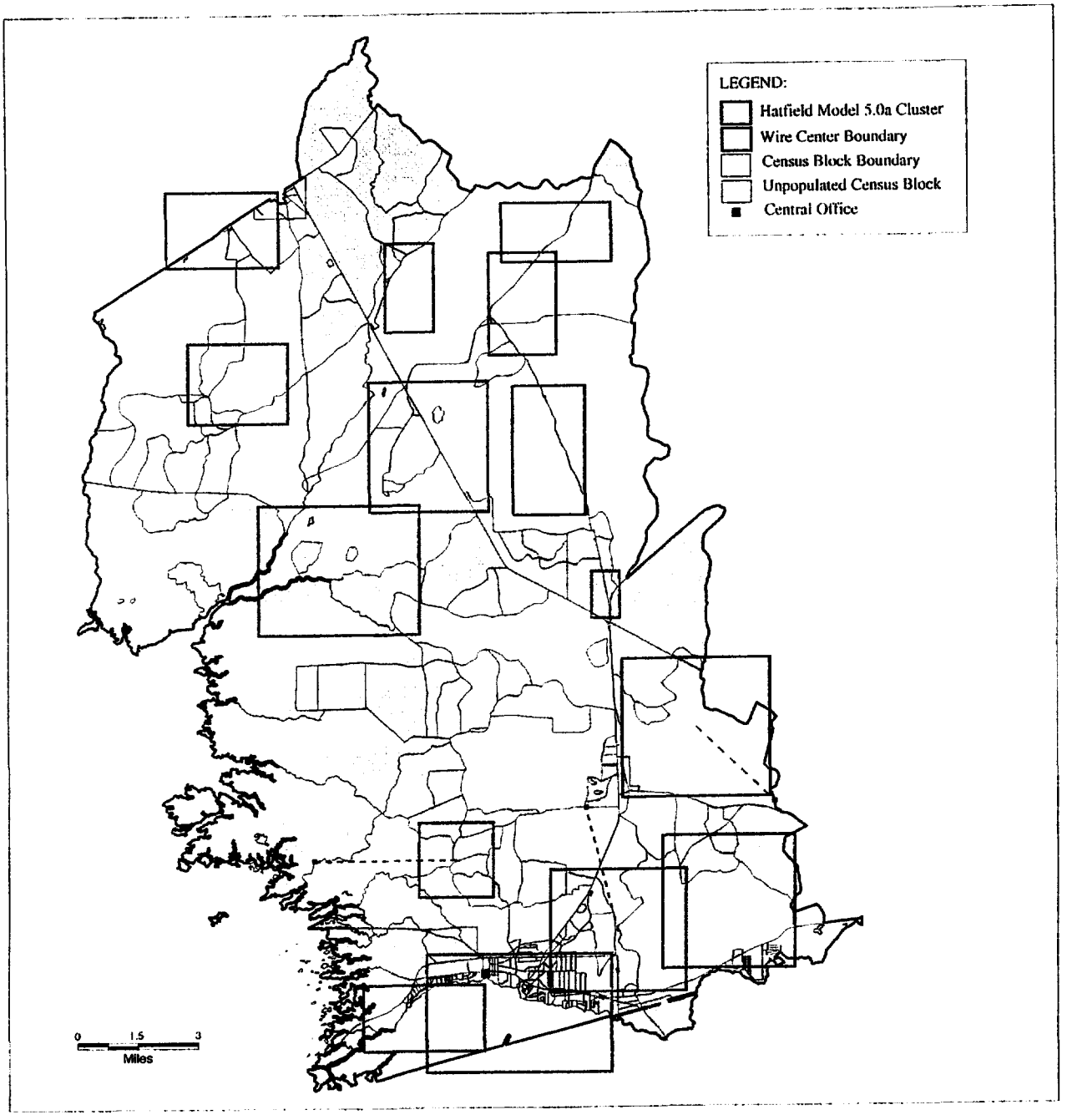


Assuming 6 locations in this cluster yields 6 plots, each 0.17 square miles in size. The HAI distribution module algorithm then assumes each lot is twice as deep as it is wide. This yields lots that are 3,048' deep and 1,524' wide.[4]

HAI 5.0a conceptually models this cluster as consisting of 2 rows of lots (East-West). Since twice the lot depth exceeds the North-South dimension

[4] Note that the HAI algorithm is not consistent with respect to the aspect ratio of lots versus the aspect ratio of the equivalent area rectangular cluster. The aspect ratio of a lot is independent of the aspect ratio of the rectangular cluster and is always 2. Thus, in this example, the sum of the lot depths (3,048' x 2 = 6,096') exceeds the "depth" of the rectangular cluster (5,900').

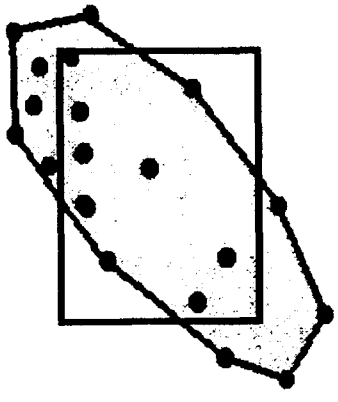
Yankeetown Wire Center
Levy County, FL
Hatfield Model 5.0a Clusters



Wire Center FL 07991 01303
CLLI YNTWFLMA
Exhibit KDD-12

Figure 2. Stylized PNR Polygon Cluster and the HAI 5.0a Equivalent Area Rectangle (Access Database)

Figure 3. Formation of the HAI 5.0a Rectangular Clusters



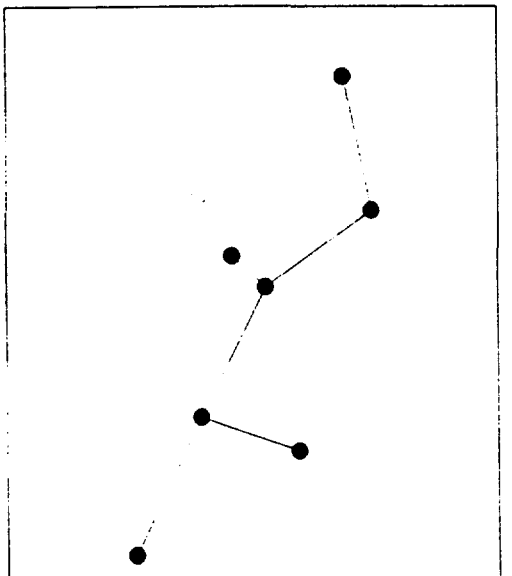
Using Minimum Spanning Trees to Estimate Subscriber Dispersion and Minimum Network Length

Phil Bolian, Stopwatch Maps
For *INDETEC* International

I. Background

A *Minimum Spanning Tree* is a construct from graph theory. It is commonly used in network design as a measure of the *dispersion* of the points to be served by a network, and as a benchmark for the *shortest possible* length of a network to serve those points.

For a set of *points* (we would say "subscriber locations"), a *Spanning Tree* is a set of straight line segments that connect *every* point (subscriber), simply drawing a line from one point to another, using no excess lines. If there are *N* points, there will necessarily be *N - 1* of these line segments.



A Minimum Spanning Tree

The *Minimum* Spanning Tree of a set of points is that set of connecting line segments whose total length is the *shortest possible* for this set of points.

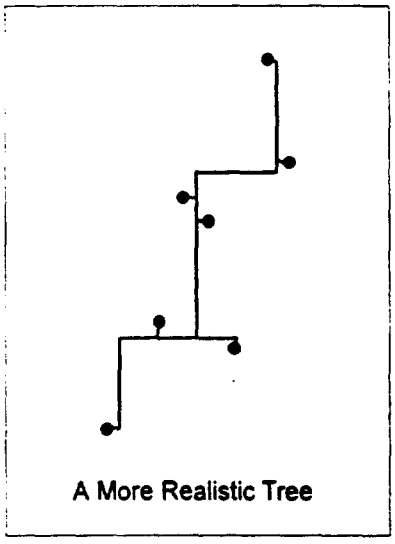
If you know the distance from every point to every other point in a set, it is not difficult to construct, and to determine the length of, the Minimum Spanning Tree of those points. The famous algorithm for calculating it, published in 1957 by R.C. Prim of Bell Labs[1], uses this simple logic:

- First, find the two points that are closest to each other and connect them
- Then repetitively, until all points have been connected, find the shortest distance between any already-connected point and any not-yet-connected point, and connect those points

As Prim pointed out in his paper, there is one and only one *shortest total length.*

[1] R. C. Prim, "Shortest Connection Matrix Network and Some Generalizations," *Bell System Technical Journal*: 36, 1389-1401, November 1957

Minimum Spanning Tree has no respect for rights-of-way, and a telephone network *must* respect them, the Minimum Spanning Tree regularly understates the minimum practical network length.



In this figure, we have constructed a more nearly realistic part of a network, running along what would be streets or roads. Even having tailored this sub-network to this exact set of points, we find the length of the tree in this figure to be 18% greater than the length of the Minimum Spanning Tree for those same points. To account for future growth, real telephone networks can not be tailored so tightly to a static set of customers, and are therefore even *less* efficient of length than in the illustration at the right.

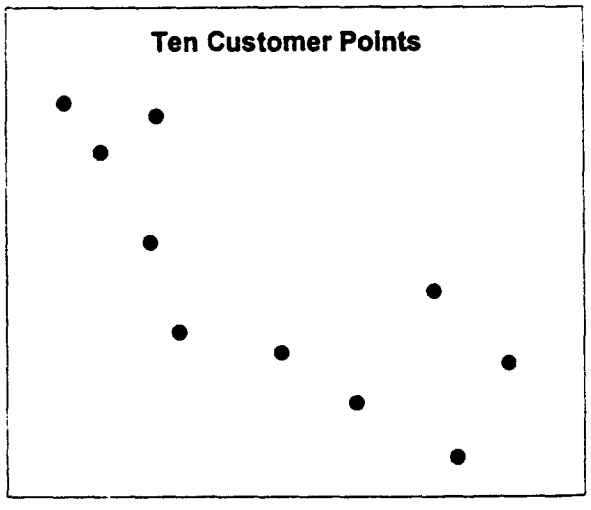
We know that a common rule-of-thumb factor used by telephone engineers to convert arbitrary straight line distances (such as are used in a Minimum Spanning Tree) to realistic cable runs is the square root of 2, or 1.414. It would be no great leap to consider that a reasonable minimum network would be something like 1.414 times the length of the Minimum Spanning Tree of the points served.

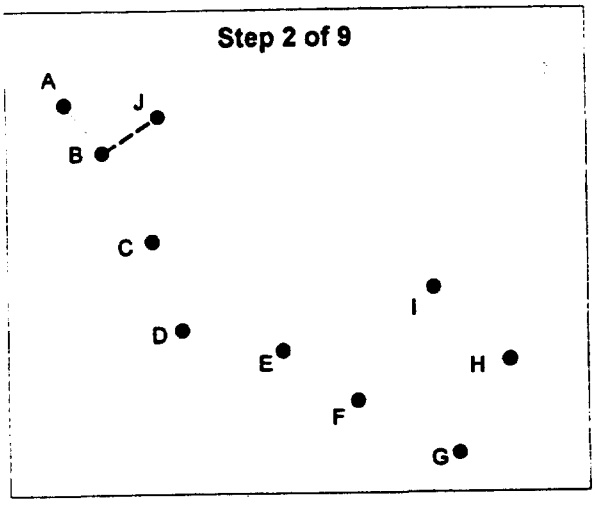
II. How a Minimum Spanning Tree Is Formed

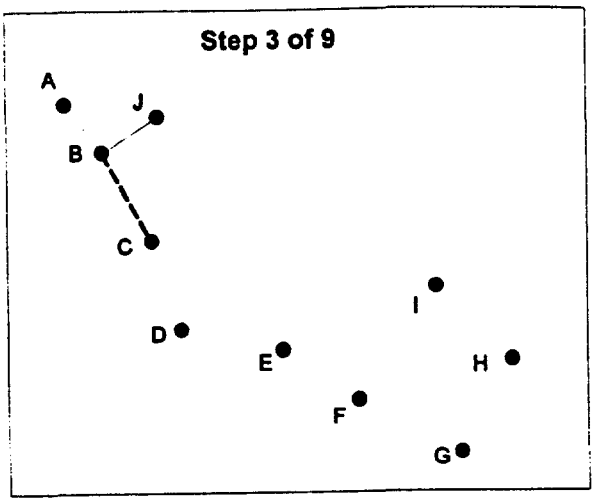
The principal reason that a Minimum Spanning Tree is so much used as a measure of dispersion of a set of points is that it is a relatively easy metric to calculate.

This section illustrates the calculation of a Minimum Spanning Tree for the ten points shown at the right, step by step.

So that we will be able to identify those points in this discussion, let's label each with a letter, as we show directly below.

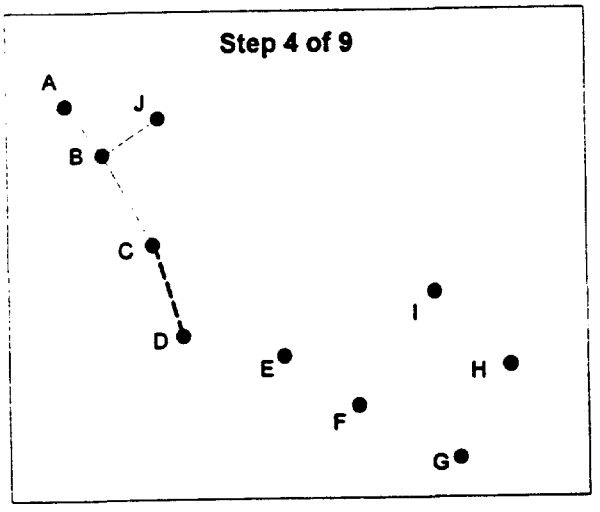


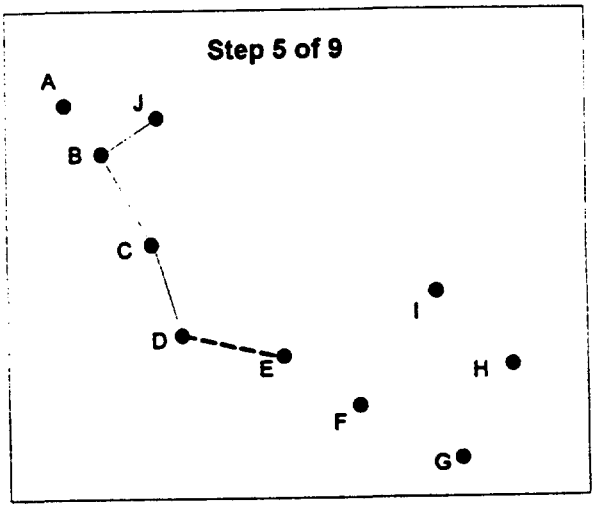




From **A**, **B**, and **J**, the shortest connection to any other point is from **B** to **C**. So we'll connect them, as seen on the right, above.

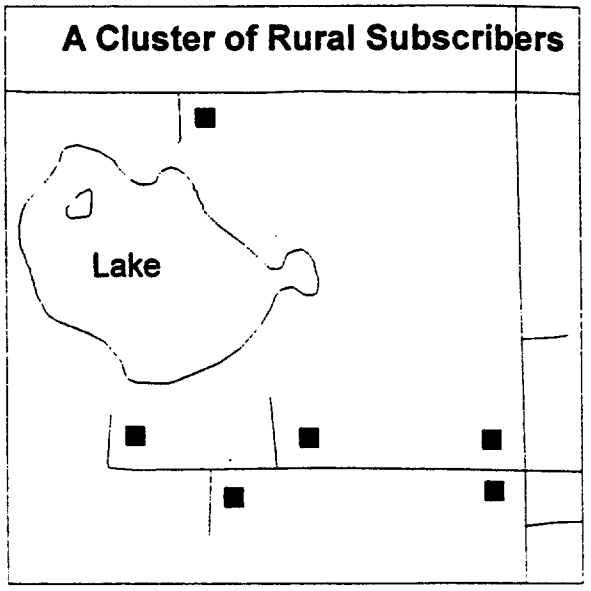
The process continues following the same rules until all points have been connected. We show the complete sequence below.





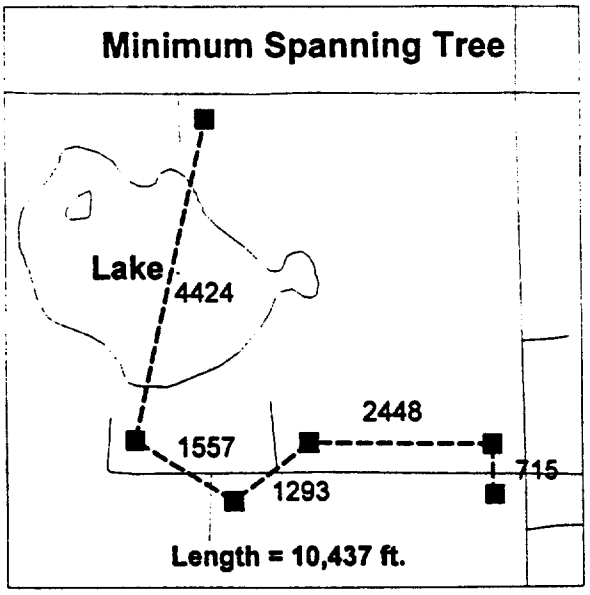
III. Minimum Spanning Tree vs. Actual Cable Route

Here is an example of the relation of Minimum Spanning Tree and a possible cable route to serve a cluster of subscribers in a rural area.



We must remember that Minimum Spanning Tree is an arbitrary, mathematical measure that has no respect for natural obstacles nor humanly restricted of rights-of-way. It simply measures the straight-line distance from one subscriber point to another, using the shortest set of straight lines possible. If that should lead through a cow pasture, a body of water, or a high mountain, the calculation does not care. And it certainly does not consider that cables basically run along roads ... the calculation makes use of nothing other than the location of each of the points, and the distance of each point from each other.

So the Minimum Spanning Tree that would be produced for this configuration of subscribers is as shown at the right. The line segments connect the points from one to another, always with a straight line, and always using the shortest set of line segments possible. The fact that several of these line segments run obliquely across a road is natural ... the calculation is not even aware of roads. And the fact that one of the segments runs across a lake is, once again, a natural result of a mathematical procedure that always seeks the shortest straight-line distances and knows nothing of obstacles.



Here we have shown the length, in feet, of each of the line segments of the Minimum Spanning Tree. The total length is 10,437 feet. We will be hard pressed to devise a realistic cabling route that can match that length, because cable routes – unlike abstract mathematical procedures – are compelled to honor natural and man-made restrictions.

The cable route is compelled to follow roads. In this case, we have run the cable along the side of the road that favors the largest number of subscriber points. We show here the length of each

The "Shorter-Than-Minimum-Spanning-Tree" Fallacy

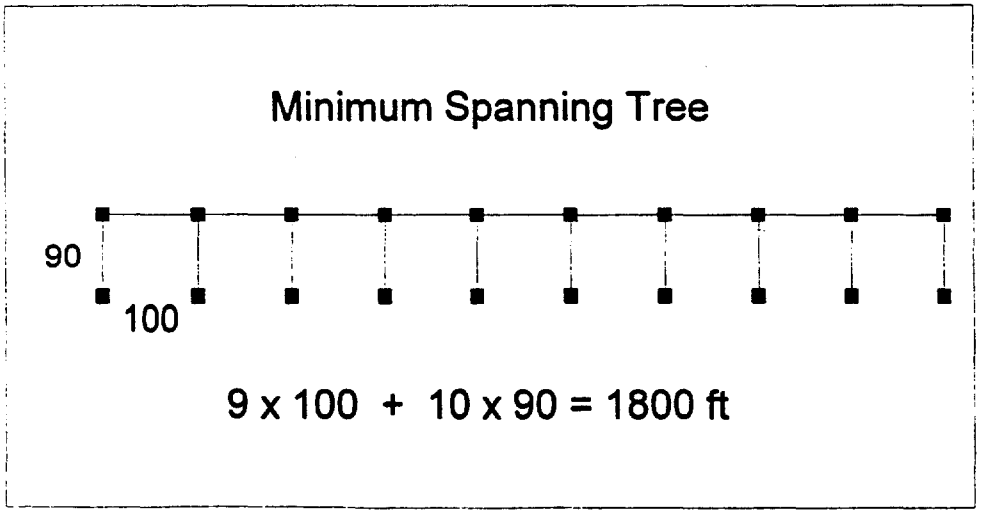
By Phil Bolian, Stopwatch Maps
For *INDETEC* International

It is certainly true that the classic Minimum Spanning Tree construct allows branches only at the existing nodes of a graph. It is also true that – in a few very special cases – the deliberate insertion of additional nodes might produce a slightly shorter tree than the Minimum Spanning Tree. In a telephone network, additional nodes may be introduced at will. Thus one might argue that it is at least *conceivable* that some cabling in a telephone network could be slightly shorter than the measure of a Minimum Spanning Tree. That argument would certainly require an example to illustrate the case. However, such examples are difficult to develop.

In a June 10, 1998 ex parte to the FCC, AT&T and MCI present an example purportedly illustrating part of a telephone network that uses less cable footage than the measure of the Minimum Spanning Tree for the subscribers to be served. The example is based on the premise that on a typical suburban street, running cable down one side (or the middle) of the street, and extend drops to each house, will yield less DRD [Distribution Route Distance] than the Minimum Spanning Tree distance.

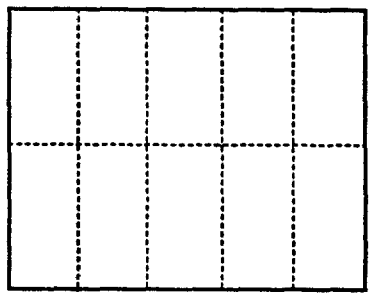
Unfortunately for AT&T and MCI, the example they cite does *not* prove their point. In fact, it proves them wrong. Let's examine the circumstances AT&T and MCI cite.

Imagine a suburban block, with ten houses on either side of the street. Imagine them evenly spaced. In this first example, let the lot sizes be 100 feet, and let the distance from the front of one house to its cross-street neighbor be 90 feet (in a later example we'll reverse those distances). The Minimum Spanning Tree length for these original locations is 1,800 feet.



Constructed Lots

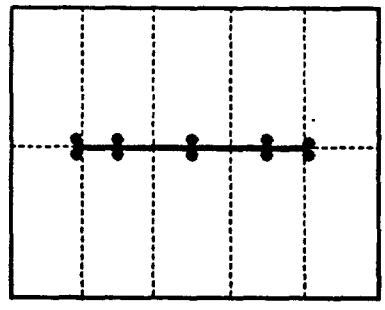
Each Height = 0.78 mi. = 4118 ft.
Each Width = 0.39 mi. = 2059 ft.



In this example, there is no backbone cable, only a branch cable. The DLC site is at the centroid of the rectangular cluster. 150-ft. drops connect to the customers.

Cabling to Serve Customers

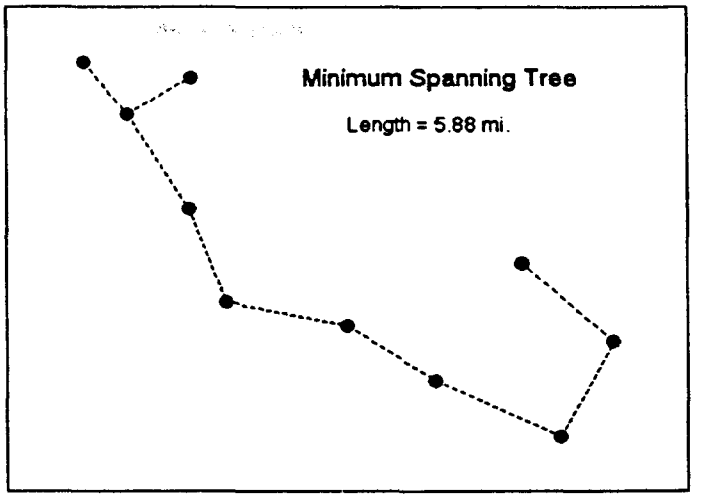
Branch Cable Length = 6177 ft.
10 Drops, each at 150 ft.



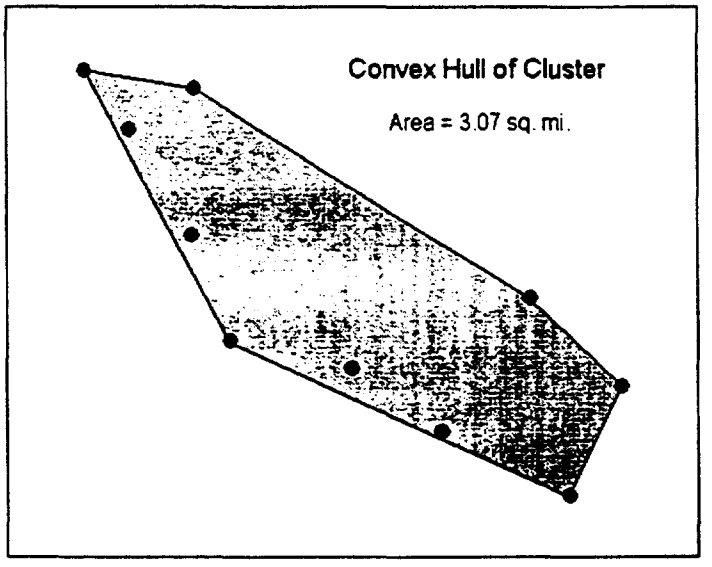
Total Cable Length = 7677 ft. = 1.45 mi.

Less than 1/4 of the Minimum Spanning Tree length!

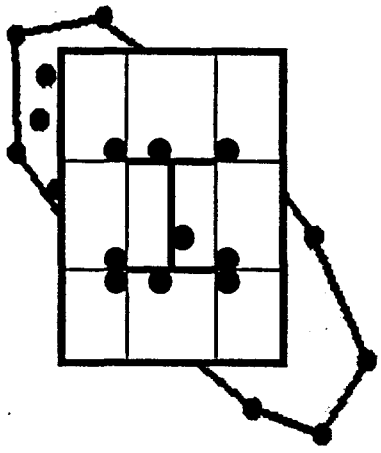
The *minimum spanning tree* for these points – the mathematically shortest connection possible for these points – is 5.88 miles.



When HAI has determined the set of points that constitute a cluster, it logically draws a *convex hull* around those points, and determines its area.



HAI then logically constructs a *minimum bounding rectangle* – oriented North-South-East-West – which exactly bounds the cluster's points. HAI then determines the *aspect ratio* of that rectangle (that is, the ratio of the rectangle's height to its width) … in this case, 0.8.



HAI 5.0a subtracts off two lot depths from the cluster North-South length to determine the length of the backbone cable. It also subtracts off two lot widths from the East-West cluster length to determine the length of the branch cable. In the figure shown above, there are two branch cables. Backbone and branch cable is laid in only the middle lot. A drop serves the house in each lot.

Since the default drop length in the lowest density area is 150 feet, the house in each lot must be 150 feet from a branch cable. That is, the houses are concentrated toward the center of the rectangular cluster as indicated in the figure.[1]

This has an important implication for whether the model is providing for a realistic amount of cable. Assume that the area of the convex hull is 15 square miles. Hence, the area of the rectangle is the same and the area of each lot is roughly 1.67 square miles. Lots are assumed to be twice as deep as they are wide. Each lot is 1.83 miles deep (9,640') and 0.91 miles wide (4,820'). Thus, the total distance of cable, including the 150' drops, in this cluster = 9,640' + 2*4,820' + 9*150' = 20,630' or 3.91 miles.

Examining the underlying convex hull cluster of geocoded and surrogate locations strongly suggests that this amount of cable is much too little to serve customers in their "actual" locations. That is, the placement of customers for determining cable lengths within the rectangular clusters is inconsistent with where PNR locates customers in the underlying polygon

[1] As modeled by HAI 5.0a, it is only the distance from the cluster center to the edge of the middle lot (in this example) that matters for determining whether multiple DLCs are needed.

Description of Alternative "Road" Surrogating Methodology

The current PNR procedure followed to determine customer locations is to:

A. Geocode each customer's location to its precise street address; and, if this is impossible

B. Assign that customer a "surrogate" geocode by assuming that unlocated customers actually are located uniformly along the perimeter of their Census Block.

In general, over 70 percent of all customer locations are determined by stage A, and the remaining 30 percent or less are determined by the "surrogate" location process.

The following routine could be employed as an alternative to stage B of the current surrogating methodology. Rather than assuming the geocodes of unlocated customers to be uniformly along the perimeter of their Census Block, it assumes that these unlocated customers are located in specified density or proportions (not necessarily uniform) along a specified selection of road types inside of and bordering their Census Block.

1) Find all the roads of the type that houses could be located on. We are currently using the same roads that are claimed to be used in BCPM3: A21, A24, A25, A28, A31, A34, A35, A38, A41, A44, A45, A48, A61, A62, A64

2) Sort by segment centroid and then remove duplicate segments. Duplicates occur because the same segment can occur twice if the road has 2 different names.

3) String together all of the segments with common TLIDs (Tiger Line Identifications).

4) Build an ordered list of all segments. This is done by ordering all of the polylines created in step 3.

5) Using boundary roads, starting in the south west, continuing along connected lines, using the remaining far south west segment when breaks occur.

6) Using interior roads, starting in the south west, continuing along connected lines, using the remaining far south west segment when breaks occur.

7) Compute the total equivalent distance of the segments.

   - Double the distance of interior roads because houses can be located on either side of these roads – as opposed to only on one side of exterior roads.

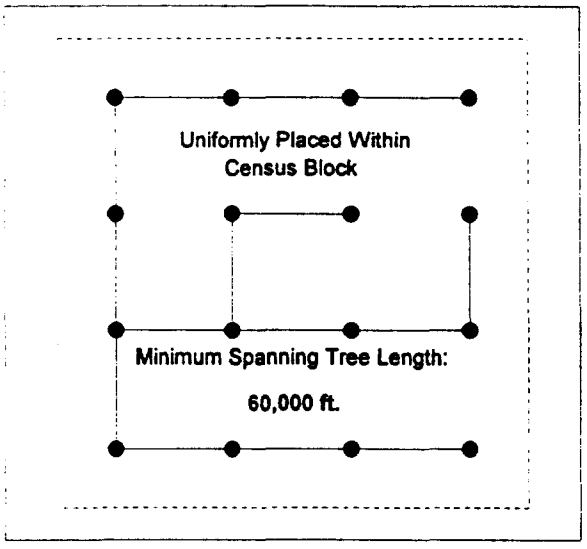
   - Inflate or deflate the assumed distance of all segments of a particular road type based on the assumed relative density of customer locations along this type of road versus the average type of road.

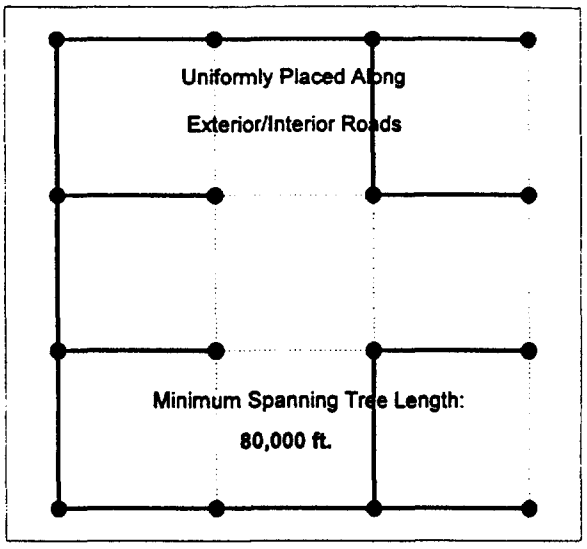
8) Based on the number of surrogate points to be inserted, compute the distance between points to be inserted as

   DBP = Number of Surrogates / Total Equivalent Road Distance.

9) Starting at a distance of DBP/2 into segment 1, and for every DBP of road thereafter, save the latitude and longitude as a surrogate customer location point.

this uniform placement *within* the square (but, of course, *every* configuration of Minimum Spanning Tree for that placement of points has exactly the same length). Surprisingly or not, it is *again* 60,000 ft.





Then what of a placement along interior as well as exterior roads? In the figure at the right, above, imagine that there are two east-west and two north-south interior roads, and that the bounds of the Census Block are also roads. Then, if we place these points uniformly along *all* roads, we find that the dispersion of the points has *grown*, not diminished. The Minimum Spanning Tree of this configuration is 80,000 ft.

In other words, the placement of surrogate points uniformly on the periphery of a Census Block is *not* a *more* dispersed configuration of points than the other two placements we have investigated here. In fact, it is *less* dispersed than the second alternative. Said yet another way, neither of the two alternative placements presented here would reduce the Minimum Spanning Tree of these points … One would even extend it.

We have examined the dispersion of uniformly placed surrogate points *in a single* Census Block, and found that the placement for surrogates used by HAI 5.0a is *not* the most conservative placement available. We do not even address the fact that if two *adjacent* Census Blocks have surrogate points placed along their peripheries, the points along a common boundary will be far closer together than if they had been spread throughout the areas of each Census Block.

1

EXHIBITS

Third, a key validation test is whether the models estimate enough distribution cable distance to at least connect customers, as the crow flies, in the locations identified by the models.

Once customers have been located and aggregated into serving areas, HAI 5.0a and BCPM use different modeling tools in the estimation of the distribution distance needed to connect customers to each other and to the network. The focus should not be on the assumptions behind these tools but on the estimated distances that result from the application of these tools. Specifically, the focus should be on whether the models estimate enough distribution cable distance to connect customers in the locations identified by the models. In the case of HAI 5.0a, these are the geocoded and surrogate locations within the PNR polygon clusters. In the case of BCPM 3.1, these are the microgrids within the ultimate grids.

The minimum spanning tree (MST) test, offered in my testimony, is a test of a model's internal consistency in this regard, i.e., whether it does what its purports to do based upon its own modeling assumptions. When applied to HAI 5.0a and BCPM 3.1, the test indicates that the HAI 5.0 contains a substantial shortfall. In the lowest density zone, the model's estimated distribution distance (including drop and connecting cable) is less than its MST distance in 87% of its main clusters. For the same density zone, BCPM 3.1's estimated distribution distance (including drop and connecting cable) is less than its MST distance in substantially fewer ultimate grids. Overall, the HAI 5.0a shortfall totals at least

CLUSTERS?

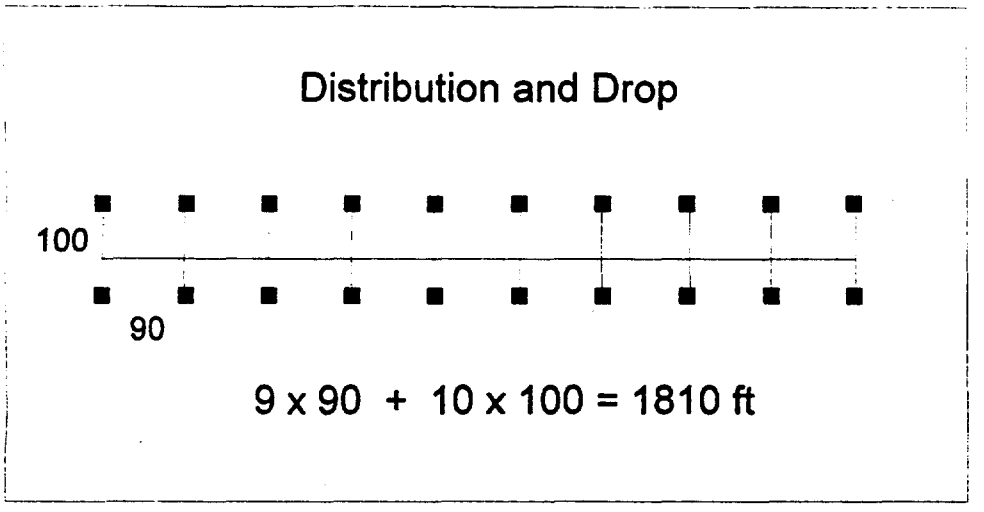
A. Not substantially. Table 11 presents the results of the HAI MST test, in the same format as Tables 9 and 10, for HAI serving areas defined in this manner. As Table 11 indicates, the addition of the outlier clusters reduces by 0.89 million feet (169 miles or 9%) the total shortage for BellSouth's Florida territory. In the lowest density zone, < 5 lines per square mile, the share of "servings areas" that are short declines from 87% to 76%. The comparable figure for BCPM 3.1 (from Table 10) is 32%. Including outliers improves the HAI model's showing in this test because the T1 road cable distance between the outliers is estimated assuming rectangular routing while the MST is the straight-line distance.

Table 11. HAI 5.0a Distribution Route Distance Understatement: Default Drop Lengths, Expanded Serving Area Definition, BellSouth Florida

Data for Only Serving Areas That Are Short

DZ	HAI SA Dist Route Feet Shortage	MST for Short SA	% Short	Number of SA Short	Number of SA in DZ	Number of SA Short in DZ (%)
< 5	2,314,677	6,789,656	34.09%	120	157	76.43%
5 - 20	4,016,334	15,756,075	25.49%	256	396	64.65%
20 - 100	1,697,531	6,980,288	24.32%	138	415	33.25%
100 - 200	295,974	1,360,514	21.75%	30	227	13.22%
200 - 650	187,645	740,964	25.32%	32	604	5.30%
650 - 850	19,973	137,864	14.49%	6	216	2.78%
850 - 2,550	250,752	1,380,601	18.16%	48	1,491	3.22%
2,550 - 5,000	80,714	661,603	12.20%	31	1,376	2.25%
5,000 - 10,000	35,165	291,621	12.06%	24	832	2.88%
> 10,000	64,757	176,762	36.64%	16	234	6.84%
	8,963,523	34,275,948	26.15%	701	5,948	11.79%

VIII. SUMMARY



Hence, it is quite difficult to improve upon the Minimum Spanning Tree distance.

Attachment A

DOCUMENT OFF-LINE

This page has been substituted for one of the following:

o An oversize page or document (such as a map) which was too large to be scanned into the ECFS system.

o Microfilm, microform, certain photographs or videotape.

o Other materials which, for one reason or another, could not be scanned into the ECFS system.

The actual document, page(s) or materials may be reviewed by contacting an Information Technician. Please note the applicable docket or rulemaking number, document type and any other relevant information about the document in order to ensure speedy retrieval by the Information Technician.

One CD-ROM