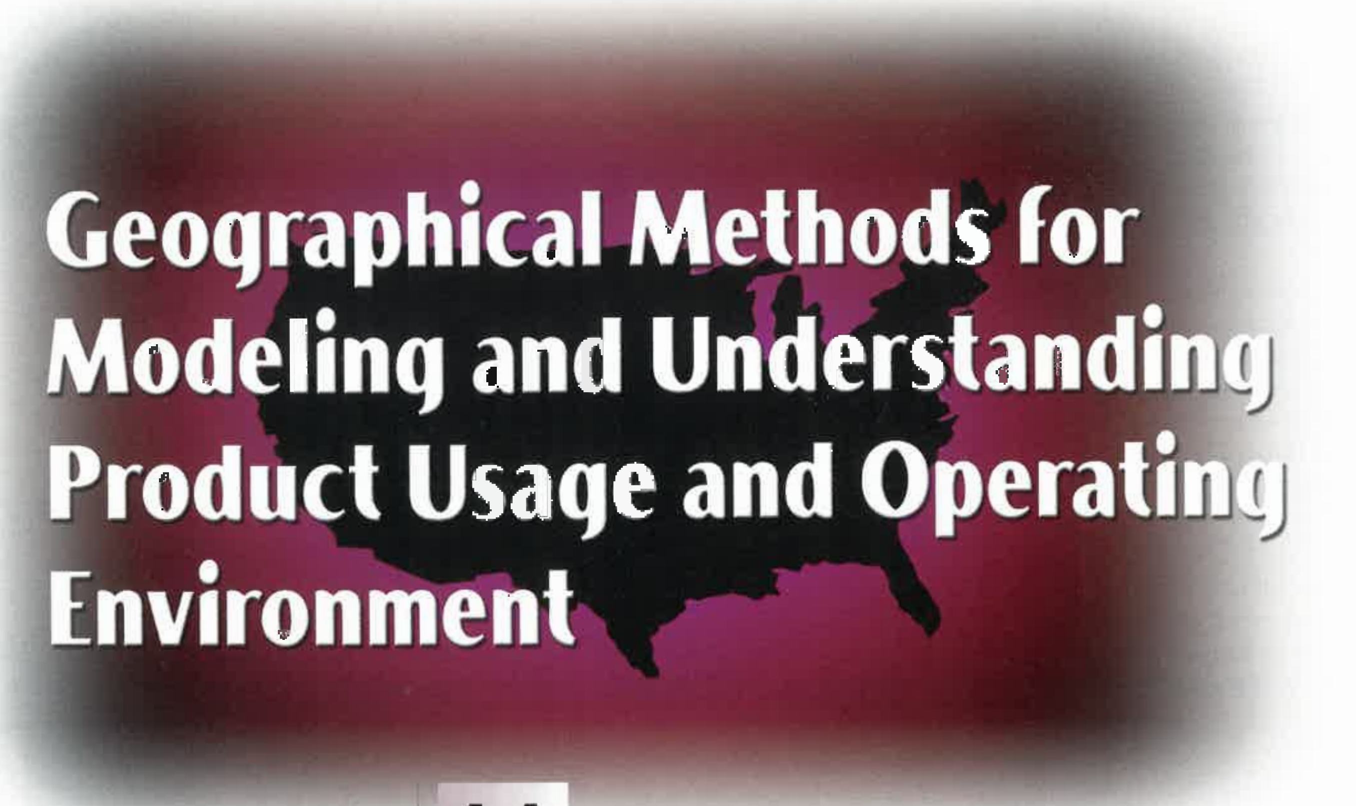


Geographical Methods for Modeling and Understanding Product Usage and Operating Environment

by **Graham S. Duthie**
Visteon Corp.*

Understanding how and under what conditions customers use products is critical to making sound engineering design decisions. Traditionally, new generation products are often based on prior designs and test specifications that may or may not subsume the range of conditions in a given set of operating environments. This article reviews geographical methods that can aid design engineers in their understanding of what the range of product operating environments will be and the proportion of products that may be in service under a given set of conditions. Ultimately, geographical methods can enable better quantification of the risk involved when making critical design decisions—specifically those decisions related to potentially marginal product function or durability within a certain geographical area.

INTRODUCTION

Knowledge of the expected operating environment of a product is needed so that credible product validation tests can be created to confirm that new product designs will meet expected field life needs. During the course of product durability test development, geographical methods and geographic information systems (GIS) can be used to enable understanding of the whole range of product operating locations, and both the customers and environments associated with those locations.

Two examples of how existing data can be analyzed using geographical methods to understand product usage in the presence of a range of environmental operating conditions are presented. The first example discusses the utilization of product warranty return information to determine potential worst case operating *locations* that later can lead to discovery of the specific conditions that may cause reduced product durability. The second example describes a weighted environmental factor model that was developed to aid in the understanding of automotive air conditioning system operating condi-

As published in *Service Life Prediction: Challenging the Status Quo*, by the Federation of Societies for Coatings Technology, Blue Bell, PA; ISBN 0-934010-60-9, 2005.
*Visteon Corporation, One Village Center Dr., Van Buren Township, MI 48111. Author is now employed by Andersen Corporation, 100 Fourth Ave., N., Bayport, MN 55003; gduthie@andersencorp.com.

tions in the United States. The weighted model method is presented as an example of how the range of operating conditions can be explored using available, geographically referenced data.

HIERARCHY OF PRODUCT USAGE AND ENVIRONMENT

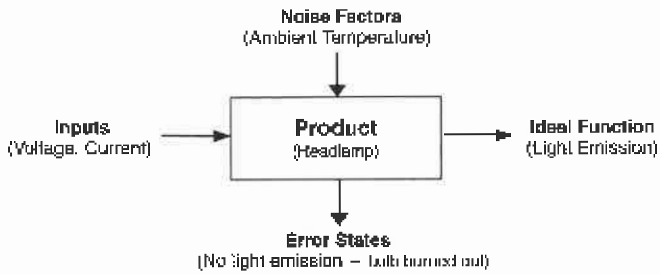
Prior to the assessment of a product's operating environment, it is useful to have a simplified model of the potential external factors that will affect a product's function and service life. The parameter diagram, or p-diagram, is a common reliability engineering tool used to depict a product, the product's function, error states, and external noise factors. The p-diagram (*Figure* 1) represents a product as a "black box" that takes inputs and performs a function that leads to expected outputs. A simple example product is an automotive headlamp—the headlamp takes in electricity and produces light. In addition to electricity, there are generally other inputs that must be understood before a robust headlamp assembly can be designed and tested. These other inputs are termed noise factors. Noise factors can consist of environmental parameters (i.e., ambient temperature), customer usage (i.e., the number of times the headlamps are turned on and off), number of car washes per year, or dimensional variation as a result of the manufacturing process. Products that are not robust to external noises may have lower reliability than expected and could enter an undesirable error state.

To augment the p-diagram's ability to represent a product in the face of external factors that may influence durability, a nested hierarchical model (*Figure* 2) of a product in the real world is proposed that suggests the connectivity of certain factors from the product itself at the lowest level up to a geographical scale at the highest level.

Also shown at the far right of the nested hierarchy is "time" to indicate the temporal aspects that exist for nearly every factor affecting product durability. At each layer in the hierarchical model it is likely that existing data for many factors can be located. The most common data needs can often be fulfilled with experimentally collected data or publicly available datasets from government sources. In the United States a wide variety of environmental and demographic data can be obtained at many spatial and temporal resolutions (addressed in *Figure* 2) as either small or large entities at fine or course grain that exhibit high or low frequency behaviors.[1]

The hierarchical product model is useful towards understanding the levels of measurement where data related to product operating parameters can be found. If a product is known to be sensitive to thermal cycling

Figure 1—Parameter diagram.



then it follows that temperature data from both the product itself (*component* level) and the local ambient temperature (*geography* level) would be needed to determine how the two temperature values relate. In an automotive environment the ambient temperature and the temperature of a component are often different at any point in time, but frequently there is a relationship between the two values that can potentially be used to model the thermal state of the component under any given ambient condition. Note that the usefulness of a model representing the range of product operating environments is dependent on obtaining georeferenced data at appropriate spatial and temporal scales. A geographic information system is the tool that makes it possible for the researcher or engineer to integrate the numerous georeferenced data sets that may be required to make a prediction of product operating states relative to external conditions or to explore product warranty data.

Figure 2—Hierarchical model of an automotive component in the real world.

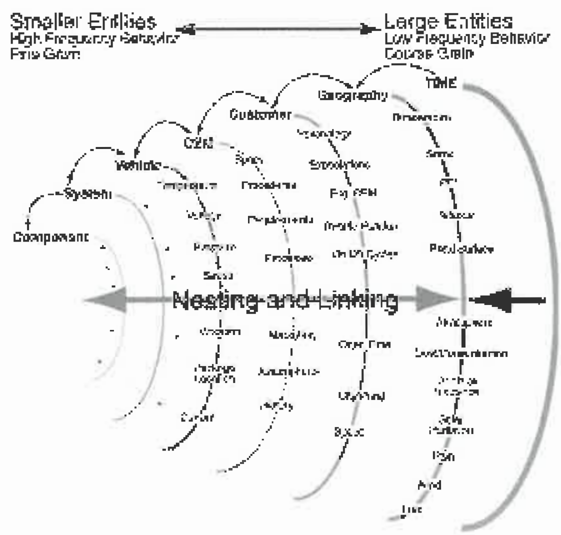


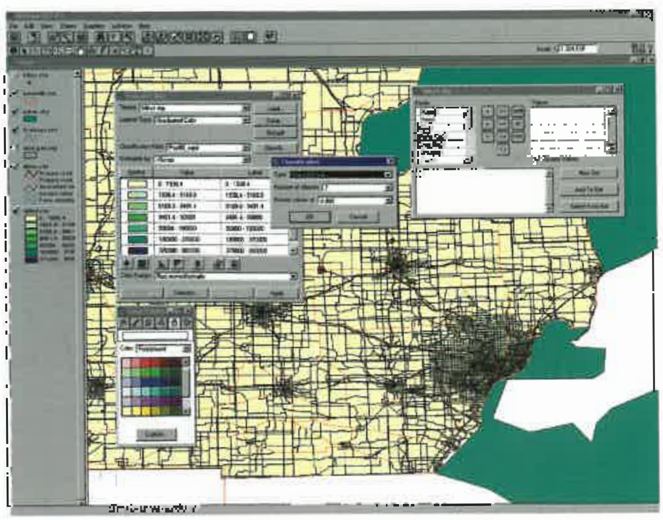
Figure 3—Three-dimensional representation of information layers and normal view of the resulting map.

A BRIEF REVIEW OF GEOGRAPHIC INFORMATION SYSTEMS

In the past, it was common to create maps by hand and use layers of transparent vellum or acetate to create overlays that would enable visualization of the relative position of one or many features to another set of features on the ground under various scenarios. Since the late 1980s growth in computing power and intensive software development has led to easy-to-use geographic information systems (GIS) that can subsume, display, and enable analysis of many types of data.

A GIS can be described most simply as data linked to a map in a computing environment. On the digital map each type of data is contained in a specific layer (*Figure* 3). For each point, line segment, or polygon within each layer, a record of data exists that contains details pertaining to the feature. For line and polygon features the GIS creates topology that describes the relative position and connectivity, if applicable, for each item. Topology enables the creation of networks (i.e., roads, pipelines, electrical grids, etc.) and provides the ability to conduct spatial queries to find items adjacent or within distance of another feature.

Figure 4—Sample screenshot from ArcView 3.2. Map shows the Detroit, MI, area.



Each feature in a GIS normally has additional attribute data. A polygon representing a building might contain the address of the structure, the owner's name, telephone number, square footage, year of construction, etc. A complete layer of all the houses in a city could be displayed along with other layers including streets, political boundaries, cell phone tower locations, sewer lines, shopping malls, flood plains, and more. The design of a GIS enables the user to perform spatial queries between multiple layers of information, such as finding all of the banks within two miles of a specific address or isolating areas where past weather conditions met a combination of user-specified criteria. The subset of features selected by the query will be highlighted in the map and can be further subset if needed. The results of a query can also be exported for use in other software packages.

Most GIS packages enable the researcher to display and interact with geographic data in many useful ways. GIS packages enable the user to apply multiple classification methods to quantities or attributes associated with the layers represented on the map. Changing the classification method alters the visual appearance of the items displayed on the map. GIS users should experiment with data classification so that the appropriate representation of the underlying distribution of values is communicated in any paper maps that are created. A sample screenshot from ArcView™ 3.2[2] is shown in *Figure* 4 to give the reader an example of a "typical" GIS display.

In the course of this article, it is not possible to provide a complete review of the capabilities and applications of GIS tools. The reader is *strongly* encouraged to investigate and evaluate the wide range of GIS software in the market; one or two applications will likely meet the reader's needs.

APPLICATIONS OF GEOGRAPHICAL METHODS AND GIS

Example 1: Geographical Analysis of Field Return Data for a Consumer Product

To effectively understand and reduce the risk that external noise factors will affect product function over time normally requires expensive physical testing of prototypes. Before any testing can be performed, a test procedure must be developed that is capable of reproducing expected field failure modes. Many products are improved versions of earlier products that have extensive warranty histories. In the case where there is warranty information available for a product, geographical techniques can be employed to help understand and

Figure 5—Map showing count of returns by location.

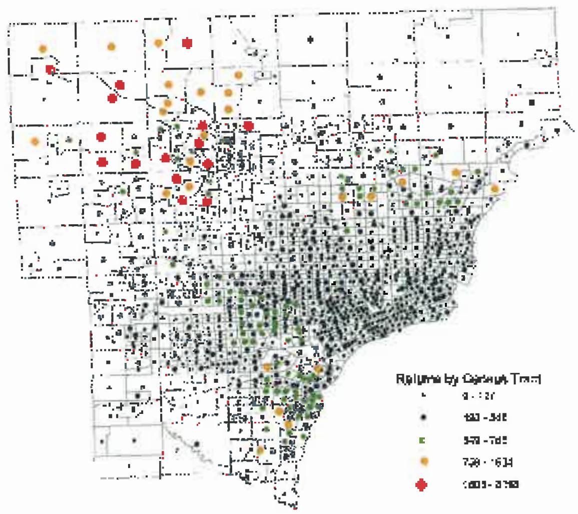
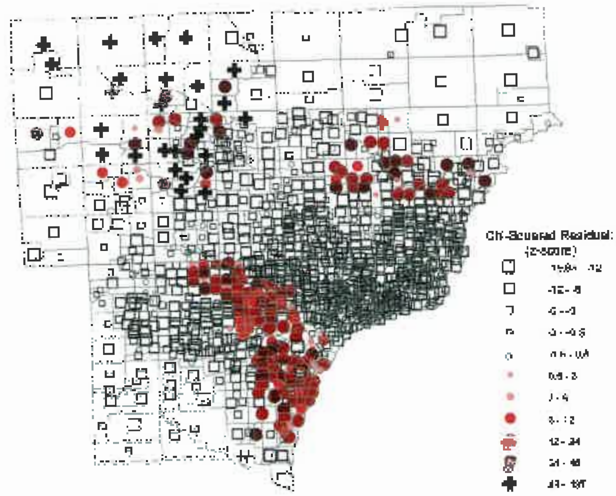


Figure 6—Map showing chi-squared residuals by census tract.



quantify environmental and customer usage factors that may have led to historic product returns from the field. Coupled with prior knowledge of the sensitivity of a product to varying environmental conditions and customer usage, the results of a geographic analysis of field return data is likely to help point the engineer or researcher toward a set of conditions associated with the locations of highest product return relative to expected returns. The base assumption in this example is that the number of returns is directly proportional to sales of the product (e.g., if 10% of the total units were sold in a given area, then 10% of the total number of product returns should come from the same area). Where it is discovered that more products were returned than expected, generally, a reason exists for the higher than expected numbers of returns—spatial randomness can be statistically tested to determine the significance of specific problem locations.

Given that the locations of both product sales counts and product return counts are known, it is possible to create a series of maps that can lead the researcher to locations where products are returned earlier or in greater proportion than expected based on product sales counts in a given area. The example presented in this article is fictional and was created using available data from the U.S. Census Bureau for the Detroit, MI area. The basic, mechanical steps to enable the creation of the maps in *Figures* 5–7 are listed below:

(1) Obtain product sales data by selling location.

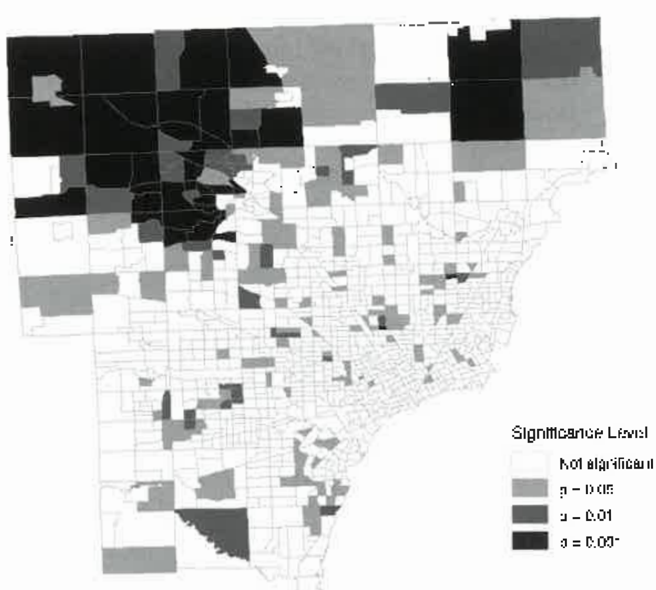
(2) Obtain returns data by return location origin. It is assumed that the return location is normally the place that the product was used. (This may not be true and should be verified if possible.)

(3) Geocode locations of sale and returns data. In the case where product sales and returns are made through a set of standardized locations, it is advised to geocode the standard locations separately, in advance of any analysis. Geocoding is a computer process that matches physical addresses with latitude and longitude coordinates.

(4) Sum the number of returns for each location.

(5) Create a map to visualize the pattern of return counts by location. Often, unusual patterns of

Figure 7—LISA analysis results showing potentially significant locations.



product returns can be detected visually at this step (see *Figure* 5).

(6) Create a set of aggregate (group) sales and return locations so that the effect of locations that sell few units is reduced. The level of aggregation should be selected based on appropriate knowledge of how a product behaves in an environment over a selected area.

(7) Sum sales and return counts by aggregate geographic area as defined in step (6).

(8) Compute chi-squared residual (referred to by Dykes as the signed chi-squared statistic[3]) statistic for the results of step (7).

$$\text{Formula 1:} \qquad csresidual = (O - E)/\sqrt{E}$$

where O = observed count of returns

E = expected count of returns

Positive chi-squared residual values indicate more returns than expected based on sales. Negative values indicate fewer returns than expected. The chi-squared residual statistic is a z-score and the values are normally distributed and centered around zero. The user of the chi-square statistic is advised that the value of the statistic often varies based on the magnitude of the observed or expected values with large values of O or E yielding large chi-squared residual values. Conclusions based on small ranges of chi-squared residual values (i.e., $-6<csresidual<6$) should be made with caution.

(9) Map and symbolize chi-squared residual values. Visually review map for "interesting" locations (see *Figure* 6).

(10) Perform local indicators of spatial association (LISA)[4] analysis of chi-squared residuals to determine the most significant clusters of high or low residuals. High residuals are coupled with locations having conditions to which the product is not robust. Low residual clusters are coupled with locations having *favorable* operating conditions (see *Figure* 7). If there are few or no significant clusters and there are many warranty returns, it is possible that the warranty claims are not related to any specific external condition or product duty cycle, but are potentially related to a manufacturing issue.

(11) Conduct physical analysis of field returns from both high and low locations.

(12) Assess environmental conditions and mechanical or electrical cycling in the locations of high residuals compared to areas of low residual clusters.

(13) Select a *best set* of environmental conditions and cycling profiles needed to perform a physical test.

It can be observed from the example maps that there are potentially three main clusters of high residual values indicating potential locations where the product has higher than expected returns. Upon analysis using Anselin's LISA method,[5] as implemented in SpaceStat™, the primary cluster of high residuals is located in the upper left of the map. The most significant cluster of high residuals would be the area with the environmental conditions or customer duty cycling that the product was least capable of tolerating. In the scenario illustrated in this example, a test incorporating the conditions and duty cycling of the "worst" area is likely to reproduce the field failure mode in a laboratory setting.

Example 2: Population and Time of Vehicle Usage Weighted Model of Temperature and Relative Humidity for Automotive Applications

This example illustrates how weighting of external operating conditions can be used to understand the normal operating envelope of temperature and relative humidity and the points of greatest usage for an automobile, given that both human population and time of human activity are unequally distributed within any given area and time-of-day.

Datasets used in the model include Typical Meteorological Year (TMY2) data (NREL), TMY2 weather station locations (NREL), human population counts by location (U.S. Census), and proportion of vehicular trips by hour of the day (*Figures* 8–11).[7-9]

The model was constructed first by loading the population counts by county (*Figure* 8) into the GIS system and then creating an additional layer of points representing the TMY2 weather stations (*Figure* 9). Next, the population points were assigned to the nearest TMY2 weather station using Thiessen polygons computed by the GIS (*Figure* 10). A Thiessen polygon represents an area around a point where all of the points within the area are closest to the surrounded point of interest than any other point of interest. Next, the human population counts were summed for each Thiessen polygon and the percent of total human population was computed. The result is a list of human population weighted weather stations.

The remaining steps of the weighting process were performed in a database using a sequence of grouping queries and calculations. The final results from the query sequence provide data that can be plotted to show the number of operation hours that will occur per year at each combination of temperature and relative humidity for a vehicle operated a total of 500 hours per year had the vehicle actually lived in every TMY2 weather station location (*Figure* 12). The plot is useful for visually assessing the combinations of relative humidity and temperature that have the greatest operating

Figure 8—Percent of total U.S. population by county, 1999 estimates.[6]

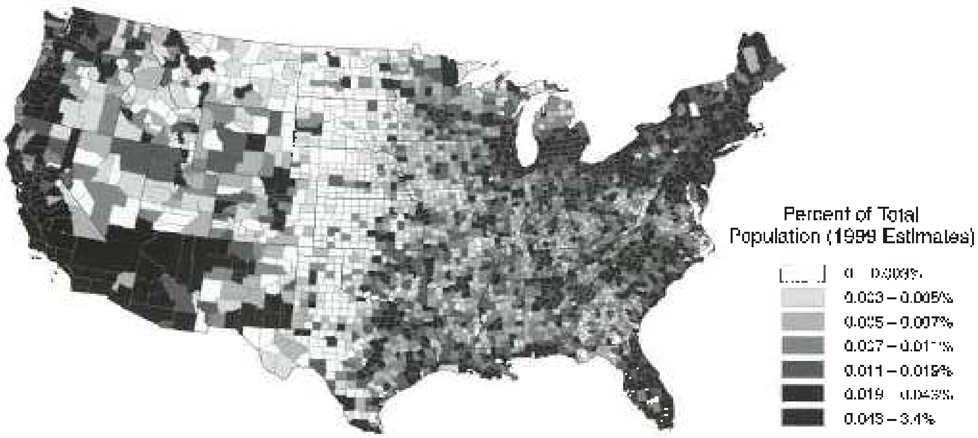


Figure 9—TMY weather station locations, continental USA.

Figure 10—Percent of total population as assigned to TMY weather stations.

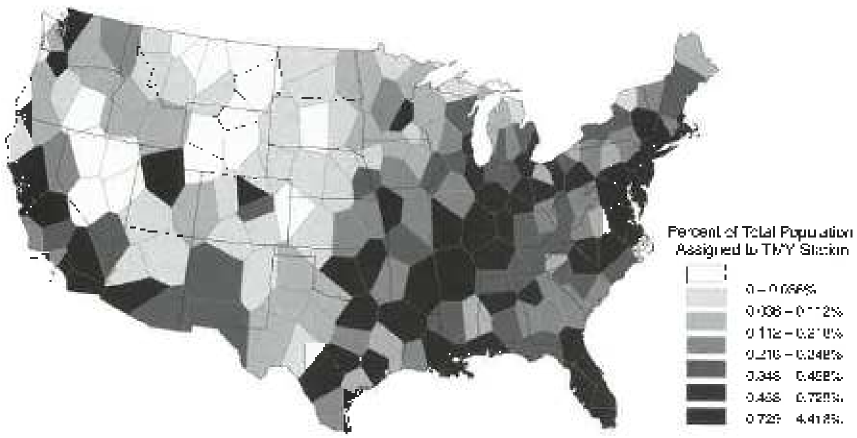
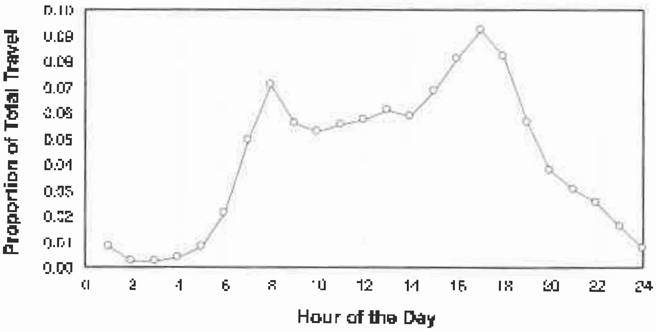


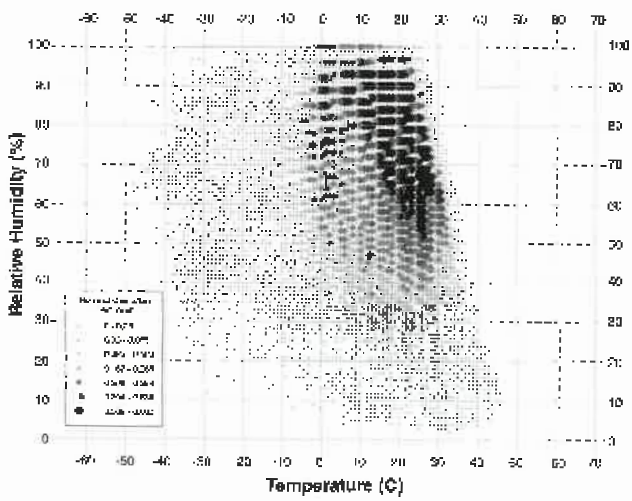
Figure 11—Proportion of total vehicular travel occurring by time of day.



time. If the product does not function properly in the peak areas of the condition map, it is possible that customer satisfaction will be affected, subsequently generating warranty claims that will be costly to the producer.

Areas of high incidence of the most common conditions can be identified for further analysis and, if additional variables are required, the model can be extended to create what is often known as a suitability

Figure 12—Population and time of vehicle usage weighted ambient temperature and relative humidity.



map. The suitability map requires that additional variables be rescaled or ranked and then weighted appropriately before including them in the model. Suitability modeling is beyond the scope of this article.

CONCLUSION

Geographical analysis methods are very powerful in their ability to help identify potential locations of good or bad product performance that otherwise might be invisible to the researcher employing nonspatial analysis of data. A GIS is a convenient and powerful tool that can be used to implement geographical methods and enable visual display and dynamic interaction of analysis results. Many different GIS packages are available to the researcher. Many spatial analysis tools are available in advanced statistics software programs and GIS packages.

By analyzing past product warranty history, it is possible to more fully understand the conditions and customer usage parameters that could lead to useful laboratory tests that will enable prediction of a product's service life. CT

References

(1) Ahl, V. and Allen, T.F.H., *Hierarchy Theory*, Columbia University Press, New York, 1996.
(2) Environmental Systems Research Institute (ESRI), *ArcView 3.2*, 1999. http://www.esri.com.
(3) Dykes, J. and Unwin, D., *Maps of the Census: A Rough Guide*, Department of Geography, University of Leicester, United Kingdom, 1998. http://www.geog.le.ac.uk/jad7/AGOCG/.
(4) Anselin, L., *Local Indicators of Spatial Association: LISA*, Geographical Analysis, Ohio State University Press, Columbus, OH, 27:93-115, 1995.
(5) Anselin, L., *SpaceStat 1.90*, Ann Arbor, MI (1998).
(6) Environmental Systems Research Institute (ESRI), *USA Population by County*, 1999. http://www.esri.com.
(7) National Renewable Energy Laboratory, *TMY2 User's Manual*, 1995. http://rredc.nrel.gov/solar/old_data/nsrdb/tmy2/.
(8) Federal Highway Administration, *Nationwide Personal Transportation Survey*, 1995. http://www-cta.ornl.gov/npts/1995/Doc/index.shtml.
(9) U.S. Dept. of Energy, *Typical Meteorological Year Data (TMY2)*, 1995. http://www.eren.doe.gov/buildings/energy_tools/energyplus/energyplus_weatherdata.html.