



About the Author



Christopher S. Carlson
FCAS, MAAA

Chris Carlson is a Consultant in Pinnacle's Columbus, Ohio office. Currently a member of the Casualty Actuarial Society Board of Directors, Chris is the incoming President-Elect of the CAS. He is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries and a graduate of the University of Oregon. He has served as a CAS Vice President and chair of the Ratemaking Research and Seminar committees. Chris is also chair of the Casualty Committee of the Actuarial Standards Board.

Since joining Pinnacle last fall, Chris has been actively involved in territorial rating and other GLM projects.

Prior to joining Pinnacle in 2005, Chris was a senior actuarial officer with Nationwide Insurance Companies. During 25 plus years there, his assignments included personal lines ratemaking projects in homeowners and personal automobile and reserving projects in the personal, commercial and assumed reinsurance lines. He was involved in defending the use of territorial rating and territorial assignment in Ohio.

Smoothing and Clustering: Advanced Techniques to Redefine Rating Territories

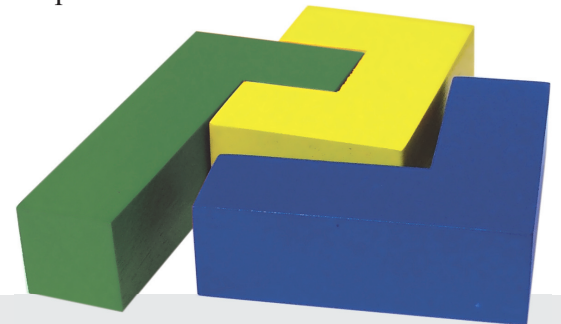
By Christopher S. Carlson, FCAS, MAAA

Insurance companies depend on territorial boundaries to identify areas with significant differences in future loss potential within a state, county or city. However, many of today's territorial definitions, established many years ago and only minimally revised, are outdated.

Thankfully, the recent emergence of predictive modeling tools allows insurers to very easily adjust their experience-based data for the impact of all non-territorial rating factors. By applying smoothing and clustering techniques to post predictive model-adjusted data, insurers can now define new and better territories to gain rating accuracy and competitive advantage.

Territorial enhancements are available to all insurers, regardless of size, through smoothing and clustering techniques. Smoothing is a process that produces a more accurate estimate of future loss potential in a given location by incorporating information from the surrounding area. For example, locations next to a specific suburban location in Chicago should have much

more similar experience, and thus offer more predictive information about loss potential, than locations in Central Illinois or even the area surrounding St. Louis, due to differing characteristics such as traffic congestion or weather patterns.



Territorial enhancements are available to all insurers, regardless of size, through smoothing and clustering techniques.

Clustering, a process that brings together, by location, risks with similar experience, can be performed contiguously or non-contiguously. Non-contiguous clustering, for example, could place the suburban areas of Buffalo, Rochester and Syracuse into one territory. A contiguous approach would develop three distinct territories for each suburban area.

Determining the Level of Detailed Data

To capture the full potential of smoothing and clustering, an insurer must identify the most detailed data available and construct an optimal set of territorial definitions from the “ground up.” This is a departure from the more traditional “top down” approach, which begins with state-level data and subdivides the state by county, city and so forth.

Typically the most detailed data is available at the ZIP code level. Although imperfect, this approach is currently the best available territorial boundary definer for most insurers. With the emergence of geo-coding, insurers will eventually have the opportunity to use smaller areas such as census blocks or other location identifiers as the basic territory building blocks. (Since the ZIP code approach remains the most common, ZIP code-based examples will be used in this monograph.)

The most effective basis for developing territorial definitions is pure premium, also known as loss costs. Before an insurer subjects detailed pure premium data to a smoothing procedure, it is best to eliminate, as much as possible, the impact of other rating variables from the historical ZIP code experience. The use of evolving generalized linear

modeling (GLM) processes can adjust the actual experience for the impact of other rating variables and overcome potential class bias. For example, bias can exist where mostly high replacement cost valued homes are in one area and low replacement cost valued homes in another. It is also helpful to exclude specific claim types, such as hurricane losses, to prevent distortions to the pure premium caused by a single event in a specific territory. Additionally, capping individual losses would prevent the distortion caused by a single large loss.

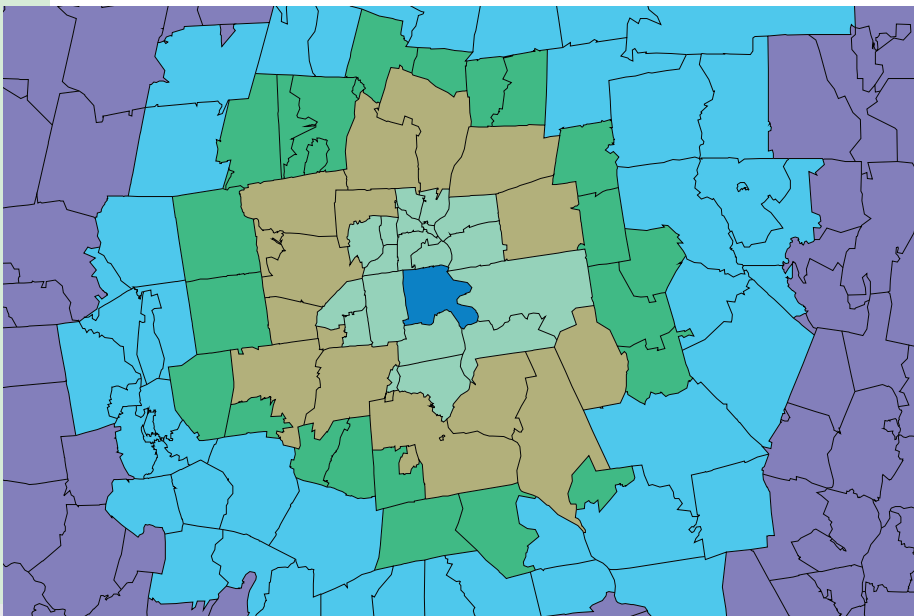
The actual experience of the insurer and any available industry-wide experience at a detailed-enough level can be used either separately or in combination. Even if the only available data is limited, smoothing is still possible. Doing so, however, requires the inclusion of data from a greater-than-optimal geographic distance.

Smoothing Techniques

Territorial building blocks, such as ZIP codes, generally do not contain sufficient policy experience to be fully credible or predictive of future loss potential on their own. Rather than using the current common process of blending local data with all of the statewide experience, smoothing includes only the experience of neighboring ZIP codes. This approach builds sufficient exposure or loss volume to achieve the desired level of credibility.

Smoothing techniques take a bull’s eye approach by bringing in data from all ZIP codes within a set radius from the center of a ZIP code. The smoothing calculations move the radius further and further away from the center until the desired credibility level is found.

A distance constraint can also be imposed to limit how far the model travels from the center of the geographic area. This limitation is valuable in states with large urban and rural areas to avoid the inclusion of experience from unlike areas. There are no set standards on the number of exposures or claims needed for smoothing or the maximum distance an insurer should allow the smoothing to go.



Smoothing techniques take a bull’s eye approach by bringing in data from all ZIP codes within a set radius from the center of a ZIP code.

The need to apply smoothing to the underlying data is heavily dependant on the volume of data in any ZIP code. Generally, the more available data per ZIP code and across all ZIP codes, the less smoothing is necessary. If in doubt, the insurer should smooth the data and review the impact of the smoothing process.

In most cases, unless working with total industry data in a confined area, such as a specific metropolitan area, smoothing will produce vastly improved accuracy compared to traditional techniques. An exception to this rule is modeled data related to a catastrophic loss which is essentially pre-smoothed by its very nature. Hurricane and severe thunderstorm models are designed, through the tremendous number of simulated years, to develop estimates that change in a continuous fashion from location to location. Therefore, catastrophic-related data is generally ready-made for clustering.

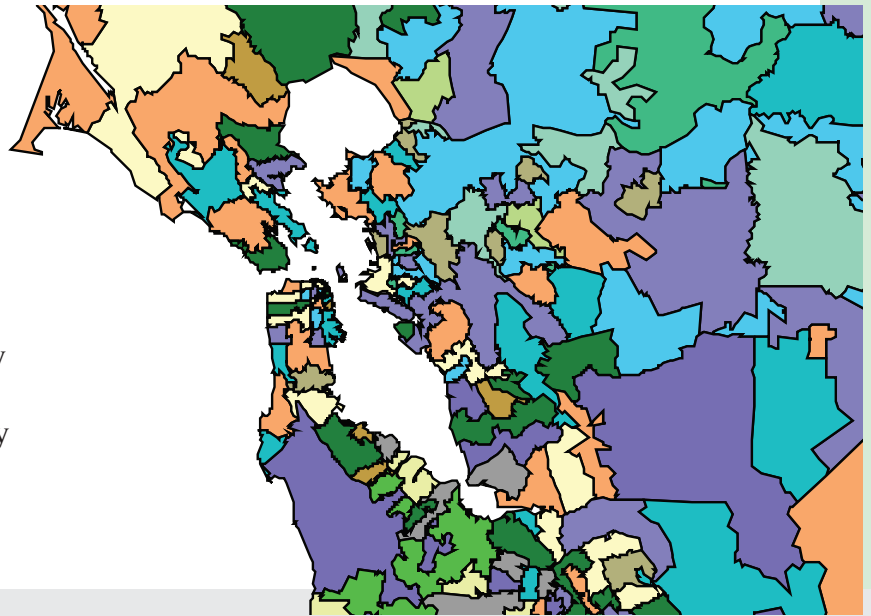
Incremental Clustering

Next, insurers can “build” territorial definitions from the detailed data by using clustering.

As previously noted, there are two approaches to clustering -- contiguous and non-contiguous. With the non-contiguous approach, ZIP codes from anywhere within the study area can be grouped together within a cluster. In the contiguous fashion, ZIP codes added together within a cluster have an additional constraint -- they have to be adjacent to one another.

Consider an example in a state that has 500 distinct ZIP codes. The two ZIP codes with the closest experience based on the data are added to each other to form a new cluster, which creates a total of one cluster with 498 single ZIP codes remaining. The data points with the closest experience can be determined based on percentage or additive differences.

The clustering process continues, adding successively determined locations with similar experience one at a time until a predetermined number of clusters remain. Or, the process can be



...insurers can build territorial definitions from the detailed data by using the clustering approach.

left unbounded until only a single cluster exists, which comprises the complete study area.

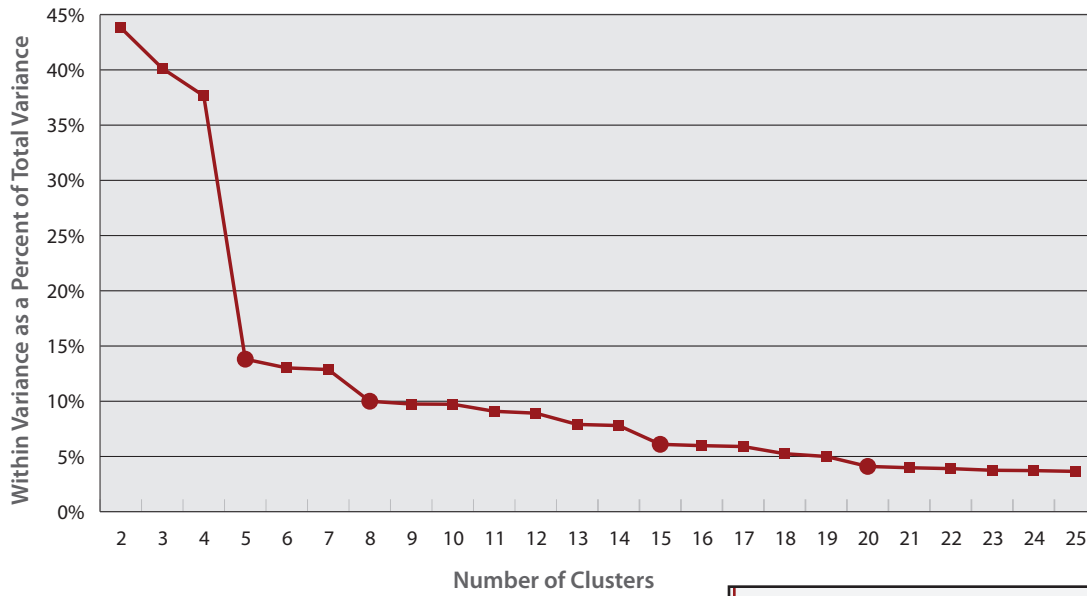
Determining the Optimum Number of Clusters

After identifying numerous potential sets of territorial definitions, the insurer identifies the most appropriate and optimum set by reviewing summarized variance statistics – often in graphical form (see graph on page 4). Territories are performing well when the loss experience within a given territory is sufficiently similar and the loss experience between two territories is sufficiently dissimilar. Therefore, the variance in experience within a given territory should be small and the variance in experience between territories should be large.

This review process is the reverse approach to clustering – top down versus ground up -- in that it can begin with a single territory solution and move toward more and more territories until reaching a balance between the number of territories and the relative difference between the experiences of the various territories. At some point, there are generally diminishing returns for adding additional territories where little additional rating accuracy is gained from the increased level of detail.

Cluster Variance Comparisons

(Total Variance = Within Variance + Between Variance)



Evaluating Results

The final test of a revised territory definition structure involves evaluating its impact on the existing portfolio. Since most of the current territorial definitions were established many years ago, definition changes and resulting rates from the clustering technique may create significant rate changes at an individual customer level. Measuring such impact is an essential part of the decision process. While the indicated definitions and rates are the most appropriate based on the data, some tempering and rate increase phase-in may be required to mitigate rating impacts.

The indicated definitions and rates should also be evaluated for marketing and competitive considerations. A useful step in the final review process entails overlaying the new definitions on top of the current definitions. Additional valuable information can be drawn from maps and/or rate examples that compare definitions and rates of major competitors with new definitions and indicated rates.

These comparisons may identify additional territorial segments (combinations of all included insurer definitions) and allow treatment of the widest rate level and competitive

The Power of Clustering

The goal of final cluster selection is to balance the number of territories with loss potential similarity. As the number of clusters increase, the within variance decreases. The clusters with the most potential are 5, 8, 15 and 20 because they are closest to the lower left-hand corner. Cluster sets just prior to these (4, 7, 14 and 19) have dramatically more within variance while the clusters immediately following (6, 9, 16 and 21) do not have significantly lower within variance to warrant the addition of one more cluster.

positions. This comparison process may also lead to a “phase-in” approach to implementing the indicated territorial definitions.

Conclusion

While often overlooked as part of an insurance company’s segmentation efforts, territories remain a key part of the insurance pricing foundation. Insurers now have the opportunity to more accurately analyze and construct their rating territories based on their own experience. Those applying the latest actuarial and technological advances will make the greatest gains from the traditional territorial rating process.

If you would like more information about this subject, please contact Chris Carlson at Pinnacle Actuarial Resources, Inc. by phone at (614) 734-8366 or by e-mail at ccarlson@pinnacleactuaries.com.