

Geospatial Analysis: Linear, Nonlinear, or Both?

Kelsey Osterloo and Weixin (Angela) Wu, State Farm Insurance

ABSTRACT

An important component of insurance pricing is the insured location and the associated riskiness of that location. Recently, we have experienced a large increase in the availability of external risk classification variables and associated risk factors by geospatial location. As additional geospatial data becomes available, it is prudent for insurers to take advantage of the new information to better match price to risk. Generalized additive models using penalized likelihood (GAMPL) have been explored as a way to incorporate new location-based information. This type of model can leverage the new geospatial information and incorporate it with traditional (non-location based) insurance rating variables in a regression-based model for rating. Our discussion demonstrates the use of the LOESS procedure as well as the GAMPL procedure in a combined solution. Both procedures are available in SAS/STAT® software. We discuss in detail how we built a local regression model and used the predictions from this model as an offset into a generalized additive model. We compare the results of the combined approach to results of each model individually.

INTRODUCTION

The riskiness of an insured household location is included in most insurance companies' rating plans. For many years, this risk was classified based on large geographic areas such as counties or zip codes. This would mean that for the location risk portion of the insured's premium, all that are insured within the same zip code would be given the same risk classification. While geographic rating with counties or zip codes has improved upon the traditional insurance rating with non-location based policyholder characteristics, it has its own disadvantage. For two policyholders that have similar characteristics but live across the street from each other in different counties/zip codes, we sometimes observe noticeable premium differences. The issue occurs because counties or zip codes are less granular than desired and neighboring counties or zip codes could have significantly different geographical risks on average. As technology and data have continued to improve, insurance companies are now able to identify an insured's location at a much more granular level. An example of this smaller level of granularity uses square grid cells based on latitude and longitude that are approximately one square kilometer in size. This level of granularity allows the insurance company to more finely segment their policyholders and therefore charge the appropriate rate. In addition to influencing the level of granularity, technology enhancements have allowed for more location-based data to become available. Our paper looks at an area within a state and discusses the approaches we used to model the area's risk using only latitude and longitude coordinates. These models are then compared and contrasted to models that incorporate new location-based data all with the goal of more accurately predicting an insured's risk.

METHODOLOGY

LOCAL REGRESSION

Local regression is a nonparametric approach to estimating regression surfaces originally proposed by William S. Cleveland. In this method, an n^{th} degree polynomial is fit at every observation in the dataset using only observations that are closest to the data point being estimated. The model is fit using weighted least squares meaning that more weight is given to points that are closest to the point of interest and points that are further away are given less weight. This logic follows the idea that points closest to each other are more likely to share similar characteristics and therefore should be given more influence in the prediction whereas points further away may not share similar characteristics with the point being predicted and therefore should be given less weight. It's important to note that the definition of distance here is not limited to the physical/geographic distance between two points. Points with similar magnitude in external risk

classification variables are also considered “close” to the point of interest. To determine the points used for fitting the observation of interest, one chooses a selected span or smoothing parameter. This input value is the proportion of the entire dataset that will be used to fit the local regression. For example, if the user’s dataset contains 100 observations and a smoothing parameter of 0.5 is chosen, the 50 nearest observations to the data point will be used to fit the local regression model. Smaller smoothing parameter values will result in models that fit more closely to the observed data. Too small of a smoothing parameter will closely match the observed data, but may result in over-fitting. On the other hand, large smoothing parameters can produce models that are too smooth and under-predict the data (Cleveland 1979). We chose to fit local regression models using PROC LOESS available in the SAS/STAT® 14.2 software (SAS Institute Inc. 2017).

There are two ways to incorporate geographic rating in LOESS with traditional (non-location based) class plan rating variables.

The first way is a single step approach in that one directly includes the traditional class plan rating variables and the location-based variables as regressors in a single LOESS model. Although this approach is simple in modeling, it significantly reduces the explainability of traditional class plan rating variables. For example, it would be difficult to isolate the effect of home age in modeling losses using a LOESS single stage model. Additionally, as more regressors enter the LOESS model, this approach becomes very computationally expensive and impractical to use in a production environment.

A second way is a two-step approach that treats the LOESS model as a residual model. In the first step, we apply a generalized linear model (GLM) to the class plan variables. This provides explainability of the class plan variables. The original target variable is adjusted for the effects of the first stage GLM and then as a second step in the process, a location-based LOESS model is fit to the adjusted target variable from the first step.

For the rest of the paper, we will only be focusing on geographic rating with or without external location-based variables. In other words, we did not consider traditional (non-location based) class plan rating variables in our models.

GENERALIZED ADDITIVE MODELS

Generalized additive models are extensions of generalized linear models in that they relax the linearity assumptions to allow for smoothing functions to enter the model. The smoothing functions can capture the nonlinear patterns of the data that may be too difficult to capture through a traditional linear parametric term. Generalized additive models can be expressed as follows:

$$g(E(Y)) = \beta_0 + f_1(x_1) + f_2(x_2, x_3) + \dots + f_m(x_n)$$

where $g(\cdot)$ is the link function and $E(Y)$ is the expected value of the target variable. The $f_m(\cdot)$ terms can either be parametric (i.e. linear terms) or non-parametric (i.e. smoothing functions). Some examples of smoothing functions are local average, running average, running line, kernel-based, and spline. Because we chose to fit these models using PROC GAMPL available in the SAS/STAT® 14.2 software (SAS Institute Inc. 2017), our smoothing functions are splines. This procedure is a high-performance procedure that fits GAMs based on low-rank regression splines such that each spline term is constructed using the thin-plate spline regression technique (Rodriguez 2016).

Similarly, there are also two ways to incorporate location-based rating variables in GAMPL with traditional (non-location based) class plan rating variables. For GAMPL, there is no significant appeal to conduct a two-step model as is the case for LOESS because regression coefficients are produced for each regressor in GAMPL and clearly interpretable.

For the rest of the paper, we will only be focusing on geographic rating with or without external location-based variables. As stated before, we did not consider traditional (non-location based) class plan rating variables in our models.

METHODOLOGY APPLICATION – INTRODUCTION

DATA STRUCTURE

As mentioned above, square grid cells are created from latitude and longitude coordinates as a way to more finely segment policyholders. These grid cells are 0.01 degree latitude by 0.01 degree longitude in size which equates to approximately one square kilometer. For every policyholder, we know the latitude and longitude coordinates of their household location. We can then assign policies to the grid cell that match their latitude and longitude coordinate (truncated to the hundredths place). For example, if policyholder A has a household location of 38.9108 degrees latitude and -75.5276 degrees longitude, we would assign them to the grid cell 38.91 latitude and -75.52 longitude. Similarly, if policyholder B has a household location at 38.9114 degrees latitude and -75.5277 longitude, they too would be assigned to grid cell 38.91 latitude and -75.52 longitude. Policyholders A and B's losses, premiums, exposures, etc. would be added together to get a summarized total of each grid cell's information. We used and compared both the policy-level and summarized grid-level data in building the models. The area we used for modeling consisted of 267,439 policies which equated to 39,516 grid cells.

MODEL SPECIFICATIONS

The following variables were used in the building of the models:

- Pure Premium (PP): The target variable for each model, defined as the amount of crime loss per exposure. The pure premium is not adjusted for the traditional class plan rating variables.
- Latitude: The latitude coordinate used as a covariate (LOESS) or part of the spline term (GAMPL).
- Longitude: The longitude coordinate used as a covariate (LOESS) or part of the spline term (GAMPL).
- Exposure (EHY): The loss exposure which for this Homeowners example is earned house years. Values of one mean the policy was in force for the entire policy term while values less than one reflect the portion earned before the policy lapsed or was terminated. This variable was used as a weight for both the local regression and generalized additive models.
- Crime Risk: An estimate of the crime risk at a particular latitude and longitude coordinate. This variable represents the type of new location-based data available in the market today and is used as an additional covariate in the models.

METHODOLOGY APPLICATION – MODELING USING LATITUDE AND LONGITUDE

LOCAL REGRESSION MODELS (PROC LOESS)

In building the local regression models, there were quite a few inputs that we wanted to test and then compare the results of each model to see which input gave us the best result. These inputs included using both the policy-level and grid-level data, looking at both the direct and interpolated fitting parameter, and varying the smoothing parameter. All of the model variations were fit using code similar to the example below:

```
proc loess data=area2_policy;  
  model PP = latitude longitude / smooth=0.05 direct;  
  weight ehv;  
  ods output OutputStatistics=area2_policyd_outstats  
             FitSummary=area2_policyd_fitsummary;  
run;
```

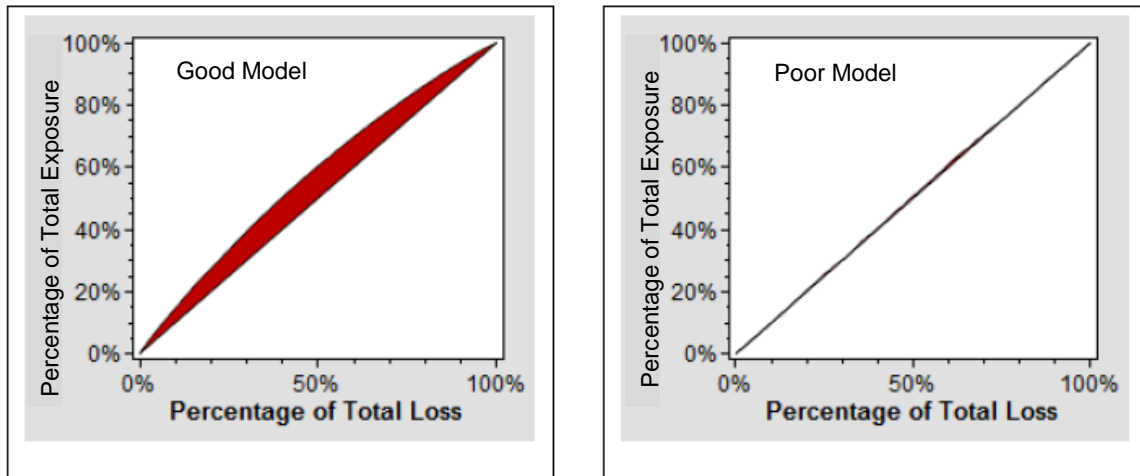
We chose to look at only three different smoothing parameters as run time became a factor in our model building. Ultimately, we ran through the following combinations of models using a local linear regression (degree = 1) for each run:

- Policy-Level Data / Direct Fitting / Smooth = 0.01, 0.05, 0.1
- Policy-Level Data / Interpolated Fitting / Smooth = 0.01, 0.05, 0.1
- GRID-Level Data / Direct Fitting / Smooth = 0.01, 0.05, 0.1

It is worth noting that we did not try interpolated fitting with the grid-level data. As you will see in the next section, the policy-level plots revealed some sharp lines instead of smooth transitions between predictions when using interpolated fitting. We chose to focus our attention on direct fitting only for the grid-level data after this was revealed.

LOESS MODEL COMPARISONS

We looked at both plots of the predictions as well as the Gini coefficient to compare model runs. The Gini coefficient is a way to measure how well the model is segmenting the data based on loss. To calculate, the data is first ranked based on model predictions. Then the cumulative percentage of total exposures (y axis) is plotted against the cumulative percentage of total loss (x axis). If the model is doing a good job such that higher percentage of losses correspond to larger scores, then the curve is concave above a 45 degree line (i.e. Good Model, below). However, for poor performing models, the curve tends to overlap the 45 degree line (i.e. Poor Model, above). Thus, a higher Gini coefficient is desired.

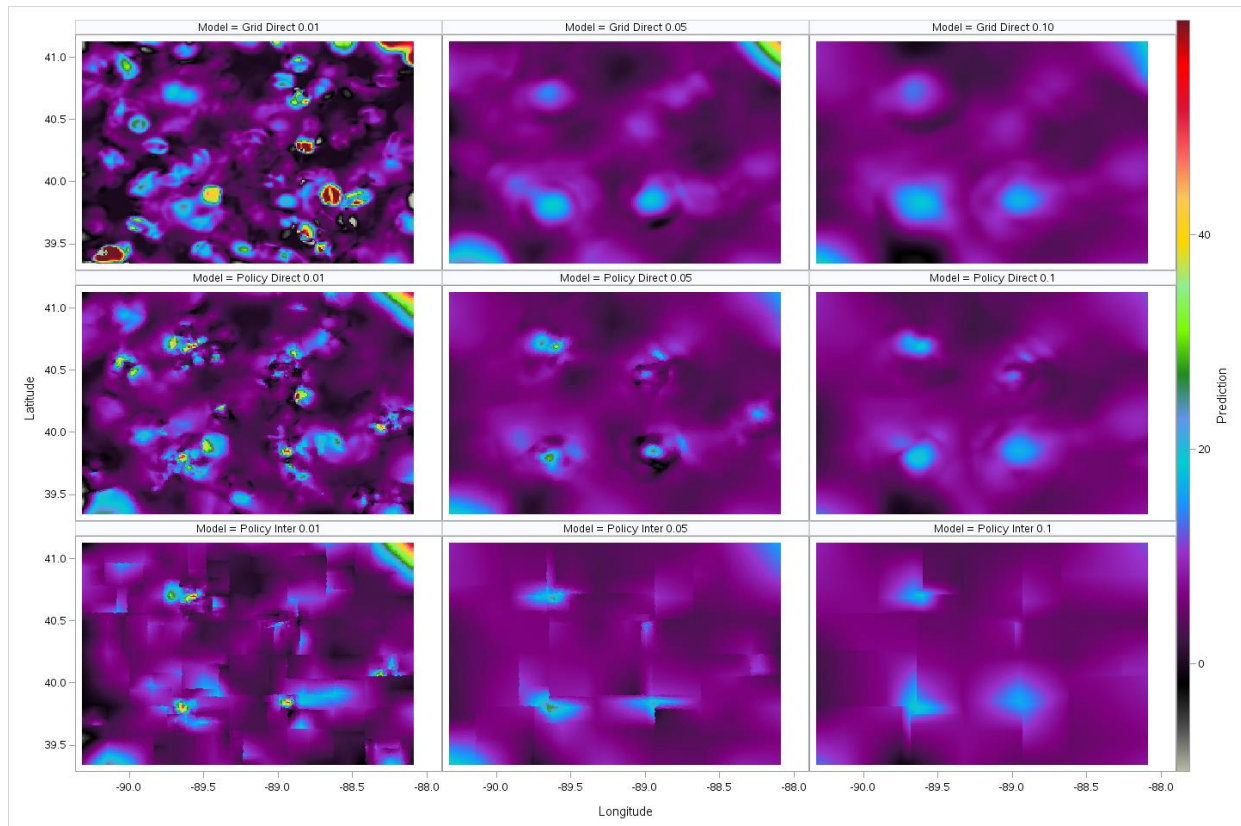


Output 1: Example of Gini Coefficient

We calculated the Gini coefficient on an out of time validation dataset and the results of each LOESS model run are as follows:

Grid Direct 0.01	Grid Direct 0.05	Grid Direct 0.1
0.267	0.245	0.235
Policy Direct 0.01	Policy Direct 0.05	Policy Direct 0.1
0.278	0.312	0.290
Policy Interpolated 0.01	Policy Interpolated 0.05	Policy Interpolated 0.1
0.284	0.291	0.284

Table 1: Gini Coefficients of LOESS Models



Output 2: LOESS Comparison Plots

First, we can see that the grid-level data performs the worst in terms of the Gini coefficient for each span selected. Using the grid-level data, a smoothing parameter = 0.05 means using 5% of the surrounding grid cells regardless of whether those grid cells are densely or sparsely populated. 5% of grid cells in urban areas include a lot more policies than 5% of grid cells in rural areas do. Based on credibility theory, the credibility standard should be selected based on the number of exposures. If we deem 13,372 policies (5% of policies) as the full credibility standard, each localized regression should be modeled with 13,372 policies. Using the grid-level data would result in different number of policies used in each localized regression, creating sub-optimal modeling results as shown in Table 1.

Second, we compare the direct fitting method versus interpolated fitting method. In initial stages of fitting models, we noticed that we obtained very strange results. Specifically, not all the observations were coming back with a predicted value. We discovered these results were due to the weight variable. Earned house years will have a value of zero if there are no policyholders in a particular grid cell. The SAS procedure will delete observations that have zero, negative, or missing weight. There are two ODS tables in PROC LOESS that provide predictions and they react differently to observations with a zero, negative, or missing weight variable when the DIRECT option is specified in the MODEL statement. The following table summarizes this idea:

	Direct	Interpolated
ScoreResults/Score Statement	Scoring not supported	Predictions for all observations
OutputStatistics	Predictions for only non-zero weight observations	Predictions for only non-zero weight observations

Table 2: Direct vs Interpolated Fitting

In order to output predictions on all observations when the DIRECT option is specified, we chose to add a small number to the EHY weight when the value was zero. Now, all observations would be used for analysis. Even after adjusting for these zero weight cases, the plots show the interpolated fit graphs to still have sharp lines between some predictions. Ideally we would like nice smooth progressions between observations as we saw with direct fitting. Even though it's more computationally expensive than interpolated fitting, we chose a direct fitting model as our champion model.

Lastly, we compared the smoothing parameter between models. Too large of a smoothing parameter would over-smooth the model and too small of a smoothing parameter would over-fit the model and does not generalize well. We want to strike a balance between over-smoothing and over-fitting. A smoothing parameter of 0.05 seems to perform the best for each method (grid direct/policy direct/policy interpolated) because they produce the highest Gini on the validation dataset.

In conclusion, we consider LOESS modeled with policy-level data, direct option specified, and smoothing parameter of 0.05 as our champion LOESS model.

GENERALIZED ADDITIVE MODELS (PROC GAMPL)

In building the generalized additive models with PROC GAMPL, there are different ways to control the smoothing. We found the tuning parameter maximum degrees of freedom to have the most significant impact to the level of segmentation on our model. Other tuning parameters we considered were setting a fixed smoothing parameter (SMOOTH=) and setting a maximum number of knots (MAXKNOTS=). We did fix the MAXKNOTS=100000, however we did not let that number vary as we did with maximum degrees of freedom. Our goal here is to accurately capture the risk of a policyholder so we are looking for models with a high degree of segmentation (i.e. higher Gini). As a practical matter, we must balance better segmentation with long run times and over-fitting the models. We varied the max degrees of freedom from 50 to 500, with each new model run adding 50 degrees of freedom. An example of the code we used to fit the models is below:

```
proc gampl data=area2_policy seed=29389;
  id GRID_ID;
  model PP_Exp = spline(latitude longitude / maxdf=50 maxknots=100000) /
    dist=gamma link=log;
  weight ehyp;
  performance nodes=all;
  output out=area2_grid_gampl pred=gampl_pred_area2_grid;
run;
```

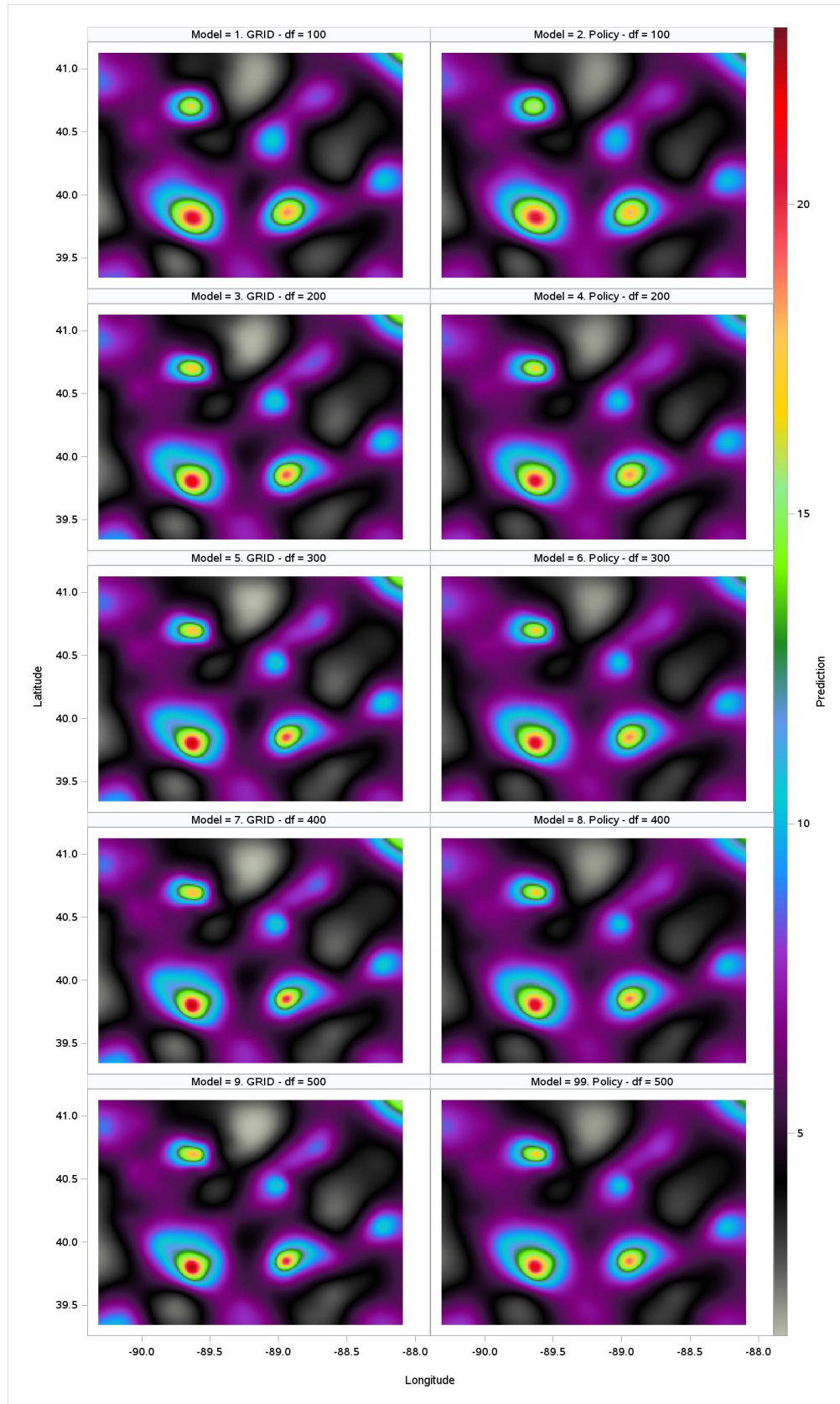
It is worth noting that our pure premium target can have many zero or missing values. When a policyholder (or any policyholders within the grid) does not have a claim, they have no loss and therefore the numerator of pure premium is zero. Additionally, if no policies are written within a certain grid cell, the denominator of pure premium is zero causing the target to be undefined (missing). When GAMPL fits the model, it will leave out the missing observations. It will also leave out the pure premium = 0 observations because we are using a log link function. Removing missing pure premium from model fitting is justified because those observations do not provide any value to the model. Removing observations with pure premium of zero, on the other hand, is problematic because they are valid observations. By leaving the zero cases out, we saw predictions that were higher than desired as there were no zero cases to 'draw down' the predicted smooth surface. We decided to fix this by adjusting the target. We added in an expense ratio (expenses/exposure) to all pure premiums greater than or equal to zero. In other words, we left the target as missing when there was no exposure, but for observations with exposure the target became pure premium plus expenses. When comparing the GAMPL results to LOESS, we removed the expense component for an apples to apples comparison.

GAMPL COMPARISON

Same as when making the LOESS model comparisons, we looked at both the Gini coefficients and plots to determine the best model out of all the GAMPL runs.

DEGREES OF FREEDOM	GRID	POLICY
50	0.243	0.223
100	0.253	0.227
150	0.258	0.232
200	0.265	0.235
250	0.268	0.238
300	0.271	0.239
350	0.274	0.241
400	0.274	0.241
450	0.275	0.242
500	0.276	0.243

Table 3: Gini Coefficients of GAMPL Models



Output 3: GAMPL Comparison Plots

In choosing the best generalized additive model, we first looked for differences between the policy-level and grid-level models. The plots look very similar between the two as we see the same color patterns in each. However, there were some differences in the predictions and the Gini coefficients. We believe this may have to do with the zero inflation problem in our target. The policy-level data is going to have many more observations where the target is zero where at a grid-level there is a high likelihood that the summarized target is going to be non-zero. Even though we added the expense ratio to all cases, our study shows the gamma distribution that we fit to both the policy-level and grid-level data fits better at a summarized grid-level. For these reasons, we chose the 500 degrees of freedom grid-level GAMPL model as the winner. It showed reasonable results in the plots and had the highest Gini coefficient across all the GAMPL models tested.

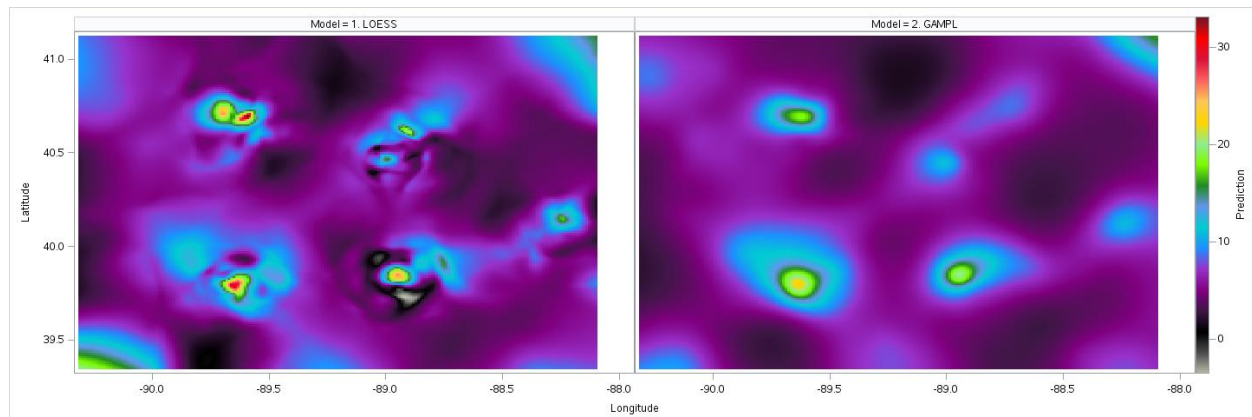
LOESS VS GAMPL COMPARISON

The last comparison we make is between our champion local regression model (policy-level data / direct fitting / smoothing parameter = 0.05) and our champion generalized additive model (500 degrees of freedom, grid-level model). First, we compare the Gini coefficients:

	LOESS	GAMPL
TRAINING DATA	0.368	0.344
VALIDATION DATA	0.312	0.276

Table 4: Gini Coefficients of Best LOESS & GAMPL Models

Second, we compare the plots of the two champion models on the same scale:



Output 4: Comparison of Chosen LOESS and GAMPL Models

Notice that the Gini coefficient calculated on the training dataset is higher than the Gini coefficient calculated on the validation for both LOESS and GAMPL. This is expected for three reasons. First, the validation dataset is an out of time sample. Loss distributions could shift slightly resulting in the models built with the training dataset to not be a perfect fit for the validation dataset. Even if the loss distribution doesn't change, the random nature of losses could result in models not performing as well in the validation data. Last but not least, our chosen models could be slightly over-fitting and thus do not generalize as well. We believe all three factors play into the difference in Gini between training and validation data.

Based on the Gini coefficient and graph comparison, LOESS seems to perform better in this case while the GAMPL model seems to over-smooth. However, it's worth noting that the chosen LOESS model (using policy-level data with DIRECT option) takes significantly longer run time compared to the high performance GAMPL models. So in practice, GAMPL is a good alternative for modeling geographic risks.

METHODOLOGY APPLICATION – MODELING USING LATITUDE AND LONGITUDE + LOCATION-BASED DATA

As mentioned above, as more and more location-based data becomes available, it will be important for insurance companies to be able to utilize its potential predictive power of policyholder's location risk. As an example of this type of data, we used a single variable (a measure of crime risk) and added it to the champion models shown above. In addition, we explore one other approach. In our third case, we will use a LOESS model (using only latitude and longitude) as an offset to a GAMPL model and compare those results to the individual approaches.

LOCAL REGRESSION MODELS (PROC LOESS)

We ran the champion LOESS model mentioned in the previous section only this time adding in an additional crime risk covariate. The crime risk covariate is measured at much larger scales than the scale of latitude and longitude. If we directly include the crime risk covariate without any standardization or rescaling, the variable would exert undue influence on the LOESS model. Based on preliminary tests, we use Crime_Risk/1000 as the additional regressor in the MODEL statement:

```
proc loess data=area2_policy;
  model PP = latitude longitude crime_risk_sca/ smooth=0.05 direct;
  weight ehy;
  ods output OutputStatistics=area2_policyd_scale_outstats
             FitSummary=area2_policyd_scale_fitsummary;
run;
```

It is worth mentioning that PROC LOESS does offer a SCALE= option in the MODEL statement. We did not include this in our testing of different standardization of the crime risk covariate because we wanted to keep the code (model) as similar to the previous champion LOESS model as possible. Using that option standardizes all variables in the MODEL statement which we did not want to do, especially with latitude and longitude.

GENERALIZED ADDITIVE MODELS (PROC GAMPL)

Similar to LOESS, we ran the champion GAMPL model with the additional crime regressor. The code changes slightly in that we now add in a PARAM() statement to specify that we want the crime variable to be a parametric term in the model:

```
proc gampl data=area2_grid seed=29389;
  id GRID_ID;
  model PP_Exp = spline(latitude longitude / maxdf=50 maxknots=100000)
             param(crime_risk_sca) / dist=gamma link=log;
  weight ehy;
  performance nodes=all;
  output out=area2_grid_scale_gampl pred=gampl_pred_area2_grid;
run;
```

USING A LOESS MODEL AS AN OFFSET INTO A GENERALIZED ADDITIVE MODEL

As a third alternative, we combined both the LOESS and GAMPL approaches. We will call this method the LOESS-offset GAMPL as the LOESS predictions are used as an offset in the GAMPL model. In choosing this approach, we note that LOESS has a tendency to over-fit the data while GAMPL has the potential to over-smooth it. LOESS-offset GAMPL may be a good compromise of the two approaches. The code below provides an example for this type of method. Results of this method are shown in the next section.

```

proc loess data=area2_policy;
  model PP = latitude longitude/ smooth=0.05 direct;
  weight eh;
  ods output OutputStatistics=area2_policyd_scale_outstats
             FitSummary=area2_policyd_scale_fitsummary;
run;

```

< Here we omitted code that summarizes area2_policyd_scale_outstats data to grid-level >

Then:

```

data offset;
  set area2_grid_scale_outstats;
  Offset_Log = log(Pred);
run;

proc gampl data=offset seed=29389;
  id GRID_ID;
  model PP_Exp = spline(latitude longitude / maxdf=500 maxknots=100000)
           param(crime_risk_sca) / offset=Offset_Log dist=gamma link=log;
  weight eh;
  performance nodes=all;
  output out=area2_gamploffset pred=gamploffset_pred;
run;

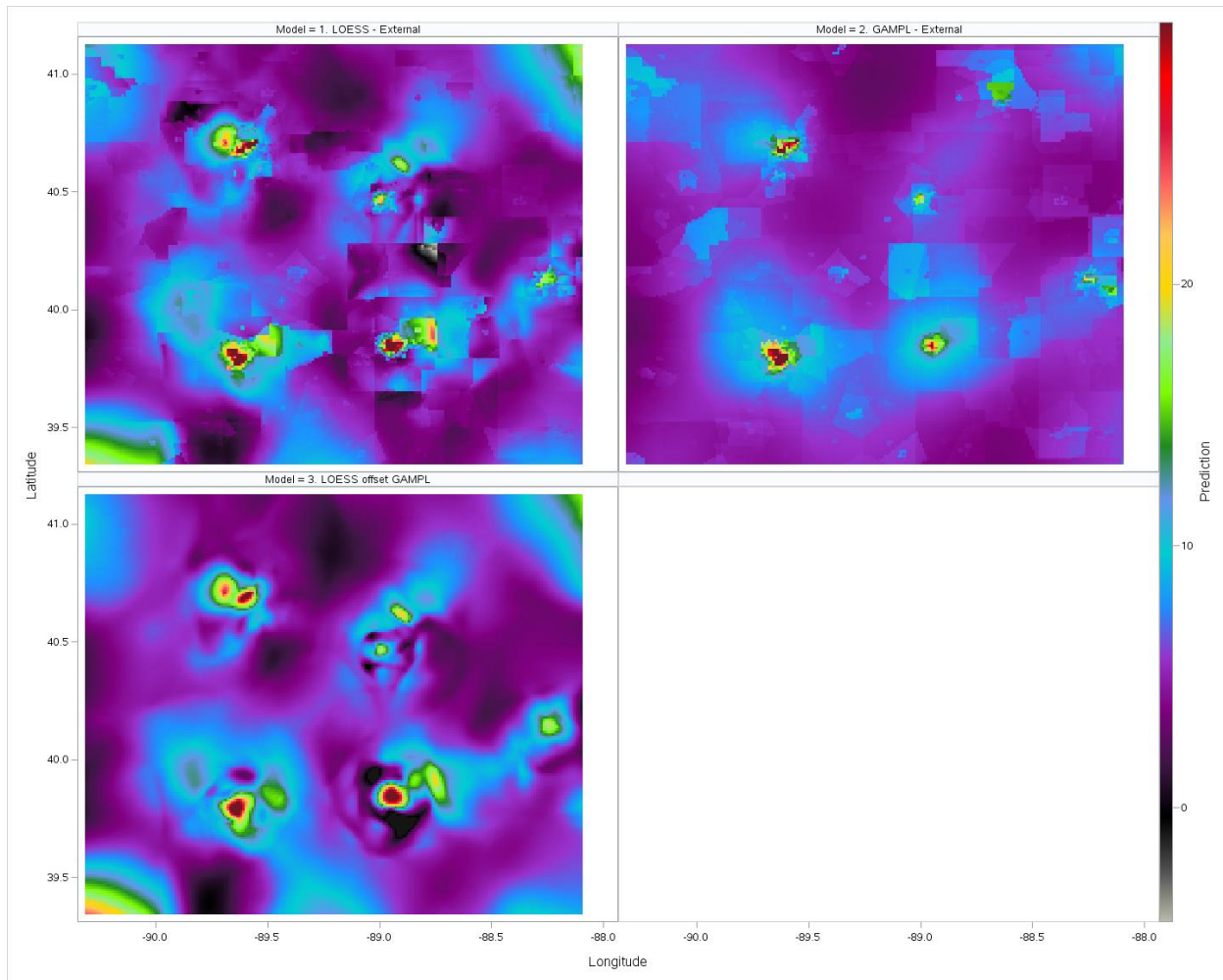
```

METHOD COMPARISON

We now look at one final comparison between LOESS, GAMPL, and the LOESS-offset GAMPL models using the added crime risk covariate in each model. Comparing the Gini coefficients first:

	LOESS	GAMPL	LOESS-OFFSET GAMPL
TRAINING DATA	0.375	0.333	0.368
VALIDATION DATA	0.319	0.320	0.309

Table 5: Comparison of Gini Coefficients for LOESS, GAMPL, and LOESS-offset GAMPL Models



Output 5: Chosen LOESS, GAMPL, and LOESS-offset GAMPL Models

We compare Gini coefficients of the three models including the crime variable and see mixed results in terms of improvement. The LOESS model shows an increase in Gini in both the training and validation data indicating that the external data did improve the model. However, the model still may be over-fitting as there is a significant difference in Gini between the training and validation data informing us that the LOESS model may not generalize well. The GAMPL model shows a decrease in the training Gini when the external crime data is added, but shows the opposite effect in the validation data. In fact, there is a very large improvement in the GAMPL model Gini coefficient for the validation data (0.276 to 0.320). This is a sign that the GAMPL model is not over-fitting the data and that the external data is a good addition. The LOESS-offset GAMPL model produces the lowest Gini in the validation data out of the three models being compared. The difference in Gini coefficients between the training and validation for that model is also significant indicating that this model also does not generalize well.

In comparing the plots, we see very similar patterns between the three. We can make out three high risk areas in all three plots telling us that all the models are capturing the same high risk. Where the plots start to differ is in their levels of segmentation. We see in the LOESS and especially in the LOESS-offset GAMPL models that there is a wider range of predictions (more black on the low end and deeper red on the high end). This is a sign of more segmentation. We also notice more 'boxy' patterns in these plots because the external crime data is less granular.

Overall, we believe there are pros and cons to each approach. The LOESS model gives us a great deal of segmentation, but is prone to over-fitting. The generalized additive model (GAMPL) does not provide as

much segmentation, but generalizes very well and seems to be utilizing the external data the best. The LOESS-offset GAMPL model also provides a great deal of segmentation shown in the plots. However, the LOESS-offset GAMPL seems to provide the greatest over-fitting of the three contenders. We could try to adjust more parameters within these models and also test additional location-based data as it becomes available to improve these models.

CONCLUSION

Insurance data has unique challenges and utilizing more location-based data is one of those challenges. There are several approaches available to model geospatial data as part of an overall class rating plan. We have demonstrated that a local linear approach like LOESS allows the modeler to tune how closely the model fits the data based on the smoothing parameter. A non-linear approach like a generalized additive model (GAMPL) offers more flexibility and allows both linear and spline terms to be modeled together. As a third alternative, we showed how to combine both a local linear and non-linear approach in a single model. All three approaches have advantages and disadvantages. For this particular analysis and set of data, we would choose the generalized additive model (GAMPL) as our champion. The GAMPL process offers the greatest modeling flexibility, is very simple to add additional location-based variables to, and utilizes high-performance technology which can improve run times. This allows us to meet our goal of accurately predicting an insured's location risk at a granular level.

REFERENCES

- Cleveland, W. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots". *Journal of the American Statistical Association*. 74(368): 829-836.
- Rodriguez, R. (2016). "Statistical Model Building for Large, Complex Data: Five New Directions in SAS/STAT® Software." In Proceedings of the SAS Global Forum 2016 Conference. Cary, NC: SAS Institute Inc. <http://support.sas.com/resources/papers/proceedings16/SAS4900-2016.pdf>
- SAS Institute Inc. (2017). "The GAMPL Procedure". In SAS/STAT® 14.2 User's Guide. Cary, NC: SAS Institute Inc. http://documentation.sas.com/?docsetId=statug&docsetVersion=14.2&docsetTarget=statug_hpgam_toc.htm&locale=en
- SAS Institute Inc. (2017). "The LOESS Procedure". In SAS/STAT® 14.2 User's Guide. Cary, NC: SAS Institute Inc. http://documentation.sas.com/?docsetId=statug&docsetVersion=14.2&docsetTarget=statug_loess_toc.htm&locale=en

ACKNOWLEDGMENTS

Special thank you to Bob Rodriguez and Weijie Cai of the SAS Institute.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Kelsey Osterloo
Property and Casualty Actuarial Department
State Farm Insurance Companies
One State Farm Plaza
Bloomington, Illinois 61710-001
kelsey.osterloo.u15h@statefarm.com

Weixin (Angela) Wu
Property and Casualty Actuarial Department
State Farm Insurance Companies
One State Farm Plaza
Bloomington, Illinois 61710-001
angela.wu.uz28@statefarm.com

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

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