

Contents

Predictive Modeling and By-Peril Analysis for Homeowners Insurance



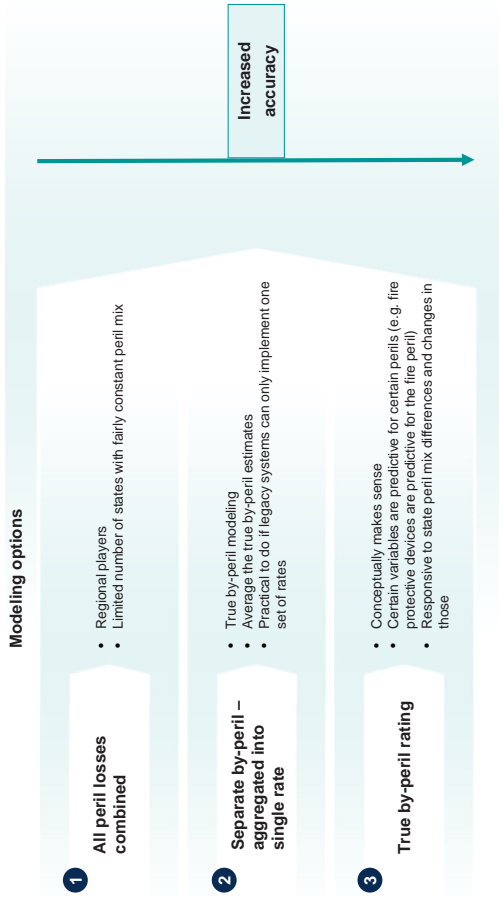
Case for by-peril modeling

By-peril model building

- Data
- Peril grouping
- Variables
- Interactions
- By-peril territories
- Model validation

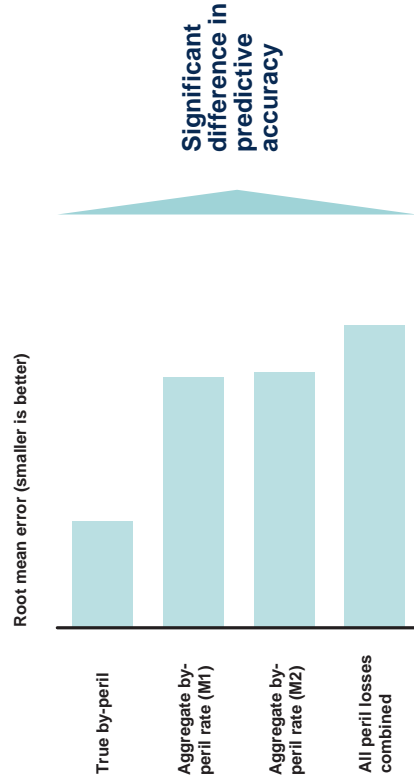
Conclusion

Case for by-peril modeling



3

Case for by-peril modeling



4

By-peril model building: data staging

Data options

Years used

- Balance volume with recency to reflect an appropriate mix of business

Data used

- Internal data used for non-catastrophes
- Simulated data for catastrophes

Data split

- Modeling/testing/validation
- Modeling/validation
- Out of time data split
- Out of sample data split

5

By-peril model building: peril grouping

Grouping or peril separation

Theft

- Disappearance/theft on premises
- Disappearance/theft off premises

Liability

- Liability/Medical payments

Fire

- Human-made fire
- Environmental fire

Water

- Weather water
- Non-weather water

Other

- Glass
- Aircraft
- Vehicles

Grouping or breaking of perils -- considerations

Availability of detailed/accurate peril codes from claims

- Is the correct cause of loss captured? Wind or hail damage to the roof?

- Are there additional peril code break-outs available?

Grouping or breaking

- Use judgment, intuition, similarity
- Electrical fire
- Grease fire from kitchen
- Fire from candles
- Fire from cigarette smoking
- Children playing with matches
- Fireplace fire
- Fire caused by electrical appliance

6

By-peril model building: data staging

Practical considerations

- Indications for appropriate rate level
- Systems cost trade off – cost benefit analysis of by-peril implementation
- Time constraints / speed to market / will dictate some of the options
- Impact on other departments: claims, systems, pricing/actuarial, financial reporting

By-peril model building: variable selection

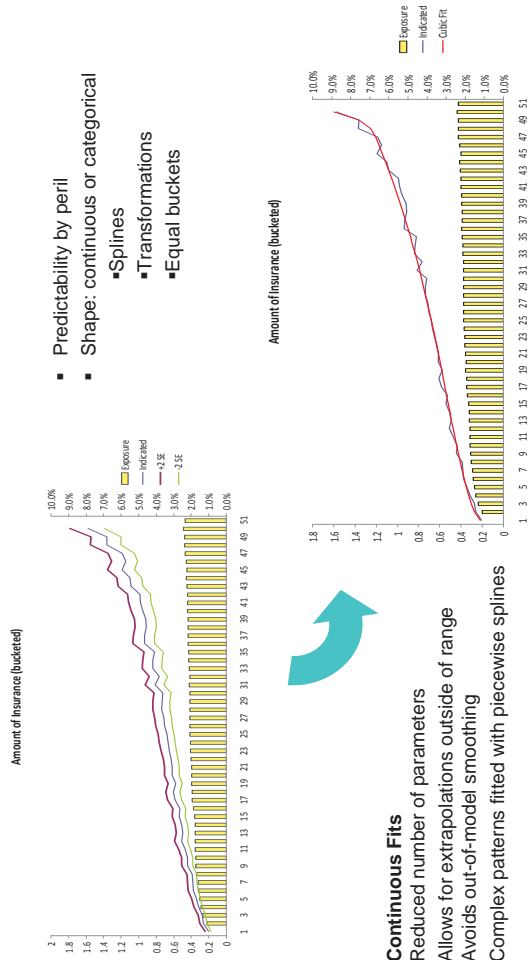
House characteristics	Occupant characteristics	Location and / or External variables
Amount of insurance	Age	Weather
Number of stories	Gender	Temperature
Number of rooms	Marital status	Precipitation
Square footage	Insurance score	Elevation
Age of electrical	Occupation/retired	Slope
Age of home	Number of occupants	Geography
Age of plumbing	Prior claim activity	Commercial business
Age of roof	Other personal lines	Protection class
Roof material	Full payment/installments	Demographics
Construction type	Billing lapse	Population Density
Protective devices	Good payer	

Which variables are predictive for which peril?

Financial variables

By-peril model building: variable selection

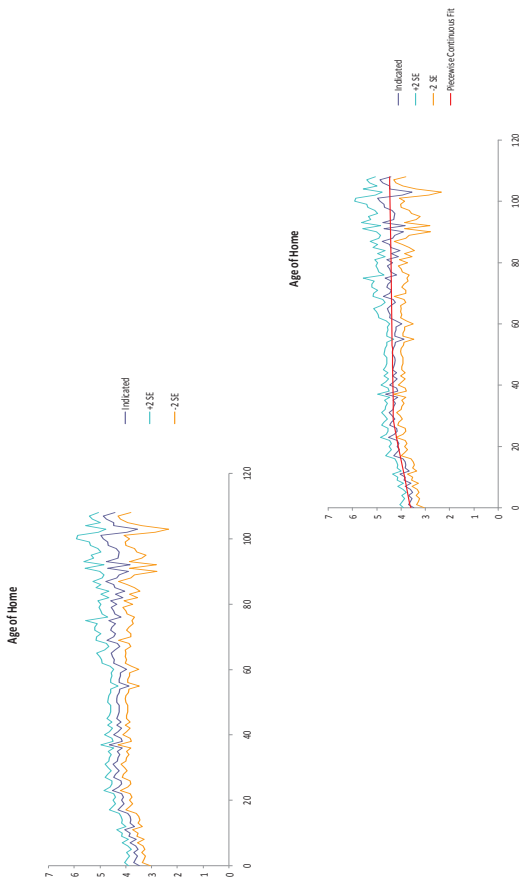
Univariate analysis



9

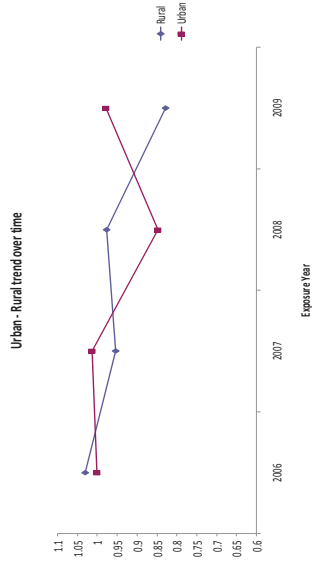
By-peril model building: variable selection

Univariate analysis



10

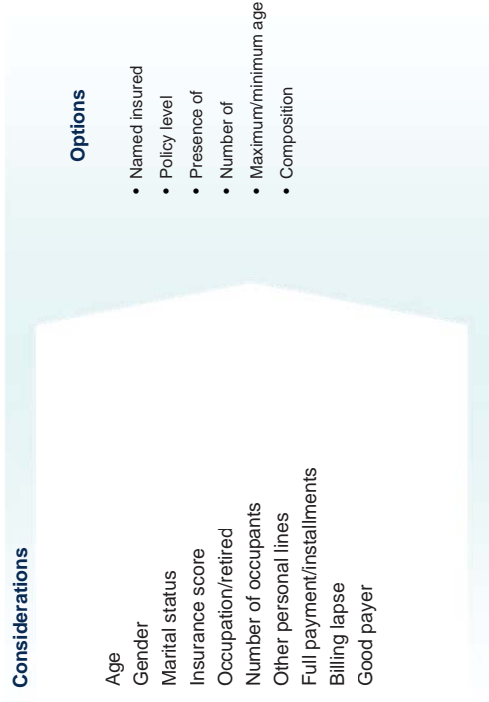
Model building: consistency over time



- Looking for a stable trend over time
- Data quality
- Correlation with other variables

11

Variable selection: occupant characteristics



12

Variable selection: occupant characteristics

Considerations

Vendor data

- House characteristics
- Insurance scoring
- Prior claim activity
- Weather
- Demographics
- Elevation

External data links

- Cost
- How often is the data updated by the vendor?
- How often does the data have to be updated?
- Regulatory support and environment

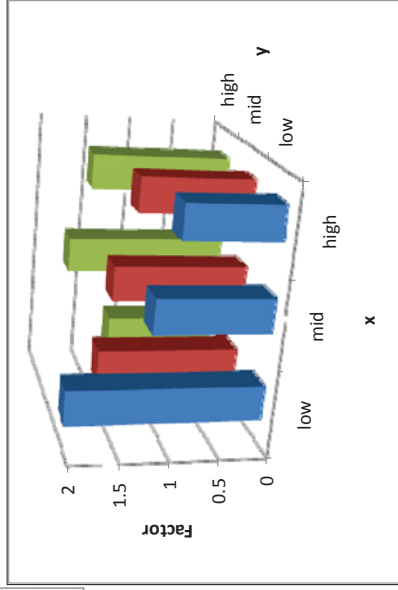
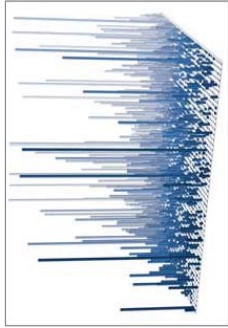
Variable selection: predictiveness by peril

Variable/Peril	Fire	Liability	Theft	Water	Wind	Other
By-Peril Territory						
Insurance Score						
Age of Home						
Protection class						
Construction material						
Amount of Insurance						
Other lines						
Full Pay						
Square Feet						
Number of Rooms						
Claim Free						
Retired Flag						
Good Payer						
Prior Claims						
Secondary Residence						
Fire Protective Device						
Theft Protective Device						
Number of Occupants						

- By-peril territory, Insurance Score, Amount of Insurance, Full Pay, Age of Home, and Claims History are consistently powerful across all perils
- Territory shows larger spread in weather perils
- Insurance Score is predictive in weather perils

Variable selection: identifying interactions

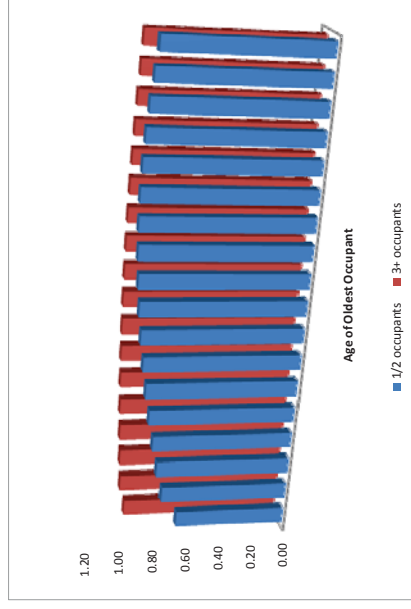
- Looking for situations where the effect of variable x differs depending on variable y
- Granularity can be a problem so grouping is often needed before testing for interactions



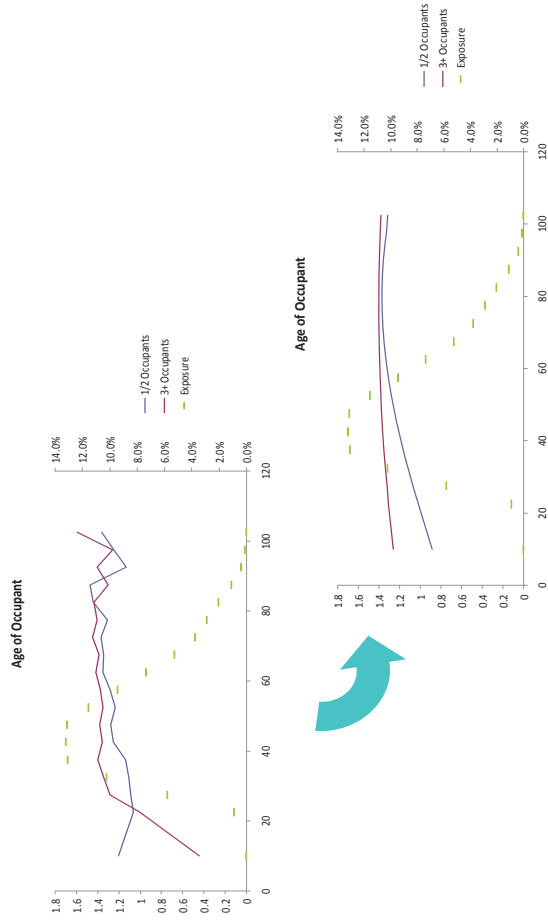
Variable selection: modeling interactions

Move from categorical*categorical interaction to categorical*continuous interaction

	young household	old household
small household	0.6	1.2
large household	0.9	1.1



Variable selection: modeling interactions



17

By-peril territories

Practical considerations

- Territories are a collection of units (5 digit postal code, 3 digit postal code, counties, puma, etc)
- Data at the unit level by-peril is noisy due to limited information in one area
- Territories are correlated with other rating variables (e.g. amount of insurance, age of home)

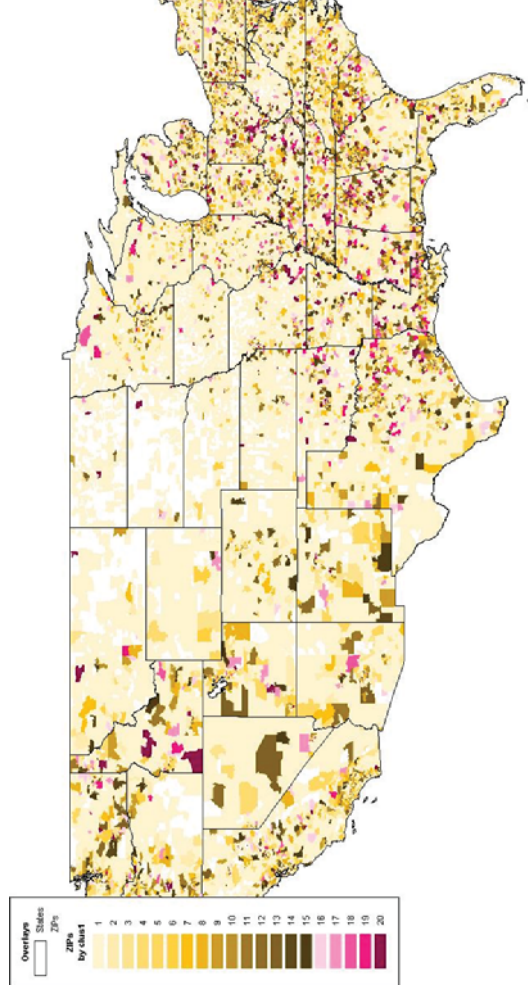
Modeling solutions

- Use territories developed by third parties using industry data
- Use residual risk based on initial models that include house information, occupant information, and external weather, demographics, geographical data
- Use residual risk based on internal data only

18

By-peril territories

Unsmoothed residual risk



19

By-peril territories: smoothing

Distance based

- Nearby units play a bigger role
 - Farther units play a smaller role
- Each unit of distance adds the same amount of risk independent of location
More appropriate for weather related

Adjacency based

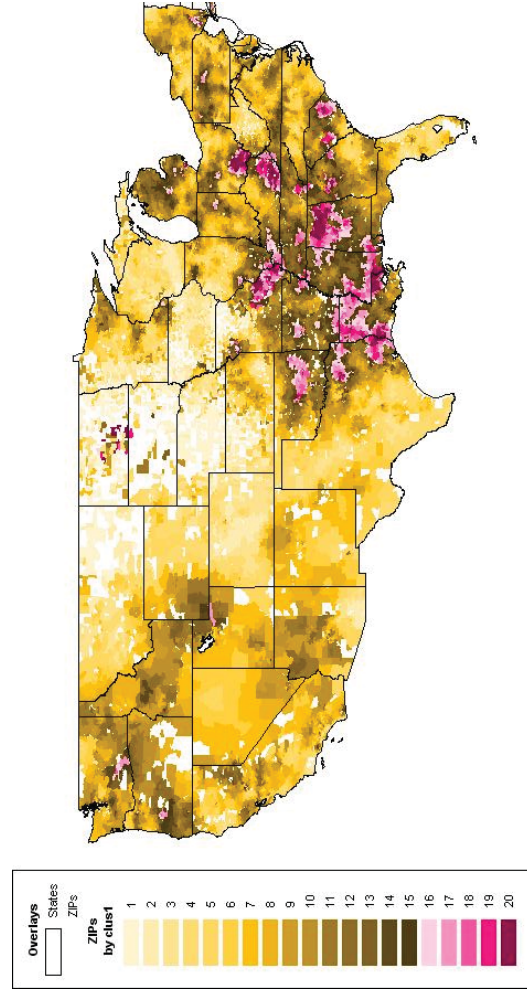
- Surrounding units play a bigger role
- Outer rings of units play a smaller role

Clustering smoothed residuals

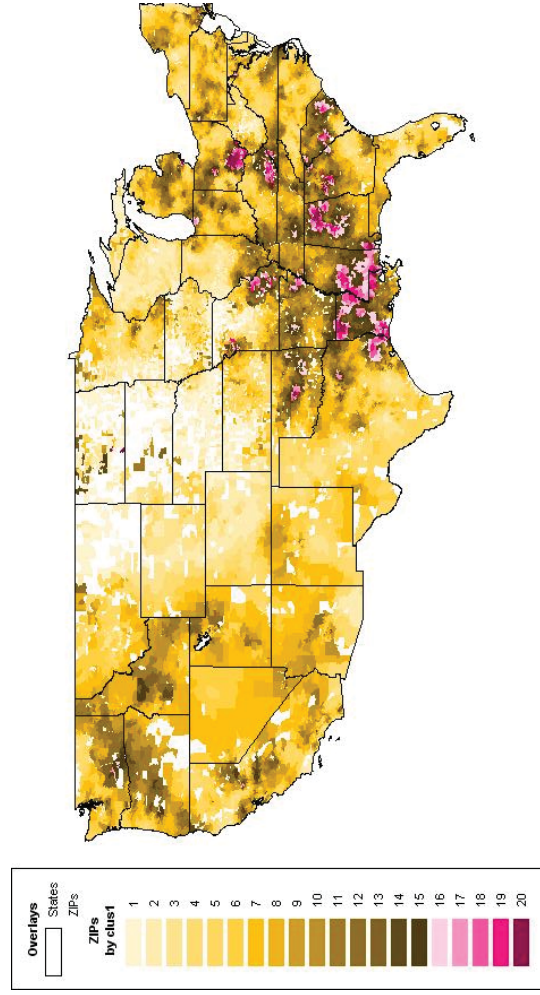
- Maximize variance between clusters
- Minimize variance within clusters

20

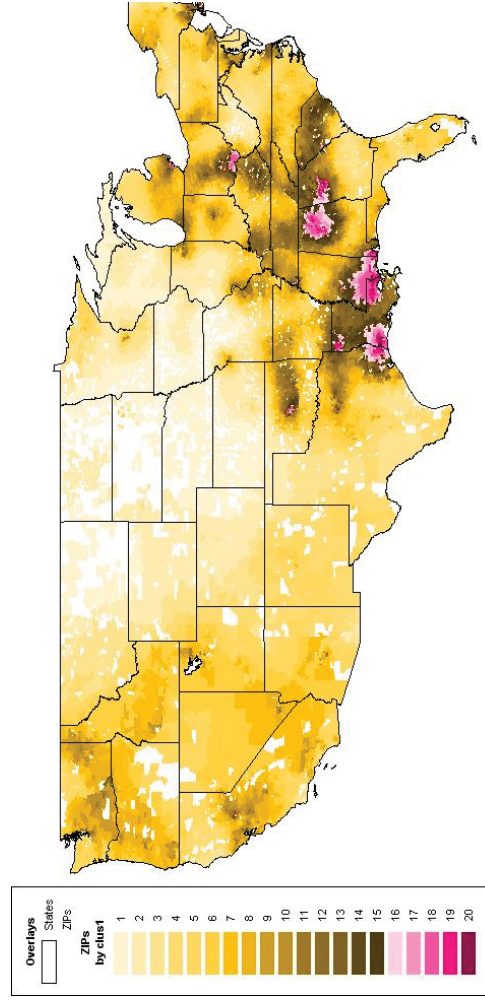
By-peril territories: how much smoothing?



By-peril territories: how much smoothing?



By-peril territories: how much smoothing?



Model validation

A couple of options...

Out of sample validation

- Traditional splitting of modeling and validation of the entire dataset may not work
- Out of sample validation might fail if the observations are not independent (weather related perils)
- The losses coming from the same "event" would be found both in the modeling and in the validation dataset

Out of time validation

- Could solve the independence issue if one year is kept aside for validation

Conclusion

- Homeowners predictive modeling could be as sophisticated and innovative as auto modeling
- By-peril modeling is an important way of achieving increased sophistication and accuracy

Contact

David R. MacInnis
Senior Predictive Modeler
Allstate Insurance Company
drmaau@allstate.com