November 6–9, 2022

Hilton Minneapolis

Minneapolis, Minnesota

ANNUAL MEETING

Applications of Gaussian Process Regression Models

Marco De Virgilis & Giulio Carnevale

Presenters

Marco De Virgilis
Senior Actuarial Data Scientist
Arch Insurance Group

Giulio E. Carnevale
Reserving Actuary
PartnerRe

Disclaimer

The views and opinions expressed in this presentation are those of the authors' and do not necessarily reflect the position of the organizations of which they are part.

Agenda

- Theoretical Background
- Covariance Functions
- Claim Data Representation
- Application
- Model Comparisons
- Conclusion

Theoretical Background

Theoretical Background

- Gaussian Processes (**GP**) are stochastic processes based on the normal distribution.
- Gaussian Process Regression (**GPR**), is a relatively lesser-known procedure based on Gaussian processes, which can be implemented in the context of machine learning.
- GPR can be defined as a supervised non-parametric machine learning technique stemming from the **Bayesian** field.
- One of the main features of GPR is the ability to produce **a probability distribution** of functions that fit a set of observations and so return predictions with **uncertainty intervals** around them.
- GPR techniques can be ideal candidates to extend traditional stochastic reserving techniques in order to quantify **reserve variability**.

Bayesian Framework

- In the context of Machine Learning a **full range of estimates** is a remarkable result, as the most famous alternatives only provide point estimates.
- **Bayesian inference** allows us to *update* our views considering the observed evidence:

Rather than fitting a model purely from data, this approach blends the external ***prior*** knowledge with **actual observation**s, resulting in a ***posterior*** estimate that synthetizes the two

Bayesian Framework

The Bayesian approach comes with advantages and disadvantages:

Advantages:

1. Incorporate in models external knowledge not present in the data, which is particularly useful for actuarial applications.
2. Predictions come naturally in a probabilistic form.

Disadvantages:

1. Inference leads almost always to intractable mathematical formulas which can be treated only via simulation.
2. It can be difficult to translate prior knowledge into suitable priors.

Gaussian Processes as regression tools

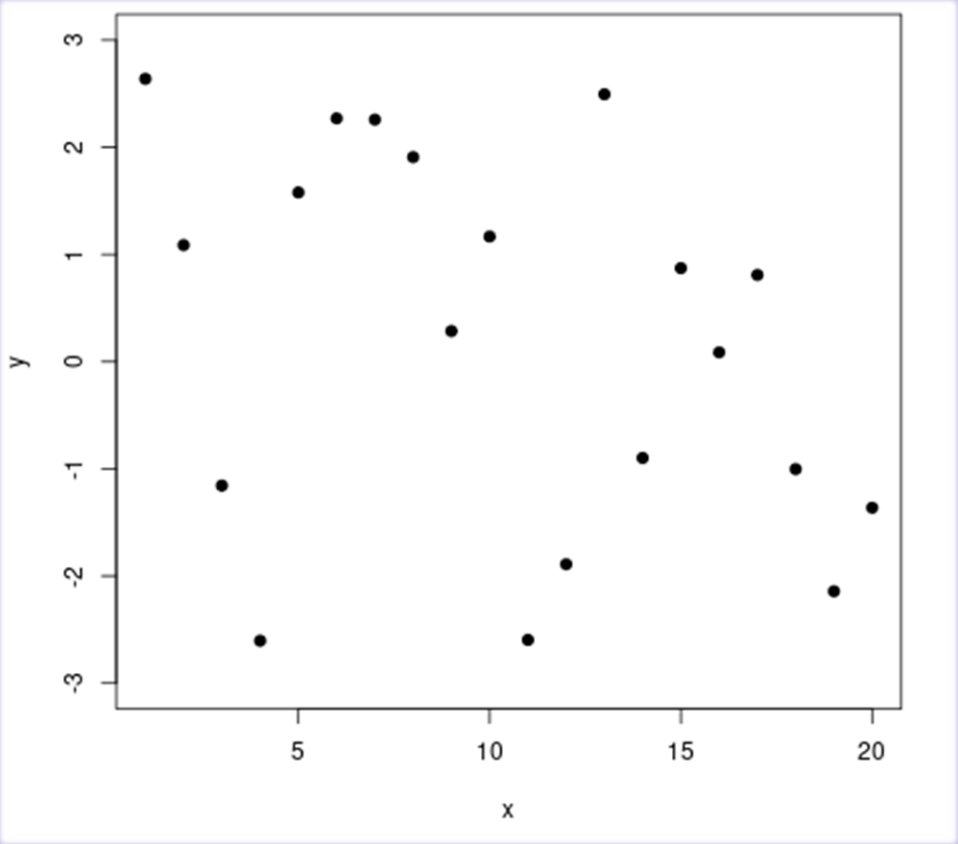
Gaussian Processes (**GP**) are stochastic processes that generalize the Multivariate Normal distribution. The generalization step lies in the parametrization:

$$\mathrm{GP}(x) = N(\boldsymbol{\mu}(x), \boldsymbol{\Sigma}(x, x'))$$

Where:

- $\boldsymbol{\mu}$(x) is the mean function
- $\boldsymbol{\Sigma}(x, x')$ is the covariance function

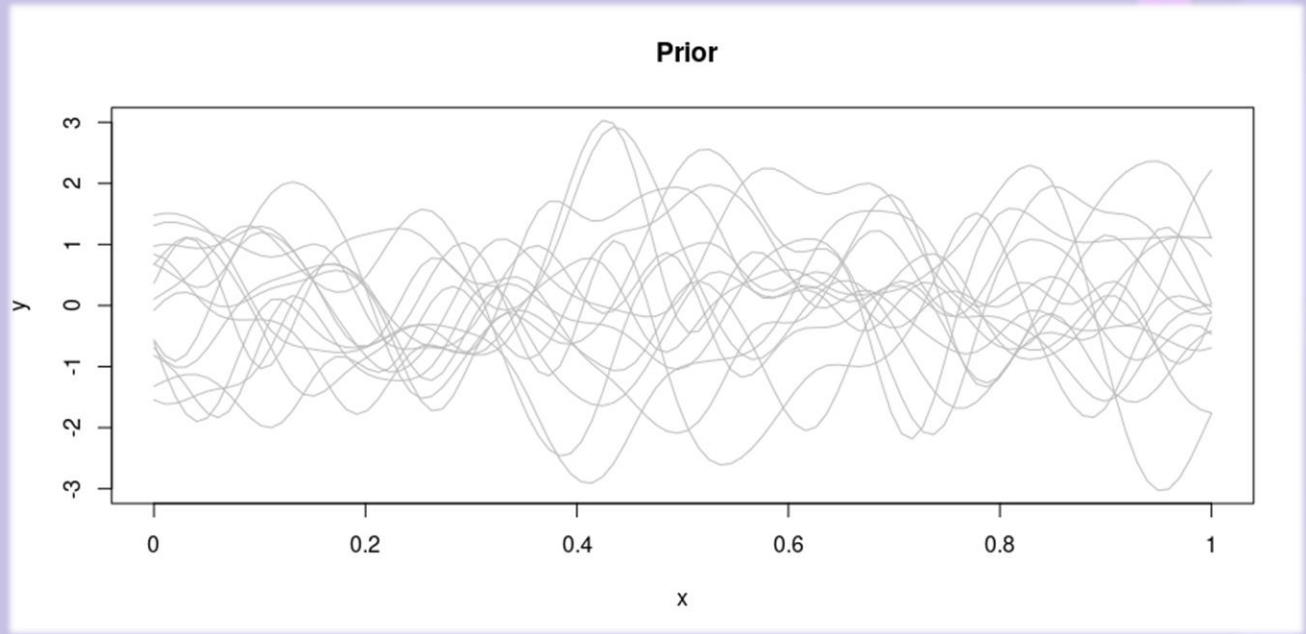
Gaussian Processes as regression tools



A GP can be thought as the process that generates the data points, i.e. the cloud of n data points can be thought as a n-dimensional realization of a GP($\boldsymbol{\mu}$(x), $\boldsymbol{\Sigma}(x, x')$):

- $\boldsymbol{\mu}$(x) is the function that describes where on average is to be found every n-th realization. (Usually set to 0)
- $\boldsymbol{\Sigma}(x, x')$ is the Covariance function or *kernel* and determines how spread are the realizations. It can be quickly interpreted as a measure of distance between the realizations.

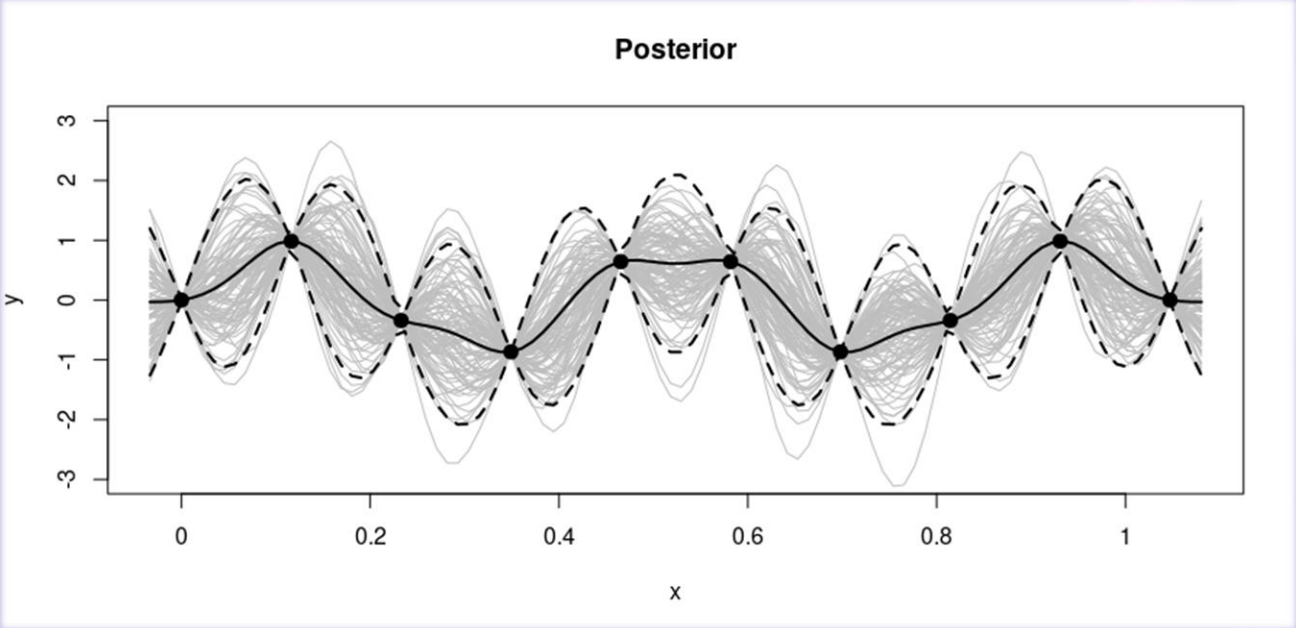
Gaussian Processes as regression tools



The functions $\boldsymbol{\mu}$(x) and $\boldsymbol{\Sigma}(x, x')$ characterize the process. From a well-defined GP we can sample as many realization as we desire.

The picture shows many realizations from a particular GP.

Gaussian Processes as regression tools



Introducing a set of observations, we can **retain** from the prior samples **only the functions** that pass through the observation points: **effectively blending prior knowledge and actual evidence.**

Covariance Functions

Covariance functions

- Covariance Functions, or *kernels,* fully specify the functional form of the **prior** and significantly impact the final form of the posterior.
- A very simple and intuitive meaning of $\boldsymbol{\Sigma}(x, x')$ is the one of **distance**: we can assume that the correlation between two points decreases the farther away they are from each other.
- For instance, if two input points are very close to each other and expected to behave **similarly**, the kernel will have a value close to **1**. Conversely, when the points are **far apart** and there's no expectation they will behave similarly, the kernel will have a value close to **0**.

Euclidean distance

The Euclidean distance is one of the most commonly used *kernels* and from this the meaning of distance can be immediately understood:

$$\boldsymbol{\Sigma}(x, x') = \exp(-||x - x'||^2)$$

The value of the output decreases exponentially with the increase of the distance between two points.

Squared Exponential

This is another commonly found kernel and has the form:

$$\boldsymbol{\Sigma}(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\theta^2}\right)$$

Where σ^2 is the scaling factor and θ the lengthscale factor. Again, the concept of distance is very evident in the definition.

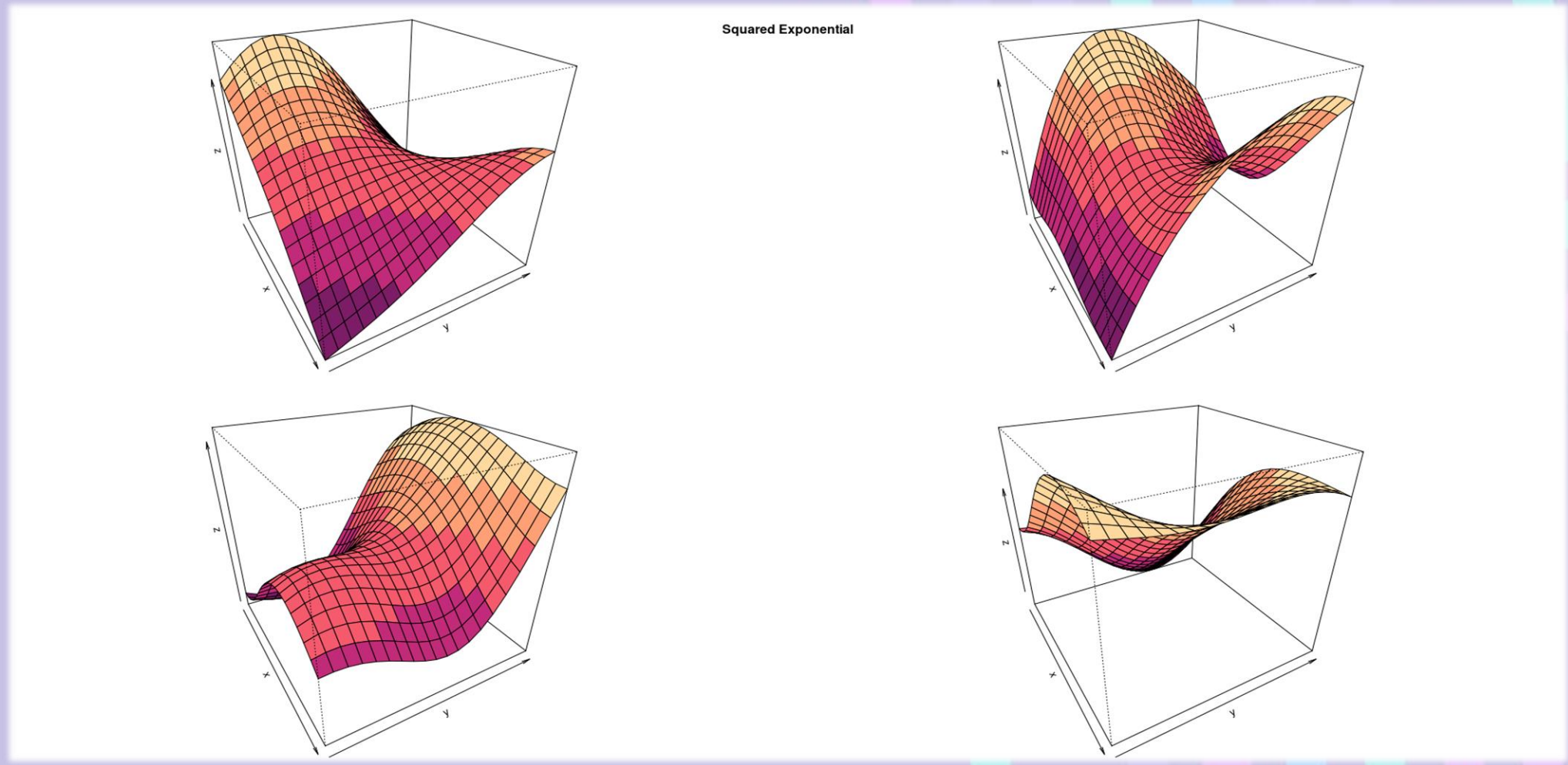
Matern 3/2 and Matern 5/2

$$\Sigma(x, x') = \sigma^2 \left(1 + \frac{\sqrt{3}|x - x'|}{\theta}\right) exp\left(\frac{-\sqrt{3}|x - x'|}{\theta}\right)$$

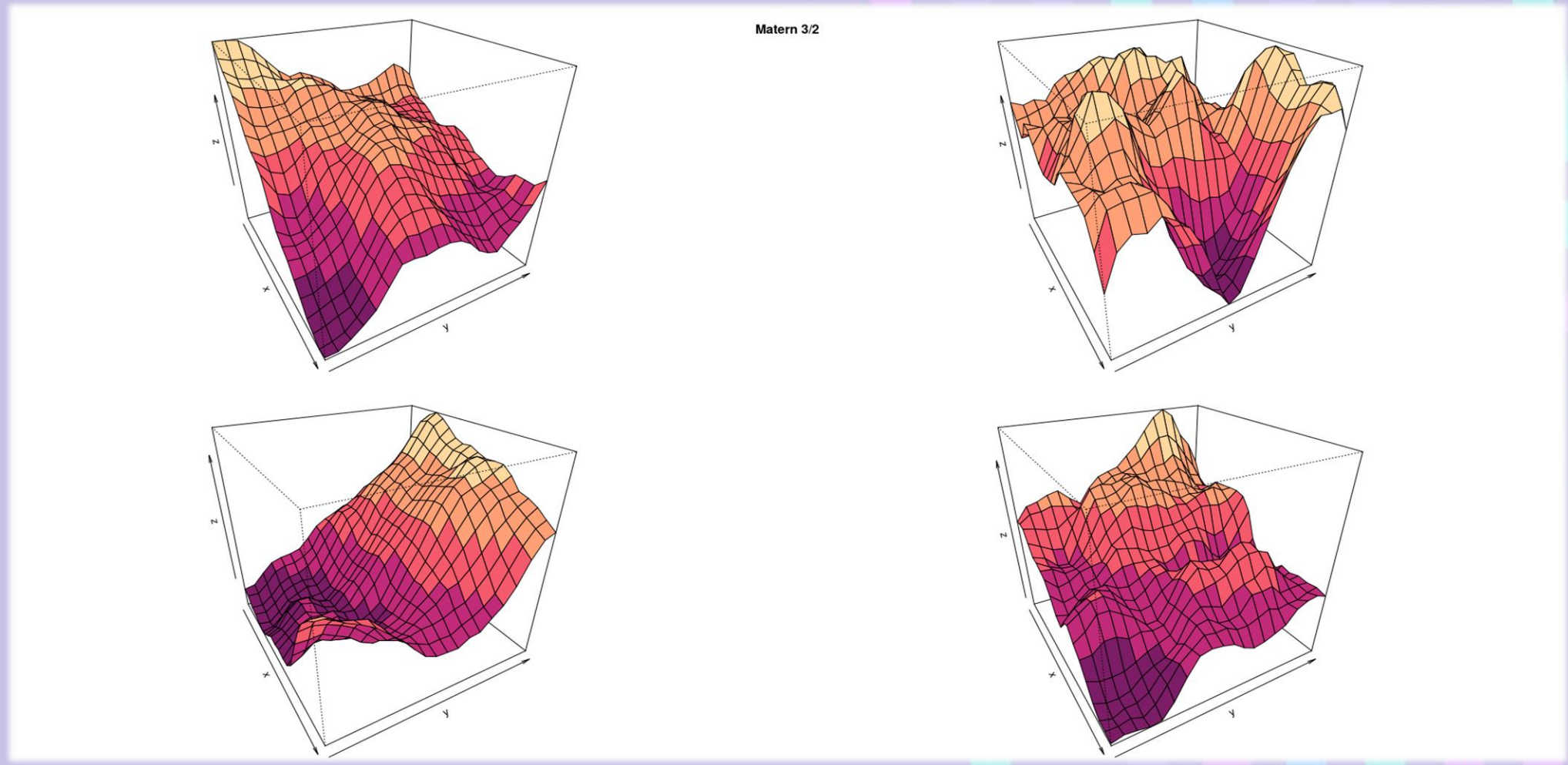
$$\Sigma(x, x') = \sigma^2 \left(1 + \frac{\sqrt{5}|x - x'|}{\theta} + \frac{5|x - x'|^2}{3\theta^2}\right) exp\left(\frac{-\sqrt{5}|x - x'|}{\theta}\right)$$

Similarly to the previous example, σ^2 is the scaling factor and θ the *lengthscale* factor. The covariance between two measurements is a function of the distance.

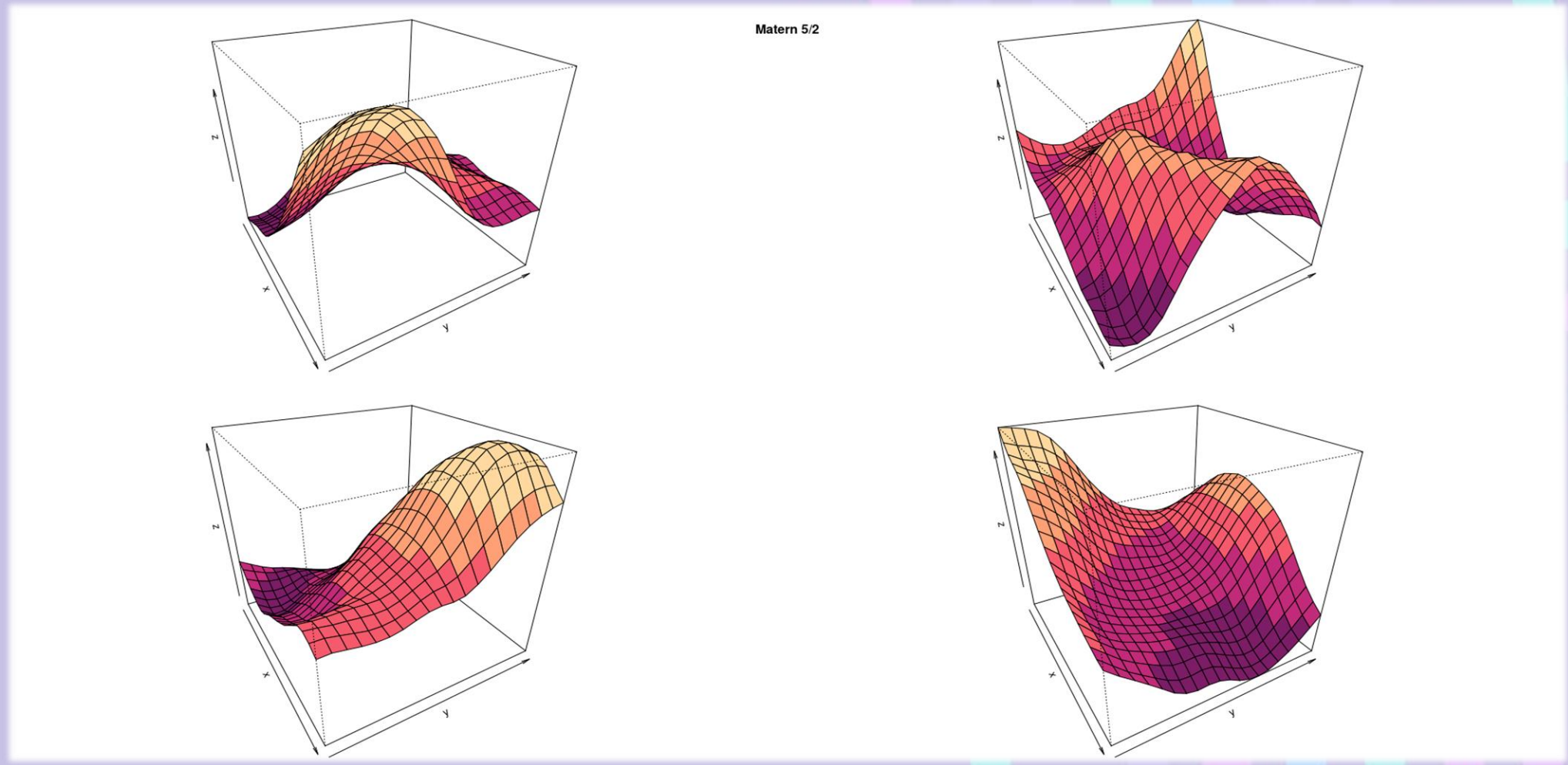
Squared Exponential - Samples



Matern 3/2 - Samples



Matern 5/2 - Samples



Claim Data Representation

Wide version of data

Actuaries have been analyzing claims in the form of run-off triangles, a matrix representation (or, equivalently, wide format table) where origin periods are reported on the rows and development periods on the column:

AY\DP	1	2	3	4	5	6	7	…	n
1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,\dots}$	$X_{1,n}$
2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,\dots}$	
3	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$		
4	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$			
5	$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$	$X_{5,5}$				
6	$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$					
7	$X_{7,1}$	$X_{7,2}$	$X_{7,3}$						
…	$X_{\dots,1}$	$X_{\dots,2}$							
n	$X_{n,1}$								

Long version of data

Instead of the usual triangle, we can think of claim data in a long format. Common practice for data scientists, but maybe not so common for actuaries:

AY	DP	Amount
1	1	$X_{1,1}$
1	2	$X_{1,2}$
1	…	$X_{1,\dots}$
1	n	$X_{1,n}$
2	1	$X_{2,1}$
2	2	$X_{2,2}$
2	…	$X_{2,\dots}$
2	n-1	$X_{2,n-1}$
…	…	$X_{\dots,\dots}$
n	1	$X_{n,1}$

Long version of data

It is still easy to see how Accident Year and Development Period represent the coordinates on a typical 3-d cartesian plane.

AY	DP	Amount
1	1	$X_{1,1}$
1	2	$X_{1,2}$
1	…	$X_{1,\dots}$
1	n	$X_{1,n}$
2	1	$X_{2,1}$
2	2	$X_{2,2}$
2	…	$X_{2,\dots}$
2	n-1	$X_{2,n-1}$
…	…	$X_{\dots,\dots}$
n	1	$X_{n,1}$

Long version of data

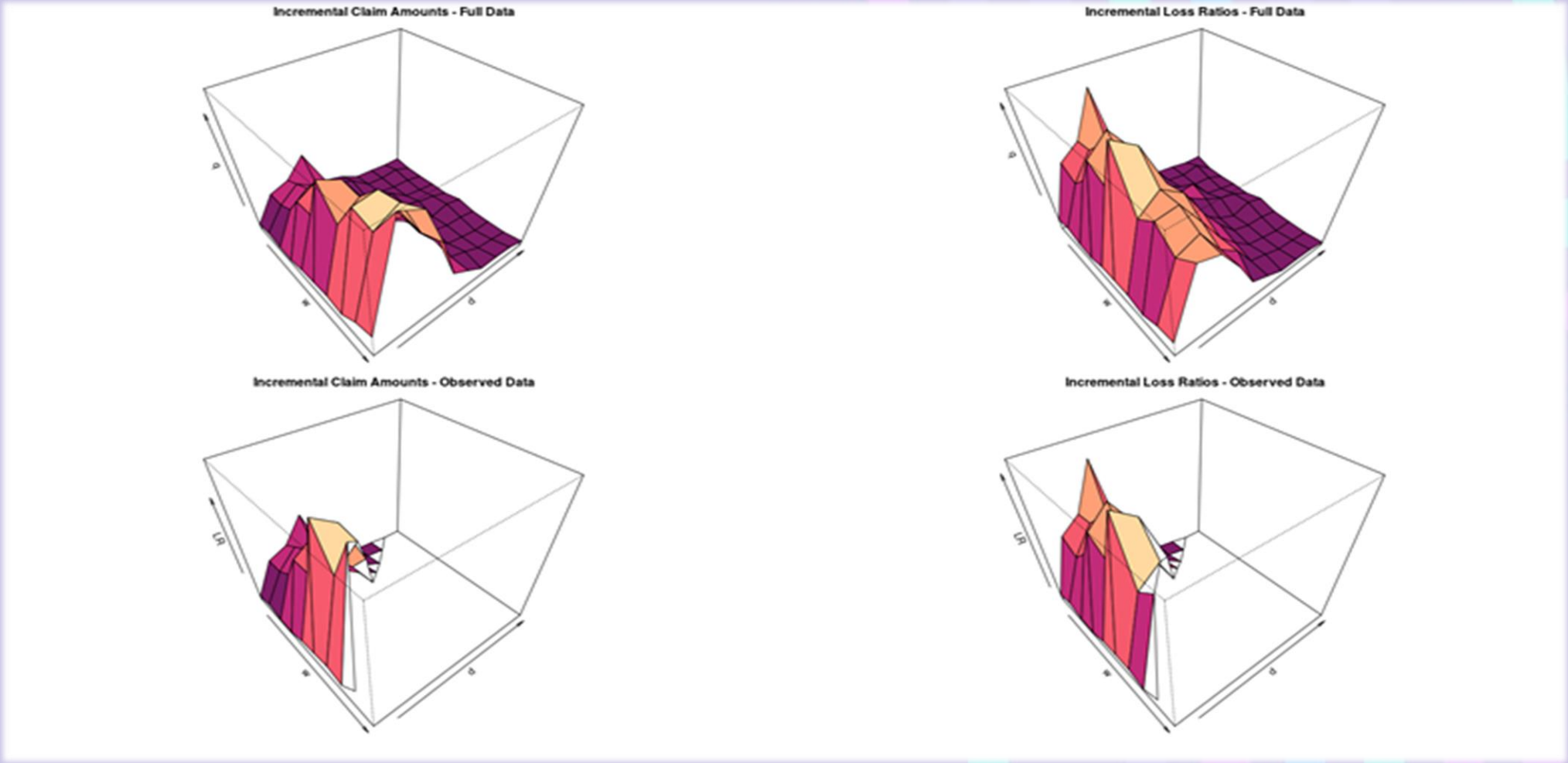
The goal of the method is to predict future observations (lower triangle) after training on history (upper triangle), in a typical ML fashion.

Effectively, estimating a function: $f(AY, DP) = X_{AY,DP}$

Where $f(.)$ is a GPR.

AY	DP	Amount
2	n	$X_{2,n}$
3	n-1	$X_{3,n-1}$
3	n	$X_{3,n}$
…	…	$X_{\ldots,\ldots}$
…	n	$X_{\ldots,n}$
n	n	$X_{n,n}$

Graphical Representation of Claim Data



Application

Application on real data

- We tested this methodology on both **Claim Amounts** and **Loss Ratios** from the publicly available triangles from the NAIC Schedule P (G. Meyers and P. Shi, 2011).
- This dataset includes major personal and commercial lines of business from many US P&C insurers.
- This dataset is made of **fully developed triangles**, allowing to test of the procedure interms of AvsE.

Application on real data

- Before fitting the models, we performed several data cleaning and data quality operations to ensure data was suitable for our purposes.
- We worked with **10x10 triangles** with no missing data points and no major business changes.
- We fit the model on the upper triangle (observed data) and estimated predictions for the lower triangle.
- **The procedure has been iterated for all the triangles in the dataset.**

Application on real data

- In order to evaluate the quality of the **point estimates** we looked at **RMSE** (Root Mean Square Error) of the observed reserve vs the estimated one.
- With respect to the **stochastic** properties, we have calculated the observed reserve percentile on the estimated distribution and then tested it with **Kolmogorov-Smirnov test** for uniformity (Meyers, 2015).
- The main idea behind this test is that **differences between Observed and Estimated data would be uniformly distributed**, implying that **on average the model is unbiased**: i.e. on average we do not have predictions distorted upwards or downwards.

Model specifications and parameters

As mentioned, we are fitting **Incremental Paid Claims** and **Incremental Paid Loss Ratios** as function of AY and DP:

$$X_{AY,DP} \sim \mathrm{GP}(0, \boldsymbol{\Sigma}(x, x'))$$

$$\frac{X_{AY,DP}}{EP_{AY}} \sim \mathrm{GP}(0, \boldsymbol{\Sigma}(x, x'))$$

Input Warping and Standardization

- Inputs are on two very different scales: AY goes from 1988 to 1997 whereas the DP goes from 1 to 10.
- To obtain stable and reliable results, a standardization is necessary, in order to remove distorsions due to scale differences.
- We standardized data according to a procedure called **Input Warping** (Snock et al., 2014)

Input Warping and Standardization

- We applied to both AY and Dev Year the ***min-max scale:***

$$\mathrm{w}' = \frac{w - w_{min}}{w_{max} - w_{min}}$$

- Then the normalized data points are taken as the input to a function of **Beta Cumulative Distribution function:**

$$\mathrm{w}'' = I_{w'}(a_w, b_w) = \frac{B(w', a_w, b_w)}{B(a_w, b_w)}$$

Where a_w and b_w are real parameters.

Input Warping and Standardization

- The procedure returns a number between **0 and 1** for both AY and DP.

- In simple terms, we can think about this procedure as a sophisticated **standardization technique** that normalizes the inputs.

Model specifications and parameters

We employed 3 prior kernels for modeling purposes:

1. Squared Exponential
2. Matern 3/2
3. Matern 5/2

Each of them for both **Incremental Paid amounts** and **Incremental Paid Loss Ratios.**

Comparison

Procedure Comparison

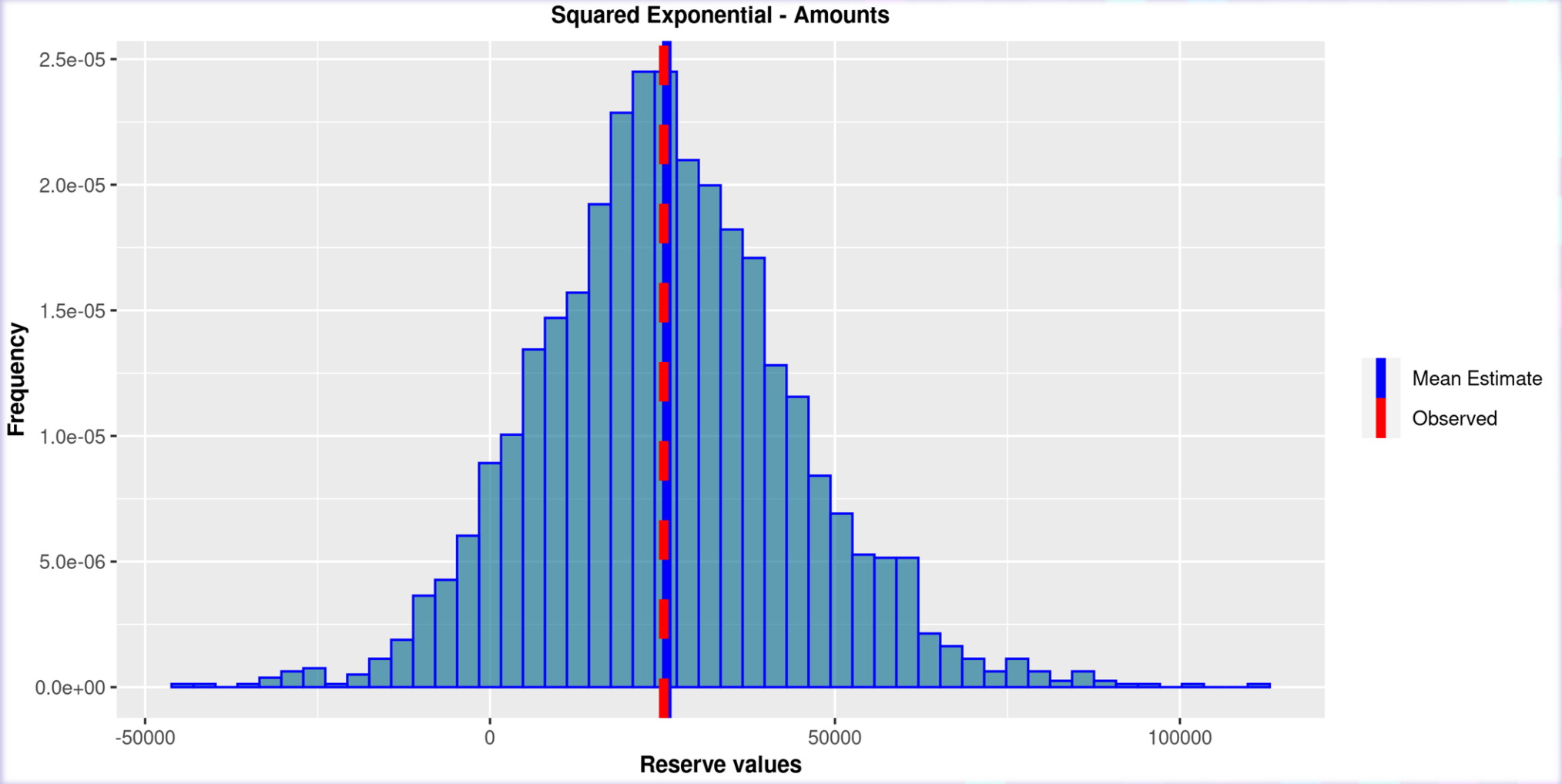
- We ran the procedure on every triangle in the dataset, and being the actual observed development claims available, we were able to **test** the results, both in terms of **point estimate** and **stochastic accuracy.**
- With respect to **predictive power** performance, we used the classical metric of **RMSE.**
- With respect to **stochastic accuracy**, we gathered the *percentile* at which the **Actual reserve** fell on the **Predictive Distribution** for all the triangle in the dataset. Subsequently, we tested that the distribution of the percentiles was **uniform** on the interval (0,1) using the Kolmogorov-Smirnov test.

Procedure Comparison

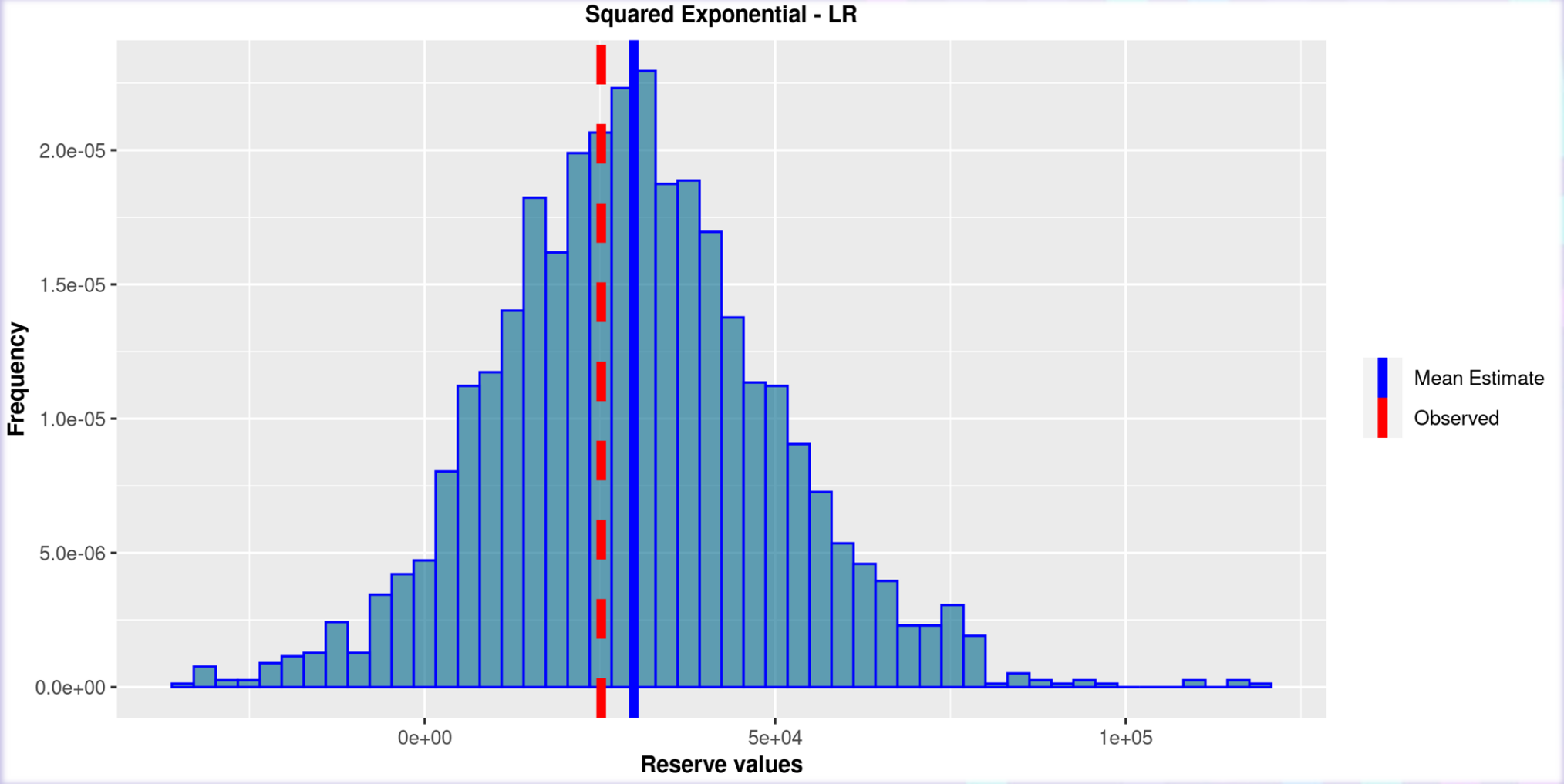
Method	RMSE	K-S Test
GPR - Squared Exponential (amounts)	770,706	0.152
GPR - Matern 3/2 (amounts)	549,021	0.103
GPR - Matern 5/2 (amounts)	**247,976**	**0.165**
GPR - Squared Exponential (LR)	449,609	0.149
GPR - Matern 3/2 (LR)	481,146	0.17
GPR - Matern 5/2 (LR)	**235,784**	**0.358**
Chain Ladder ODP	558,320	0.001
Cape Cod ODP	416,693	0.001

RMSE: the lower, the better
K-S Test: the higher, the better.

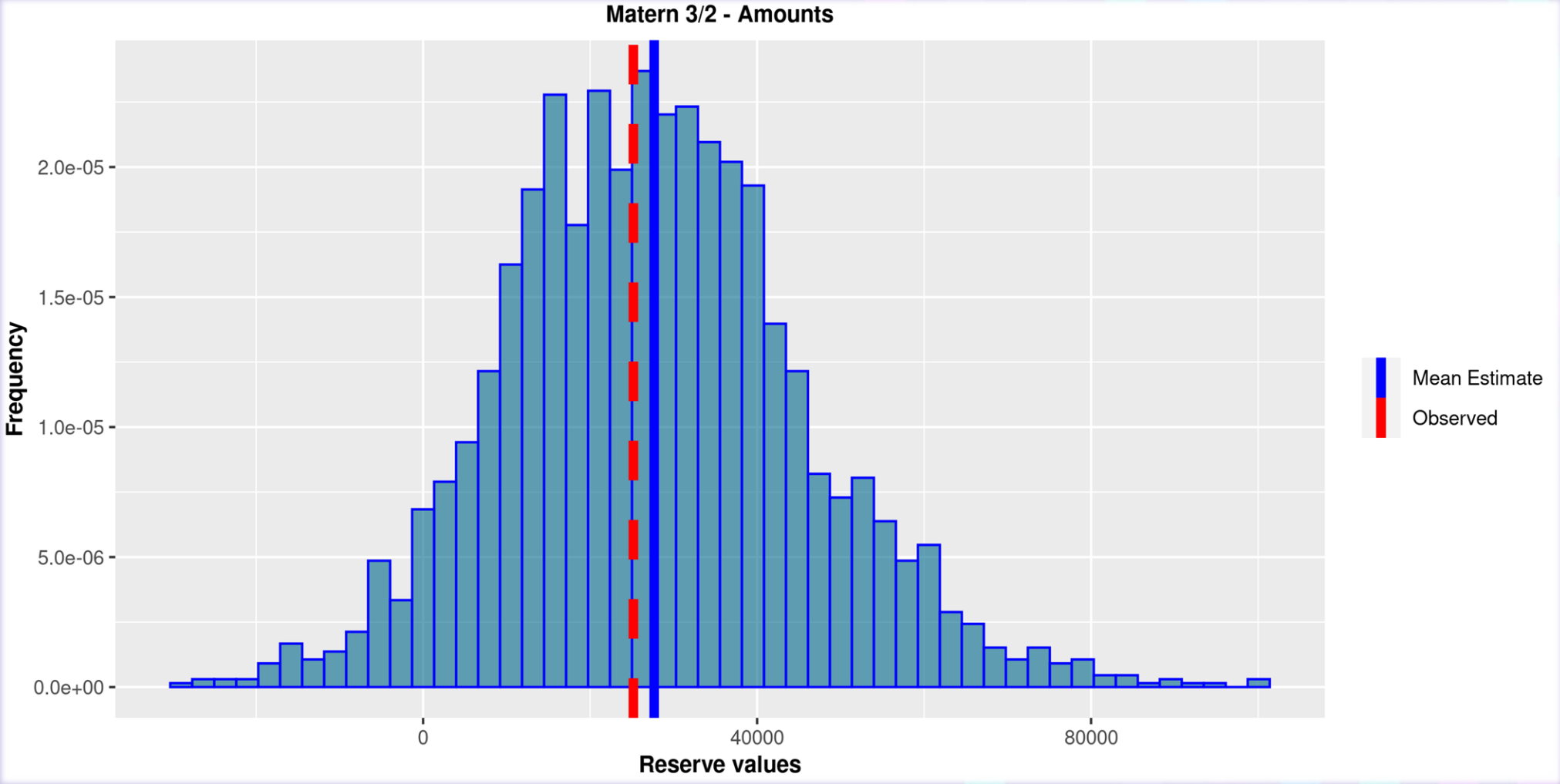
Ultimate Amounts



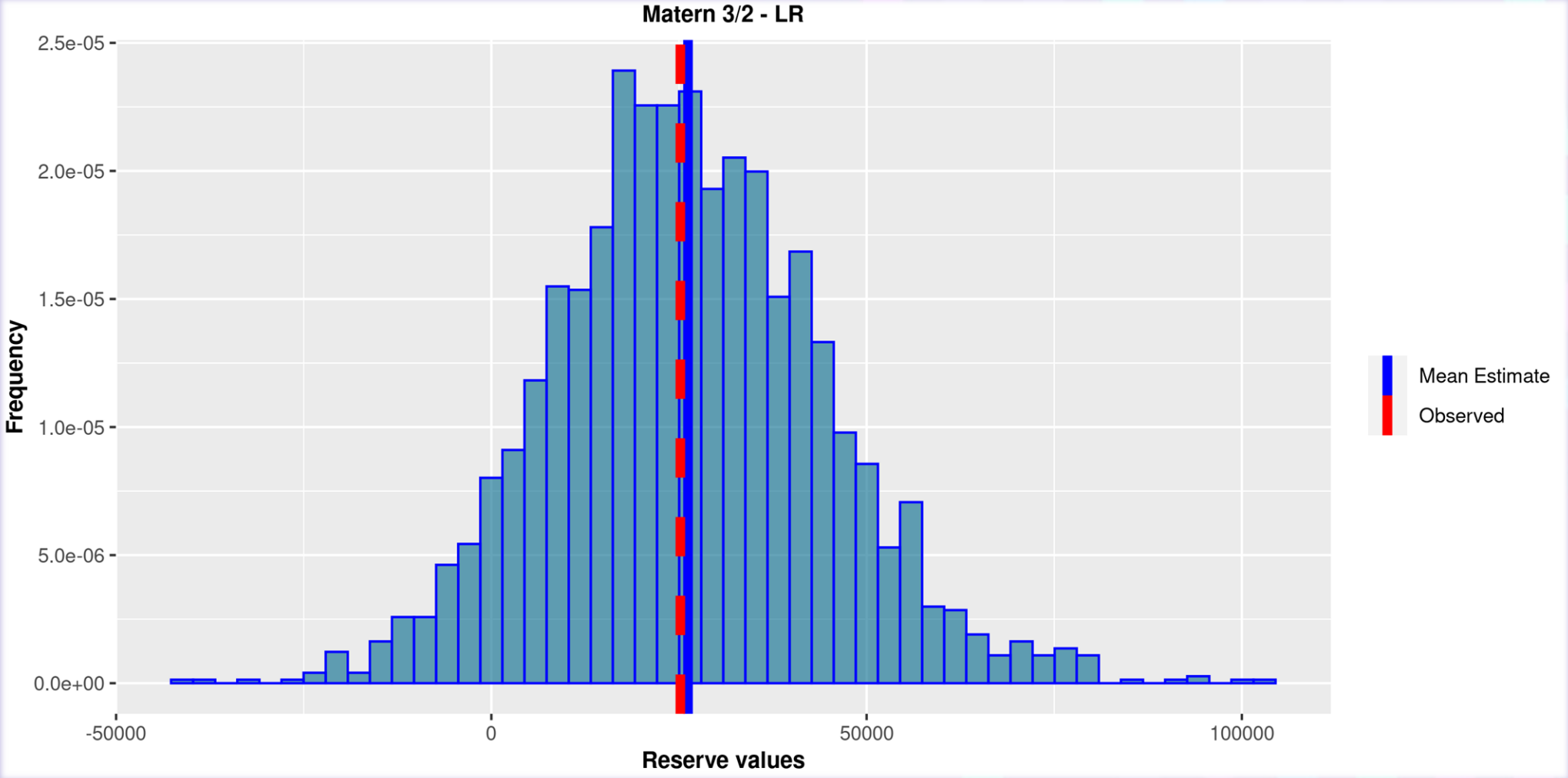
Ultimate Amounts



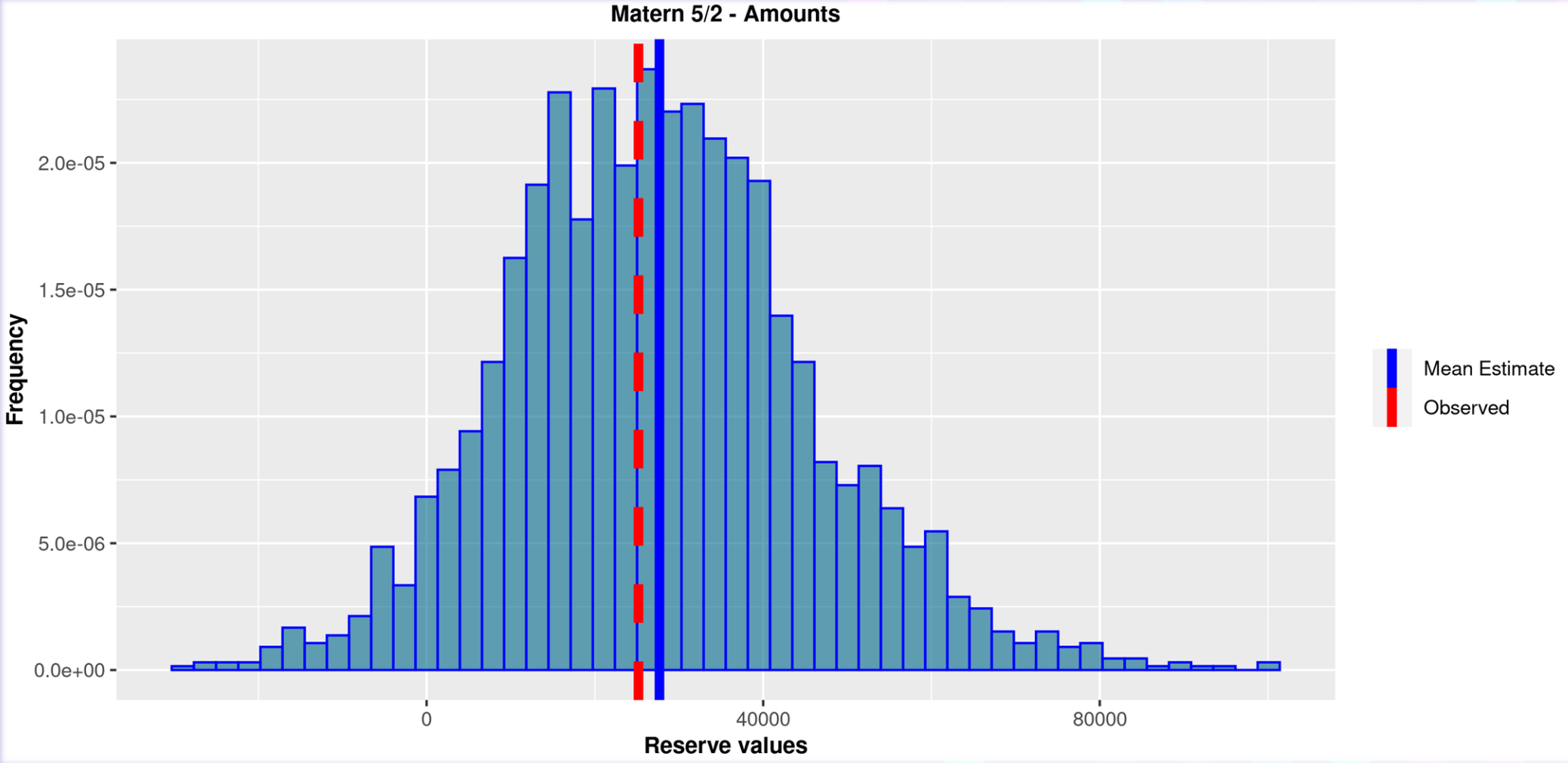
Ultimate Amounts



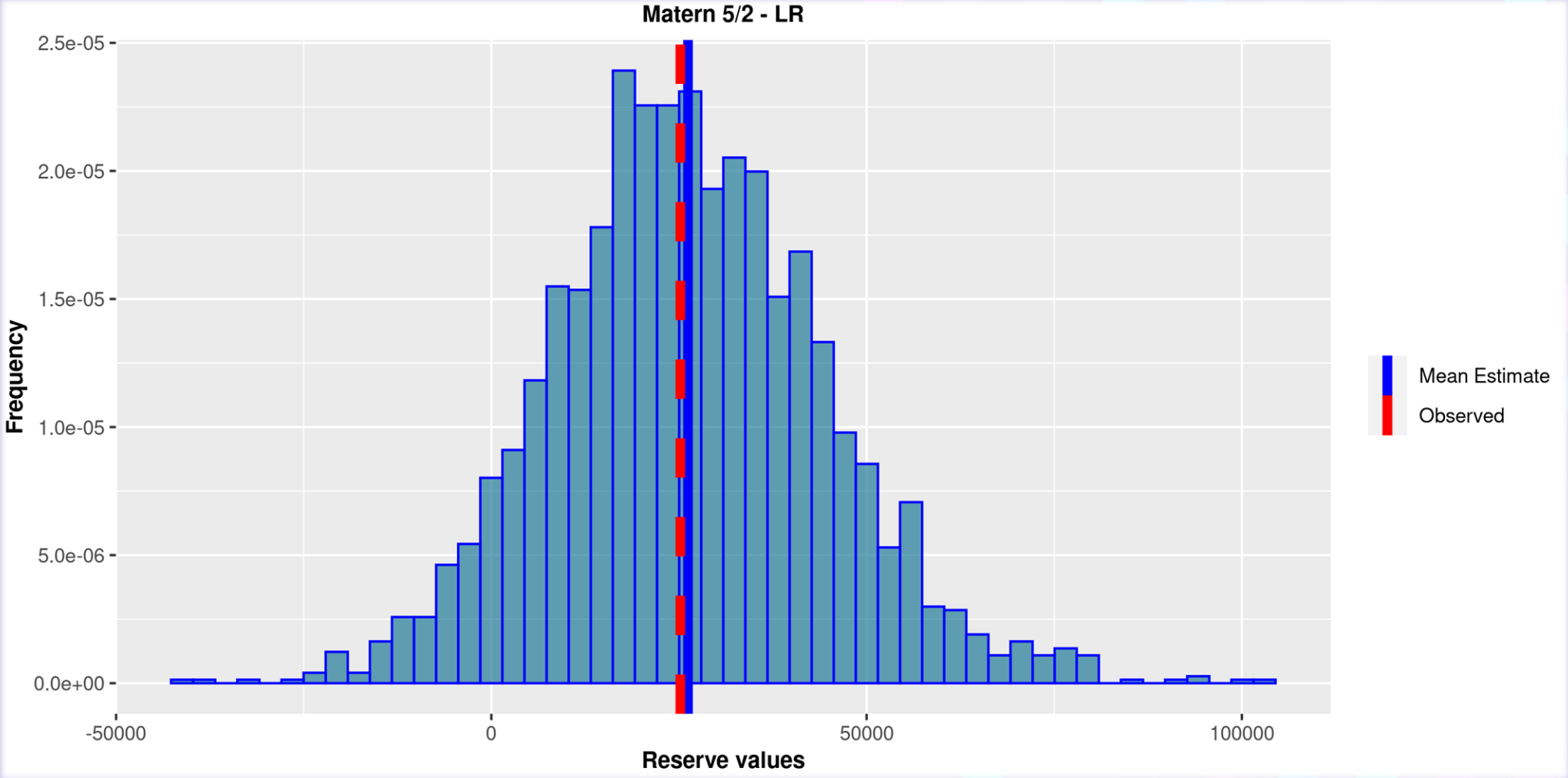
Ultimate Amounts



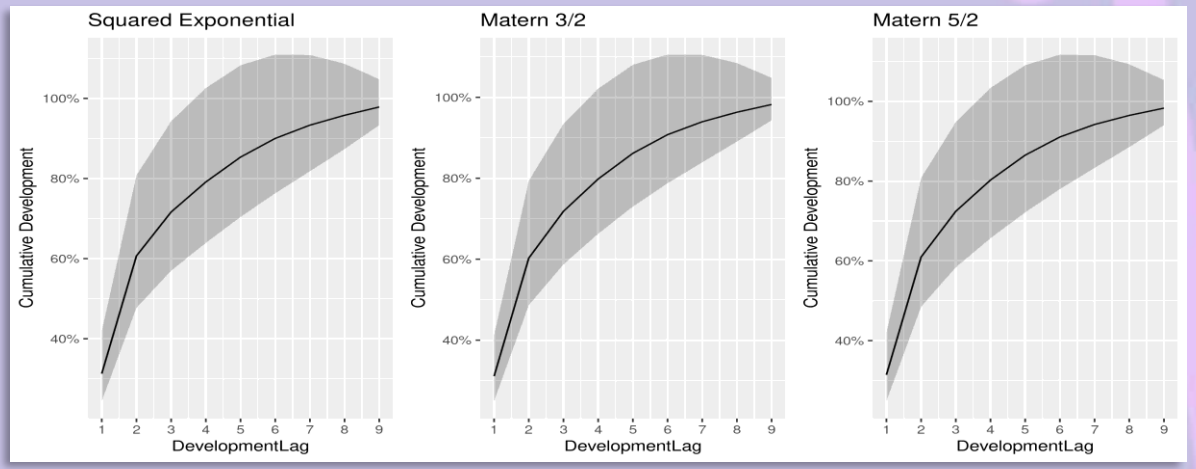
Ultimate Amounts

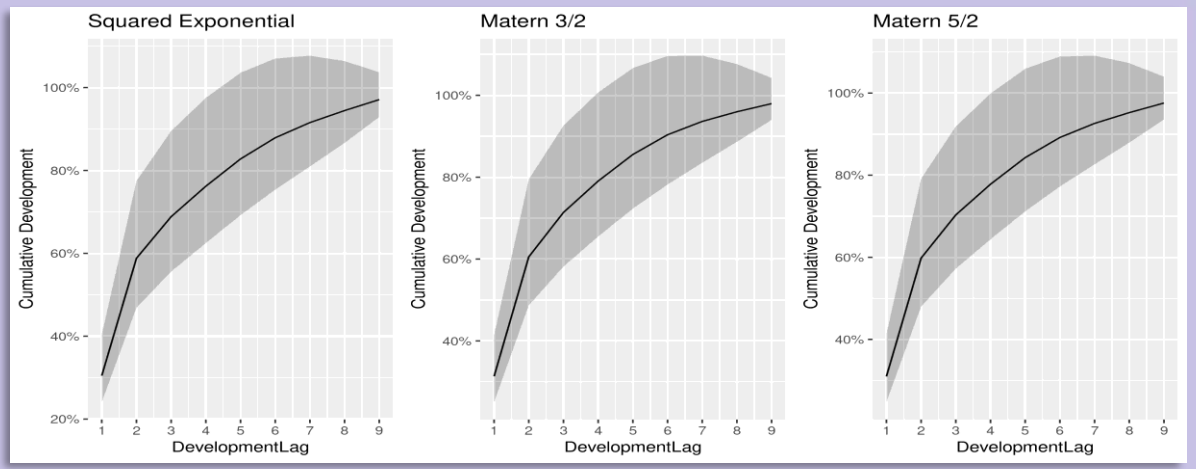


Ultimate Amounts



Procedure Comparison





- Having a full range of estimates, we can produce the implied **probability distribution of the Loss Development Factors.**

- This feature can turn out particularly useful for practitioners as it makes possible to **pick a selection** according to a specific **risk appetite**.

Variance Decomposition

The variance of the predictive distribution can be decomposed in its two main components: **Expected Value of Process Variance** (EVPV) and **Variance of Hypothetical Mean** (VHM).

Method	Mean Reserve	Total Variance	EVPV	VHM
Squared Exponential (amounts)	25,621	104,892,875	104,785,405	107,470
Matern 3/2 (amounts)	27,672	88,379,162	88,291,744	87,418
Matern 3/2 (amounts)	24,752	93,356,182	93,263,299	92,883
Squared Exponential (LR)	29,835	94,209,270	94,115,054	94,216
Matern 3/2 (LR)	26,202	88,015,296	87,927,915	87,380
Matern 5/2 (LR)	27,182	90,852,410	90,755,527	96,883

Variance Decomposition

- The decomposition of variance is the starting point for further research on claims reserve variability on a **n-years** time horizon, which extends the applications of the methodology to **capital modelling**.
- Bayesian models seem to be **particularly promising** in this respect, as variance is easily decomposed: the variability of parameters (so the estimation variability) is described by the **posterior probability distribution**. On the other hand, the process variance is simulated via the **predictive distribution.**

Conclusions

Conclusions

- GPR models are able to **outperform traditional techniques** from a **deterministic**, as well as a **stochastic**, point of view.
- Moreover, due to the nature of the method, we are also able to further investigate the **distributions** of the model parameters in order to further **study the variability** of the claim estimates.
- In a context where the **stochastic behavior** of the reserve estimate is valuable, such an approach can definitely bring a hedge against more simplistic methodologies, e.g. ODP.

Conclusions

- The methodology, as presented, could definitely be improved and refined. Further studies could be focused on the implementation of **different Covariance functions** such as the Periodic Covariance function or higher orders of the Matern formulation.
- Furthermore, it would be interesting to define similar **multivariate** processes that are **non-Gaussian** in nature: a good candidate for insurance claim analysis may be the Wishart distribution, i.e. a multidimensional generalization of the **gamma distribution** (Eaton 2007).
- Another area of improvement could be the introduction of **further dimensions**. In our study we had the AY and the DP as inputs, however in some contexts where **inflation** is an important factor to consider (e.g. structured settlements analysis) having the **Calendar Year** as an additional predictor could bring additional value.

Contacts:

Marco De Virgilis
devirgilis.marco@gmail.com

Giulio Carnevale
giulio.carnevale1@gmail.com

Casualty Actuarial Society
4350 North Fairfax Drive, Suite 250
Arlington, Virginia 22203

www.casact.org