KOALA: Estimating coalition probabilities in multi-party electoral systems

KOALA
KoalitionsAnalyse

Alexander Bauer[1], Andreas Bender[1], André Klima[1], Helmut Küchenhoff[1]

[1] *Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Germany*

Alexander.Bauer@stat.uni-muenchen.de

Data and Research Question

Setting

We analyze artificial earthquake data derived from large-scale computer simulations based on a 1994 earthquake in Northridge (USA). In each of 135 simulations, the (isotropic) **absolute ground velocity** [m/s] was measured at 6146 virtual seismograms with a temporal resolution of 2Hz.

This project marks the first time that physics-based simulations of earthquakes are combined with modern statistical methods. Apart from gaining new insights in the geophysical processes regression models could in future be used to predict expected ground movements in earthquake regions.

Main research question

How do the physical conditions at an earthquake fault affect the surficial ground velocity measured over time?

Challenges

- Very high-dimensional data
- Spatio-temporal functional data

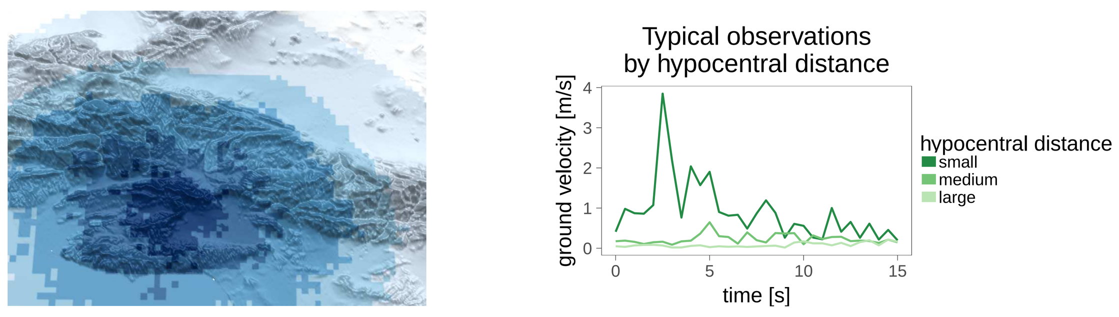

Figure 1: *Left: Categorized mean absolute ground velocity in one simulation over the area under study, darker colours correspond to increased velocity. Right: Typical observations of absolute ground velocity over time. The initial peak of the ground velocity is delayed and smaller as hypocentral distance increases.*

Simulation setup

The artificial earthquake data were generated using the open-source software SeisSol (`www.seissol.org`). In each simulation:

1. Five $\boxed{\text{simulation parameters}}$ were pre-set, and
2. absolute ground velocities were simulated solving elastic wave equations coupled to frictional failure at the earthquake fault.

Influence parameters

The influence parameters are all **constant over time**.

- soil material ({*rock*, *sediment*})
- linear slip weakening distance [m]
- static coefficient of friction
- dynamic coefficient of friction
- direction of tectonic background stress [°]

- hypocentral distance of seismometer [m]
- elevation of seismometer [m]
- landform at seismometer* ({*ridge*, *plain*, *valley*, ...})
- moment magnitude [Nm]

*Categorization into landforms was performed using the Topographic Position Index (TPI) of Weiss (2001)

① Modeling process

We use a **Generalized Functional Additive Model** (GFAM) (see Scheipl et al., 2016) which is an extension of the GAM model class.
In our case, only the response is functional and we use a Gamma model with log-link.

$$y_i(t_l) \sim F(\mu_{il}, \boldsymbol{\nu}) \quad \text{with} \quad g(\mu_{il}) = \beta_0(t) + \sum_{r=1}^{R} f_r(\boldsymbol{\mathcal{X}}_{ri}, t_l), \qquad i = 1, \ldots, n$$

- $y_i(t_l)$: Value of functional response observed at time point t_l
- $F(\mu_{il}, \boldsymbol{\nu})$: Conditional distribution of $y_i(t_l)$ with conditional expectation μ_{il} and shape parameters $\boldsymbol{\nu}$
- $g(\cdot)$: Link function
- $\beta_0(t)$: Functional intercept
- $f_r(\boldsymbol{\mathcal{X}}_{ri}, t_l)$: One of R additive effects with associated covariates $\boldsymbol{\mathcal{X}}_{ri}$ and potentially varying over the functional time domain t
- n: number of functional observations

We use a **highly performant estimation algorithm** from Wood et al. (2016) to make estimation of this complex model on such large data feasible. Major advances are:

- a block-wise Cholesky decomposition
- a compressed representation of marginal spline bases

A prediction error based approach was used for tuning basis sizes, resulting smooth effects were estimated using (tensor product) P-splines.

② Covariate effects

The hypocentral distance and the dynamic frictional resistance have by far the strongest effects, with higher values leading to decreased ground velocities for both.

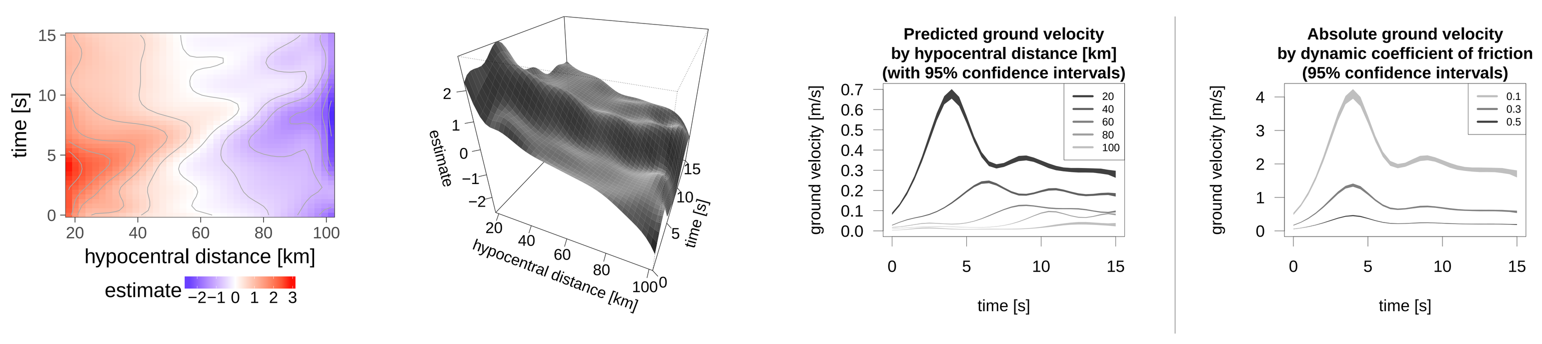

Figure 2: *Left: Nonlinear, time-varying effect of hypocentral distance as heatmap and 3D surface, and predictions based on varying hypocentral distances, while other covariates are held constant at realistic values. Right: Predictions based on varying values of the dynamic coefficient of friction, which has a linear, time-constant effect of -5.48*

③ Model evaluation

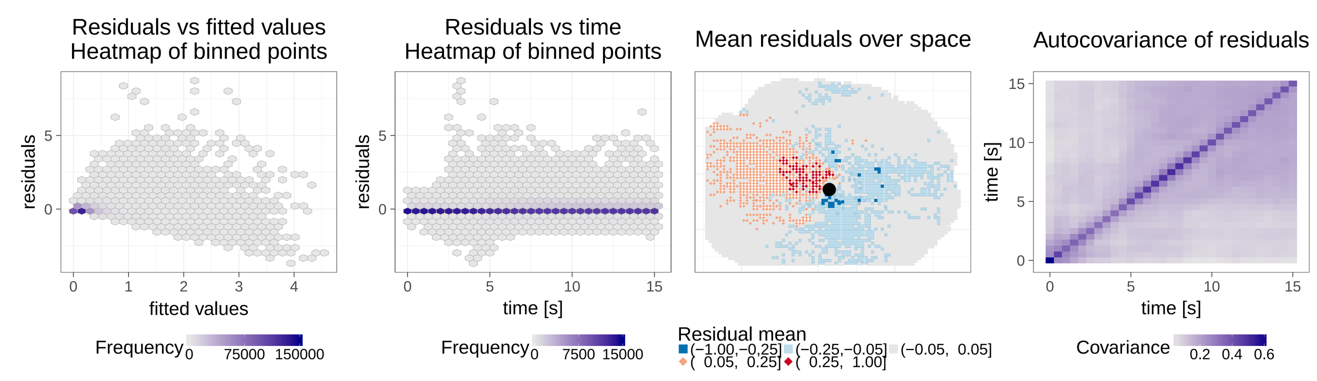

Figure 3: *From left to right: Residuals vs fitted values, residuals vs the time domain, residuals vs space, autocovariance of residuals over the functional domain. The black dot in the third plot marks the epicenter.*

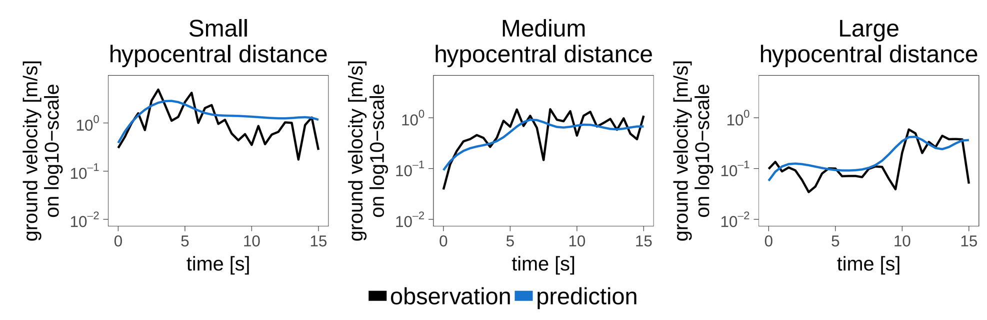

Figure 4: *Comparison of model predictions and raw observations for typical observations with different hypocentral distances.*

⇒ Spatial residual structure remaining
⇒ Predictions in general behave well (70.7% explained null deviance)

④ Conclusion & Outlook

Functional additive regression models are a promising approach for modeling surficial ground velocity.

Our model

- allows a better understanding of the observed seismological patterns
- adds value to the current seismological discussion of how important precise determination of specific physical parameters is
- offers predictions which could in future replace computer-intensive earthquake simulations

Secondary finding

Moment magnitude can be predicted very well using the simulation parameters (98.2% explained null deviance)

Future research

The model will be refined further, e.g. by explicitly modeling spatial correlation and by relaxing the strict assumption of the hypocenter as fixed point source for all earthquakes. Furthermore model performance will be examined for additional earthquakes.

References

Bauer, A. (2016). *Auswirkungen der Erdbebenquelldynamik auf den zeitlichen Verlauf der Bodenbewegung*. MA thesis. Ludwig-Maximilians-Universität, Munich, Germany. Available: https://epub.ub.uni-muenchen .de/31976/

Scheipl, F., Gertheiss, J., Greven, S. (2016). Generalized functional additive mixed models. *Electronic Journal of Statistics*, **10.1**, 1455–1492.

Weiss, A. (2001). Topographic position and landforms analysis. *Poster presentation, ESRI user conference*, San Diego, CA, **200**.

Wood, S.N. et al. (2016). Generalized additive models for gigadata: modelling the UK black smoke network daily data. *Journal of the American Statistical Association*. DOI: 10.1080/01621459.2016.1195744.