

# A FACTOR MODEL FOR THE ESTIMATION OF MULTIVARIATE EMPIRICAL GROUND-MOTION MODELS

NICOLAS KUEHN\*<sup>1</sup>

<sup>1</sup>*University of California, Los Angeles*

## Abstract

A Bayesian latent factor to estimate multivariate empirical ground-motion models is proposed. In the factor model, event terms, site terms, and within-event/within-site residuals for each response oscillator period are modeled as linear combinations of a small number of latent factors, thus introducing inter-period correlation in the model. The factor model provides a lower-dimensional approximation to the full residual correlation structure. The factor model is demonstrated on Italian ground-motion data. Results indicate that the factor model leads to similar behavior as a traditional ground-motion model where each target is estimated separately.

## 1 Introduction

Empirical ground-motion models (GMMs) are an indispensable tool for probabilistic seismic hazard analysis (PSHA). A GMM provides an estimate of the distribution of a ground-motion parameter (such as peak ground acceleration (PGA) or the response spectrum (PSA)), conditioned on a particular source/site scenario. Empirical GMMs are typically estimated via regression, which is usually carried out separately for each target variable. Correlations between response spectral periods are useful for ge.g. generating conditional mean spectra (*Baker, 2011*) or joint hazard (*Bazzurro and Cornell, 2002*). Correlation models are typically estimated from residuals (e.g. *Baker and Bradley, 2017*; *Baker and Jayaram, 2008*; *Bradley, 2011*; *Huang and Galasso, 2019*; *Kotha et al., 2017*) to empirical GMMs.

Some studies (e.g. *Arroyo and Ordaz, 2010a,b*; *Kuehn and Scherbaum, 2015*) have proposed methods to estimate multivariate GMMs, where all ground-motion parameters are regressed simultaneously together with their full covariance. Modeling the full spectrum during regression can be useful, such as when taking into account uncertainty in predictor variables (*Kuehn and Abrahamson, 2018*) or for truncated regression (*Kuehn et al., 2020b*) due to triggering.

Here, we model the residual structure in empirical GMMs with a factor model, which accounts for inter-period correlation by modeling the relation between periods via a small number of

---

\*[kuehn@ucla.edu](mailto:kuehn@ucla.edu)

latent factors. Factor models are widely used to model the covariance structure in multivariate data via dimensionality reduction, dating back to *Spearman (1904)* and *Thurstone (1935)*. In the next section, we describe the general model and how it can be used to model the covariance structure of event terms, site terms, and residuals in empirical GMMs. We then estimate the factor model GMM on Italian ground-motion data, and compare the results with a traditional GMM which is estimated separately for each target variable.

## 2 Factor Model GMM

The goal of an empirical GMM is to model the distribution of several ground-motion parameters of interest, given earthquake source and site related parameters such as magnitude, distance, or time-averaged shear wave velocity in the upper 30m,  $V_{S30}$ . In the following, let us assume a number of  $P$  ground-motion target variables. Typically, empirical GMMs are estimated separately for each target variable. Each target is usually modeled as lognormally distributed, which means that its logarithm is normally distributed:

$$\begin{aligned}\mu_p &= f_{GMM}(\vec{c}_p; \vec{x}) + \delta B_p + \delta S_p \\ y_p &\sim N(\mu_p, \phi_{SS,p})\end{aligned}\tag{1}$$

where  $p$  is an index for the target variable, and  $y_p$  is the logarithm of the  $p$ th target variable.  $\vec{x}$  is a vector of predictor variables (e.g. magnitude, distance,  $V_{S30}$ ),  $\vec{c}_p$  is a vector of coefficients for each target variable, and  $f_{GMM}(\vec{c}_p; \vec{x})$  describes the functional dependence of the target variable on the predictor variables.  $\delta B$  and  $\delta S$  are event and site term, respectively, and describe a constant offset for each record from individual events/stations (*Al-Atik et al., 2010*).  $\phi_{SS}$  is the standard deviation of the within-event/within-site aleatory variability. The event terms and site terms are typically modeled as independent random variables, normally distributed with mean zero and standard deviations  $\tau_p$  and  $\phi_{S2S,p}$ , respectively (*Al-Atik et al., 2010*).

One can write the model as a multivariate normal distribution

$$\vec{y} \sim MN(\vec{\mu}, \Phi_{SS})\tag{2}$$

where  $\Phi_{SS}$  is the covariance matrix between the different target variables. When estimating empirical GMMs, the models are typically estimated separately for each target variable, which means that  $\Phi_{SS}$  is a diagonal matrix with entries  $\Phi_{SS,pp} = \phi_{SS,p}^2$ . Similarly, event and site terms across periods can be modeled as multivariate normal distributions with a mean vector of zeros and diagonal covariance matrices  $T$  and  $\Phi_{S2S}$ .

One can relax the assumption of independence, and estimate the joint distribution of all target variables during model fitting. This requires to perform a regression for all target periods at the same time, with estimation of full covariance matrices  $\Phi_{SS}$ ,  $T$ , and  $\Phi_{S2S}$ . The truncated regression model proposed in *Kuehn et al. (2020b)* relies on modeling the joint distribution of different target variables.

Previous studies of multivariate GMM regressions (e.g. *Arroyo and Ordaz, 2010a,b*; *Kuehn and Abrahamson, 2018*; *Kuehn et al., 2020b*) modeled the covariance matrices as **full-rank**, which can be computationally expensive. Here, we propose a GMM factor model, where structure across periods is introduced by a linear combination of latent (unobserved) factors. The factors are of lower dimension than the targets.

In the following, we briefly describe the factor model, following the exposition given in [Papastamoulis and Ntzoufras \(2022\)](#). In a factor model, the target variables are modeled as a linear combination of the latent factors

$$\vec{y} = \vec{\mu} + \Lambda\vec{\eta} + \vec{\epsilon} \quad (3)$$

where  $\Lambda$  is a  $P \times D$  factor loadings matrix,  $\vec{\eta}$  are  $D$  latent factors, an  $\vec{\epsilon}$  are  $P$  independent errors, typically called *idiosyncratic errors*. For  $D < P$ , this is an approximation of the full problem with a lower-dimensional problem. The errors are independent from the factors, and independent normal random variables

$$\vec{\epsilon} \sim N(0, \Sigma) \quad (4)$$

where  $\Sigma$  is a diagonal matrix. The distribution of  $\vec{y}$  conditional on the latent factors is

$$\vec{y}|\vec{\eta} \sim N(\vec{\mu} + \Lambda\vec{\eta}, \Sigma) \quad (5)$$

and the marginal distribution is

$$\vec{y} \sim N(\vec{\mu}, \Lambda\Lambda^T + \Sigma) \quad (6)$$

The term  $\Lambda\Lambda^T$  describes the *shared variation* between targets ([Heaps, 2022](#)). Each factor is assumed to be independent and normally distributed with standard deviation one

$$\lambda_p \sim N(0, 1) \quad (7)$$

For the factor GMM, we model event terms, station terms, and within-event/within-station residuals across multiple periods with a factor model. For event terms and station terms, the model is

$$\delta\vec{B} = \Lambda_E\vec{\eta}_E \quad (8)$$

$$\delta\vec{S} = \Lambda_S\vec{\eta}_S$$

$\eta_E$  and  $\eta_S$  are the latent factors for events and stations, and  $\Lambda_E$  and  $\Lambda_S$  are the associated factor loadings matrices. We omit the idiosyncratic errors for event and site terms, since we found that their values were very small, and they hindered model estimation. The standard deviations of the event and site terms for each period can be calculated as

$$\vec{\tau} = \sqrt{\text{diag}(\Lambda_E\Lambda_E^T)} \quad (9)$$

$$\vec{\phi}_{S2S} = \sqrt{\text{diag}(\Lambda_S\Lambda_S^T)}$$

i.e. as the square-root of the diagonal of the shared variation matrix.

The within-event/within-site residuals are modeled as

$$\delta\vec{W}S = \Lambda_R\vec{\eta}_R + \vec{\epsilon}_R \quad (10)$$

where subscript  $R$  stands for record, and the other terms are defined as before. The idiosyncratic errors  $\vec{\epsilon}$  are each normally distributed with mean zero and standard deviation  $\phi_{0,p}$ . Thus the total within-event standard deviation can be calculated as

$$\vec{\phi}_{SS} = \sqrt{\vec{\phi}_\lambda^2 + \vec{\phi}_0^2} \quad (11)$$

$$\vec{\phi}_\lambda^2 = \text{diag}(\Lambda_R\Lambda_R^T)$$

## 2.1 Implementation

The parameters of the models are estimated with Bayesian inference ([Gelman et al., 2013](#)) via Markov Chain Monte Carlo (MCMC) sampling. The models are implemented in the probabilistic programming language Stan ([Carpenter et al., 2017](#); [Stan Development Team, 2022](#)) and run from the computing environment R ([R Core Team, 2022](#)) using the package `cmdstanr` ([Gabry and Češnovar, 2021](#)).

Bayesian inference of factor models with MCMC is not straightforward ([Papastamoulis and Ntzoufras, 2022](#)). The factor model is not uniquely identified, as any rotation of the factor loadings matrix and the factors leads to the same model. To avoid problems with rotation, the factor loadings matrix is often constrained such that  $\lambda_{ij} = 0$  for  $j > i$  and  $\lambda_{ii} > 0$  ([Geweke and Zhou, 1996](#); [Lopes, 2014](#); [Lopes and West, 2004](#)); thus, the upper triangle is zero. Here, we order the target variables from short periods to long periods, and constrain the factor loadings matrix such that the lower triangle is zero. The reason is that at longer periods, there is less data due to filtering, which makes the association to the factor loadings harder to detect.

We illustrate the model using a slightly modified version of the model and data of the ITA18 GMM [Lanzano et al. \(2019a\)](#). The model is simple (linear in coefficients), and the number of data points is not too high. We illustrate the model using 18 periods (mainly a computational reason) and eight factors. For comparison, we also fit a Bayesian model where each target variable is estimated separately. We run two chains with 500 warm-up samples and 500 estimation samples each, resulting in 1000 posterior samples.

## 3 ITA18 Model and Data

The ITA18 GMM ([Lanzano et al., 2019a](#)) is a model developed for Italian ground-motion data. The data is taken from the Engineering Strong Motion (ESM) flatfile [Lanzano et al. \(2019b\)](#), enhanced by some small regional Italian records from the Italian Accelerometric Archive (ITACA) database ([Luzi et al., 2008, 2017](#); [Pacor et al., 2011](#)), as well as 12 worldwide events in the magnitude range 6.1 to 8. In total, ITA18 is based on 5607 records from 146 events and 1657 stations.

Here, we use the subset of ITA18 from the ESM data base, which consists of 4541 Italian records from 137 events and 886 stations. The magnitude-distance scatterplot is shown in [Figure 1](#).

The functional form of ITA18 relating the target variable to the predictors is

$$\begin{aligned}
 y = & a + b_1(M_W - M_h) 1_{(M_w \leq M_h)} + b_2(M_W - M_h) 1_{(M_w > M_h)} \\
 & + [c_2 + c_1(M_W - M_h)] \log_{10} \sqrt{R_{JB}^2 + h^2} + c_3 \sqrt{R_{JB}^2 + h^2} \\
 & + k \left[ \log_{10} \frac{V_{S30}}{800} 1_{(V_{S30} \leq 1500)} + \frac{1500}{800} 1_{(V_{S30} > 1500)} \right] \\
 & + \delta B + \delta S + \delta WS
 \end{aligned} \tag{12}$$

where  $y$  is the target variable (in log10 units), and  $a$ ,  $b_1$ ,  $b_2$ ,  $c_1$ ,  $c_2$ ,  $c_3$ , and  $k$  are coefficients

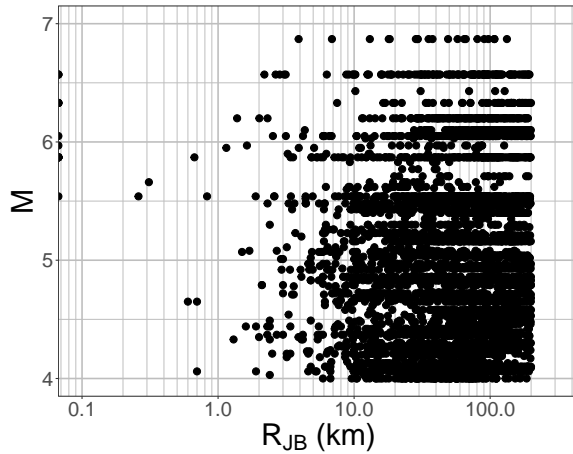


Figure 1: Magnitude-distance scatterplot of the data.

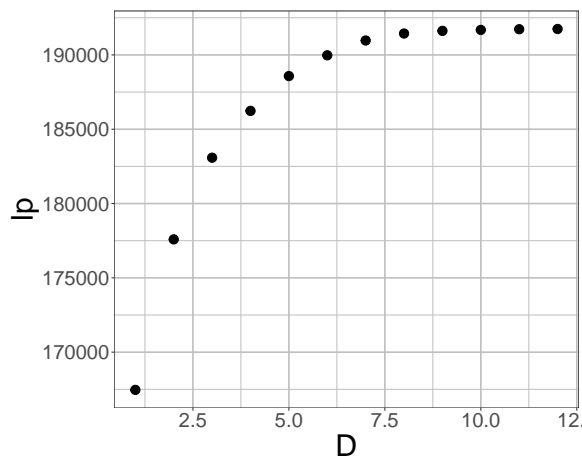


Figure 2: MAP log density versus number of factors, estimated for within-event/within-station residuals.

that are estimated during the regression. The predictor variables are moment magnitude  $M_W$ , Joyner-Boore distance  $R_{JB}$ ,  $V_{S30}$ . The original ITA18 model includes two coefficients describing the dependence of  $y$  on fault type; we omitted these parameters  $M_h$ ,  $M_{ref}$ , and  $h$  are parameters that are fixed before model estimation.

ITA18 was developed for PGA, PGV, and 36 response spectral periods from  $T = 0.01s$  to  $T = 10.s$ . In this work, we estimate the different models for PGA and periods from 17 periods from  $T = 0.01 - 3$ , i.e.  $P = 18$ . A longer periods, data drops out due to filtering. We use  $D = 8$  for the number of latent factors. The number is chosen as follows: we take the within-event/within-station residuals of the separate model, and estimate a simple factor model like Equation (3) via maximum-a-posteriori estimation for different numbers of factors. Figure 2 shows the value of the maximised log-density ( $lp$ ) versus the number of factors. At about  $D = 8$ , the log-density reaches a plateau.

For the factor model, we also provide reasonable starting values for the MCMC algorithm, as this can to avoid label switching in the entries of the factor loadings matrix ([Deschamps, 2018](#)). For the coefficients, the starting values are taken from the posterior distribution of the separate model, while starting values for the the factor loadings are generated by randomizing the values obtained from optimization based on within-event/within-station residuals. Starting values for the latent factors are generated by sampling from the prior (a standard normal distribution).

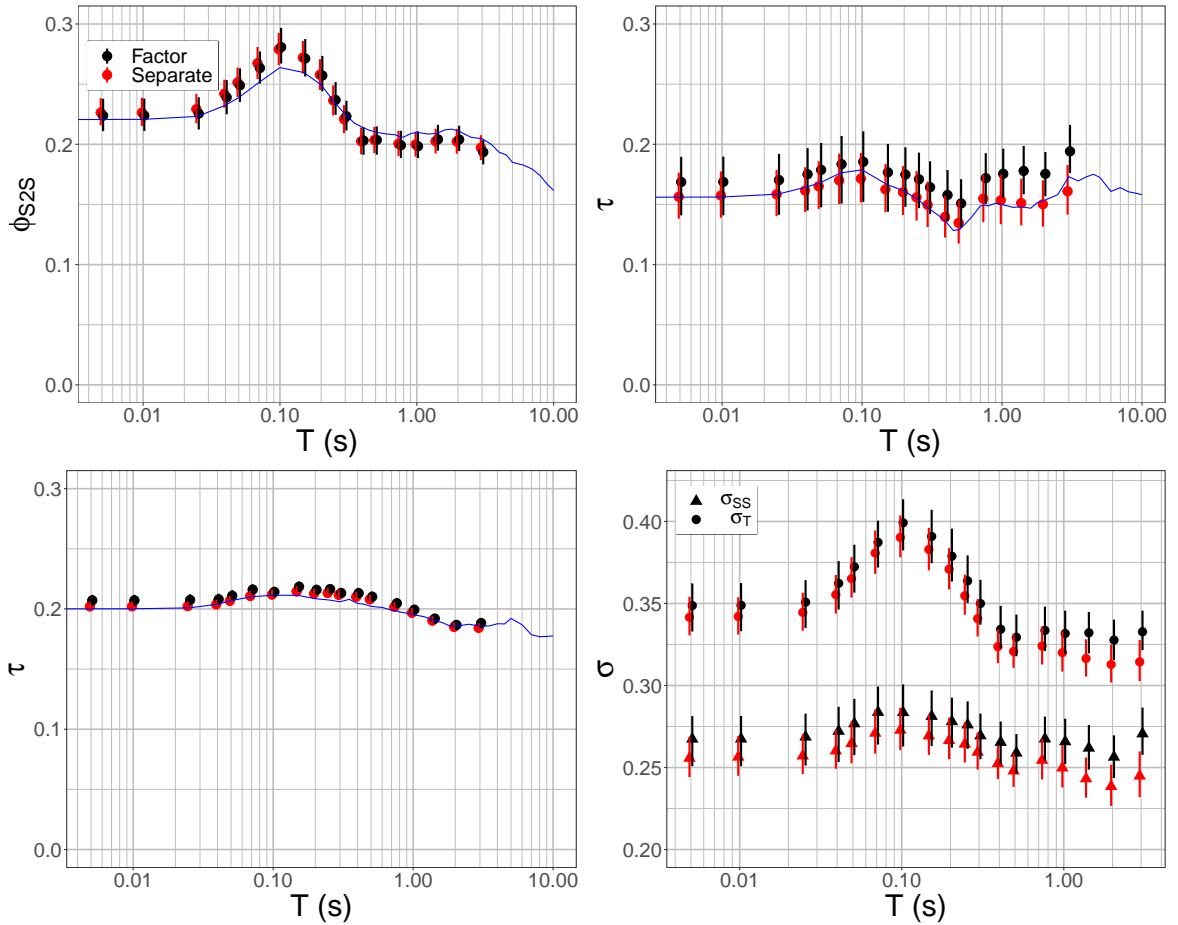


Figure 3: Estimated standard deviations vs period. Plotted is the mean of the posterior distribution, error bars indicate 5% to 95% intervals of the posterior. The periods have been slightly moved to avoid plotting points at the same place. The blue line shows the value of ITA18.

## 4 Comparison between Factor and Separate Model

In this section, we show comparisons of the estimated factor model GMM as well as the model where each target variable is estimated separately. For the factor model, there are some warnings that the tree-depth of the NUTS algorithm (*Hoffman and Gelman, 2014*) has reached the maximum limit, and some values of the  $\hat{R}$ -statistic (*Vehtari et al., 2021*) are larger than 1.05 for some parameters. Thus, the results for the factor model should be interpreted with care, but we believe that the estimated values are reasonable.

Figure 3 shows the estimated standard deviations  $\phi_{S2S}$ ,  $\tau$ ,  $\phi$ . In addition, we show total standard deviation  $\sigma_T^2 = \phi_{S2S}^2 + \tau^2 + \phi_{SS}^2$ , and single-station standard deviation  $\sigma_{SS}^2 = \tau^2 + \phi_{SS}^2$ . In general, these are very similar between the models. The factor model leads to some slightly larger values. Due to the modeled correlation, the individual event/site terms and residuals have to obey some additional constraints.

Figure 4 shows the values of the components of the within-event standard deviation (cf. Equation (11)) for the full factor model. Most of the within-event/within-station variability is explained by the shared variation, with a larger share of idiosyncratic errors at longer periods.

The estimated coefficients are shown in Figure 5. In general, the coefficients are very

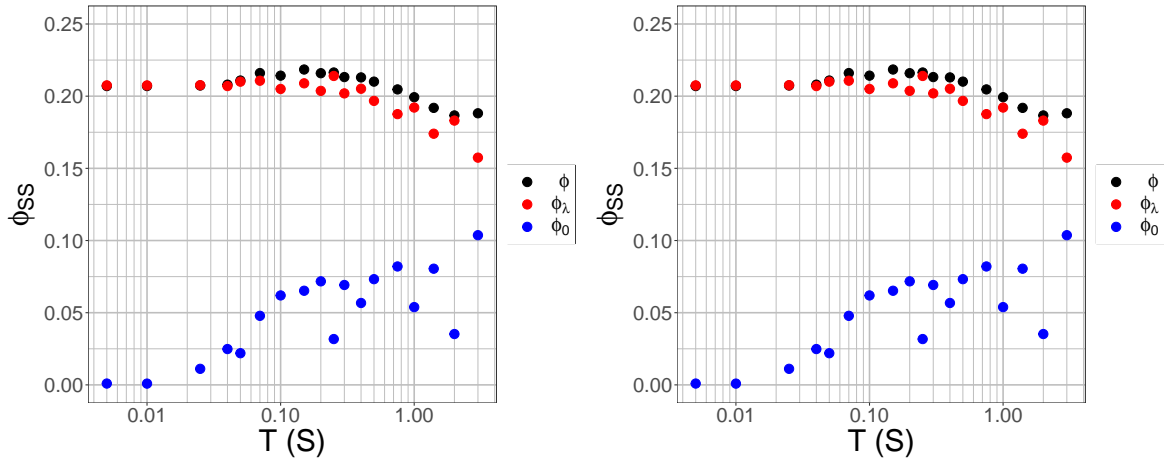


Figure 4:  $\phi_{SS}$  and its components for the factor model.

similar between models. The coefficient  $b_2$ , which controls the large magnitude scaling, shows the largest difference. The reason for this is probably that there are not many large events, and the modeled correlation between event terms enforces some consistent behavior across periods.

Figure 6 shows the empirical correlation between the posterior samples of some periods across periods, i.e. we calculate the correlation at two periods  $T_1$  and  $T_2$

$$\rho(a_{T_1}, a_{T_2}) \quad (13)$$

in this case for the intercept  $a$ . Figure 6 shows the correlations with respect to  $T = 0.5s$ , for the intercept  $a$ , the large magnitude scaling  $b_2$ , and the anelastic attenuation  $c_3$ . While the priors for each coefficient are independent for both the separate and factor model, there is strong correlation between the coefficients at different periods in the factor model. By contrast, the correlation is (expected) close to zero for the separate model. This would have a consequence if one calculates joint predictions for several periods and wants to take the epistemic uncertainty in the correlations into account.

Figure 7 shows correlations between different periods, calculated from the shared variation  $\Lambda\Lambda^T$ , for event terms, site terms, and within-event/within-site residuals. For comparison, we also calculate empirical correlations from the estimated terms for the separate model, and show the correlation model of *Baker and Jayaram (2008)* (note that this model is for correlation between total residuals). The correlations are all quite similar, which is in line with previous research that has found that period-o-period correlations of the response spectrum are stable across different regions and ranges of predictor variables (*Baker and Bradley, 2017*).

## 5 Discussion and Conclusions

We have presented a factor model GMM, to estimate empirical GMMs for multiple target variables at the same time, while also accounting for inter-period correlation in event terms, site terms, and within-event/within-site residuals. The factor model GMM is a low-rank approximation to a full covariance matrix, and thus models the residual structure using a lower-dimensional space.

In general, results from the factor model GMM and a separate model (i.e. all targets are estimated without accounting for correlation) are similar, both in terms of the estimated coeffi-

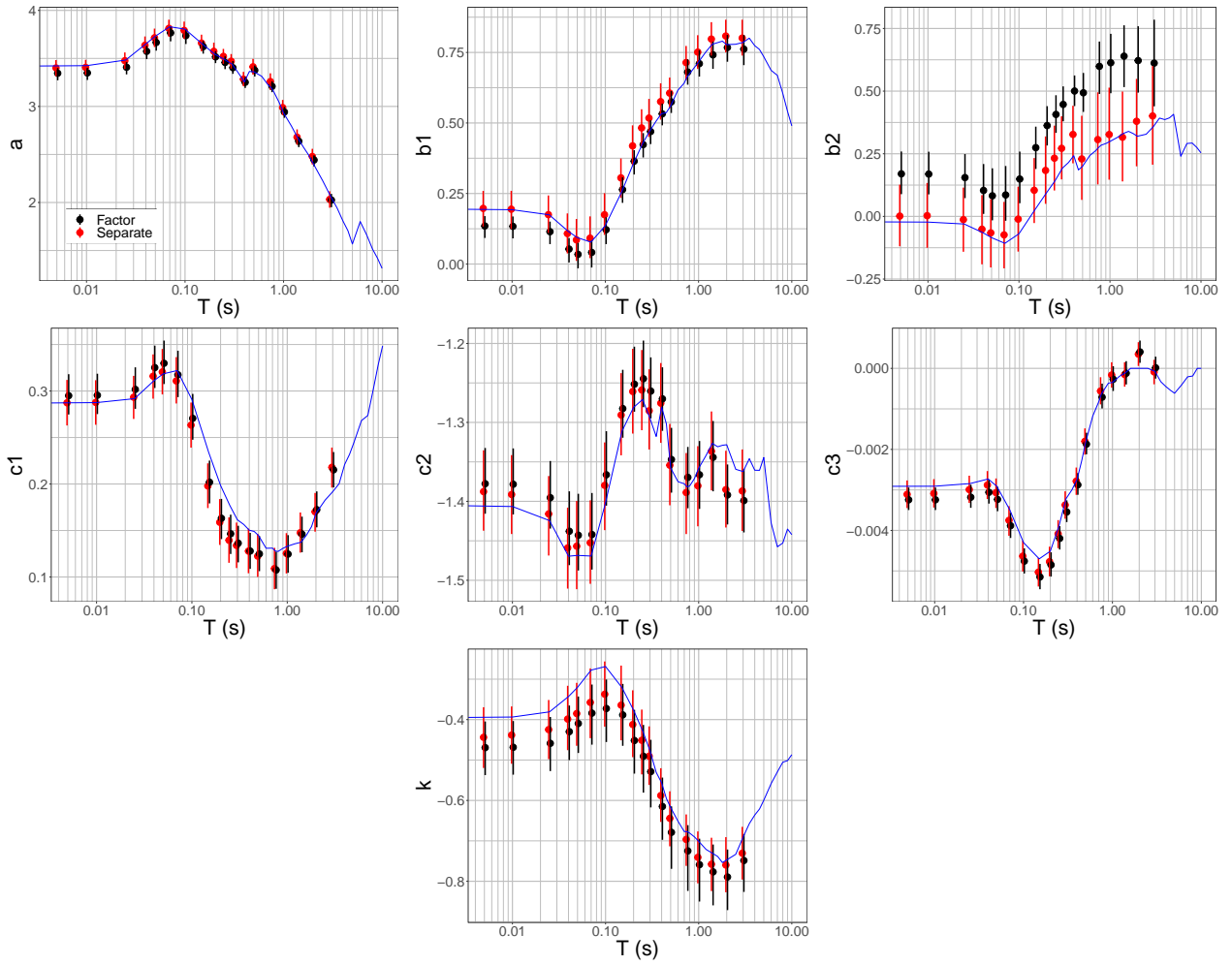


Figure 5: Estimated coefficients vs. period. Dots show the mean of the posterior distribution, the uncertainty interval comprises the 5% and 95% quantiles of the posterior distribution. The blue line shows the coefficients of ITA18.

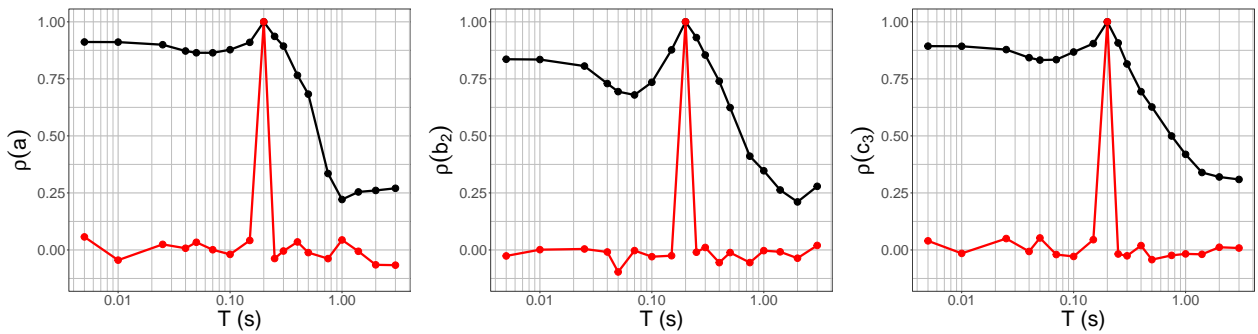


Figure 6: Empirical correlation between constant ( $a$ ), large magnitude scaling ( $b_2$ ), and anelastic attenuation ( $c_3$ ) coefficients, conditioned on  $T = 0.2s$ , estimated from the posterior samples.

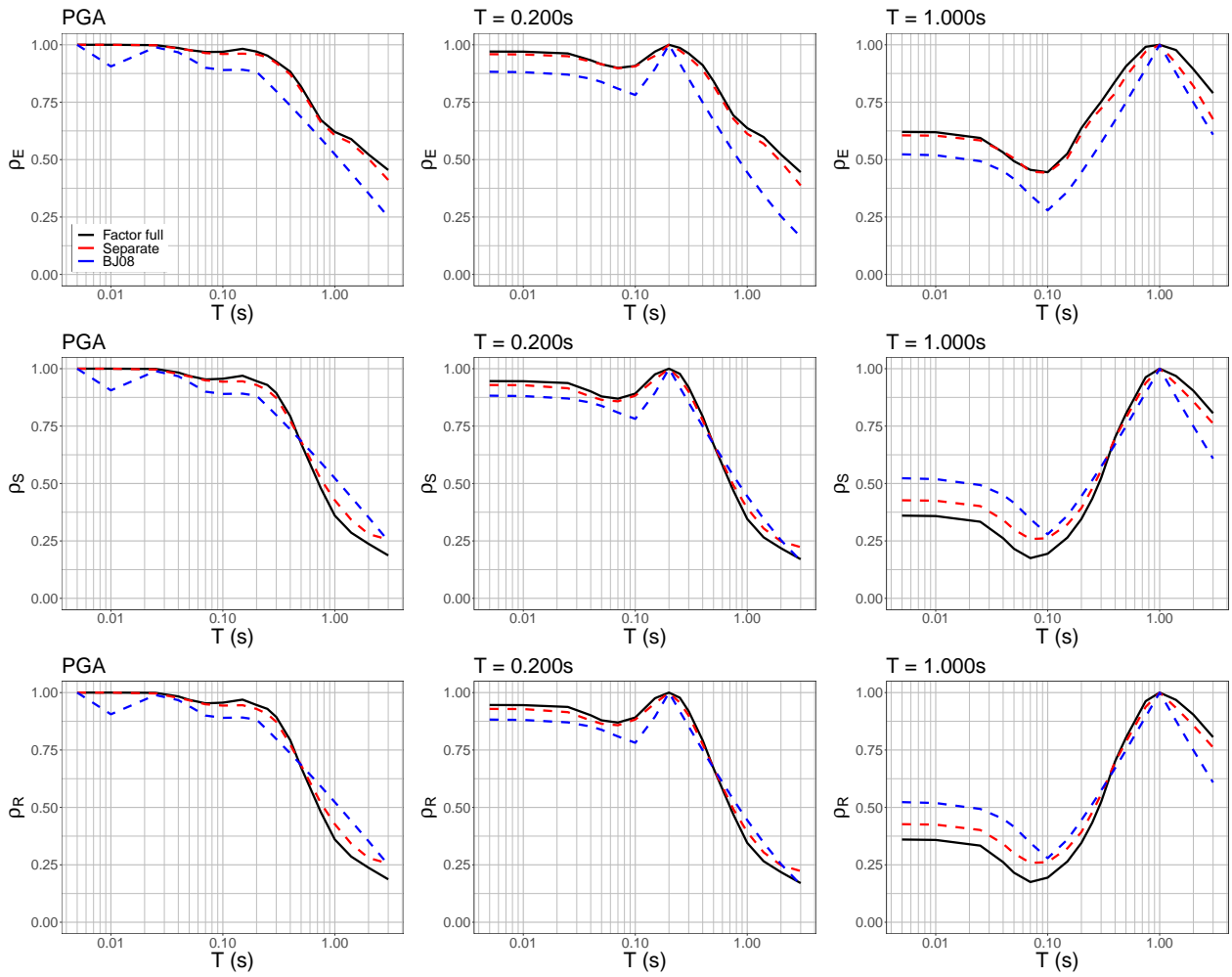


Figure 7: Estimated correlations, and comparisons with model of *Baker and Jayaram (2008)*. Dashed indicates correlations are estimated from residuals.

coefficients and standard deviations. This raises the question whether it is useful to account for the inter-period correlation structure during regression. While results are quite similar, there are use cases, such as truncated regression (*Kuehn et al., 2020b*), or accounting for uncertainty in predictor variables (*Kuehn and Abrahamson, 2018*), which require modeling the full spectrum.

In the factor model GMM, we account for correlations in the residuals, while the coefficients are modeled separately. The factor model induces correlation in the posterior distribution of the coefficients (cf, Figure 6), but no structure or smoothing is applied for the behavior of the periods across periods. One way forward could be to impose a smooth relation of the coefficients. *Bortolotti et al. (2022)* describe the use of functional data analysis to obtain smooth coefficients. One could also model the functional dependency of the coefficients on periods using a Gaussian process, as in *Kuehn et al. (2020a)*.

The factor model requires choosing the number of latent factors. Often, the maximum number of factors is taken to be the Lederman bound

$$\phi(P) = \frac{2P + 1 - \sqrt{(8P + 1)}}{s} \quad (14)$$

since  $\Sigma$  (the covariance matrix of the idiosyncratic errors) is globally identifiable when  $P < \phi(P)$  (*Bekker and Ten Berge, 1997; Heaps, 2022; Papastamoulis and Ntzoufras, 2022*). While there are methods proposed to identify the number of latent factors during model estimation (*Fruehwirth-Schnatter and Lopes, 2018; Heaps, 2022*), here we fixed the number based on the maximum estimated log density when optimizing for a smaller problem (cf. Figure 2) for simplicity.

The components of variability and uncertainty in GMMs ( $\tau, \phi_{S2S}, \phi_{SS}$ ) are sometimes modeled as dependent predictor variables; in particular the between-event variability  $\tau$  is often assumed to decrease for large magnitudes. In the factor model, the value of  $\tau$  is the square root of the diagonal of  $\Lambda_E \Lambda_E^T$ . Thus, dependence of  $\tau$  on magnitude needs to be modeled through dependence on the factor loadings matrix, which would lead to a magnitude-dependent correlation structure. Alternatively, one could model the distribution of the latent factors as magnitude dependent.

Here, we have used the factor model to estimate an empirical GMM, but there are other applications in the context of ground-motion modeling and hazard/risk analysis. The latent factors for events and sites ( $\eta_E$  and  $\eta_S$ ) are assumed to be independent, but they could be modeled as spatially correlated (*Qiu et al., 2018; Thorson et al., 2015a,b*). This opens an avenue for modeling multi-variate nonergodic GMMs (*Lavrentiadis et al., 2022*). Similarly, a spatial factor model could be used for modeling cross-correlations of ground-motion parameters, which is useful for regional seismic risk analysis (.g. *Miller and Baker, 2015*). Such a model provides a lower dimensional approximation to the full cross-covariance, which is similar in spirit to the model of *Markhvida et al. (2018)*.

In general, we believe that the factor model provides a convenient tool for modeling the full response spectrum in empirical GMMs in case this is needed.

## 6 Data and Code

Data is available from (*Lanzano et al., 2019b*) and *Lanzano et al. (2019a)*. Stan and R code to run the factor and separate models is available at <https://github.com/nikuehn/>

## Acknowledgements

The regressions are carried out in R (*R Core Team*, 2022) using package `cmdstanr` (*Gabry and Češnovar*, 2021). Plots are made with `ggplot2` (*Wickham*, 2016), and data organization is done using `tidyverse` (*Wickham et al.*, 2019) and `posterior` (*Bürkner et al.*, 2022).

Partial support of Pacific Gas & Electric Company and California Department of Transportation are gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect those of the sponsors.

## References

- Al-Atik, L., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn, The Variability of Ground-Motion Prediction Models and Its Components, *Seismological Research Letters*, 81(5), 794–801, doi:10.1785/gssrl.81.5.794, 2010.
- Arroyo, D., and M. Ordaz, Multivariate Bayesian Regression Analysis Applied to Ground-Motion Prediction Equations, Part 1: Theory and Synthetic Example, *Bulletin of the Seismological Society of America*, 100(4), 1551–1567, doi:10.1785/0120080354, 2010a.
- Arroyo, D., and M. Ordaz, Multivariate Bayesian Regression Analysis Applied to Ground-Motion Prediction Equations, Part 2: Numerical Example with Actual Data, *Bulletin of the Seismological Society of America*, 100(4), 1568–1577, doi:10.1785/0120090320, 2010b.
- Baker, J. W., Conditional Mean Spectrum: Tool for Ground-Motion Selection, *Journal of Structural Engineering*, 137(3), 322–331, doi:10.1061/(ASCE)ST.1943-541X.0000215, 2011.
- Baker, J. W., and B. A. Bradley, Intensity Measure Correlations Observed in the NGA-West2 Database, and Dependence of Correlations on Rupture and Site Parameters, *Earthquake Spectra*, 33(1), 145–156, doi:10.1193/060716eqs095m, 2017.
- Baker, J. W., and N. Jayaram, Correlation of Spectral Acceleration Values from NGA Ground Motion Models, *Earthquake Spectra*, 24(1), 299, doi:10.1193/1.2857544, 2008.
- Bazzurro, P., and C. Cornell, Vector-valued probabilistic seismic hazard analysis (VPSHA), in *Proceedings of the 7th US national conference on Earthquake Engineering*, 2002.
- Bekker, P. A., and J. M. Ten Berge, Generic global Identification in factor analysis, *Linear Algebra and Its Applications*, 264(1-3), 255–263, doi:10.1016/S0024-3795(96)00363-1, 1997.
- Bortolotti, T., R. Peli, G. Lanzano, and S. Sgobba, Weighted functional data analysis for the calibration of ground motion models in Italy Weighted functional data analysis for the calibration of ground motion models in Italy, *Tech. Rep. 31*, MOX-Report No. 31/2022, 2022.
- Bradley, B. A., Empirical correlation of PGA, spectral accelerations and spectrum intensities from active shallow crustal earthquakes, *Earthquake Engineering & Structural Dynamics*, 40(15), 1707–1721, doi:10.1002/eqe.1110, 2011.
- Bürkner, P.-C., J. Gabry, M. Kay, and A. Vehtari, posterior: Tools for working with posterior distributions, r package version 1.3.1, 2022.
- Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, Stan : A Probabilistic Programming Language, *Journal of Statistical Software*, 76(1), 1–32, doi:10.18637/jss.v076.i01, 2017.
- Deschamps, L., Fitting Generalized Bayesian Factor Models with stan, (June), 1–17, doi:10.13140/RG.2.2.36670.59208/1, 2018.
- Fruehwirth-Schnatter, S., and H. F. Lopes, Sparse Bayesian Factor Analysis when the Number of Factors is Unknown, *arXiv*, doi:10.48550/arXiv.1804.04231, 2018.
- Gabry, J., and R. Češnovar, *cmdstanr: R Interface to 'CmdStan'*, <https://mc-stan.org/cmdstanr>, <https://discourse.mc-stan.org>, 2021.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, February, 1–646 pp., Chapman and Hall/CRC, doi:10.1201/b16018, 2013.

- Geweke, J., and G. Zhou, Measuring the Pricing Error of the Arbitrage Pricing Theory, *Review of Financial Studies*, 9(2), 557–587, doi:10.1093/rfs/9.2.557, 1996.
- Heaps, S. E., Structured prior distributions for the covariance matrix in latent factor models, *arXiv*, pp. 1–17, 2022.
- Hoffman, M. D., and A. Gelman, The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, *Journal of Machine Learning Research*, 15, 1593–1623, 2014.
- Huang, C., and C. Galasso, Ground-motion intensity measure correlations observed in Italian strong-motion records, *Earthquake Engineering and Structural Dynamics*, 48(15), 1634–1660, doi:10.1002/eqe.3216, 2019.
- Kotha, S. R., D. Bindi, and F. Cotton, Site-Corrected Magnitude- and Region-Dependent Correlations of Horizontal Peak Spectral Amplitudes, *Earthquake Spectra*, 33(4), 1415–1432, doi:10.1193/091416eqs150m, 2017.
- Kuehn, N., Y. Bozorgnia, K. W. Campbell, and N. Gregor, Partially Non-Ergodic Ground-Motion Model for Subduction Regions using the NGA-Subduction Database, *Tech. Rep. August*, Pacific Earthquake Engineering Research Center, 2020a.
- Kuehn, N. M., and N. A. Abrahamson, The Effect of Uncertainty in Predictor Variables on the Estimation of Ground-Motion Prediction Equations, *Bulletin of the Seismological Society of America*, 108(1), 358–370, doi:10.1785/0120170166, 2018.
- Kuehn, N. M., and F. Scherbaum, Ground-motion prediction model building: a multilevel approach, *Bulletin of Earthquake Engineering*, 13(9), 2481–2491, doi:10.1007/s10518-015-9732-3, 2015.
- Kuehn, N. M., T. Kishida, M. AlHamaydeh, G. Lavrentiadis, and Y. Bozorgnia, A Bayesian model for truncated regression for the estimation of empirical ground-motion models, *Bulletin of Earthquake Engineering*, 18(14), 6149–6179, doi:10.1007/s10518-020-00943-8, 2020b.
- Lanzano, G., L. Luzi, F. Pacor, C. Felicetta, R. Puglia, S. Sgobba, and M. D’Amico, A Revised Ground-Motion Prediction Model for Shallow Crustal Earthquakes in Italy, *Bulletin of the Seismological Society of America*, 109(2), 525–540, doi:10.1785/0120180210, 2019a.
- Lanzano, G., S. Sgobba, L. Luzi, R. Puglia, F. Pacor, C. Felicetta, M. D’Amico, F. Cotton, and D. Bindi, The pan-European Engineering Strong Motion (ESM) flatfile: compilation criteria and data statistics, *Bulletin of Earthquake Engineering*, 17(2), 561–582, doi:10.1007/s10518-018-0480-z, 2019b.
- Lavrentiadis, G., N. A. Abrahamson, K. M. Nicolas, Y. Bozorgnia, C. A. Goulet, A. Babič, J. Macedo, M. Dolšek, N. Gregor, A. R. Kottke, M. Lacour, C. Liu, X. Meng, V.-B. Phung, C.-H. Sung, and M. Walling, Overview and introduction to development of non-ergodic earthquake ground-motion models, *Bulletin of Earthquake Engineering*, doi: 10.1007/s10518-022-01485-x, 2022.
- Lopes, H. F., Modern Bayesian Factor Analysis, in *Bayesian Inference in the Social Sciences*, 1904, pp. 115–153, doi:10.1002/9781118771051.ch5, 2014.
- Lopes, H. F., and M. West, Bayesian model assessment in factor analysis, *Statistica Sinica*, 14(1), 41–67, 2004.

- Luzi, L., S. Hailemichael, D. Bindi, F. Pacor, F. Mele, and F. Sabetta, ITACA (Italian Accelerometric Archive) : A web portal for the dissemination of Italian strong-motion data, *Seismological Research Letters*, 79(5), 716–722, doi:10.1785/gssrl.79.5.716, 2008.
- Luzi, L., F. Pacor, and R. Puglia, Italian Accelerometric Archive, doi:10.13127/ITACA.2.3, 2017.
- Markhvida, M., L. Ceferino, and J. W. Baker, Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics, *Earthquake Engineering and Structural Dynamics*, 47(5), 1107–1123, doi:10.1002/eqe.3007, 2018.
- Miller, M., and J. Baker, Ground-motion intensity and damage map selection for probabilistic infrastructure network risk assessment using optimization, *Earthquake Engineering & Structural Dynamics*, 44(7), 1139–1156, doi:10.1002/eqe.2506, 2015.
- Pacor, F., R. Paolucci, G. Ameri, M. Massa, and R. Puglia, Italian strong motion records in ITACA: Overview and record processing, *Bulletin of Earthquake Engineering*, 9(6), 1741–1759, doi:10.1007/s10518-011-9295-x, 2011.
- Papastamoulis, P., and I. Ntzoufras, On the identifiability of Bayesian factor analytic models, *Statistics and Computing*, 32(2), 23, doi:10.1007/s11222-022-10084-4, 2022.
- Qiu, Q., J. Sung, W. Davis, and R. Tchernis, Using spatial factor analysis to measure human development, *Journal of Development Economics*, 132(December 2017), 130–149, doi:10.1016/j.jdeveco.2017.12.007, 2018.
- R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.
- Spearman, C., "General Intelligence," Objectively Determined and Measured, *The American Journal of Psychology*, 15(2), 201–292, 1904.
- Stan Development Team, Stan modeling language users guide and reference manual, version 2.31.0, 2022.
- Thorson, J. T., M. D. Scheuerell, A. O. Shelton, K. E. See, H. J. Skaug, and K. Kristensen, Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range, *Methods in Ecology and Evolution*, 6(6), 627–637, doi:10.1111/2041-210X.12359, 2015a.
- Thorson, J. T., H. J. Skaug, K. Kristensen, A. O. Shelton, E. J. Ward, J. H. Harms, J. A. Benante, and B. D. Inouye, The importance of spatial models for estimating the strength of density dependence, *Ecology*, 96(5), 1202–1212, doi:10.1890/14-0739.1, 2015b.
- Thurstone, L. L., *The Vectors of Mind*, The University of Chicago Press, Chicago, Illinois, 1935.
- Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner, Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC (with Discussion), *Bayesian Analysis*, 16(2), 1–23, doi:10.1214/20-BA1221, 2021.
- Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York, 2016.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Golemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, Welcome to the tidyverse, *Journal of Open Source Software*, 4(43), 1686, doi:10.21105/joss.01686, 2019.