Ground motion models for Italian seismic data

Tutored by Teresa Bortolotti

POLITECNICO MILANO 1863

Erica Casassa – Martina Fervari – M. Giovanna Latorraca – R. Elias Roux – Edoardo Vitale

Index

Data Exploration

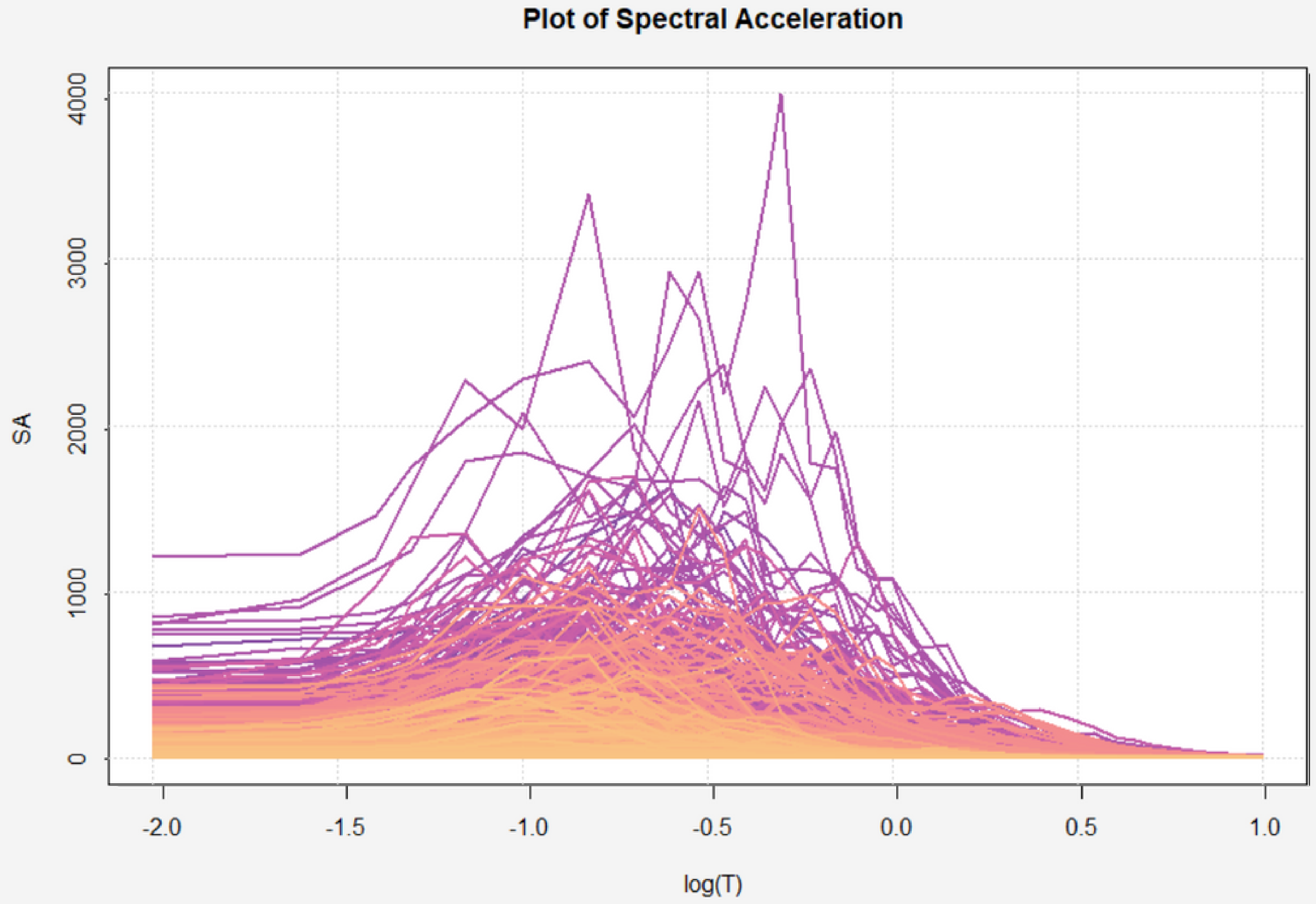
event_id	event_lat	event_lon	station_id	station_lat	station_lon	magnitude	distance	sof	vs30	SA_0	…		SA_9	SA_10	ev_nation_code
EMSC-20080613_0000091	39.02980	140.8807	AKT18	39.81470	140.5790	6.9	73.43	TF	475.0000	44.4896		380850	2.2439700	1.7926500	JP
EMSC-20080613_0000091	39.02980	140.8807	AKT19	39.66340	140.5721	6.9	58.68	TF	620.0000	70.4333		079600	3.2924150	2.5976700	JP
EMSC-20080613_0000091	39.02980	140.8807	AKT1A	39.22230	140.1283	6.9	60.43	TF	320.0000	74.9125		060950	3.0667800	2.5171700	JP
EMSC-20080613_0000091	39.02980	140.8807	AKT1C	39.07180	140.3185	6.9	38.02	TF	829.0000	77.6330		309400	1.7523900	1.2728350	JP
EMSC-20080613_0000091	39.02980	140.8807	AKT1D	38.98010	140.4952	6.9	19.72	TF	455.0000	178.4850		375100	2.8700000	3.1730200	JP

The flat-file of 'ITA18' dataset includes **5778 records**, relative to **146 earthquakes** and **1657 stations**, mostly registered in Italy. The most important features for our analysis are:

- <u>Magnitude</u>
- Joyner-Boore distance (<u>Distance</u>) defined as the distance to the projection on the surface of the rupture plane of the fault
- Style of Faulting (<u>SoF</u>)
- Speed of shear waves at a depth of 30 meters (<u>VS30</u>)
- Spectral Acceleration (<u>SA</u>), registerd at different periods (our target)

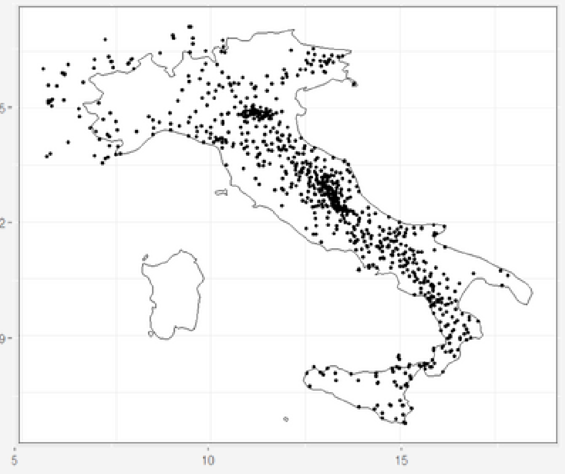
Data Exploration

<u>**GOAL:**</u> Predict **SA** at unseen locations taking into account **spatial correlation.**

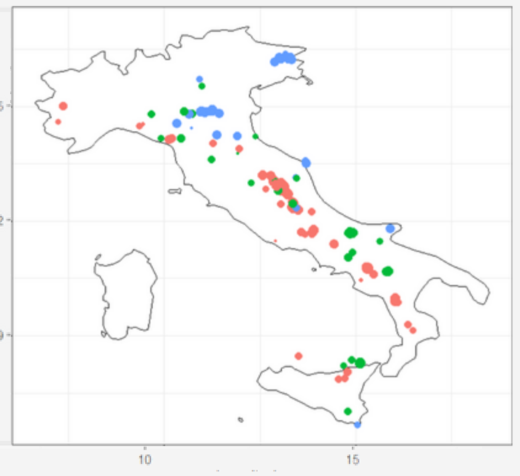


The **Spectral Acceleration** (SA) is a unit measured in g (the acceleration due to Earth's gravity) that describes the <u>maximum acceleration in an earthquake on an object</u>.

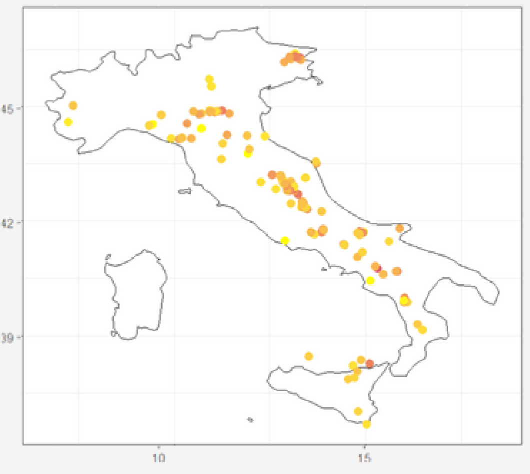
Geographical distribution of the stations



Magnitude according to type of faulting

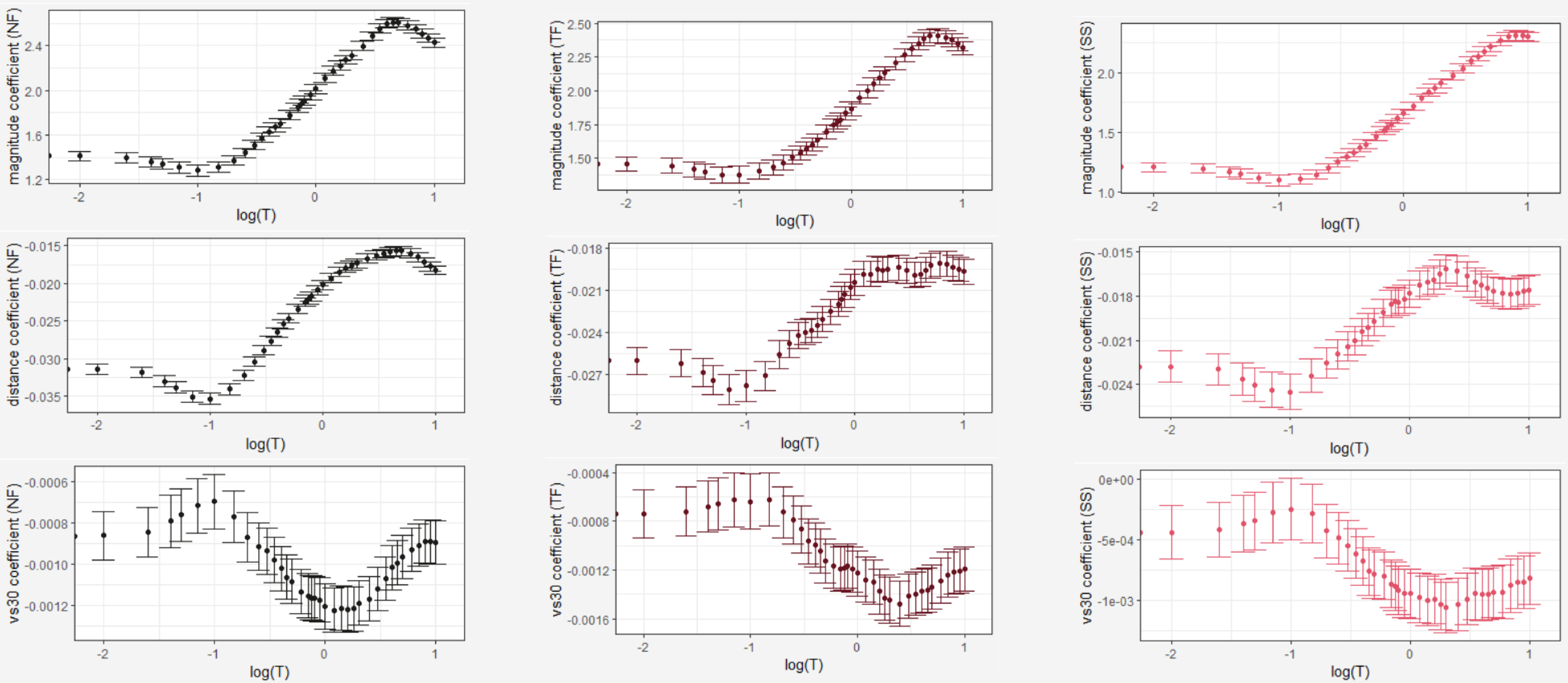


Magnitude of the Italian events



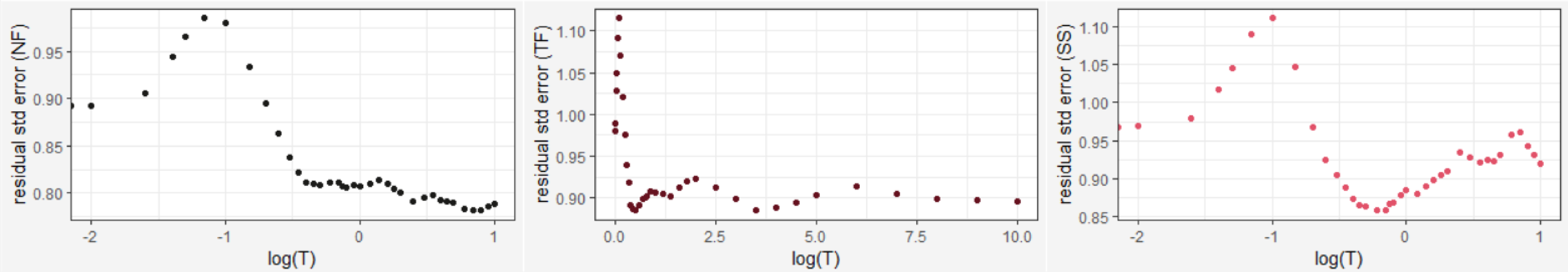
First Linear Regression Model

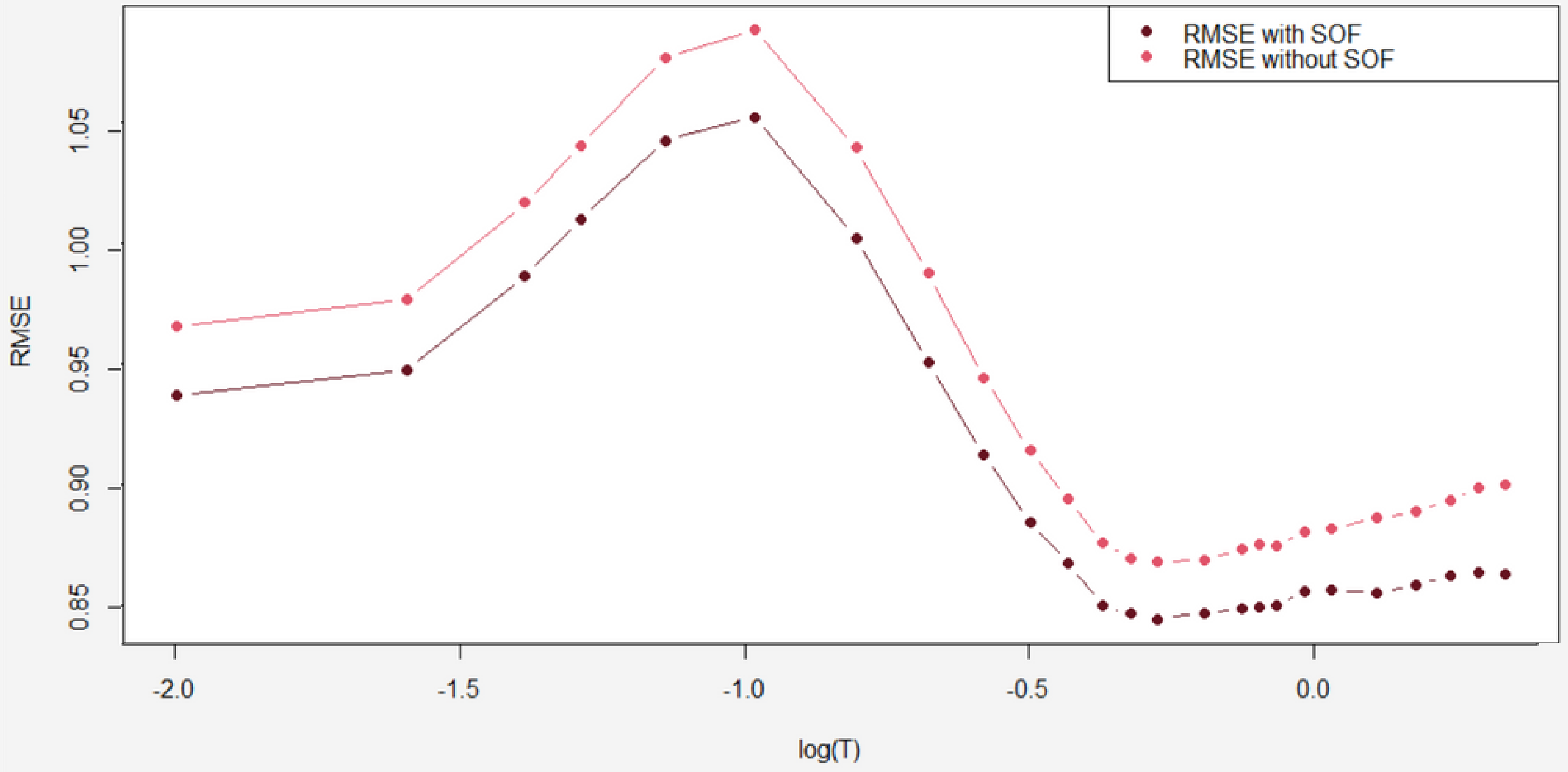
$$\log(SA_T) = \beta_0 + \beta_{magnitude} * M_w + \beta_{distance} * D + \beta_{vs30} * VS_{30} + \varepsilon$$



First Linear Regression Model

$$\log(SA_T) = \beta_0 + \beta_{magnitude} * M_w + \beta_{distance} * D + \beta_{vs30} * VS_{30} + \varepsilon$$



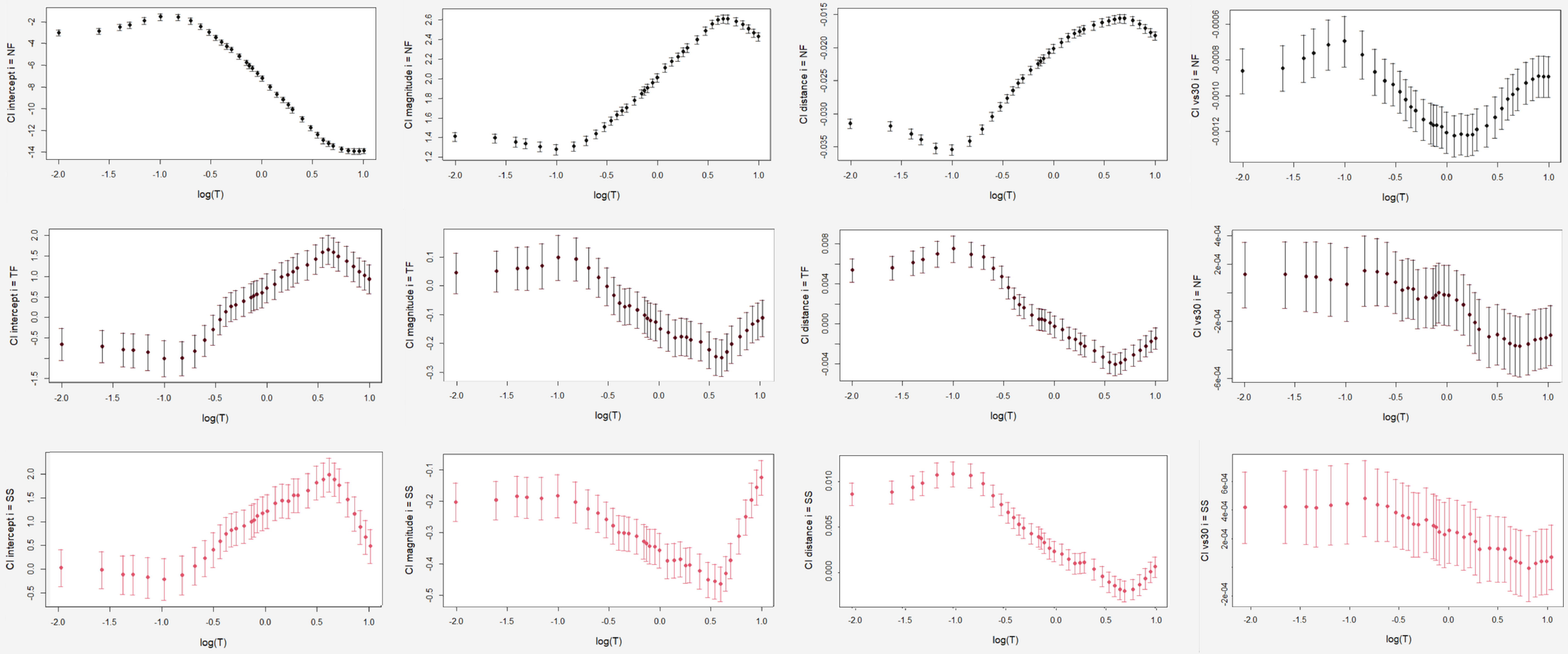


We <u>split data</u> based on the **Style of Faulting** and fit the linear model using just magnitude, distance and vs30. Then, we introduce SoF as a new regressor inside the model.

To choose the best model, we perfor a **K-fold cross validation.** Since the MSE of the model with SoF (as regressor) is smaller with respect to the MSE of the initial model, the <u>model with SoF is considered as the reference</u>.

Second Linear Regression Model

$$\log(SA_T) = \beta_{0,i} + \beta_{magnitude,i} * M_w + \beta_{distance,i} * D + \beta_{vs30,i} * VS_{30} + \varepsilon \quad i = \{TF, NF, SS\}$$



ITA18 Ground Motion Model

ITA18 GMM is the "state-of-art" seismological model used to predict the intensity of spectral acceleration along the Italian territory:

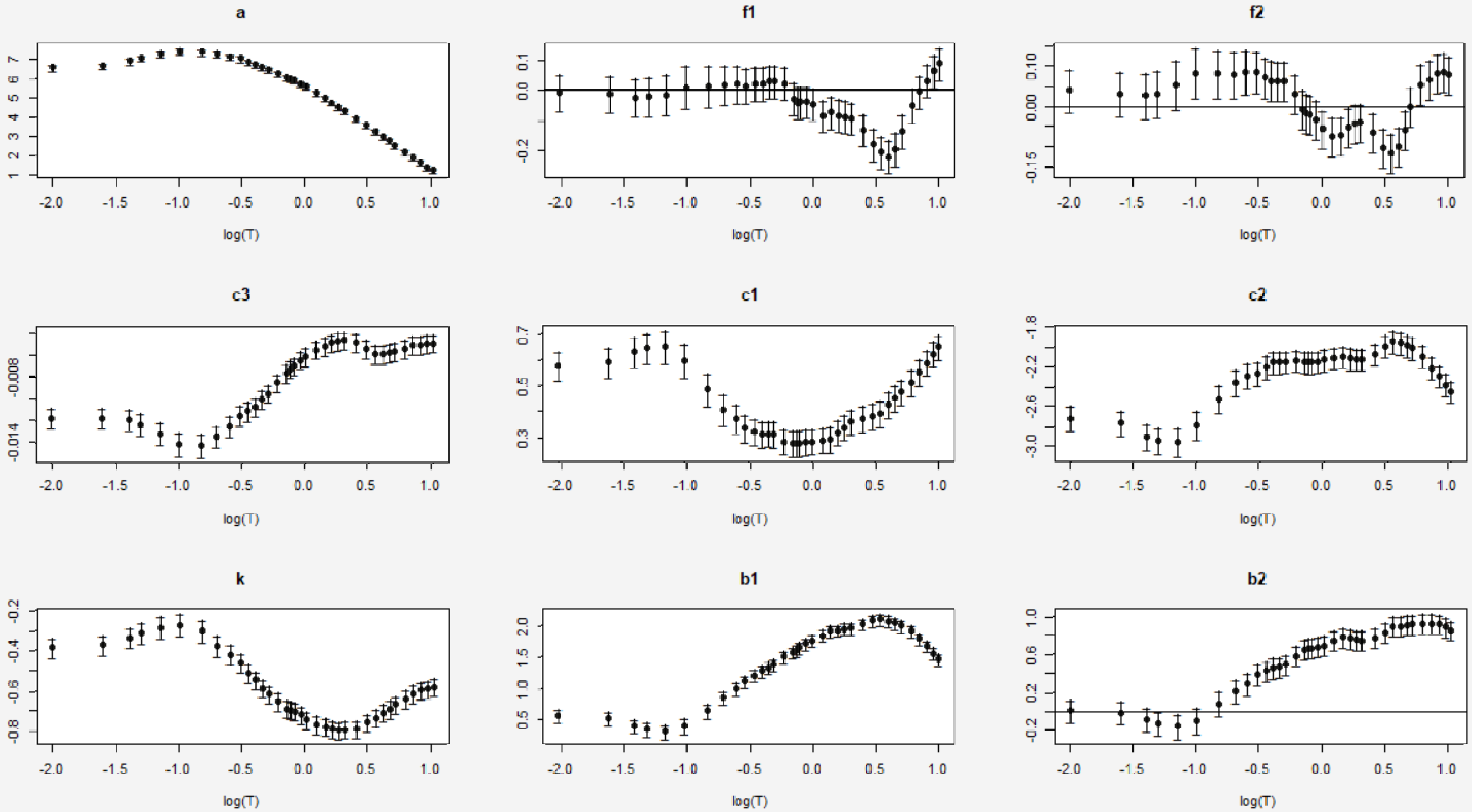
$$\log(SA_T) = a + F_M(M_w, SoF) + F_D(M_w, D) + F_S(VS_{30}) \; + \varepsilon$$

where:

- Source term: $F_M(M_W, SoF) = f_j * SoF_j + \begin{cases} b_1 * (M_W - M_h) & if\ M_W \leq M_h \\ b_2 * (M_W - M_h) & if\ M_W > M_h \end{cases}, with\ M_h = 5{,}7$
- Path term: $F_D(M_W, D) = \left(c_1 * \left(M_W \; - M_{ref}\right) + c_2\right) * \log_{10} D + c_3 * D, with\ M_{ref} = 4{,}5$
- Site term: $F_S(VS_{30}) = k * \log_{10}\left(\frac{V_0}{800}\right),\ with\ V_0 = VS_{30}\ if\ VS_{30} \leq \ 1500\ m/s\ otherwise\ 1500\ m/s$

ITA18 Ground Motion Model

$$\log(SA_T) = a + F_M(M_w, SoF) + F_D(M_w, D) + F_S(VS_{30}) + \varepsilon$$



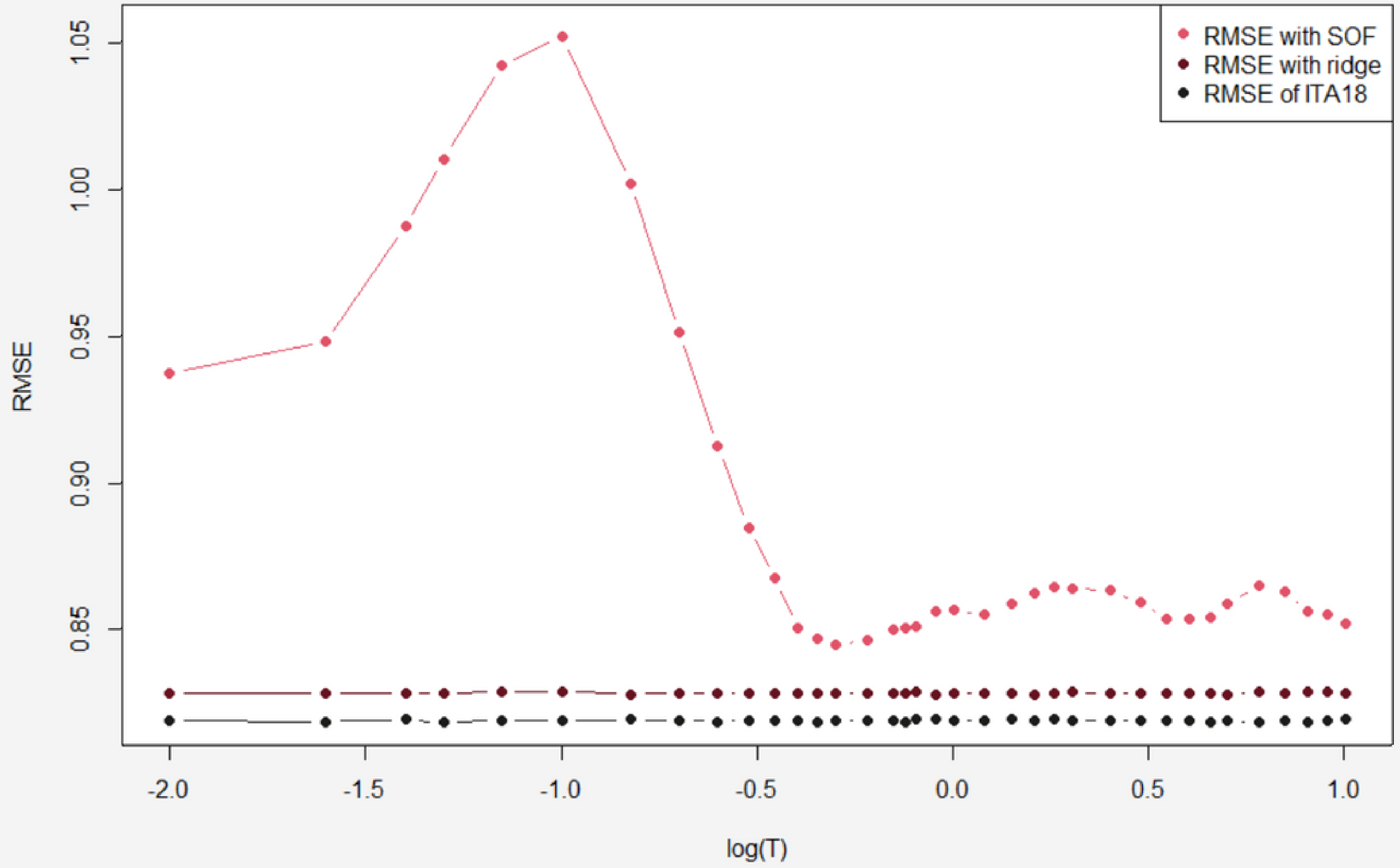
ITA18 Ground Motion Model

$$\log(SA_T) = a + F_M(M_w, SoF) + F_D(M_w, D) + F_S(VS_{30}) + \varepsilon$$

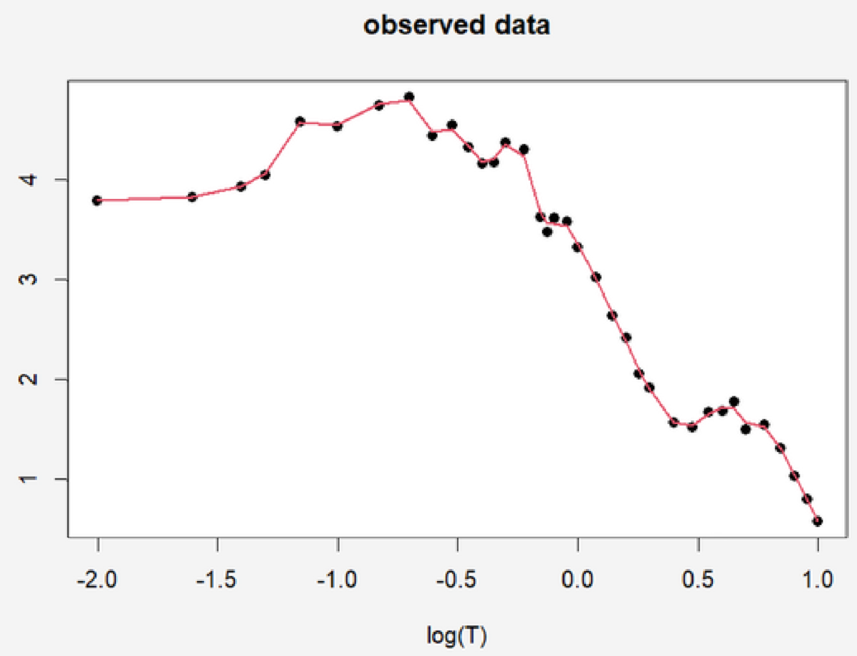
The above model is <u>physically robust and well interpretable</u> but presents some issues due to **collinearity among regressors,** thus we try to solve this problem by using a **Ridge Regression.**

However, as we can see from the plot on the right, the <u>cross validation results</u> show **the original model, as the optimal.**

Summing up, by comparing **ITA18**, its penalized version and the linear model with SoF, the first still represents the **best scalar model** among all of them.

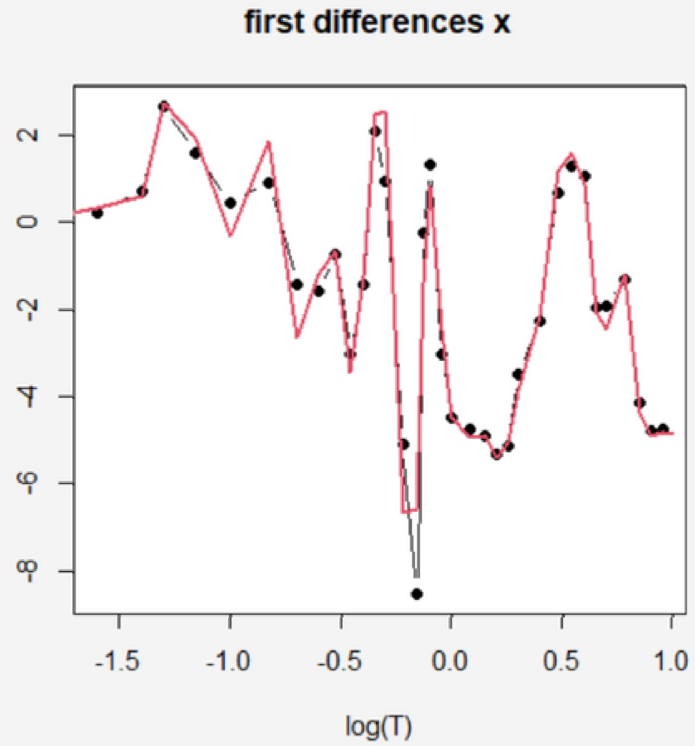


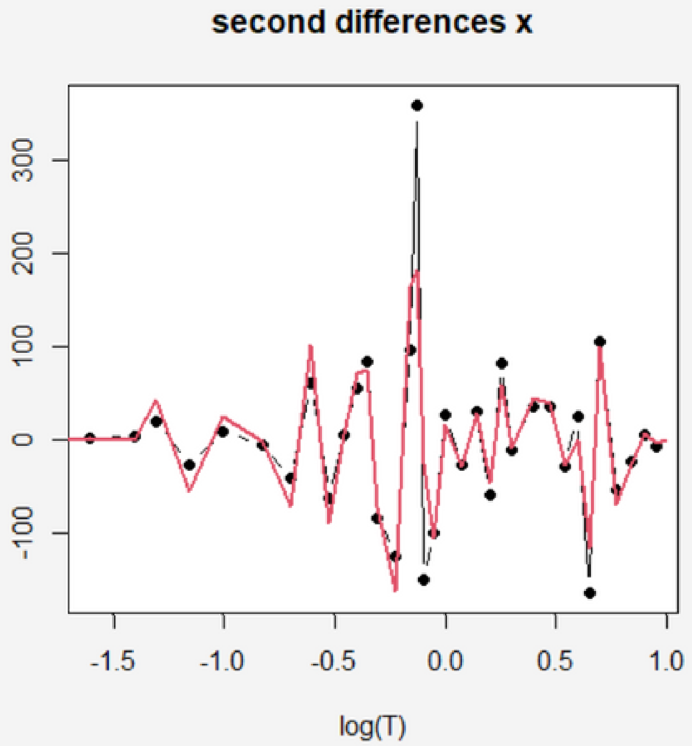
Functional ITA18 Model

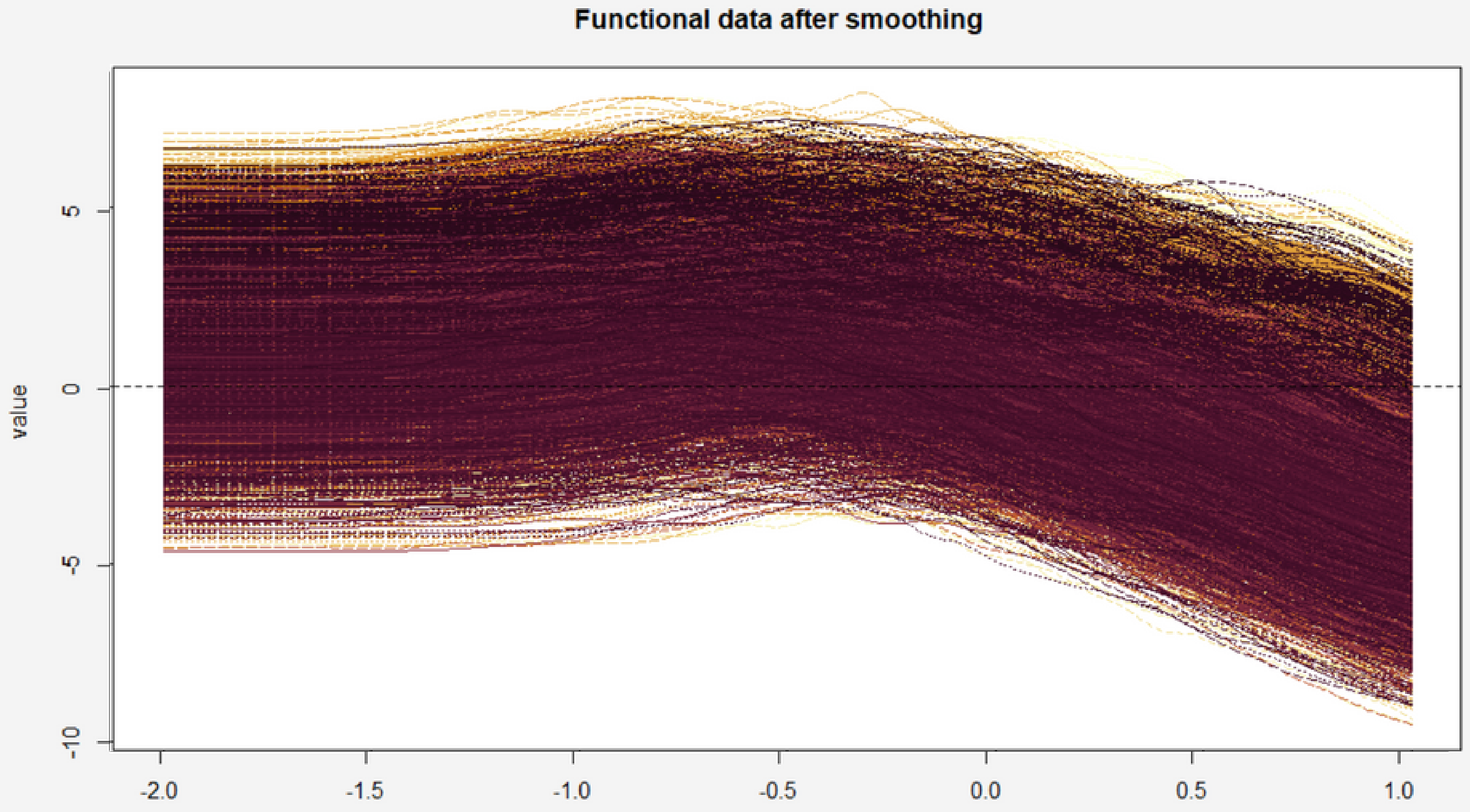


As a further step, we try to improve the scalar ITA18 model taking it into a functional framework.

Firstly, the data are smoothed using **cubic Bspline basis** with coefficient of penalization **λ=10^(-5)** and then, the functional model is fitted by a linear regression.

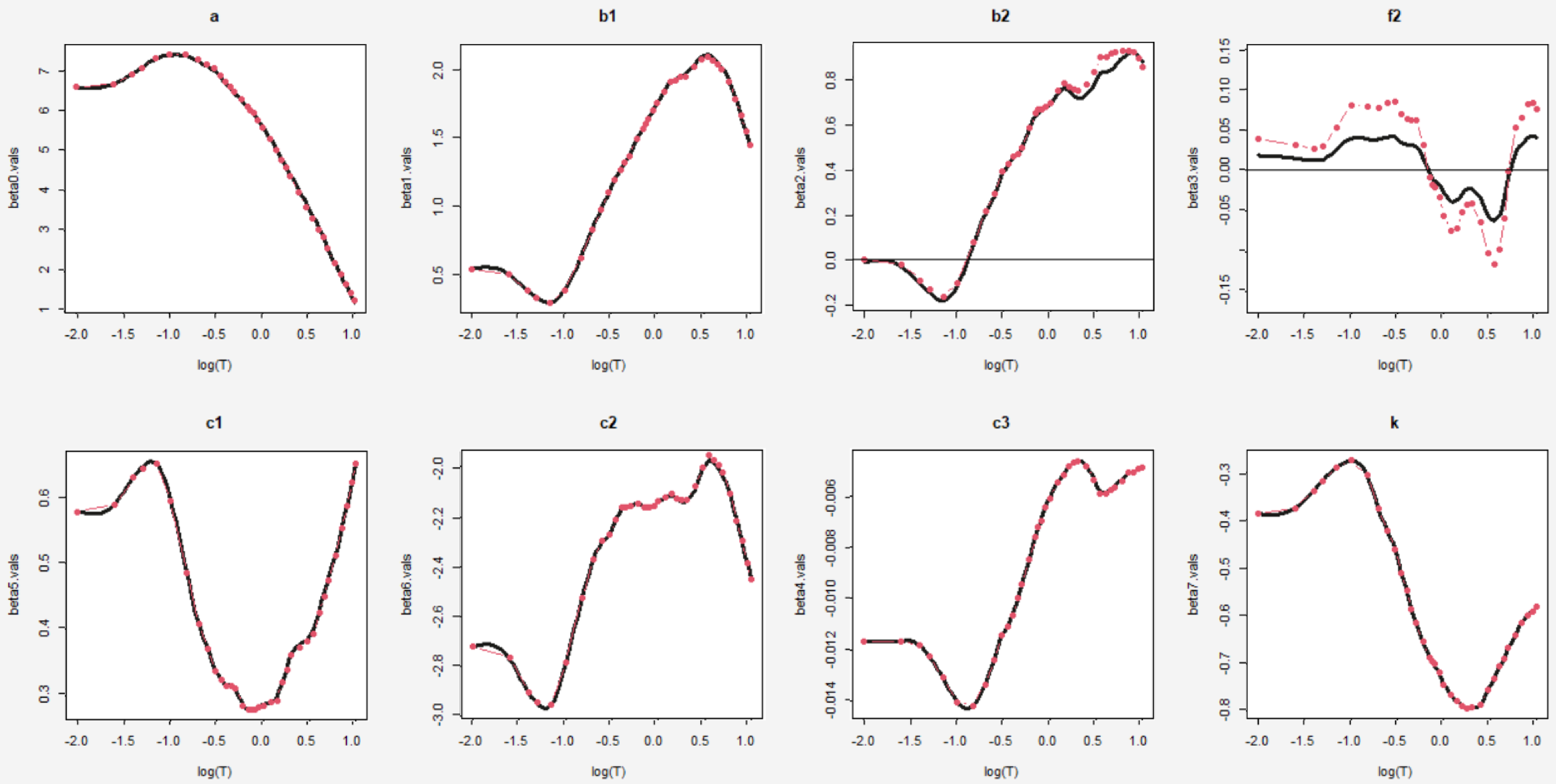






Functional ITA18 Model

$$\log_{10}(SA)(T) = a + F_M(M_w, SoF; T) + F_D(M_w, D; T) + F_S(VS_{30}; T) + \varepsilon$$



Functional ITA18 Model

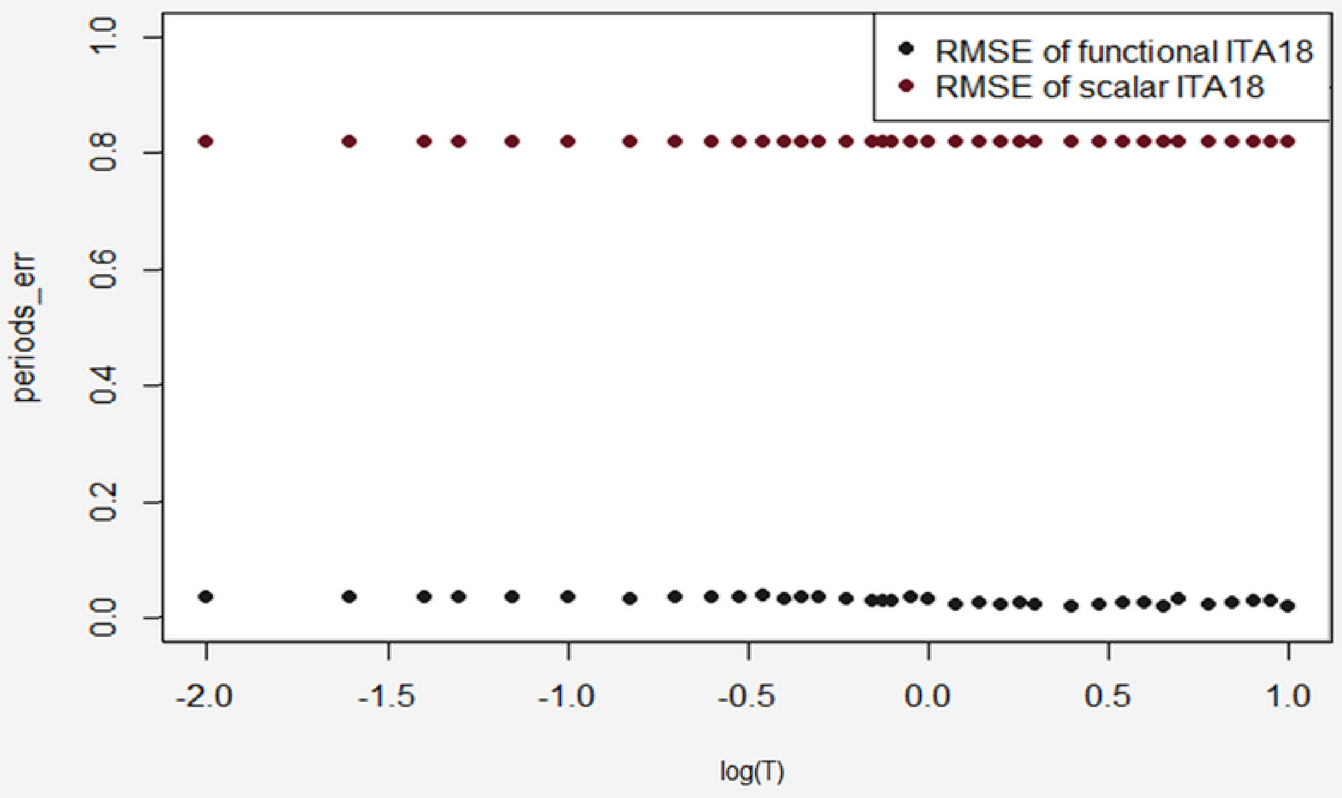
$$\log_{10}(SA)(T) = a + F_M(M_w, SoF; T) + F_D(M_w, D; T) + F_S(VS_{30}; T) + \varepsilon$$

From the **cross validation** results, we have that the MSE of the functional model is always much lower than the MSE of the scalar ITA18, with a mean difference of almost 0.8.

Moreover, functional models have already many advantages, such as:

- **continuous estimation** of the target variable along an entire interval
- **correlation between different time steps** (periods)
- inherent **regularization** of the model

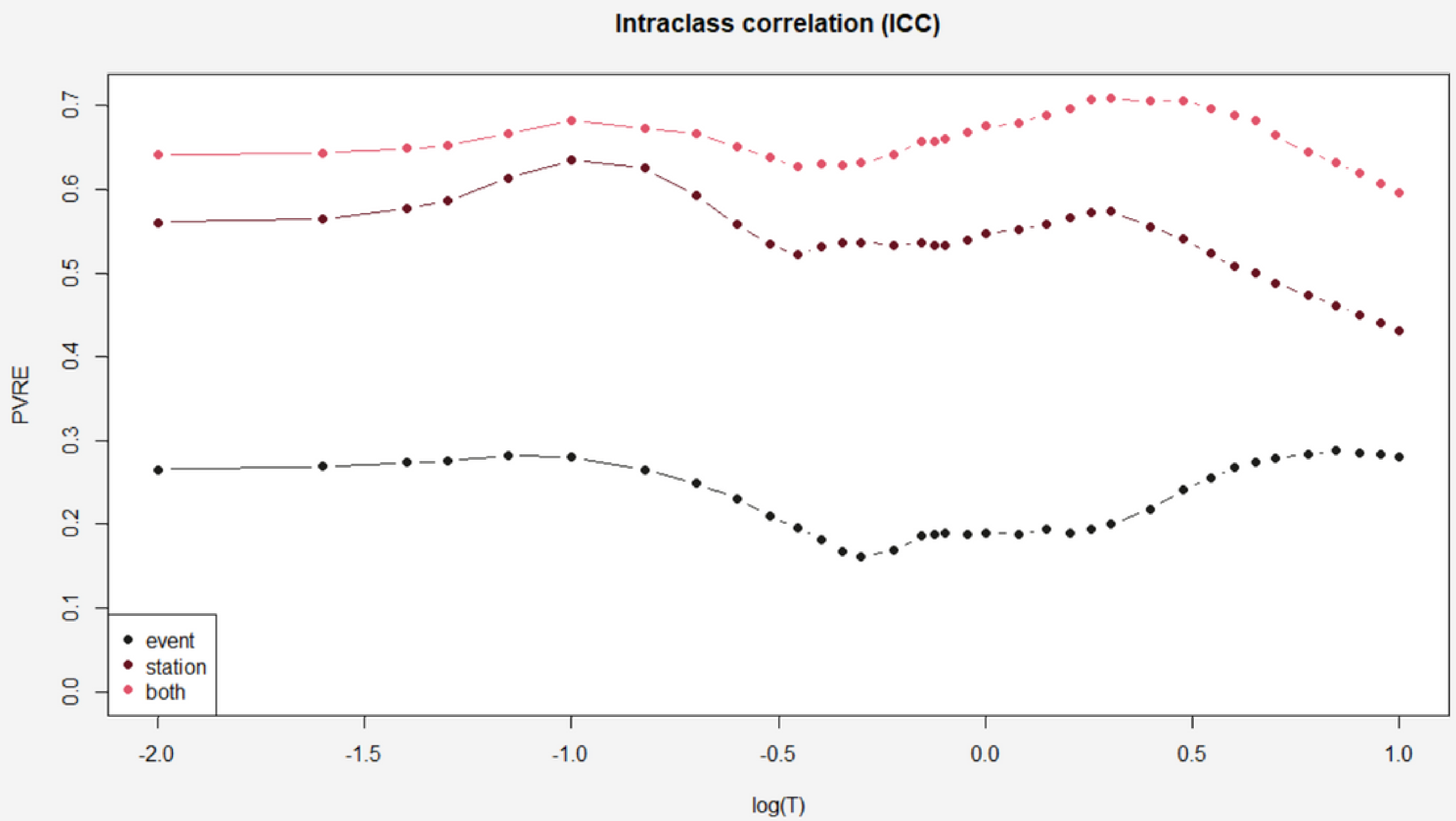
The proposed functional model represents a **robust and significant tool**, which allows one to predict the whole spectrum of ground motion intensity, simultaneously on a continuous range of periods.



Linear Mixed Effects Model

$$\log(SA_T) = a + F_M(M_W, SoF) + F_D(M_W, D) + F_S(VS_{30}) + rand\_effect + \varepsilon$$

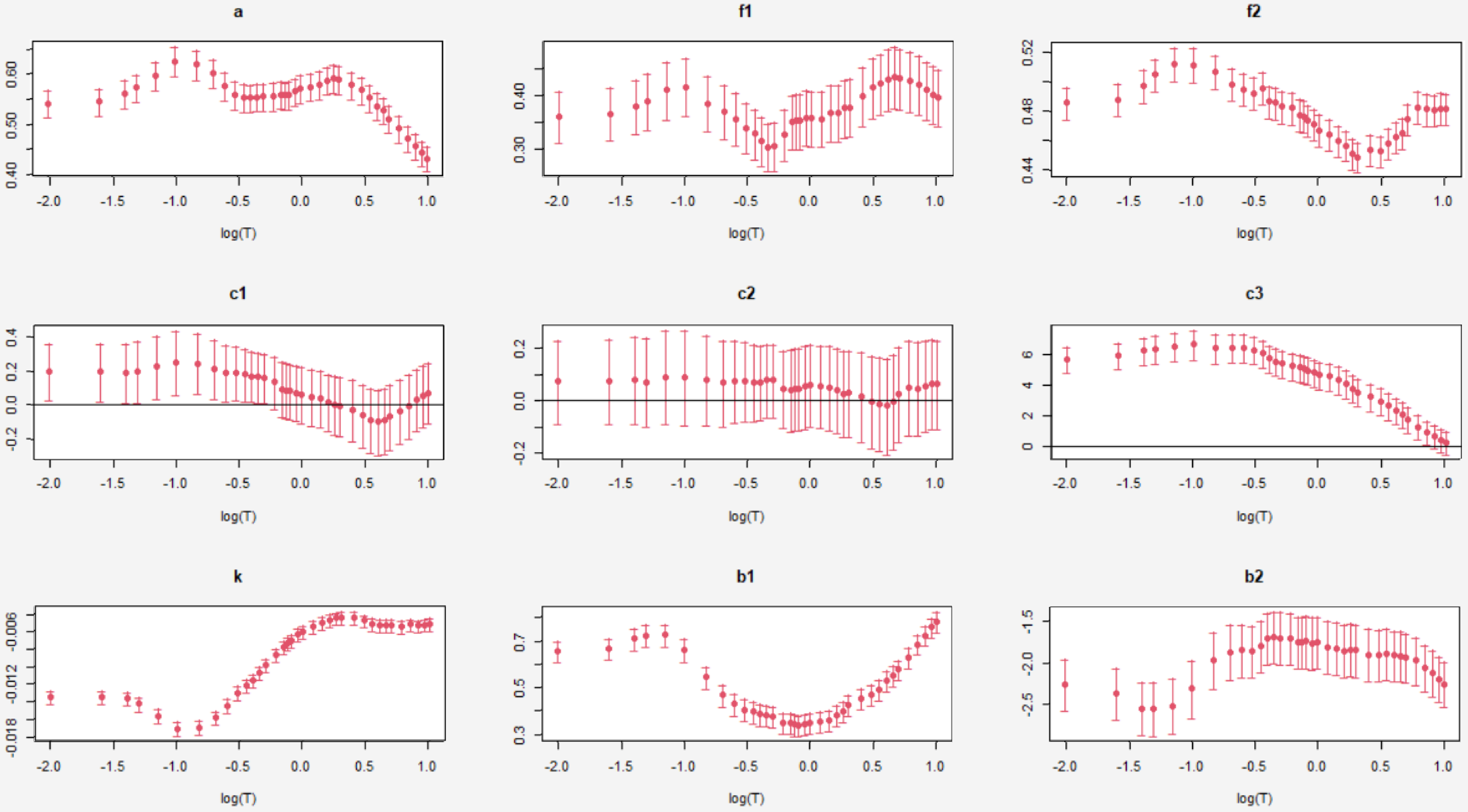
To better capture the **internal variability of the ITA18 model,** we introduce random effects given by the station of measurement, the event and both of them simultaneosly.



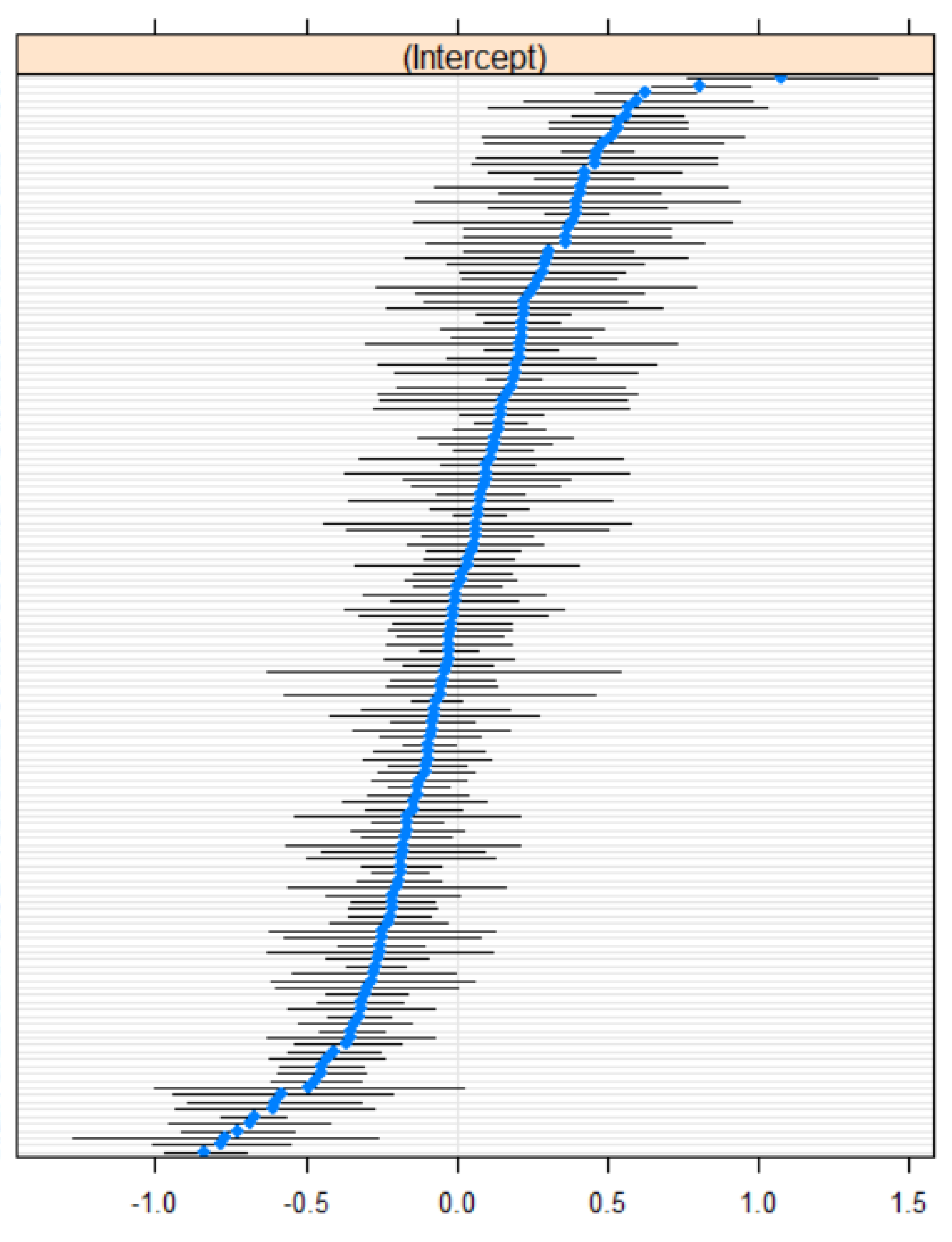
After testing the 3 random effect separately, the best model to describe the internal dependence among units is the one with both, **station_id and event_id as random effects**.

Linear Mixed Effects Model

$$\log(SA_T) = a + F_M(M_W, SoF) + F_D(M_W, D) + F_S(VS_{30}) + rand\_effect + \varepsilon$$

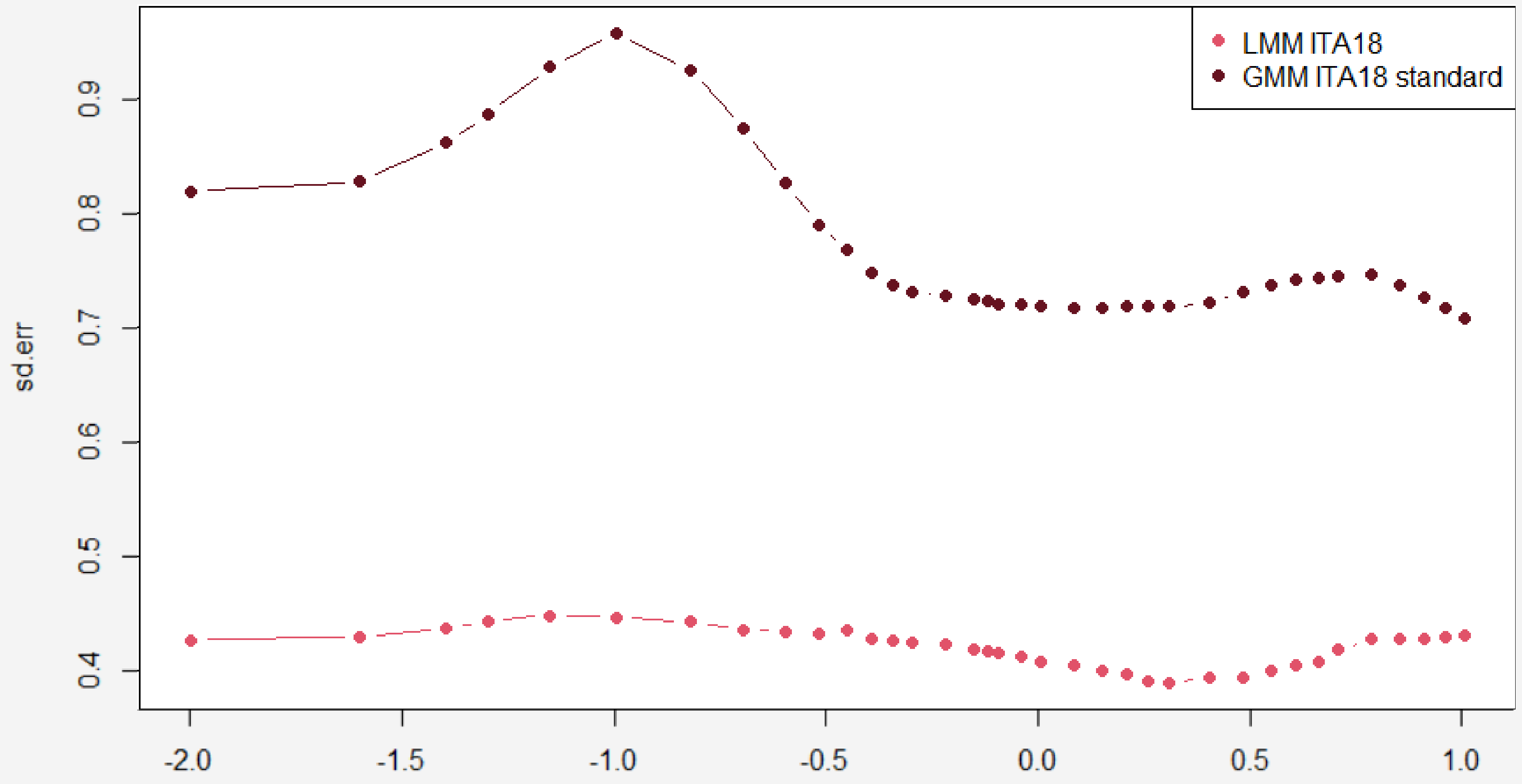


Linear Mixed Effects Model



As we can observe from the dotplot on the right, the random effect has a significant impact, thus we expect the standard deviation of the residuals in the Linear Mixed Effects Model (LMM) to be lower compared to the standard model.

This intuition is confirmed by the graph below, where the standard deviation of the residuals in the model with random effect is consistently lower than in the other model across all periods.

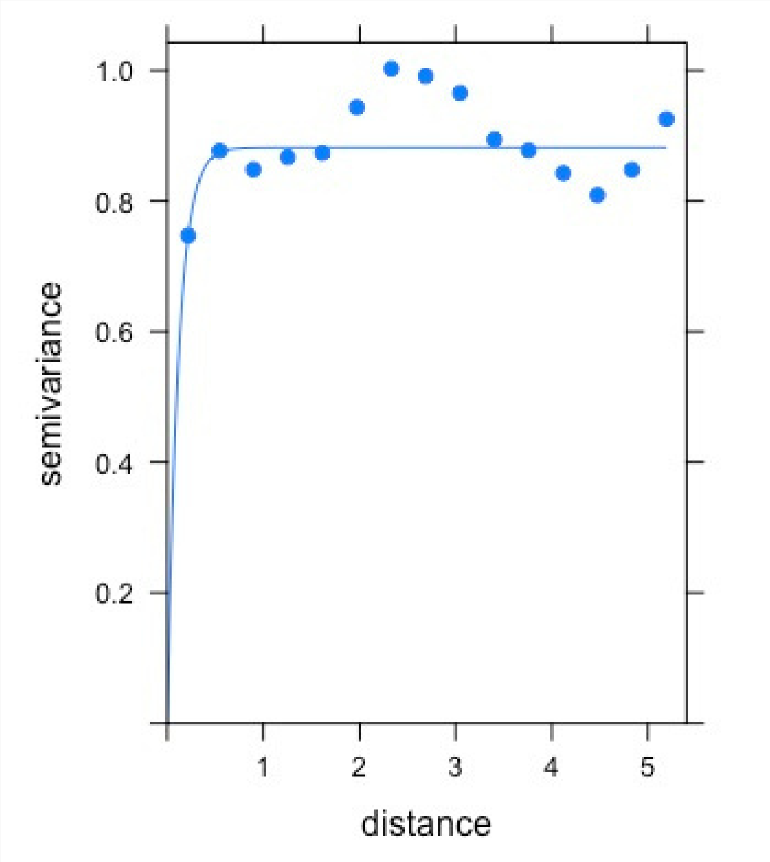


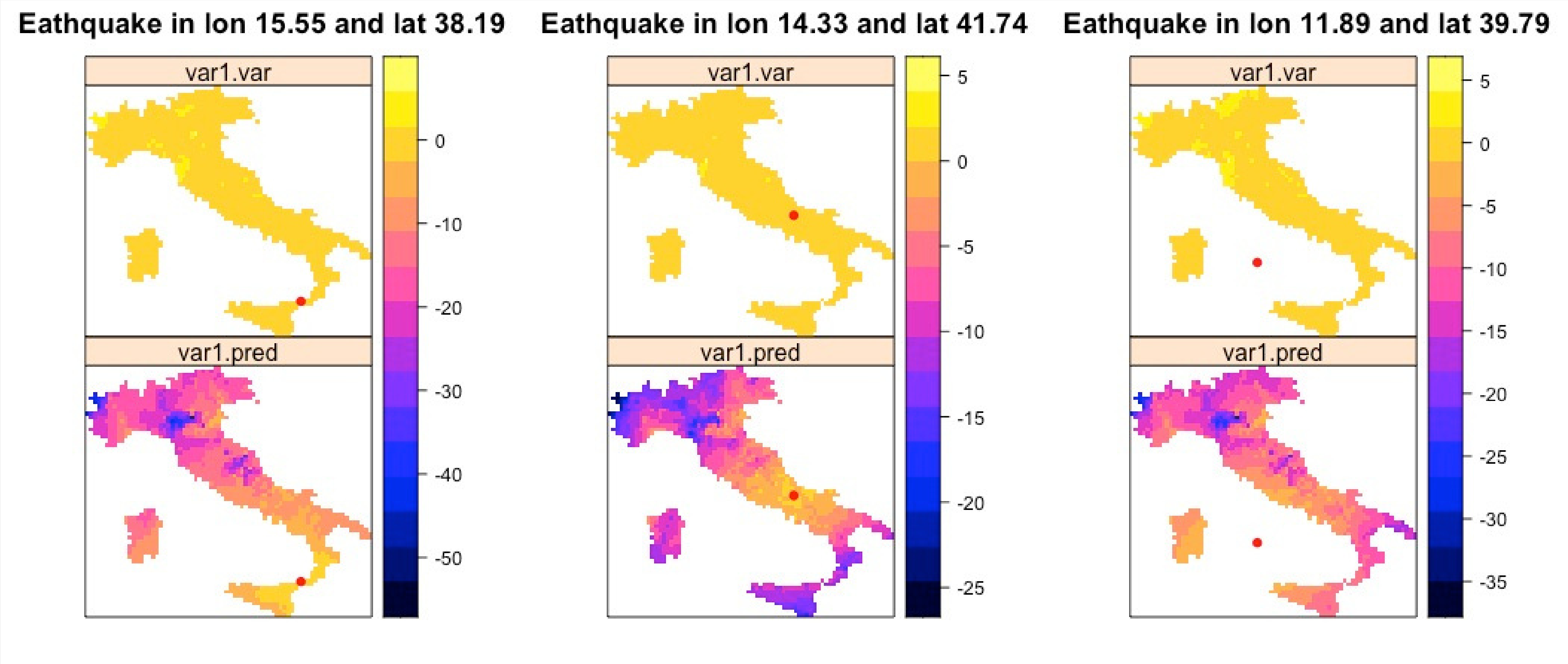
Geostatistics Analysis

At the beginning we tried to make a spatial visualization of our sampling stations on the italian territory.
We used the WSG84 coordinates of those stations and colored them according to the different VS30 value, since it is an indication of the kind of soil

Geostatistics Analysis

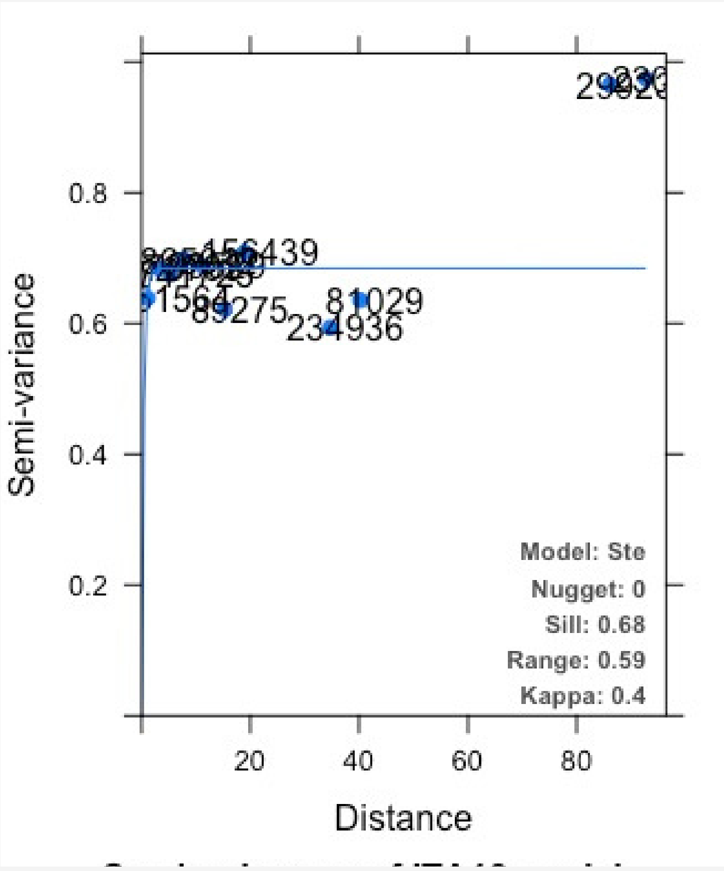
First draft : we tried to explain our variable of interest (logSA_0.01) with a simple model via universal kriging.
Hence we used log(SA_0.01)~ distance + vs30 + magnitude

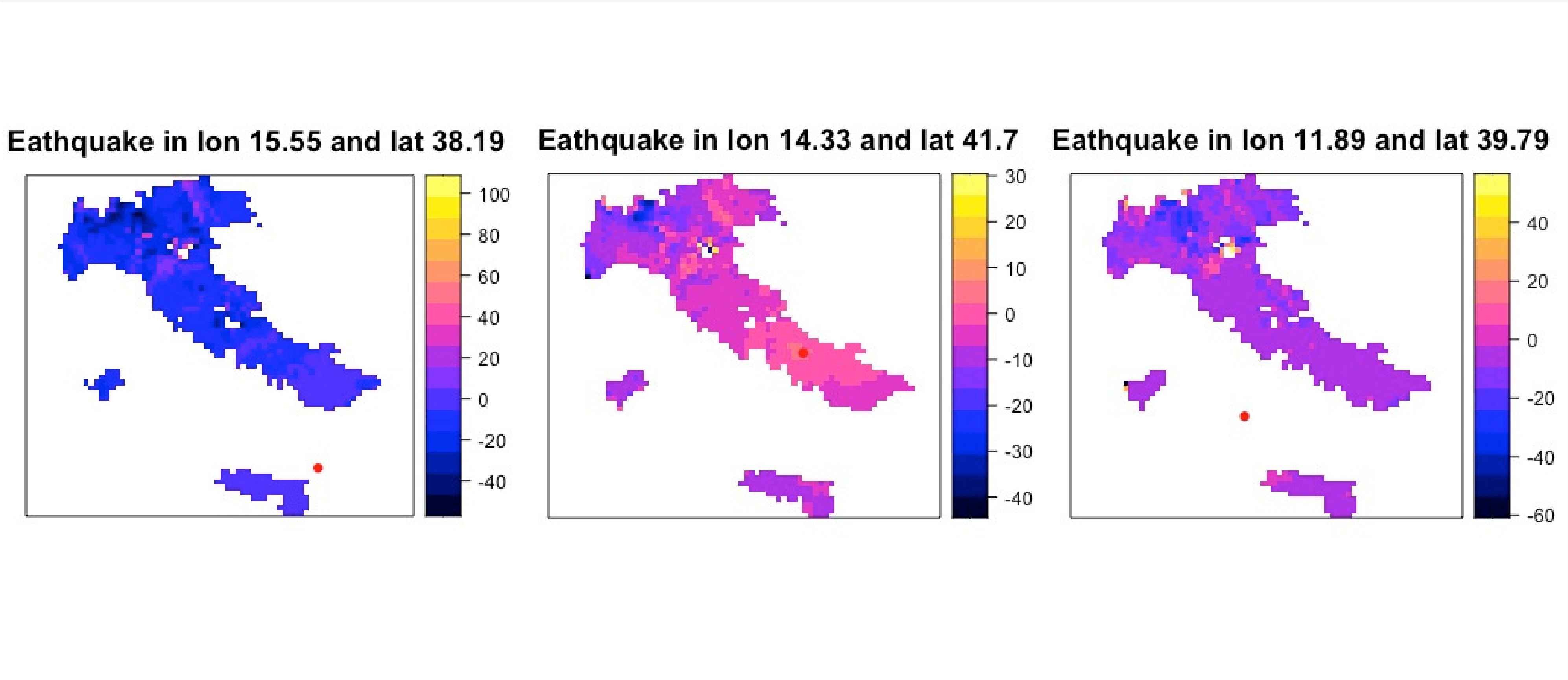




Geostatistics Analysis

Second draft : we used the official ITA18 ground motion model which has been proved to be better than the simpler one in previous analysis. Moreover it is more accurate in theory since it is trained on a large majority of stations (eg also in the states or japan).
DA MOSTRARE ALLA TUTOR PER CHIARIMENTI





Geostatistics Analysis

References

- ITA18 strong-motion flat-files (https://shake.mi.ingv.it/ita18-flatfile/)

- T. Bortolotti, A. Menafoglio "Weighted functional data analysis for partially observed seismic data: An application to ground motion modelling in Italy"