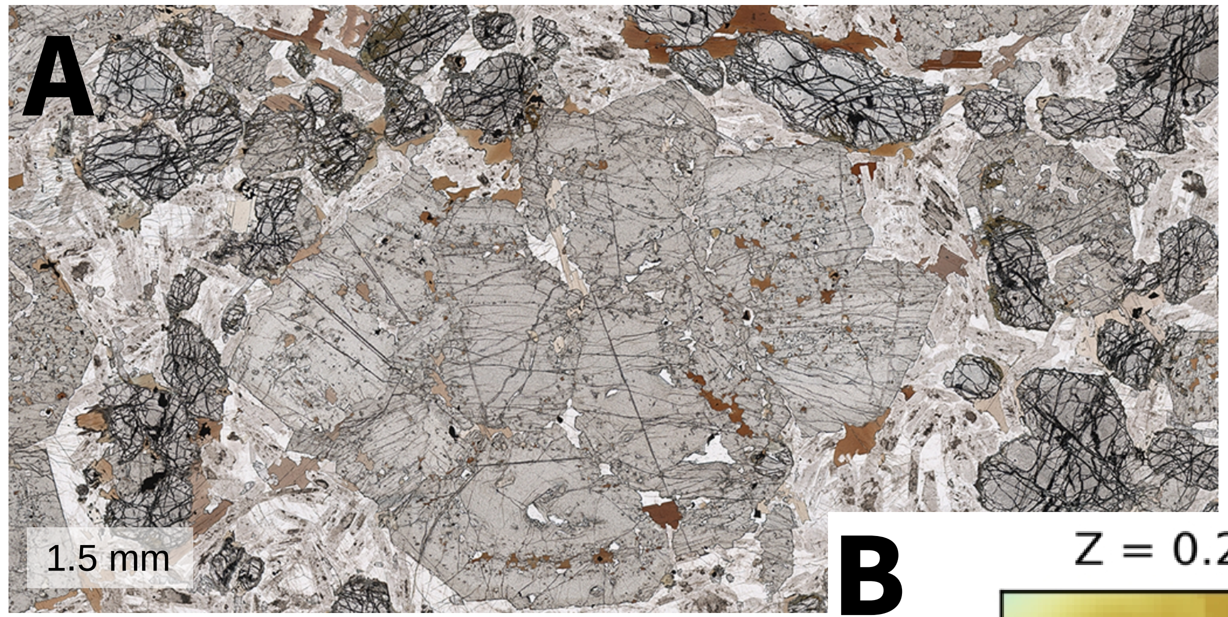
Small-scale heterogeneity characterization using Bayesian inference

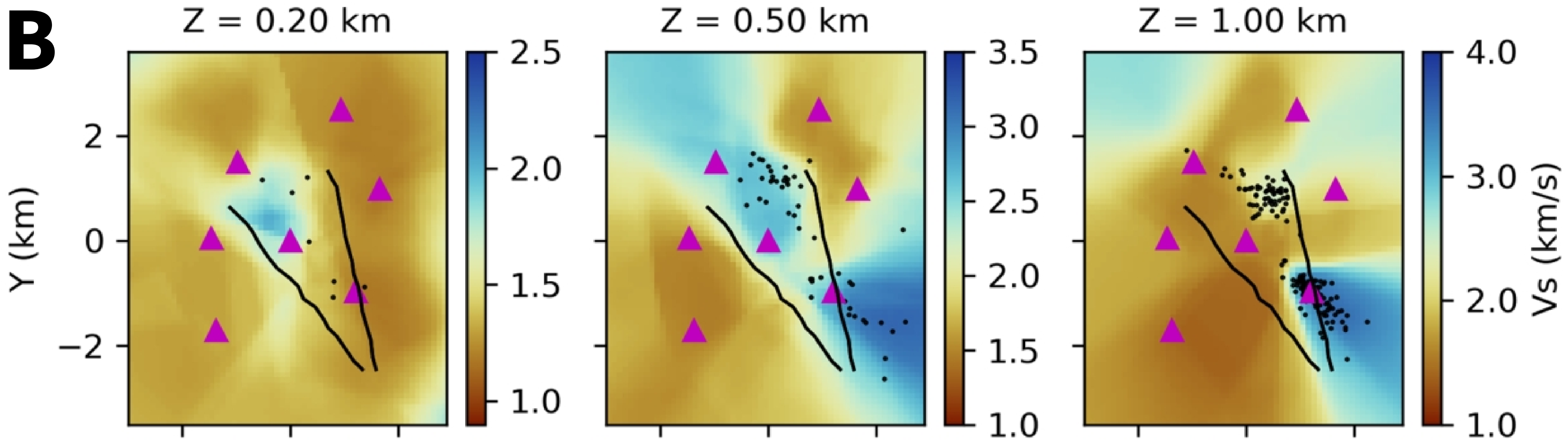
Itahisa González Álvarez[1,*], Sebastian Rost[1], Andy Nowacki[1], Neil Selby[2]
[1]School of Earth and Environment, University of Leeds, [2]AWE Blacknest
eeinga@leeds.ac.uk @seismowaves

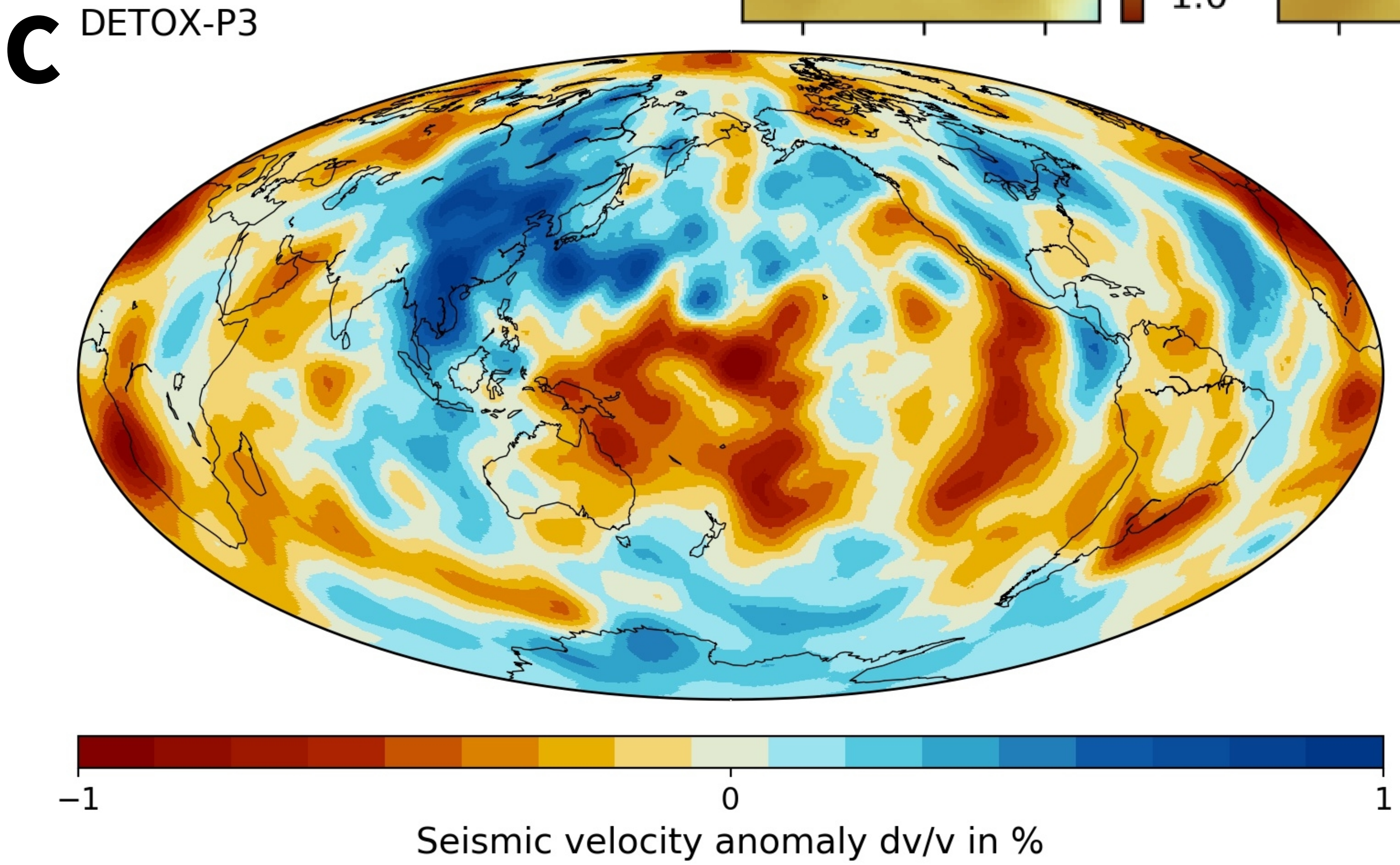
Introduction



A *Kentallenite thin section, from University of Leeds teaching collection at www.virtualmicroscope.org.*

B *Horizontal slices through the 3-D shear velocity model of New Ollerton mining site, from joint body and surface wave inversion. From Zhang et al. (2020).*





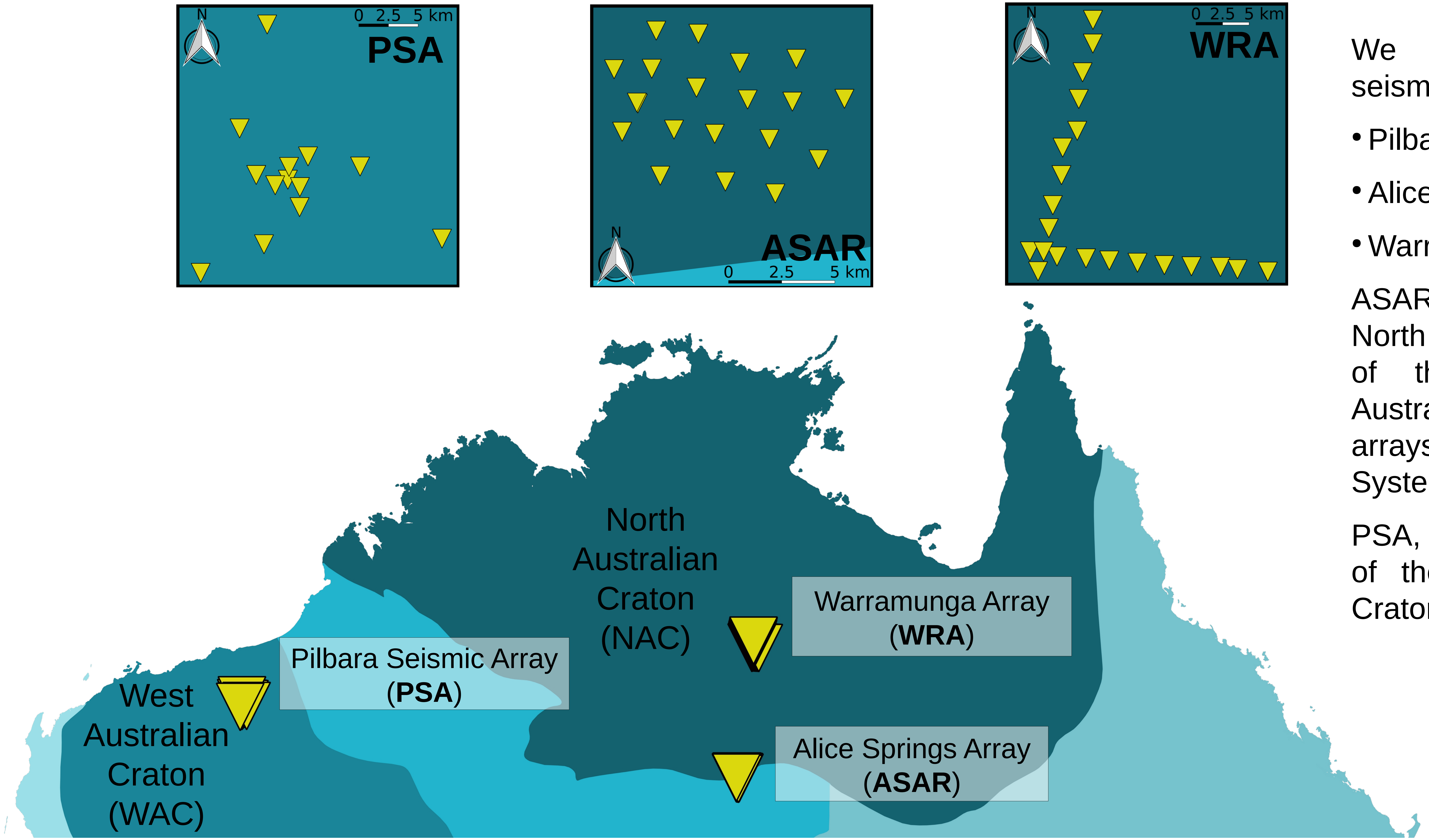
C *Global tomography depth slice at the core-mantle boundary (~2890 km), from the DETOX-P3 model by Hosseini et al. (2020).*

Our planet is highly heterogeneous at multiple scale lengths, yet these inhomogeneities are not always included in models of the Earth's interior.

In particular, seismic codas are caused by the scattering of the wavefield at near-receiver, small-scale structures. Scattering and attenuation affect signals recorded at the surface and can, therefore, affect any measurements and predictions we obtain from them.

Thorough understanding of these phenomena and their causes is key. In addition to providing additional information about the tectonic history of the region, small-scale characterizations can help us calculate correction factors or develop a technique that effectively removes the effect of inhomogeneities from seismic signals.

Data



We analysed data from three seismic arrays in Australia:

- Pilbara Seismic Array (PSA)
- Alice Springs Array (ASAR)
- Warramunga Array (WRA)

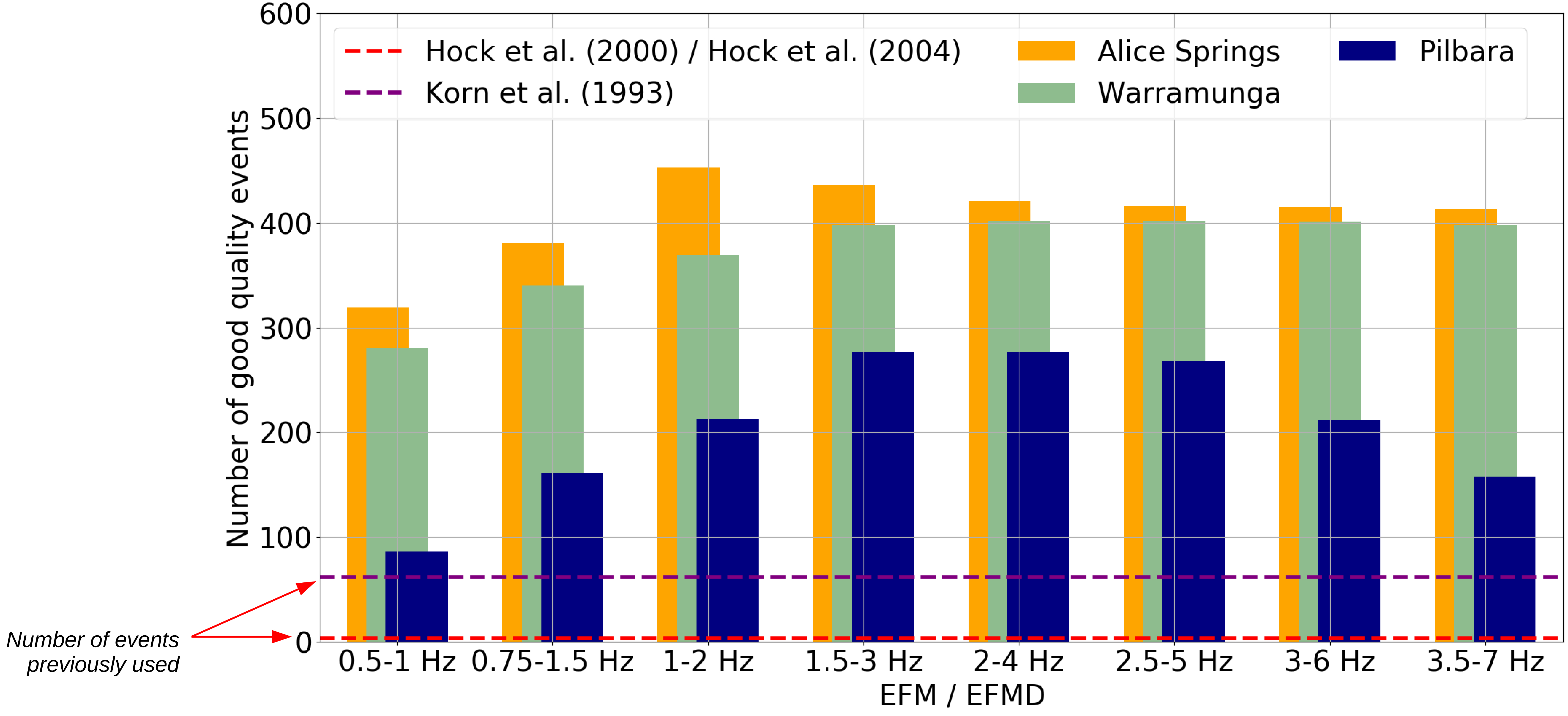
ASAR and WRA are located on the North Australian Craton (NAC), one of the Proterozoic cratons in Australia, and are primary seismic arrays of the International Monitoring System.

PSA, on the other hand, sits on top of the Archaean West Australian Craton (WAC).

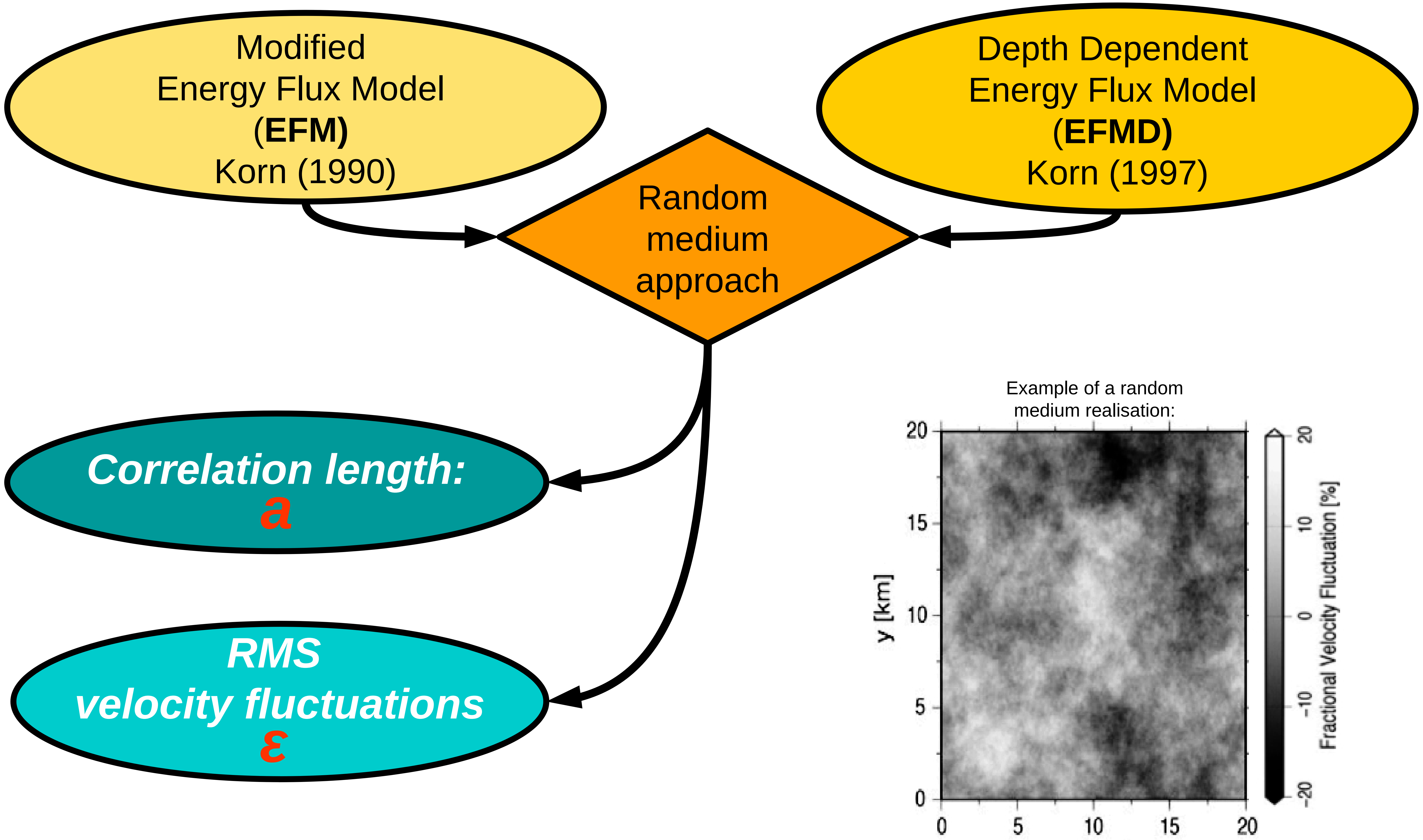
Data

Our dataset is, to our knowledge, the largest dataset ever used in a study using energy flux models. This dataset includes, for each array, earthquakes:

- from January 2012 to December 2018
- 30 to 80 degrees epicentral distance from the arrays
- minimum depth of 200 km
- magnitudes 5 to 7



Methods



Modified from Sato et al. (2012), Chapter 2, Figure 2.6

We use two different methods to model the small-scale structure of the lithosphere beneath seismic arrays or stations:

- Energy Flux Model (EFM): it does not allow any layering in the structure.
- Depth Dependent Energy Flux Model (EFMD): can be used with multiple scattering layers.

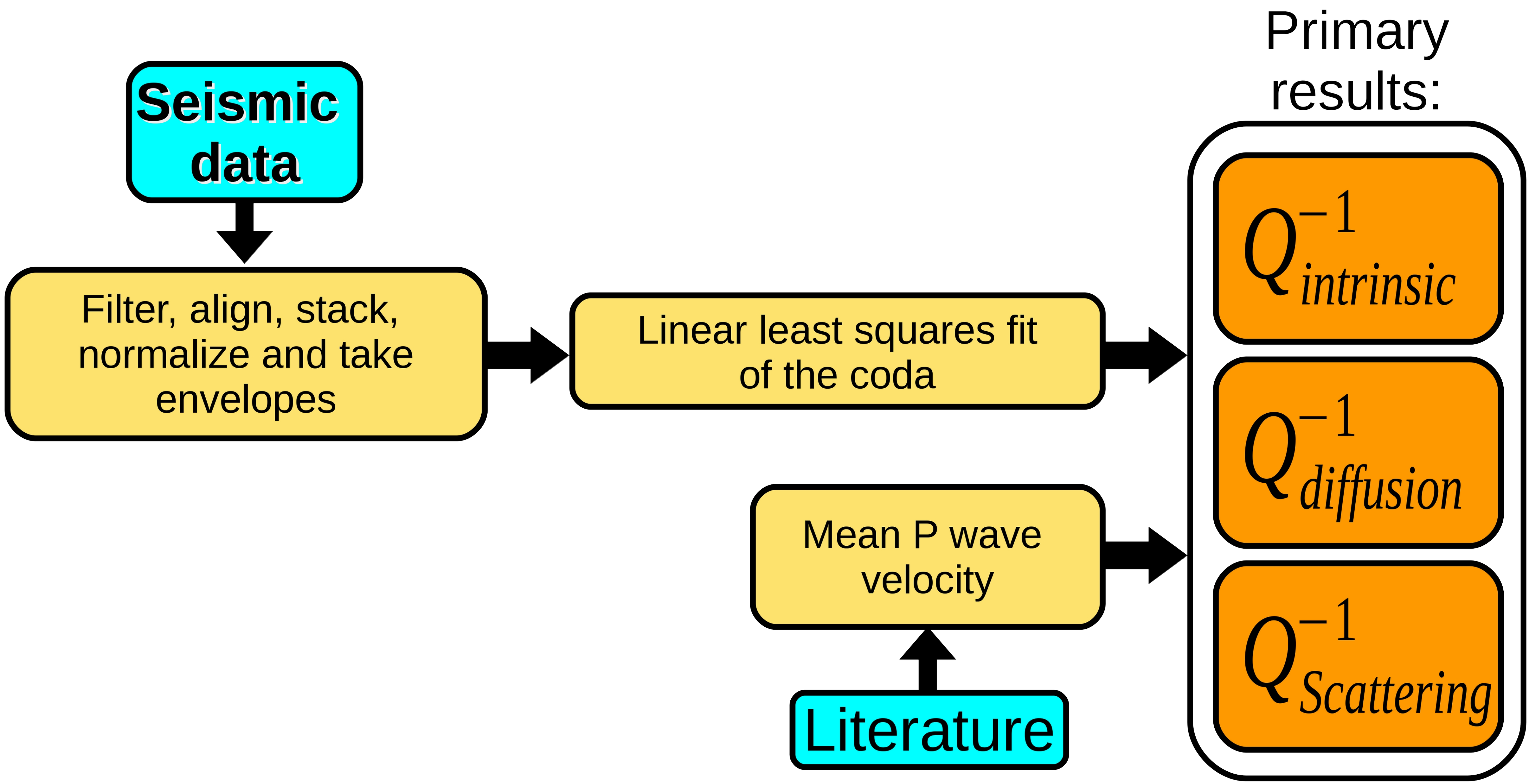
They both use the random medium approach and are stochastic methods, thus meaning that they do not aim at deterministically imaging the lithosphere but at characterising it statistically.

Each layer in the model is characterized by an autocorrelation function which depends on two parameters: ***a*** is the correlation length or characteristic scale length of the heterogeneity, and **ε** is the RMS velocity fluctuations, which represents the strength of the heterogeneity.

These methods:

- Use the single scattering approximation.
- Are acoustic models.
- Do not describe scattering interactions in detail, but treat them statistically.
- Are computationally highly effective.

The single-layer Energy Flux Model (EFM)

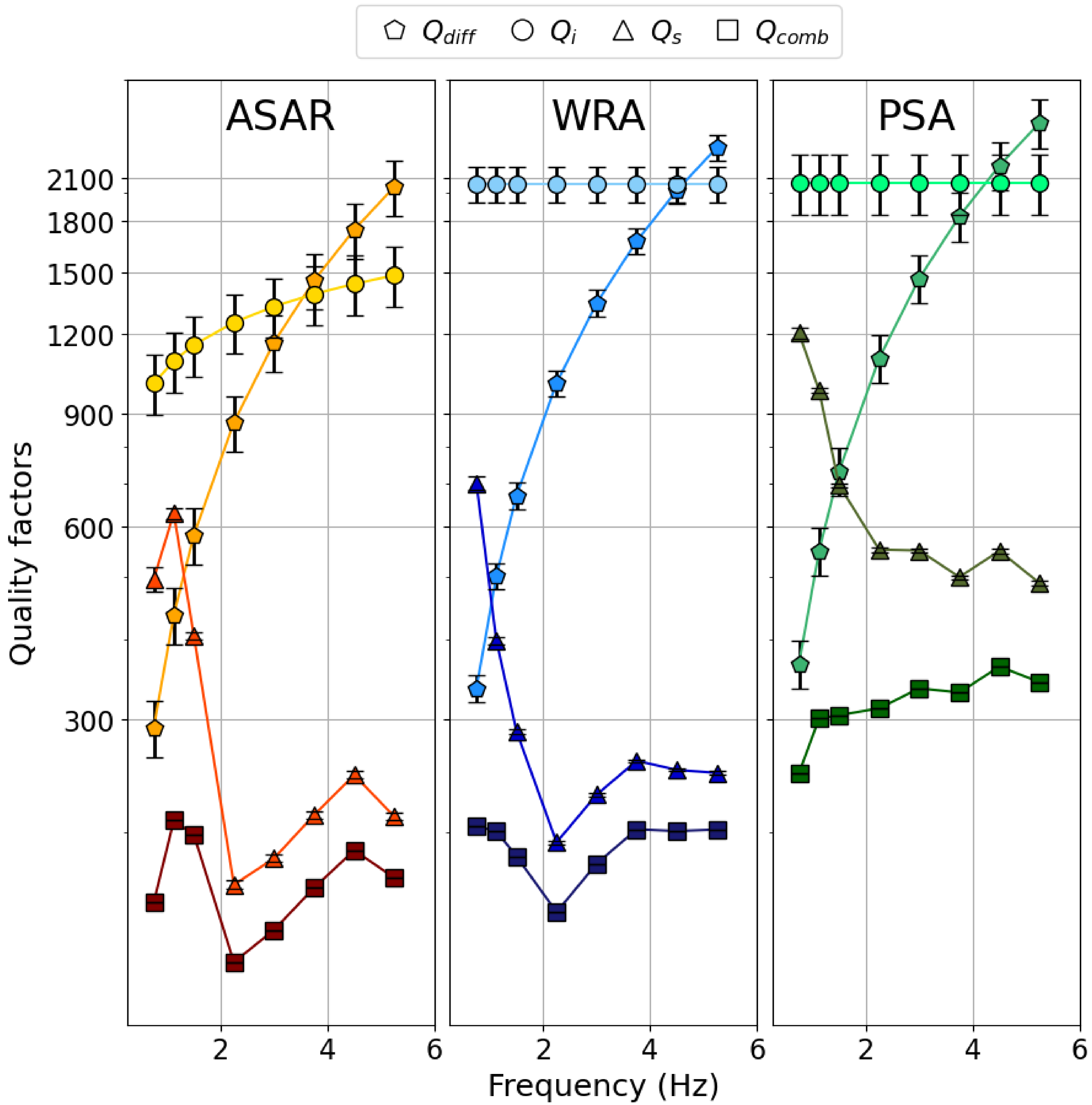


An advantage of the EFM, is that it allows us to compare between different attenuation mechanisms:

- scattering
- anelasticity (intrinsic attenuation)
- diffusion (leakage out of the heterogeneous layer)

To do this, the first step is to calculate the coda envelopes as described in Korn (1990). The coda of the logarithm of the squared envelope can be fitted with linear least squares. For each frequency band, we obtain two parameters from the linear fit. These are used to calculate the scattering, intrinsic and diffusion quality factors for each frequency band. A background P wave velocity of the area is required for the calculation.

The single-layer Energy Flux Model (EFM)



Results for the Australian arrays

For each array, and assuming an exponential autocorrelation function, we calculated the value and frequency dependence of the intrinsic (Q_i), diffusion (Q_{diff}) and scattering (Q_s) quality factors.

The combined quality factor, Q_{comb}, summarises the effects of attenuation in the coda:

$$Q_{comb}^{-1} = Q_s^{-1} + Q_i^{-1} + Q_{diff}^{-1}$$

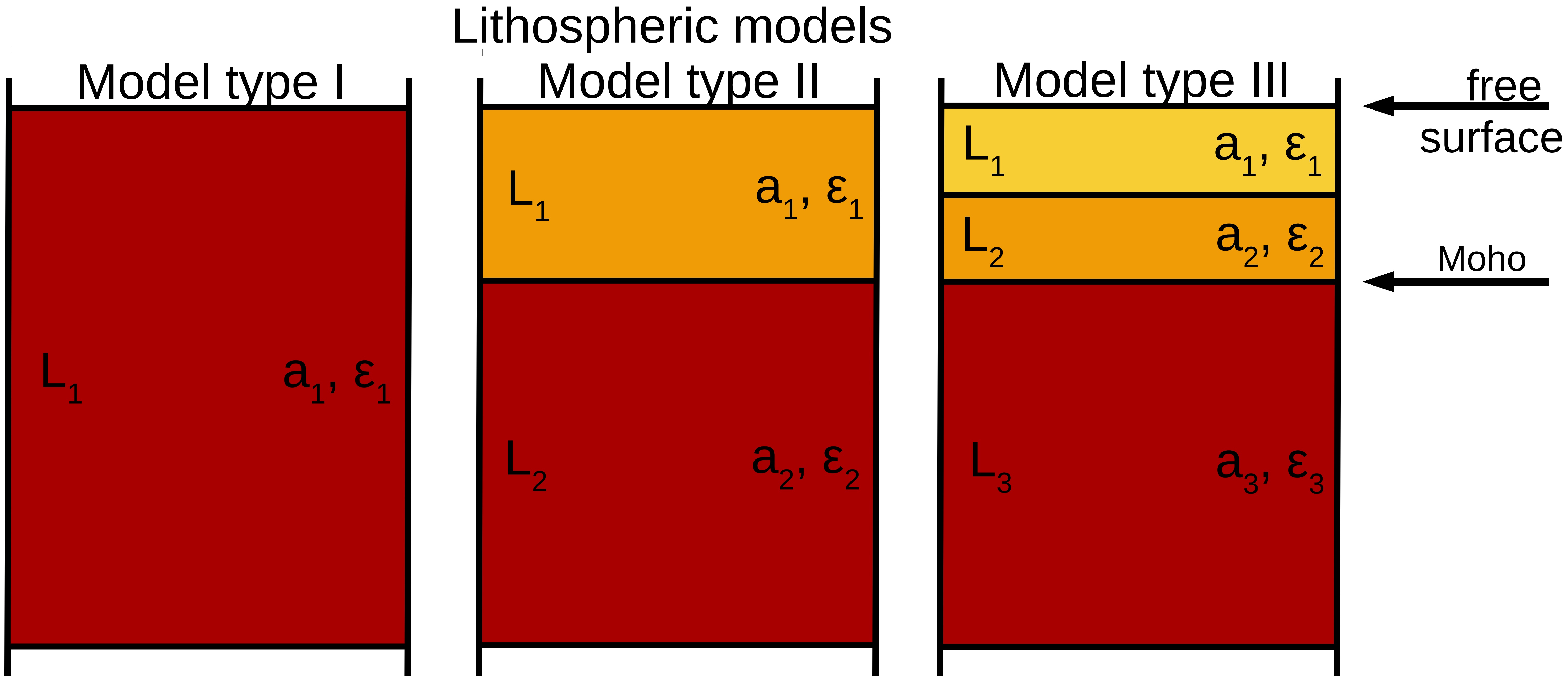
Our results:

- indicate intrinsic and diffusion attenuation are strongest for ASAR.
- suggest scattering and total attenuation are similarly strong for ASAR and WRA.
- show that the lithosphere beneath PSA is less attenuating and heterogeneous than for ASAR or WRA.
- agree with previous studies and with the tectonic histories of the regions the arrays are located on. Lower quality factors are related to areas with recent or intense tectonic histories. ASAR is on an area widely affected by the accretionary processes that took place during the assembly of the Australian continent, as well as two orogens. WRA, closer to the centre of the NAC, is on an area with a less active tectonic history. The WAC, where PSA is, has been located on passive tectonic margins for most of its history, which can be related to the much higher quality factors obtained for this array.

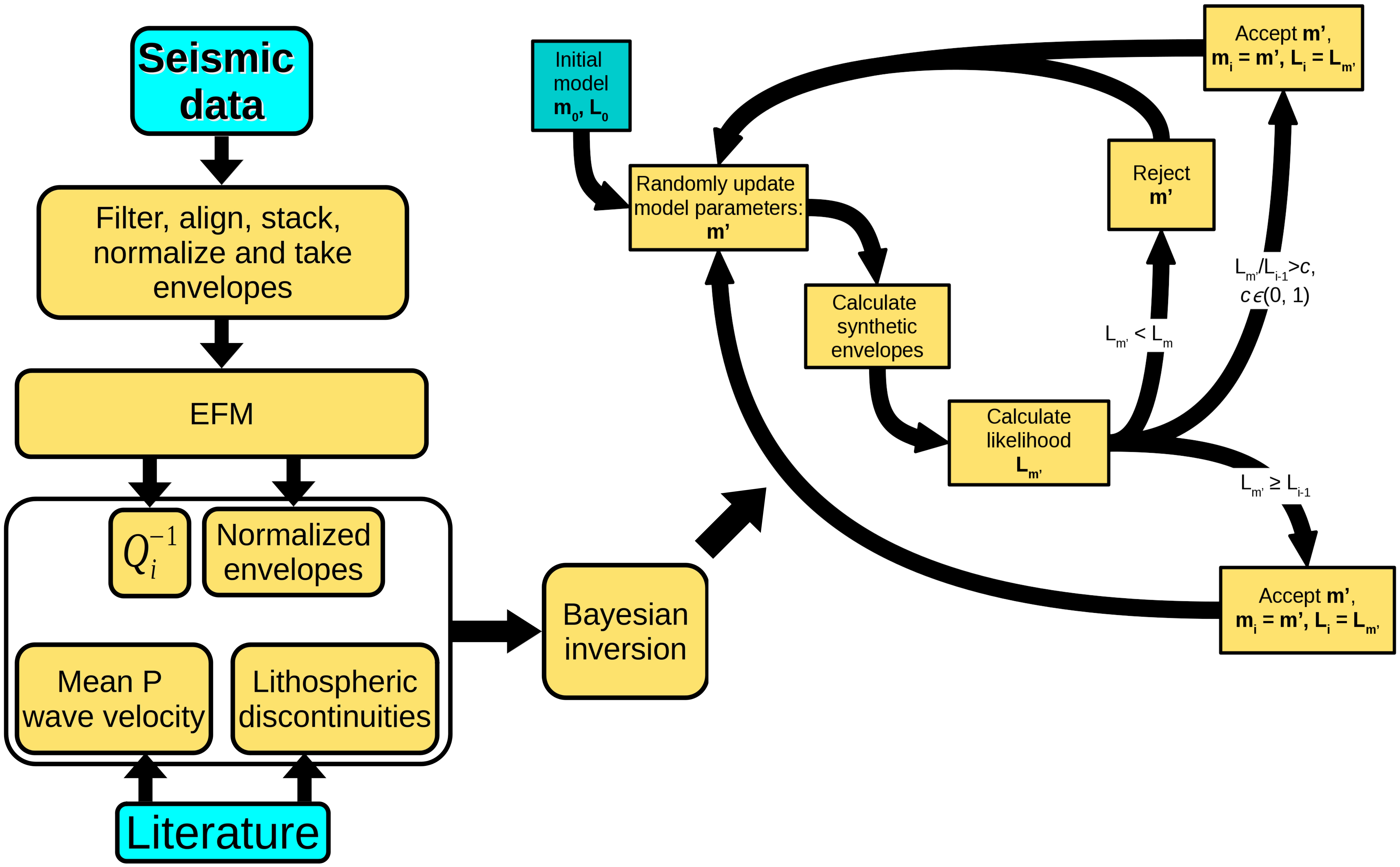
The depth-dependent Energy Flux Model (EFMD)

For the EFMD, we tested three different lithospheric models with increasing complexity:

- Model type I: single scattering layer that encompasses the entire lithosphere
- Model type II: two horizontal layers, which represent the crust and lithospheric mantle respectively
- Model type III: includes two equally thick layers in the crust and the same bottom layer as model type II for the lithospheric mantle.



The depth-dependent Energy Flux Model (EFMD)

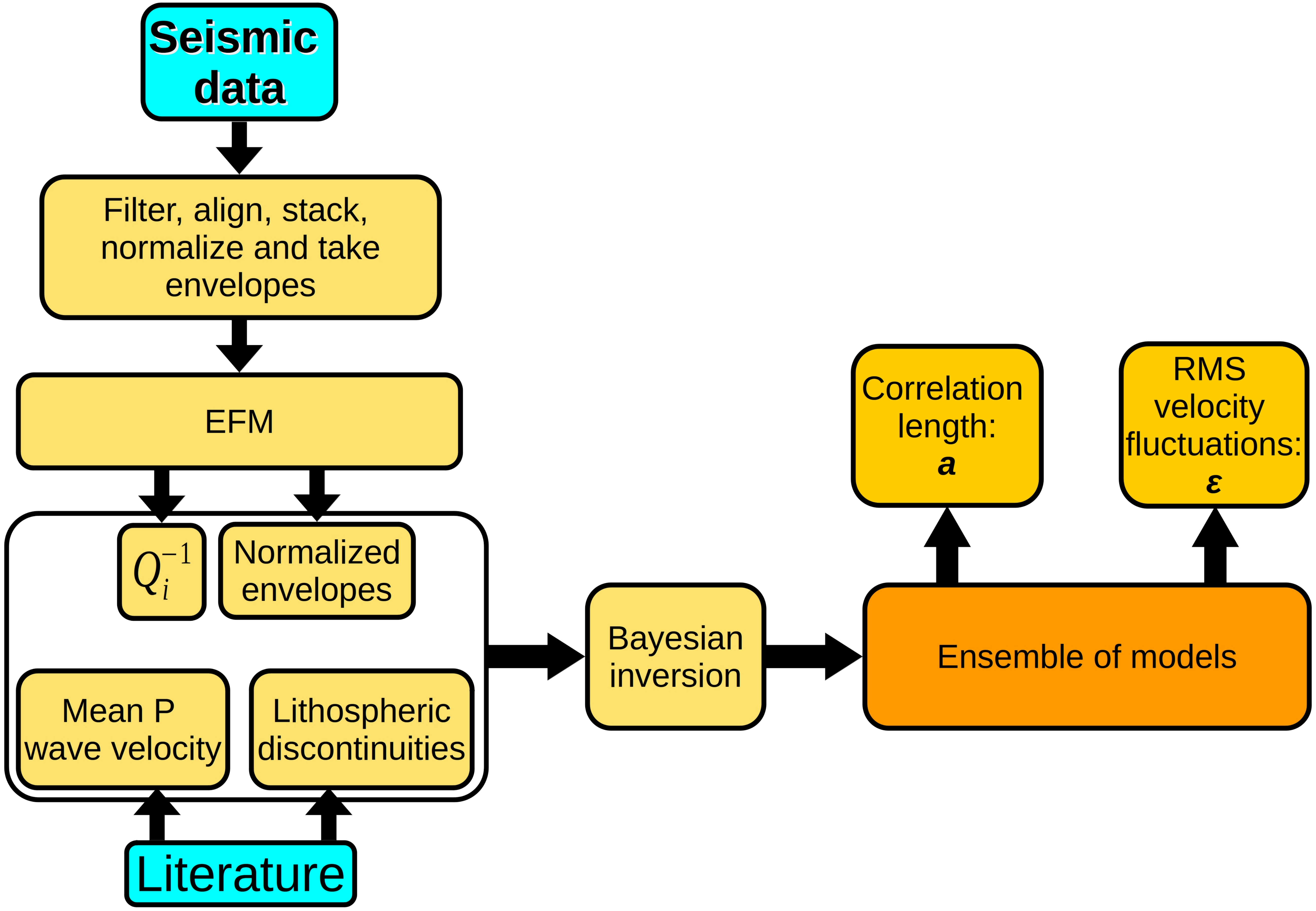


The Bayesian EFMD requires:

- the normalised coda envelopes calculated in the first step of the EFM.
- the intrinsic quality factor value obtained from the EFM inversion.
- a background P wave velocity for each layer in the model.
- crustal and lithosphere-asthenosphere transition depths.

Our Bayesian EFMD uses a Markov chain Monte Carlo (McMC) algorithm: one of the scattering parameters (correlation length or RMS velocity fluctuations) in the initial model is randomly chosen and randomly updated. Synthetic envelopes for the new model are compared with those for the previous model using the likelihood. The new model is accepted if the likelihood is improved, or if the likelihood ratio is larger than a random number in the (0, 1) range. Else, it will be rejected and the algorithm will go back to the previous model.

The depth-dependent Energy Flux Model (EFMD)



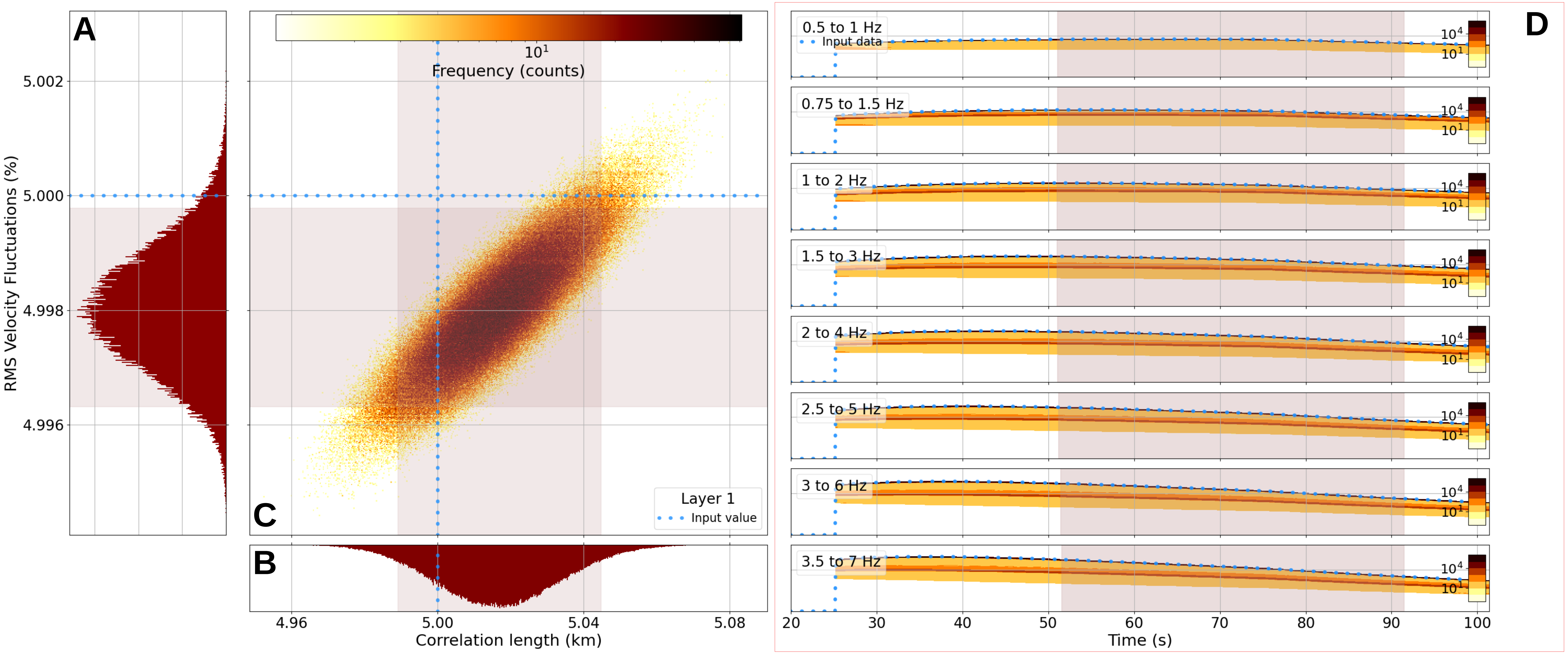
From our ensemble of models, we can obtain the posterior density functions (PDFs) for each parameter and layer.

These PDFs allow us to obtain detailed information about the trade offs and uncertainties in the determination of the heterogeneity parameters. Instead of extracting a single value of each parameter for each layer, we calculate the 5 – 95 percentile range (PR) and use it as our solution in each case.

To test our Bayesian inversion algorithm, we tried to recover the input parameters for one 1-layer model, three 2-layer models and one 3-layer model. For each one, we ran 3 parallel chains, each 1, 2 or 3 million iterations (models tested) long for models with 1, 2 or 3 layers respectively. The same number of chains and iterations were used for our real data inversions.

The depth-dependent Energy Flux Model (EFMD)

Synthetic tests results: model type I

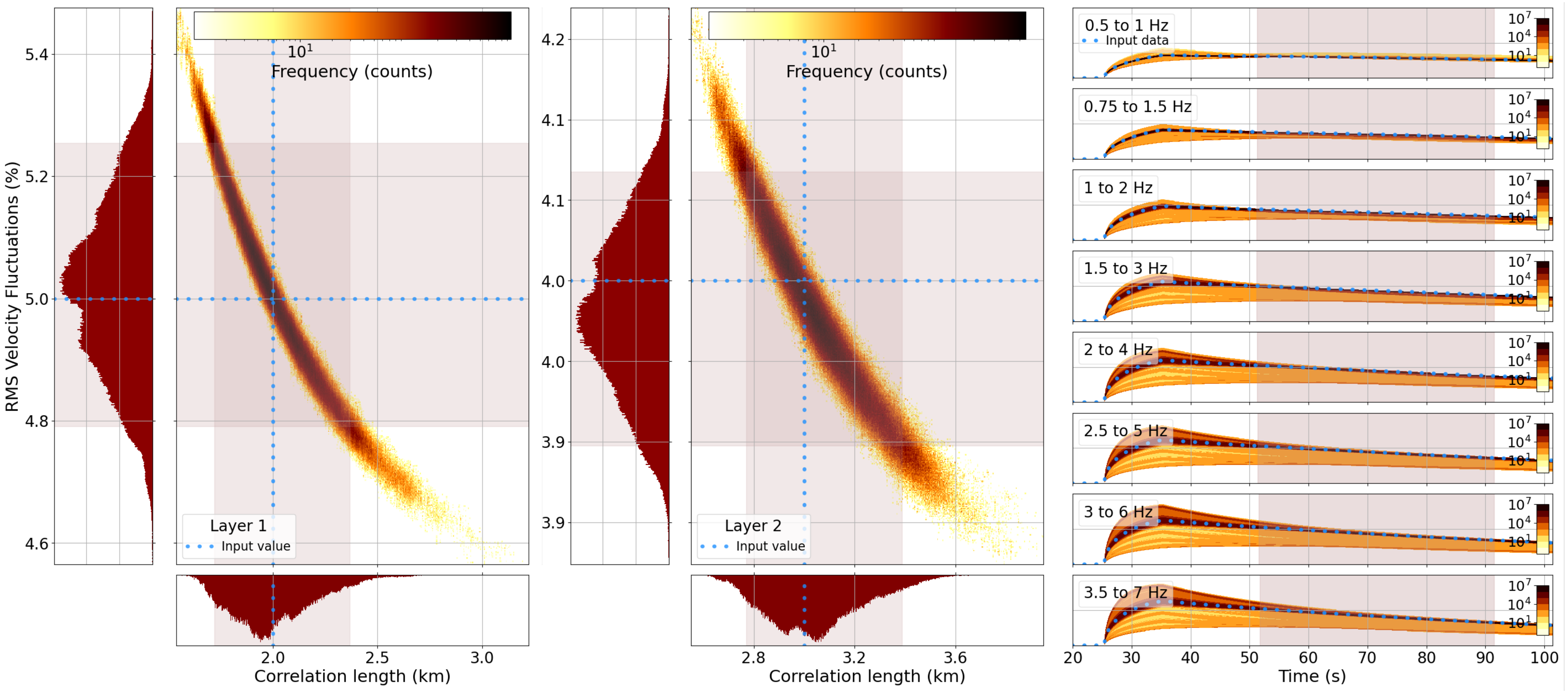


Results for our synthetic test of model type I (1-layer model).

Panel content in all EFMD results figures is as in this one. A-B represent the posterior density functions (PDFs) of the RMS velocity fluctuations (ε) and correlation length (a) respectively, while C shows the joint PDF of both parameters. Dotted blue lines mark the input parameter values and shaded area the 5-95 percentile range (PR). The PDFs are nearly Gaussian, symmetric and extremely narrow, which points to the range of suitable values of the parameters being very well defined. Panels on the right (D) show the input synthetic envelopes (blue dotted lines) together with histograms of the ensemble of synthetic envelopes obtained from all models accepted by our Bayesian inference algorithm. Shaded area represents the time window used for the fit. In all cases, the highest density of envelopes is found in a very narrow zone around the input values, showing that fits to the input data are very good despite the very slight overestimation of a and underestimation of ε (<0.4% of the true value in both cases).

The depth-dependent Energy Flux Model (EFMD)

Synthetic tests results: model type II (similar heterogeneity strength in both layers)

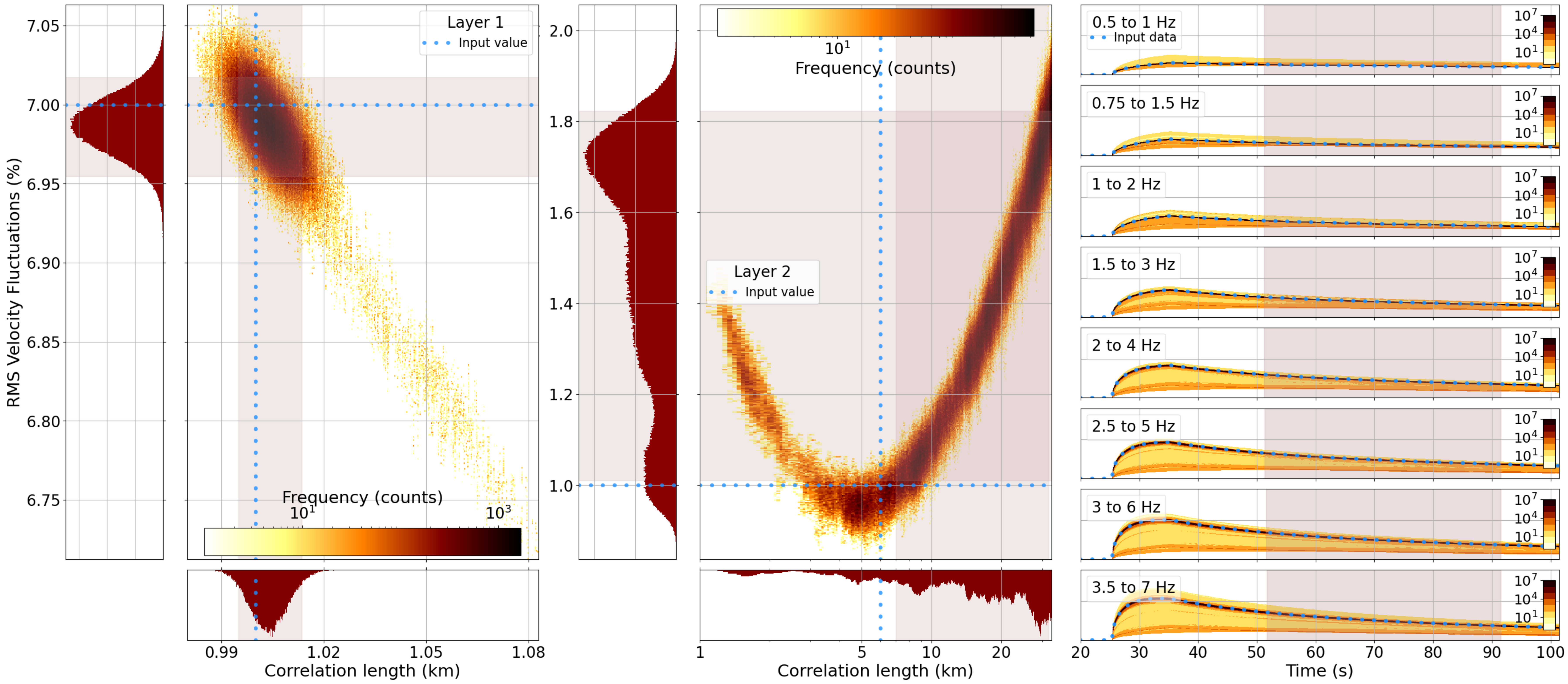


Results for our synthetic test of model type II (2-layer model) with similar heterogeneity strength on both layers.

PDFs for the parameters in both layers are narrow (5–95 PR is < 0.7 km wide for *a* and < 0.5% for *ε*) and approximately centred around the input values, even if they are not Gaussian and show some local maxima. The true values of the parameters lie within the 5–95 PR in all cases, near the centre of the joint PDFs, and the maximum difference between the input values and the absolute maxima of the PDFs is ~2%. Panels on the right indicate fits to the synthetic data are very good, since they show again that the largest concentration of synthetic envelopes for all frequencies coincides with the input data envelopes.

The depth-dependent Energy Flux Model (EFMD)

Synthetic tests results: model type II (strong crustal heterogeneity, weak lithospheric mantle heterogeneity)

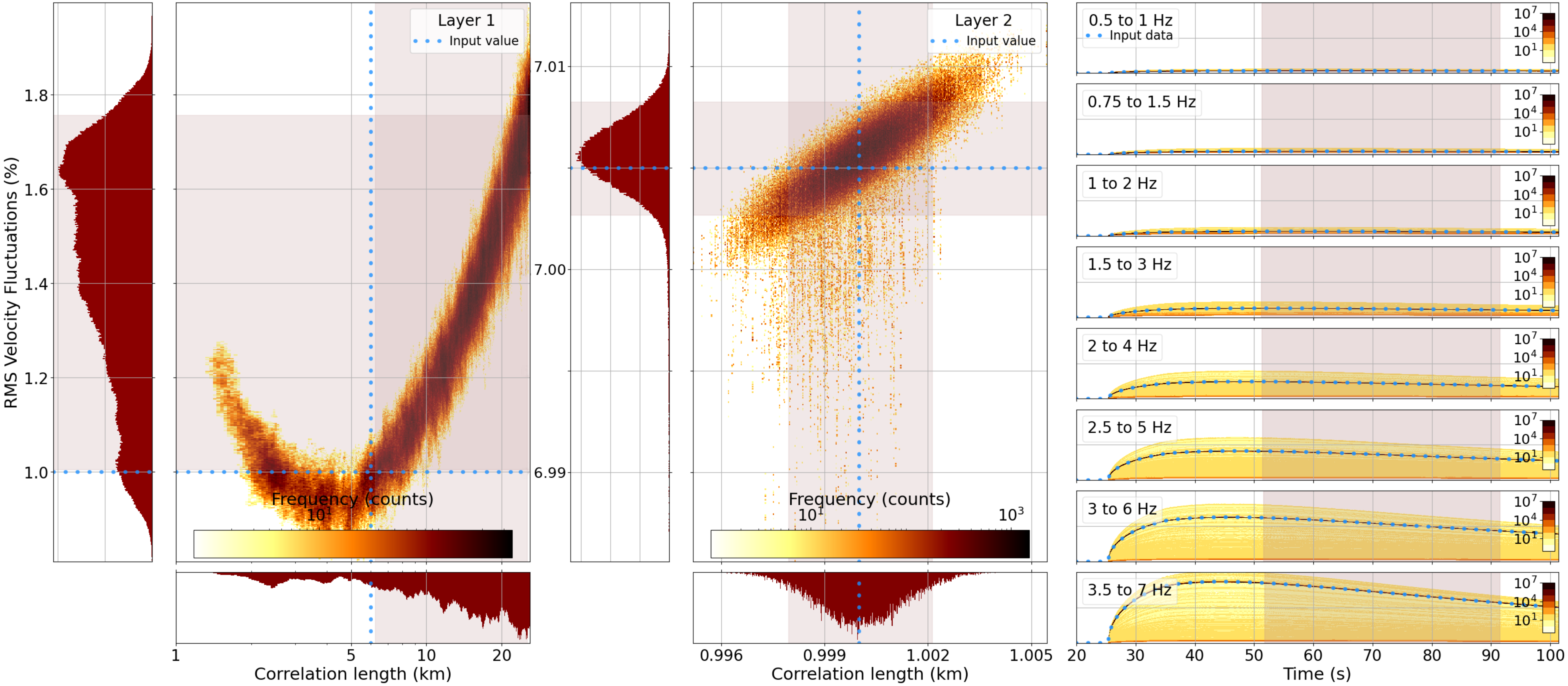


Results for our synthetic test of model type II (2-layer model) and our first high contrast setting (strong crustal heterogeneity vs. weak upper mantle heterogeneity).

PDFs for the parameters in layer 1 are very narrow (< 0.01 km wide for *a* and < 0.07% for ε) and approximately Gaussian. However, solutions for layer 2 are not unique and PDFs have complicated shapes. The algorithm tends to favour higher values of the parameters in this layer, with the true values of the parameters lying slightly outside the 5-95 PRs. Still, panels on the right show that the synthetic envelopes for the vast majority of the accepted models provide really good fits to the input data.

The depth-dependent Energy Flux Model (EFMD)

Synthetic tests results: model type II (weak crustal heterogeneity, strong lithospheric mantle heterogeneity)

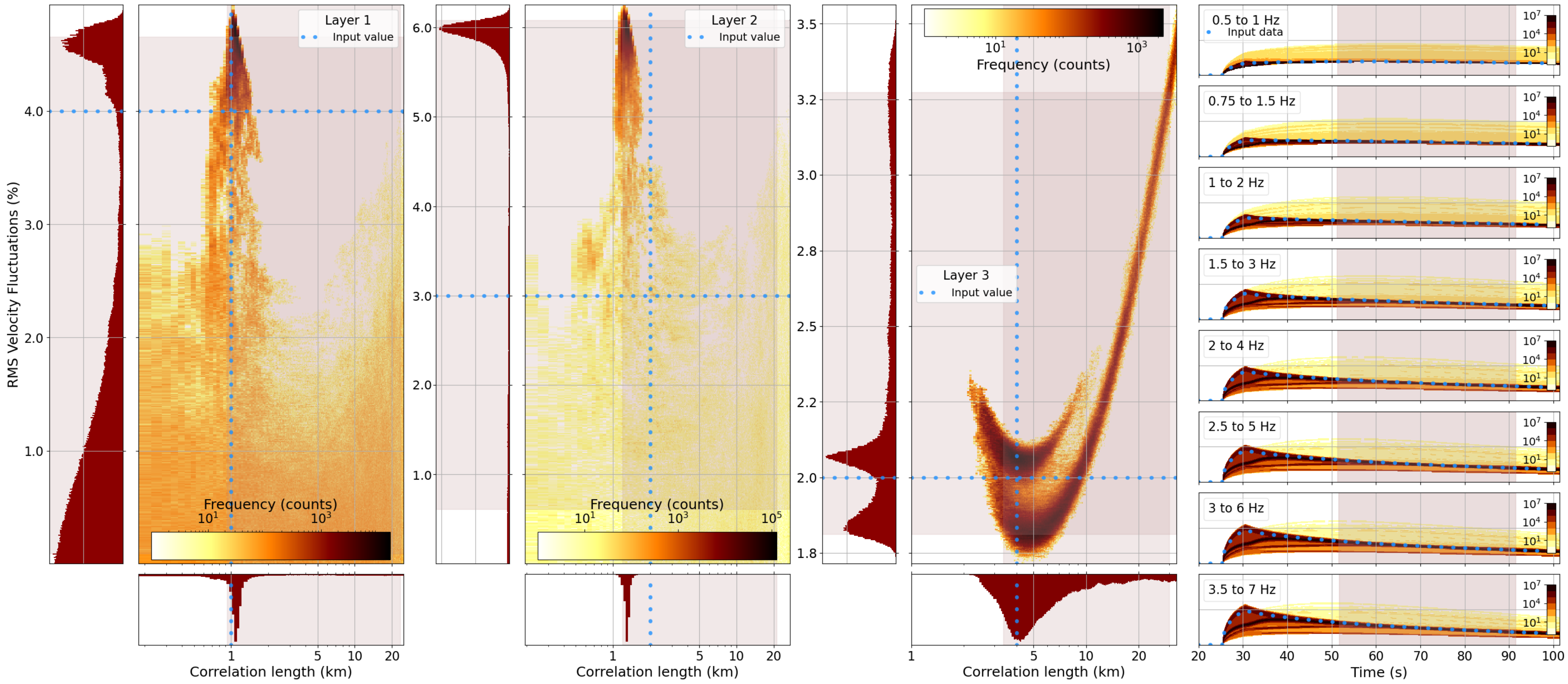


Results for our synthetic test of model type II (2-layer model) and our second high contrast setting (weak crustal heterogeneity vs. strong upper mantle heterogeneity).

PDFs for the parameters in layer 2 are very narrow (~ 0.004 km wide for *a* and ~ 0.004% for ε) and approximately Gaussian. As in our previous test, solutions for layer 1 are not unique and PDFs have complicated shapes. The algorithm favours higher values of the parameters in this layer. Still, panels on the right show that the synthetic envelopes for the vast majority of the accepted models provide really good fits to the input data.

The depth-dependent Energy Flux Model (EFMD)

Synthetic tests results

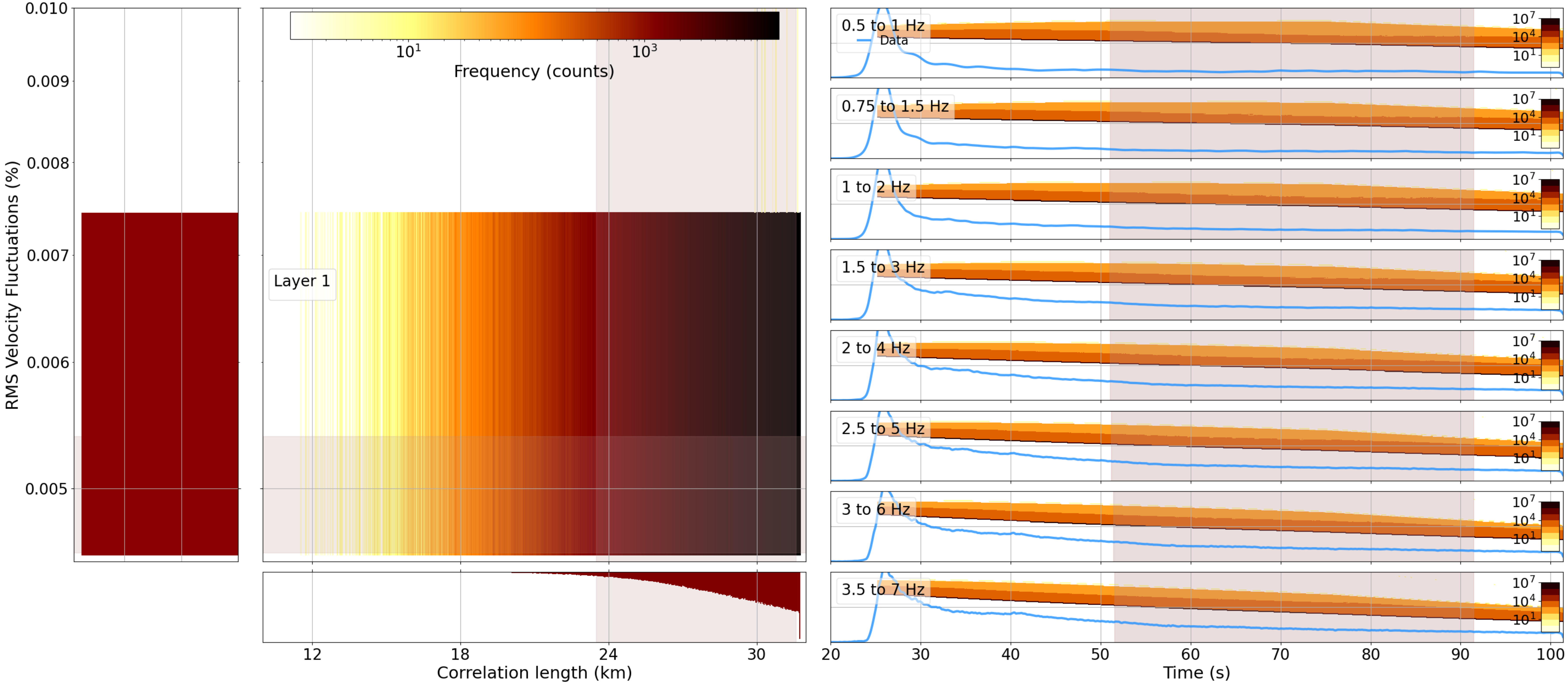


Results for our synthetic test of model type III (3-layer model), combination of the results from three independent chains with a total of 15 million models tested.

PDFs are non-Gaussian and have complex shapes, which widens the 5–95 PR and increases the range of suitable values of the parameters. Correlation length PDFs show clearly defined maxima near the true values of the parameter in all layers. RMS velocity fluctuations PDFs are more complex and do not show clear maxima near the input parameter values. These results show a strong trade-off between parameter values in different layers of the model, especially the two crustal layers, and allows us to identify two independent sets of parameters that provide equally good fits to the data, even if neither of them fully match the input parameter values. This difference between input and recovered parameter values made us decide not to use model type III in our real data inversions.

The depth-dependent Energy Flux Model (EFMD)

Pilbara Seismic Array (PSA) results

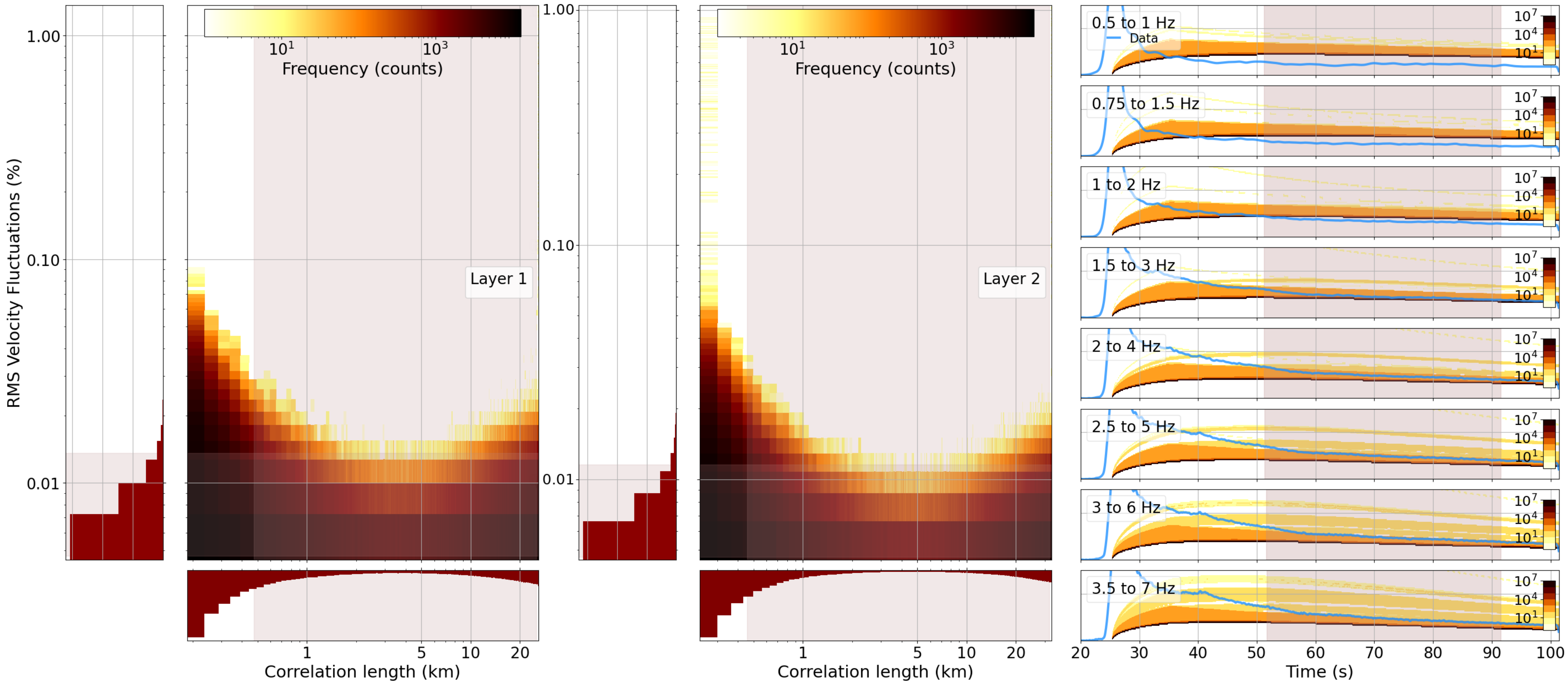


Results for PSA and model type I (1-layer model), combination of the results from three independent chains with a total of 3 million models tested.

This model generates high amplitude codas that very slowly decay over time. The algorithm favoured very high correlation lengths (>23 km) and very low RMS velocity fluctuation (<0.008%) values, which would represent a weakly heterogeneous lithosphere. However, the very low likelihood values obtained during the inversion and the poor fits to the data observed on the panels on the right, make these results unreliable and this model too simple to explain the lithospheric structure beneath this array. Tests of this model type on ASAR data, for which coda amplitudes are higher, yielded similar results.

The depth-dependent Energy Flux Model (EFMD)

Pilbara Seismic Array (PSA) results

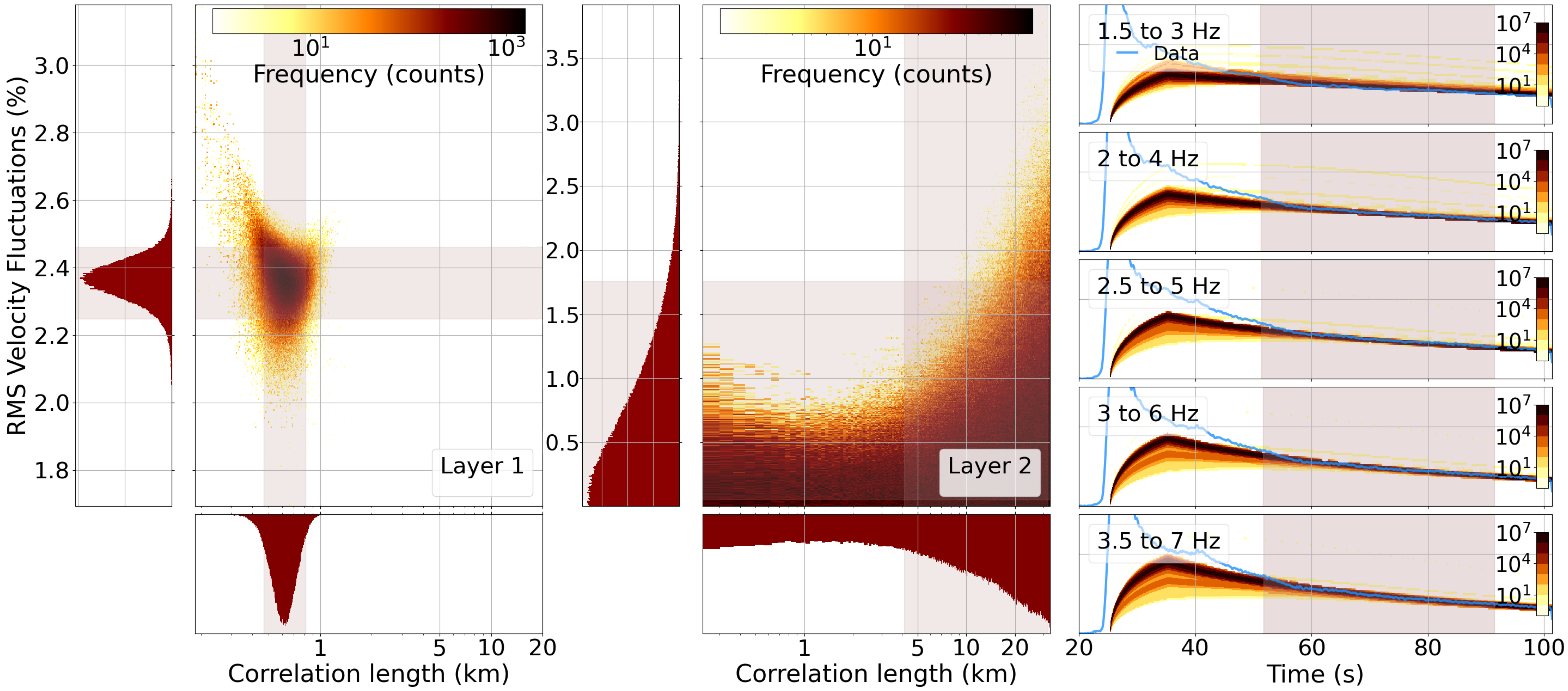


Results for PSA and model type II (2-layer model), combination of the results from three independent chains with a total of 9 million models tested.

Eight frequency bands were used in this inversion. PDFs point to very low ε values in both layers and either low (<1 km) or very large (>10 km) *a* values. Coda fits for the lowest three frequency bands (right panels) are poor, while the algorithm can fit our envelopes at higher frequencies. These results, combined with the very low likelihood values obtained for this case, led us to hypothesize that codas at those frequencies may be affected by large-scale heterogeneities and not composed only of energy scattered at small-scale structures. **Therefore, we used only the five highest frequency bands for the rest of our inversions.**

The depth-dependent Energy Flux Model (EFMD)

Pilbara Seismic Array (PSA) results

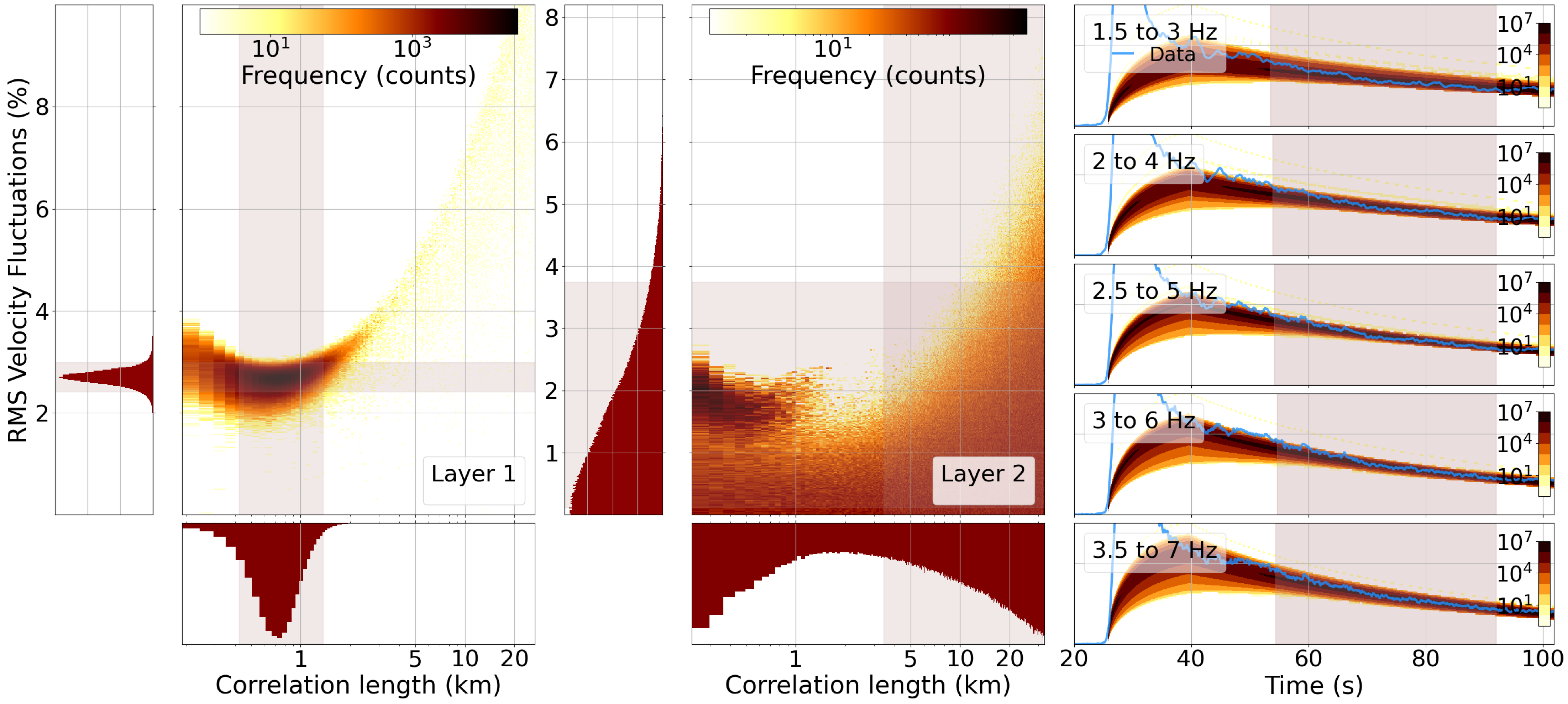


Results for PSA and model type II (2-layer model), and the five highest frequency bands. Combination of three independent chains, 9 million models tested.

PDFs in layer 1 are gaussian, narrow and mostly symmetric, which points to clearly defined parameter values. Correlation lengths are in the 0.5-0.8 km range, while ε take values from 2.3-2.5 %. Parameter values for layer 2 are not unique. The lithospheric mantle appears to be mostly homogeneous, with low values of ε (<2%) being favoured and a slight prevalence of high (>5 km) *a* values, even if any value of this parameter within our range of interest is capable of fitting our data. Likelihood values of all models accepted in this inversion are high, which is also obvious from the good fits to the data shown on the panels on the right of this figure.

The depth-dependent Energy Flux Model (EFMD)

Alice Springs Array (ASAR) results

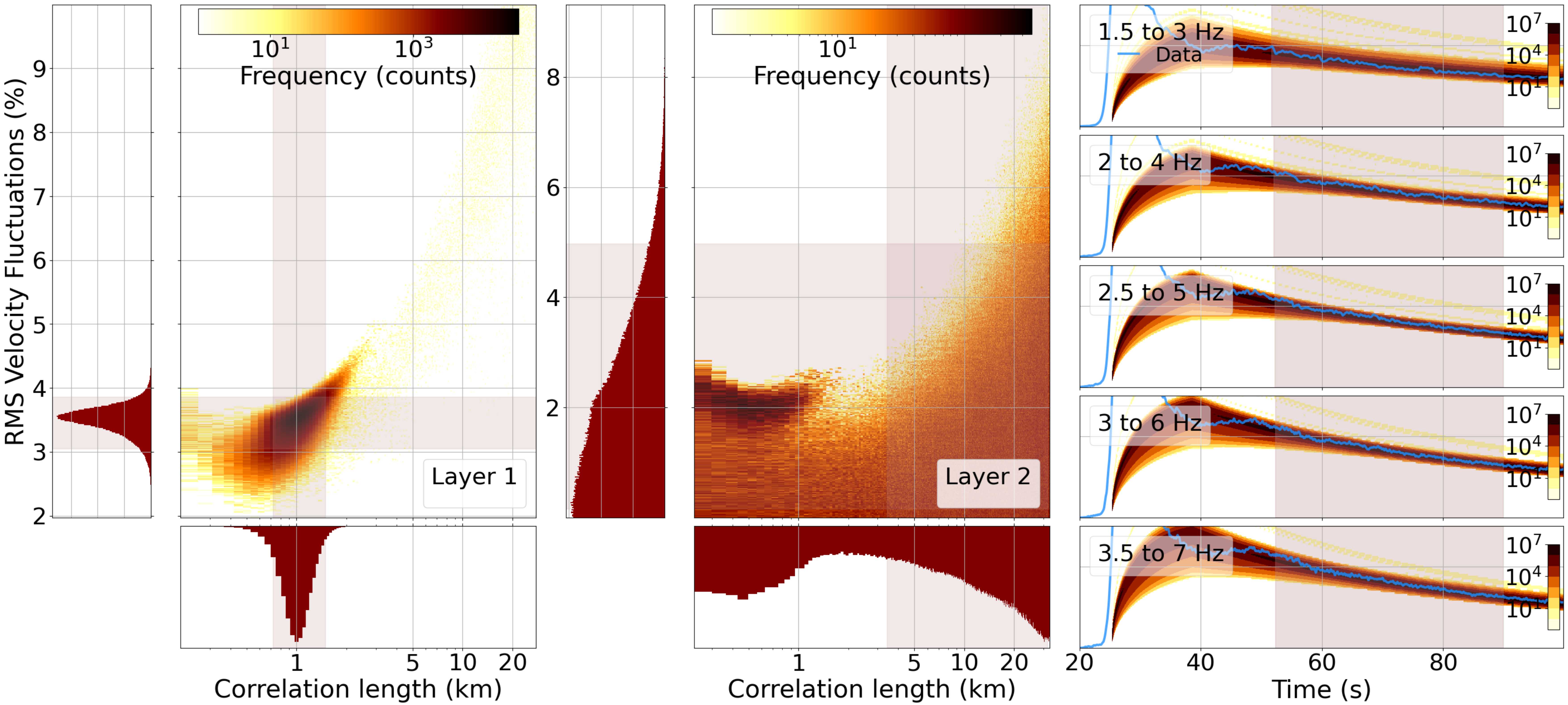


Results for ASAR and model type II (2-layer model), and the five highest frequency bands. Combination of three independent chains, 9 million models tested.

PDFs in layer 1 are nearly gaussian and narrow. Correlation lengths are in the 0.2-1.4 km range, while RMS velocity fluctuations take values from 2.4-3.0 %. As before, solutions are not unique for the lithospheric mantle. ε values below 3.7% are paired with *a* values throughout the entire range considered here, even if the PDF shows similarly high peaks on both ends (<1 km and >5 km). Fits to the data are good for these frequencies and likelihood values for this inversion are high.

The depth-dependent Energy Flux Model (EFMD)

Warramunga Array (WRA) results



Results for WRA and model type II (2-layer model), and the five highest frequency bands. Combination of three independent chains, 9 million models tested.

PDFs for the crust are nearly gaussian and narrow. Correlation lengths are in the 0.7-1.5 km range, while RMS velocity fluctuations take values from 3.1-3.9 %. For layer 2, solutions are not unique. Any correlation length within our range of interest can provide good fits to the data, even if there's a slight preference for high (>5 km) values over lower values. The 5-95 PR for ε in this layer extends up to 5%. High likelihood values obtained from this inversion demonstrate the quality of the fits to our data, shown on the panels on the right.

Conclusions

We combined energy flux models with a new Bayesian inference algorithm and applied them to a large, high-quality dataset for three seismic arrays in Australia (ASAR, WRA and PSA). Our results:

- show that the structure beneath ASAR is the most attenuating and heterogeneous one.
- show that the heterogeneity and overall attenuation structure for ASAR and WRA is comparable and different to PSA.
- show that scattering is the dominant attenuation mechanism above ~2Hz for all arrays.
- suggest the crust is more heterogeneous than the lithospheric mantle for all arrays. Correlation lengths in the crust vary from ~0.2-1.5 km and RMS velocity fluctuations take values in the 2-4 % range. The lithospheric mantle structure is more complex and solutions for this layer are not unique.
- are consistent with the tectonic histories of the areas the arrays are located on.
- agree with previous studies in these areas using different techniques.

Conclusions

The combination of the EFM and Bayesian EFMD is an effective and useful tool for the characterization of the near receiver, small-scale heterogeneity structure of the lithosphere. We have shown that this approach:

- allows us to compare different attenuation mechanisms.
- can be used both for seismic arrays and single seismic stations.
- can be used for seismically quiet areas, since it requires teleseismic data.
- can be used for weak and strong scattering regimes.
- allows us to use a Bayesian inference algorithm. The resulting PDFs of the heterogeneity parameters (correlation length and RMS velocity fluctuations), provide detailed information about the parameter space, as well as the trade offs and uncertainties in the determination of the parameters.

Want to know more?

Check out our vPICO session on Monday, April 26, from 9:00 – 10:30

or

Have a look at our preprint, submitted and currently under review with Geophysical Journal International, at https://doi.org/10.31223/X5S89Q