

Bayesian analysis of the chemical composition of float glass

Joe Zemmels

4/13/2020

Introduction

Determining whether two fragments of glass originate from the same pane is a useful endeavor in the field of forensics. For example, matching a fragment of glass taken from a crime scene to one found on a subject would provide evidence that the suspect was at the scene of the crime at one point in time. It is possible to measure the elemental composition of a particular fragment of glass. The elemental composition can be measured for a variety of elements (Sodium (NA), Titanium (TI), etc.). It is widely believed in the forensics community that two fragments of glass can be “matched” to the same pane based on the similarity of their elemental compositions. Park et al. (2018, 2020) collected data from two glass manufacturing companies on the elemental compositions of panes of float glass (a particular method of manufacturing glass) sampled over a period of a few weeks. They used the data to demonstrate the efficacy of supervised learning methods to classify glass fragments as matching or non-matching.

Park & Carriquiry (2018) briefly discuss the apparent time-dependent structure of the elemental concentrations for the panes of glass sampled from the two companies. For example, for many of the elements for which the concentration was measured, panes sampled closer together in time appear to have more similar concentrations than panes sampled further apart in time. The analysis in this report will focus on describing the variability of the elemental concentrations over time for one of the two companies.

Data

The data come from two float glass manufacturers, referred as Company A and Company B in the dataset, taken over a period of 3 weeks (January 3rd - January 24, 2017) and 2 weeks (December 5th - December 16th, 2016). On nearly every day during these two periods, two panes of glass were collected from both sides of a glass “ribbon,” which is the form that float glass is commonly manufactured into. This resulted in a total of 31 panes from Company A, denoted $\{AA, AB, \dots, AAR\}$ in the data, and 17 panes from Company B, denoted $\{BA, \dots, BR\}$. Many of the Company B panes were sampled at disparate times. An example of the recorded concentrations of the element iron (^{57}Fe) for panes sampled from Company B is shown in Figure 1 (with analogous plots of other elements looking similar). We can see that the sampling rate of the panes is infrequent and the number of panes sampled per sampling event unbalanced. Thus, for the sake of the analysis, we will only focus on the elemental concentrations for panes sampled from Company A.

Each pane was broken into fragments and 24 fragments were randomly selected. These are indexed 1-24 in the data. For 21 of these 24 fragments, the elemental composition was repeatedly measured 5 times. The remaining 3 fragments had their elemental composition measured 20 times. Thus for each pane of glass, we have 165 elemental concentration measurements. Each of these “reps” are represented as a row in the dataset. The elemental concentrations were recorded in parts/million (micrograms/gram) for the 18 elements. These 18 elements are: lithium (^7Li), sodium (^{23}Na), magnesium (^{25}Mg), aluminum (^{27}Al),

silicon (^{29}Si), potassium (^{39}K), calcium (^{42}Ca and ^{43}Ca), titanium (^{49}Ti), manganese (^{55}Mn), iron (^{57}Fe), rubidium (^{85}Rb), strontium (^{88}Sr), zirconium (^{90}Zr), tin (^{118}Sn), barium (^{137}Ba), lanthanum (^{139}La), cerium (^{140}Ce), neodymium (^{146}Nd), hafnium (^{178}Hf) and lead (^{208}Pb).

Park & Carriquiry (2018) consider the average log-transformed elemental concentration measurement values for each fragment when constructing features for their supervised learning classifiers. That is, they first log-transform the measurements and then take an average per fragment. The same will be done for this analysis. A plot of densities of the log-transformed averaged elemental concentrations for the fragments from three of the 18 total elements are in Figure 2. We will focus on only three elements for this analysis because considering all 18 was deemed too computationally intensive (for my computer, at least). The plots in Figure 2 show the distribution of elemental concentrations for each pane. This plot is meant to convey that the concentrations of certain elements vary considerably between panes. For modeling purposes, it would be interesting to determine how time-dependent these concentrations are.

Plotting the log-transformed elemental concentrations by pane uncovers the temporal structure for some of these elements discussed in Park & Carriquiry (2018) and Park & Tyner (2019). Figure 4 shows a scatterplot of the log-transformed mean elemental concentrations for each of the 24 fragments sampled from each pane against the date in which these fragments were sampled. The solid curves connect the within-date averages. For certain elements, such as Hafnium (^{178}Hf) and Zirconium (^{90}Zr), we can see what appears to be a trend. This may indicate, for example, that the availability of these elements changed over the period of time for which samples were taken from Company A.

Methods

As discussed above, there may be some time dependency for some of the elemental concentrations. A temporal model, such as an autoregressive model, for the elemental concentrations may be useful for explaining the behavior. We can compare these results to a model fit without assuming a temporal structure (similar to what is used in Park & Carriquiry (2018)). We will thus divide our analysis into an atemporal, “pane-specific” model and an autoregressive model. Note that we do not have observations from every day of the 3 week period during which samples were taken. We will discuss how to deal with these missing observations below. Samples from the posterior will be simulated using the `rstan` package. We will assess convergence using standard tools such as \hat{R} and effective sample size as well as diagnostics available for the specific “No U-Turn Sampler” (NUTS, Hoffman and Gelman, 2014) used by Stan if divergent transitions occur.

Pane-specific model

Adapting notation from Park & Carriquiry (2018), we will let $\bar{y}_{ijk\cdot}$ be the mean log of the observed concentration of the i th element across all measurements of the k th fragment from the j th pane, $i = 1, 2, 3$ (Iron, Hafnium, Zirconium, respectively), $j = 1, \dots, 31$, $k = 1, \dots, 24$, where the dot indicates averaging over a fourth index associated with measurement (which we’re not considering for this analysis). We will

associate with the mean log concentrations a collection of random variables $\{\bar{Y}_{ijk}\}$ where for $k = 1, \dots, 24$,

$$\begin{aligned}\bar{Y}_{ijk}|\mu_{ij}, \sigma_i^2 &\stackrel{ind}{\sim} N(\mu_{ij}, \sigma_i^2) \\ \mu_{ij} &\stackrel{d}{=} \eta_i + \tau_i \theta_{ij} \\ \eta_i &\stackrel{ind}{\sim} p(\eta_i) \propto 1 \\ \tau_i^2 &\stackrel{ind}{\sim} p(\tau_i^2) \propto I(\tau_i > 0) \\ \theta_{ij} &\stackrel{iid}{\sim} N(0, 1) \\ \sigma_i^2 &\stackrel{iid}{\sim} \text{Inv} - \chi^2(1, 1).\end{aligned}$$

The interpretation of the data model specified is that the log mean concentration of each fragment differs by location based on the element/pane combination considered, yet will differ in scale only by which element is considered (constant variance within-element). Note that μ_{ij} is defined in this way because another model in which $\mu_{ij} \stackrel{ind}{\sim} N(\eta_i, \tau_i^2)$ was explored yet was found to result in divergent transitions during the Hamiltonian Monte Carlo sampling procedure. Specifically, the phenomenon known as “Neal’s funnel” was clearly observed when plotting τ_1^2 against any μ_{1j} for $j = 1, \dots, 31$. The parameterization used above was found to be more successful.

Autoregressive model

We are interested in comparing how the “pane-specific” model described above compares to a model in which elemental concentrations depend on each other through time. Tweaking the notational conventions introduced for the pane-specific model, we will now let $\bar{y}_{ijk}(t)$ denote the log concentration of the observed concentration of the i th element averaged over all measurements of the k th fragment from the j th pane on the t th day, $i = 1, 2, 3$, $j = 1, \dots, 34$, $k = 1, \dots, 24$, $t = 1, \dots, 22$. Note that j is now indexed from 1 to 34 because, as discussed below, data will be simulated for days in which no panes were sampled. We will associate with these observed mean log concentrations a set of random variables, $\{\bar{Y}_{ijk}(t)\}$ in which we assume for $k = 1, \dots, 24$, $t = 2, \dots, 22$

$$\begin{aligned}\bar{Y}_{ijk}(t)|\mu_{ij}(t), \sigma_i^2 &\stackrel{ind}{\sim} N(\mu_{ij}(t), \sigma_i^2) \\ \mu_{ij}(t) &\stackrel{d}{=} \eta_i + \gamma_i(t) + \tau_i \theta_{ij} \\ \eta_i &\stackrel{ind}{\sim} p(\eta_i) \propto 1 \\ \gamma_i(t) &\stackrel{d}{=} \alpha_i \gamma_i(t-1) + \epsilon_i(t) \\ \alpha_i &\stackrel{iid}{\sim} p(\alpha_i) \propto 1 \\ \epsilon_i(t) &\stackrel{iid}{\sim} N(0, 1) \\ \tau_i &\stackrel{ind}{\sim} p(\tau_i) \propto I(\tau_i > 0) \\ \theta_{ij} &\stackrel{iid}{\sim} N(0, 1) \\ \sigma_i^2 &\stackrel{iid}{\sim} \text{Inv} - \chi^2(1, 1).\end{aligned}$$

For $t = 1$, since we don’t have a previous observation to condition on, we will assume that $\mu_{ij}(1) \stackrel{d}{=} \eta_i + \epsilon_i(1) + \tau_i \theta_{ij}$. Note that this model is effectively equivalent to the pane-specific model if $\alpha_i = 0$, $i = 1, 2, 3$; that is, there is no dependence on time. This model is meant to reflect the observation made that the expected elemental concentrations for particular days may depend on each other through time. As noted above, there are 3 days, 1/11, 1/12, and 1/23, during which a pane was not sampled. We will simulate a

pane-worth of data ($24 \cdot 3$ observations) for each of these days from a normal with expectation equal to the linearly interpolated mean at these days and variance based on the posterior mean σ_i^2 values estimated from the pane-specific model. After some exploration, it was determined that the estimated posterior mean σ_1^2 (i.e., the iron concentration variance) was inflated due to a handful of extreme observations that we can see in the time series plot in Figure 3. Thus, we scaled the variances down for the estimated posterior mean σ_1^2 value until the simulated values “looked like” the observed iron concentration values (which is rather hacky, so I would be interested in determining a better way to deal with this).

Results

Pane-specific model

For the pane-specific model a total of 4 chains were run each with 5000 burn-in and 5000 inferential samples. After re-parameterizing the model, no divergent transitions occurred indicating that traversal of the parameter space was successful. We can see from Figure 5 that the scale reduction factor values are all well-behaved indicating that the chains mixed well. Traceplots (not shown for sake of space) provided no clear indication of convergence failure.

The effective sample sizes for some of the parameters, specifically those in Table 1, indicate some autocorrelation among the inferential samples. We can see in particular that η_2 and η_3 , which act as the expectations on the data model means for the Hafnium and Zirconium concentrations, have a small effective sample size. However, these do still “pass” the rule-of-thumb discussed by Gelman et al. in BDA3 pg. 288 of running a chain until the effective sample size of all parameters of interest exceed $10c$ for c chains. We can also see based on the ACF plots in for η_2, η_3, τ_2 , and τ_3 shown in Figure 6 that the sample autocorrelation for these parameters does not decay as quickly as we would like. The Discussion section below discusses this further.

The estimated densities of the μ_{ij} terms based on the inferential samples are shown in Figure 7. Given the uninformative priors placed on the parameters governing the μ_{ij} s, it makes sense that the densities are located similar to those of the observed data for each pane as shown in Figure 3. The estimated densities for the “element-specific” parameters (i.e., parameters indexed only by $i \in \{1, 2, 3\}$) are shown in Figure 8. While improper uniform priors were placed on η_i and $\tau_i, i = 1, 2, 3$, we can see that the shapes of the distribution are very much unimodal (although an odd “bump” can be seen in the plot of τ_1 near 0).

Autoregressive model

Similar to the pane-specific model, 4 chains with 5000 burn-in and 5000 inferential samples were used to fit the AR(1) model. No divergent transitions occurred and, as we can see by Figure 9, the scale reduction factors all look passable. Again, traceplots did not indicate convergence failure. Table 2 shows which parameters had an associated effective sample size ratio less than .1. The effective sample sizes for the other model parameters were at least greater than .01 and all parameters again passed the rule-of-thumb for effective sample size discussed in BDA3. This indicates that samples for certain parameters may not be particularly “trustworthy.” This will be expanded upon in the Discussion section below. Figure 10 shows the posterior element/pane-specific credible intervals over time. Again, we see that the μ_{ij} s tend to be located near their associated element/pane-specific observations.

Figure 11 shows the distributions of the 3 elements’ AR(1) parameters, α_i . Recall that these were constrained to be between -1 and 1 (inducing a stationary AR(1) process). We can see that the distribution of the AR(1) parameter associated with Iron is fairly diffuse on this interval while those of Hafnium and Zirconium are quite concentrated near 1. This makes sense if we interpret the trajectory of the mean concentrations shown in Figure 3 as that of a near-random walk temporal process. These results indicate that there *is* some

temporal dependence for some elemental concentrations, which answers our question of interest. We also now have the ability to better describe this dependence through use of the autoregressive model. Whether such a model generalizes to other elements would be interesting to explore further.

Discussion

As discussed above, the ACF plots for the Hafnium and Zirconium low-level parameters under the pane-specific model do not decay as quickly as we would like. Alongside the fact that τ_2 and τ_3 also have relatively low effective sample sizes, we may take this as a sign that the μ_{ij} s are misspecified under the pane-specific model, at least for $i = 2, 3$. We can clearly see in the time series plots in Figure 4 that Hafnium and Zirconium concentrations exhibit some pattern in time, so it makes sense that enforcing a common mean structure in η_2 and η_3 may not “agree” with the observed Hafnium and Zirconium data. Similarly, for the autoregressive model, the effective sample size seems to be particularly small for parameters associated with data that aren’t “agreeable” with what the parameter represents. To illustrate this, consider Table 2 containing the only parameters for which the effective sample size ratio was less than .1. We can see that they are associated with Iron which, if you recall, exhibited the least obvious temporal pattern out of the three elements considered. Again, the effective sample size may be an indication of data behavior that our model fails to adequately describe.

As noted above, the AR(1) coefficients for Hafnium and Zirconium seem to indicate extreme positive dependence between mean concentrations over time. We could, however, try to also explain these data using, say, a quadratic regression on time (such a model as well as a hybrid AR(1) and regression model were attempted, but found to be too computationally taxing for my computer). The key difference between these two models, I believe, is what we want the “data generating mechanism” to be represented as. The method by which these *float glass* panes are manufactured yields a continuous “ribbon” of glass that is eventually cut into various sizes. Perhaps the raw materials fed into the furnace that produces this ribbon contained a considerable amount of Hafnium and Zirconium at the beginning of the sampling period that began to “run out” near the end of the sampling period. It might make sense then to use an AR model representing the idea that, for example, using up Hafnium-rich raw material on one day may *cause* there to be less Hafnium in manufactured panes on the next day. Regression parameters with respect to time may not have as simple a physical interpretation but may “explain” the data better overall. Exploring other temporal models may be interesting future work. Additionally, as noted in Park & Carriquiry (2018), the concentrations of certain elements appear to be highly correlated with each other (e.g., Neodymium tends to show up in high concentrations with Hafnium). Another interesting area for exploration might be to model this between-element dependence using, say, a vector autoregression (the `bvar` package performs Bayesian vector autoregression).

Lastly, while we have determined and described some of the temporal structure in the data, we haven’t *really* determined which proposed model actually fits the data better. Determining Bayes Factors from MCMC output isn’t particularly straightforward, but the `bridgesampling` package does allow for some estimation based on `rstan` (among other packages) output. This may be interesting to look into as well to make more formal, hypothesis test-based conclusions about which model fits the data best.

References

- Gelman, A, J. Carlin, et al. (2013). Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis. ISBN: 978-1-4398-4095-5.
- Homan, M. D. and A. Gelman (2014). “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo”. In: J. Mach. Learn. Res. 15.1, p. 1593–1623.

- Park, S. and A. Carriquiry (2019). “Learning algorithms to evaluate forensic glass evidence”. In: Ann. Appl. Stat. 13.2, pp. 1068-1102. DOI: 10.1214/18-AOAS1211.
- Park, S, A. Carriquiry, et al. (2020). “A database of elemental compositions of architectural float glass samples measured by LA-ICP-MS”. In: Data in Brief 30, p. 105449. DOI: <https://doi.org/10.1016/j.dib.2020.105449>.
- Park, S. and S. Tyner (2019). “Evaluation and comparison of methods for forensic glass source conclusions”. In: Forensic Science International 305, p. 110003. DOI: <https://doi.org/10.1016/j.forsciint.2019.110003>.

Tables

Table 1: Parameters with low effective sample size under pane-specific model.

| diagnostic | parameter | value | rating | description |
|------------|-----------|-----------|--------|----------------------------|
| neff_ratio | eta[3] | 0.0666882 | low | $N[\text{eff}]/N \leq 0.1$ |
| neff_ratio | eta[2] | 0.0720064 | low | $N[\text{eff}]/N \leq 0.1$ |
| neff_ratio | tau[2] | 0.1286298 | ok | $N[\text{eff}]/N \leq 0.5$ |
| neff_ratio | tau[3] | 0.1358744 | ok | $N[\text{eff}]/N \leq 0.5$ |
| neff_ratio | lp____ | 0.1577378 | ok | $N[\text{eff}]/N \leq 0.5$ |
| neff_ratio | tau[1] | 0.2204662 | ok | $N[\text{eff}]/N \leq 0.5$ |

Table 2: Parameters with low effective sample size under autoregressive model.

| param | description | n |
|-------------|----------------------------|----|
| epsilonIron | $N[\text{eff}]/N \leq 0.1$ | 28 |
| etaIron | $N[\text{eff}]/N \leq 0.1$ | 1 |
| gammaIron | $N[\text{eff}]/N \leq 0.1$ | 34 |

Figures

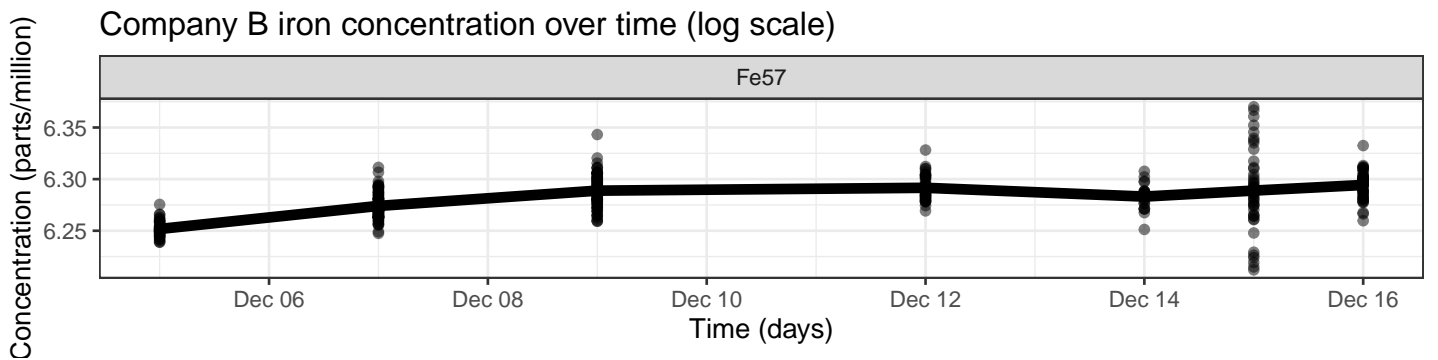


Figure 1: Iron concentrations from Company B over time

Elemental concentration distributions for company A (log scale)

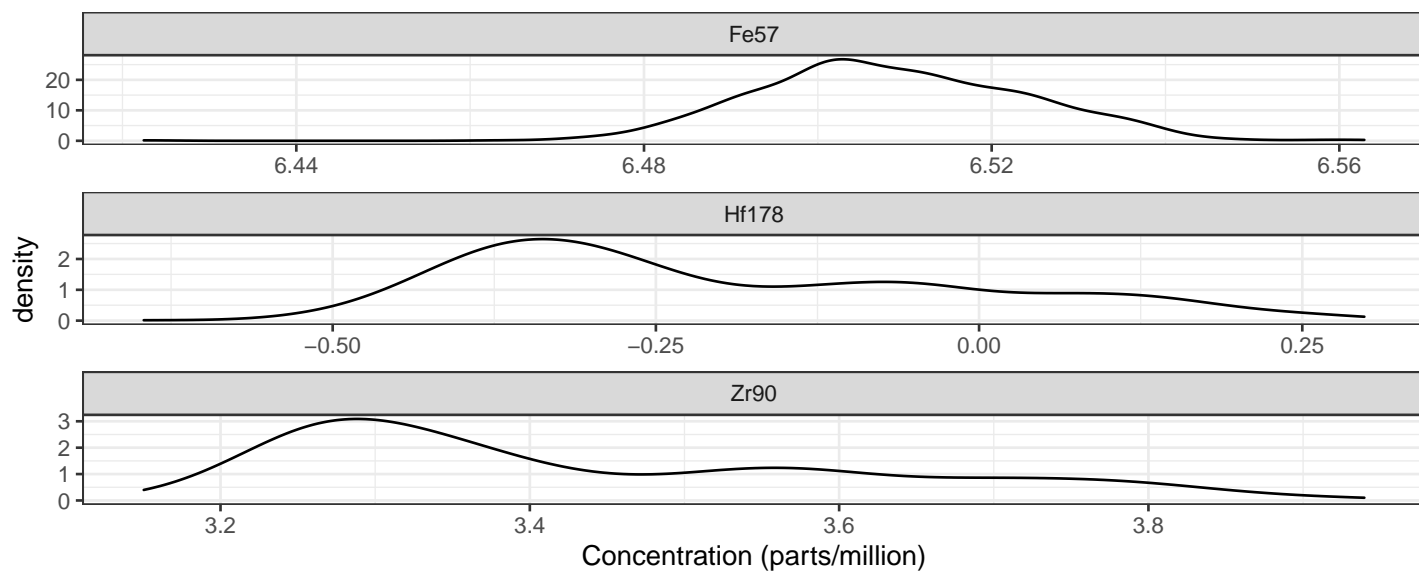


Figure 2: Company A element-specific concentration distributions for three elements.

Company A elemental conc. distribution per pane (log scale)

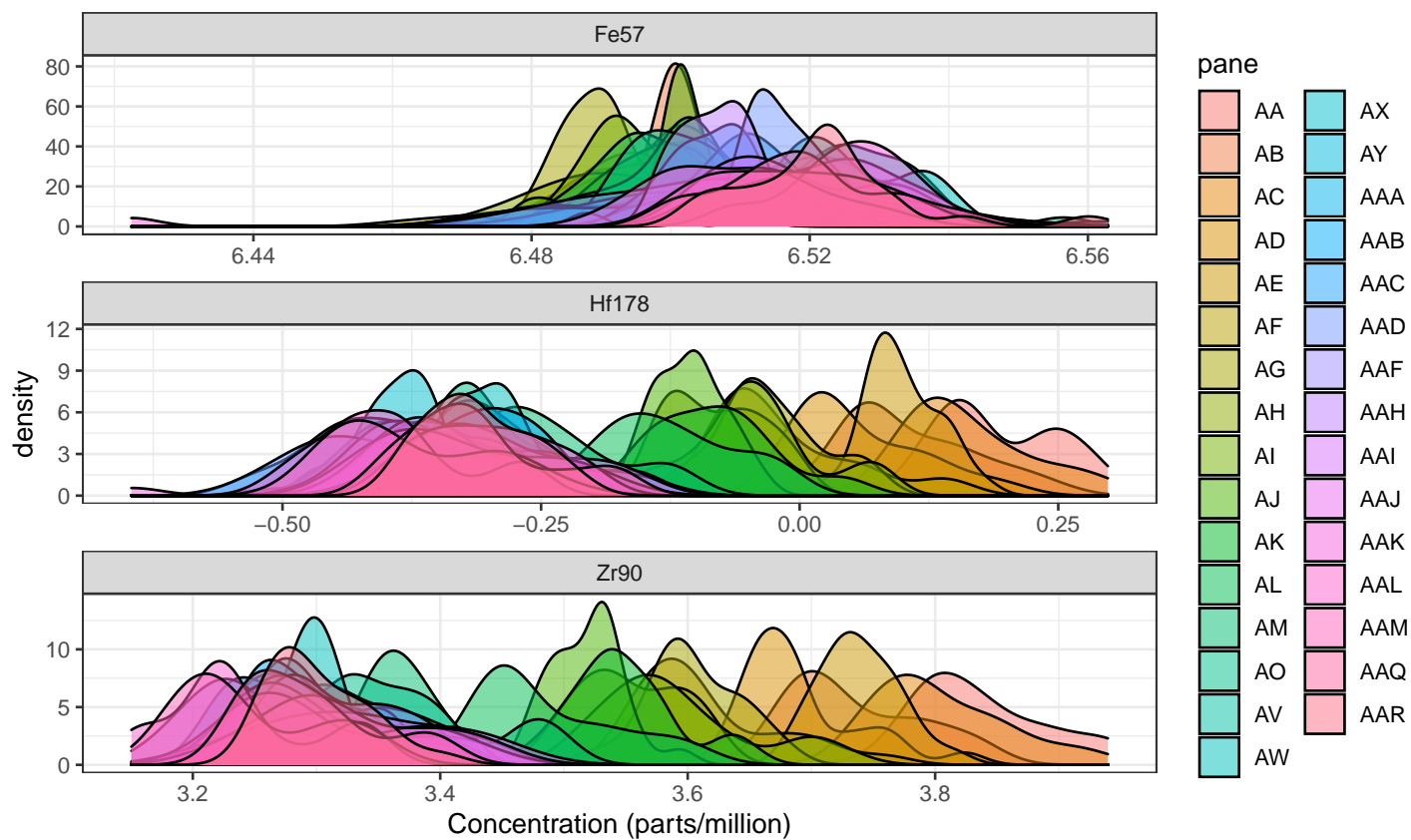


Figure 3: Company A pane-specific concentration distributions for three elements.

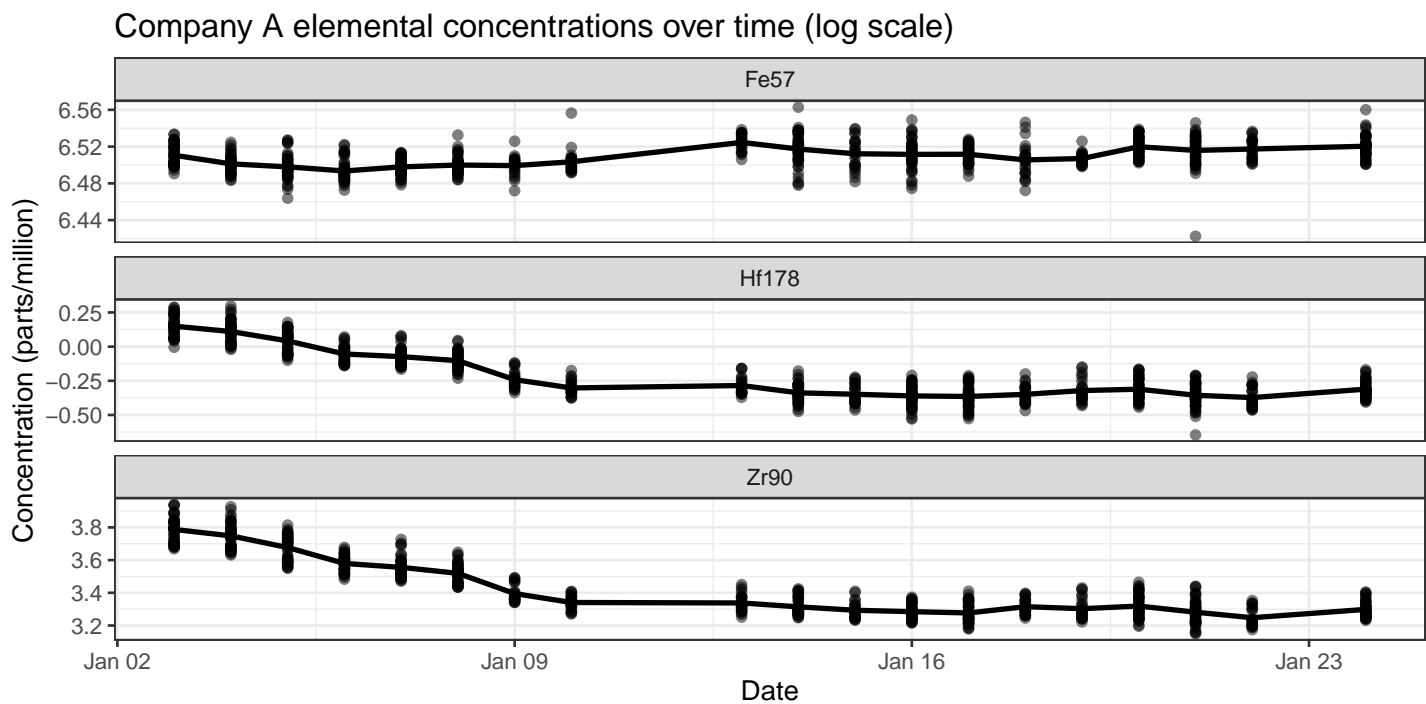


Figure 4: Company A concentrations over time for three elements.



Figure 5: Scale reduction factor values for parameters in pane-specific model.

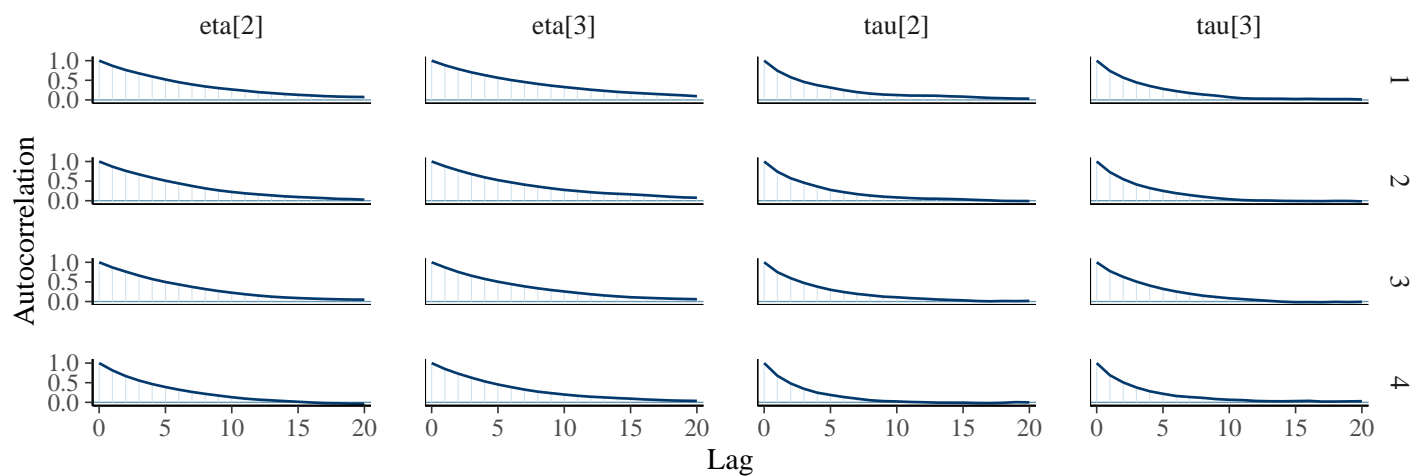


Figure 6: ACF plots for low-level Hafnium and Zirconium parameters in pane-specific model.

Posterior distributions for μ_{ij} under pane-specific model

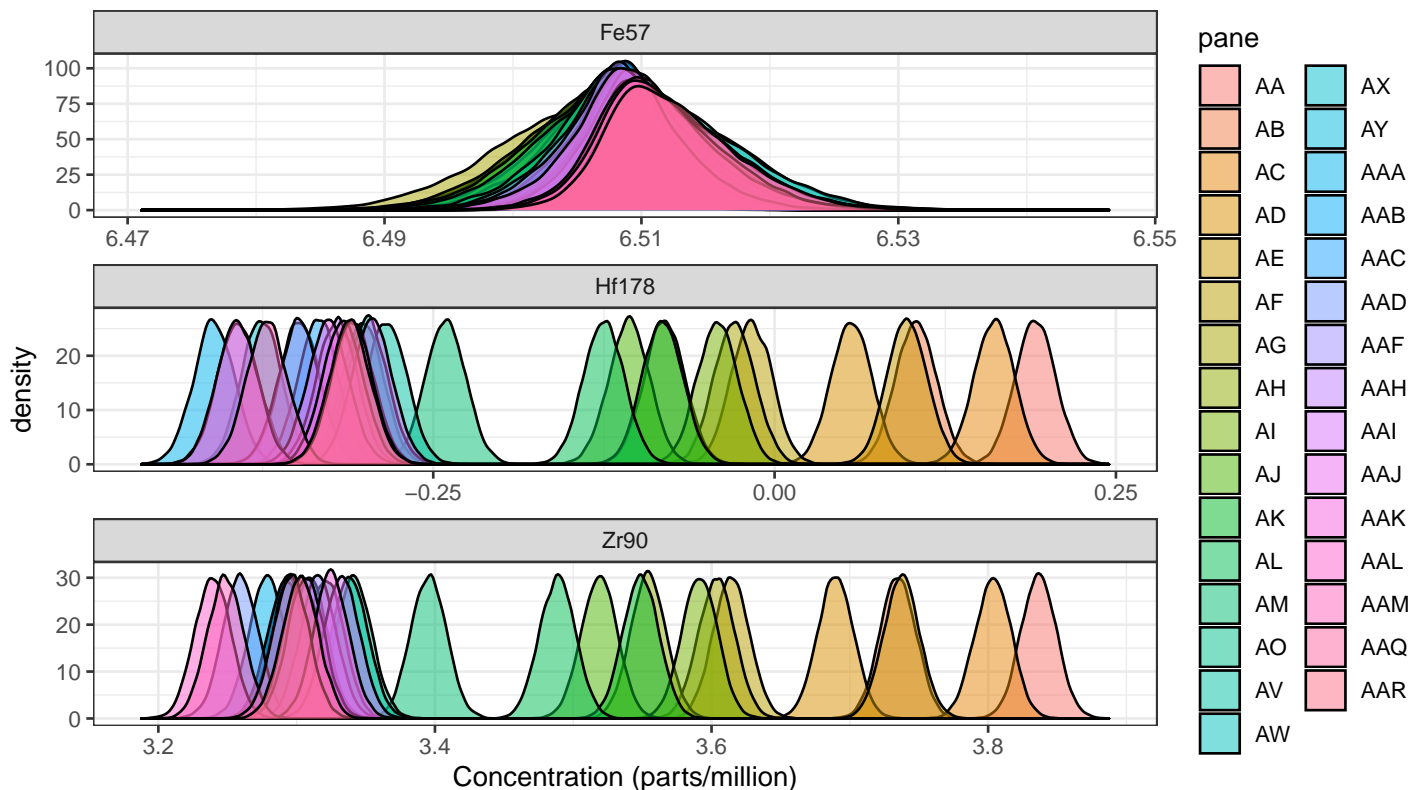


Figure 7: Estimated posterior densities for element/pane-specific means under pane-specific model.

Posterior distributions of element-specific parameters

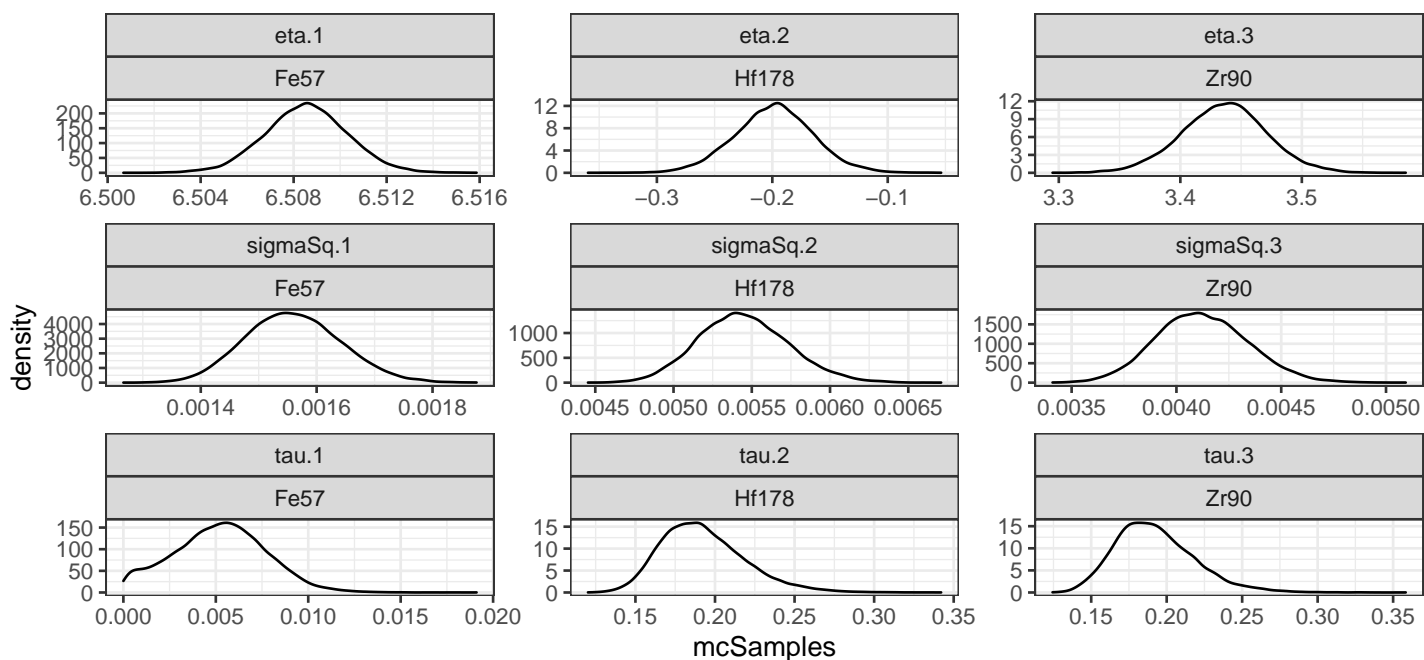


Figure 8: Estimated posterior densities for element-specific parameters under pane-specific model.

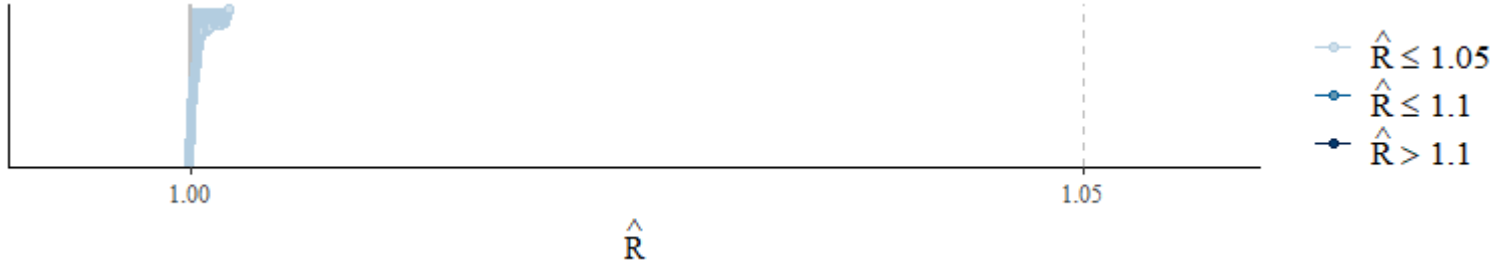


Figure 9: Scale reduction factor values for parameters in autoregressive model.

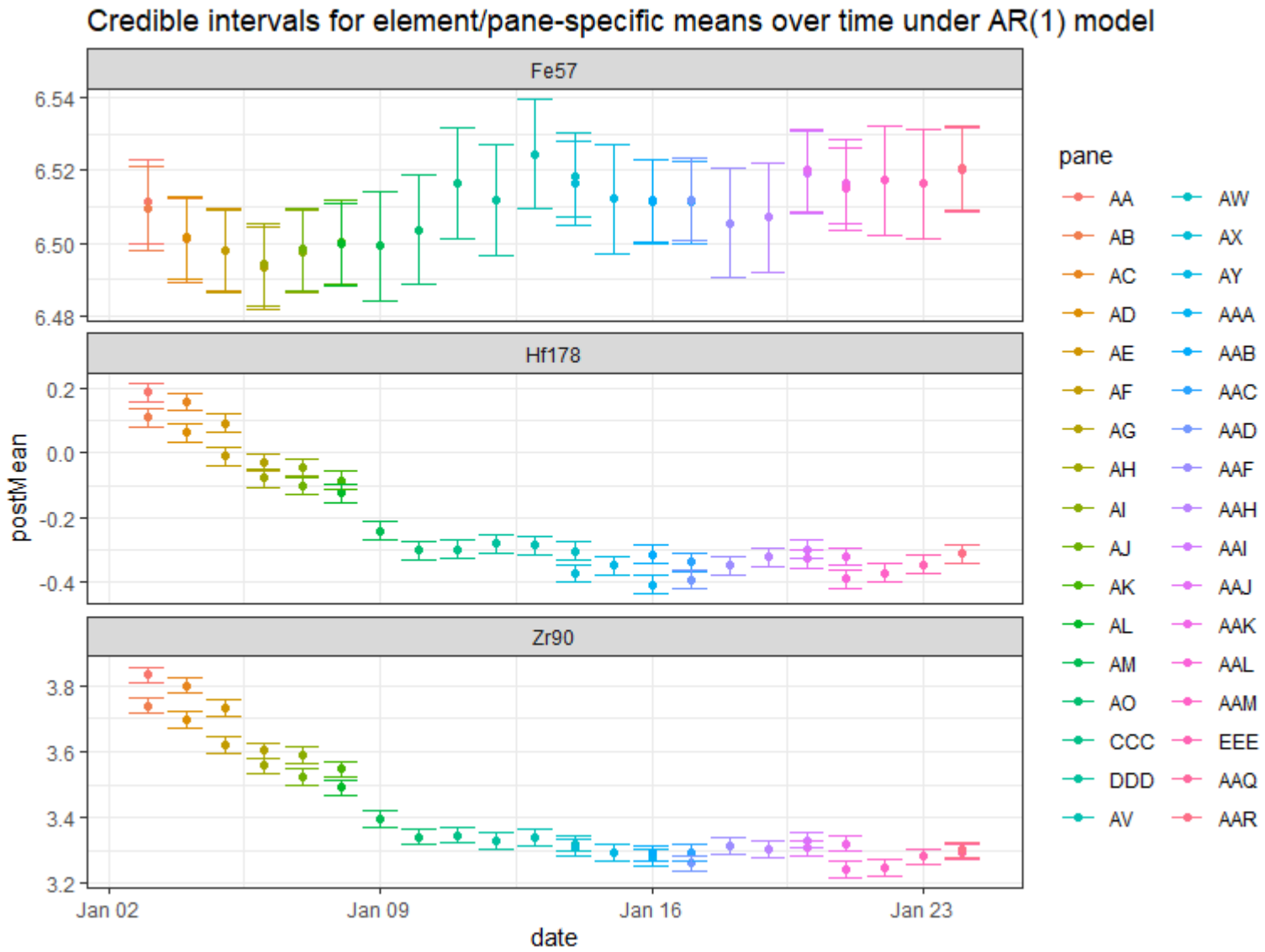


Figure 10: Posterior element/pane-specific credible intervals for data model expectations under autoregressive model.

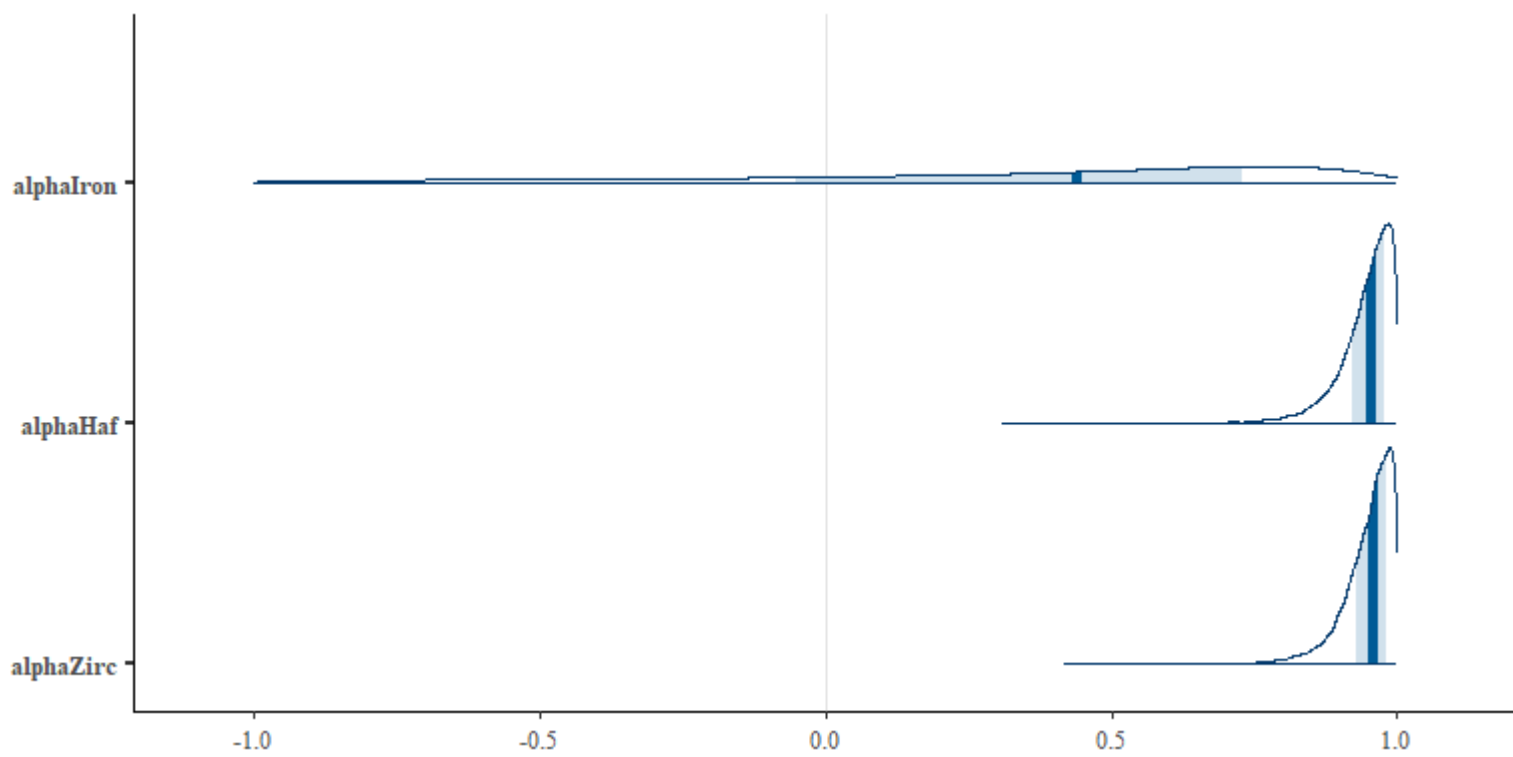


Figure 11: Posterior distributions of the AR(1) coefficients for the 3 elements