Probabilistic Programming for temperature reconstructions in the geological past to improve our understanding of greenhouse climates

Msc thesis

M. P. Geurtsen (5896347)

2022-12-19 05:58:08

Contents

1 Abstract

This thesis concerns the building of a Bayesian model for a geological temperature proxy based on oxygen isotope fractionation in carbonates. The major improvement over original regressions is a better quantification of uncertainty and the use of partial pooling. Multiple models are compared for further improvements.

2 Introduction

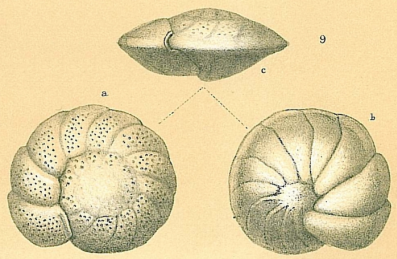

Figure 1: An illustration of the foraminifera Cibicidoides pachyderma. Image source: Brady, H.B. (1884) Pl. 94 - CC BY-NC-SA 4.0

With the ever increasing urgency of the problem of man-made global warming, it is more important than ever that we have accurate forecasts of climate and temperature in the future in order to guide our actions on climate change mitigation. Our current models are based on our understanding of the temperature in the past. As such, it is also very important to have an accurate understanding of the temperature of our geological past. This improves our climate models for the future, however it also increases our understanding of the geological past in general. Current models take the history of ocean level and temperature into account in order to project into the future [1]. Past ocean temperature is studies with a wide variety of methods. One such method is the oxygen isotope temperature proxy.

This thesis is focused on the calibration of this temperature proxy by studying its application to fossils of benthic foraminifera like the *Cibicidoides pachyderma* in figure 1. The differences and similarities between species, composition and other variables is explored, leading to an examination of the resulting uncertainty in the temperature estimation. The intended improvement is to produce a method to quantify uncertainty around temperature estimates which can be taken into account in future climate change models instead of using point estimates from classical linear regression analysis.

The calibration of the oxygen isotopes proxy is performed through a linear regression, with the temperature as a dependent variable, and the δ^{18}O isotope ratio as independent variables. This results in a coefficient describing the relationship between the two. By combining many such estimates with dating methods, we can reconstruct the temperature throughout various geological periods. This analysis was done in a paper by Westerhold et al. [24], which produced a plot showing a temperature timeline based on δ^{18}O and δ^{13}C measurements reproduced in figure 2 below.

The focus of this thesis will be on improving the calibration of this temperature proxy through probabilistic programming. We will be using multiple different sources of measurements of $\delta^{18}O_c$ and $\delta^{18}O_w$, where the main focus is to break down the differences in the temperature relationship for different sources of $\delta^{18}O_c$ and quantifying the variability in the measured relationship. Along the way we will encounter many sources of variability

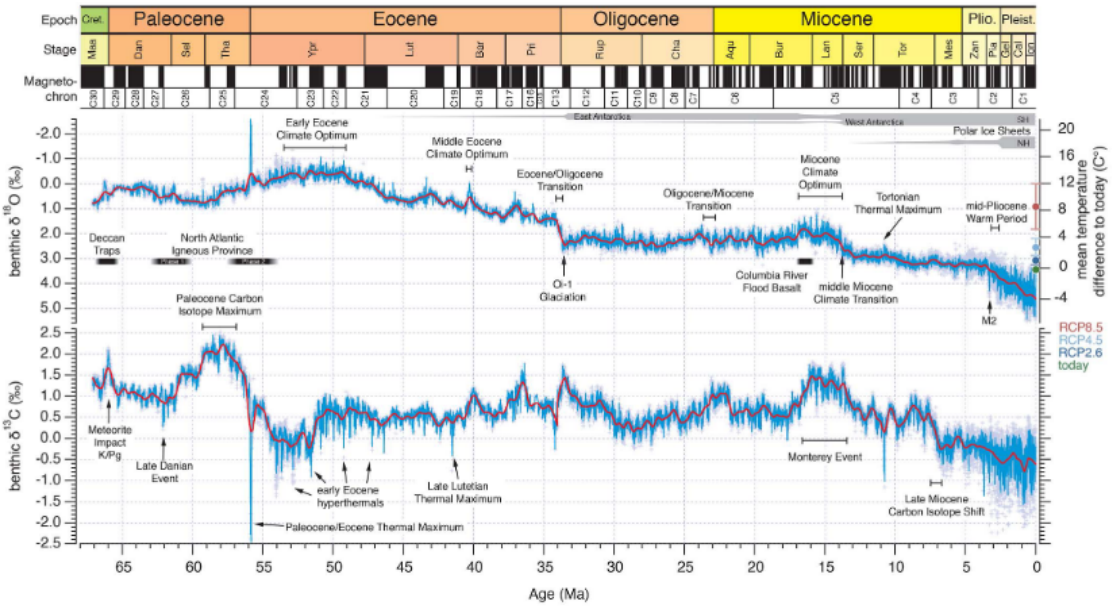

Figure 2: A temperature timeline by Westerhold et al. [24]. The blue and red lines are smoothings over the binned data by 20 000 years and 1 million years respectively. The erratic nature of the blue line (with less smoothing) gives some idea of the temperature dispersion of the timeline, which we mean to explicitly provide.

and measurement difficulties in the δ^{18}O proxy, which we will each in turn incorporate into the model.

We will go over a brief overview of assumptions and measurement difficulties. As mentioned before, one of our assumptions is that the oxygen isotope originates predominately from the environmental water. If this assumption does not hold, then our relationship falls into the water. Just as important is that the carbonate is formed under equilibrium isotopic fractionation, as described above. While this is a completely reasonable assumption for inorganically precipitated carbonate, it might not be completely correct in organically formed foraminifera shells.

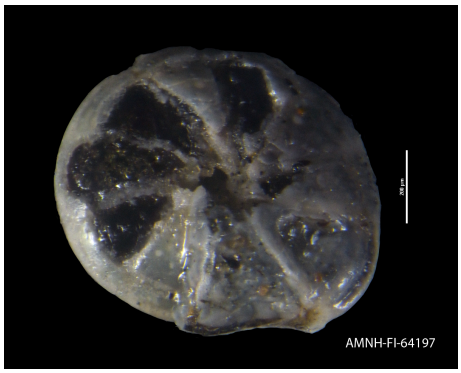

Figure 3: A photograph of the foraminifera Hoeglundina elegans. Image source: marinespecies.org, CC BY-NC-SA 4.0

Some further difficulties are that some organisms of foraminifera live in specific places and most organisms have a preferred temperature range, so each species is hard to compare. Different organism also form the shell's carbonate into different polymorphs, different

structures of the crystal making up the shell [12]. For example, some of the species we're analyzing are taxa that make calcitic shells, such as *Cibicidoides pachyderma* (shown in figure 1), while other taxa make aragonitic shells, like *Hoeglundina elegans* (shown in figure 3). Furthermore each measurement is dependent on the specific procedure that is performed and and any number of other laboratory differences that affect the validity of calibrations. One such important difference which we will return to is which reference is used, e.g. SMOW vs VSMOW, PDB vs VPDB, and the offsets associated with different species.

To derive this relationship as accurately as possible, we will collect data from various studies involving both inorganically precipitated carbonate as well as organically formed foraminifera shells of various species. Some sources will be from lab grown foraminifera, while most will be from drilling cores in the ocean under many different conditions, in the tropics and in the arctic. The wish is to end up with a model that captures the full range of possibilities where the oxygen isotope proxy may be used.

Figure 4: An aragonite crystal cluster from Spain. The shape of the crystal is different from calcite as a result of a different crystal lattice. Over millions of years aragonite turns to the more stable calcite. Attribution: Ivar Leidus - Own work, commons.wikimedia.org CC BY-NC-SA 4.0,

The problem of learning optimally from data that is organized at many different levels suggests as a solution a hierarchical model in Bayesian statistics, which is a very powerful and suitable tool for such a task [7], as it allows the model to learn it's own priors from the data. (these types of models are also known as multilevel or partial pooled models.) We will start by constructing a hierarchical linear model based on different organisms, lab processes and mineral structures. This will allow us to see the impact of each of the sources of variability, outlined above, on the resulting estimates.

The Bayesian hierarchical model is a very powerful method of analysis, which is a good reason to choose it. There is however a more important reason. For this project, we want to accurately propagate the calibration uncertainty throughout our entire analysis, so that in the end we have a complete idea of how certain we are of the estimated relationship. In Bayesian statistics we get this information "for free", as it is already an integral part of specifying the model and getting a joint posterior distribution to capture the data

generation process.

After a more in depth description of the temperature proxy in section 3.1 and the statistics in section 3.2, we will turn to the proposed new model in section 4, and explore more possible improvements. The findings will be discussed in section 5, finally ending in a discussion including conclusions and ideas for future work in section 6.

3 Background

3.1 geological background

Let us first briefly introduce the oxygen temperature proxy. The proxy works by comparing the isotopic composition of carbonate, such as aragonite or calcite, and that of its environment.

The oxygen atoms can have different atomic masses, corresponding to different isotopes, which can be measured by using a mass spectrometer. We are interested in the ratio between the ^{18}O isotope and the more common ^{16}O isotope. The ratio is called $\delta^{18}O$, expressed in the following formula:

$$\delta^{18}O_x = \frac{([^{18}O]/[^{16}O])_x}{([^{18}O]/[^{16}O])_{reference}} \cdot 1000‰$$

When this ratio is measured in carbonate and standardized against a reference, we can compare it to the same ratio, but measured in the environment, and check the difference between the two. As we measure ^{18}O in carbon and the surrounding water, we refer to these as $\delta^{18}\mathrm{O_c}$ and $\delta^{18}\mathrm{O_w}$ respectively. The resulting quantity of interest is the difference $(\delta^{18}O_c - \delta^{18}O_w)$.

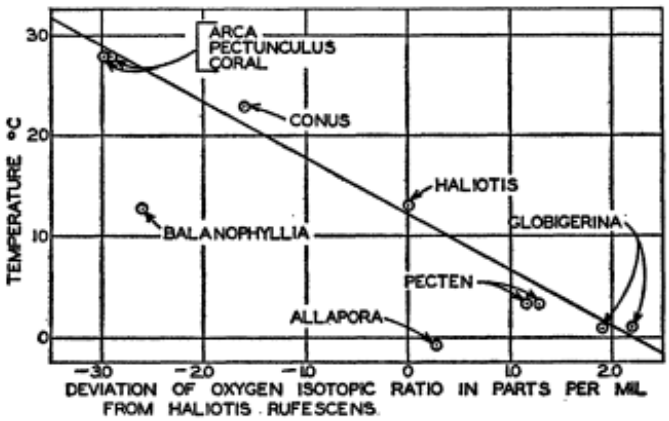

Figure 5: The relationship between temperature and isotope fractionation ratio for various samples, by Urey [22].

The isotopic composition of the carbonate, $\delta^{18}\mathrm{O_c}$, is dependent on the isotopic composition of the water, $\delta^{18}\mathrm{O_w}$, and of the environmental temperature. This relationship can be derived by regression, as shown in figure 5. Specifying this relationship as precise as possible will lead to precise temperature estimates. The reason we can assume a precise relationship between the difference $(\delta^{18}O_c - \delta^{18}O_w)$ and the environmental temperature is because we assume that the carbonate was formed in the water under equilibrium isotopic fractionation.

Isotopic fractionation is an isotopic exchange between two phases. In this case we are describing the process of forming carbonate in exchange with water. The carbonate can be either organic or inorganic, however in both cases we require it to have come about at equilibrium fractionation. Equilibrium fractionation means the chemical process includes an exchange of isotopes that approaches some balance in the ratio of heavier and lighter isotopes between two phases, in this case the carbonate phase and the water phase. The balance leads to a fractionation factor α; the ratio of the two isotopes in one phase divided by the corresponding ratio for the other phase. This fractionation factor is dependent on the temperature of the environment. In the case of foraminifera, the exchange process between carbonate ion and water adheres to the following formula (only marking oxygen isotopes):

$$\mathrm{Ca^{2+}C\,^{16}O_3^-} + \mathrm{H_2\,^{18}O} \rightleftharpoons \mathrm{CaC\,^{18}O\,^{16}}O_2 + \mathrm{H_2\,^{16}O}$$

$$\alpha = \frac{\frac{3[C\,^{18}O_3^{2-}]+2[C\,^{16}O\,^{18}O_2^{2-}]+[C\,^{16}O_2\,^{18}O^{2-}]}{3[C\,^{18}O_3^{2-}]+2[C\,^{16}O_2\,^{18}O^{2-}]+[C\,^{16}O\,^{18}O_2^{2-}]}}{\frac{[H_2\,^{18}O]}{[H_2\,^{16}O]}}$$

The counterpart of equilibrium fractionation is kinetic fractionation, in which case the process is incomplete and unidirectional, which means it does not obey a fractionation factor. If the carbonate we are studying was formed under kinetic fractionation, we cant depend on the isotope ratio to be a reliable thermometer. [21]

Figure 6: A calcite crystal from Irai, Brazil. Calcite is a stable polymorph of carbonate, the crystal is often white. Attribution: Rob Lavinsky, iRocks.com – CC-BY-SA-3.0

Oxygen isotope fractionation has been a significant object of study since at least 1949, when Urey published a seminal paper on isotope concentrations in calcium carbonate [22]. Urey and others have also quantified the temperature dependency of δ^{18}O on non biological sources of carbonate.

Urey adapted isotope fractionation study to oxygen isotopes and worked out a method of researching the oxygen isotopes ^{18}O and^{16}O in fossils. They considered lead-208 based isotope dating, an example of isotope fractionation research, based on which an upper limit can be placed on the age of the earth. The same principle governing isotope dating might be turned to temperature Isotope fractionation had been studied in lab conditions, and

experimentation yielded the discovery that algae plants contain less^{13}C than the solution that the plants were grown in.

The research is based on equilibrium isotope fractionation of oxygen isotope composition in the sea water and in the shells formed in that seawater. As explained in the introduction, the primary source of the oxygen atoms in the shell is the oceanic water, therefore if the calcium carbonate in that shell was produced in equilibrium with the surrounding water, the difference $(\delta^{18}O_c - \delta^{18}O_w)$ allows us to derive the temperature. Urey states that the differences in the δ^{18}O ratio for 1 °C is only 0.0176%, so the measurement of the two isotopes must be very precise. The preparation and mass spectrometer methods by Urey were precise enough to measure at that scale. They procured a number of specimens and created the first comparison graph of δ^{18}O and temperature, shown in figure 5.

In 1953, Epstein et al. [6] improved the method of purifying the samples, and eliminated a source of extraneous oxygen. They derived an improved temperature ^{18}O regression[6], leading to the formula:

$$t(^\circ C) = 16.5 - 4.3\delta + 0.14\delta^2 \pm 0.5^\circ C$$

Where δ is the difference $(\delta^{18}O_c - \delta^{18}O_w)$ of the sample to a reference gas.

McCrea calculated on theoretical grounds precisely what the influence of temperature should be on the fractionation ratio based on the assumptions that the exchange of isotopes happens at equilibrium, then studied the isotopic composition of foraminifera grown under laboratory conditions.

McCrea derives a fractionation factor α which takes into account all possible configurations of the two oxygen isotopes of interest both in the carbonate shell and in the water.

Exchange of O^{16} and O^{18} between	Temperature, °C 0.0	8.0	17.0	25.0
$CO_3^=$(aq.)−O	1.1046_3	1.0998_4	1.0948_3	1.0907_2
H_2O(g)−O	1.0741_3	1.0716_2	1.0689_8	1.0667_7
H_2O(l)−H_2O(g)	1.0103_5	1.0095_9	1.0089_9	1.0083_0
$CO_3^=$(aq.)−H_2O(l)	1.0176_1	1.0162_6	1.0148_6	1.0137_6
Calcite−H_2O(l)	1.0254_8	1.0237_8	1.0210_1	1.0205_4

Figure 7: table of fractionation factors by McCrea [16].

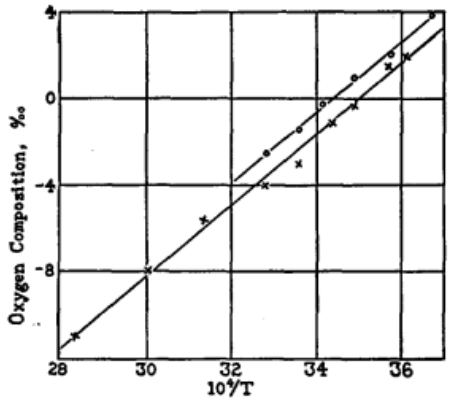

Figure 8: figure of McCrea of oxygen composition as a function of temperature with a linear relation drawn through the points.

$$\alpha = \frac{\frac{3[C\,^{18}O_3^{2-}]+2[C\,^{16}O\,^{18}O_2^{2-}]+[C\,^{16}O_2\,^{18}O^{2-}]}{3[C\,^{18}O_3^{2-}]+2[C\,^{16}O_2\,^{18}O^{2-}]+[C\,^{16}O\,^{18}O_2^{2-}]}}{\frac{[H_2\,^{18}O]}{[H_2\,^{16}O]}}$$

McCrea derives fractionation factors at various temperatures, where are here shown as a table in figure 7 The isotopic composition resulting from the precipitated calcite of McCrea follows a linear regression very well, as shown in figure 8. Nevertheless McCrea used a quadratic term as well, which is quite common in these regressions [2].

The calcite was precipitated slowly, with an average half-time of 28 hours. The resulting temperature to δ^{18}O relationship was given as

$$t = 16.0 - 5.17(\delta^{18}O_c - \delta^{18}O_w) + 0.092(\delta^{18}O_c - \delta^{18}O_w)^2$$

This can be compared with the relationship of biological origin, given by Urey. Before comparison this relationship must be corrected; the standard used by Urey was -0.95 ‰ different from that used by McCrea, leading to the equation

$$t = 18 - 5.44(\delta^{18}O_c - \delta^{18}O_w)$$

Both equations have resulting temperatures within one °C temperature difference and are thus said to be in agreement according to McCrea.

In 2014, Marchitto et al. published a paper combining new core top measurements with previously published data to rederive the relationship between δ^{18}O and temperature. Core-top measurements are measurements made at the top of a drilling core, which are necessarily the most recent measurements.

Marchitto et al. underscore the uncertainty still surrounding the temperature to δ^{18}O relationship in benthic foraminifera, as well as the inconsistent use of the derived paleotemperature equations: various species are used interchangeably, species which may not be at equilibrium are used as well as inorganically precipitated calcite at equilibrium. Furthermore the temperature sensitivities may themselves be dependent on the temperature as warm-water foraminiferal sensitivities are smaller than very cold waters. Marchitto et al. study three groups: *Cibicidoides* and *Planulina*, *Uvigerina*, and *Hoeglundina elegans*. These measurements were combined with measurements from earlier studies.

Marchitto et al. refer to the paleotemperature equations of Shackleton and of Lynch-Stieglitz, Curry, and Slowey, preferring the latter as it is better constrained, based on *Cibicidoides* and *Planulina* (between 4 °C and 26 °C). The difference in the two equations is quite big: Shackleton has a slope of -0.25 ‰ per °C while Lynch-Stieglitz, Curry, and Slowey is only -0.21 ‰ per °C. The difference comes to about 2 °C.

Duplessy, Labeyrie, and Waelbroeck [5] showed that regressions based on *Cibicidoides* core top measurements, when adjusted using Shackleton's +0.64 ‰ adjustment, visually agree well with Shackleton's regressions.

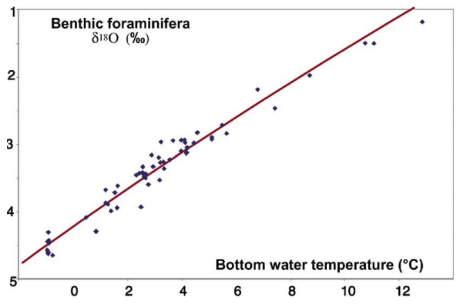

Figure 9: Duplessy, Labeyrie, and Waelbroeck's new measurements visually aligning with Shackleton's regression line.

Marchitto et al.'s new measurements are from five taxa:

- *Cibicidoides pachyderma*,
- *Planulina ariminensis*
- *Planulina foveolata*
- *Uvigerina peregrina*
- *Hoeglundina elegans*

The first four are calcitic, while the last is aragonitic.

These are core-top samples from 31 multicores in the Florida Straits at temperatures of 5.8 - 19.0 °C. The seawater $\delta^{18}O$ is well constrained.

A second source of data is of *Cibicidoides wuellerstorfi* in Arctic Ocean core tops, at temperatures below 0 °C. This data is combined with previously unpublished Little bahama Bank sediments used by Lynch-Stieglitz, Curry, and Slowey [14]

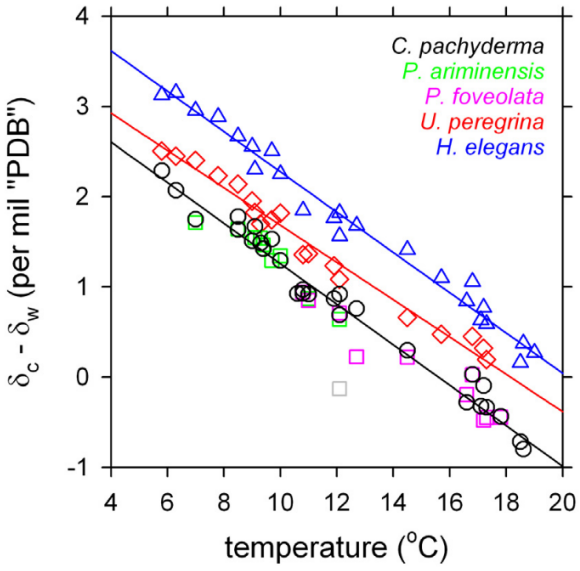

Figure 10: The temperature dependency in the Florida Straits benthic foraminifera, Marchitto et al. [15]. The lines are RMA regression, grey points are excluded 3σ outlier.

The regressions between $\delta^{18}O$ difference in the Florida Straits are shown in figure 10. Marchitto et al. colored the data points to species, showing the offset difference. In these regressions, the slope visually appears almost identical, particularly between *H. elegans* and *C. pachyderma*.

Marchitto et al. used Reduced Major Axis regression (RMA) as an alternative to the traditional ordinary least squares regression (OLS). RMA, also known as total least squares regression, is preferred over OLS by Marchitto et al. as the regression error is quantified over both axis instead of only over the y axis.

The timeline shown in figure 2 comes from a 2020 paper by Westerhold et al., who constructed a continuous temperature timeline going back 66 million years, based on benthic foraminifera isotope fractionation records. Temperature timelines are important to identity the Earth's past climate, and therefore the current climate and climate change. Their measurements are timed to an accuracy of ± 100 thousand years for the Paleocene (66 mya - 56 mya) and Eocene, and to ± 10 thousand years for the late Miocene (23 Ma - 5.3 Ma) to Pleistocene (2.5Ma-11kyr), which were collected from 14 ocean drilling cores, mostly from the species *Cibicidoides* and *Nuttallides*.

The timeline was constructed by binning the measurements on average every 2 thousand years and smoothing by a locally weighted function. The red line is smoothed over 1 million years, while the blue line is smoothed over 20 thousand years.

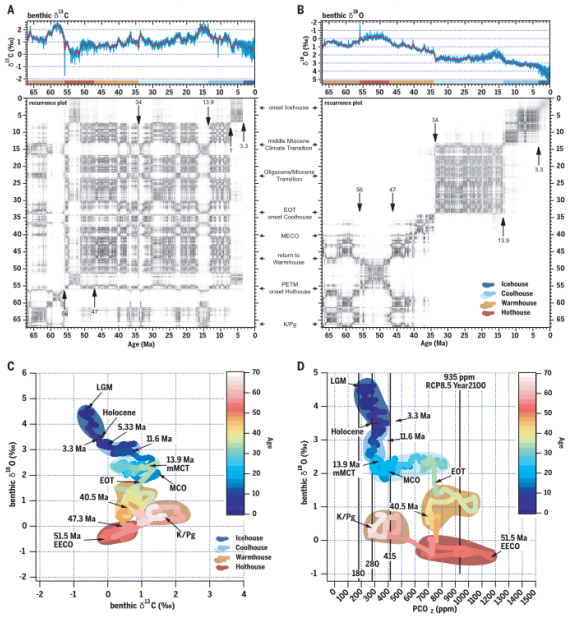

Figure 11: climate states of the Cenozoic as set out by Westerhold et al., who have determined through recurrence analysis that four distinct climate states emerge corresponding to specific temperature bands.

The results were then analyzed using recurrence analysis (RA), which is a technique from chaos theory to analyze dynamical systems. It allows creating visualizations of the state of the system in terms of the transitions from one state to another, which is called a phase space trajectory. Westerhold et al.'s resulting Recurrence plots, copied in figure 11, wherein they identified four broad climate states which they names Hothouse, Warmhouse, Coolhouse, and Icehouse according to mean temperatures. A recurrence plot should be read by diagonal lines representing phase shifts to a different phase, and vertical lines representing staying stationary in the same phase (the plots are symmetrical, so choosing

vertical over horizontal lines is purely by convention).

When comparing δ^{18}O to a reference, there is a choice of what reference to be used, which impacts the measured difference. The reference of choice depends on what we're measuring. For oxygen from carbonate, the Pee Dee Belemnite (PDB) standard is used, which was established based on Belemnite fossils found in South Carolina along the Pee Dee river. The original supply has run out, so now a reference value of a hypothetical Vienna Pee Dee Belemnite is used another standard from Vienna, VPDB, is used instead. For oxygen isotope ratios in the water, the reference Vienna Standard Mean Ocean Water (VSMOW) is used, which, like VPDB, was preceded by a standard from America, called SMOW. It is unclear which reference exactly were used for the earliest studies like Urey [22] and McCrea [16], while later studies usually referenced the latest referents of PDB or SMOW or their Vienna equivalents.

3.2 statistical background

The relationship between δ^{18}O and temperature has usually been determined with linear regressions, sometimes with a quadratic term. In this case the temperature is the outcome variable, and we're interested in the impact of the δ^{18}O predictor term, which is given by its coefficient. The regression has usually been done using ordinary least squares (OLS) regression, which is a technique for determining a linear relationship between predictor variables and an outcome variable. It works by minimizing the quadratic of the residual errors, which are the difference between the predicted point and the real point in the y direction. As indicated in the section discussing Marchitto et al. [15], there are alternative forms of regression which minimize different errors. The Bayesian framework is quite different.

To compare Bayesian modeling to OLS, we will need to think of OLS in terms of parameter estimation of a probability distribution. We can describe OLS in terms of maximum likelihood estimation (MLE), which is a method to choose parameters that best describe the outcome variable in terms of a probability distribution, in the case of OLS it is the normal distribution. The idea is quite logical: since we know the data, we're going to figure out which distribution makes these data the most probable. In other words, choose the parameters θ from the set of possible parameters Θ that maximize the likelihood L of the set of data Y:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \widehat{L}_n(\theta; Y)$$

.

Our first Bayesian model will be a simple linear regression, like OLS. But unlike OLS, the model will require prior distributions for all parameters. Prior probability distributions allow us to inform our model which combination of values is plausible, and which are not. This way, priors allow for statistical inference when there is not enough data for frequentist methods, provided there is sufficient background knowledge to specify informative priors. Usually priors should have enough information for the task of identifying reasonable values, but in theory anything can be a prior. A possible prior is the flat prior, where each possibility is assigned the same value. (in the case of continuous outcomes, this is not a proper distribution, meaning it does not integrate to one. This also means it is not a probability distribution. Nevertheless in principle it can be used as a prior.) Bayesian methods result in a full posterior probability distribution instead of resulting in the most likely parameter values. This comes with certain advantages, but for the purposes of comparison we will focus on taking the mean of this distribution as our point value, and

the standard deviation as our measure of variability, mirroring a typical regression with a coefficient and associated error. If Bayesian regression is done with a flat prior, all the information shaping the posterior distribution must come from the data, so we end up getting a mean at the parameter values which make the data we're seeing most probable: this is equivalent to Maximum Likelihood Estimation.

Markov Chain Monte Carlo and Hamiltonian Monte Carlo

All Bayesian models conclude on a posterior distribution which fully described the model's information on all parameters. The trick is how to get that posterior distribution. Bayesian statistics relies on Markov Chain Monte Carlo (MCMC) to get to the posterior distribution.

MCMC is essentially a way to explore such a posterior distribution by walking in the parameter space. The first MCMC algorithm, the Metropolis-Hastings algorithm, was discovered by Metropolis et al. [18] in 1953. To describe the metropolis algorithm by analogy, envision a landscape with hills and a person, Mark, who will be walking along the hills. Each time Mark is about to take a step in a random direction, they measure how much higher or lower their foot will be than it currently is. If it's higher, Mark wants to make the step, but if it's lower, they are not really sure what they want. In that case, Mark chooses to make the step randomly according to how big the difference is. Mark will end up on the tops of the hills more often than in the valleys, but they will still occasionally end up in the valleys, according to how deep they are.

Figuring out the posterior distribution would be possible if we could draw random draws from it. This would be like Mark teleporting randomly around the terrain, each time noting how high they are. Teleporting is hard, so MCMC instead allows draws to be correlated with each other, just like Mark makes steps from one place to another. The metropolis algorithm has Mark walk along the hills defining the parameter probability for a set number of steps, and records where Mark has been, and how often they have been there. If Mark keeps going for a long time, eventually Mark will walk throughout all the hills, and the algorithm knows the entire probability space. Usually MCMC models will feature multiple chains: various Marks starting their walks at different points.

Stan uses an implementation of MCMC that is a variant of Hamiltonian Monte Carlo (HMC) called the No U Turn sampler (NUTS) [11]. Hamiltonian Monte Carlo is a complicated algorithm that improves greatly over more straightforward methods to fit Bayesian models.

Hamiltonian Monte Carlo runs a physics simulation of a particle gliding, much like Mark walked, across a landscape of hills. The particle speeds up when the hills are more steep, and slows down when the hills are more level. The particle travels throughout the entire space before it has been back where it's started. HMC uses the particle to find points to propose, analogous to where Mark places his next step. But because it follows the particle, HMC's samples are less correlated and explore the hills more efficiently. Because of the physics simulation each step takes longer to compute, and all parameters must be continuous: you could imagine Mark stepping up and down stairs, but the particle cannot glide over bumpy stairs.

A regression's equation

We will now turn to creating our first model of the temperature to δ^{18}O relationship and comparing it to previously established regression equations. There have been many studies of this relationship, and consequently there are as many regressions. Bemis et al. [2] has collected a table of commonly used temperature to δ^{18}O relationships as of 1998, which

can be viewed in figure 12. Bemis et al. conveniently made all equations conform to the following formula:

$$T(^\circ C) = a + b(\delta^{18}O_c - \delta^{18}O_w) + c(\delta^{18}O_c - \delta^{18}O_w)^2$$

Some of the regressions were recalculated using original data, some were approximated as the original formula was of the form $10^3 \ln \alpha$ instead. The coefficients of most interest are b and c, respectively for the linear and the quadratic dependence on the oxygen fractionation ratio in the source.

Table 1. Comparison of Commonly Used Calcite Temperature:δ^{18}O Relationships With Equations Developed in This Study

		$T(°C) = a + b(\delta_c - \delta_w) + c(\delta_c - \delta_w)^2$			δ_w Correction	T(°C) Offset From LL Equation (1)			T(°C) Offset From HL Equation (2)		
Reference	Source	*a*	*b*	*c*	(VSMOW to VPDB)	0°C	15°C	25°C	0°C	15°C	25°C
McCrea [1950]	Inorganic	16.0	- 5.17	0.09	- 0.20‰	- 0.2	- 0.2	+ 0.7	+ 1.4	+ 1.5	+ 2.5
Epstein et al. [1953]	Mollusk	16.5	- 4.3	0.14	- 0.20‰	+ 3.6	+ 0.5	- 0.1	+ 4.7	+ 1.9	+ 1.5
Craig [1965]	Mollusk (modified from *Epstein et al.* [1953])	16.9	- 4.2	0.13	- 0.20‰	+ 4.2	+ 0.9	+ 0.1	+ 5.3	+ 2.3	+ 1.6
O'Neil et al. [1969]	Inorganic*	16.9	- 4.38	0.1	- 0.20‰	+ 3.3	+ 0.8	+ 0.3	+ 4.5	+ 2.3	+ 1.9
Horibe and Oba [1972]	Mollusk	17	- 4.34	0.16	- 0.20‰	+ 4.2	+ 1.0	+ 0.6	+ 5.3	+ 2.4	+ 2.2
Shackleton [1974]	BF: (modified from *O'Neil et al.* [1969]; calibrated with *Uvigerina*)	16.9	- 4.0		- 0.20‰	+ 3.4	+ 0.9		+ 4.8	+ 2.3	
Erez and Luz [1983]	PF: *Globigerinoides sacculifer* (laboratory)	17.0	- 4.52	0.03	- 0.22‰	+ 2.0	+ 0.8	+ 0.3	+ 3.5	+ 2.3	+ 1.9
Bouvier-Soumagnac and Duplessy [1985]	PF: *Orbulina universa* (laboratory)	16.4	- 4.67		- 0.20‰	+ 0.7	+ 0.3	+ 0.0	+ 2.2	+ 1.8	+ 1.6
Bouvier-Soumagnac and Duplessy [1985]	PF: *O. universa* (Indian Ocean)	15.4	- 4.81		- 0.20‰	- 0.8	- 0.8	- 0.7	+ 0.8	+ 0.8	+ 0.9
Kim and O'Neil [1997]	Inorganic*	16.1	- 4.64	0.09	- 0.27‰	+ 1.2	- 0.3	- 0.4	+ 2.6	+ 1.2	+ 1.3
This study, (1)	PF: *O. universa* (LL)	16.5	- 4.80		- 0.27‰						
This study, (2)	PF: *O. universa* (HL)	14.9	- 4.80		- 0.27‰						
This study, (3)	PF: *Globigerina bulloides* (11-chambered shell)	12.6	- 5.07		- 0.27‰						
This study, (4)	PF: *G. bulloides* (12-chambered shell)	13.2	- 4.89		- 0.27‰						
This study, (5)	PF: *G. bulloides* (13-chambered shell)	13.6	- 4.77		- 0.27‰						

The abbreviations are defined as follows: BF, benthic foraminifera; PF, planktonic foraminifera; LL, low light (20-30 μEinst m^{-2} s^{-1}); HL, (>380 μEinst m^{-2} s^{-1}); and VSMOW, Vienna SMOW; and VPDB, Vienna Peedee belemnite. Temperature offsets of published equations from (1) and (2) are calculated for 0°, 15°, and 25°C. Asterisks denote a conversion from "$10^3 \ln \alpha$" notation using a quadratic approximation (see text).

Figure 12: table showing temperature relationships from different studies, including 5 new relationships from Bemis et al. [2].

$$
\begin{array}{lcll}
a & \sim & \text{Normal}(16, 4) & \text{prior for the intercept} \\
b & \sim & \text{Normal}(-4, 2) & \text{prior for linear term coefficent} \\
c & \sim & \text{Normal}(0.1, 0.5) & \text{prior for quadratic term coefficient} \\
[b]\ \sigma & \sim & \text{Normal}_+(0, 4) & \text{prior for the standard deviation} \\
 & & & \\
\mu_i & = & a + b(\delta^{18}O_c - \delta^{18}O_w) + c(\delta^{18}O_c - \delta^{18}O_w)^2 & \text{linear model} \\
T_i & \sim & \text{Normal}(\mu, \sigma) & \text{likelihood}
\end{array}
\tag{1}
$$

As described above, our outcome variable will be temperature, and we will have two prediction variables; the difference $(\delta^{18}O_c - \delta^{18}O_w)$, and its square, resembling Bemis et al.'s equation in figure 12. This model is described in math block 1. Note the linear model near the bottom: it specifies the same equation as Bemis et al. It is preceded by priors for each parameter.

The likelihood shows that we're assuming our error will be normally distributed around some mean, which is not an uncommon assumption for linear regression. The priors are weakly informative, as they have been set in the range we would expect given the prior work shown in figure 12, but each has deliberately been given a wide standard deviation to allow for the data to overwhem the priors. The prior for σ has a plus sign in subscript,

which denotes that the distribution has to have positive values. A standard deviation naturally can not be negative.

This model can be fitted in Stan, which is statistical modelling based on the Hamiltonian Monte Carlo algorithm. Stan uses the Stan language for model specification. The Stan language is c-like, so in addition to specifying our priors, likelihood and linear model, we need to tell Stan about the types to use for the variables. Stan uses various blocks to define the program, with the most important one being the model block. The Stan code for this model is shown in listing 1. Note especially the model block; it is virtually identical to the specification as written in model block 1 above.

A linear regression model which was ran on a subset of the final dataset will be reproduced here. This does not represent actual results but is merely meant to illustrate the process. At the time this model was fitted, the data processing had not yet finished. The model was fitted to the data that is collected from Marchitto et al. [15], which yields the results in listing 2. The parameters can be directly compared to those of Bemis et al.; this will not be the case for the actual models after our data preprocessing steps as set out in section 4.1. First we will look at $\hat{R}$. $\hat{R}$ is a ratio of variance in and between chains. As the variance between chains goes down, $\hat{R}$ approaches 1 from above. A rule of thumb is that a $\hat{R}$ above 1.01 indicates some kind of fitting problem which should be investigated. The inverse is not necessarily true: a good $\hat{R}$ value does not imply there are no problems. Next we will look at the effective sample size. This metric is a calculation showing how many random samples a chain effectively has. This is always lower than the number of samples each chain has, as the actual samples are autocorrelated. `ess_bulk` and `ess_tail` are respectively estimates of how many random samples there are in the sections of the posterior with high probability and in the sections with low probability (the tails). The $\hat{R}$ and `ess_bulk` look reasonable in this case, so we can cautiously conclude the mean can be trusted.

The results are not terribly surprising: all parameters are well within the prior standard deviations of the prior means. It is interesting that c is a negative term here, while the table from Bemis et al. only has positive quadratic terms. The data that was used for fitting this model came wholly from Marchitto et al. [15], which thus was not available for Bemis et al. to add to their table, providing an independent source.

When modelling, there are a number of ways to check whether the model makes sense.One way is to turn the model around and use it to generate data instead of analyzing data. This is a good way to read the model definition in math block 1: We follow our assumptions to have random variables behave as defined by our priors, we add these together in a linear model and we get a certain value as a result. We can have Stan generate this result in a `generated quantities` block, shown in listing 5.

This block is used to generate variables from the model, and is run once after the model has been fitted. In this case, it is generating two variables: `Y_sim` and `log_lik`. The latter is primarily handy for model comparison, to which we will return in a later section. The former is a variable holding fake data, which is generated according to our model. To see how well our model is capturing the process forming the data, we can compare `Y_sim` to the input data, for example simply by plotting both. When done after fitting, this is called a posterior predictive check. When done before fitting the model, it is a prior predictive check. Both are useful to evaluate the performance of our model: we can check whether our priors are actually reasonable, and we can check whether the likelihood makes sense.

Figure 13 plots eight possible distributions of fake data, overlayed over the real data. Our choices for priors do not match the data very well, some distributions are extremely narrow, and some are misplaced. Figure 14 mirrors figure 13 but shows draws from the

```
data {
  int N;
  vector[N] Y;
  vector[N] b1;
  vector[N] b2;
}

parameters {
  real a;
  real b;
  real c;
  real<lower = 0> sigma;
}

model {
  vector[N] mu;

  a     ~ normal(16,4);        // prior for the intercept
  b     ~ normal(-4,2);        // prior for linear term coefficient
  c     ~ normal(0.1,0.5);     // prior for quadratic term coefficient
  sigma ~ normal(0,4);         // prior for the standard deviation
  mu    = a + b * b1 + c * b2; // linear model
  Y     ~ normal(mu, sigma);   // likelihood
}
```

Listing 1: initial Stan model.

```
   variable      mean    median     sd    mad        q5      q95 rhat ess_bulk ess_tail
 lp__        -206.59  -206.26  1.44   1.19  -209.41 -204.93 1.00      813     1104
 a             16.38    16.38  0.28   0.28    15.94   16.84 1.00     1032      984
 b             -3.73    -3.74  0.34   0.33    -4.28   -3.16 1.00      539      770
 c             -0.21    -0.21  0.10   0.11    -0.38   -0.05 1.00      560      856
 sigma          2.28     2.28  0.13   0.13     2.08    2.50 1.01     1241     1263
```

Listing 2: initial results.

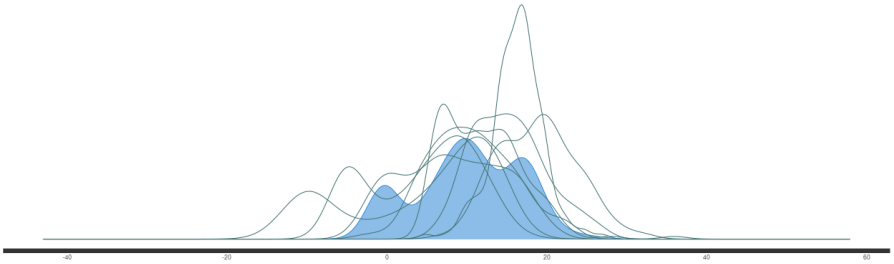

Figure 13: prior predictive check of the model. The eight lines correspond to fake data, while the colored region shows the real data distribution. The fake data shows both distributions that are very wide and distributions that are very narrow relative to the real dataset.

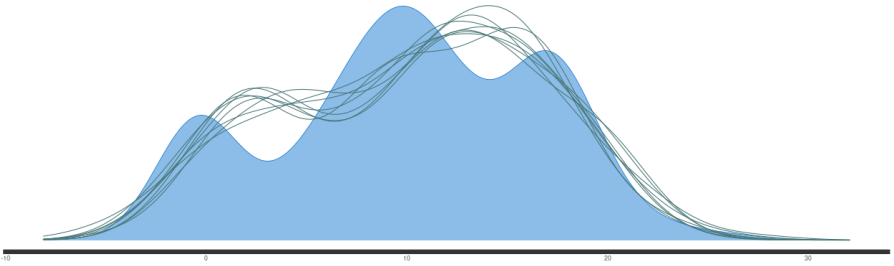

Figure 14: posterior predictive check of the model.

posterior of the model. As can be expected, once the model has seen the real data, the resulting fake data distribution is visually much more like the real data distribution, though it does not completely capture it.

Hierarchical models

The goal is to expand this simple model to incorporate multiple levels.

The original model is of a type called a fully pooled model. this means it incorporates the assumption that all data points come from the same distribution. By using a hierarchical model, we can relax this assumption, while still allowing the model to use the whole dataset to learn each parameter.

Firstly we want to add a level according to species, as it is expected from the literature

```
generated quantities {
  vector[N] log_lik;
  vector[N] Y_sim;
  vector[N] mu;
  for (i in 1:N){
    mu = a + b * b1 + c * b2;
    log_lik = normal_lpdf(Y[i] | mu,sigma);
    Y_sim[i] = normal_rng(mu[i],sigma);
  }
}
```

Listing 3: a block to generate fake data from the model.

that each species has a slightly different temperature relationship. This makes the Stan code slightly more complicated, as Stan now needs a lot more information: it needs to be told how many groups there are and which group each observation is assigned. In the model in math block 3 (from section main model: hierarchical linear model), the relationship is pooled for all three parameters; the intercept, the linear relationship and the quadratic relationship.

The choice of priors has been changed: we're using the posterior values of the initial model as priors for this model. Hierarchical models need a lot more informative priors, as the geometry that Stan has to explore has become a lot more complicated and multidimensional. By utilizing the outcome of the previous model, we allow Stan to fit this model adequately. As the parameters a, b and c are now pooled, the initial values of these parameters are priors, making the normal distributions we're using for them priors for those priors. These are usually called hyperpriors.

The Stan code to fit this model is shown in listing 4. It is a more complicated model, where a, b and c are now vectors with one value per group. In the linear model on line 28, each parameter uses the value given by `species[i]`. `species` assigns each observation in `Y` an index corresponding with its group, in this case the species that observation belongs to.

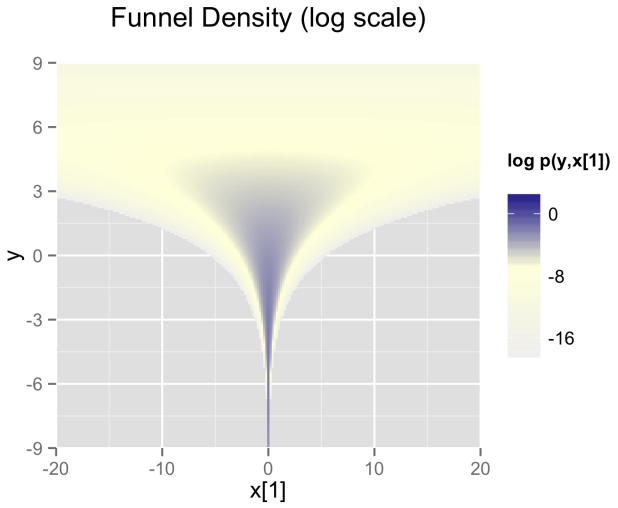

Figure 15: Neal's funnel in log probability density.

Note that the earlier model is a centered model, which we could reparameterize to create a non-centered model. Centered models are more straightforward in that the parameters that are fitted are the parameters of interest, which is easier to interpret. Non-centered models are models which do not model the parameters of interest directly but model latent variables from which the parameters of interest can be retrieved. Having a non-centered model is preferable as it captures the impact of each latent variable relative to the population, instead of capturing the absolute impact as centered models do. This eliminates one potential source of funnels, which are regions where log-probability changes rapidly. Funnels make it harder for HMC to sample from the posterior probability. Figure 15 shows a funnel discussed in the Stan users guide [20]. To effectively sample the broad area of relatively lower probability (called the body in the user guide), the sampler needs a big step size, while in the very small area of high probability, the neck, requires a very small step size. Transforming parameters, or reparameterization, is an effective way to

eliminate some funnels. Usually fully pooled models, like the first model here, do not come up against problems like Neal's funnel. Partially pooled models frequently do however, which means tricks like reparameterization are essential.

Bayesian workflow when developing a model

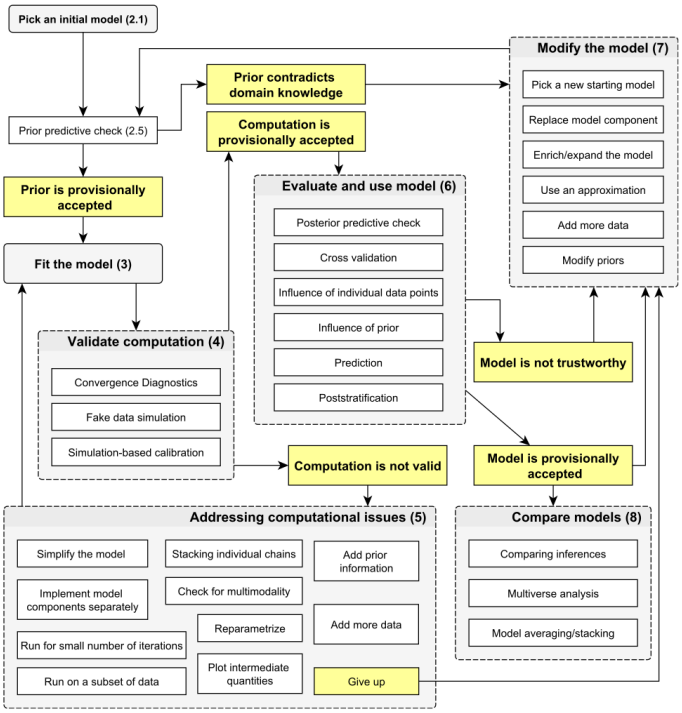

Figure 16: Gelman et al.'s diagram of iterative model improvement. Numbers indicate chapters in Gelman et al. [8].

In developing the hierarchical model we have already gone through an iteration improving the simple regression model. This demonstrates the type of iteration that will be used throughout this thesis, based on Gelman et al. [8]. Figure 16 shows a diagram from Gelman et al. of iterative model improvement which captures the workflow. Starting at the upper left corner, we've created an initial model and fitted the model. Validating computation and evaluation came in the form of fake data simulation and posterior predictive check. We noted the convergence diagnostics of $\hat{R}$ and `ess_bulk`, but in the posterior predictive check we noted that visually, the fake distributions did not seem to capture the true data exactly. We chose to expand the model into a multilevel model. This hierarchical model is where we ended up now. Now we would like to compare these models. This can be done by leave one out cross validation (LOO). In LOO, we estimate for each datapoint what the posterior distribution would be like if we left that point out. Then we check how well the model would predict that point. By performing this check for both models, we can see which performs best. The R package `loo` conveniently performs this check, reporting a score called `elpd_diff`. Without describing the meaning in detail yet, the important takeaway from this comparison is whether there is a significant difference between the two models given the standard error `loo` reports for each model. In this case the difference is more than an order of magnitude the se, so we can conclude there is a significant difference between the two models.

```
data {
  int N;
  int K; // number of sources
  vector[N] Y;
  vector[N] b1;
  vector[N] b2;
  array[N] int<lower=1, upper=K> species; // source group assignments
  int<lower=0, upper=1> prior_only;
}

parameters {
  vector[K] a;
  vector[K] b;
  vector[K] c;
  real<lower = 0> sigma;
}

model {
  vector[N] mu;

  a ~ normal(16,0.3);      // hyperprior for the species intercept
  b ~ normal(-3,0.3);      // hyperprior for species slope
  c ~ normal(-0.2,0.1);    // hyperprior for species quadratic term slope
  sigma ~ normal(2.2,0.1); // prior for the standard deviation

  if (!prior_only) {
    for (i in 1:N) {
      mu = a[species[i]] + b[species[i]] * b1 + c[species[i]] * b2;
      Y[i] ~ normal(mu, sigma);
    }
  }
}

generated quantities {
  vector[N] log_lik;
  vector[N] Y_sim;
  vector[N] mu;

  for (i in 1:N){
      mu = a[species[i]] + b[species[i]] * b1 + c[species[i]] * b2;
    log_lik[i] = normal_lpdf(Y[i] | mu,sigma);
    Y_sim[i] = normal_rng(mu[i],sigma);
  }
}
```

Listing 4: A hierarchical model based on the first model. Note that instead of using individual parameters a or b in the linear model, the model instead uses values from a vector (list) of a and b, one value for each group. This allows the hierarchical model to learn about the whole group the same way the initial model learns about the whole dataset.

```
        elpd_diff se_diff
model2      0.0       0.0
model1 -10778.4     761.7
```

Listing 5: the result of using R package `loo` *to compare the two models. The preferred model is model2, the hierarchical models.*

4 Methods

4.1 data processing

The dataset is a combination from four different source papers. It includes biological data from Grossman and Ku [9], Herguera, Jansen, and Berger [10], and Marchitto et al. [15]. These are the only papers that were under consideration for inclusion which included δ^{18}O data separately for the carbonate and the water, temperature measurements with data specified per species. Originally a dataset from Bouvier-Soumagnac and Duplessy [3] was going to be included as well, however it was ultimately excluded because the data as represented in the paper's tables differed from that in the appendix, which led to uncertainty on the part of this thesis' writer as to which was the correct data to include. In order to compare the biological data to inorganically precipitated carbonate, a 2009 study by Kim et al. was included, leading to in total four data sources.

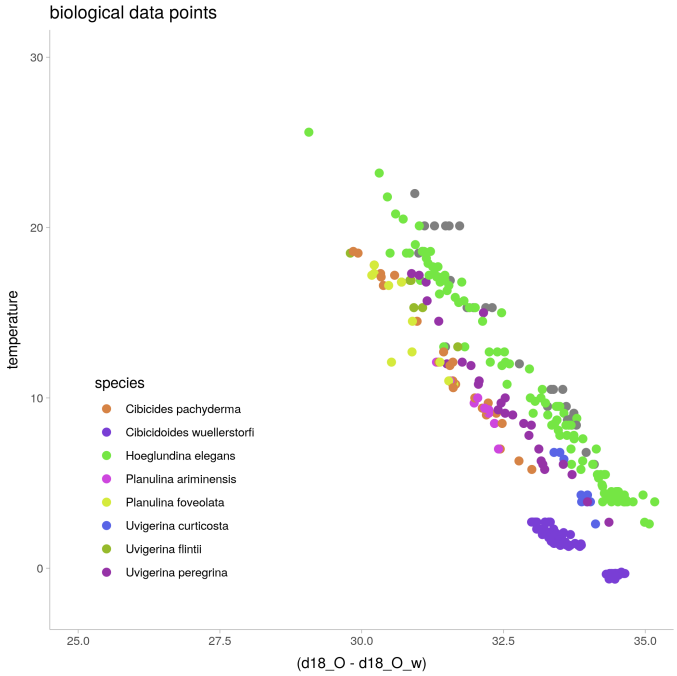

Figure 17: The datapoints from foraminifera carbonate. Visually, the groupings already have a clear suggestion of slope and dispersion.

For the Deming regression model we require uncertainty values for the δ^{18}O measurements. measurement errors for $\delta^{18}O_c$ were available for Grossman and Ku, Herguera, Jansen, and Berger, and Marchitto et al., but not for Kim et al. For $\delta^{18}O_w$, the uncertainty values were only available for Marchitto et al. Where these values were unavailable we had

to perform data imputation. Because this concerns uncertainty estimates, we chose to impute by taking the maximum value in our dataset and double it. This lets our model knows that we're more uncertain about these values than about those in other papers because we don't know the measurement error.

The dataset includes 295 datapoints from foraminifera, 28 datapoints from other biological carbonate (Gastropods and Scaphopods, from Grossman and Ku [9]) and 87 datapoints from abiological carbonate for a total of 410 datapoints. The datapoints are visualized in figures 17 and 18.

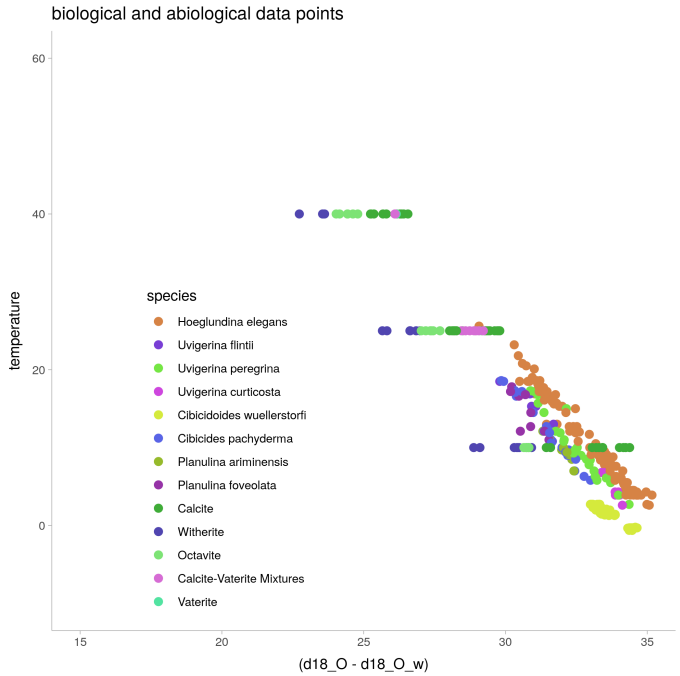

Figure 18: All datapoints (note changed Y-axis). The abiological datapoints were precipitated at different rates for comparison, as such the groups are more Rather than less spread out than the biological points. This visualization also includes biological datapoints in different phylums (Gastropods and Scaphopods).

At this point the form of the input data must briefly be addressed. As can be seen in the data table in the appendix, section 8, we have decided to have $\delta^{18}\mathrm{O_c}$ and $\delta^{18}\mathrm{O_w}$ both expressed relative to the VSMOW standard, instead of VPDB and VSMOW respectively for $\delta^{18}\mathrm{O_c}$ and $\delta^{18}\mathrm{O_w}$, as is done in the regressions from Bemis et al. [2] . This is one of the possible expressions of these values recommenced by IUPAC [4]. This is preferred simply because we want the comparison our model is learning to be as simple as possible. (hereafter, $\delta^{18}\mathrm{O_c}$ and $\delta^{18}\mathrm{O_w}$ values will always both be expressed relative to VSMOW unless otherwise noted). To subsequently use the draws from the model, the input $\delta^{18}\mathrm{O_x}$ values need to both be in VSMOW format. The preferred translation formula (which is also used by our data preprocessing) is the one recommended by IUPAC [4]:

$$\delta^{18}O_{\mathrm{x/VSMOW}} = 1.031 \cdot \delta^{18}O_{\mathrm{x/VPDB}} + 29.99$$

4.2 complete pooling models

This thesis includes 8 models in total. The most important change when compared to the original OLS models is using partial pooling across species. The models will be discussed in turn, starting with a fully pooled linear regression.

fully pooling Bayesian linear regression

This simple Bayesian model fits to the same equation, with the exception of the quadratic component. It therefore has only linear components, which makes it much easier to fit.

$$T(^{\circ}C) = a + b \cdot (\delta^{18}O_c - \delta^{18}O_w)$$

This is a very simple model, as such the stan code is shown here in its entirety except the `generated quantities` block. It being a complete pooling model, meaning there are only population level parameters to fit, so it can only give us the most general idea of our entire data set. It will necessarily have higher uncertainty than a model that fits on specific subgroups, as the next category of models will do.

```
data {
  int N;
  vector[N] Y;
  vector[N] x1;
  int prior_only;
}

parameters {
  real a;
  real b;
  real<lower = 0> sigma;
}

model {
  vector[N] mu;
  a     ~ normal(120,50);      // prior for the intercept
  b     ~ normal(-4,1);        // prior for linear term coefficient
  sigma ~ normal(0,2);         // prior for the standard deviation

  if (! prior_only) {
    mu    = a + b .* x1;       // linear model
    Y     ~ normal(mu, sigma); // likelihood
  }
}
```

Listing 6: The stan code for the linear regression.

This model was originally run with flat priors, those being the default option in stan when no priors are given. As can be seen in the stan code in listing 6, there are weakly informative priors now. Apriori we should expect this model to fit quite fast: it's a simple

regression model with only two parameters in the linear model. with enough data, the posterior will overwhelm any sane prior. The results section will set out a posterior predictive check for the priors shown here.

fully pooling Bayesian linear regression using a quadratic component

This Bayesian model will fit the original formula defined by Bemis et al.:

$$T(^{\circ}C) = a + b \cdot (\delta^{18}O_c - \delta^{18}O_w) + c \cdot (\delta^{18}O_c - \delta^{18}O_w)^2$$

This model features priors for each regression component, a linear model, and a likelihood.

$$\begin{array}{lcll} a & \sim & ? & \text{prior for the intercept} \\ b & \sim & ? & \text{prior for linear term coefficent} \\ c & \sim & ? & \text{prior for quadratic term coefficient} \\ \sigma & \sim & \text{Normal}_+(0, 2) & \text{prior for the standard deviation} \\ & & & \\ \mu_i & = & a + b(\delta^{18}O_c - \delta^{18}O_w) + c(\delta^{18}O_c - \delta^{18}O_w)^2 & \text{linear model} \\ T_i & \sim & \text{Normal}(\mu, \sigma) & \text{likelihood} \end{array} \tag{2}$$

we call this a simple quadratic model: It is a linear regression with a linear and a quadratic component. As discussed in the background section, we have doubts on whether there is a theoretical interpretation for a quadratic component, and there are computational reasons for leaving it out, so in later models we will run comparable regressions without the quadratic component and compare the models.

Later models in this thesis will be iterative improvements on these early fully pooled models. During development we have a rough idea what shape the posterior for newer models will take: we have seen the output of previous, simpler models with a range of priors, and there is enough data to overwhelm the prior. With both of those facts in hand, defining priors for later models is more straightforward than it sounds at first glance.

To define a prior for this model we will have to discuss our current data. A reasonable start, given the amount of research in the δ^{18}O temperature reconstruction, is to use the previously established values as inspiration. We could then construct normal distributions around those that show our model which values we find acceptable (we want our prior to encapsulate the knowledge that a value of -4.23 for our slope is pretty reasonable, while a value of -42.3 is not so reasonable).

The one parameter we've already given a prior is the likelihood's σ. This sigma denotes the final dispersion around the linear model when fit to the temperature. This dispersion should incorporate the uncertainty of the temperature measurement. The temperature uncertainty varies over different temperature measurements. Our half-normal prior peaks at 1.6, which overshoots all reported measurement uncertainty while not taking too much probability mass away from the region below 1. This prior will feature more in later models as well for similar reasons.

The shift in input data discussed in section 4.1 has an impact on the interpretation and values of our parameters, as the intercept and slope are now expressed on a VSMOW-VSMOW δ^{18}O ratio instead of a VPDB-VSMOW ratio. assuming we would take in $\delta^{18}O_c$ values on the VPDB scale in order to facilitate comparison to the Bemis et al. [2] regressions, the model would look like the following equation:

$$T(^{\circ}C) = a + b \cdot ((\delta^{18}O_{c_{\text{VPDB}}} \cdot 1.031 + 29.99) - \delta^{18}O_{w_{\text{VSMOW}}}) +$$

$$c \cdot ((\delta^{18}O_{c_{\text{VPDB}}} \cdot 1.031 + 29.99) - \delta^{18}O_{w_{\text{VSMOW}}})^2$$

So defining reasonable priors is now challenging. This is compounded by the choice of including a quadratic component: fitting this model with too wide priors will result in a very hard likelihood for stan to explore. Nevertheless, arithmetic shows the transformation will have an effect on the a parameter but not on the b parameter. Ultimately, the priors that were chosen correspond to the posterior of the first fully pooled model, which are the ones shown in the stan code in listing 6 and in the model statement of the main model, in section main model: hierarchical linear model.

4.3 partial pooling models, 2 level

main model: hierarchical linear model

This model is the natural extension of the fully pooled model by using partial pooling across species. The slope and intercept are assumed to be independent from each other, however the slopes and intercepts of the different species are assumed to be similar to each other, and fit together using partial pooling. This model explores the similarities and differences between species. The model definition is shown below. This model has a lot more moving parts to it, however when viewed from a distance, it still concerns just two parameters: a and b, leaving out the quadratic component c at least for the moment. The main difference in the linear model is that these are now group specific.

This is where the hyperpriors come in. Each $a_{species}$ is drawn from a prior distribution defined by the population level parameter a and σ_a. These form the link between the species. Because stan will fit all parameters together, the common reliance on the global parameter makes the group parameters tend together, even though they have individual room to fit.

This model once again does away with the quadratic component, mirroring the simpler fully pooled models. Having all four models allows making a number of interesting comparisons concerning the quadratic component:

- does the quadratic parameter lead to better fits?
- does it still lead to better fits when species are partially pooled?
- what is the computational cost of using a quadratic component?

Apart from the quadratic component, this model is ideally suited for studying the different values given to different species.

This model is the basis for further exploration in models, as such a full model description is given below.

$$
\begin{array}{lcll}
a & \sim & \text{Normal}(120, 50) & \text{hyperprior for the intercept} \\
b & \sim & \text{Normal}(-4, 1) & \text{hyperprior for linear term coefficent} \\
 & & & \\
\sigma_a & \sim & \text{Normal}_+(0, 50) & \text{sigma hyperprior for the intercept} \\
\sigma_b & \sim & \text{Normal}_+(0, 5) & \text{sigma hyperprior for linear term coefficent} \\
 & & & \\
\sigma & \sim & \text{Normal}_+(0, 2) & \text{prior for the standard deviation} \\
 & & & \\
a_{species} & \sim & \text{Normal}(a, \sigma_a) & \text{prior for the intercept} \\
b_{species} & \sim & \text{Normal}(b, \sigma_b) & \text{prior for linear term coefficent} \\
 & & & \\
\mu_i & = & a_{species} + b_{species}(\delta^{18}O_c - \delta^{18}O_w) & \text{linear model} \\
T_i & \sim & \text{Normal}(\mu, \sigma) & \text{likelihood}
\end{array}
\tag{3}
$$

This model has a lot more priors. These priors are partially based on the posteriors of the simpler complete pooling linear model. As such, we have a prior for the intercept just above 100, with a sizable standard deviation of 50. Our linear slope parameter also follows the previous model, and is much in line with the slope we see in the literature. Sigma's are all halfnormal distributions, which is recommended by McElreath.

hierarchical model using a quadratic component

This model is very similar to the main model, except it incorporates the quadratic component again, allowing for the comparison between a linear model with and without a quadratic component to be made on models with two levels as well as on the simpler fully pooled models.

$$
\begin{array}{lcll}
a & \sim & \text{Normal}(120, 50) & \text{hyperprior for the intercept} \\
b & \sim & \text{Normal}(-4, 1) & \text{hyperprior for linear term coefficent} \\
c & \sim & \text{Normal}(0, 5) & \text{hyperprior for quadratic term coefficient} \\
 & & & \\
\sigma_a & \sim & \text{Normal}_+(0, 50) & \text{sigma hyperprior for the intercept} \\
\sigma_b & \sim & \text{Normal}_+(0, 5) & \text{sigma hyperprior for linear term coefficent} \\
\sigma_c & \sim & \text{Normal}_+(0, 5) & \text{sigma hyperprior for quadratic term coefficient} \\
 & & & \\
\sigma & \sim & \text{Normal}_+(0, 2) & \text{prior for the standard deviation} \\
 & & & \\
a_{species} & \sim & \text{Normal}(a, \sigma_a) & \text{prior for the intercept} \\
b_{species} & \sim & \text{Normal}(b, \sigma_b) & \text{prior for linear term coefficent} \\
c_{species} & \sim & \text{Normal}(c, \sigma_c) & \text{prior for quadratic term coefficient} \\
 & & & \\
\mu_i & = & a_{species} + b_{species}(\delta^{18}O_c - \delta^{18}O_w) + & \\
 & & c_{species}(\delta^{18}O_c - \delta^{18}O_w)^2 & \text{linear model} \\
T_i & \sim & \text{Normal}(\mu, \sigma) & \text{likelihood}
\end{array}
\tag{4}
$$

The priors of this model are very similar to the ones presented for the previous model. The quadratic parameter c is assumed to be close to 0, but the chosen prior is deliberately quite wide to allow the model to explore alternate ways of fitting to the data.

hierarchical model with separate δ^{18}O parameters

This model pulls apart the $(\delta^{18}O_c - \delta^{18}O_w)$ variable to separate $\delta^{18}O_c$ and $\delta^{18}O_w$ variables with their own parameters. This allows us to see whether a regression will recover the theoretically grounded $(\delta^{18}O_c - \delta^{18}O_w)$ term, or whether it will see some other way to fit the data.

$$
\begin{array}{lcll}
a & \sim & \text{Normal}(120, 50) & \text{hyperprior for the intercept} \\
b_1 & \sim & \text{Normal}(0, 10) & \text{hyperprior for linear term coefficent} \\
b_2 & \sim & \text{Normal}(0, 10) & \text{hyperprior for linear term coefficent} \\
 & & & \\
\sigma_a & \sim & \text{Normal}_+(0, 50) & \text{sigma hyperprior for the intercept} \\
\sigma_{b1} & \sim & \text{Normal}_+(0, 5) & \text{sigma hyperprior for linear term coefficent} \\
\sigma_{b2} & \sim & \text{Normal}_+(0, 5) & \text{sigma hyperprior for linear term coefficent} \\
 & & & \\
\sigma & \sim & \text{Normal}_+(0, 2) & \text{prior for the standard deviation} \\
 & & & \\
a_{species} & \sim & \text{Normal}(a, \sigma_a) & \text{prior for the intercept} \\
b_{species} & \sim & \text{Normal}(b, \sigma_b) & \text{prior for linear term coefficent} \\
 & & & \\
\mu_i & = & a_{species} + & \\
 & & b_{1species}\delta^{18}O_c + b_{2species}\delta^{18}O_w & \text{linear model} \\
T_i & \sim & \text{Normal}(\mu, \sigma) & \text{likelihood}
\end{array}
\tag{5}
$$

The $b1$ and $b2$ priors are interesting. Because the model is intended to be let free to interpret the data, we center a normal distribution on 0, with tails twice as wide as the typical slope parameter value in the posterior of our main hierarchical model.

hierarchical model with deming regression

This model attempts to capture the uncertainty inherent in the measurement of the $\delta^{18}O_c$ and $\delta^{18}O_w$ variables. Instead of fitting a linear model on the input variables directly, the linear model is fed faux data which is created point by point by drawing from a normal distribution centered around the $\delta^{18}O_x$ as a mean, with a standard deviation given by $\sigma_{\delta^{18}O_x}$

$$
\begin{array}{lcll}
a & \sim & \text{Normal}(120, 50) & \text{hyperprior for the intercept} \\
b & \sim & \text{Normal}(-4, 1) & \text{hyperprior for linear term coefficent} \\
 & & & \\
\sigma_a & \sim & \text{Normal}_+(0, 50) & \text{sigma hyperprior for the intercept} \\
\sigma_b & \sim & \text{Normal}_+(0, 5) & \text{sigma hyperprior for linear term coefficent} \\
 & & & \\
\sigma & \sim & \text{Normal}_+(0, 2) & \text{prior for the standard deviation} \\
 & & & \\
a_{species} & \sim & \text{Normal}(a, \sigma_a) & \text{prior for the intercept} \\
b_{species} & \sim & \text{Normal}(b, \sigma_b) & \text{prior for linear term coefficent} \\
 & & & \\
\eta_{\delta^{18}O_c} & \sim & \text{Normal}(\delta^{18}O_c, \sigma_{\delta^{18}O_c}) & \text{draw of the carbonate ratio} \\
\eta_{\delta^{18}O_w} & \sim & \text{Normal}(\delta^{18}O_w, \sigma_{\delta^{18}O_w}) & \text{draw of the water ratio} \\
 & & & \\
\mu_i & = & a_{species} + b_{species}(\eta_{\delta^{18}O_c} - \eta_{\delta^{18}O_w}) & \text{linear model} \\
T_i & \sim & \text{Normal}(\mu, \sigma) & \text{likelihood}
\end{array}
\tag{6}
$$

The deming regression incorporates a form of deviation which we can give by data input, thereby improving uncertainty estimation elsewhere in the model.

The priors of this model are the same as the main hierarchical model.

hierarchical model with a correlation matrix

Up to now, all models have had the assumption baked in to their structure that there is no correlation between slope and intercept. By using a correlation matrix we can allow our model to explore that assumption. Our model structure now looks very different. Instead of defining a prior and a parameter per line, we are now defining parameters in terms of each other. It is best to start once again with the likelihood and the linear model. These have not changed. looking ahead to our priors, we see that the prior for our temperature deviation also has not changed.

$$
\begin{array}{lcll}
\begin{bmatrix} a \\ b \end{bmatrix} & \sim & \text{MVNormal}(\begin{bmatrix} 120, -4 \end{bmatrix}, \begin{bmatrix} 50 & 1 \\ 1 & 5 \end{bmatrix}) & \text{prior for population parameters} \\
\sigma^a_{species} & \sim & \text{Normal}(0, 50) & \text{sigma prior for species intercept} \\
\sigma^b_{species} & \sim & \text{Normal}(0, 5) & \text{sigma prior for species slope} \\
\text{R} & \sim & \text{LKJcorr}(1) & \text{prior for correlation matrix} \\
a_{species} & \sim & \text{Normal}_+(a, \sigma^a_{species}) & \text{prior for the species intercept} \\
b_{species} & \sim & \text{Normal}_+(b, \sigma^b_{species}) & \text{prior for the species slope} \\
 & & & \\
\text{S} & = & \begin{pmatrix} \sigma^a_{species} & 0 \\ 0 & \sigma^b_{species} \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma^a_{species} & 0 \\ 0 & \sigma^b_{species} \end{pmatrix} & \text{construct covariance matrix} \\
 & & & \\
\sigma & \sim & \text{Normal}_+(0, 2) & \text{prior for the standard deviation} \\
 & & & \\
\mu_i & = & a_{species} + b_{species}(\delta^{18}O_c - \delta^{18}O_w) & \text{linear model} \\
T_i & \sim & \text{Normal}(\mu, \sigma) & \text{likelihood}
\end{array}
\tag{7}
$$

Starting with the population parameters, a and b are defined in terms of a multivariate normal distribution. . The sigma priors are similar to previous models. Then we come to the LKJ prior, which defines a prior for the correlation matrix. The LKJ prior is defined over a correlation matrix R and parameterized by a single argument η, which denotes the degree to which every species and parameter is going to be correlated. High levels of η means lower correlation.

partial pooling model with a third level of composition

Extending the model to a third level allows for the model to have partial pooling in multiple levels. We chose to add a layer for composition, which means there are three layers of parameters in this model:

1. population
2. composition
3. species

Each composition group is also based on the a and b population parameters, and each species is based on the a and b parameters of it's composition. Now the model knows that species that have the same composition are more similar than species that don't have the same composition.

$$
\begin{array}{lcll}
\begin{bmatrix} a \\ b \end{bmatrix} & \sim & \text{MVNormal}(\begin{bmatrix} 120, -4 \end{bmatrix}, \begin{bmatrix} 50 & 1 \\ 1 & 5 \end{bmatrix}) & \text{prior for population parameters} \\
\sigma^a_{species} & \sim & \text{Normal}(0, 50) & \text{sigma prior for species intercept} \\
\sigma^b_{species} & \sim & \text{Normal}(0, 5) & \text{sigma prior for species slope} \\
\text{R} & \sim & \text{LKJcorr}(1) & \text{prior for correlation matrix} \\
a_{species} & \sim & \text{Normal}_+(a, \sigma^a_{species}) & \text{prior for the species intercept} \\
b_{species} & \sim & \text{Normal}_+(b, \sigma^b_{species}) & \text{prior for the species slope} \\
a_{composition} & \sim & \text{Normal}_+(a, \sigma^a_{composition}) & \text{prior for the composition intercept} \\
b_{composition} & \sim & \text{Normal}_+(b, \sigma^b_{composition}) & \text{prior for the composition slope} \\
\\
\text{S} & = & \begin{pmatrix} \sigma^a_{species} & 0 \\ 0 & \sigma^b_{species} \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma^a_{species} & 0 \\ 0 & \sigma^b_{species} \end{pmatrix} & \text{construct covariance matrix} \\
\\
\sigma & \sim & \text{Normal}_+(0, 2) & \text{prior for the standard deviation} \\
\\
\mu_i & = & a_{species} + b_{species}(\delta^{18}O_c - \delta^{18}O_w) & \text{linear model} \\
T_i & \sim & \text{Normal}(\mu, \sigma) & \text{likelihood}
\end{array}
\tag{8}
$$

5 Results

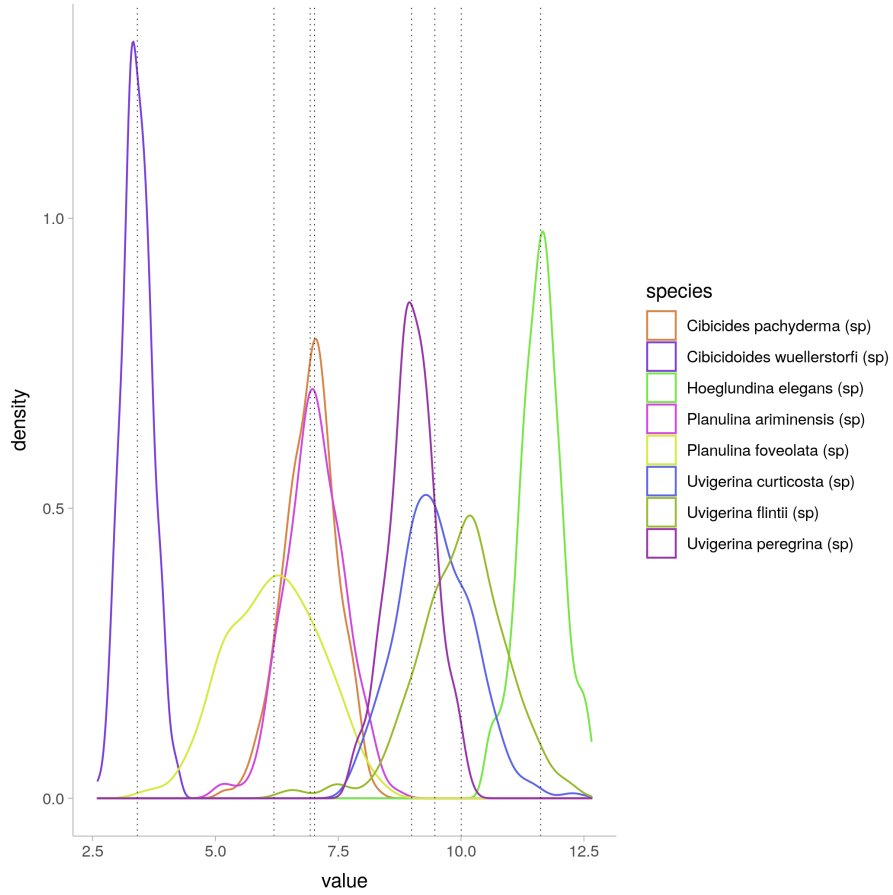

Figure 19: A visualization of temperature distributions for various species in the main model, given a $\delta^{18}O_c$ of 33, a $\delta^{18}O_w$ of 0.3 (note: both are in VSMOW), and a respective standard deviation of 0.08 and 0.7 for those measurements. This figure shows the corresponding uncertainty for each species, as well as the means.

Each model was run at least once, however the main model and three level model have been run more than once with expanded datasets, leading to 11 model fits to cover in the results section. Each model has a corresponding section in the appendix with regression visualizations per species and composition where applicable, posterior predictive check figures, model summaries and loo values. Some sections will feature autocorrelation plots and traceplots. The species in the model summaries will be numbered. To check which species a number corresponds to, check the species name tables in section 8.1.

Unless otherwise noted, all models have good $\hat{R}$ values ($\hat{R} \leq 1.01$), effective sample sizes (> 1000 effective samples) and low Monte Carlo standard errors for all parameters (all MCMC se are around two orders of magnitude lower than the estimated standard deviation of the parameters), as well as good visual checks in trace plots, autoregression plots, and posterior predictive checks.

Also unless otherwise noted, all model fits are run with 4 chains on 1000 warmup samples and 1000 samples for 4000 post-warmup samples total. Almost all models benefited from a higher maximum treedepth at 15, and all were run with an adapt_delta setting of 0.99, excepting the fully pooled models and the main model with foraminifera data, which

had a setting of 0.95 due to lower complexity.

Due to the inclusion of 11 model fits, the total results are quite large. Instead of including all long model summaries, visual checks and regression lines, the results section will feature the most notable ones, with references at appropriate points to the appendix which holds all visualizations, visual checks, model summaries, and visual checks.

5.1 fully pooled linear regression (foraminifera data)

This is the simplest model, which compares quite closely to the original regressions, excepting that it has no quadratic component. It fits extremely fast (2 seconds), and has a reasonable fit to the data. The posterior predictive check shows it doesn't capture the shape of the data: rather it simply puts most of its probability density in the center of the data mass. This model has an a parameter for population intercept, a b parameter for population slope and a σ parameter for residual dispersion on the temperature.

The population linear regression shown in figure 25 has deceptively narrow bounds when compared to population linear regressions with later partially pooled models. This makes sense when looking at the model summaries: later summaries have comparatively lower values for the residual standard deviation, as they capture the data structure much better. As we have not included the residual standard deviation in the population regression, these figures significantly under report temperature uncertainty.

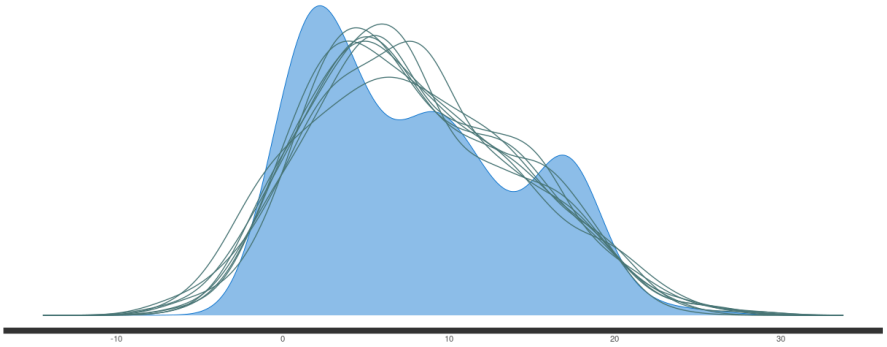

Figure 20: posterior predictive check for the fully pooled linear regression model. Showing the data (colored blue) and a sample of simulated data based on the draws of the model (lines).

The trace plots of this model (shown in picture 21, in this case of the sigma parameter) look healthy. There are 4 chains producing lines, which look like a healthy "fuzzy caterpillar": there is good mixing, no chains spend more time in one part than in another part.

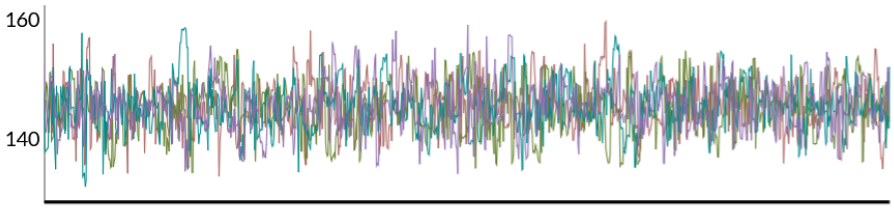

Figure 21: trace plot of the sigma parameter of the fully pooled linear model.

5.2 fully pooled linear regression using a quadratic component (foraminifera data)

The fully pooled model with a quadratic component captures the formula which includes a third component, the c parameter which relates to the square of the $\delta^{18}O$ difference. This model was run to compare its fit to the simplest model. The results are mostly unsurprising: the fit to data as seen by the posterior predictive check in figure 22 is the same as the previous model, and the model summary in section 8.4 in the appendix shows these two first models to be virtually identical in parameter estimates.

This model was compared to the first fully pooled model by use of a leave-one-out cross validation (loo) approximation known as PSIS [23], which gives an estimate of leave one out log likelihood scores without fitting the model again for each datapoint. The estimated log likelihood values for this model and the previous model are -730.2 and -730.6 respectively, with an estimated standard deviation of 7.0 for both. This means that from the point of view of out of sample predictive performance, these models perform exactly the same: the difference is an order of magnitude smaller than the standard deviation on the log likelihood. However, there are some reasons we prefer the first model to this one.

Firstly, there does not seem to be a theoretical reason to presume temperature will depend quadratically on the $\delta^{18}O$ ratio, so the first model seems more theoretically grounded. Secondly, this model fits much slower than the first, because there are much more ways to produce a fit to the data with both a linear and a quadratic component rather than with only a linear component. It's effectively the first step on the polynomial ladder: we might as well add a cubed or even a quartic polynomial component to increase our fit to data. In the absence of a theoretical reason to include a quadratic component, we prefer to make our models as simple as possible.

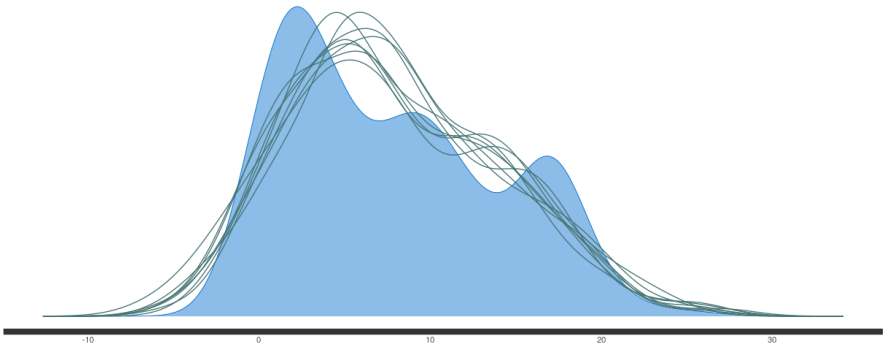

Figure 22: posterior predictive check of the fully pooled model with quadratic component.

5.3 main model: hierarchical linear model (foraminifera data)

This model allows for the use of partial pooling in order to improve on previous regressions. It has a partial pooling level for species, as we know each species has idiosyncrasies which will set it apart from other species and lead to different intercepts and slopes, while at the same time no species of foraminifera is so distant that knowing it's parameters imparts no knowledge of the other species.

This model fits relatively fast, and it fits the data very well, as can be seen in the posterior predictive check in figure 23. Note also the lower estimation for sigma, the standard deviation of the residual. Because it captures population parameter uncertainty separately in $sigma_a$ and $sigma_b$ for intercept and slope respectively, it has a lower residual

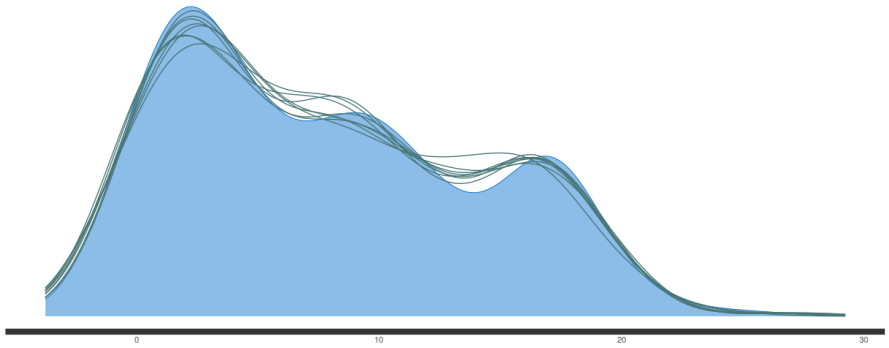

Figure 23: posterior predictive check for the main model. The shape of the predictive distributions follows that of the data fairly directly.

error. Another possible reason the residual error is lower is because the structure of this model mirrors the structure of the data much better than the fully pooled models: the fully pooled models assume that all data points are described by one regression, while the main model allows each species their own parameters.

There are psis-loo values for all models, which were run with models converted from cmdstanr to rstan, however these values cannot be used, as psis-loo reports diagnostic $\hat{K}$ values which are higher than 0.7. psis-loo therefore is uncertain of the value of the Monte Carlo standard error of the log likelihood, which means the point estimate and standard error it reports cannot be trusted.

In the case of this model, we have worked around this limitation by running the model wholly in rstan instead of converting it afterwards, which leads to extremely similar estimations from loo, except that high $\hat{K}$ values could be worked around by utilizing moment matching, resulting in very low Monte Carlo standard errors. The resulting log likelihood value compares very favorably to those of the fully pooled models with a value of -346.7 and a standard error of 24.4 (complete output of the loo package can be seen in the appendix textblock 30. Also see a visualization of the $\hat{K}$ values for the loo results over this fit in figure 35).

Later models will not be compared on loo values as they could not be run in rstan due to syntax differences between stan versions. A rewrite was not completed in time.

5.4 main model: hierarchical linear model: (all biological data)

These results are from the same model, except we're including biological data from Scaphopods and Gastropods from the Grossman and Ku study. The focus of this study is foraminifera data, so these results are only of secondary importance. We include data from these other phylums for comparison of model fit. This model takes longer to fit due to the higher number of parameters and the larger dataset. Nevertheless the results appear close: take for example the intercept of the first species, *Cibicides pachyderma*, which has a mean at 146.96 (2.5%: 137.36, 97.5%: 156.80) in the main model and a mean at 144.66 (2.5%: 134.88, 95.5%: 154.55) in this model, both with a standard deviation higher than 5.

(note when reading the model summaries that the numbers correspond to species in their respective species name list, which are 1 and 6 in respectively the earlier fit and this fit for *Cibicides pachyderma*, see section 8.1 for species name tables).

5.5 hierarchical model using a quadratic component (foraminifera data)

This model fit is over the original dataset again, with a hierarchical model that includes a quadratic component. Just like in the fully pooled case, including a quadratic component increases required sampling time. When looking at the model summary and regression lines it is clear that in most cases, the uncertainty is higher because the quadratic parameter has an outsized impact. For one example see figure 66, showing the linear regression for *Cibicidoides wuellerstorfi*, where the data is very tightly grouped. In most cases, the c parameter is within one standard deviation from 0, meaning a type S error is quite possible. This makes interpretation difficult. This is one more reason why we dislike using a quadratic component.

5.6 hierarchical model with separate oxygen ratio parameters (foraminifera data)

This model doesn't have a theoretical basis: no temperature could be derived from only the $\delta^{18}O_c$ value, or from only the $\delta^{18}O_w$ value, as it is the difference between the two that is correlated with temperature. Nevertheless, the results show a good fit to data, and a reasonable run time. Apriori we might have expected the model to retrieve the same fit as the main model: effectively, the b1 and b2 parameters should cancel out given enough data. Instead, the model fits well with an alternative fit with different estimations for intercept and slope. Instead of validating our belief that our main model is the best representation of reality, this evokes the question whether there are alternate ways to explain the data. It is curious that this model finds a good fit to data with a completely different set of parameters. While this is an interesting result, it is not easy to provide a theoretical understanding of it as there is no theoretical basis underlying this model.

5.7 hierarchical model with deming regression (foraminifera data)

Deming regression allows the model to take in estimates on the measurement uncertainty of the $\delta^{18}O_c$ and $\delta^{18}O_w$ values, which allows the model to more accurately describe the residual uncertainty. This leads to a better estimate of uncertainties. Unfortunately this structure leads to autocorrelation, which means to get the same effective number of parameters this model has to be run with twice the number of sampling iterations (and twice the number of warmup iterations). See figure 76 for the autocorrelation plot, and the healthy autocorrelation plot for the main model in figure 31 for a comparison. However with a high number of samples we don't see a problem getting good enough effective samples, and the

traceplot nor the $\hat{R}$ diagnostics show cause for concern. As expected, this model has lower residual dispersion than the main model.

5.8 hierarchical model with a correlation matrix (foraminifera data)

All models up to this point assume there is zero correlation between slope and intercept. We can relax this assumption to see whether this impacts our fit to data, in effect checking whether the assumption holds. This was done in this model. the parameters in the model summary in look slightly different. Each parameter is of the form `ab_sp[` $species_{nr}$, *parameter* `]`, leading to the two parameters per species, 1 being intercept, 2 being slope. The posterior predictive check looks the same, and the parameters are once again all well within one standard deviation of each other. This is evidence that the zero correlation assumption doesn't hamper the main model to find a good fit, as we can retrieve virtually the same fit from this model, even though it is more general. The correlation matrix, generated after the model was run, shows that there is actually a very high anticorrelation of -0.96 between slope and intercept, meaning a higher intercept is associated with a lower slope and vice versa.

5.9 partial pooling model with a third level of composition (foraminifera data)

The main model is an improvement over the fully pooled model because the partial pooling model structure informs the model better of the data generation process. This model increases that model structure by also telling the model the differences in composition between species, as calcitic foraminifera will be more alike than calcitic and aragonitic foraminifera. We want to know if this leads to a better fit to data, and to better estimates of $\delta^{18}O$ uncertainty. Secondarily we want to include abiological data in the model, which doesn't have a species, but does have a composition. By including a level of composition between species and population we can combine the biological data and the abiological data at the appropriate level, namely that of the composition. While the abiological data from the lab do not have a species to speak of, in the stan code it will have a species corresponding to it's composition. These "species" indicate that it is abiological data from the lab.

Because the abiological data is from the laboratory, this model requires special care in uncertainty estimation of the temperature residuals. As laboratory measurements have a different measurement uncertainty, we're using `individual_sigmas` as a vector of uncertainties, which for each datapoint points either to the `sigma_obs` parameter if biological, being fit as before, or if from the laboratory it is set to `sigma_T_lab`, a model input that is set to the measurement uncertainty given by the lab.

Parameter estimates are remarkably similar to main model, again falling well within 1 standard deviation.

5.10 partial pooling model with a third level of composition (biological data)

This model has a divergence that could not be mitigated as of yet. It is possible the non-foraminiferal data is too dissimilar to the foraminiferal data to run within the same partially pooled layer.

5.11 partial pooling model with a third level of composition: (foraminifera and lab data)

This model includes the foraminifera data and laboratory data. As discussed in section 5.9, the parameter `sigma_obs` is now only used for the foraminifera data, while the laboratory data has a set uncertainty. This results in a higher estimation for `sigma_obs` of 1.41 as opposed to the typical value of 0.76 when compared to the main model or even the previous three level models. There is an anomalously low standard deviation on the parameters for the Vaterite samples: the intercept stands at 173.78 (sd: 1.35, 2.5%: 171.08, 95.5%: 176.41), and the slope at -5.12 (sd: 0.05, 2.5%: -5.21 , 95.5%: -5.02). The deviation of especially the slope is much tighter than other species, while there are only 2 samples for vaterite. The model must be drawing a lot of information from the priors here: there is virtually no information in the likelihood from which to draw that information form. A visual prior predictive check compared to a posterior predictive check could have confirmed this but was not performed.

This three level model results in an interesting artifact in two figures in the appendix: figure 137 shows the model's fit at the higher level of composition for the Calcite-Vaterite mixtures, while figure 138 shows the model's fit at the species level for the same mixtures. As discussed in section 5.9, due to the construction of the model the abiological laboratory data has been included into the species level as well. Figure 138 shows a tighter fit: the model is quite certain of the appropriate values. But at the higher level, there is effectively but one datapoint: that single datapoint of the "species" of Calcite-Vaterite, which was fit at the level of species, and is now input for the level of composition. Correspondingly, the standard deviation is much larger, and the model is free to have an intercept which visually doesn't agree very well with the corresponding data as it is partially pooled with the other compositons, which are pulling the parameter values closer to those for the other groups in the composition level.

6 Discussion

6.1 quadratic fit

The base regression formula includes a quadratic component (which we call c) which is useful mainly to increase fit to data. In the case of the fully pooled models we note in section 5.2 no significant change to the parameters a and b, while the quadratic parameter c is estimated at 0.01 with a standard deviation of 0.02. This leads us to believe the quadratic parameter is not adding anything to the model. Assuming for the moment that the quadratic fully pooled model is a faithful model, the quadratic parameter of the real world could very well fall a standard deviation away from the mean of our posterior, in which case the results of our model as fit currently would have a postive c parameter while it should have a negative parameter, a type S error. It is unclear what impact that would have on our understanding of the temperature proxy, as the c parameter has no theoretical basis.

Furthermore, the loo-psis comparison in the fully pooled case shows there is virtually no difference in out of sample prediction accuracy when including the quadratic component, resulting in an estimated log probability density of -730.2 for the case with a quadratic component, and of -730.6 in the case without. The standard error is estimated to be 7.0, completely swamping the difference in predictive accuracy. In the partial pooling case the quadratic component fares no better as the standard deviation for each $c_{species}$ is still close to the mean value, or in the case of *Uvigerina peregrina* is even an order of magnitude

larger. The same questions as with the fully pooled model's high standard deviation arise again.

Another drawback of the quadratic term is that it increases the complexity of the model. In the course of this thesis the priors were developed iteratively based on the posterior of simpler models. For the first fully pooled model and the main model the data overwhelm the priors: both can be fit to the same posterior summary statistics with a wide variety of priors, including flat priors. This is not the case for the more complex quadratic models. Furthermore, these models take much longer to fit: whereas the simpler model takes 2 seconds, this model takes 42 seconds (see section 8.2) to fit. This model takes even longer when the priors are less informative, as the quadratic parameter wildly impacts the behavior of the posterior. In the case of the partially pooled models the quadratic model still takes an order of magnitude longer to fit than the main model (722 seconds and 69 seconds respectively) For all these reasons we reject the use of a quadratic parameter and recommend the main model with the formula

$$T \sim \text{Normal}(\ a + b(\delta^{18}O_c - \delta^{18}O_w)\ ,\ \sigma)$$

6.2 peeling back residual dispersion

The structure of the hierarchical models allows for a better understanding of where the uncertainty regarding our temperature estimates comes from. Firstly, the difference between the fully pooled models and the partially pooled models can not be overstated. the residual dispersion of three models is compared in table 2. For the fully pooled model the residual dispersion is estimated at 2.8. Such a large value makes inferences from the model suspect, and draws into question whether it accurately captures the data generation process. This is already known not to be the case: there is a material difference between the species of the samples. By allowing the model to incorporate this knowledge, we arrive at the main model, which has a residual dispersion of 0.79. This on its own implies that the species have a significant impact on the distribution of the temperature.

One consequence that flows from fitting a partially pooled model is that we can infer from the species that are fit already something about new species that have no regression fit yet: we can use the population level as an estimate for temperature from those species which have not been processed yet. This is advantageous, however as pointed out, a specimen's species has a significant impact on the distribution of the temperature, which has the consequence that the population parameters have a much higher standard deviation. A visualization including the population has been included in figure 24. Compare with the earlier temperature distribution visualization in figure 19, where the differences between the distributions for the species are easier to see. It is clear that for a proper estimation the species of the specimen should be known.

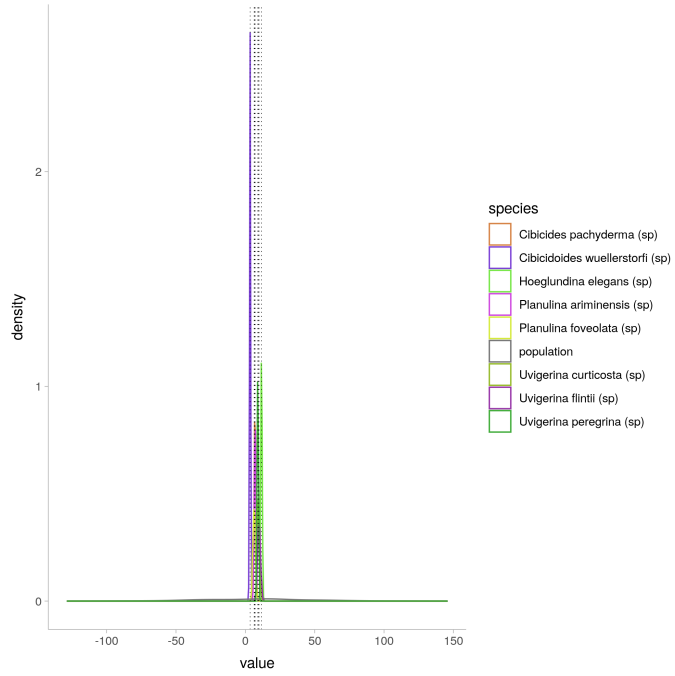

Figure 24: A visualization of temperature distributions similar to the earlier visualization, with the same values. This visualization includes the population distribution.

In general, the residual dispersion incorporates many sources of uncertainty. Some sources are readily quantifiable: the measurement error over both the dependent and the independent variables can be quantified and put into the model, as is done in the Deming model for the independent variables. This model has a residual dispersion that is much lower than the main model because the structure incorporates uncertainty better: as it includes the measurement error of the $\delta^{18}O_c$ and $\delta^{18}O$, the residual less than half the value of the main model.

model	mean	sd
fully pooled model	2.8	0.1
main model (fora data)	0.76	0.03
deming model	0.36	0.05

Table 2: comparison of the residual dispersion of three models.

As of yet there is no model that also incorporates known uncertainty over the dependent variable, although this is a potential improvement to be discussed in section 6.3. Such an enhancement would come at the cost of a much more complicated model with potential identifiability issues. It also depends on knowledge of the measurement uncertainty. This is largely unknown for the current dataset, as such it has not been attempted in this thesis.

The residual dispersion includes more sources of uncertainty which are not represented yet in the current models. this may include the impact of the environment on the foraminifera, the seasonality of their migration and precise location. To peel back the dispersion more of these variables could be studied and eventually included into models such as these. This will improve the temperature estimation.

For a most conservative aim at our posterior distribution of a temperature for a given value of $\delta^{18}O$ difference, we should include the residual as well. But this includes the scatter

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
mu_a	1.00	1577	130.34	12.17	0.31	106.01	154.34
mu_b	1.00	1451	-3.79	0.36	0.01	-4.56	-3.07
sigma_a	1.00	1902	31.61	10.20	0.23	17.30	57.03
sigma_b	1.00	1572	0.97	0.35	0.01	0.51	1.88
sigma	1.00	4983	0.76	0.03	0.00	0.70	0.82
log-posterior	1.00	969	-70.13	4.20	0.14	-79.32	-62.86
a[Cibicides pachyderma]	1.00	4776	146.96	5.02	0.07	137.36	156.80
a[Cibicidoides wuellerstorfi]	1.00	3973	75.39	5.34	0.08	64.95	85.87
a[Hoeglundina elegans]	1.00	3920	142.80	1.64	0.03	139.51	146.00
a[Planulina ariminensis]	1.00	4238	129.01	15.86	0.24	97.70	160.01
a[Planulina foveolata]	1.00	4961	144.99	11.90	0.17	122.43	169.44
a[Uvigerina curticosta]	1.00	2913	148.69	20.40	0.38	109.84	190.99
a[Uvigerina flintii]	1.00	4585	115.95	13.59	0.20	87.91	142.36
a[Uvigerina peregrina]	1.00	4784	146.21	5.09	0.07	136.47	156.14
b[Cibicides pachyderma]	1.00	4793	-4.28	0.16	0.00	-4.59	-3.98
b[Cibicidoides wuellerstorfi]	1.00	3975	-2.20	0.16	0.00	-2.51	-1.89
b[Hoeglundina elegans]	1.00	3919	-4.01	0.05	0.00	-4.11	-3.91
b[Planulina ariminensis]	1.00	4242	-3.73	0.50	0.01	-4.70	-2.75
b[Planulina foveolata]	1.00	4970	-4.24	0.39	0.01	-5.03	-3.51
b[Uvigerina curticosta]	1.00	2904	-4.26	0.60	0.01	-5.51	-3.11
b[Uvigerina flintii]	1.00	4579	-3.24	0.44	0.01	-4.09	-2.33
b[Uvigerina peregrina]	1.00	4777	-4.20	0.16	0.00	-4.50	-3.90

Table 1: main hierarchical linear model summary. This version has species names filled in.

of data in our input, and is necessarily an overestimate of the temperature dispersion. A visualization of the posterior distribution of temperature for a number of species has been included in figure 19. This visualization does not include the residual, but only the fitted values of the parameters `a`, `b`, and `sigma` per species, for a given input of δ^{18}O values.

Compare the visualization with the model summary block in table 1. The population distribution is omitted as its standard deviation is far too large to maintain readability for these distributions. This visualization shows how far apart the means are for different species,and how the distributions differ. The species with more datapoints in the dataset have a correspondingly tighter distribution, while those with less datapoints will have a larger standard deviation. This figure is produced by taking the resulting draws from the stan object and running the regression equation to a given input of δ^{18}O values.

As noted in section 5.8, the models with correlation matrices allow us to explore the impact of relaxing the assumption that slope and intercept are uncorrelated. Instead of recovering precisely the same fit, these models show a high anticorrelation between slope and intercept (in the case of the two level model with correlation matrix, a correlation value of -0.96, and with a standard devation of 0.08, so -1 is well within one sd). Nevertheless a posterior predictive check, see figure 88, shows this model performs adequately, and by inspecting the model summary in the same section we can see that the mean parameters are quite similar.

6.3 Future work

One immediate improvement to these models would simply be to train them on a dataset incorporating more data from the species that were included, and to include more species. A new dataset could be constructed. To further the work of disambiguating the residual dispersion, this dataset should include measurement errors for both the dependent and independent variables; as could be seen in the Deming regression model, the residual dispersion incorporates much uncertainty that can and should be categorized at the proper parameters to aid more precise measurements of the posterior distribution of temperature when using the oxygen temperature proxy.

This model could be used to perform a reassessment of the temperature timeline in Westerhold et al. [24]. To perform this analysis, a dataset must be constructed naturally including as much δ^{18}O measurements as possible, including a categorization by species, without any traditional offsets which are frequently used in this type of analysis: the species differences are handled more effectively by our models internally because of the partial pooling. Therefore the dataset from Westerhold that is generously made available is not immediately sufficient: not only does it not include the measurement errors (which would allow the use of the deming model), but it does not include species specification at all. To perform this analysis with the same dataset, the species must be inferred.

By using these datapoints in combination with the resulting models from this thesis, we can recover a timeline series comparable to the original timeline series presented by Westerhold et al. [24] as shown in figure 2. Just like that study, the data would require smoothing, for which we could use a Gaussian process. This would allow for capturing signal in the noise, while the use of the Bayesian models trained in this thesis allows the propagation of uncertainty all the way from the training dataset to the final temperature timeline estimations and visualizations. It is very likely we will observe wider bands of uncertainty than the current visualization suggests, as we can visualize much more of the sources of uncertainty, rather than only the scatter of the data.

One more improvement barring the direct use of measurement error could be a more sophisticated data imputation mechanism. Currently, unknown measurement errors are set at a multiple of two times the highest measurement error, but a better imputation could be to fit a parameter on this uncertainty. This does lead to a much more complicated model that could have identifiability issues, which is the reason it was not attempted in this thesis. While this is a possible improvement it is much preferred to obtain a more complete dataset for reasons laid out above.

There are more variations on the models which can be attempted; firstly there are possibilities for a third level other than composition: firstly when including non foraminifera data a third level could be a level specifying genus, which could make the model incorporate this data much better. Alternatively a level specifying the functional group could be attempted with a dataset that included more planktic datapoints.

Finally the development of tooling around the use of these models, this code, and this dataset could be very useful. One idea is to setup a web application where researchers can upload δ^{18}O measurements and get back a full posterior distribution of the estimation of temperature for further use.

6.4 Conclusion

In conclusion, the use of probabilistic programming in this thesis has lead to an improved understanding of the oxygen isotope temperature proxy in the following ways: It has provided a species specific regression line complete with uncertainty estimation per species

and for the entire population. The development of a partially pooled model, our main model above, allows for a quantification of uncertainty which has as of yet not been achieved for the oxygen isotope temperature proxy. The model is robust to a variety of priors including flat priors and has good diagnostic values. Alternate model specifications recover very similar parameter values, leading us to trust that this model is a good representation of the data generation process that led to the data in our dataset. The main model is proposed for use in estimating temperature values from paleoclimate records.

7 bibliography

references

[1] Paola Arias et al. "Climate Change 2021: The Physical Science Basis. Contribution of Working Group14 I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Technical Summary". In: (2021).

[2] Bryan E Bemis et al. "Reevaluation of the oxygen isotopic composition of planktonic foraminifera: Experimental results and revised paleotemperature equations". In: *Paleoceanography* 13.2 (1998), pp. 150–160.

[3] Yael Bouvier-Soumagnac and Jean-Claude Duplessy. "Carbon and oxygen isotopic composition of planktonic foraminifera from laboratory culture, plankton tows and recent sediment; implications for the reconstruction of paleoclimatic conditions and of the global carbon cycle". In: *The Journal of Foraminiferal Research* 15.4 (1985), pp. 302–320.

[4] Willi A Brand et al. "Assessment of international reference materials for isotope-ratio analysis (IUPAC Technical Report)". In: *Pure and Applied Chemistry* 86.3 (2014), pp. 425–467.

[5] Jean-Claude Duplessy, Laurent Labeyrie, and Claire Waelbroeck. "Constraints on the ocean oxygen isotopic enrichment between the Last Glacial Maximum and the Holocene: Paleoceanographic implications". In: *Quaternary Science Reviews* 21.1-3 (2002), pp. 315–330.

[6] Samuel Epstein et al. "Revised carbonate-water isotopic temperature scale". In: *Geological Society of America Bulletin* 64.11 (1953), pp. 1315–1326.

[7] Andrew Gelman et al. *Bayesian data analysis*. CRC press, 2013.

[8] Andrew Gelman et al. "Bayesian workflow". In: *arXiv preprint arXiv:2011.01808* (2020).

[9] Ethan L Grossman and Teh-Lung Ku. "Oxygen and carbon isotope fractionation in biogenic aragonite: temperature effects". In: *Chemical Geology: Isotope Geoscience Section* 59 (1986), pp. 59–74.

[10] JC Herguera, E Jansen, and WH Berger. "Evidence for a bathyal front at 2000-M depth in the glacial Pacific, based on a depth transect on Ontong Java Plateau". In: *Paleoceanography* 7.3 (1992), pp. 273–288.

[11] Matthew D Hoffman, Andrew Gelman, et al. "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." In: *J. Mach. Learn. Res.* 15.1 (2014), pp. 1593–1623.

[12] Sang-Tae Kim and James R O'Neil. "Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates". In: *Geochimica et cosmochimica acta* 61.16 (1997), pp. 3461–3475.

[13] Sang-Tae Kim et al. "Oxygen isotope fractionation between synthetic aragonite and water: Influence of temperature and Mg2+ concentration". In: *Geochimica et Cosmochimica Acta* 71.19 (2007), pp. 4704–4715.

[14] Jean Lynch-Stieglitz, William B Curry, and Niall Slowey. "A geostrophic transport estimate for the Florida Current from the oxygen isotope composition of benthic foraminifera". In: *Paleoceanography* 14.3 (1999), pp. 360–373.

[15] TM Marchitto et al. “Improved oxygen isotope temperature calibrations for cosmopolitan benthic foraminifera”. In: *Geochimica et Cosmochimica Acta* 130 (2014), pp. 1–11.

[16] John Morden McCrea. “On the isotopic chemistry of carbonates and a paleotemperature scale”. In: *The Journal of Chemical Physics* 18.6 (1950), pp. 849–857.

[17] Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, 2018.

[18] Nicholas Metropolis et al. “Equation of state calculations by fast computing machines”. In: *The journal of chemical physics* 21.6 (1953), pp. 1087–1092.

[19] Nicholas Shackleton. “Oxygen isotope analyses and Pleistocene temperatures reassessed”. In: *Nature* 215.5096 (1967), pp. 15–17.

[20] stan development team. *stan user guide*. 2020.

[21] Manish Tiwari, Ashutosh K. Singh, and Devesh K. Sinha. “Chapter 3 - Stable Isotopes: Tools for Understanding Past Climatic Conditions and Their Applications in Chemostratigraphy”. In: *Chemostratigraphy*. Ed. by Mu. Ramkumar. Oxford: Elsevier, 2015, pp. 65–92. ISBN: 978-0-12-419968-2. DOI: https://doi.org/10.1016/B978-0-12-419968-2.00003-0. URL: https://www.sciencedirect.com/topics/earth-and-planetary-sciences/isotopic-fractionation.

[22] Harold C Urey. “Oxygen isotopes in nature and in the laboratory”. In: *Science* 108.2810 (1948), pp. 489–496.

[23] Aki Vehtari, Andrew Gelman, and Jonah Gabry. “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC”. In: *Statistics and Computing* 27.5 (Aug. 2016), pp. 1413–1432. DOI: 10.1007/s11222-016-9696-4. URL: https://doi.org/10.1007%2Fs11222-016-9696-4.

[24] Thomas Westerhold et al. “An astronomically dated record of Earth’s climate and its predictability over the last 66 million years”. In: *Science* 369.6509 (2020), pp. 1383–1387.

8 Appendix

8.1 species names

number	species name
1	Cibicides pachyderma
2	Cibicidoides wuellerstorfi
3	Hoeglundina elegans
4	Planulina ariminensis
5	Planulina foveolata
6	Uvigerina curticosta
7	Uvigerina flintii
8	Uvigerina peregrina

Table 3: species names in the foraminifera data.

number	species name
1	Acteocina harpa
2	Alvania acuticostata
3	Benthonellania precipitata
4	Cadulus
5	Caecum crebricinctum
6	Cibicides pachyderma
7	Cibicidoides wuellerstorfi
8	Dentalium
9	Eratoidea hematita
10	Hoeglundina elegans
11	Melanella Bowdich
12	Melanella polita
13	Planulina ariminensis
14	Planulina foveolata
15	Seguenzia
16	Turbonilla
17	turridae
18	Uvigerina curticosta
19	Uvigerina flintii
20	Uvigerina peregrina

Table 4: species names in biological data.

number	species name
1	Calcite
2	Calcite-Vaterite Mixtures
3	Aragonite
4	Octavite
5	Vaterite
6	Witherite
7	unknown

Table 6: composition names in the data with combination of foraminifera data and non biological data.

number	species name
1	Hoeglundina elegans
2	Uvigerina flintii
3	Uvigerina peregrina
4	Uvigerina curticosta
5	Cibicidoides wuellerstorfi
6	Cibicides pachyderma
7	Planulina ariminensis
8	Planulina foveolata
9	Calcite
10	Witherite
11	Octavite
12	Calcite-Vaterite Mixtures
13	Vaterite

Table 5: species names in models run with a combination of foraminifera data and non biological data.

8.2 model run times

model	model run time (s)
fully pooled linear regression (foraminifera data)	2.11
fully pooled linear regression with quadratic component (foraminifera data)	41.68
main model: hierarchical linear model (foraminifera data)	68.82
main model: hierarchical linear model: (all biological data)	112.26
hierarchical model using a quadratic component (foraminifera data)	722
hierarchical model with separate oxygen ratio parameters (foraminifera data)	326
hierarchical model with deming regression (foraminifera data)	816
hierarchical model with a correlation matrix (foraminifera data)	133
partial pooling model with a third level of composition (foraminifera data)	950
partial pooling model with a third level of composition (biological data)	463
partial pooling model with a third level of composition: (foraminifera and lab data)	262

Table 7: run times of all models in seconds. These times are from the last run of all the models. They are representative, though no rigorous testing has been performed as it is not the focus of this thesis. Run time may vary based on hardware.

8.3 fully pooled linear regression (foraminifera data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
a	1.0	1054	145.0	4.1	0.1	136.7	153.3
b	1.0	1054	-4.2	0.1	0.0	-4.4	-3.9
sigma	1.0	1437	2.8	0.1	0.0	2.6	3.1
log-posterior	1.0	1117	-461.9	1.2	0.0	-465.0	-460.4

Table 8: fully pooled linear model's summary statistics.

model run time: 2.11

Computed from 4000 by 295 log-likelihood matrix

```
         Estimate   SE
elpd_loo   -730.6  7.0
p_loo         2.2  0.1
looic      1461.3 14.0
------
Monte Carlo SE of elpd_loo is 0.0.

All Pareto k estimates are good (k < 0.5).
See help('pareto-k-diagnostic') for details.
```

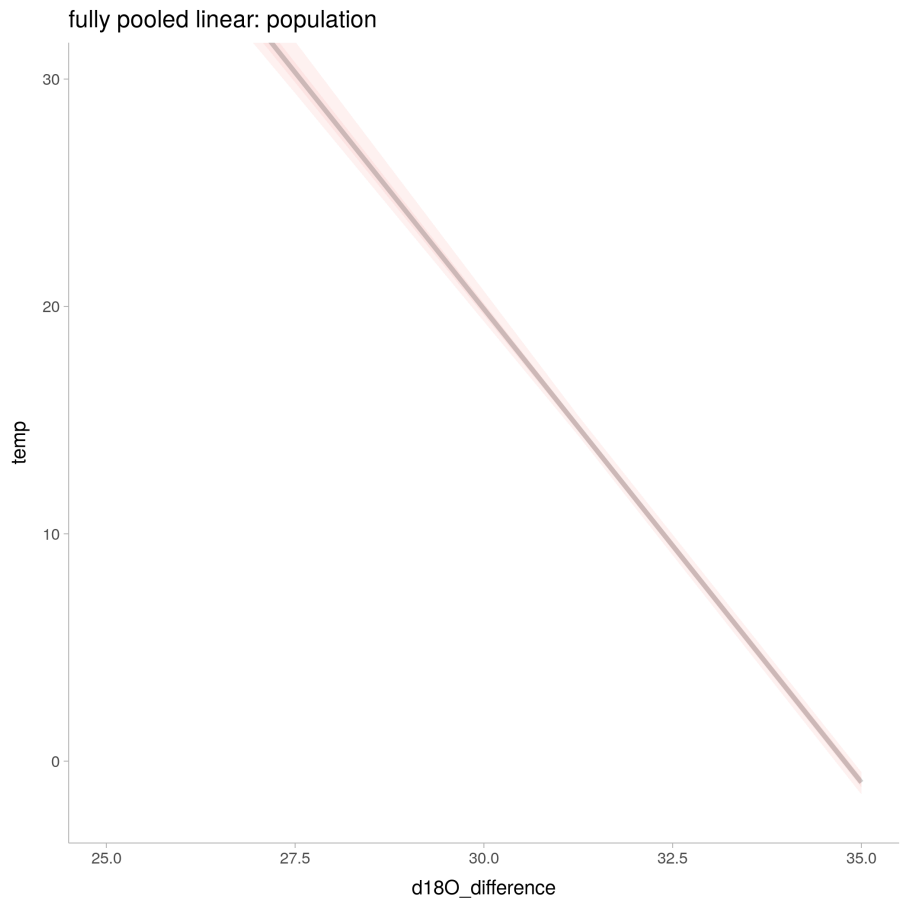

Figure 25: linear regression of the population. Remember: the uncertainty is underreported here.

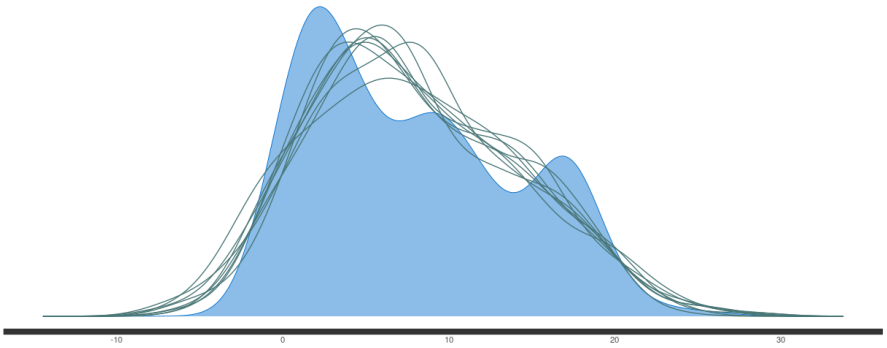

Figure 26: posterior predictive check of the fully pooled linear model.

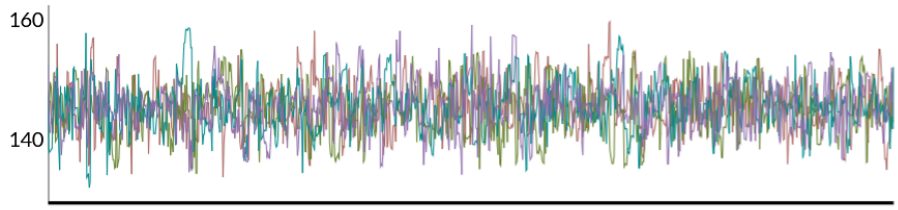

Figure 27: traceplot of the fully pooled linear model.

8.4 fully pooled linear regression with quadratic component (foraminifera data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
a	1.01	826	156.55	26.17	0.91	108.88	210.57
b	1.01	820	-4.89	1.61	0.06	-8.24	-1.95
c	1.01	820	0.01	0.02	0.00	-0.03	0.06
sigma	1.00	1366	2.84	0.11	0.00	2.63	3.07
log-posterior	1.00	983	-462.23	1.40	0.04	-465.74	-460.49

Table 9: fully pooled quad model.

model run time: 41.68

```
fully pooled linear regression using a quadratic component (foraminifera data):
          Estimate   SE
 elpd_loo   -730.2  7.0
 p_loo         2.2  0.1
 looic      1460.5 13.9

fully pooled linear regression (foraminifera data):
          Estimate   SE
 elpd_loo   -730.6  7.0
 p_loo         2.2  0.1
 looic      1461.3 14.0
```

```stan
data {
  int N_obs;
  int K;
  vector[N_obs] temperature;
  matrix[N_obs, K] x;
  int prior_only;
}

parameters {
  real a;
  real b;
  real<lower = 0> sigma;
}

model {
  vector[N_obs] mu;

  a     ~ normal(120,50);      // prior for the intercept
  b     ~ normal(-4,1);        // prior for linear term coefficient
  sigma ~ normal(0,2);         // prior for the standard deviation

  if (! prior_only) {
    mu  = a + rep_vector(b, N_obs) .* (x[,1] - x[,2]); // linear model
    temperature   ~ normal(mu, sigma);   // likelihood
  }

}


generated quantities {

  vector[N_obs] log_lik;
  vector[N_obs] temperature_sim;
  vector[N_obs] mu;

  mu = a + rep_vector(b, N_obs) .* (x[,1] - x[,2]);
  for (i in 1:N_obs){

    log_lik[i] = normal_lpdf(temperature[i] | mu[i],sigma);
    temperature_sim[i] = normal_rng(mu[i],sigma);
  }
}
```

Listing 7: The stan code for the model fully pooled linear regression (foraminifera data) .

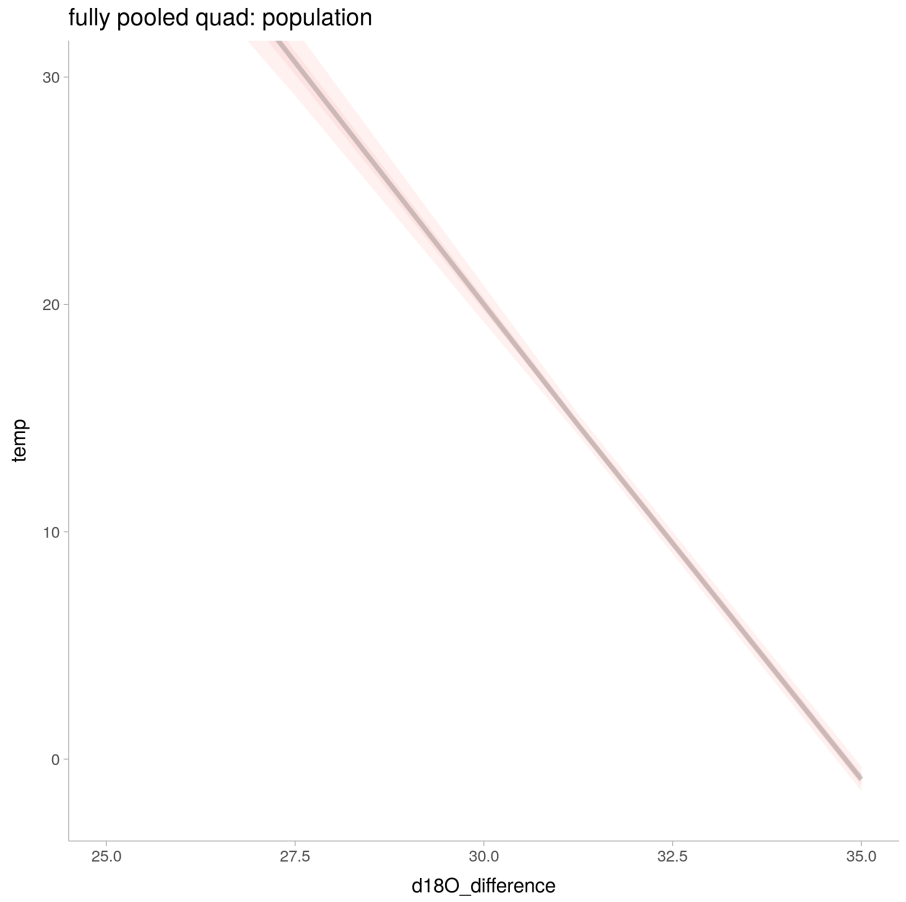

Figure 28: linear regression of the population in the fully pooled quadratic model.

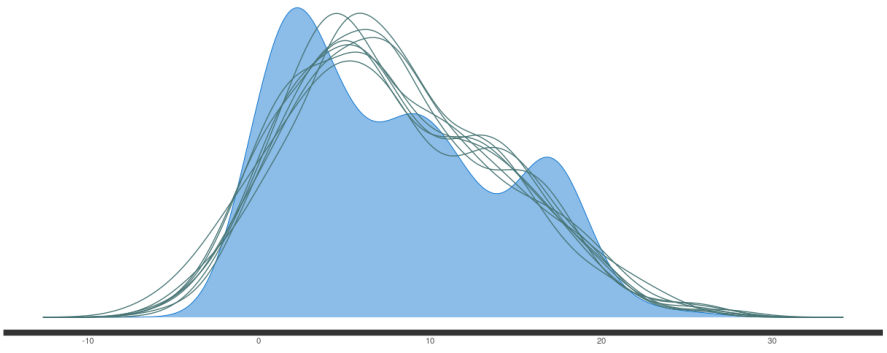

Figure 29: posterior predictive check of the fully pooled model with quadratic component.

```
data {
  int N_obs;
  int K;
  vector[N_obs] temperature;
  matrix[N_obs, K] x;
  int prior_only;
}

parameters {
  real a;
  real b;
  real c;
  real<lower = 0> sigma;
}

model {
  vector[N_obs] mu;

  a     ~ normal(100,50); // prior for the intercept
  b     ~ normal(-4,2);   // prior for linear term coefficient
  c     ~ normal(0.1,2);  // prior for quadratic term coefficient
  sigma ~ normal(0,1);    // prior for the standard deviation
  if (! prior_only) {
    mu    = a + rep_vector(b, N_obs) .* (x[,1] - x[,2]) + rep_vector(c, N_obs) .* (x[,1] - x[,2])^2; // line
    temperature     ~ normal(mu, sigma);   // likelihood
  }
}


generated quantities {
  vector[N_obs] log_lik;
  vector[N_obs] temperature_sim;
  vector[N_obs] mu;

  mu = a + rep_vector(b, N_obs) .* (x[,1] - x[,2]) + rep_vector(c, N_obs) .* (x[,1] - x[,2])^2;
  for (i in 1:N_obs){
    log_lik[i] = normal_lpdf(temperature[i] | mu[i],sigma);
    temperature_sim[i] = normal_rng(mu[i],sigma);
  }
}
```

Listing 8: The stan code for the model fully pooled linear regression with quadratic component (foraminifera data).

8.5 main model: hierarchical linear model (foraminifera data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
mu_a	1.00	1577	130.34	12.17	0.31	106.01	154.34
mu_b	1.00	1451	-3.79	0.36	0.01	-4.56	-3.07
sigma_a	1.00	1902	31.61	10.20	0.23	17.30	57.03
sigma_b	1.00	1572	0.97	0.35	0.01	0.51	1.88
sigma	1.00	4983	0.76	0.03	0.00	0.70	0.82
log-posterior	1.00	969	-70.13	4.20	0.14	-79.32	-62.86
a[1]	1.00	4776	146.96	5.02	0.07	137.36	156.80
a[2]	1.00	3973	75.39	5.34	0.08	64.95	85.87
a[3]	1.00	3920	142.80	1.64	0.03	139.51	146.00
a[4]	1.00	4238	129.01	15.86	0.24	97.70	160.01
a[5]	1.00	4961	144.99	11.90	0.17	122.43	169.44
a[6]	1.00	2913	148.69	20.40	0.38	109.84	190.99
a[7]	1.00	4585	115.95	13.59	0.20	87.91	142.36
a[8]	1.00	4784	146.21	5.09	0.07	136.47	156.14
b[1]	1.00	4793	-4.28	0.16	0.00	-4.59	-3.98
b[2]	1.00	3975	-2.20	0.16	0.00	-2.51	-1.89
b[3]	1.00	3919	-4.01	0.05	0.00	-4.11	-3.91
b[4]	1.00	4242	-3.73	0.50	0.01	-4.70	-2.75
b[5]	1.00	4970	-4.24	0.39	0.01	-5.03	-3.51
b[6]	1.00	2904	-4.26	0.60	0.01	-5.51	-3.11
b[7]	1.00	4579	-3.24	0.44	0.01	-4.09	-2.33
b[8]	1.00	4777	-4.20	0.16	0.00	-4.50	-3.90

Table 10: base hierarchical linear model.

model run time: 68.82

```
Computed from 4000 by 295 log-likelihood matrix

         Estimate   SE
elpd_loo   -346.7 24.4
p_loo        18.3  4.0
looic       693.4 48.8
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     292   99.0%   807
 (0.5, 0.7]   (ok)         2    0.7%   126
   (0.7, 1]   (bad)        1    0.3%   196
   (1, Inf)   (very bad)   0    0.0%   <NA>
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are too high.
```

```
See help('pareto-k-diagnostic') for details.
```

Figure 30: original loo estimate for model converted to rstan. Note the diagnostic values: there is one data point with a pareto $\hat{k}$ value above the cutoff of 0.7, which means we have a theoretical reason to question the validity of our derived elpd estimate. The loo estimate is almost identical in the model that was fit completely in rstan, but the Monte Carlo SE is not defined here.

```
Computed from 4000 by 295 log-likelihood matrix

         Estimate   SE
elpd_loo   -346.9 24.5
p_loo        18.6  4.1
looic       693.8 48.9
------
Monte Carlo SE of elpd_loo is 0.1.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     292   99.0%   483
 (0.5, 0.7]   (ok)         3    1.0%   139
   (0.7, 1]   (bad)        0    0.0%   <NA>
   (1, Inf)   (very bad)   0    0.0%   <NA>

All Pareto k estimates are ok (k < 0.7).
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are slightly high.
 See help('pareto-k-diagnostic') for details.
```

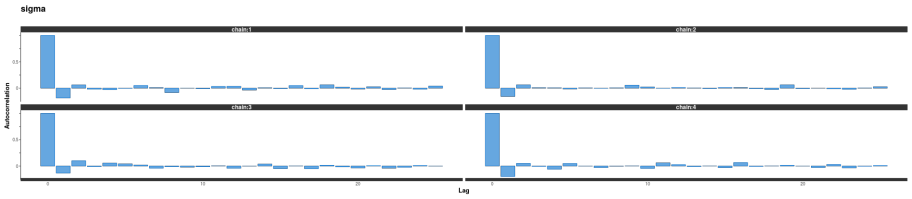

Figure 31: autocorrelation plot of the main model.

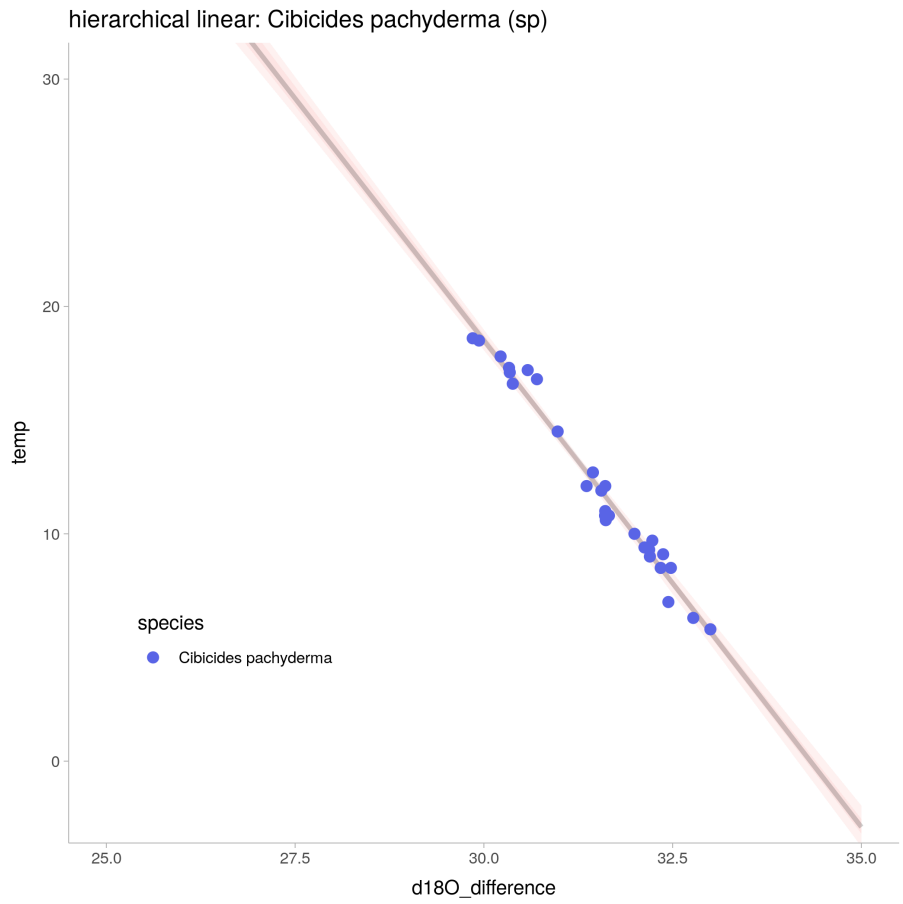

Figure 32: linear regression line of the main model: Cibicides pachyderma (sp).

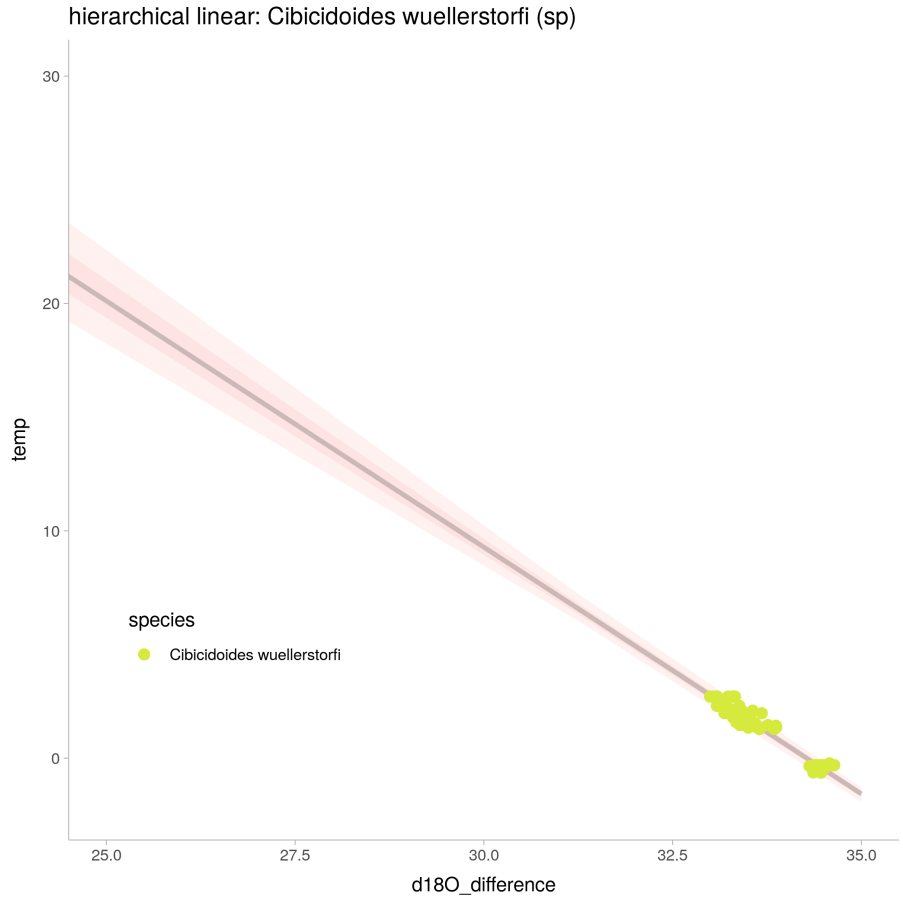

Figure 33: linear regression for main model: Cibicidoides wuellerstorfi (sp).

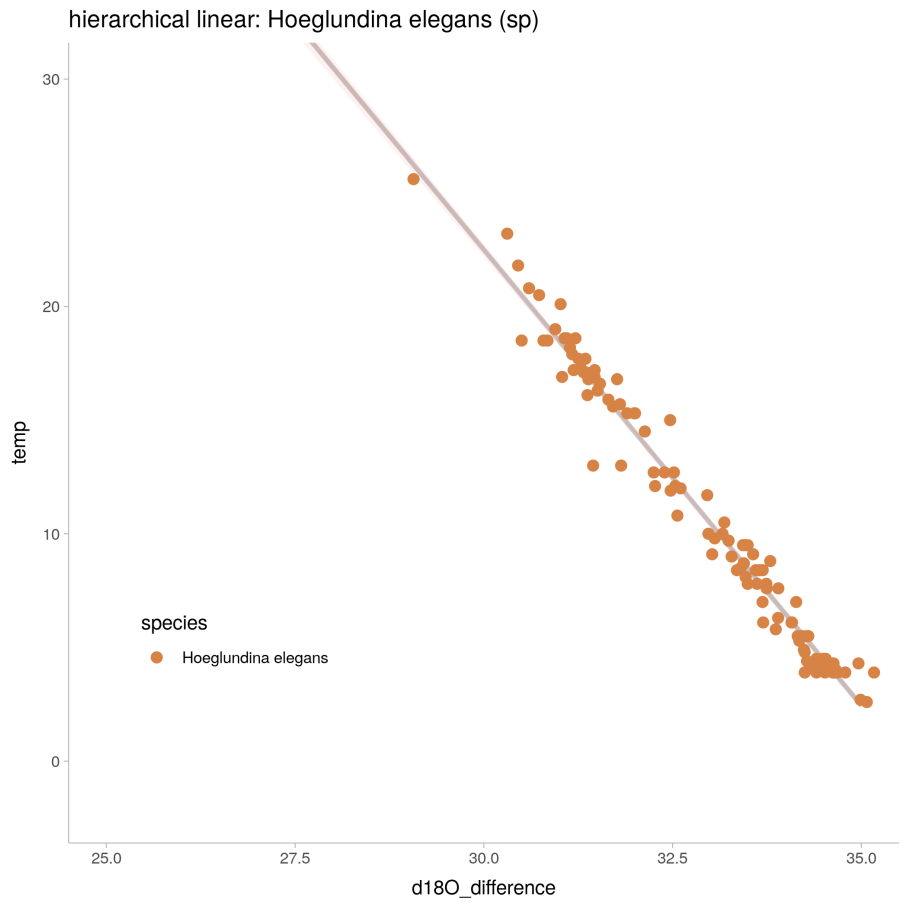

Figure 34: linear regression for main model: for Hoeglundina elegans (sp).

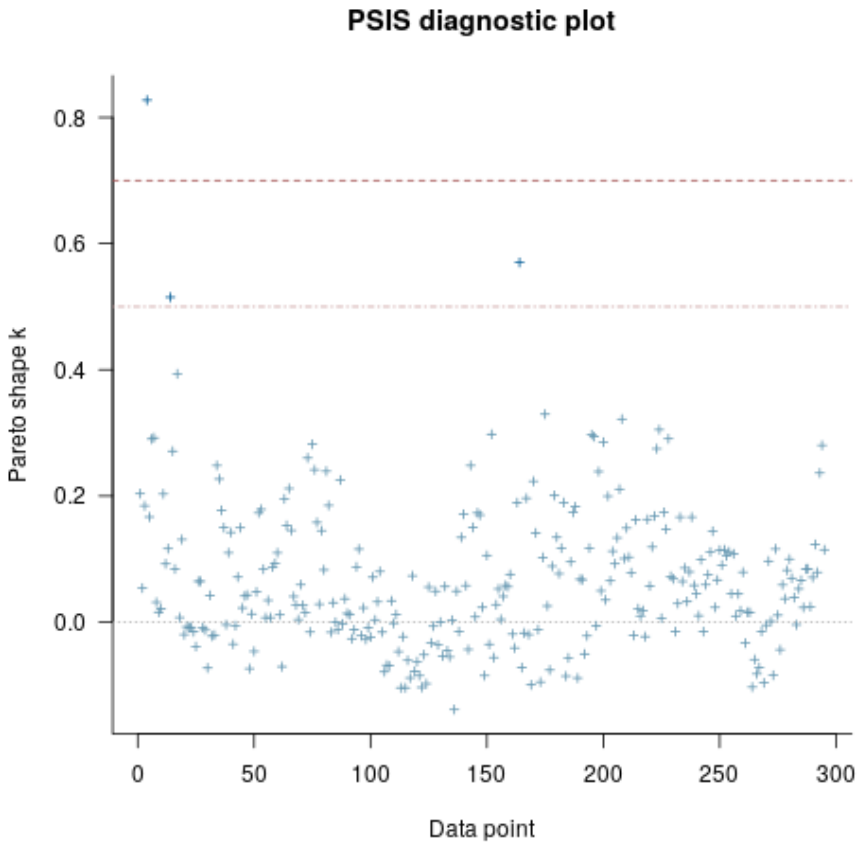

Figure 35: pareto k values of the data for the main model, note the data point above the 0.7 line. These values are scaled down when ran from Rstan, leading to a loo value that is identical but which has a proper sampling se, so we are sure we can trust it.

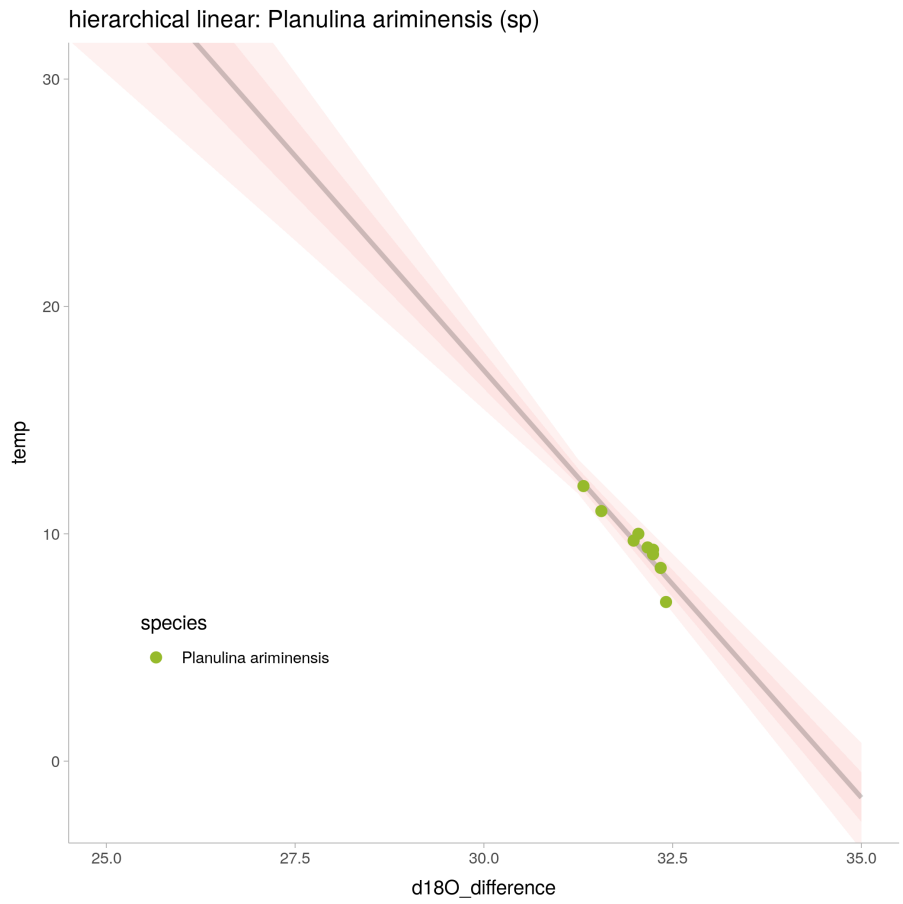

Figure 36: linear regression for main model: for Planulina ariminensis (sp).

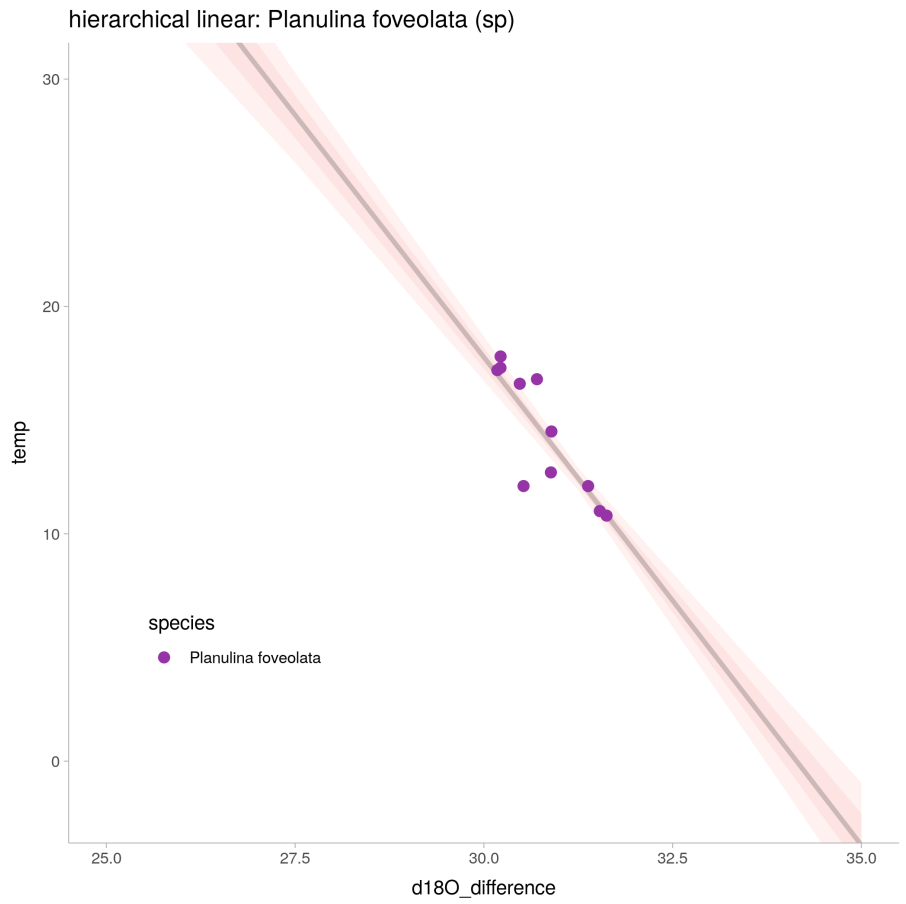

Figure 37: linear regression for main model: for Planulina foveolata (sp).

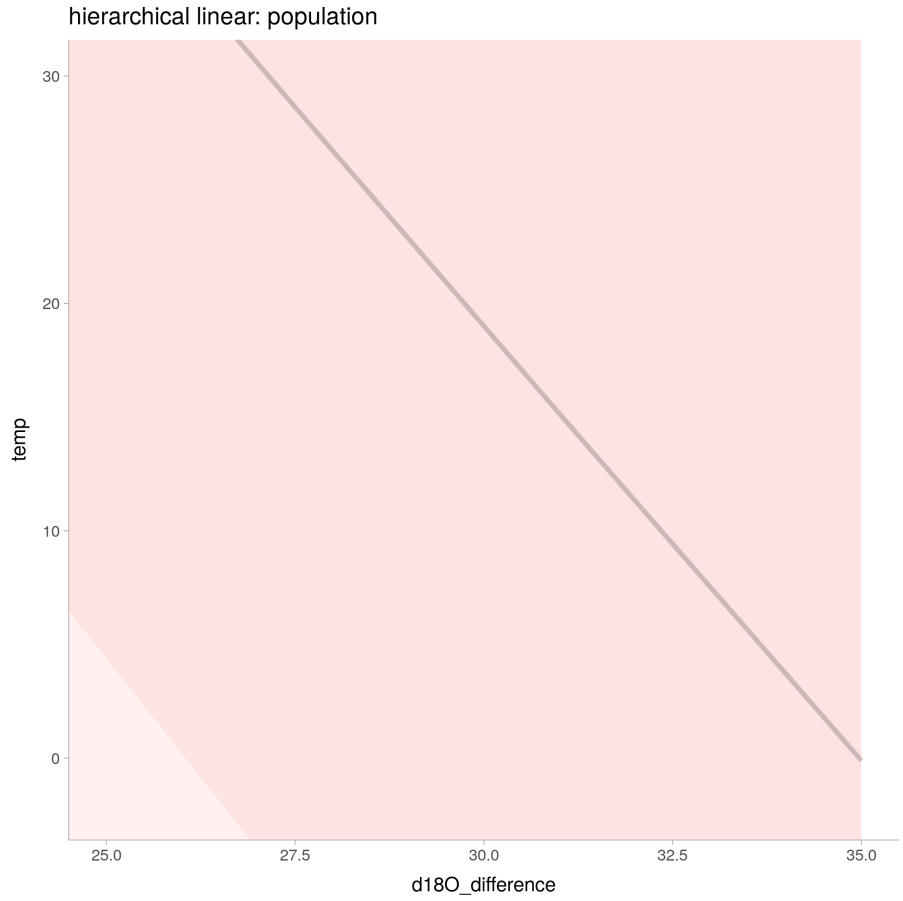

Figure 38: linear regression for main model: for population. Whereas in the fully pooled model the standard deviation is artificially low because stan puts its uncertainty in the residual sigma (which is very high compared to the hierarchical models).

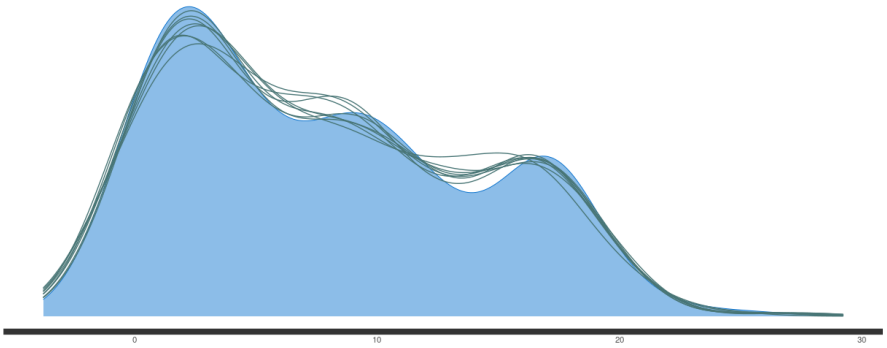

Figure 39: posterior predictive check for the main model.

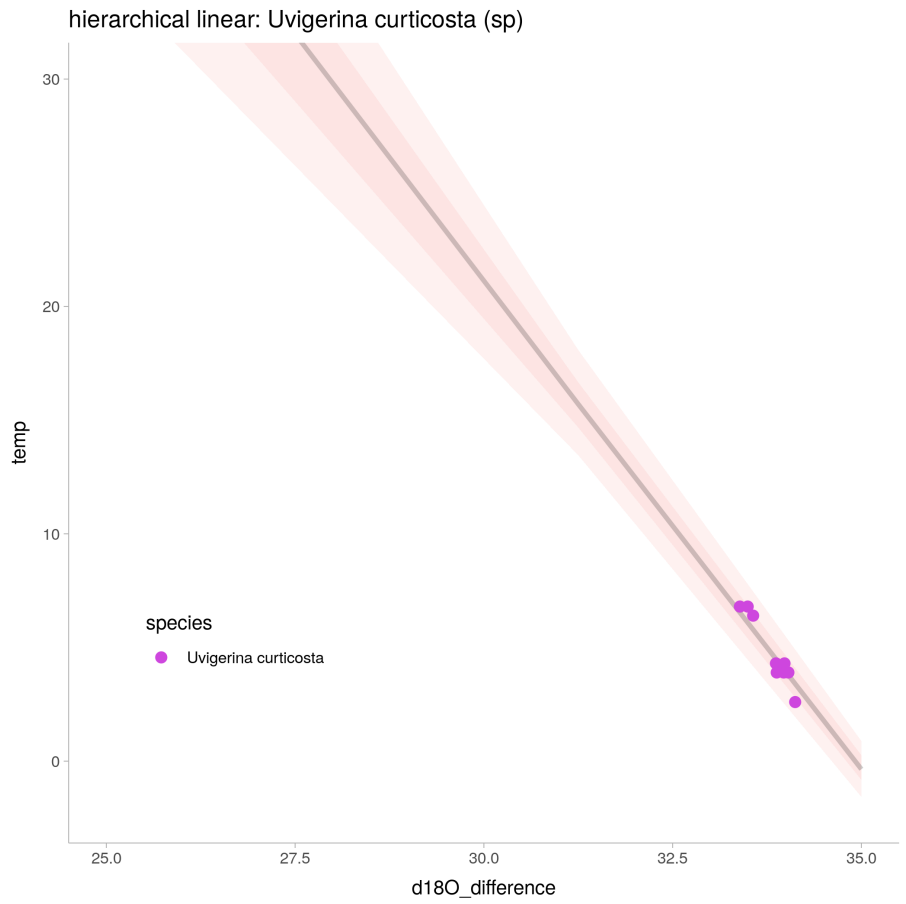

Figure 40: linear regression for main model: for Uvigerina curticosta (sp).

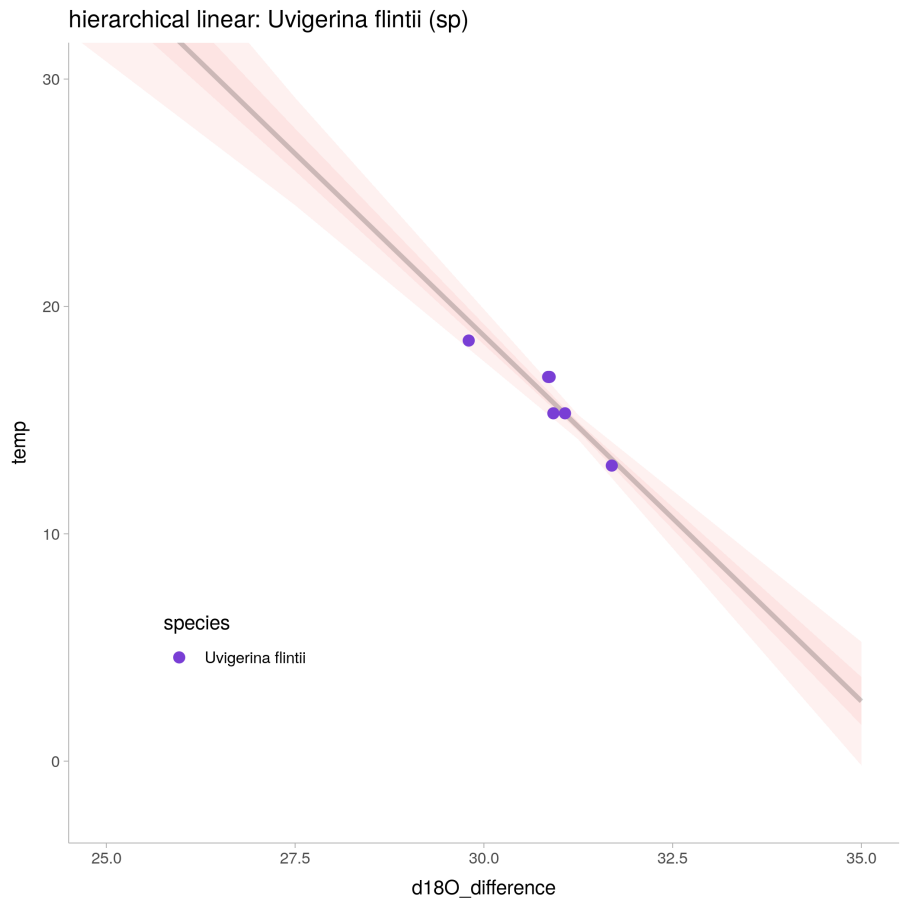

Figure 41: linear regression for main model: for Uvigerina flintii (sp).

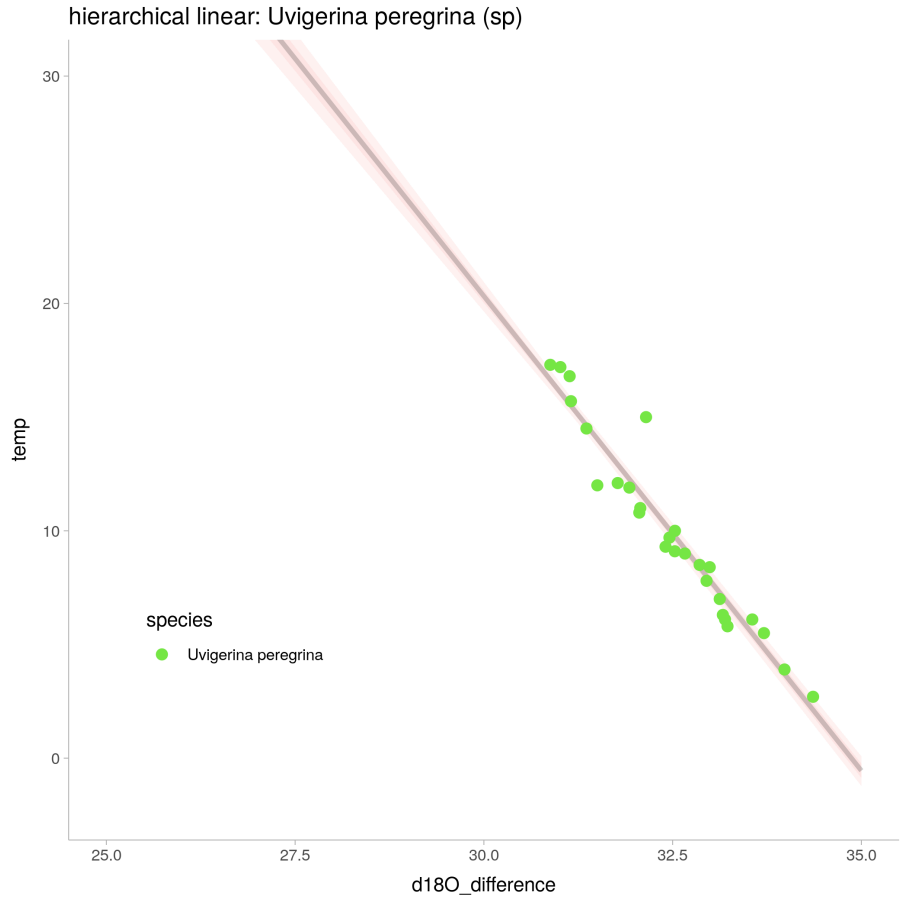

Figure 42: linear regression for main model: for Uvigerina peregrina (sp).

```
blah
data {
  int N_obs;           /* nr of observations */
  int N_sp;           /* nr of groups */
  int K;           /* nr of features */
  vector[N_obs] temperature;
  matrix[N_obs, K] x;

  array[N_obs] int<lower=0,upper=1> is_lab;

  array[N_obs] int<lower=1, upper=N_sp> species;

  int<lower=0, upper=1> prior_only;

}

parameters {
  real<lower = 0> sigma;

  real mu_a;
  real mu_b;

  real<lower = 0> sigma_a;
  real<lower = 0> sigma_b;

  vector<offset = mu_a, multiplier = sigma_a>[N_sp] a;
  vector<offset = mu_b, multiplier = sigma_b>[N_sp] b;
}

model {
  vector[N_obs] mu;

  mu_a ~ normal(120, 50);
  mu_b ~ normal(-4, 1);

  sigma_a ~ normal(0, 50);
  sigma_b ~ normal(0, 5);

  sigma ~ normal(0, 2);

  a ~ normal(mu_a, sigma_a);
  b ~ normal(mu_b, sigma_b);

  if (!prior_only) {
    mu = a[species] + b[species] .* (x[,1] - x[,2]);
    temperature ~ normal(mu, sigma);
  }
}
```

```
generated quantities {
  vector[N_obs] log_lik;
  vector[N_obs] temperature_sim;
  vector[N_obs] mu;
  mu = a[species] + b[species] .* (x[,1] - x[,2]);
  for (i in 1:N_obs){
    log_lik[i] = normal_lpdf(temperature[i] | mu[i],sigma);
    temperature_sim[i] = normal_rng(mu[i],sigma);
  }
}
```

Listing 9: The stan code for the model main model: hierarchical linear model (foraminifera data)

8.6 main model: hierarchical linear model: (all biological data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
sigma	1.00	4221	0.82	0.03	0.00	0.76	0.88
mu_a	1.00	2207	133.46	5.89	0.13	121.71	144.90
mu_b	1.00	2598	-3.75	0.18	0.00	-4.10	-3.41
sigma_a	1.00	1294	17.09	4.01	0.11	10.60	26.32
sigma_b	1.00	1406	0.50	0.12	0.00	0.30	0.78
log-posterior	1.00	766	-113.38	6.96	0.25	-128.59	-101.22
a[1]	1.00	4057	133.90	12.83	0.20	107.61	159.40
a[2]	1.00	4682	133.59	12.13	0.18	110.09	158.20
a[3]	1.00	4505	134.24	12.16	0.18	108.59	158.23
a[4]	1.00	5329	139.15	9.23	0.13	120.91	158.17
a[5]	1.00	4782	136.11	12.60	0.18	110.86	160.81
a[6]	1.00	4468	144.66	5.06	0.08	134.88	154.55
a[7]	1.00	1635	84.02	6.76	0.17	71.33	98.01
a[8]	1.00	5228	133.59	13.30	0.18	107.31	159.27
a[9]	1.00	4788	136.31	12.72	0.18	111.66	161.57
a[10]	1.00	3994	142.62	1.77	0.03	139.18	146.06
a[11]	1.00	4671	134.20	12.82	0.19	108.52	159.96
a[12]	1.00	4947	134.93	12.47	0.18	110.16	160.05
a[13]	1.00	5524	130.61	11.15	0.15	109.74	152.73
a[14]	1.00	4551	138.89	9.62	0.14	121.14	158.02
a[15]	1.00	4947	135.55	12.73	0.18	110.23	161.23
a[16]	1.00	5005	138.43	11.15	0.16	116.78	160.96
a[17]	1.00	4487	134.08	12.87	0.19	109.06	160.45
a[18]	1.00	4016	138.19	12.70	0.20	115.06	164.90
a[19]	1.00	4977	125.00	10.69	0.15	102.97	145.42
a[20]	1.00	5025	144.26	5.20	0.07	133.94	154.59
b[1]	1.00	4065	-3.74	0.39	0.01	-4.51	-2.95
b[2]	1.00	4698	-3.71	0.38	0.01	-4.49	-2.96
b[3]	1.00	4526	-3.73	0.39	0.01	-4.50	-2.90
b[4]	1.00	5311	-3.89	0.28	0.00	-4.46	-3.33
b[5]	1.00	4780	-3.77	0.38	0.01	-4.51	-3.01

b[6]	1.00	4470	-4.21	0.16	0.00	-4.52	-3.90
b[7]	1.00	1636	-2.46	0.20	0.00	-2.87	-2.08
b[8]	1.00	5205	-3.74	0.39	0.01	-4.49	-2.98
b[9]	1.00	4800	-3.68	0.40	0.01	-4.48	-2.89
b[10]	1.00	4004	-4.00	0.05	0.00	-4.11	-3.90
b[11]	1.00	4669	-3.72	0.38	0.01	-4.48	-2.96
b[12]	1.00	4991	-3.72	0.39	0.01	-4.49	-2.95
b[13]	1.00	5510	-3.78	0.35	0.00	-4.47	-3.13
b[14]	1.00	4558	-4.04	0.31	0.00	-4.66	-3.46
b[15]	1.00	4962	-3.69	0.41	0.01	-4.51	-2.89
b[16]	1.00	4967	-3.78	0.35	0.01	-4.49	-3.09
b[17]	1.00	4488	-3.72	0.39	0.01	-4.52	-2.96
b[18]	1.00	4006	-3.95	0.38	0.01	-4.73	-3.26
b[19]	1.00	4978	-3.53	0.35	0.00	-4.19	-2.82
b[20]	1.00	5038	-4.14	0.16	0.00	-4.45	-3.82

Table 11: hierarchical linear model with bio data

model run time: 112.26

```
Computed from 4000 by 323 log-likelihood matrix

         Estimate   SE
elpd_loo   -413.2 27.0
p_loo        33.5  7.4
looic       826.4 53.9
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     310   96.0%   418
 (0.5, 0.7]   (ok)         2    0.6%   911
   (0.7, 1]   (bad)       11    3.4%   25
   (1, Inf)   (very bad)   0    0.0%   <NA>
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are too high.
See help('pareto-k-diagnostic') for details.
```

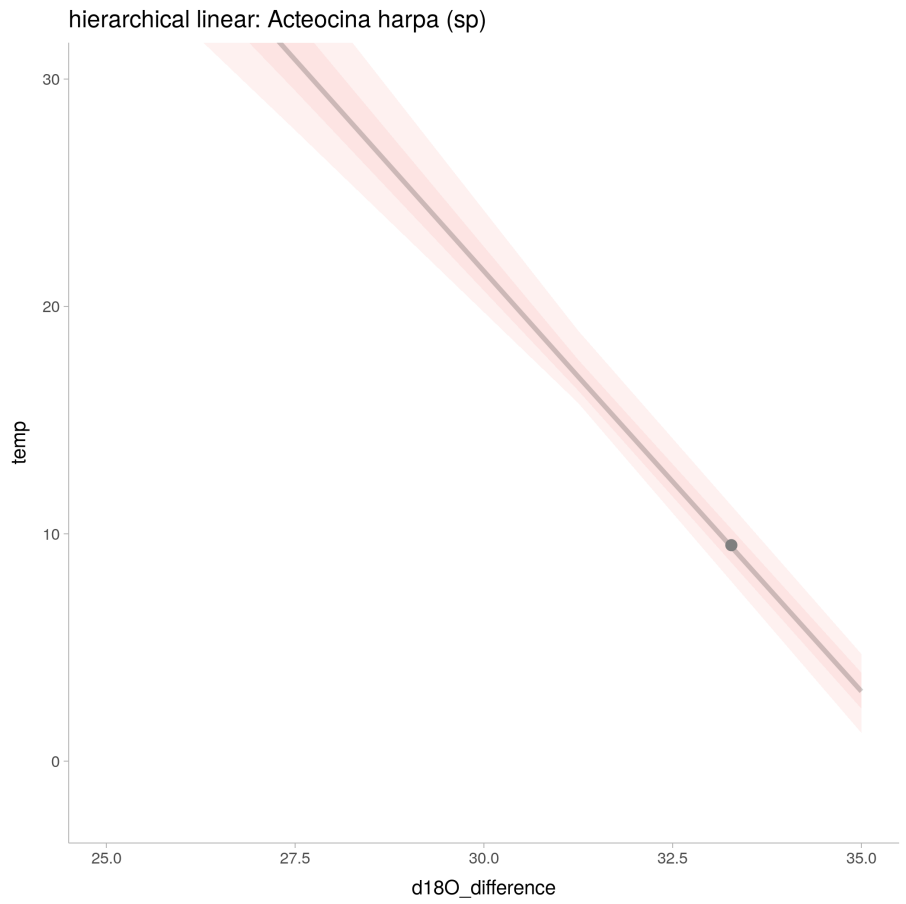

Figure 43: linear regression for main model run with all biological data: Acteocina harpa (sp).

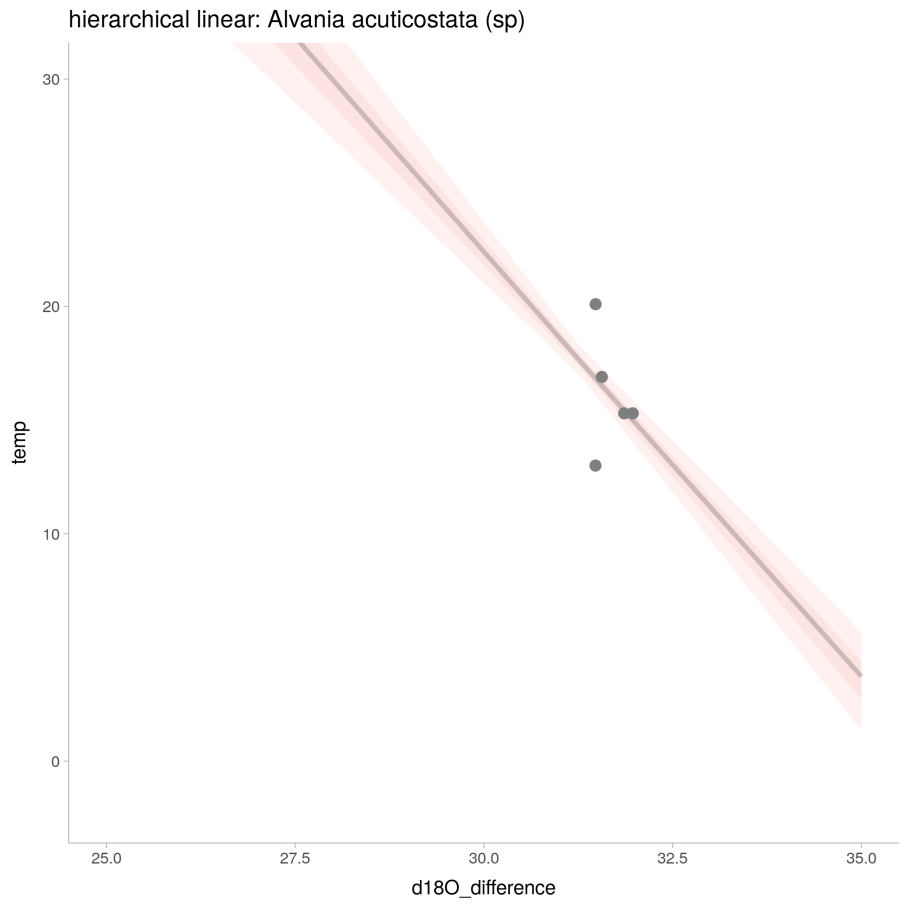

Figure 44: linear regression for main model run with all biological data: Alvania acuticostata (sp).

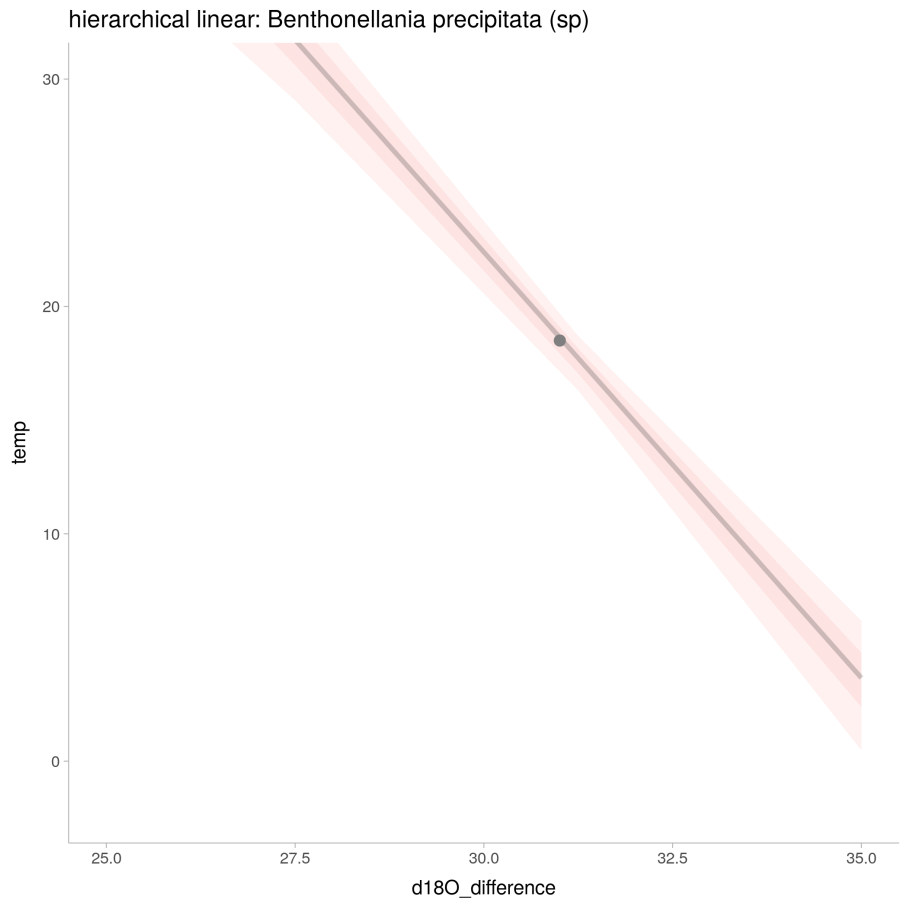

Figure 45: linear regression for main model run with all biological data: Benthonellania precipitata (sp).

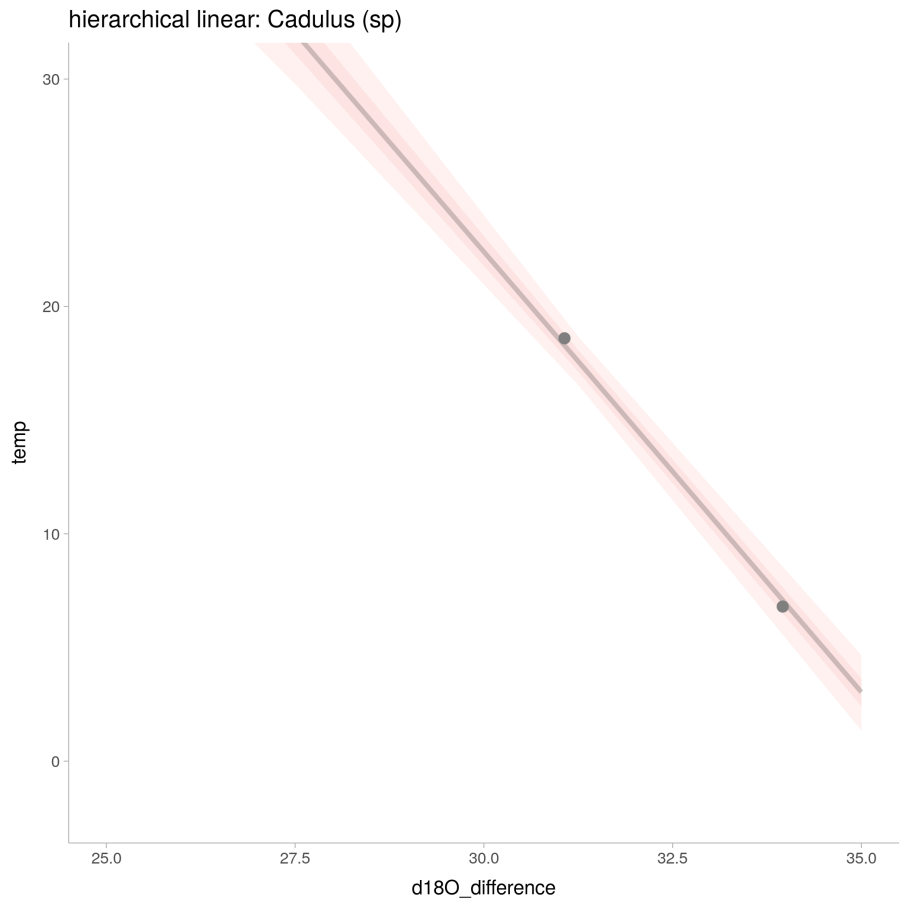

Figure 46: linear regression for main model run with all biological data: Cadulus (sp).

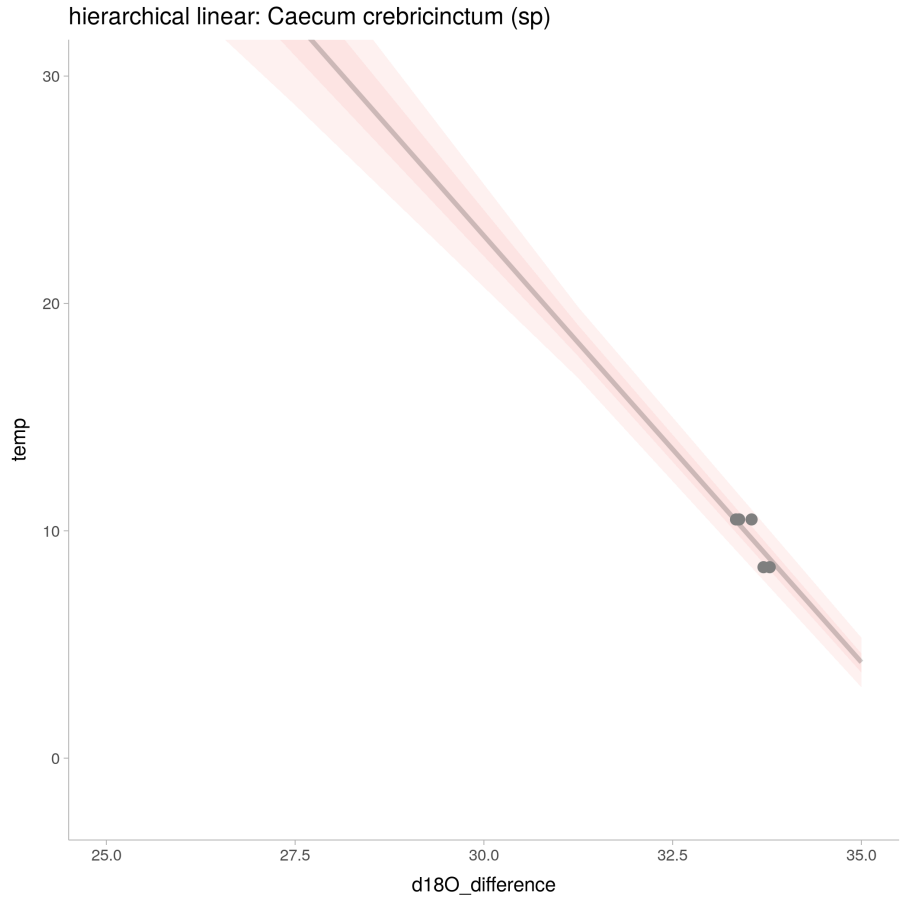

Figure 47: linear regression for main model run with all biological data: Caecum crebricinctum (sp).

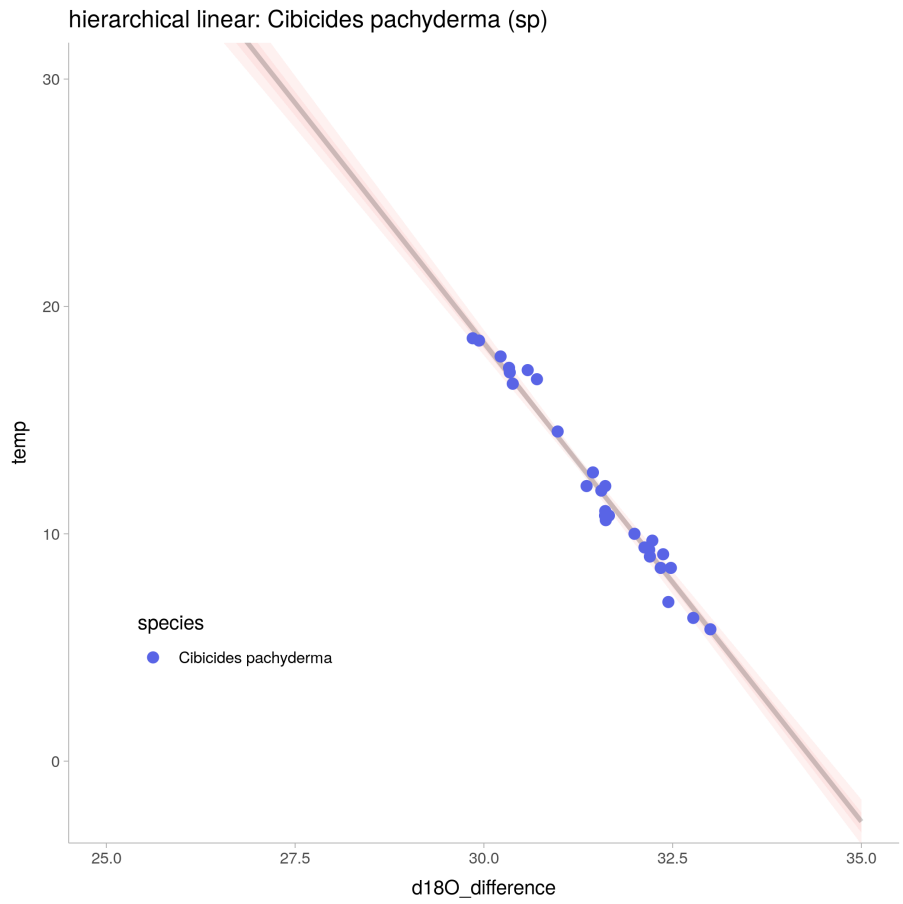

Figure 48: linear regression for main model run with all biological data: Cibicides pachyderma (sp).

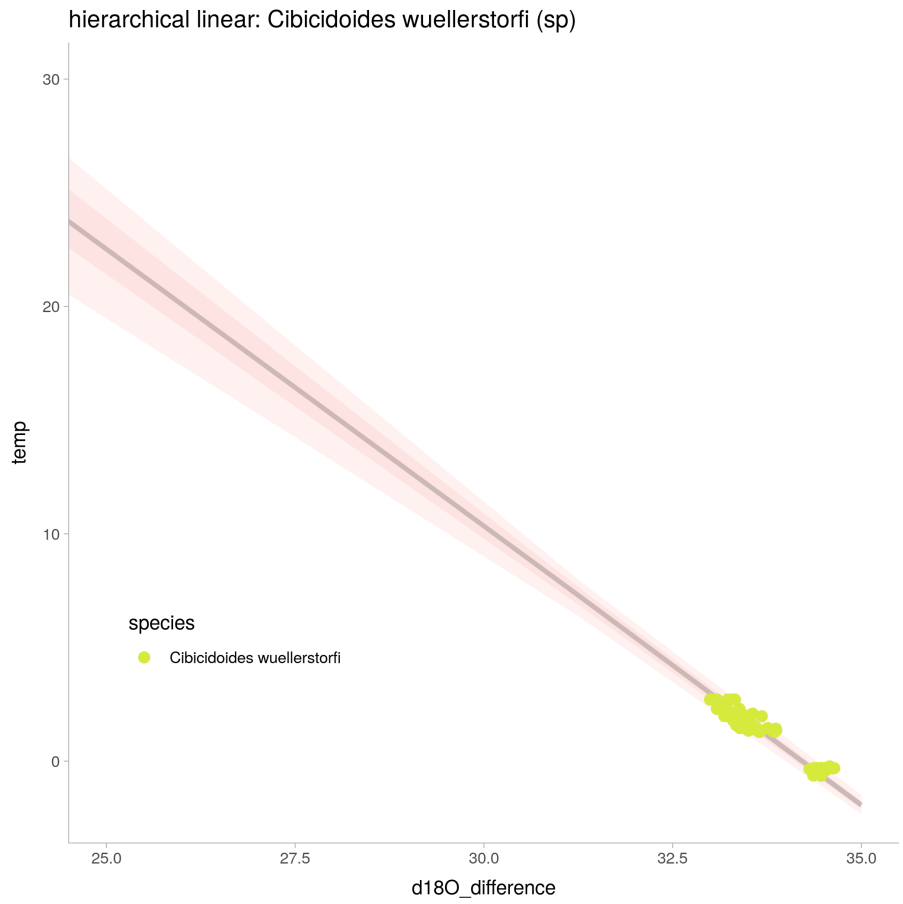

Figure 49: linear regression for main model run with all biological data: Cibicidoides wuellerstorfi (sp).

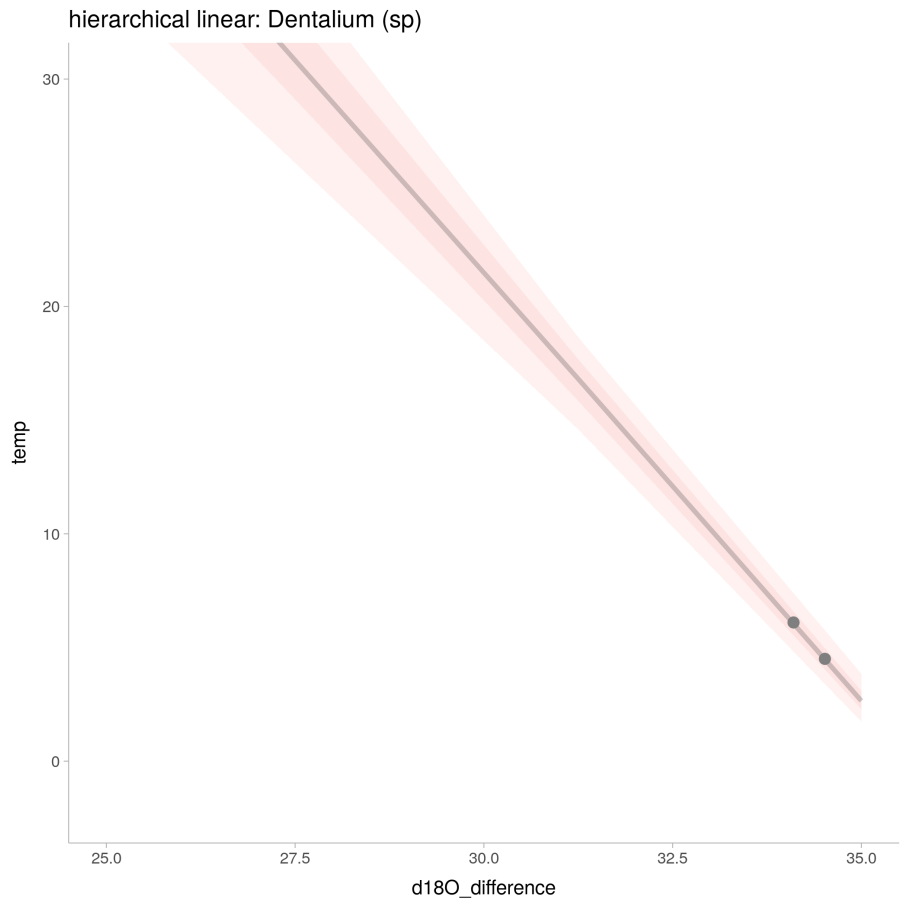

Figure 50: linear regression for main model run with all biological data: Dentalium (sp).

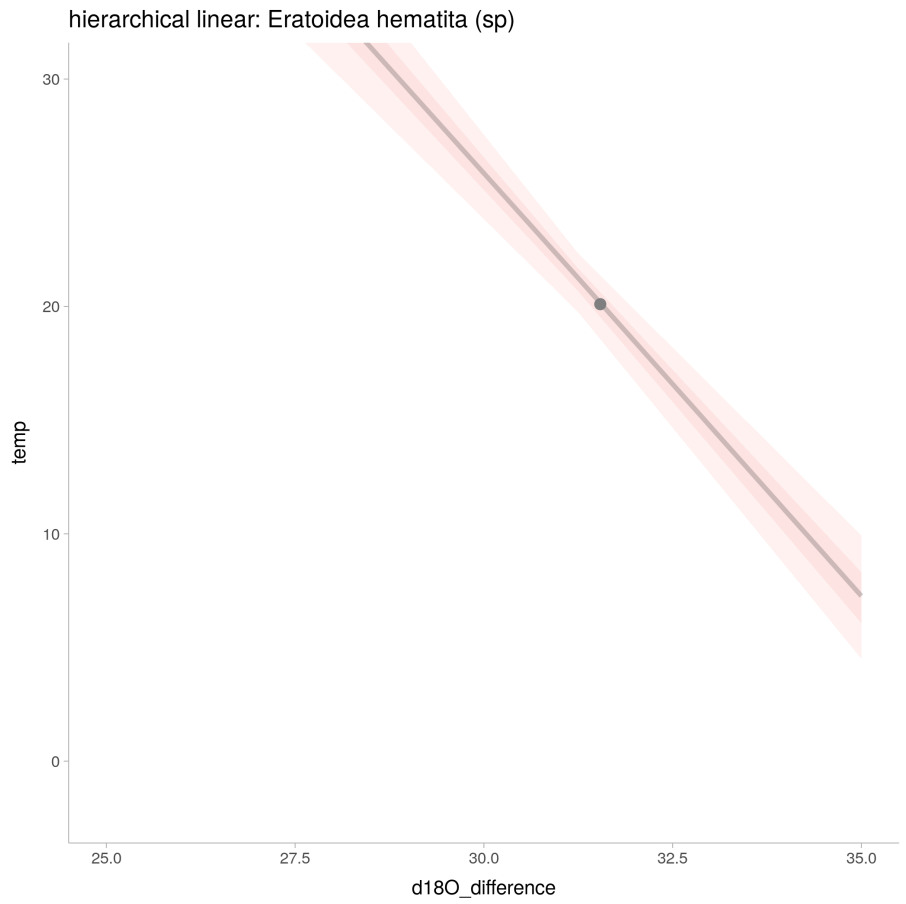

Figure 51: linear regression for main model run with all biological data: Eratoidea hematita (sp).

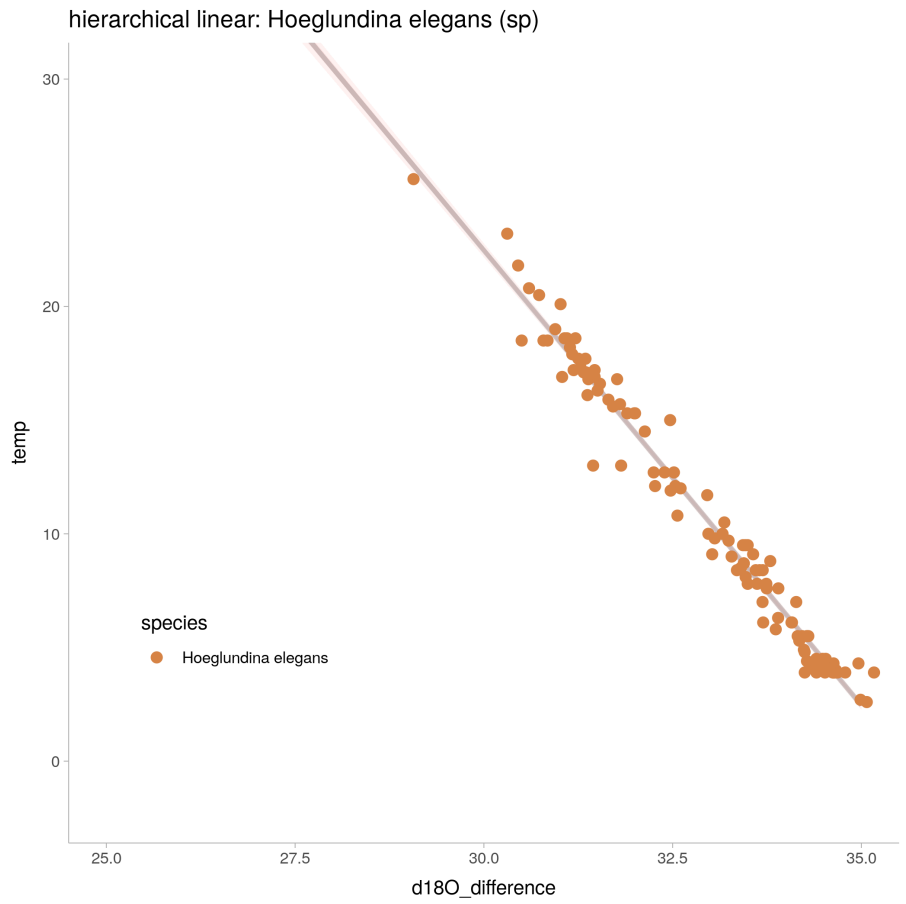

Figure 52: linear regression for main model run with all biological data: Hoeglundina elegans (sp).

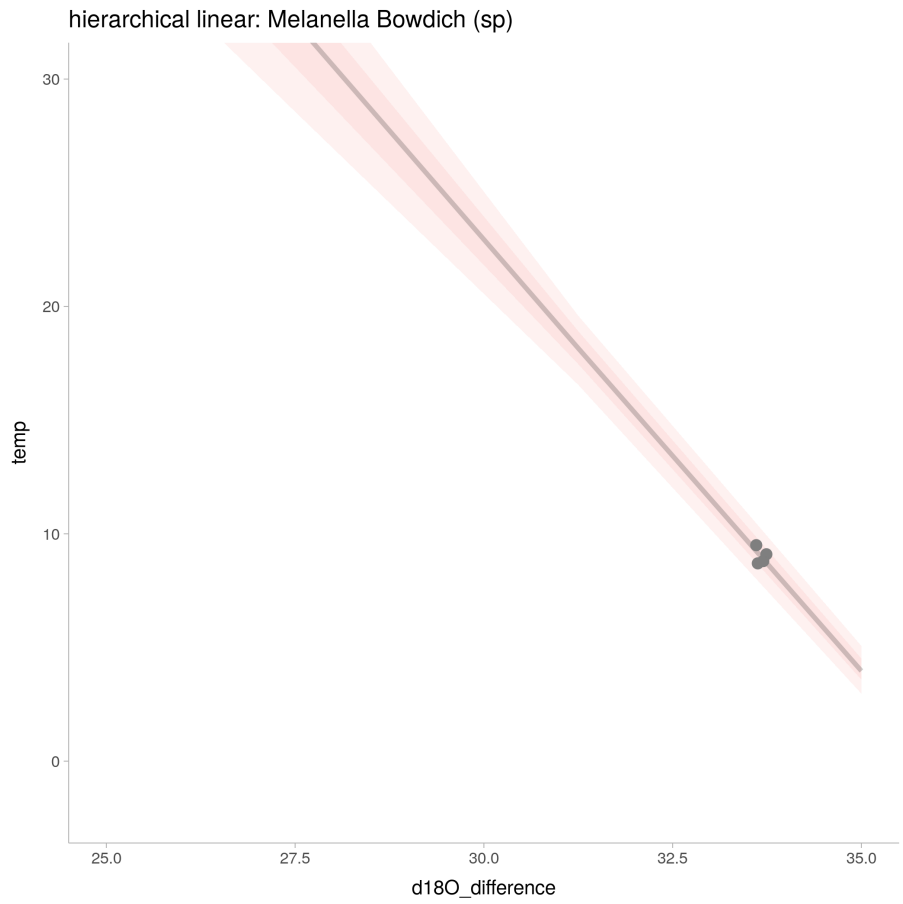

Figure 53: linear regression for main model run with all biological data: Melanella Bowdich (sp).

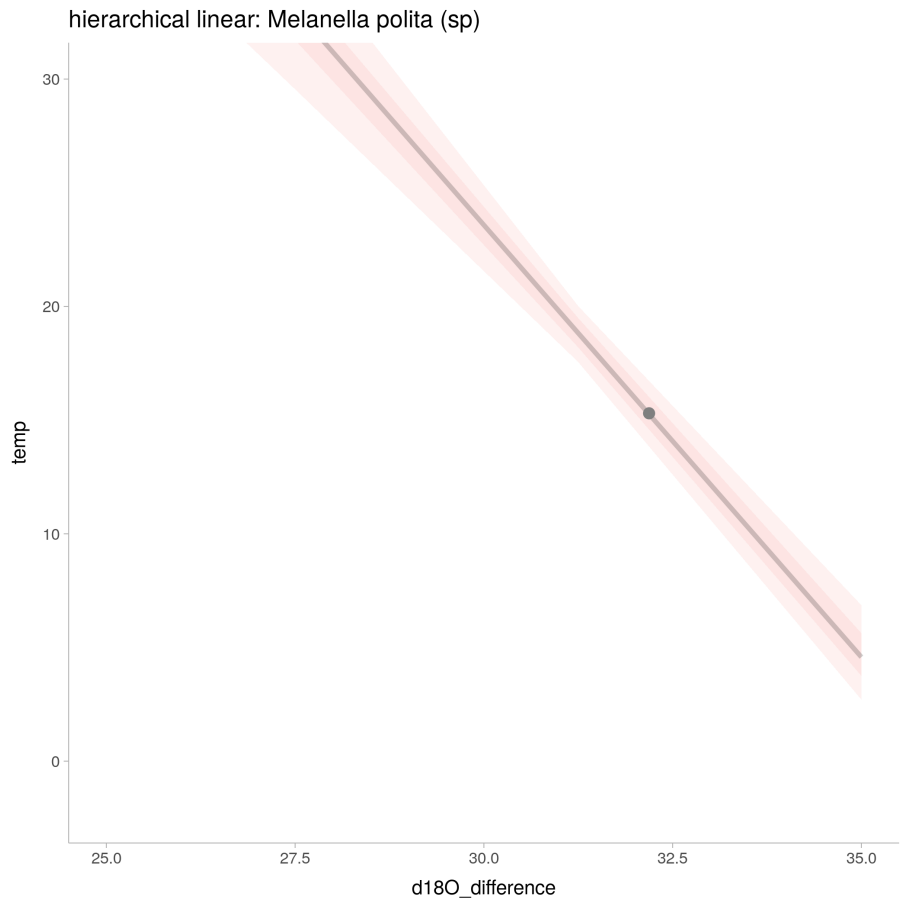

Figure 54: linear regression for main model run with all biological data: Melanella polita (sp).

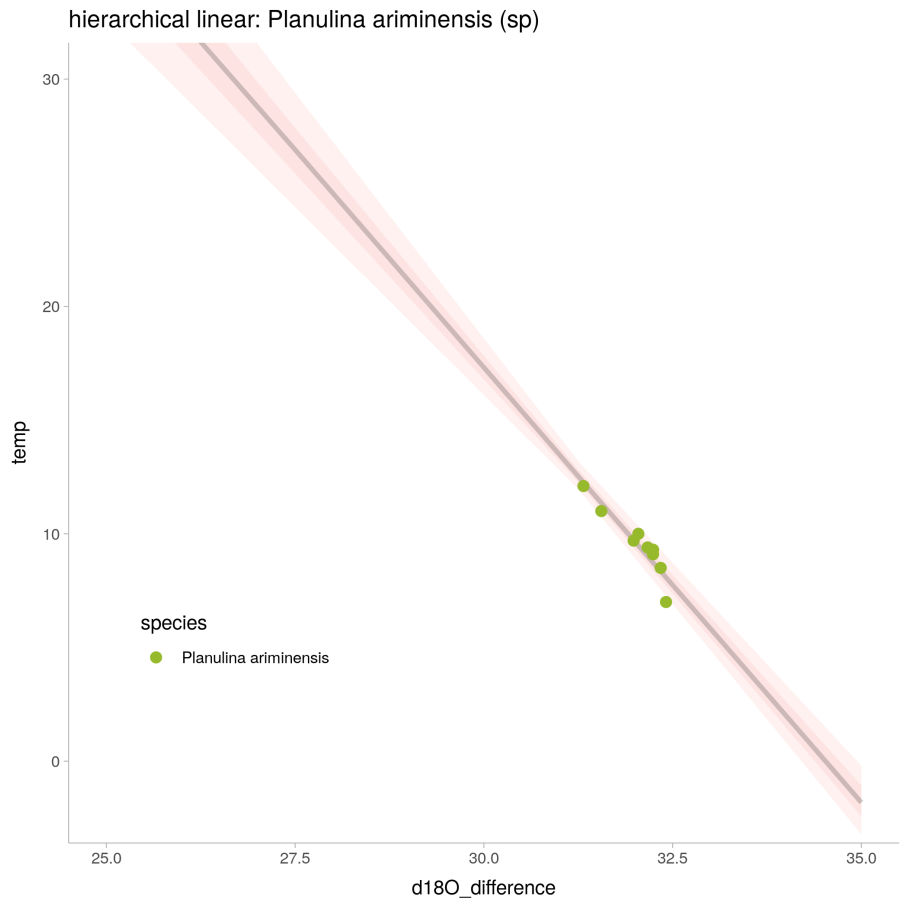

Figure 55: linear regression for main model run with all biological data: Planulina ariminensis (sp).

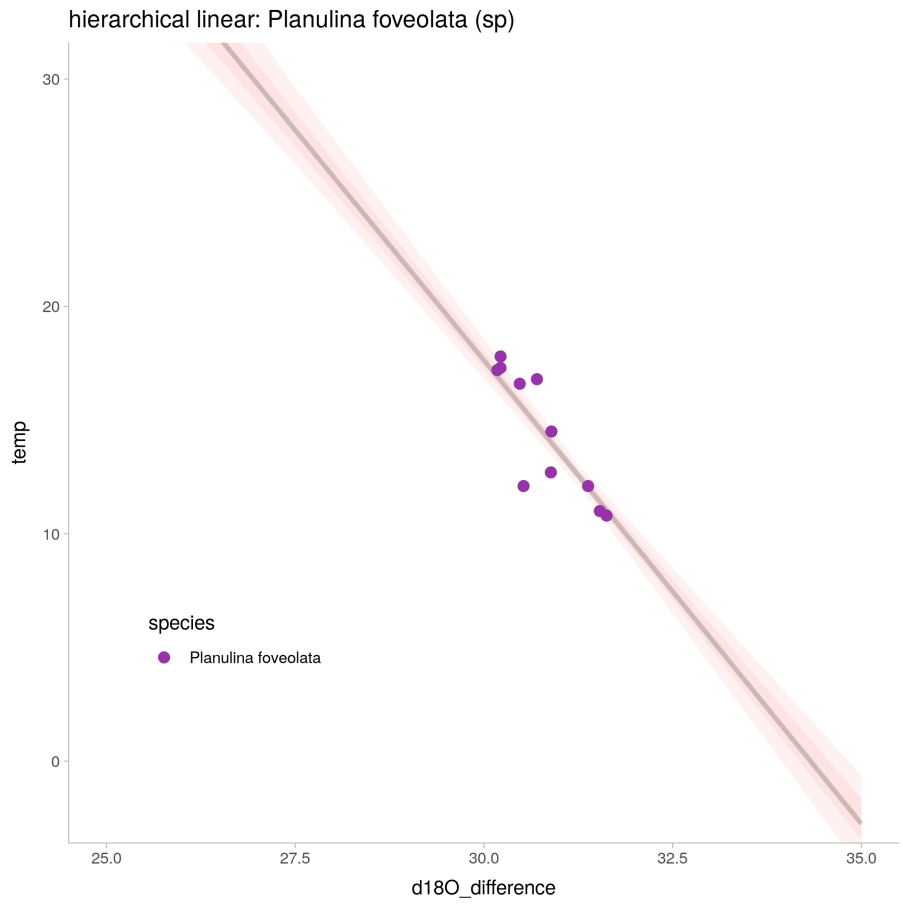

Figure 56: linear regression for main model run with all biological data: Planulina foveolata (sp).

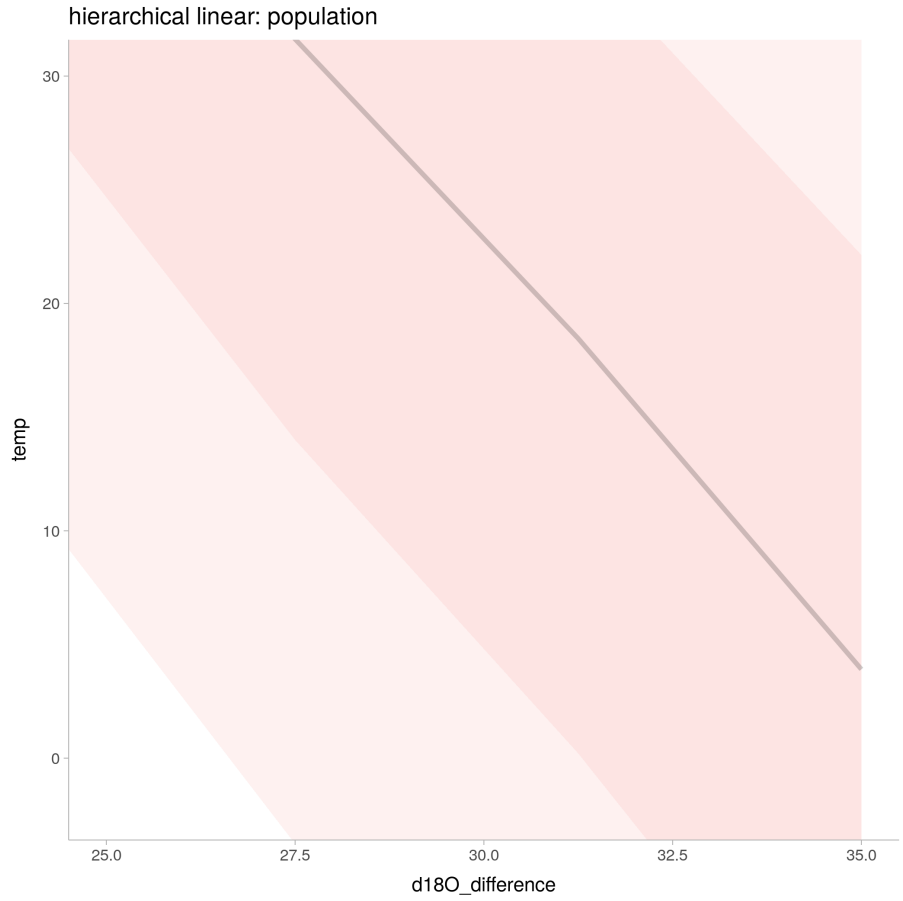

Figure 57: linear regression for main model run with all biological data: population.

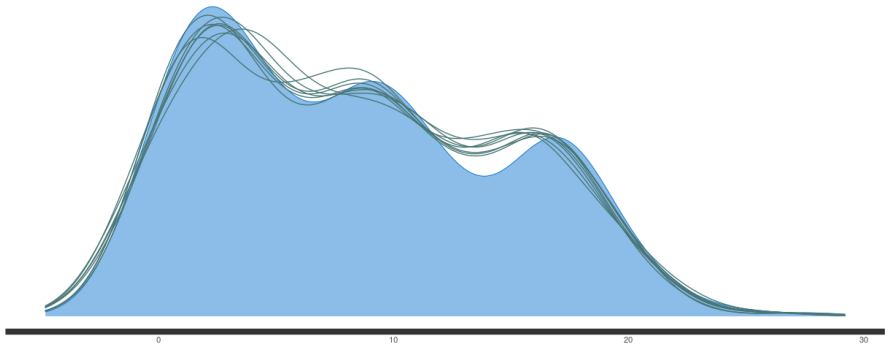

Figure 58: linear regression for main model run with all biological data: ppc

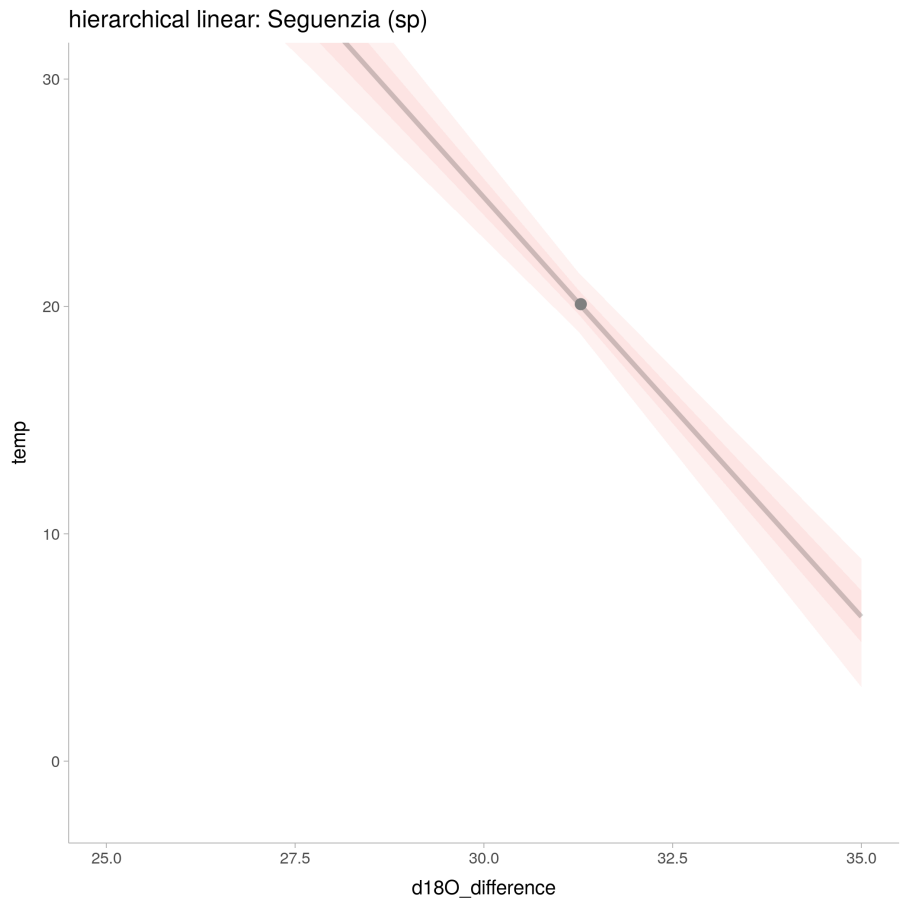

Figure 59: linear regression for main model run with all biological data: Seguenzia (sp).

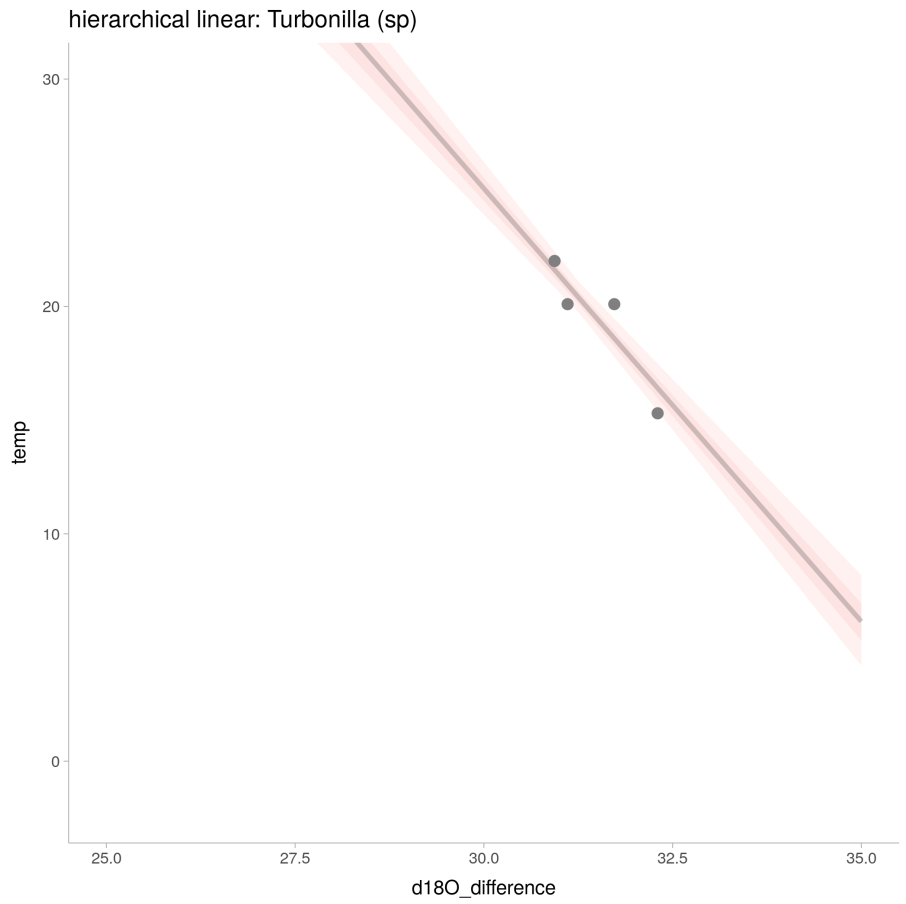

Figure 60: linear regression for main model run with all biological data: Turbonilla (sp).

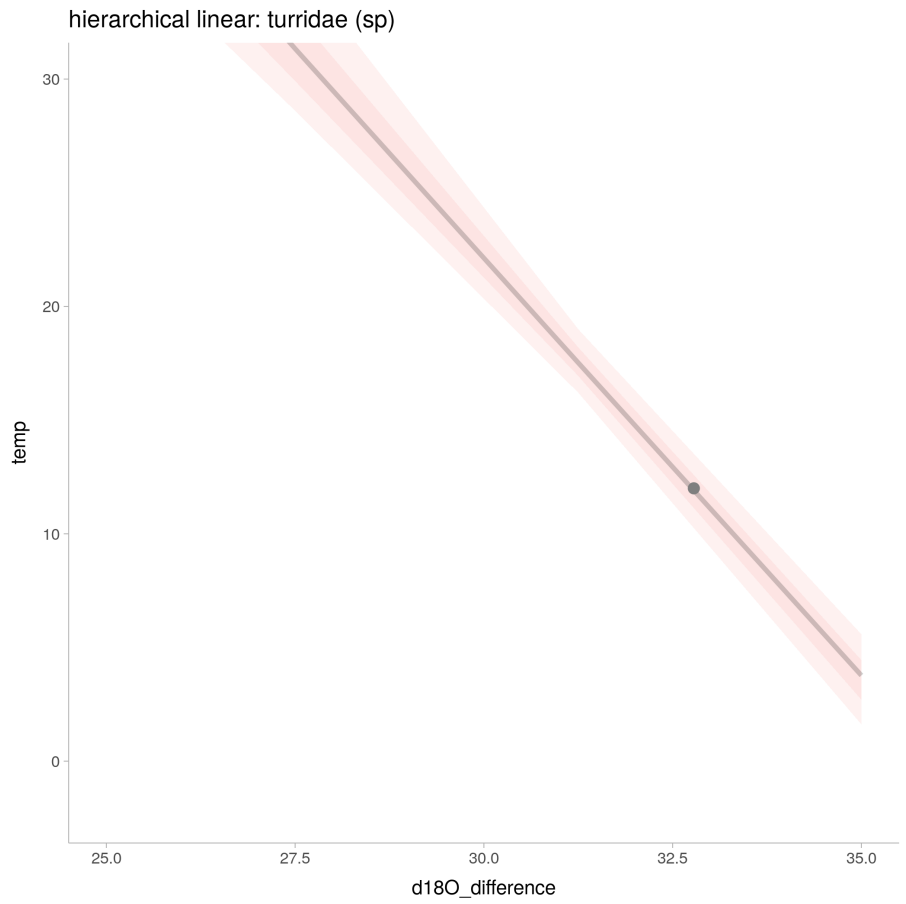

Figure 61: linear regression for main model run with all biological data: turridae (sp).

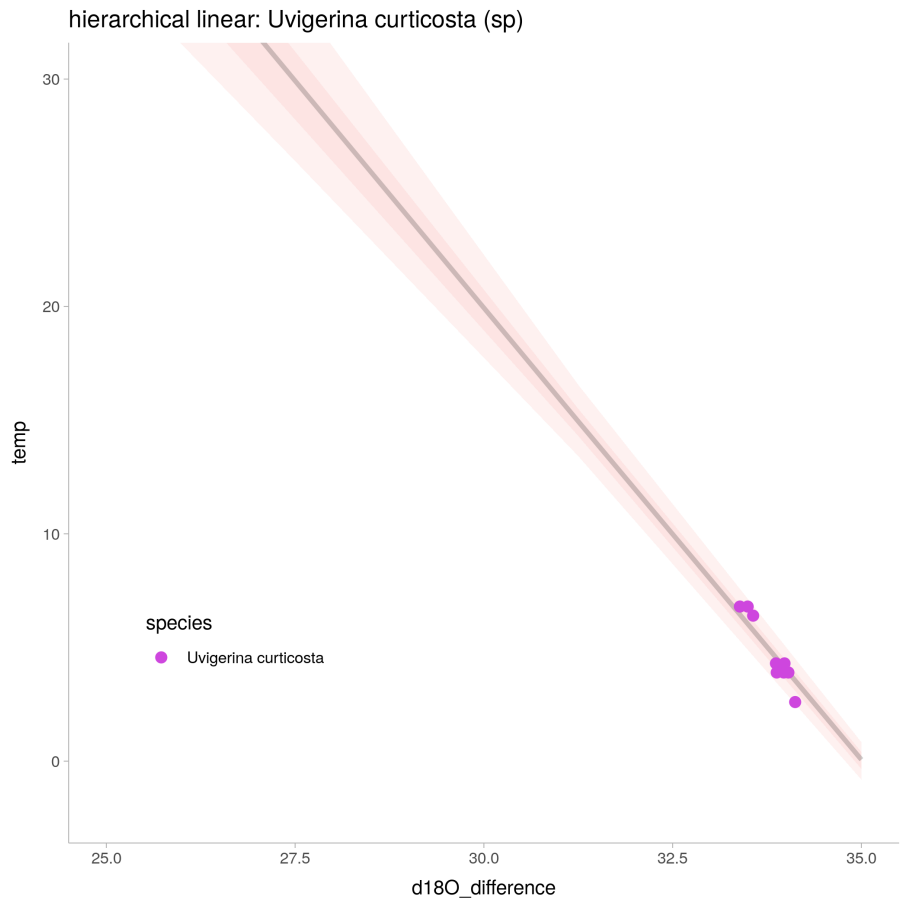

Figure 62: linear regression for main model run with all biological data: Uvigerina curticosta (sp).

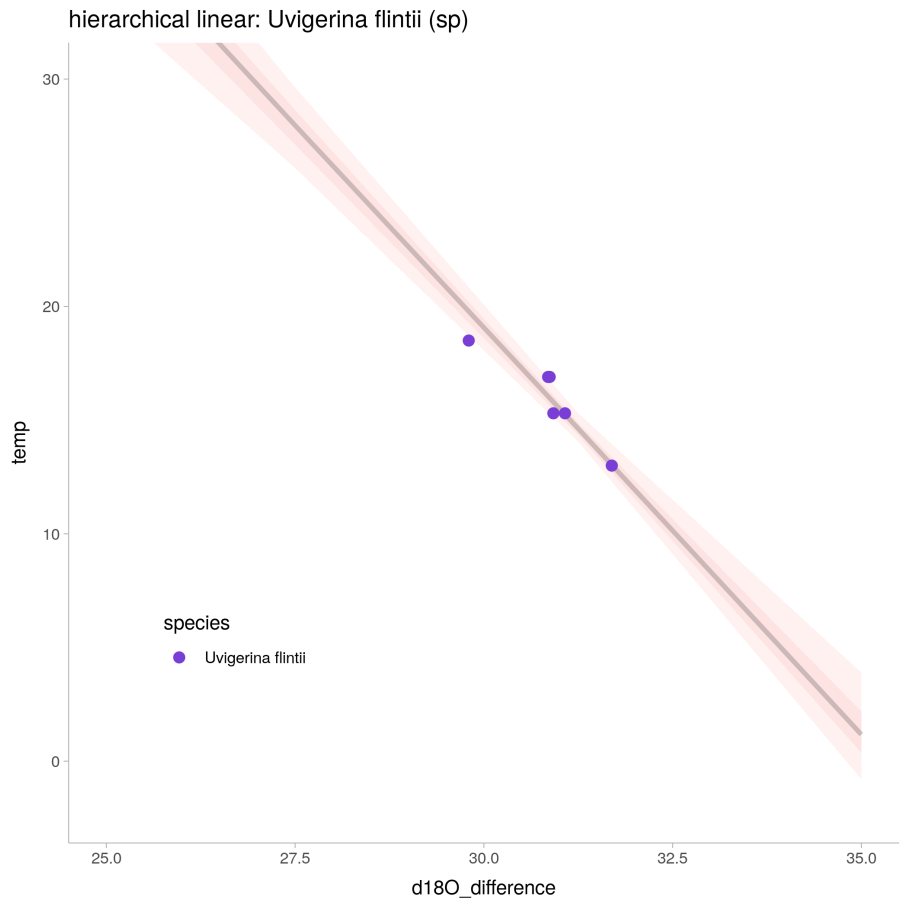

Figure 63: linear regression for main model run with all biological data: Uvigerina flintii (sp).

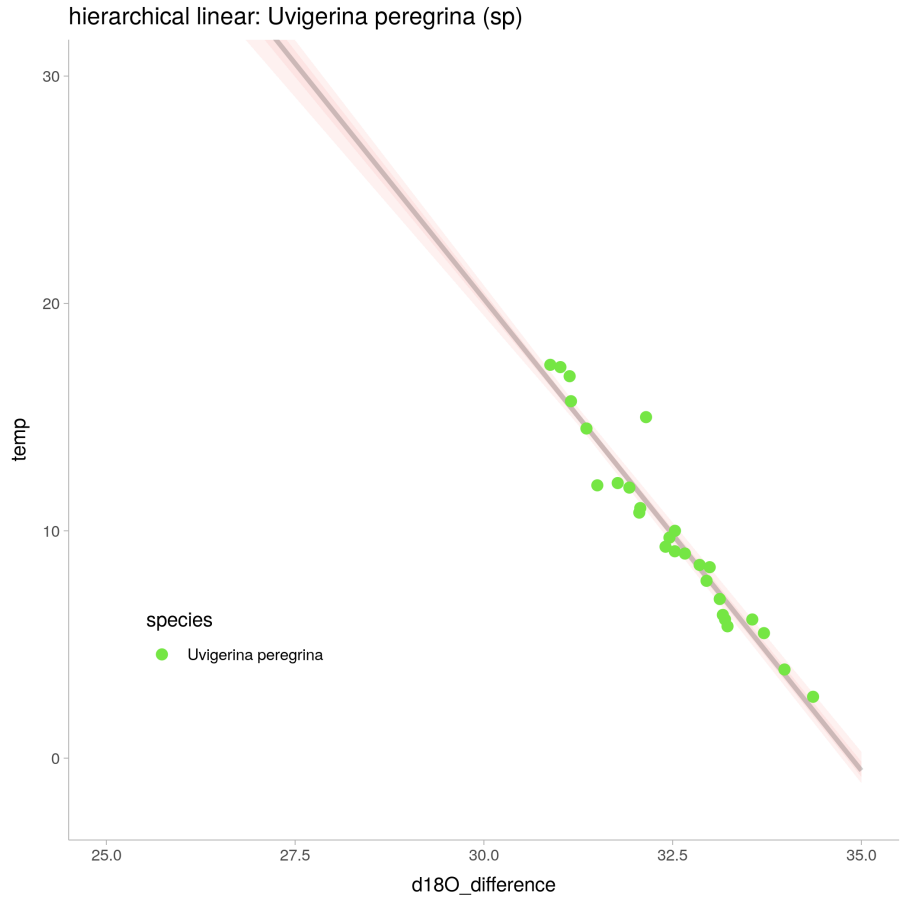

Figure 64: linear regression for main model run with all biological data: Uvigerina peregrina (sp).

8.7 hierarchical model using a quadratic component (foraminifera data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
sigma	1.00	6322	0.76	0.03	0.00	0.70	0.83
mu_a	1.00	2338	141.92	15.63	0.32	110.70	171.59
mu_b	1.00	2741	-4.43	0.87	0.02	-6.11	-2.71
mu_c	1.00	2247	0.01	0.02	0.00	-0.02	0.04
sigma_a	1.00	1663	18.67	7.83	0.19	8.18	38.18
sigma_b	1.01	1080	0.34	0.31	0.01	0.01	1.18
sigma_c	1.00	1462	0.02	0.01	0.00	0.01	0.04
log-posterior	1.01	814	-79.80	5.01	0.18	-90.24	-70.66
a[1]	1.00	2564	149.77	14.67	0.29	120.10	177.67
a[2]	1.00	1998	112.79	17.03	0.38	77.90	146.20
a[3]	1.00	2765	149.94	15.02	0.29	120.52	178.74
a[4]	1.00	3341	140.38	17.27	0.30	104.92	173.65
a[5]	1.00	3059	148.89	15.56	0.28	117.83	179.38
a[6]	1.00	3137	152.86	18.77	0.34	116.68	191.16
a[7]	1.00	3046	133.65	16.32	0.30	101.00	164.86
a[8]	1.00	2693	150.58	15.34	0.30	120.15	180.65
b[1]	1.00	2510	-4.45	0.92	0.02	-6.21	-2.64
b[2]	1.00	2025	-4.43	0.99	0.02	-6.37	-2.41
b[3]	1.00	2757	-4.45	0.92	0.02	-6.20	-2.63
b[4]	1.00	2773	-4.43	0.93	0.02	-6.24	-2.59
b[5]	1.00	2774	-4.46	0.92	0.02	-6.21	-2.66
b[6]	1.00	2759	-4.43	0.93	0.02	-6.24	-2.56
b[7]	1.00	2689	-4.40	0.94	0.02	-6.20	-2.53
b[8]	1.00	2673	-4.46	0.93	0.02	-6.28	-2.59
c[1]	1.00	2527	0.00	0.01	0.00	-0.03	0.03
c[2]	1.00	2092	0.03	0.01	0.00	0.00	0.06
c[3]	1.00	2755	0.01	0.01	0.00	-0.02	0.03
c[4]	1.00	3056	0.01	0.02	0.00	-0.02	0.04
c[5]	1.00	2916	0.00	0.02	0.00	-0.03	0.03
c[6]	1.00	2903	0.00	0.02	0.00	-0.03	0.03
c[7]	1.00	2778	0.02	0.02	0.00	-0.01	0.05
c[8]	1.00	2711	0.00	0.01	0.00	-0.03	0.03

Table 12: hierarchical model with quadratic component

model run time: 721.54

```
Computed from 4000 by 295 log-likelihood matrix

         Estimate   SE
elpd_loo   -347.3 24.6
p_loo        19.5  4.3
looic       694.7 49.1
------
Monte Carlo SE of elpd_loo is NA.
```

```
Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     291   98.6%   661
 (0.5, 0.7]   (ok)         2    0.7%   359
   (0.7, 1]   (bad)        2    0.7%   86
   (1, Inf)   (very bad)   0    0.0%   <NA>
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are too high.
See help('pareto-k-diagnostic') for details.
```

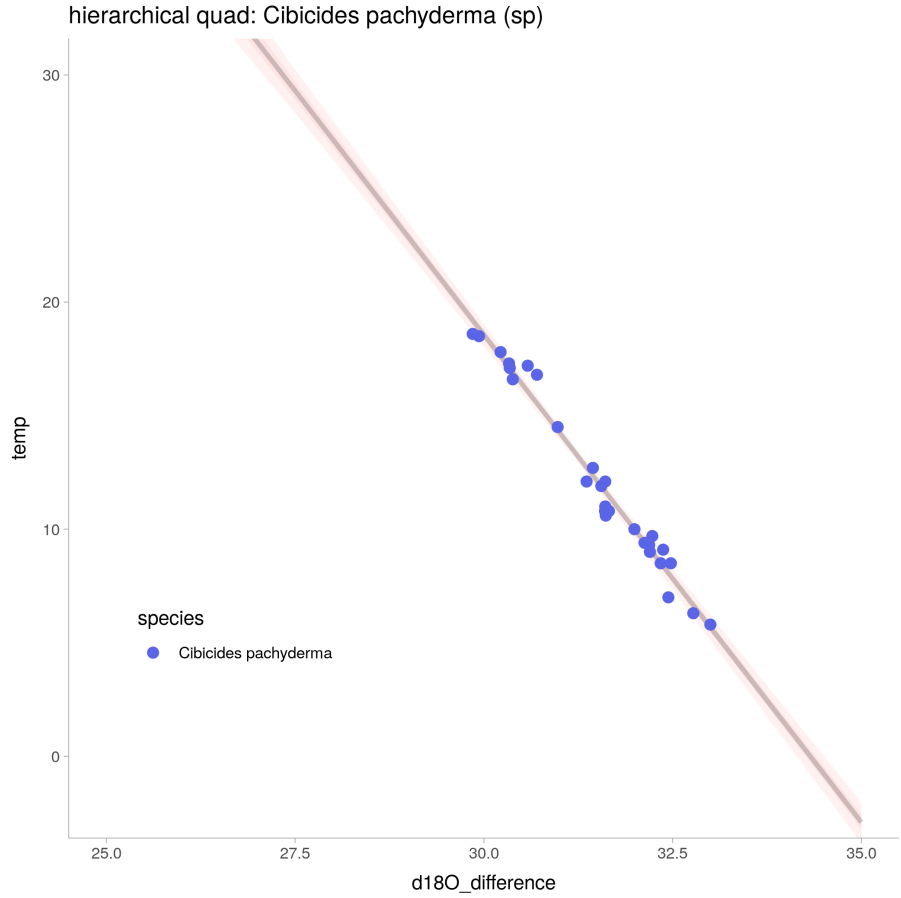

Figure 65: linear regression: hierarchical quad for Cibicides pachyderma (sp).

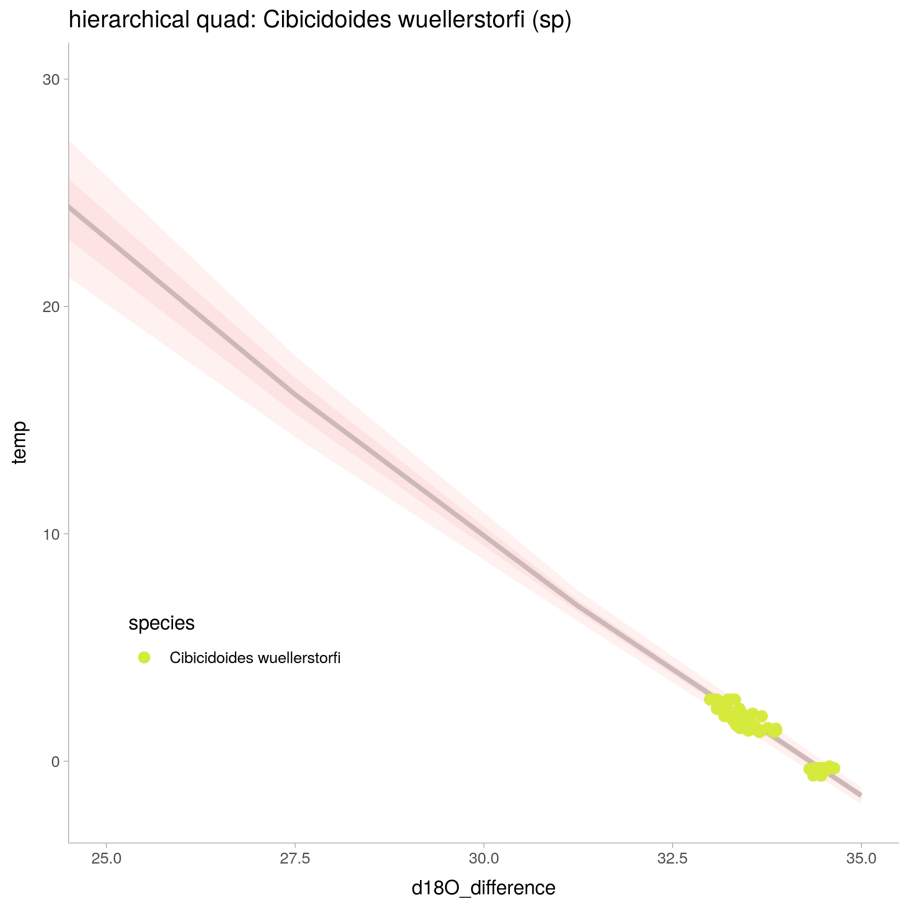

Figure 66: linear regression: hierarchical quad for Cibicidoides wuellerstorfi (sp).

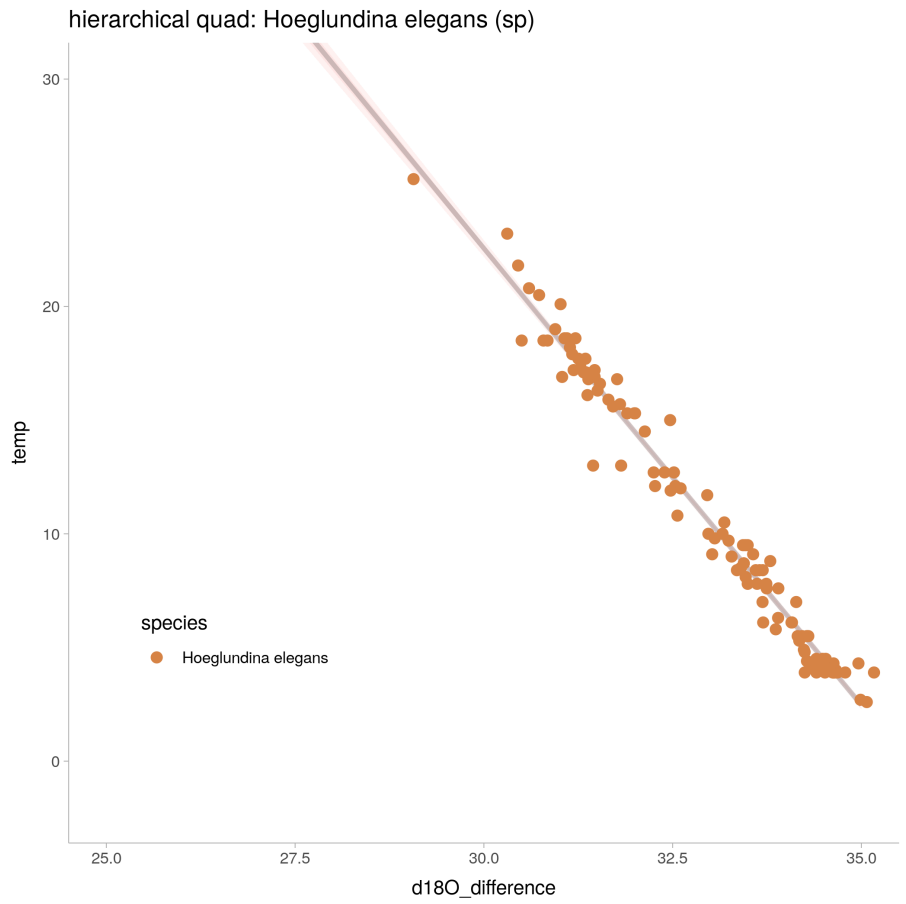

Figure 67: linear regression: hierarchical quad for Hoeglundina elegans (sp).

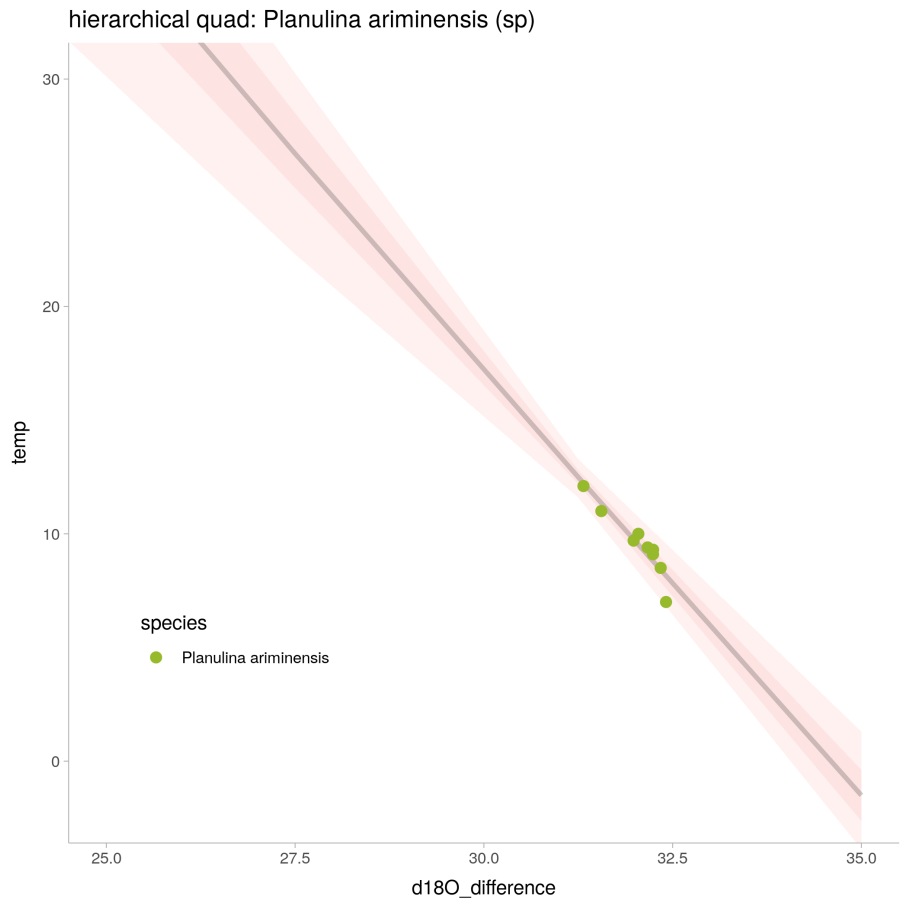

Figure 68: linear regression: hierarchical quad for Planulina ariminensis (sp).

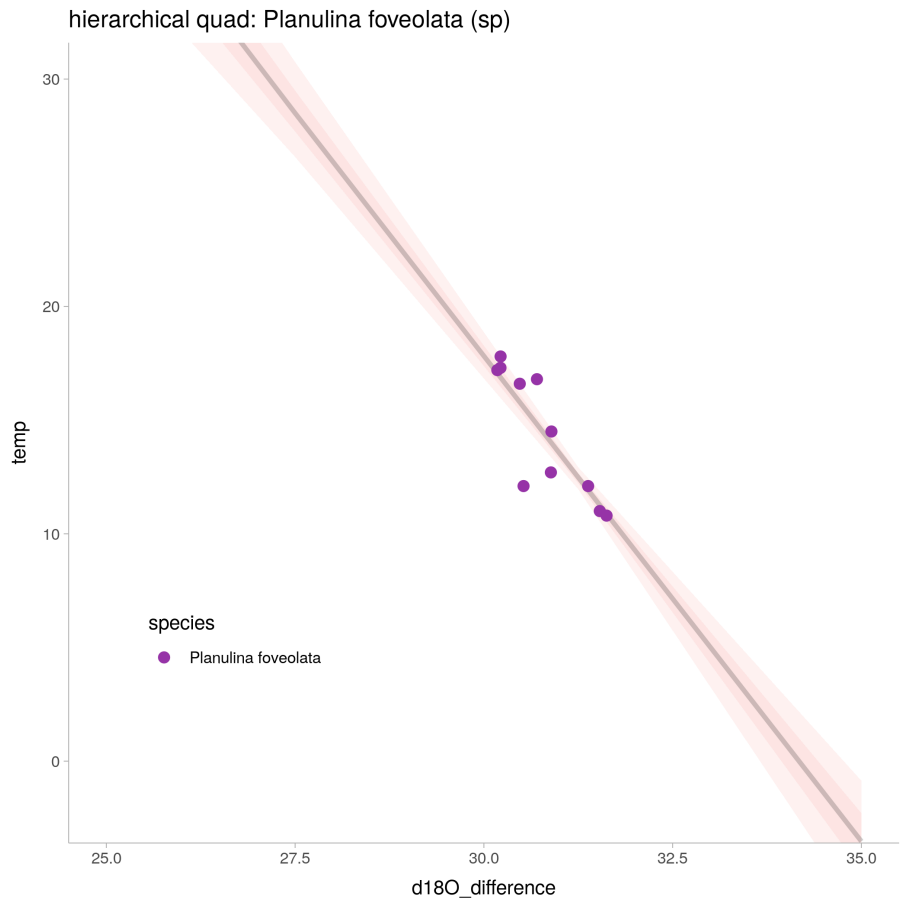

Figure 69: linear regression: hierarchical quad for Planulina foveolata (sp).

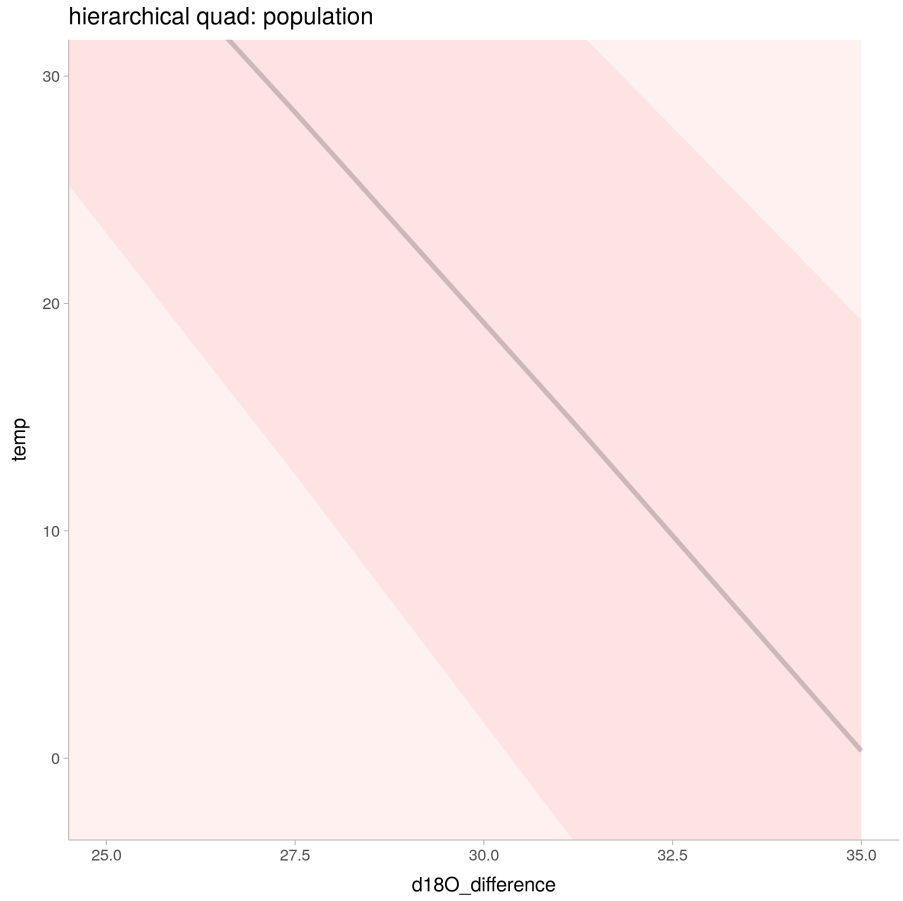

Figure 70: linear regression: hierarchical quad for population

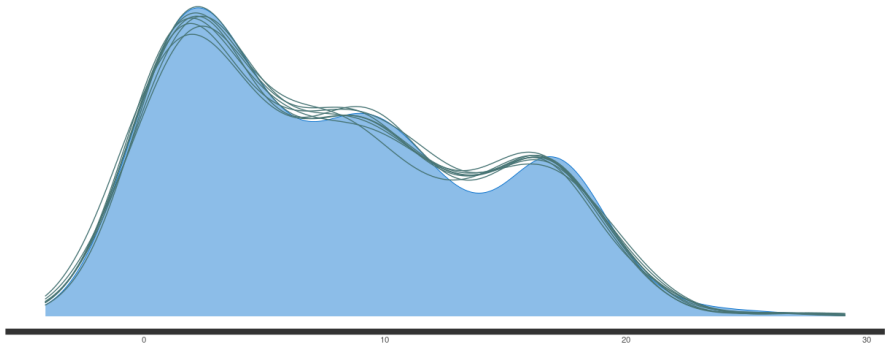

Figure 71: posterior predictive check for the hierarchical model with quadratic component

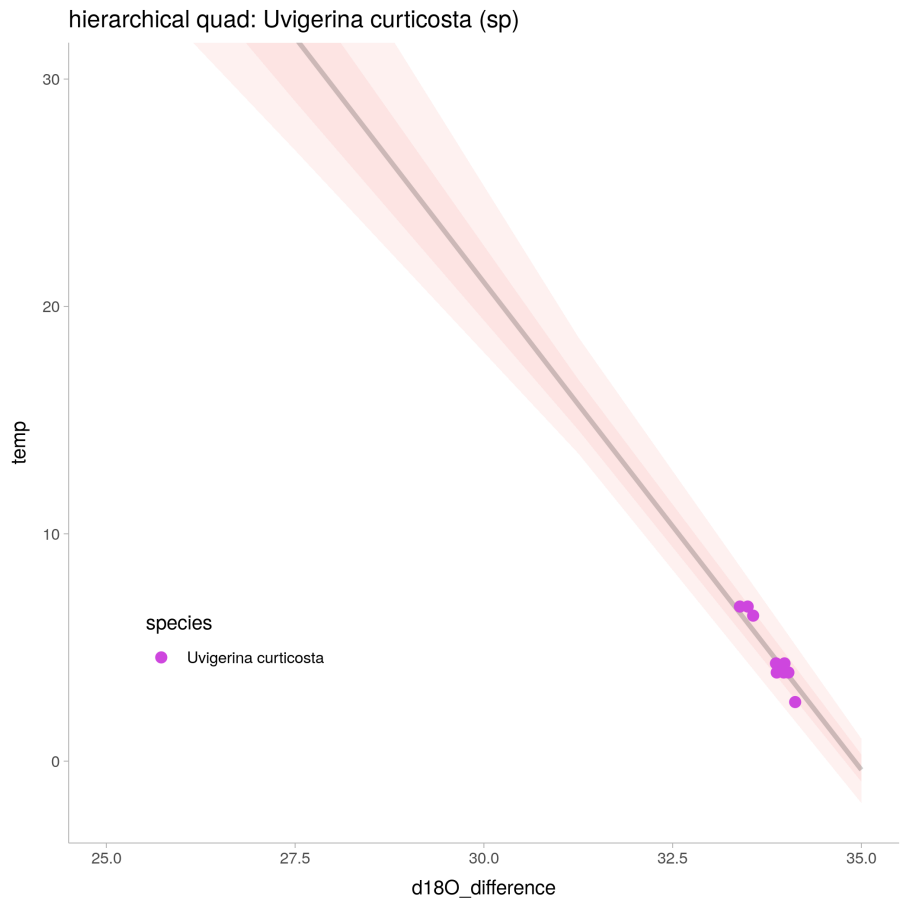

Figure 72: linear regression: hierarchical quad for Uvigerina curticosta (sp).

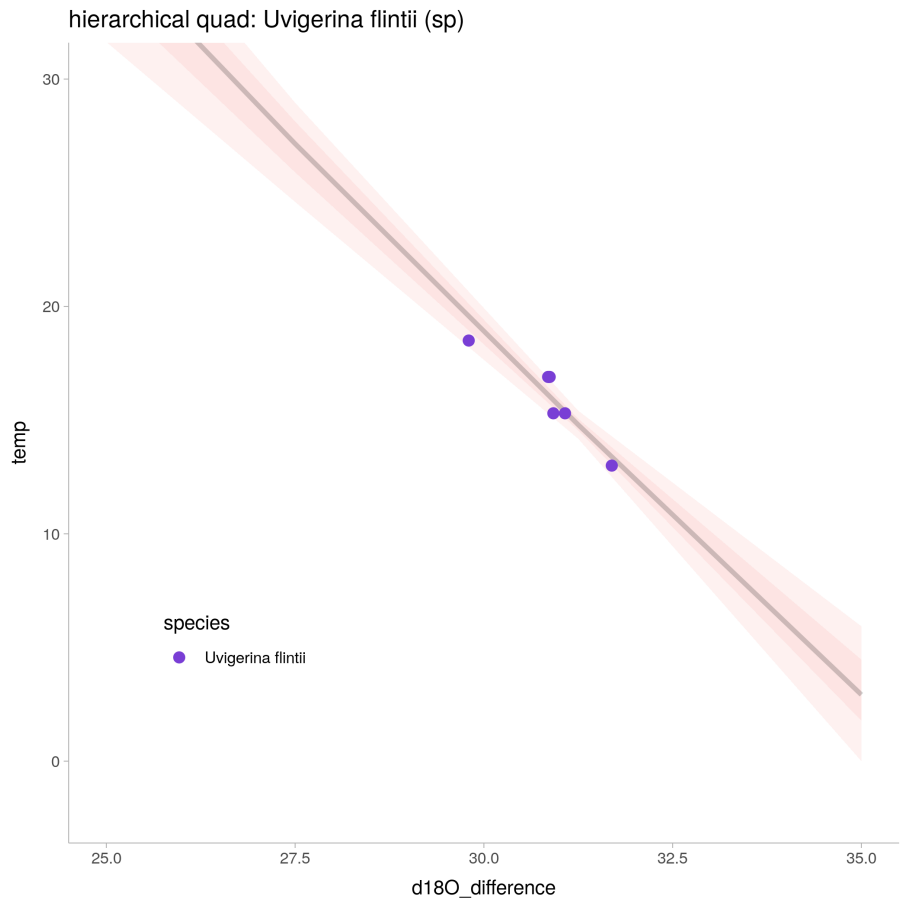

Figure 73: linear regression: hierarchical quad for Uvigerina flintii (sp).

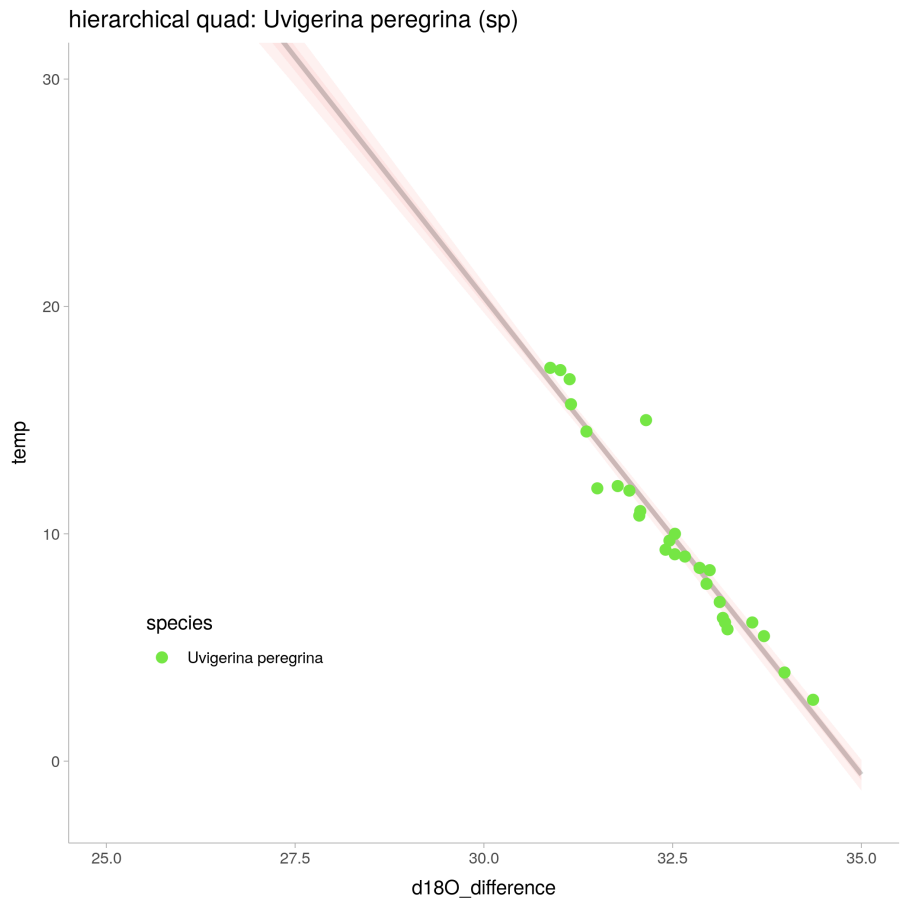

Figure 74: linear regression: hierarchical quad for Uvigerina peregrina (sp).

```
data {
  int N_obs;          /* nr of observations */
  int N_sp;           /* nr of groups */
  int K;          /* nr of features */
  vector[N_obs] temperature;
  matrix[N_obs, K] x;

  array[N_obs] int<lower=1, upper=N_sp> species;
  int<lower=0, upper=1> prior_only;
}

parameters {
  real<lower = 0> sigma;
  real mu_a;
  real mu_b;
  real mu_c;
  real<lower = 0> sigma_a;
  real<lower = 0> sigma_b;
  real<lower = 0> sigma_c;

  vector<offset = mu_a, multiplier = sigma_a>[N_sp] a;
  vector<offset = mu_b, multiplier = sigma_b>[N_sp] b;
  vector<offset = mu_c, multiplier = sigma_c>[N_sp] c;
}

model {
  vector[N_obs] mu;

  mu_a ~ normal(120,50);
  mu_b ~ normal(-4,1);
  mu_c ~ normal(0,1);
  sigma_a ~ normal(0,50);
  sigma_b ~ normal(0,5);
  sigma_c ~ normal(0,2);

  a ~ normal(mu_a, sigma_a);
  b ~ normal(mu_b, sigma_b);
  c ~ normal(mu_c, sigma_c);

  sigma ~ normal(0, 2);

  if (!prior_only) {
    for (i in 1:N_obs) {
      mu[i] = a[species[i]] + b[species[i]] * (x[i,1] - x[i,2]) + c[species[i]] * (x[i,1] - x[i,2])^2;
    }
   temperature ~ normal(mu, sigma);
  }
}
```

```
49
50 generated quantities {
51   vector[N_obs] log_lik;
52   vector[N_obs] temperature_sim;
53   vector[N_obs] mu;
54
55   for (i in 1:N_obs){
56     mu[i] = a[species[i]] + b[species[i]] * (x[i,1] - x[i,2]) + c[species[i]] * (x[i,1] - x[i,2])^2;
57     log_lik[i] = normal_lpdf(temperature[i] | mu[i],sigma);
58     temperature_sim[i] = normal_rng(mu[i],sigma);
59   }
60 }
```

Listing 10: The stan code for the model main model: hierarchical linear model: (all biological data)

8.8 hierarchical model with separate oxygen ratio parameters (foraminifera data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
sigma	1.00	4837	0.66	0.03	0.00	0.60	0.71
mu_a	1.00	1899	92.02	15.31	0.35	61.34	122.28
mu_b1	1.00	2087	-2.52	0.49	0.01	-3.51	-1.51
mu_b2	1.00	1798	6.62	1.53	0.04	3.14	9.47
sigma_a	1.00	1845	38.75	13.29	0.31	19.53	70.53
sigma_b1	1.00	1410	1.21	0.50	0.01	0.55	2.46
sigma_b2	1.00	1758	3.52	1.36	0.03	1.57	6.84
log-posterior	1.00	813	-30.56	5.71	0.20	-43.21	-20.43
a[1]	1.00	4320	84.27	16.83	0.26	50.19	116.36
a[2]	1.00	3802	74.23	6.07	0.10	62.01	85.81
a[3]	1.00	3324	137.11	3.81	0.07	129.69	144.39
a[4]	1.00	3063	84.98	25.22	0.46	32.65	133.16
a[5]	1.00	3255	57.29	19.81	0.35	16.91	94.24
a[6]	1.00	2455	138.45	25.81	0.52	92.70	193.12
a[7]	1.00	3752	76.76	18.33	0.30	39.31	111.09
a[8]	1.00	3944	64.83	14.32	0.23	36.95	92.84
b1[1]	1.00	4307	-2.40	0.51	0.01	-3.37	-1.38
b1[2]	1.00	3826	-2.17	0.19	0.00	-2.54	-1.78
b1[3]	1.00	3320	-3.85	0.11	0.00	-4.06	-3.63
b1[4]	1.00	3082	-2.42	0.76	0.01	-3.87	-0.83
b1[5]	1.00	3260	-1.61	0.61	0.01	-2.75	-0.37
b1[6]	1.00	2458	-3.95	0.76	0.02	-5.57	-2.60
b1[7]	1.00	3849	-2.12	0.55	0.01	-3.14	-0.99
b1[8]	1.00	3945	-1.80	0.42	0.01	-2.63	-0.98
b2[1]	1.00	4486	8.15	1.02	0.02	6.16	10.20
b2[2]	1.00	4145	3.31	2.60	0.04	-2.14	7.98
b2[3]	1.00	3376	4.41	0.25	0.00	3.91	4.92
b2[4]	1.00	3778	7.32	2.28	0.04	2.79	11.96

b2[5]	1.00	3837	10.00	1.03	0.02	7.98	12.08
b2[6]	1.00	3040	4.20	3.42	0.06	-3.18	10.29
b2[7]	1.00	4839	6.82	1.87	0.03	3.18	10.59
b2[8]	1.00	4012	9.86	0.95	0.01	8.01	11.69

Table 13: hierarchical model with separate parameters

model run time: 326.12

```
Computed from 4000 by 295 log-likelihood matrix

         Estimate   SE
elpd_loo   -306.3 23.1
p_loo        21.6  4.1
looic       612.6 46.3
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     288   97.6%   1226
 (0.5, 0.7]   (ok)         5    1.7%   207
   (0.7, 1]   (bad)        2    0.7%   113
   (1, Inf)   (very bad)   0    0.0%   <NA>
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are too high.
See help('pareto-k-diagnostic') for details.
```

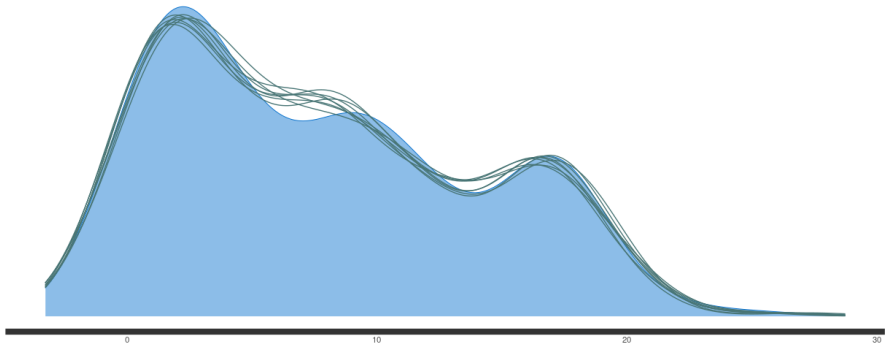

Figure 75: posterior predictive check for the hierarchical model with separate parameters

```
data {
  int N_obs;    /* nr of observations */
  int N_sp; /* nr of groups */
  int K;    /* nr of features */
  vector[N_obs] temperature;
  matrix[N_obs, K] x;

  array[N_obs] int<lower=0, upper=1> is_lab;

  array[N_obs] int<lower=1, upper=N_sp> species;
  int<lower=0, upper=1> prior_only;

  real<lower=0> sigma_T_lab;
}

parameters {
  real<lower = 0> sigma;

  real mu_a;
  real mu_b1;
  real mu_b2;

  real<lower = 0> sigma_a;
  real<lower = 0> sigma_b1;
  real<lower = 0> sigma_b2;

  vector<offset = mu_a, multiplier = sigma_a>[N_sp] a;
  vector<offset = mu_b1, multiplier = sigma_b1>[N_sp] b1;
  vector<offset = mu_b2, multiplier = sigma_b2>[N_sp] b2;
}

model {
  vector[N_obs] mu;

  mu_a ~ normal(120, 50);
  mu_b1 ~ normal(0, 10);
  mu_b2 ~ normal(0, 10);

  sigma_a ~ normal(0, 50);
  sigma_b1 ~ normal(0, 5);
  sigma_b2 ~ normal(0, 5);

  sigma ~ normal(0, 2);

  a ~ normal(mu_a, sigma_a);
  b1 ~ normal(mu_b1, sigma_b1);
  b2 ~ normal(mu_b2, sigma_b2);

```

```
  if (!prior_only) {
    mu = a[species] + b1[species] .* x[,1] + b2[species] .* x[,2];
    temperature ~ normal(mu, sigma);
  }
}

generated quantities {
  vector[N_obs] log_lik;
  vector[N_obs] temperature_sim;
  vector[N_obs] mu;

  for (i in 1:N_obs){
    mu[i] = a[species[i]] + b1[species[i]] * x[i,1] + b2[species[i]] * x[i,2];
    log_lik[i] = normal_lpdf(temperature[i] | mu[i],sigma);
    temperature_sim[i] = normal_rng(mu[i],sigma);
  }
}
```

Listing 11: The stan code for the model hierarchical model with separate oxygen ratios parameters (foraminifera data)

8.9 hierarchical model with deming regression (foraminifera data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
sigma	1.01	476	0.36	0.05	0.00	0.26	0.45
mu_a	1.00	3090	125.63	7.89	0.14	109.80	140.89
mu_b	1.00	1706	-3.89	0.38	0.01	-4.69	-3.14
sigma_a	1.00	2371	35.90	10.79	0.22	20.78	62.09
sigma_b	1.00	2554	1.10	0.35	0.01	0.62	1.92
log-posterior	1.01	414	-179.89	30.72	1.51	-234.12	-114.01
a[1]	1.00	6362	147.83	3.92	0.05	140.20	155.58
a[2]	1.00	3631	69.69	3.09	0.05	63.60	76.00
a[3]	1.00	4374	145.53	1.49	0.02	142.61	148.50
a[4]	1.00	5120	129.46	13.31	0.19	105.09	157.14
a[5]	1.00	2580	168.14	11.08	0.22	147.65	190.95
a[6]	1.00	4089	142.75	17.36	0.27	111.04	178.76
a[7]	1.00	4686	116.86	13.75	0.20	92.55	145.92
a[8]	1.00	3828	153.99	4.68	0.08	145.31	163.33
b[1]	1.00	6360	-4.31	0.12	0.00	-4.55	-4.07
b[2]	1.00	3647	-2.03	0.09	0.00	-2.22	-1.85
b[3]	1.00	4277	-4.09	0.05	0.00	-4.18	-4.01
b[4]	1.00	5114	-3.74	0.42	0.01	-4.61	-2.98
b[5]	1.00	2589	-4.99	0.36	0.01	-5.73	-4.33
b[6]	1.00	4082	-4.08	0.51	0.01	-5.15	-3.15
b[7]	1.00	4696	-3.27	0.45	0.01	-4.21	-2.48
b[8]	1.00	3823	-4.44	0.14	0.00	-4.73	-4.17

Table 14: hierarchical model with deming regression

model run time: 815.63

```
Computed from 4000 by 295 log-likelihood matrix

         Estimate     SE
elpd_loo  -6821.2  550.1
p_loo      6525.9  526.1
looic     13642.3 1100.3
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)       4    1.4%   1043
 (0.5, 0.7]   (ok)        13    4.4%   625
   (0.7, 1]   (bad)       20    6.8%   32
   (1, Inf)   (very bad) 258   87.5%   1
See help('pareto-k-diagnostic') for details.
Warning message:
```

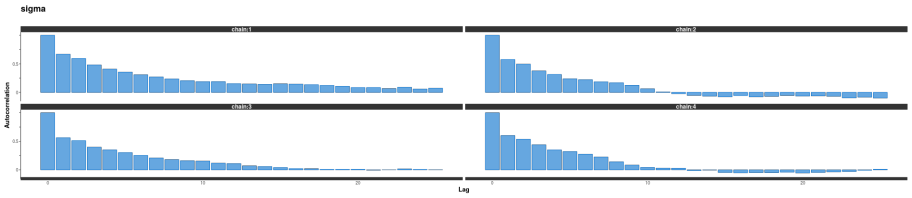

Figure 76: autocorrelation plot for the hierarchical model with deming regression. Even though this model gets mostly the same output, the increased set of parameters means the model fits much more slowly and has a hard time getting rid of its autocorrelation

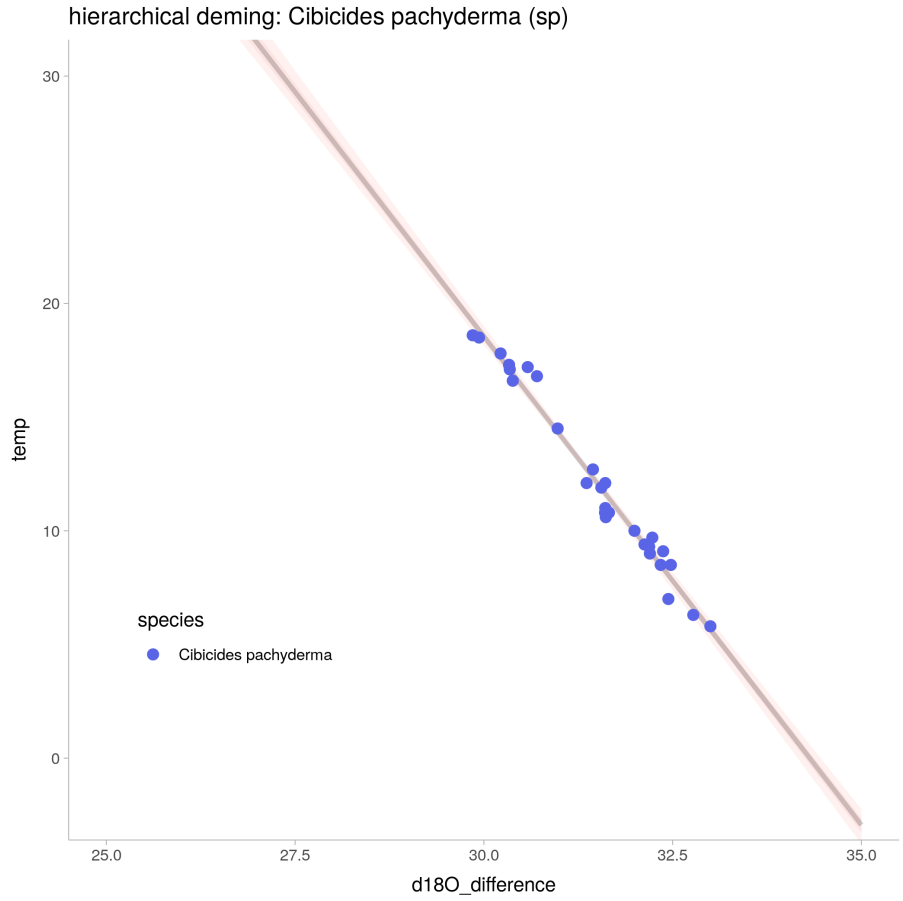

Figure 77: linear regression: hierarchical deming for Cibicides pachyderma (sp).

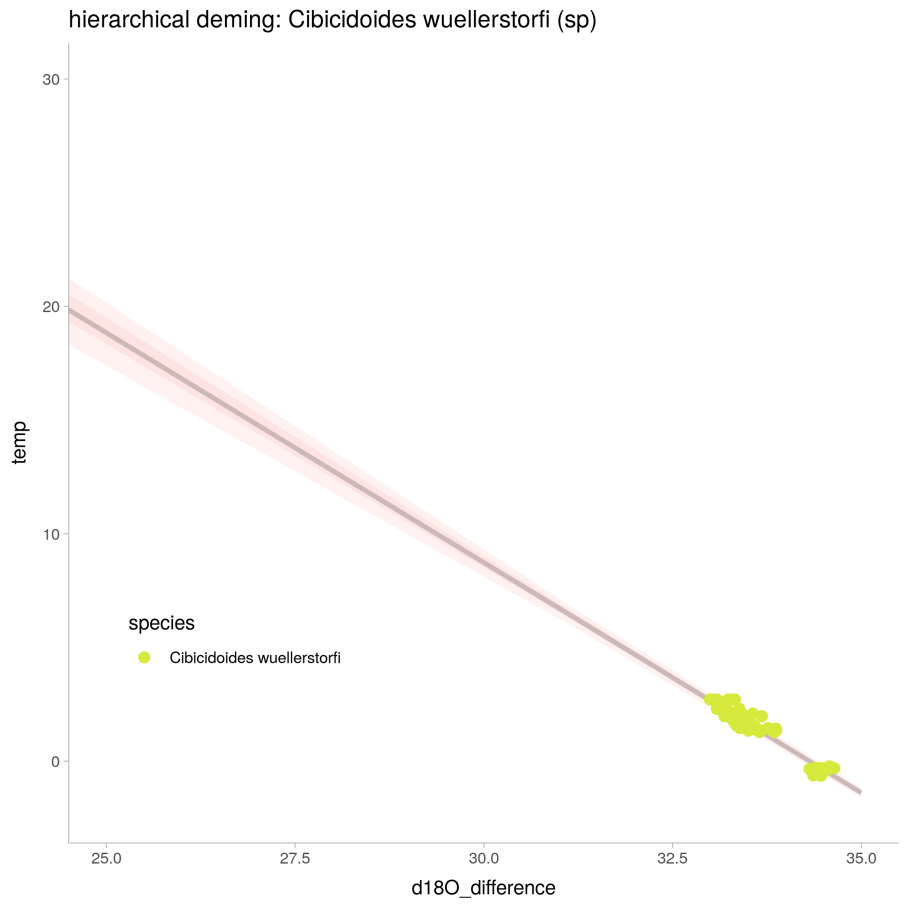

Figure 78: linear regression: hierarchical deming for Cibicidoides wuellerstorfi (sp).

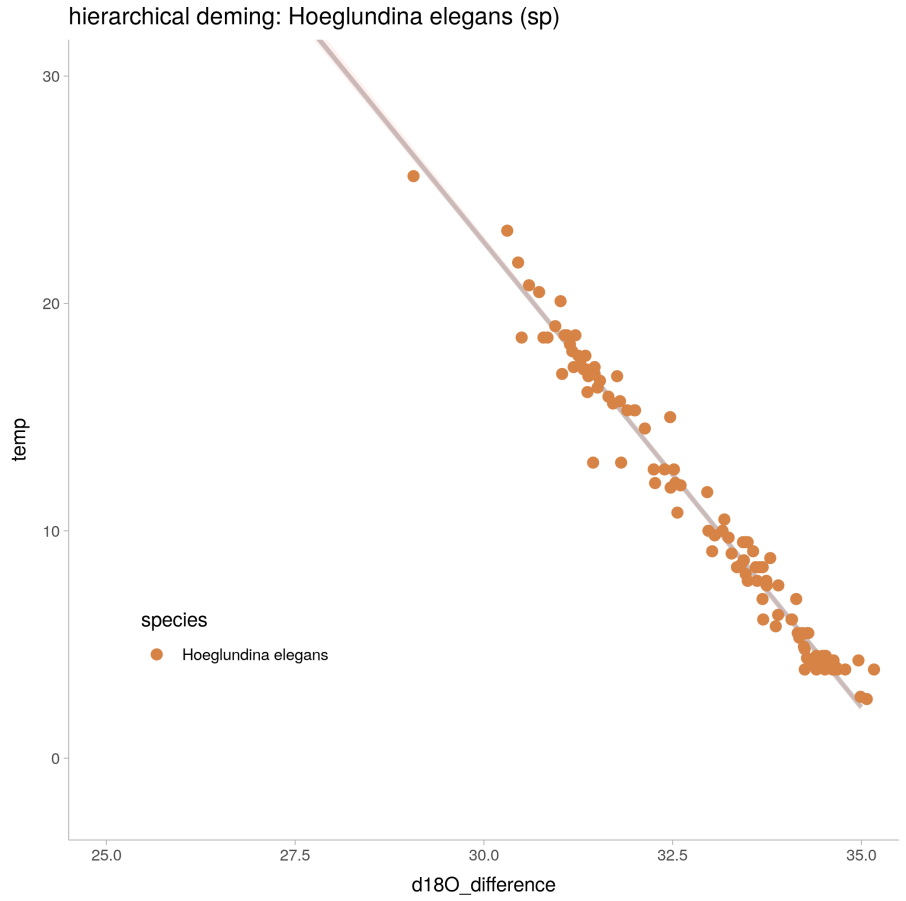

Figure 79: linear regression: hierarchical deming for Hoeglundina elegans (sp).

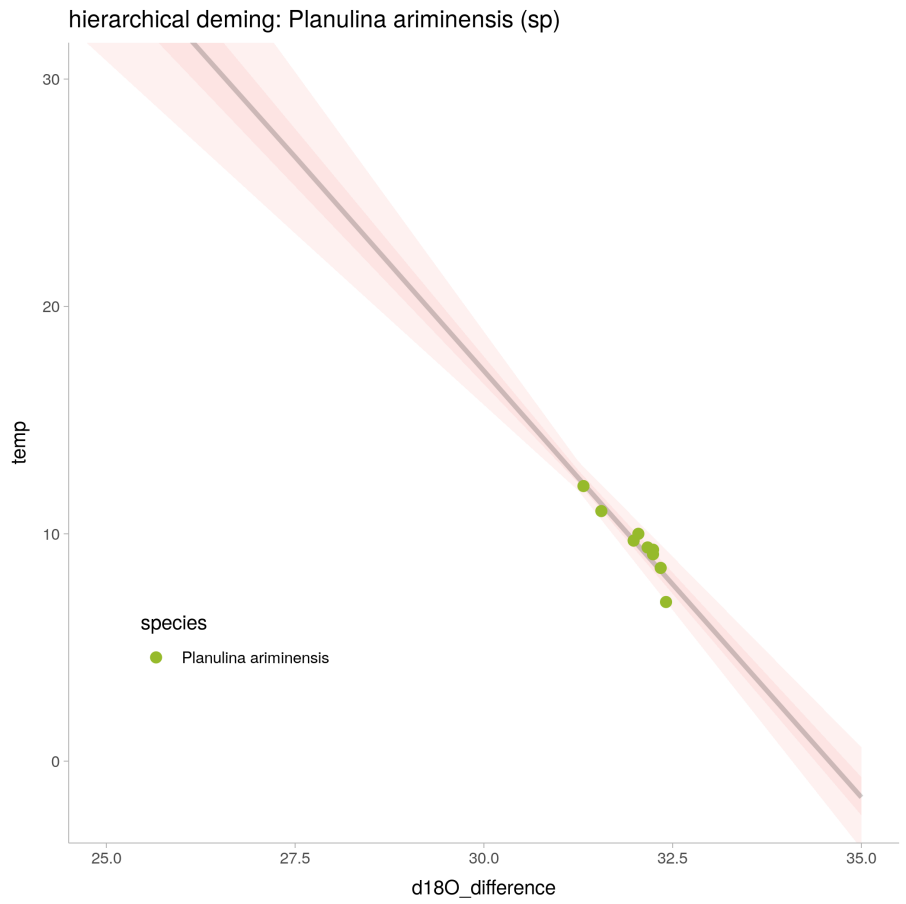

Figure 80: linear regression: hierarchical deming for Planulina ariminensis (sp).

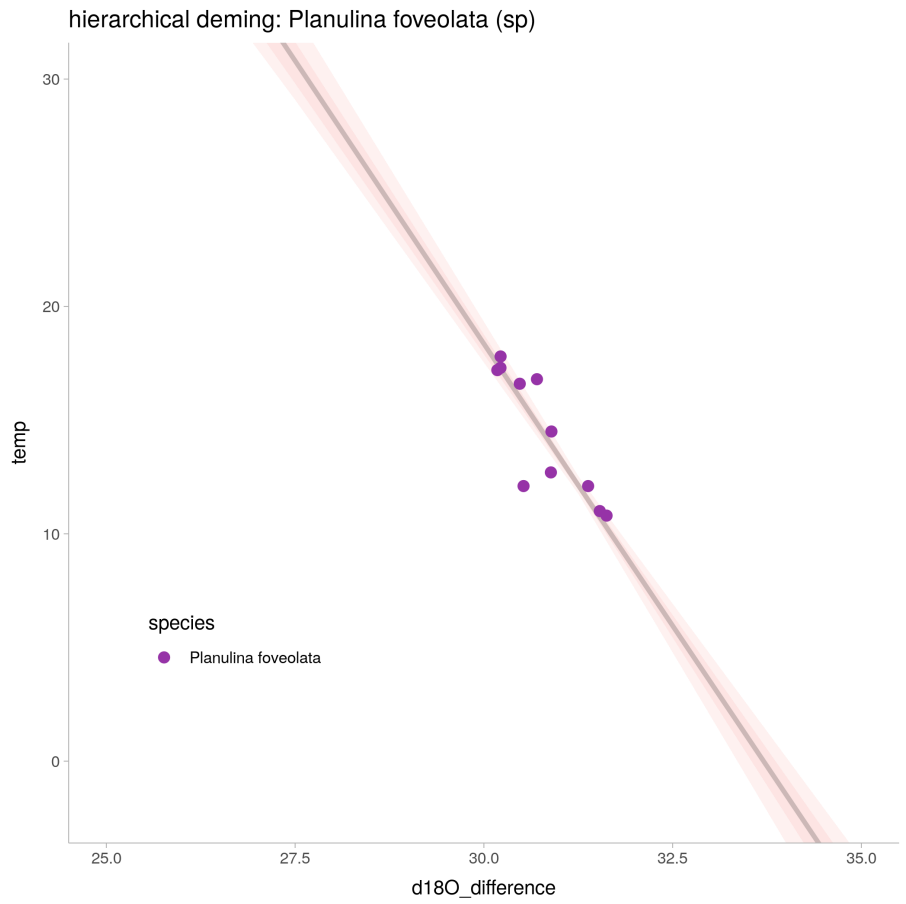

Figure 81: linear regression: hierarchical deming for Planulina foveolata (sp).

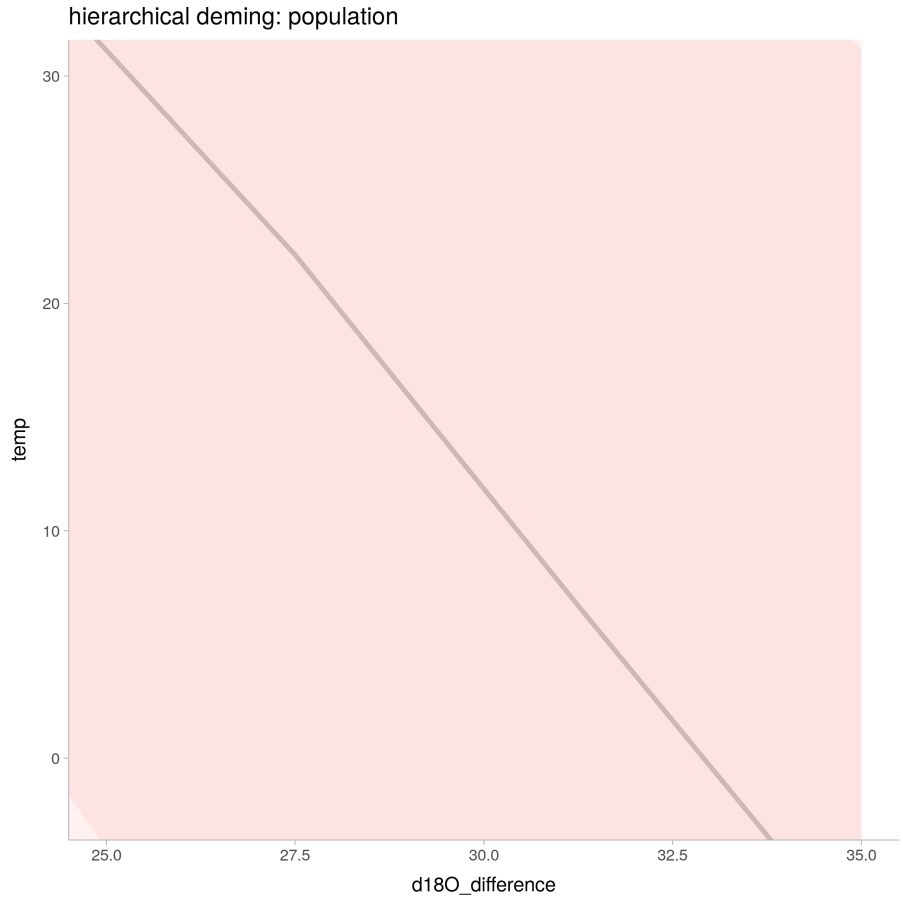

Figure 82: linear regression: hierarchical deming for population

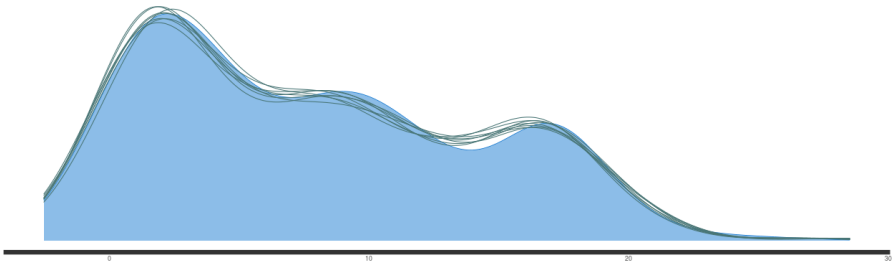

Figure 83: posterior predictive check for the hierarchical deming model

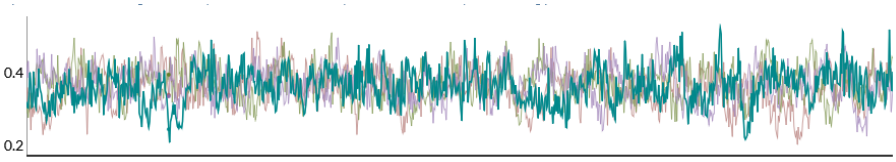

Figure 84: trace plot for the hierarchical deming model

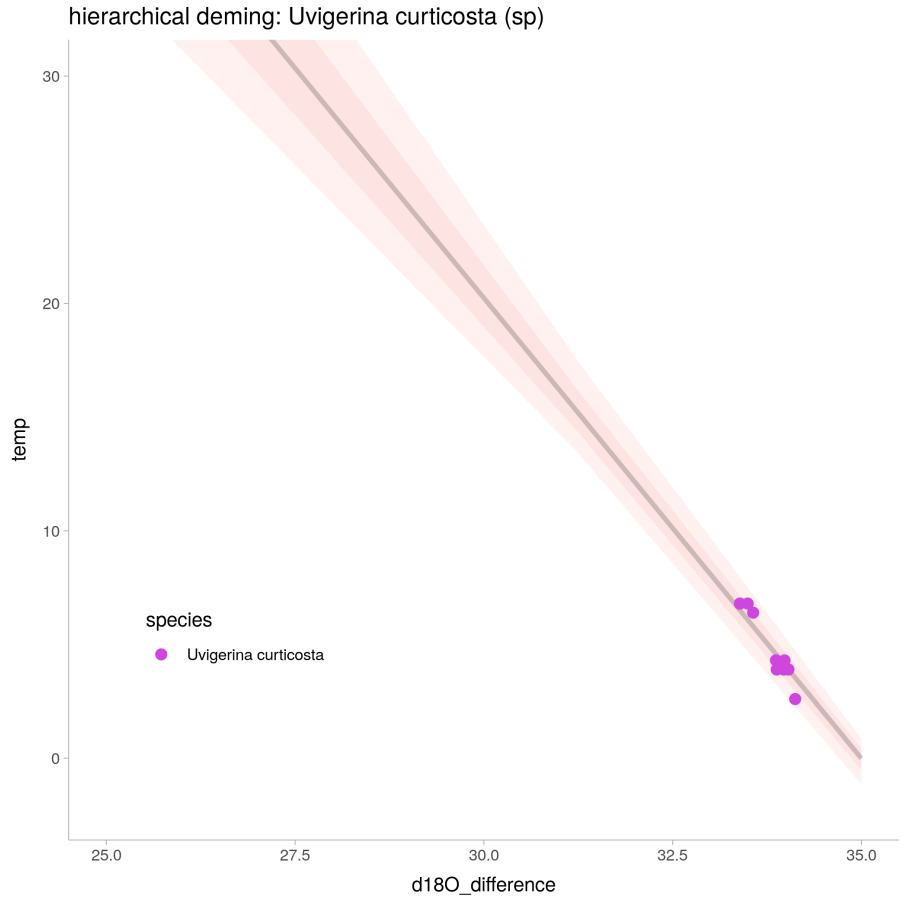

Figure 85: linear regression: hierarchical deming for Uvigerina curticosta (sp).

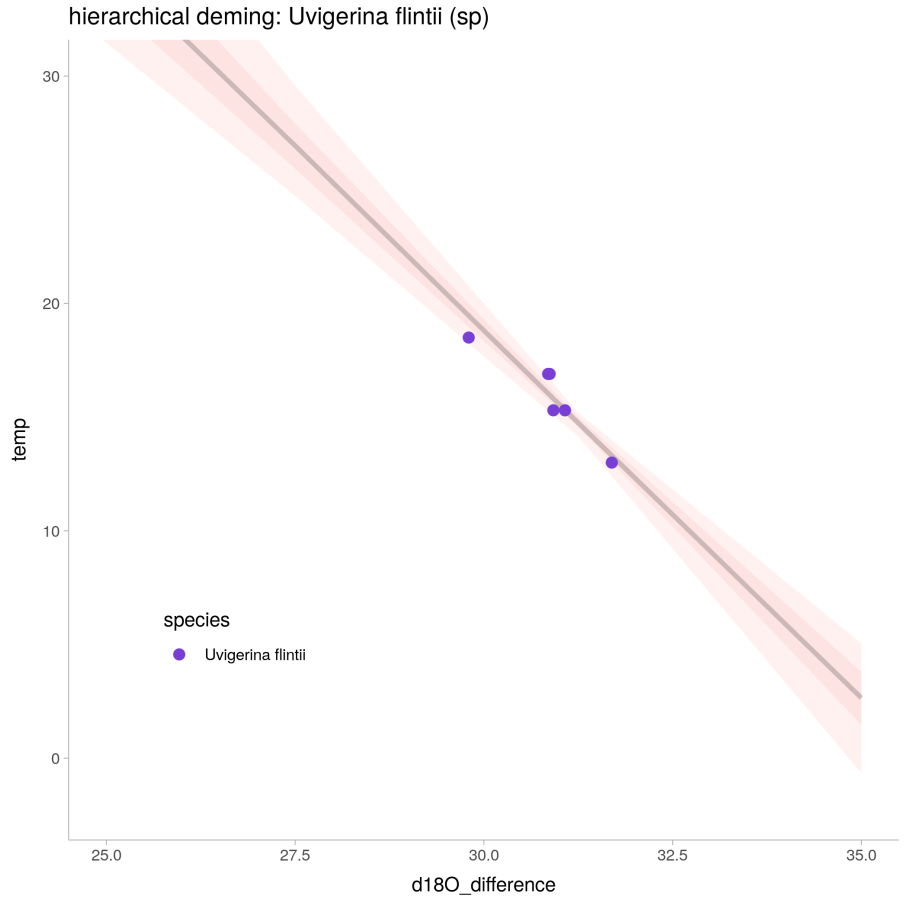

Figure 86: linear regression: hierarchical deming for Uvigerina flintii (sp).

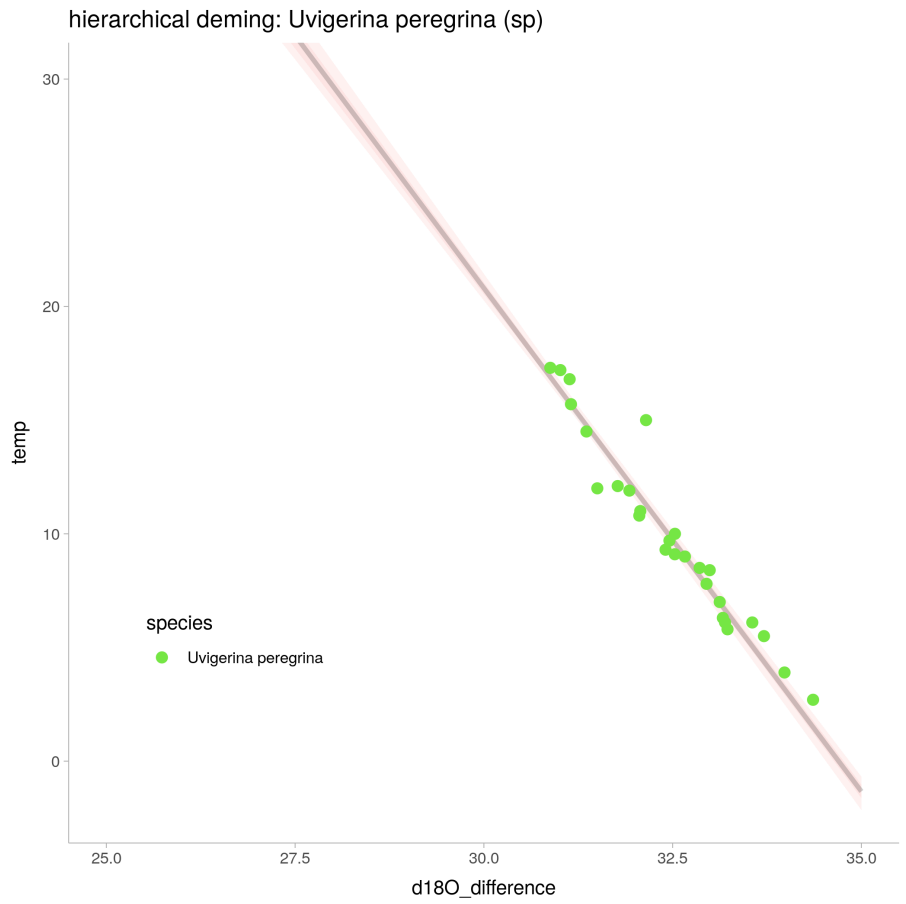

Figure 87: linear regression: hierarchical deming for Uvigerina peregrina (sp).

```
data {
  int N_obs;          /* nr of observations */
  int N_sp;          /* nr of groups */
  int K;          /* nr of features */
  vector[N_obs] temperature;
  matrix[N_obs, K] x;

  array[N_obs] int<lower=1, upper=N_sp> species;
  int<lower=0, upper=1> prior_only;
}

parameters {
  real<lower = 0> sigma;

  real mu_a;
  real mu_b;

  real<lower = 0> sigma_a;
  real<lower = 0> sigma_b;

  vector<offset = mu_a, multiplier = sigma_a>[N_sp] a;
  vector<offset = mu_b, multiplier = sigma_b>[N_sp] b;

  vector[N_obs] d18_O_c;
  vector[N_obs] d18_O_w;
}
model {

  vector[N_obs] mu;

  mu_a ~ normal(120, 10);
  mu_b ~ normal(-4, 1);

  sigma_a ~ normal(0, 50);
  sigma_b ~ normal(0, 2);

  a ~ normal(mu_a, sigma_a);
  b ~ normal(mu_b, sigma_b);

  sigma ~ normal(0, 2);

  if (!prior_only) {
    d18_O_c ~ normal(x[,1], x[,3]);
    d18_O_w ~ normal(x[,2], x[,4]);

    mu = a[species] + b[species] .* (d18_O_c - d18_O_w);

    temperature ~ normal(mu, sigma);
```

```
  }
}

generated quantities {
  vector[N_obs] log_lik;
  vector[N_obs] temperature_sim;
  vector[N_obs] mu;

  mu = a[species] + b[species] .* (d18_O_c - d18_O_w);
  for (i in 1:N_obs){
    log_lik[i] = normal_lpdf(temperature[i] | mu,sigma);
    temperature_sim[i] = normal_rng(mu[i],sigma);
  }
}
```

Listing 12: The stan code for the model hierarchical model with deming regression (foraminifera data)

8.10 hierarchical model with a correlation matrix (foraminifera data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
log-posterior	1.00	613	-77.49	4.95	0.20	-87.65	-68.25
ab_pop[1]	1.00	1384	144.25	6.25	0.17	132.05	157.14
ab_pop[2]	1.00	1094	-4.14	0.21	0.01	-4.55	-3.74
sigma_obs	1.00	4027	0.76	0.03	0.00	0.70	0.82
sd_sp[1]	1.00	987	28.51	7.99	0.25	16.86	47.18
sd_sp[2]	1.01	1026	0.83	0.24	0.01	0.48	1.39
ab_sp[1,1]	1.00	5817	147.50	4.98	0.07	137.88	157.35
ab_sp[2,1]	1.00	5809	-4.30	0.16	0.00	-4.61	-3.99
ab_sp[1,2]	1.00	3331	74.74	5.65	0.10	63.61	85.90
ab_sp[2,2]	1.00	3322	-2.18	0.17	0.00	-2.51	-1.85
ab_sp[1,3]	1.00	3885	142.85	1.64	0.03	139.66	146.05
ab_sp[2,3]	1.00	3908	-4.01	0.05	0.00	-4.11	-3.92
ab_sp[1,4]	1.00	3869	132.47	16.63	0.27	98.91	165.53
ab_sp[2,4]	1.00	3864	-3.84	0.52	0.01	-4.87	-2.79
ab_sp[1,5]	1.00	3765	146.98	11.63	0.19	124.29	170.27
ab_sp[2,5]	1.00	3752	-4.30	0.38	0.01	-5.06	-3.57
ab_sp[1,6]	1.00	2907	161.53	21.87	0.41	120.48	207.47
ab_sp[2,6]	1.00	2906	-4.64	0.65	0.01	-5.99	-3.42
ab_sp[1,7]	1.00	3528	120.11	14.18	0.24	91.56	146.83
ab_sp[2,7]	1.00	3522	-3.37	0.46	0.01	-4.24	-2.45
ab_sp[1,8]	1.00	6375	146.99	5.01	0.06	137.22	156.84
ab_sp[2,8]	1.00	6445	-4.22	0.15	0.00	-4.52	-3.92
Sigma[1,1]			1.00	0.00		1.00	1.00
Sigma[2,1]	1.01	596	-0.96	0.08	0.00	-1.00	-0.77
Sigma[1,2]	1.01	596	-0.96	0.08	0.00	-1.00	-0.77
Sigma[2,2]			1.00	0.00		1.00	1.00

Table 15: hierarchical model with correlated parameters

model run time: 132.63

```
Computed from 4000 by 295 log-likelihood matrix

         Estimate   SE
elpd_loo   -346.9 24.6
p_loo        19.0  4.3
looic       693.9 49.3
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     293   99.3%   400
 (0.5, 0.7]   (ok)         1    0.3%   134
   (0.7, 1]   (bad)        1    0.3%   43
   (1, Inf)   (very bad)   0    0.0%   <NA>
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are too high.
See help('pareto-k-diagnostic') for details.
```

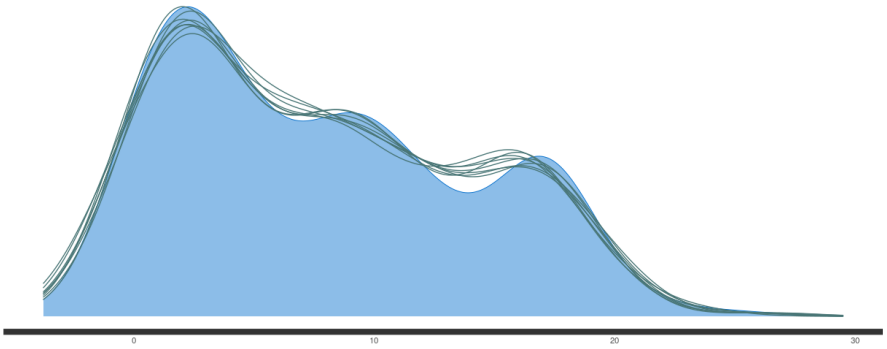

Figure 88: linear regression: corr matrix posterior predictive check

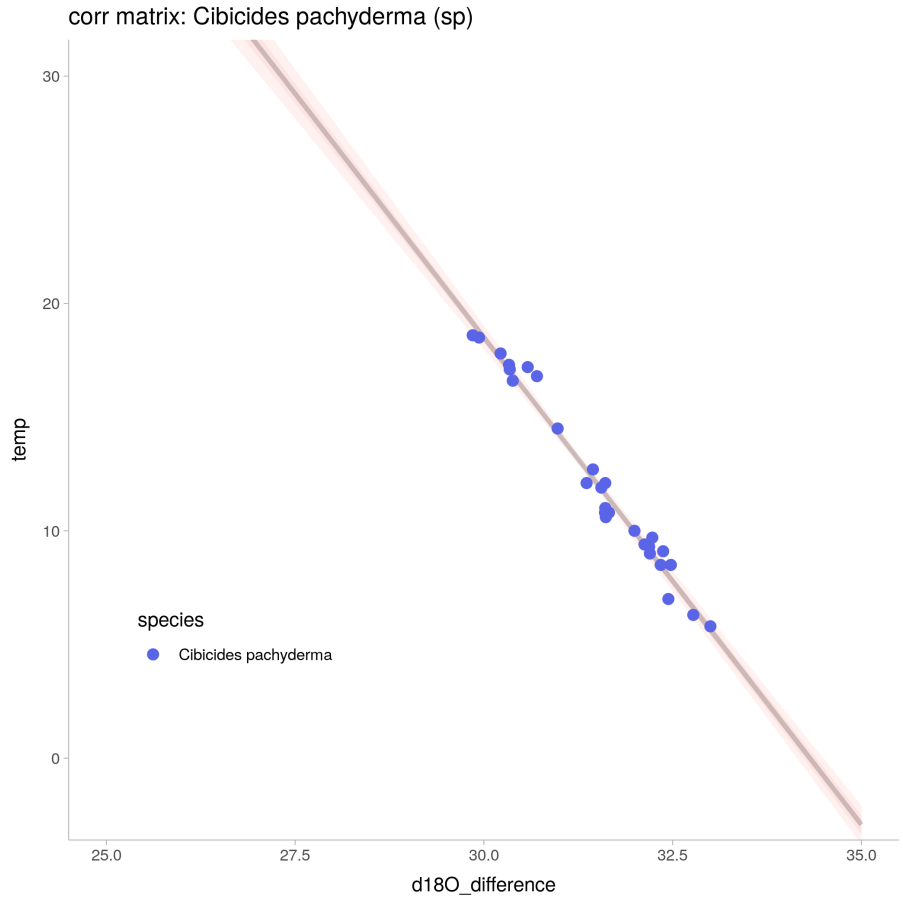

Figure 89: linear regression: corr matrix for Cibicides pachyderma (sp).

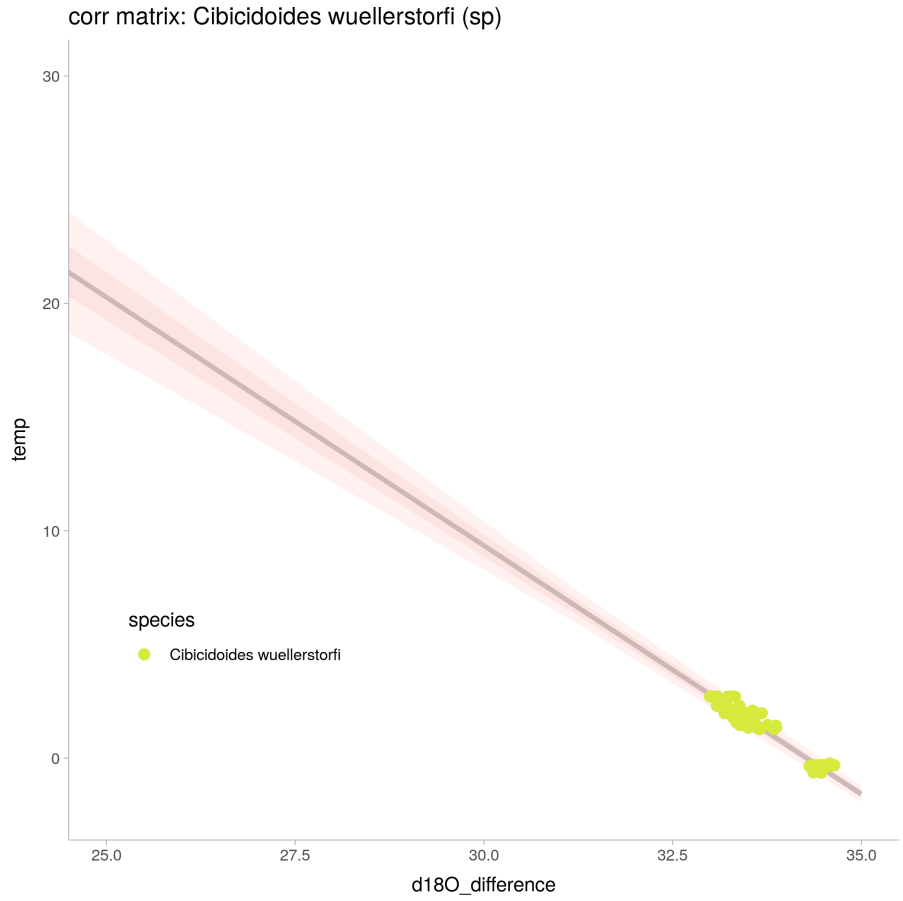

Figure 90: linear regression: corr matrix for Cibicidoides wuellerstorfi (sp).

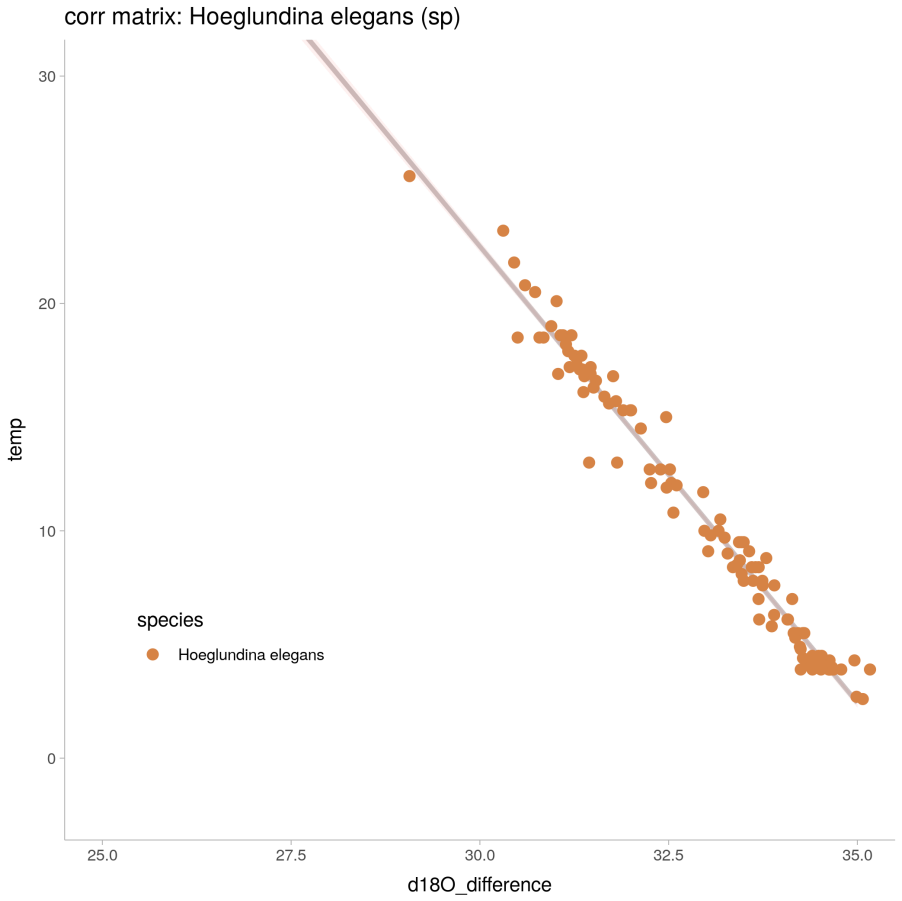

Figure 91: linear regression: corr matrix for Hoeglundina elegans (sp).

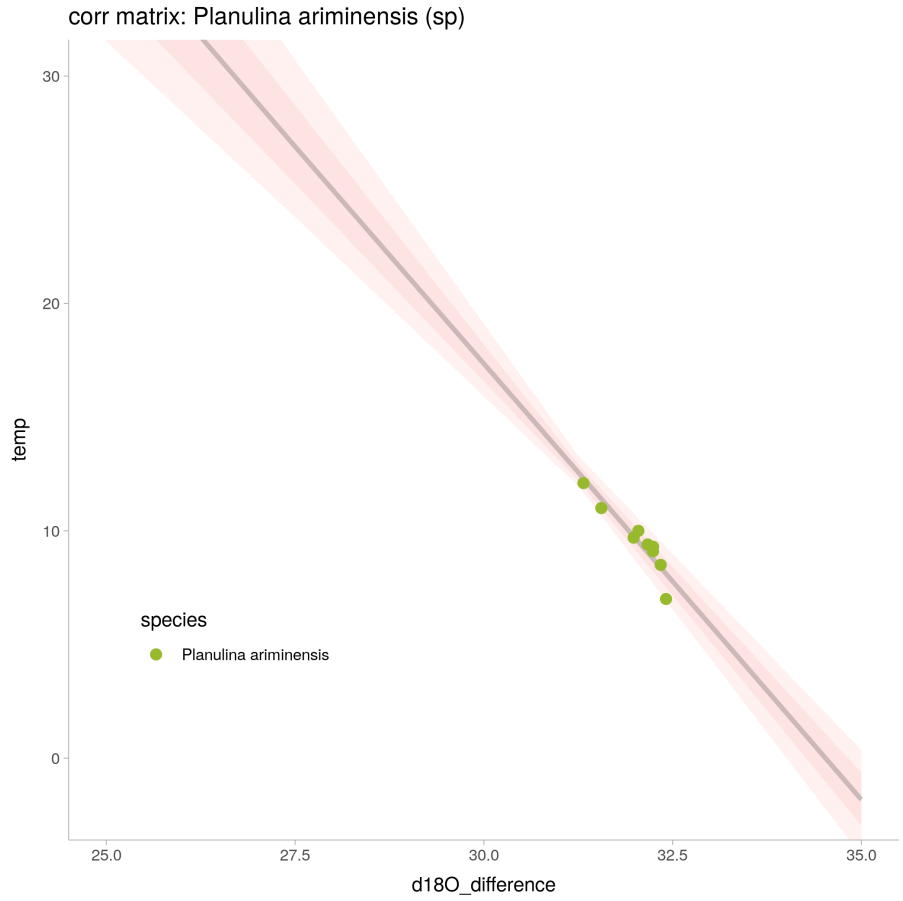

Figure 92: linear regression: corr matrix for Planulina ariminensis (sp).

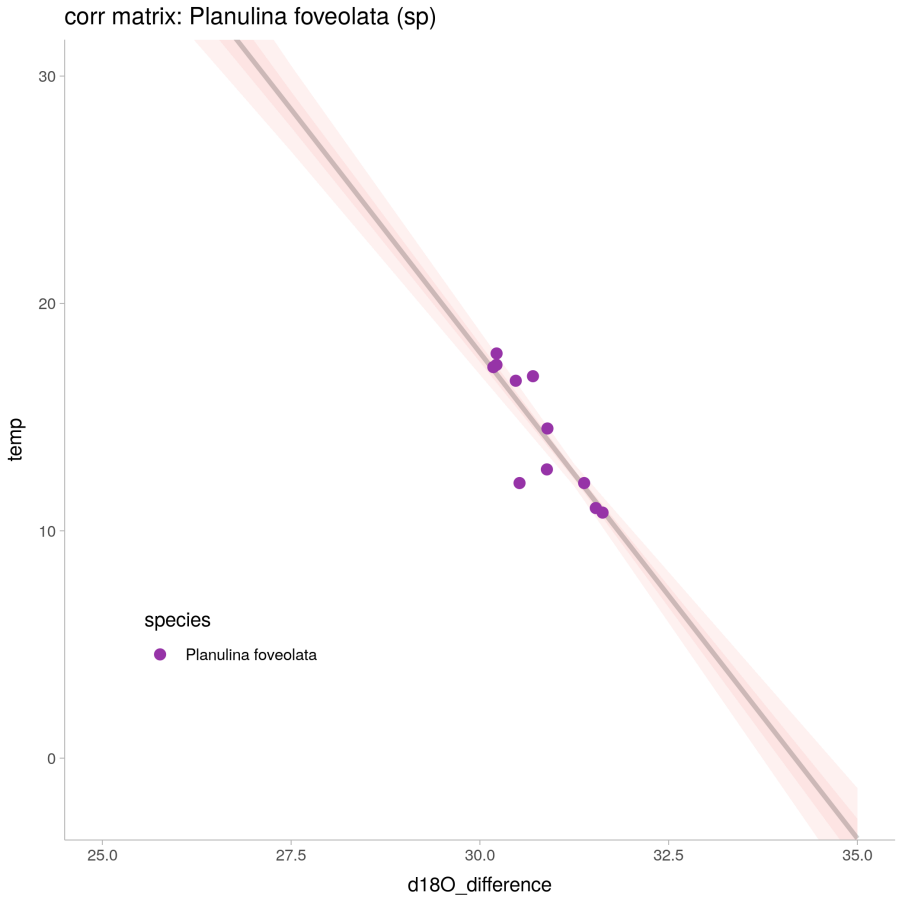

Figure 93: linear regression: corr matrix for Planulina foveolata (sp).

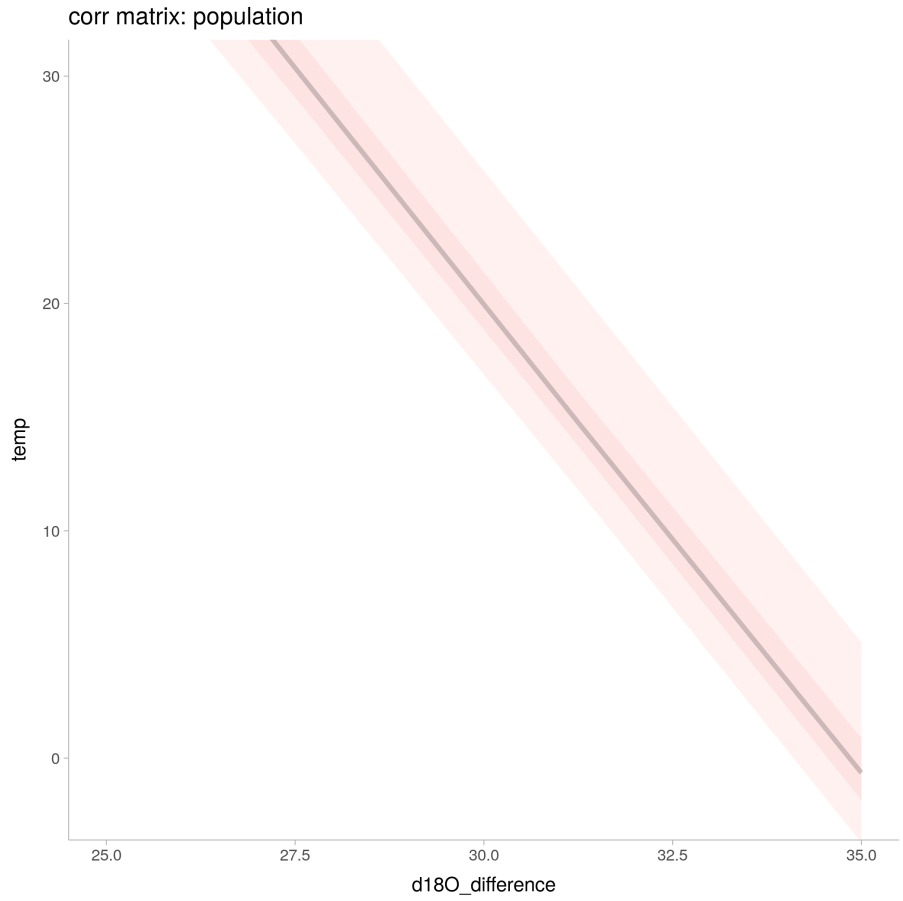

Figure 94: linear regression: corr matrix for population

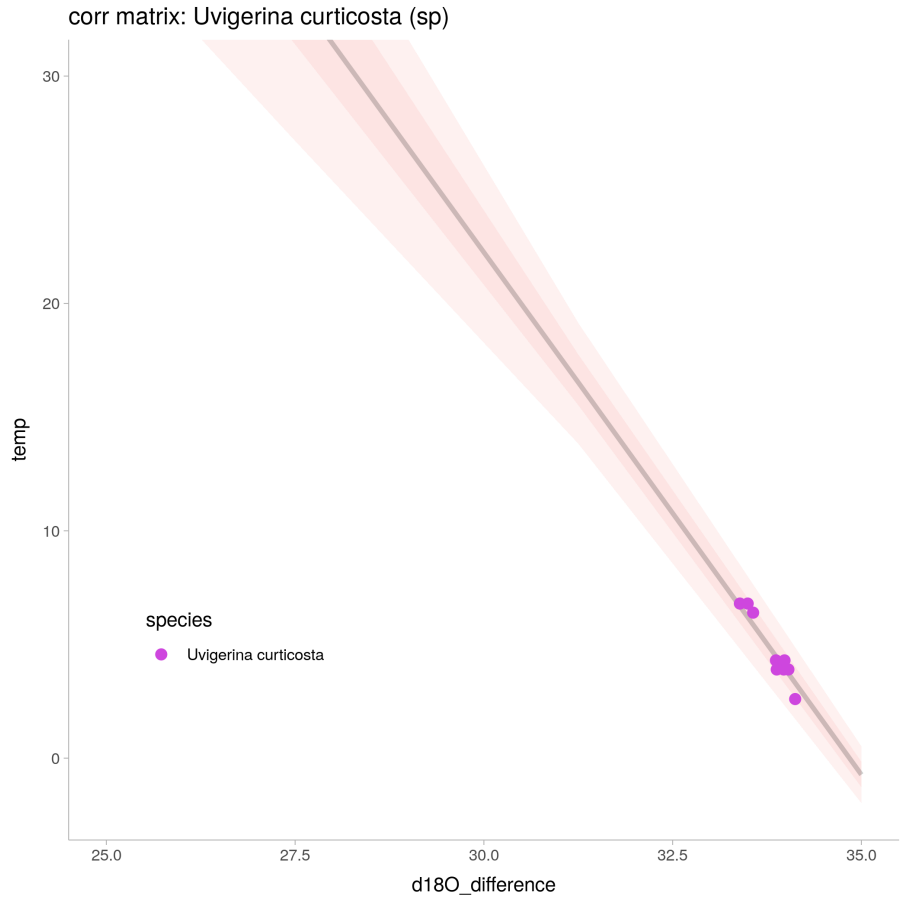

Figure 95: linear regression: corr matrix for Uvigerina curticosta (sp).

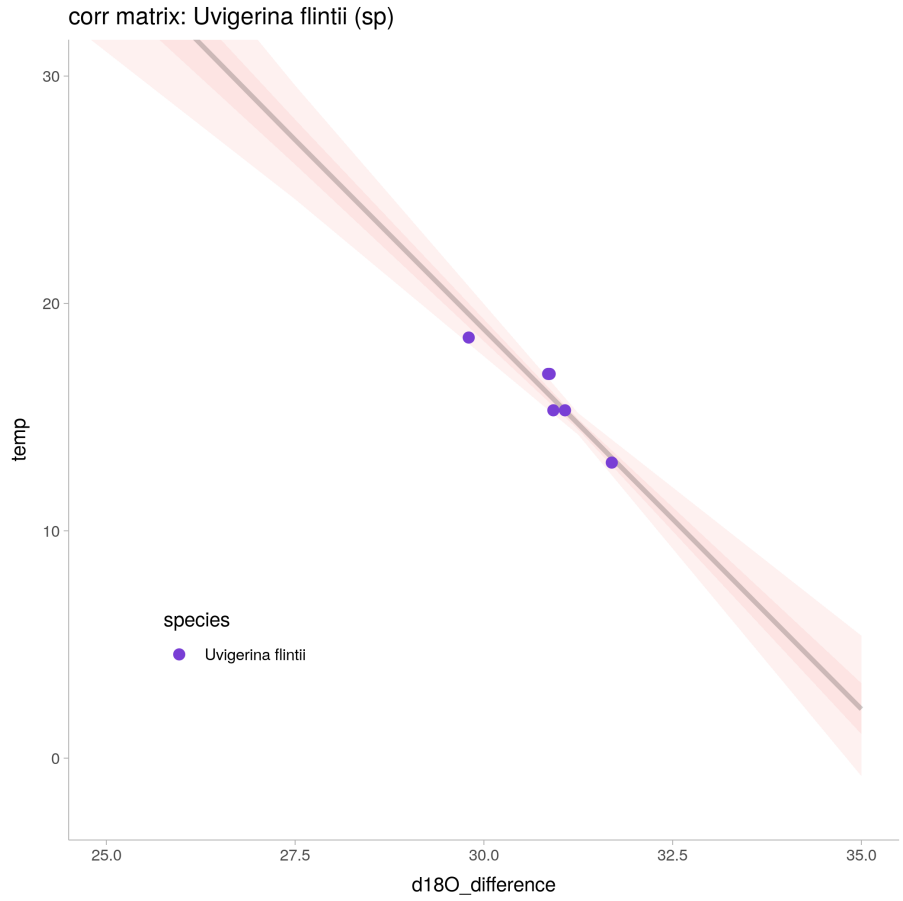

Figure 96: linear regression: corr matrix for Uvigerina flintii (sp).

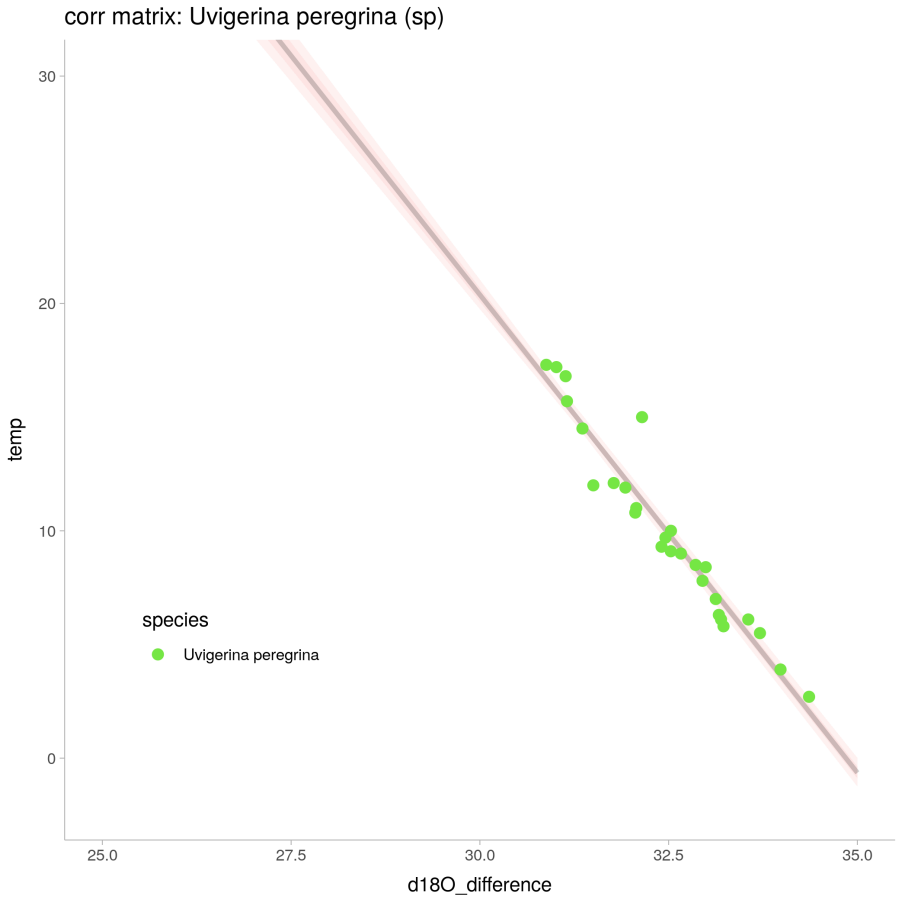

Figure 97: linear regression: corr matrix for Uvigerina peregrina (sp).

```
data {
  int N_obs;
  int N_sp;
  array[N_obs] int<lower=1, upper=N_sp> species;
  vector[N_obs] temperature;
  int K;
  matrix[N_obs, K] x;
  int prior_only;
}
parameters {
  vector[2] ab_pop;
  matrix[2, N_sp] ab_sp_std;
  cholesky_factor_corr[2] corr_sp;
  vector<lower=0>[2] sd_sp;
  real<lower=0> sigma_obs;
}
transformed parameters {
  matrix[2, N_sp] ab_sp = diag_pre_multiply(sd_sp, corr_sp) * ab_sp_std;
  for (i in 1: N_sp)
    ab_sp[, i] += ab_pop;
}
model {
  real a_pop = ab_pop[1];
  real b_pop = ab_pop[2];
  vector[N_sp] a_sp = transpose(ab_sp[1,]);
  vector[N_sp] b_sp = transpose(ab_sp[2,]);
  ab_pop ~ multi_normal([120, -4], [[50, 10],
                                    [10, 5]]');


  for (i in 1: N_sp)
    for (j in 1:2)
      ab_sp_std[j, i] ~ std_normal();
  sd_sp[1] ~ normal(0, 50);
  sd_sp[2] ~ normal(0, 5);
  corr_sp ~ lkj_corr_cholesky(2);
  sigma_obs ~ normal(0, 1);

  if (! prior_only) {
    temperature ~ normal(a_sp[species] + b_sp[species] .* (x[,1] - x[,2]), sigma_obs);
  }
}

generated quantities {
  vector[N_obs] log_lik;
  vector[N_obs] temperature_sim;
  vector[N_obs] mu;

```

```
49    corr_matrix[2] Sigma;
50    Sigma = multiply_lower_tri_self_transpose(corr_sp);
51
52    for (i in 1:N_obs){
53      mu[i] = ab_sp[1, species[i]] + ab_sp[2, species[i]] * (x[,1] - x[,2])[i];
54
55      log_lik[i] = normal_lpdf(temperature[i] | mu[i],sigma_obs);
56      temperature_sim[i] = normal_rng(mu[i],sigma_obs);
57    }
58  }
```

Listing 13: The stan code for the model hierarchical model with a correlation matrix (foraminifera data)

8.11 partial pooling model with a third level of composition (foraminifera data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
ab_pop[1]	1.00	4497	122.03	5.07	0.08	112.00	131.78
ab_pop[2]	1.00	1630	-3.64	0.44	0.01	-4.61	-2.80
sigma_obs	1.00	5476	0.76	0.03	0.00	0.70	0.83
log-posterior	1.00	794	-77.03	5.10	0.18	-87.81	-68.05
sd_sp[1]	1.00	1620	27.95	7.88	0.20	16.62	46.45
sd_sp[2]	1.00	1642	0.80	0.23	0.01	0.47	1.35
corr_co[1,1]			1.00	0.00		1.00	1.00
corr_co[2,1]	1.00	4644	-0.11	0.60	0.01	-0.98	0.96
corr_co[1,2]			0.00	0.00		0.00	0.00
corr_co[2,2]	1.00	1905	0.76	0.25	0.01	0.17	1.00
ab_sp[1,1]	1.00	4452	147.24	5.03	0.08	137.43	157.21
ab_sp[2,1]	1.00	4463	-4.29	0.16	0.00	-4.61	-3.98
ab_sp[1,2]	1.00	3833	74.07	5.43	0.09	63.63	84.40
ab_sp[2,2]	1.00	3837	-2.16	0.16	0.00	-2.47	-1.85
ab_sp[1,3]	1.00	3815	142.84	1.64	0.03	139.65	146.11
ab_sp[2,3]	1.00	3817	-4.01	0.05	0.00	-4.11	-3.91
ab_sp[1,4]	1.00	3704	124.29	15.33	0.25	93.63	155.61
ab_sp[2,4]	1.00	3701	-3.58	0.48	0.01	-4.56	-2.63
ab_sp[1,5]	1.00	3670	140.13	12.08	0.20	118.09	165.36
ab_sp[2,5]	1.00	3671	-4.08	0.39	0.01	-4.90	-3.36
ab_sp[1,6]	1.00	4299	150.46	21.30	0.32	112.13	194.89
ab_sp[2,6]	1.00	4296	-4.31	0.63	0.01	-5.63	-3.17
ab_sp[1,7]	1.00	3659	118.66	13.73	0.23	88.93	143.32
ab_sp[2,7]	1.00	3649	-3.33	0.44	0.01	-4.12	-2.37
ab_sp[1,8]	1.00	4685	146.07	5.09	0.07	135.94	155.96
ab_sp[2,8]	1.00	4656	-4.19	0.16	0.00	-4.49	-3.88
ab_co[1,1]	1.00	1422	126.13	8.52	0.23	110.60	144.90
ab_co[2,1]	1.00	1467	-3.61	0.25	0.01	-4.16	-3.15
ab_co[1,2]	1.00	1884	126.88	11.94	0.28	106.66	155.27
ab_co[2,2]	1.00	1963	-3.71	0.35	0.01	-4.55	-3.12

Table 16: hierarchical model with correlated parameters

model run time: 950.48

```
Computed from 4000 by 295 log-likelihood matrix

         Estimate   SE
elpd_loo   -347.2 24.3
p_loo        18.7  4.1
looic       694.4 48.6
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     293   99.3%   412
 (0.5, 0.7]   (ok)         1    0.3%   167
   (0.7, 1]   (bad)        1    0.3%   134
   (1, Inf)   (very bad)   0    0.0%   <NA>
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are too high.
See help('pareto-k-diagnostic') for details.
```

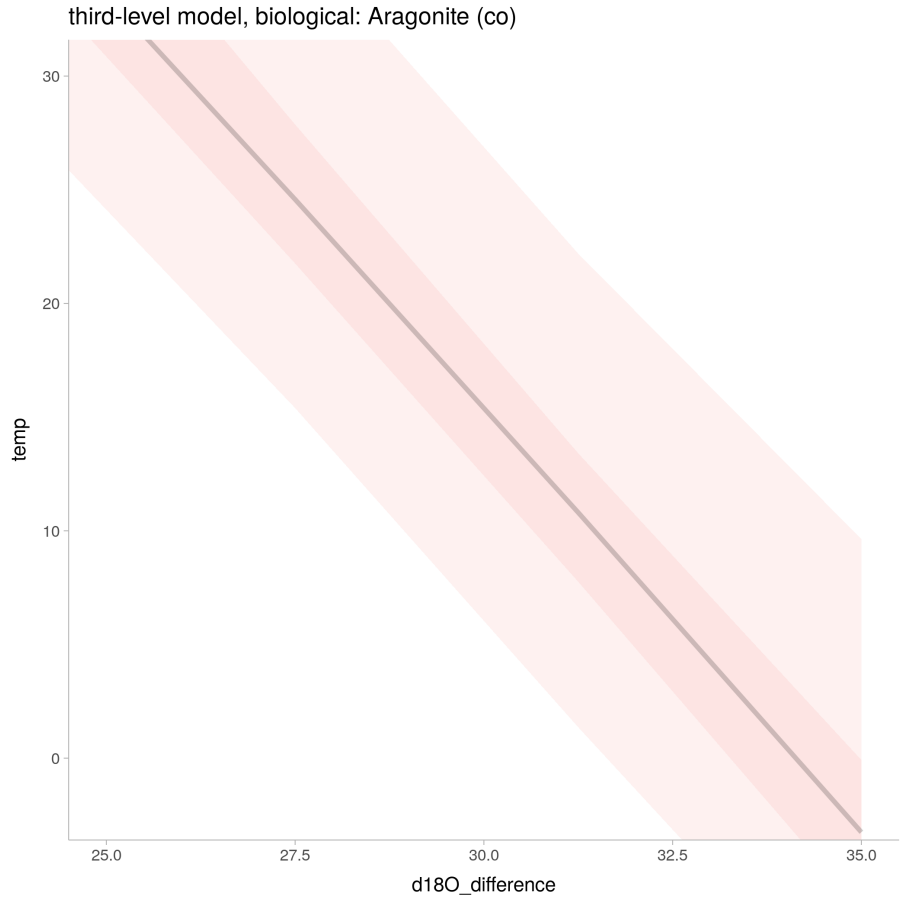

Figure 98: linear regression: three level model (data: fora) for Aragonite (co). (datapoints excluded)

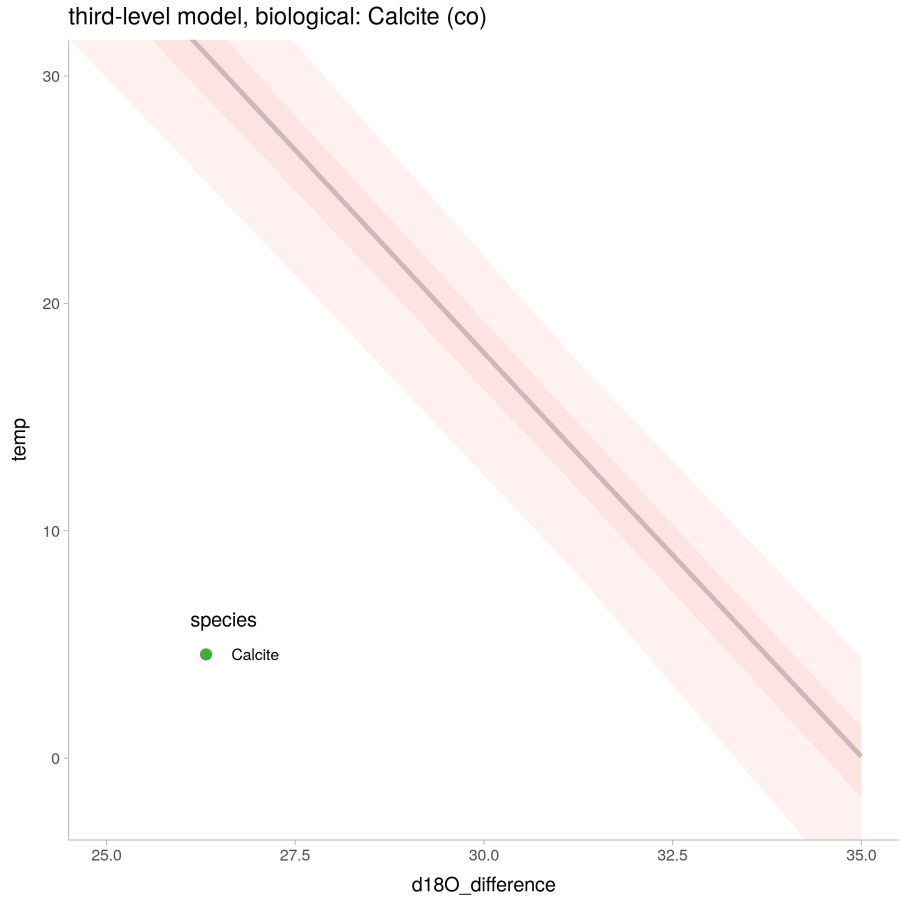

Figure 99: linear regression: three level model (data: fora) for Calcite (co). (datapoints excluded)

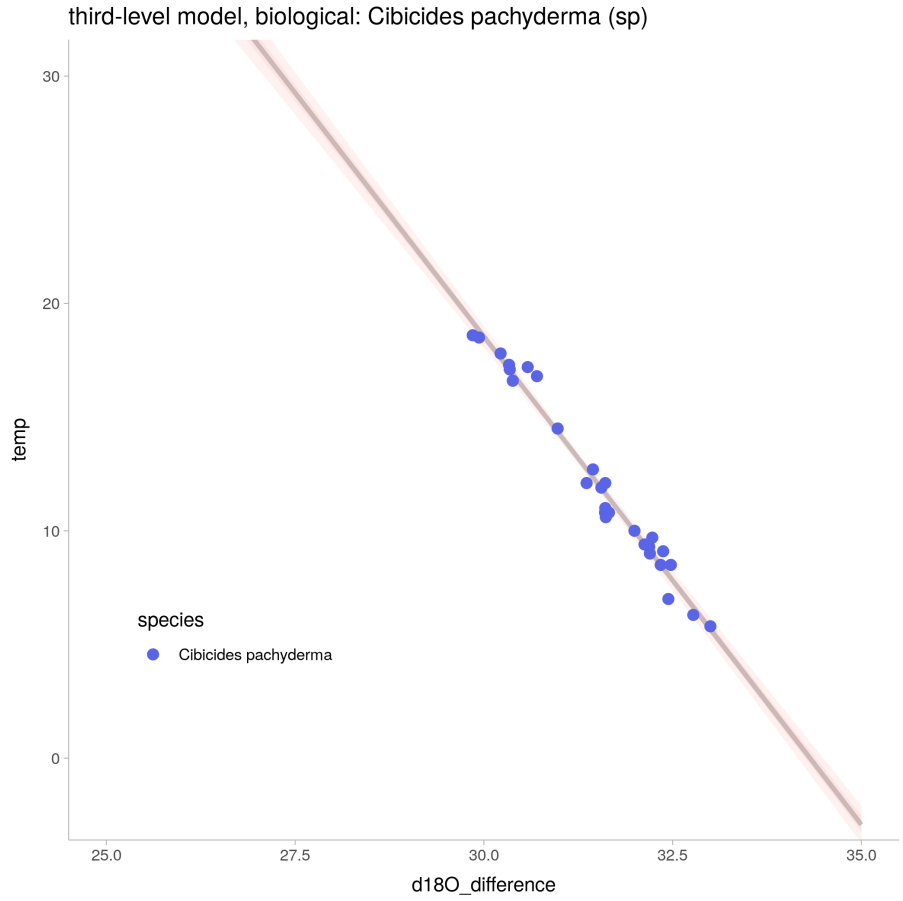

Figure 100: linear regression: three level model (data: fora) for Cibicides pachyderma (sp).

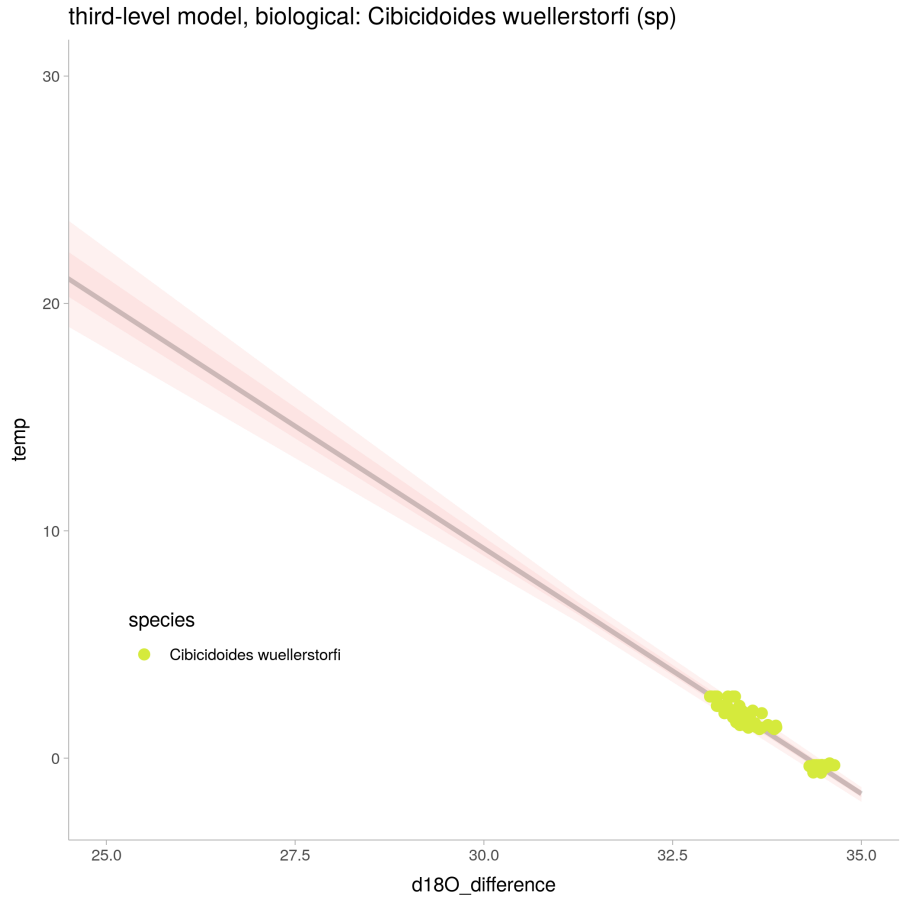

Figure 101: linear regression: three level model (data: fora) for Cibicidoides wuellerstorfi (sp).

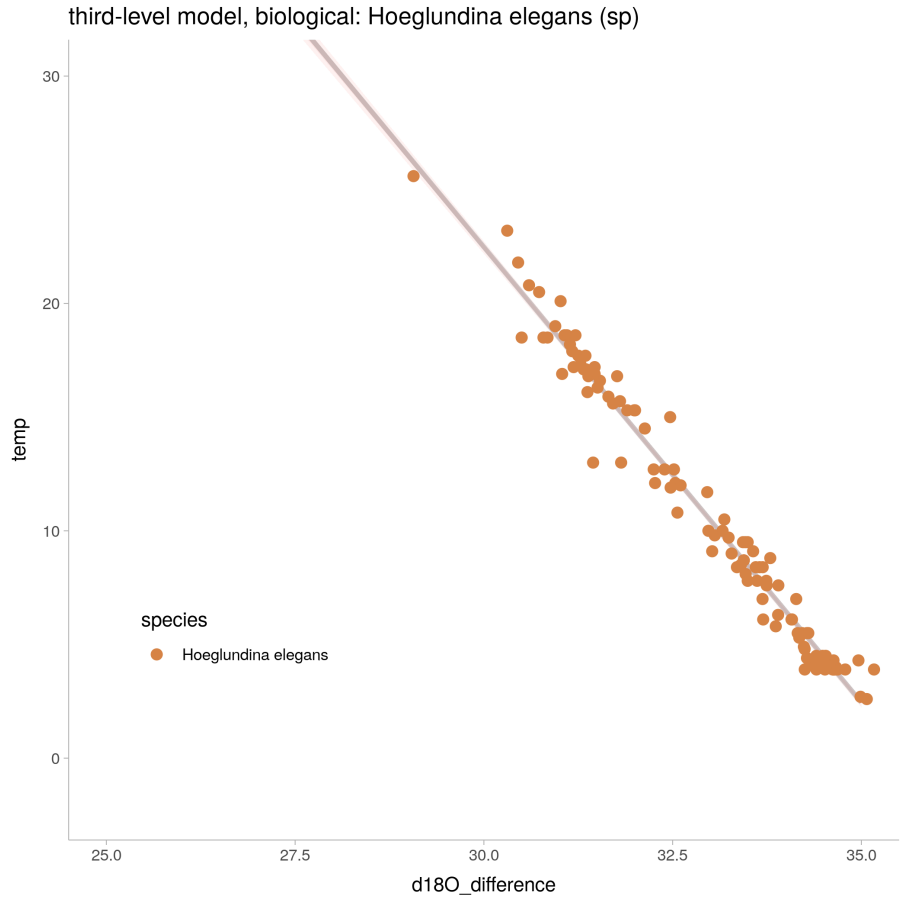

Figure 102: linear regression: three level model (data: fora) for Hoeglundina elegans (sp).

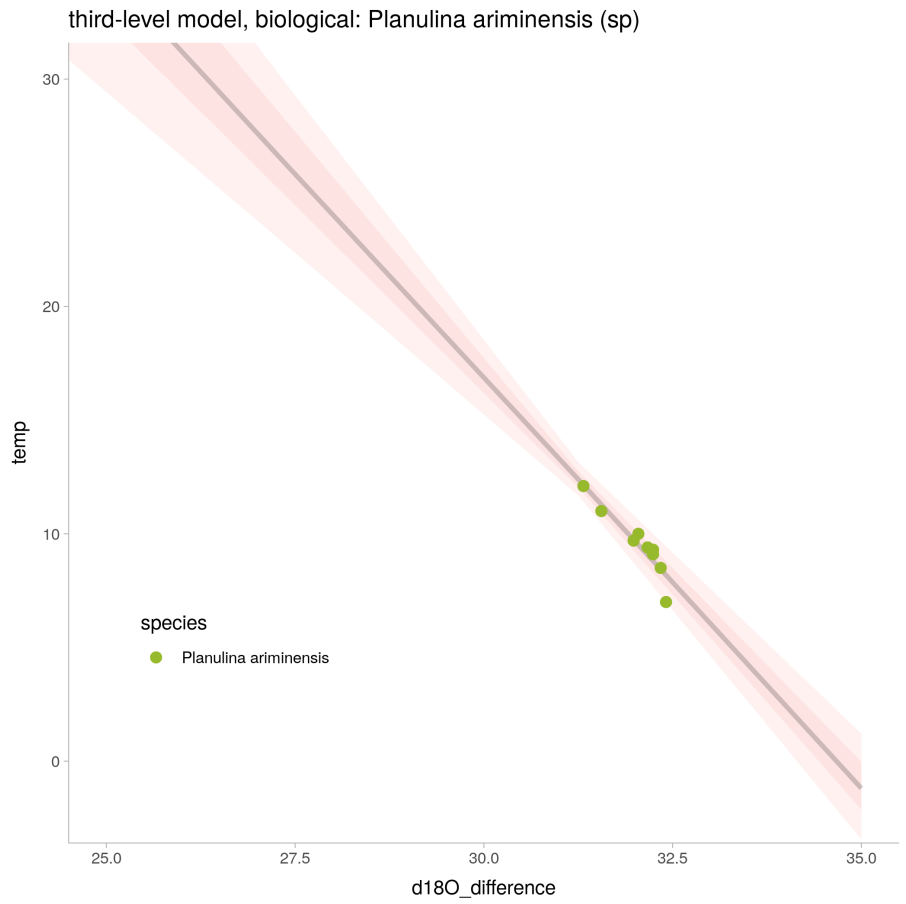

Figure 103: linear regression: three level model (data: fora) for Planulina ariminensis (sp).

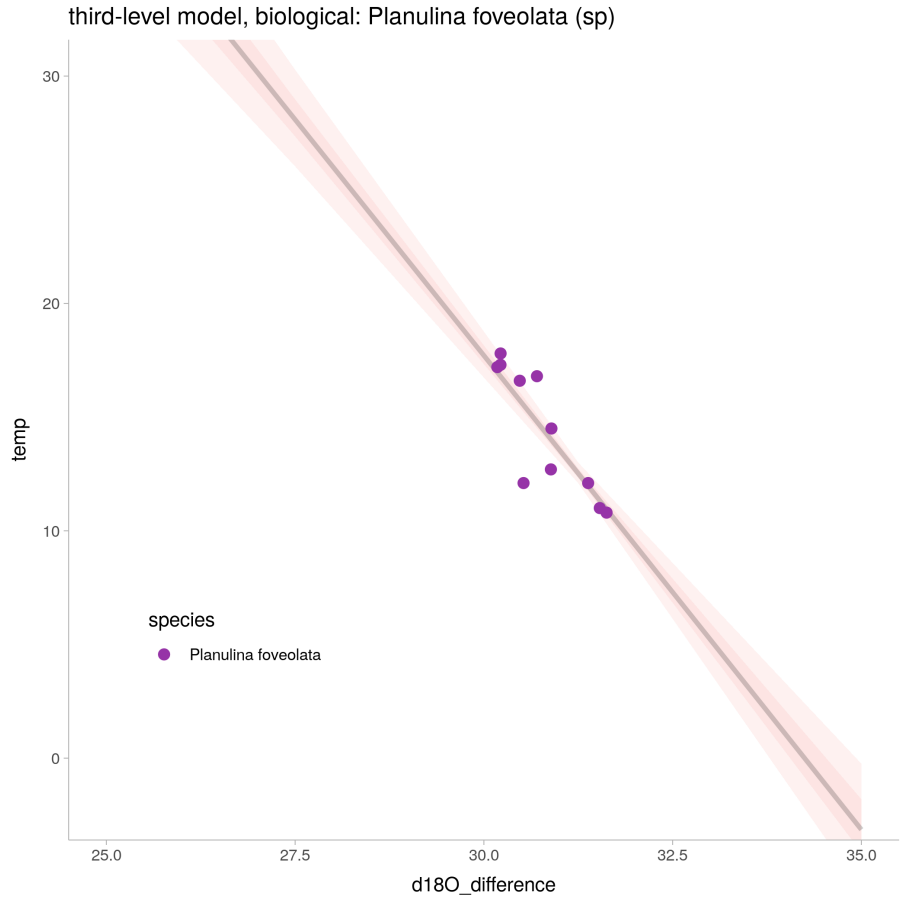

Figure 104: linear regression: three level model (data: fora) for Planulina foveolata (sp).

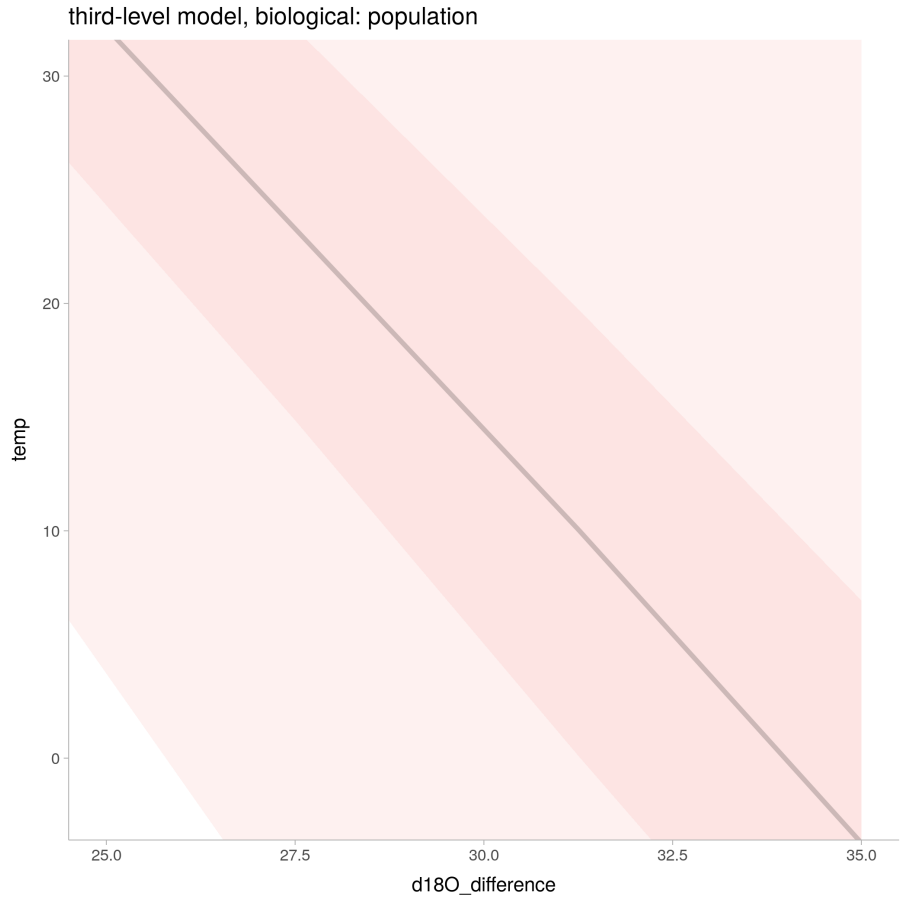

Figure 105: linear regression: three level model (data: fora) for population

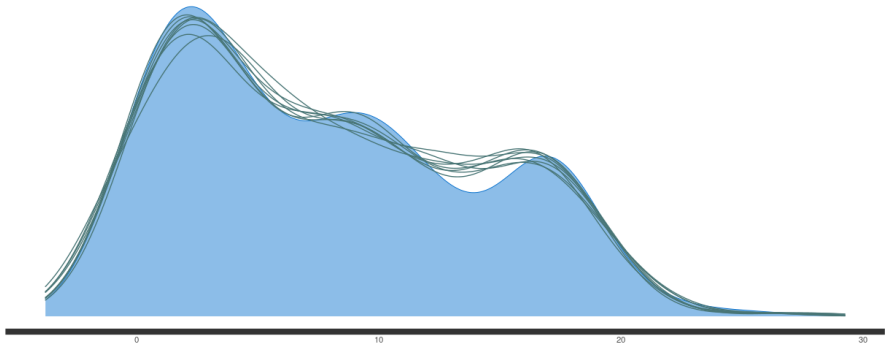

Figure 106: posterior predictive check for the three level model with foraminifera data

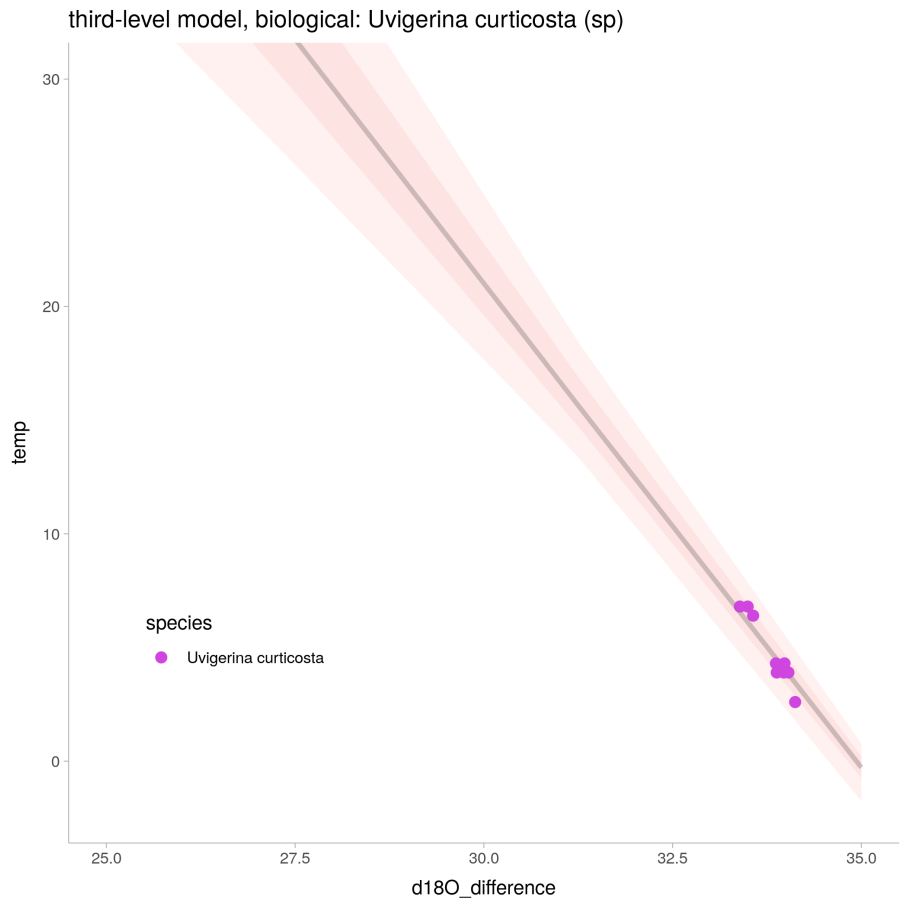

Figure 107: linear regression: three level model (data: fora) for Uvigerina curticosta (sp).

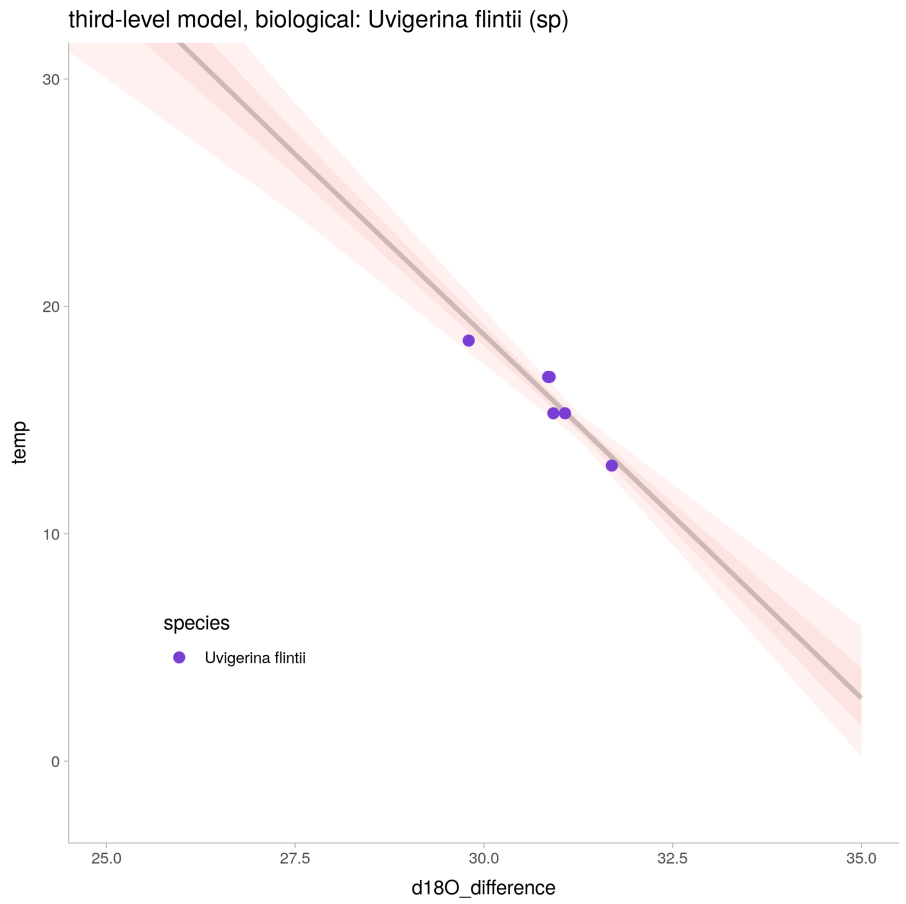

Figure 108: linear regression: three level model (data: fora) for Uvigerina flintii (sp).

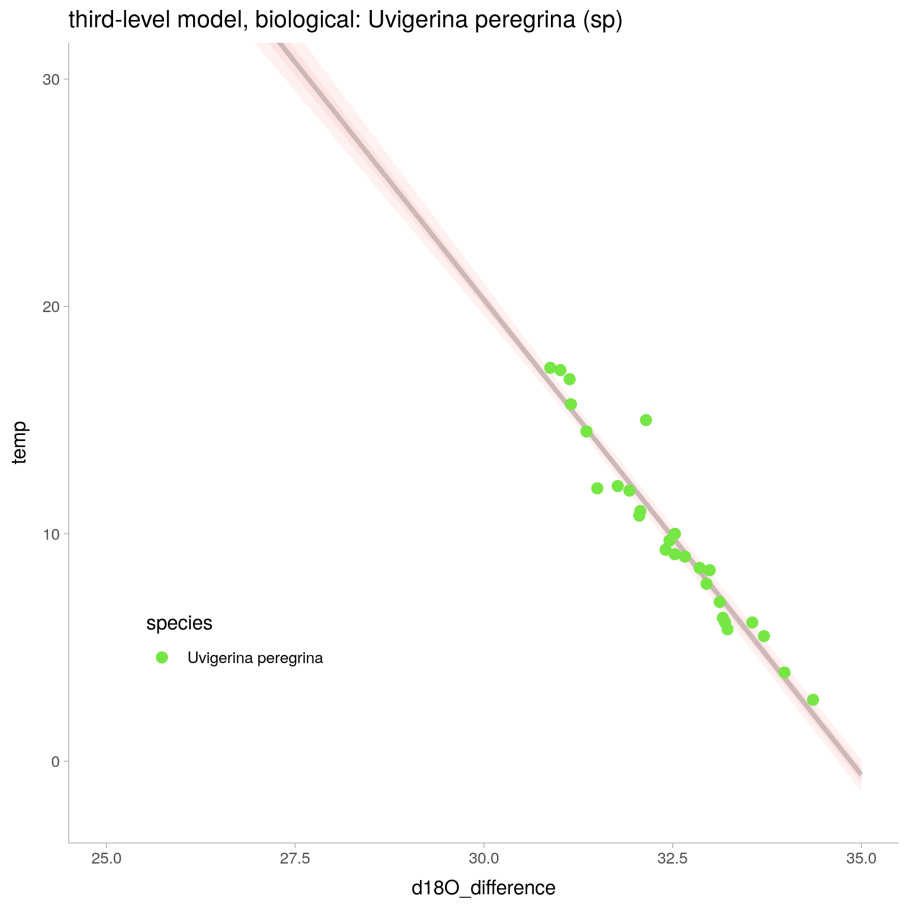

Figure 109: linear regression: three level model (data: fora) for Uvigerina peregrina (sp).

```
data {
  int N_obs;
  int N_co;
  int N_sp;
  array[N_sp] int<lower=1, upper=N_co> comps;
  array[N_obs] int<lower=1, upper=N_sp> species;
  vector[N_obs] temperature;
  int K;
  matrix[N_obs, K] x;
  int prior_only;
  real<lower=0> sigma_T_lab;
}

parameters {
  vector[2] ab_pop;
  matrix[2, N_co] ab_co_std;
  cholesky_factor_corr[2] corr_co;
  vector<lower=0>[2] sd_co;
  matrix[2, N_sp] ab_sp_std;
  cholesky_factor_corr[2] corr_sp;
  vector<lower=0>[2] sd_sp;
  real<lower=0> sigma_obs;
}
transformed parameters {
  matrix[2, N_co] ab_co = diag_pre_multiply(sd_co, corr_co) * ab_co_std;
  matrix[2, N_sp] ab_sp = diag_pre_multiply(sd_sp, corr_sp) * ab_sp_std;
  for (i in 1: N_co)
    ab_co[, i] += ab_pop;
  for (i in 1: N_sp)
    ab_sp[, i] += ab_co[, comps[i]];

  vector[N_obs] individual_sigmas;
  for (i in 1:N_obs) {
    if (species[i] == 13) {
      individual_sigmas[i] = sigma_T_lab;
    } else {
      individual_sigmas[i] = sigma_obs;
    }
  }
}
model {
  real a_pop = ab_pop[1];
  real b_pop = ab_pop[2];
  vector[N_co] a_co = transpose(ab_co[1,]);
  vector[N_co] b_co = transpose(ab_co[2,]);
  vector[N_sp] a_sp = transpose(ab_sp[1,]);
  vector[N_sp] b_sp = transpose(ab_sp[2,]);

```

```
  for (i in 1: N_sp)
    for (j in 1:2)
      ab_sp_std[j, i] ~ std_normal();

  ab_pop ~ multi_normal([120, -4], [[50, 10],
                                    [10, 5]]');
  for (i in 1: N_co)
    for (j in 1:2)
      ab_co_std[j, i] ~ std_normal();

  sd_co[1] ~ normal(0, 50);
  sd_co[2] ~ normal(0, 5);
  corr_co ~ lkj_corr_cholesky(2);

  sd_sp[1] ~ normal(0, 50);
  sd_sp[2] ~ normal(0, 5);
  corr_sp ~ lkj_corr_cholesky(2);
  sigma_obs ~ normal(0, 1);

  if (!prior_only) {
    temperature ~ normal(a_sp[species] + b_sp[species] .* (x[,1] - x[,2]), individual_sigmas);
  }
}

generated quantities {
  vector[N_obs] log_lik;
  vector[N_obs] temperature_sim;
  vector[N_obs] mu;

  corr_matrix[2] Sigma_sp;
  Sigma_sp = multiply_lower_tri_self_transpose(corr_sp);

  corr_matrix[2] Sigma_co;
  Sigma_co = multiply_lower_tri_self_transpose(corr_co);

  for (i in 1:N_obs){
    mu[i] = ab_sp[1,species][i] + ab_sp[2,species][i] * (x[,1] - x[,2])[i];
    log_lik[i] = normal_lpdf(temperature[i] | mu[i], individual_sigmas[i]);
    temperature_sim[i] = normal_rng(mu[i], individual_sigmas[i]);
  }
}
```

Listing 14: The stan code for the partial pooling model with a third level of composition (foraminifera data)

8.12 partial pooling model with a third level of composition (biological data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%

ab_pop[1]	1.00	3487	124.81	5.19	0.09	114.56	134.91
ab_pop[2]	1.00	1588	-3.62	0.31	0.01	-4.35	-3.08
sigma_obs	1.00	5167	0.82	0.03	0.00	0.75	0.89
log-posterior	1.01	687	-119.64	6.81	0.26	-134.20	-107.31
corr_co[1,1]			1.00	0.00		1.00	1.00
corr_co[2,1]	1.00	4054	-0.25	0.60	0.01	-0.99	0.91
corr_co[1,2]			0.00	0.00		0.00	0.00
corr_co[2,2]	1.00	2318	0.71	0.26	0.01	0.12	1.00
sd_sp[1]	1.00	1422	25.21	6.29	0.17	15.57	39.91
sd_sp[2]	1.00	1423	0.73	0.19	0.01	0.44	1.18
ab_co[1,1]	1.00	1343	130.60	8.25	0.23	116.49	149.43
ab_co[2,1]	1.00	1326	-3.67	0.25	0.01	-4.25	-3.24
ab_co[1,2]	1.00	1209	131.79	9.14	0.26	117.04	153.17
ab_co[2,2]	1.00	1186	-3.70	0.28	0.01	-4.34	-3.24
ab_co[1,3]	1.00	1963	130.45	11.80	0.27	109.35	158.19
ab_co[2,3]	1.00	2002	-3.62	0.35	0.01	-4.44	-2.97
ab_sp[1,1]	1.00	3553	133.94	25.97	0.44	81.09	185.26
ab_sp[2,1]	1.00	3546	-3.74	0.78	0.01	-5.28	-2.16
ab_sp[1,2]	1.00	2505	135.15	20.24	0.40	93.67	172.79
ab_sp[2,2]	1.00	2503	-3.76	0.64	0.01	-4.94	-2.45
ab_sp[1,3]	1.00	3062	137.36	20.41	0.37	95.26	177.09
ab_sp[2,3]	1.00	3065	-3.84	0.66	0.01	-5.11	-2.49
ab_sp[1,4]	1.00	6048	142.83	11.45	0.15	120.58	165.17
ab_sp[2,4]	1.00	6051	-4.00	0.35	0.00	-4.69	-3.32
ab_sp[1,5]	1.00	4765	138.42	26.24	0.38	88.46	192.59
ab_sp[2,5]	1.00	4759	-3.84	0.78	0.01	-5.45	-2.35
ab_sp[1,6]	1.00	4432	145.70	5.38	0.08	135.25	156.40
ab_sp[2,6]	1.00	4427	-4.24	0.17	0.00	-4.58	-3.91
ab_sp[1,7]	1.00	3527	74.78	5.89	0.10	63.36	86.33
ab_sp[2,7]	1.00	3531	-2.18	0.17	0.00	-2.52	-1.84
ab_sp[1,8]	1.00	4298	133.54	25.41	0.39	82.97	182.89
ab_sp[2,8]	1.00	4309	-3.74	0.74	0.01	-5.17	-2.27
ab_sp[1,9]	1.00	2265	149.27	24.31	0.51	101.05	198.75
ab_sp[2,9]	1.00	2252	-4.10	0.77	0.02	-5.66	-2.57
ab_sp[1,10]	1.00	3986	142.81	1.80	0.03	139.33	146.36
ab_sp[2,10]	1.00	3978	-4.01	0.05	0.00	-4.12	-3.91
ab_sp[1,11]	1.00	3355	134.29	26.98	0.47	81.16	187.65
ab_sp[2,11]	1.00	3354	-3.72	0.80	0.01	-5.31	-2.13
ab_sp[1,12]	1.00	2849	140.64	24.12	0.45	96.46	189.88
ab_sp[2,12]	1.00	2843	-3.90	0.75	0.01	-5.42	-2.51
ab_sp[1,13]	1.00	4582	121.84	17.44	0.26	87.90	157.07
ab_sp[2,13]	1.00	4574	-3.50	0.54	0.01	-4.61	-2.45
ab_sp[1,14]	1.00	3912	137.51	12.10	0.19	114.65	161.89
ab_sp[2,14]	1.00	3894	-4.00	0.39	0.01	-4.79	-3.26
ab_sp[1,15]	1.00	2369	146.81	22.45	0.46	100.11	187.88
ab_sp[2,15]	1.00	2364	-4.06	0.72	0.01	-5.35	-2.57
ab_sp[1,16]	1.00	3353	149.48	17.08	0.30	114.68	181.70
ab_sp[2,16]	1.00	3329	-4.13	0.54	0.01	-5.15	-3.03
ab_sp[1,17]	1.00	3625	136.69	25.01	0.42	86.47	186.92

ab_sp[2,17]	1.00	3626	-3.81	0.76	0.01	-5.33	-2.26
ab_sp[1,18]	1.00	3985	147.25	21.46	0.34	107.00	192.02
ab_sp[2,18]	1.00	3989	-4.21	0.63	0.01	-5.53	-3.02
ab_sp[1,19]	1.00	5316	115.88	13.71	0.19	87.85	141.44
ab_sp[2,19]	1.00	5311	-3.24	0.44	0.01	-4.06	-2.32
ab_sp[1,20]	1.00	4795	145.78	5.45	0.08	134.93	156.75
ab_sp[2,20]	1.00	4800	-4.18	0.17	0.00	-4.52	-3.85

Table 17: hierarchical model with correlated parameters

model run time: 463.18

```
Computed from 4000 by 323 log-likelihood matrix

         Estimate   SE
elpd_loo   -413.6 27.6
p_loo        35.7  8.5
looic       827.1 55.3
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     308   95.4%   709
 (0.5, 0.7]   (ok)         6    1.9%   181
   (0.7, 1]   (bad)        7    2.2%   25
   (1, Inf)   (very bad)   2    0.6%   12
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are too high.
See help('pareto-k-diagnostic') for details.
```

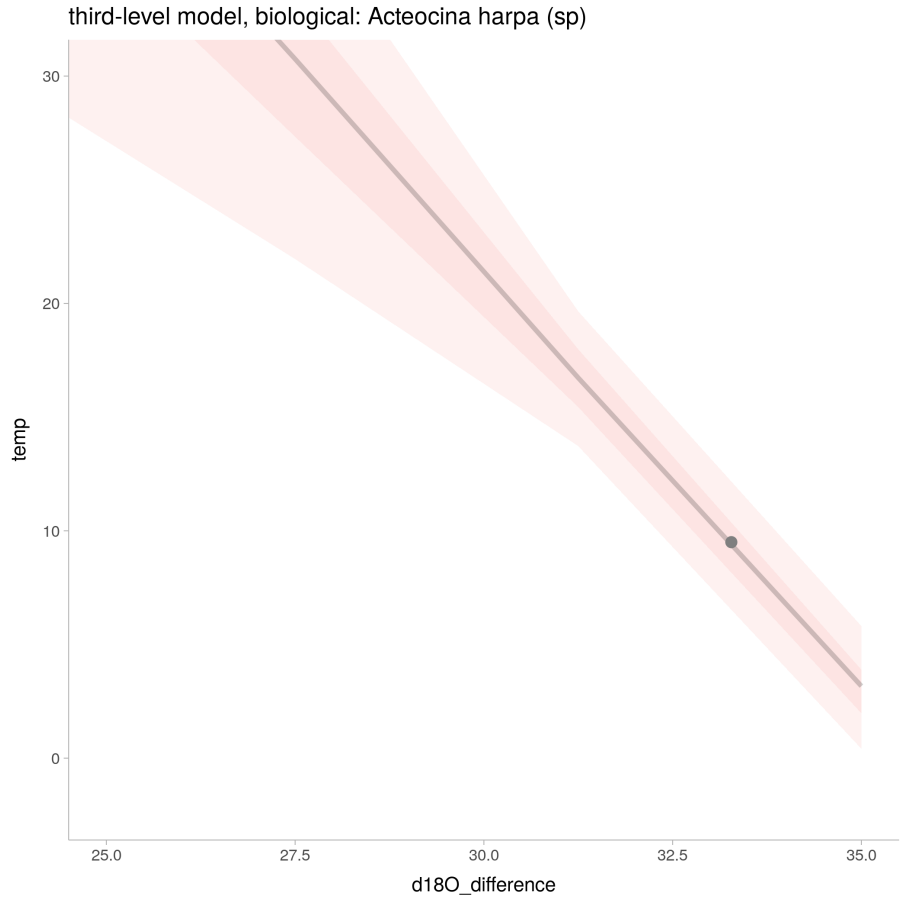

Figure 110: linear regression: three level model (data: bio) for Acteocina harpa (sp).

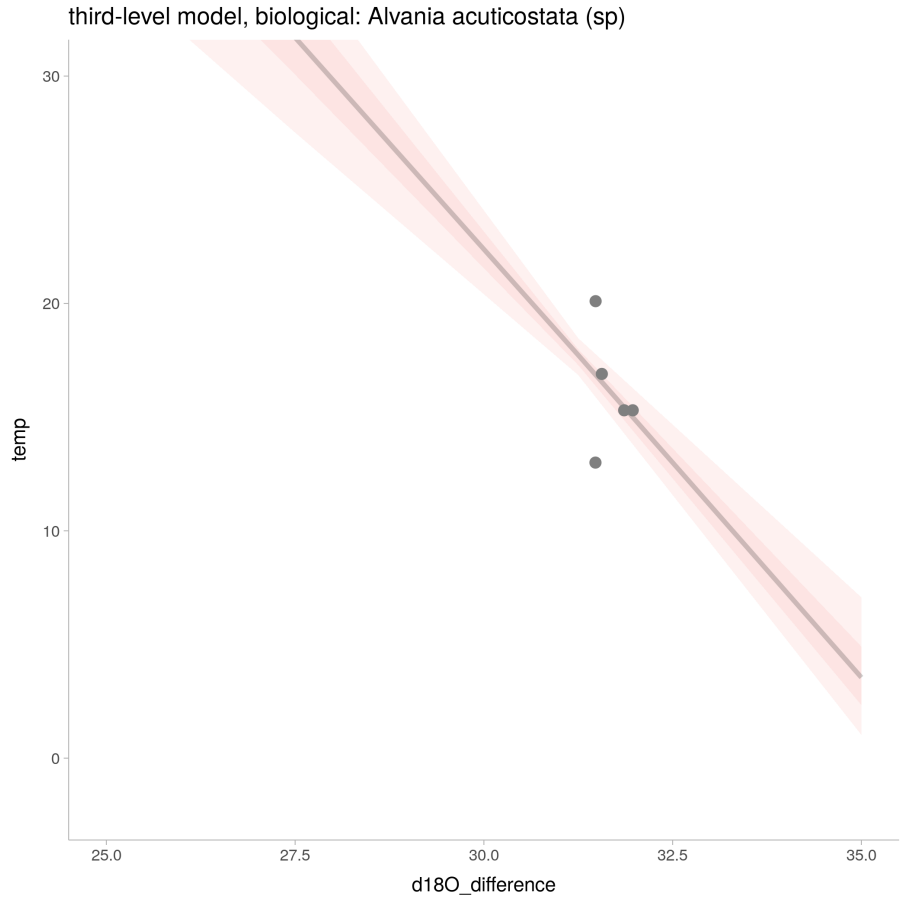

Figure 111: linear regression: three level model (data: bio) for Alvania acuticostata (sp).

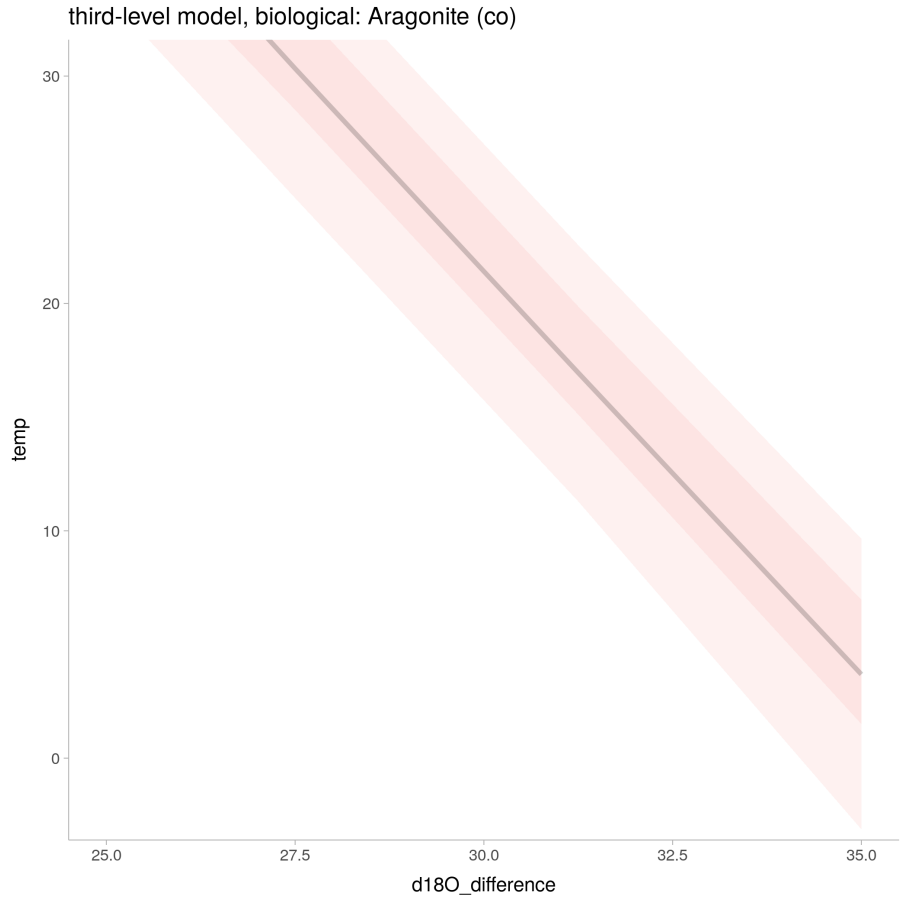

Figure 112: linear regression: three level model (data: bio) for Aragonite (co). (datapoints excluded)

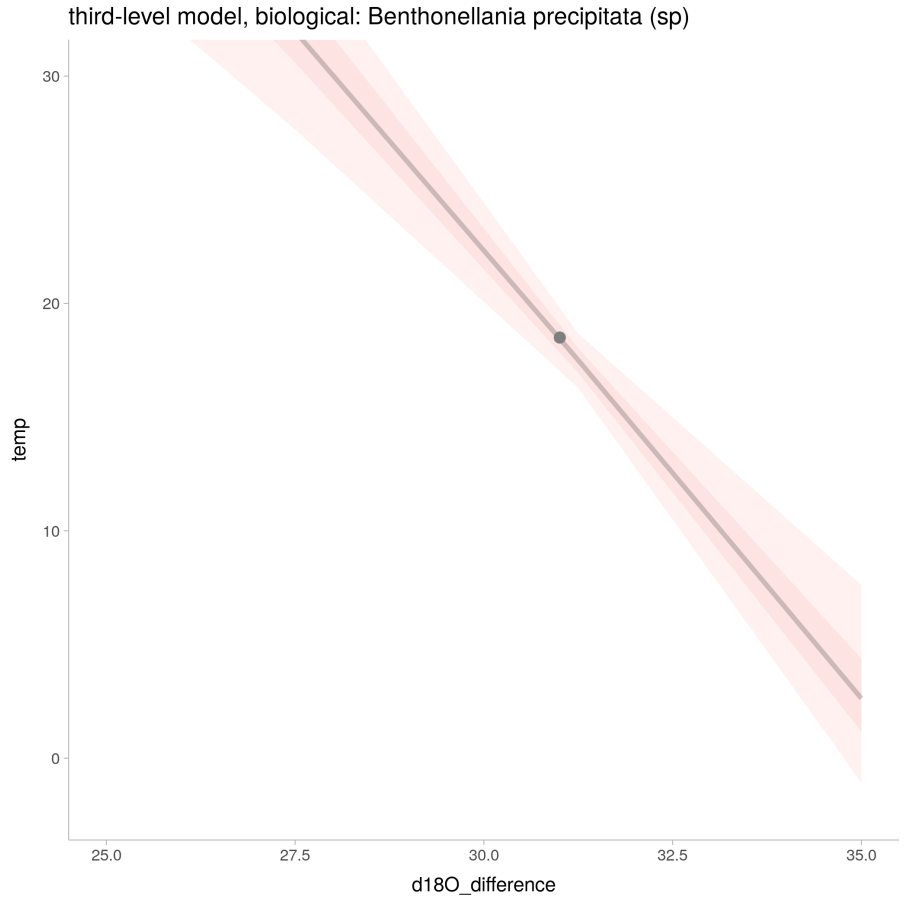

Figure 113: linear regression: three level model (data: bio) for Benthonellania precipitata (sp).

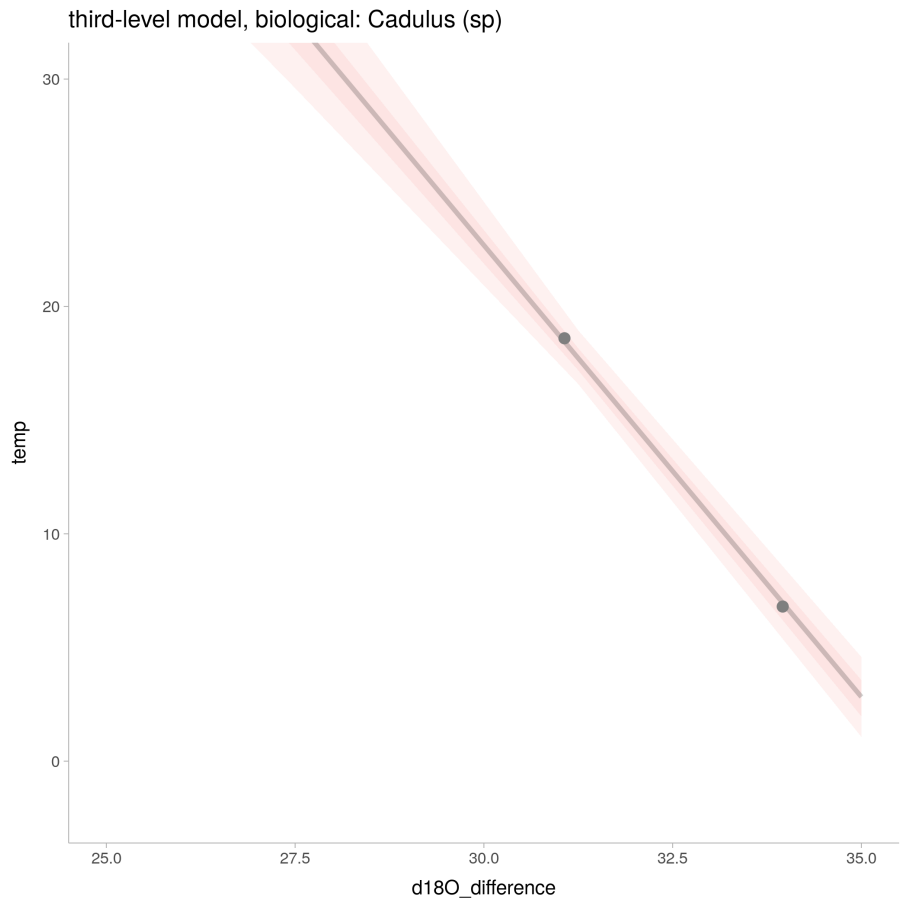

Figure 114: linear regression: three level model (data: bio) for Cadulus (sp).

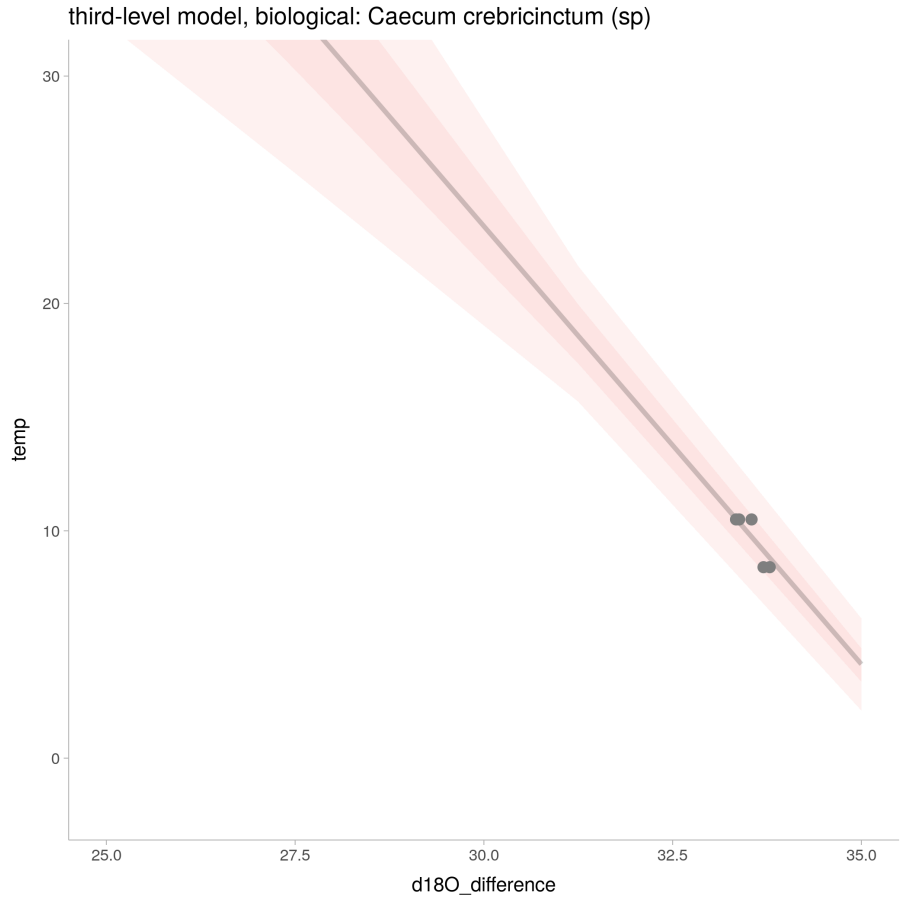

Figure 115: linear regression: three level model (data: bio) for Caecum crebricinctum (sp).

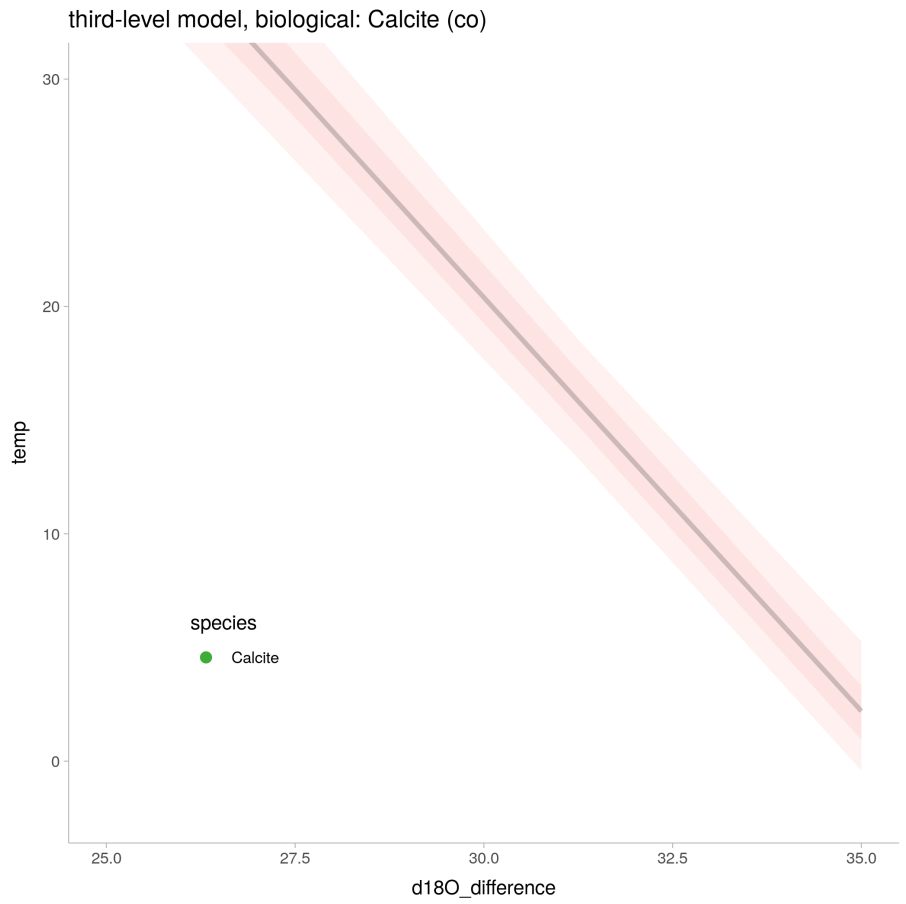

Figure 116: linear regression: three level model (data: bio) for Calcite (co).

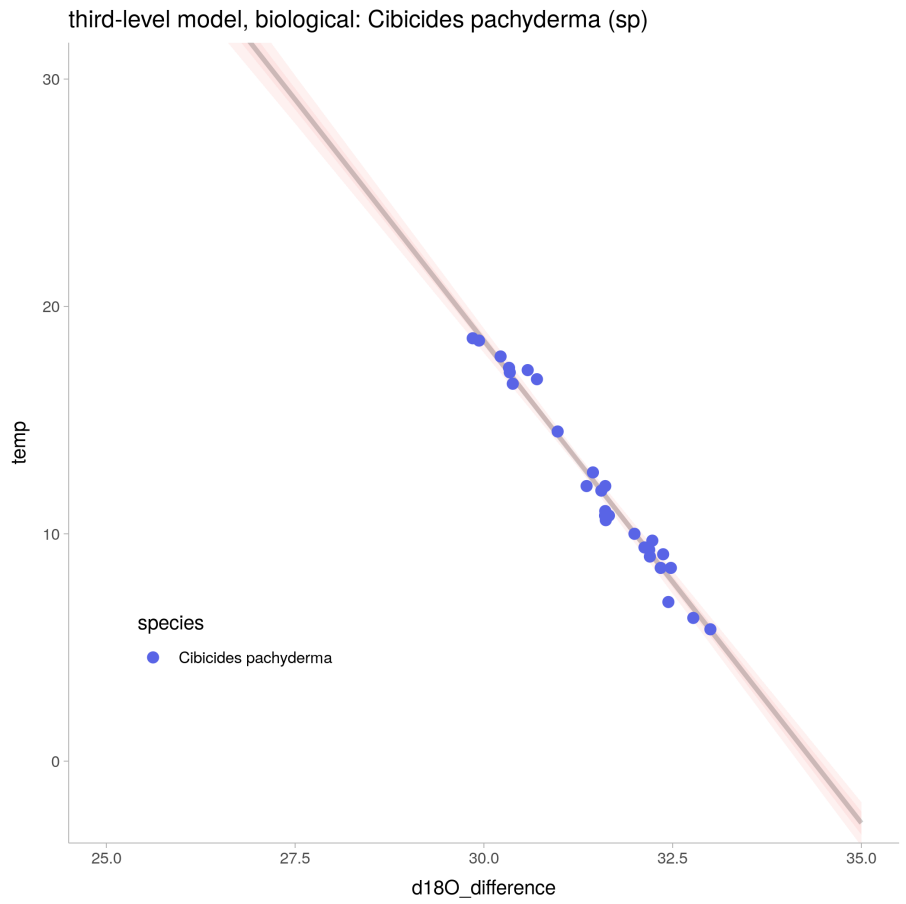

Figure 117: linear regression: three level model (data: bio) for Cibicides pachyderma (sp).

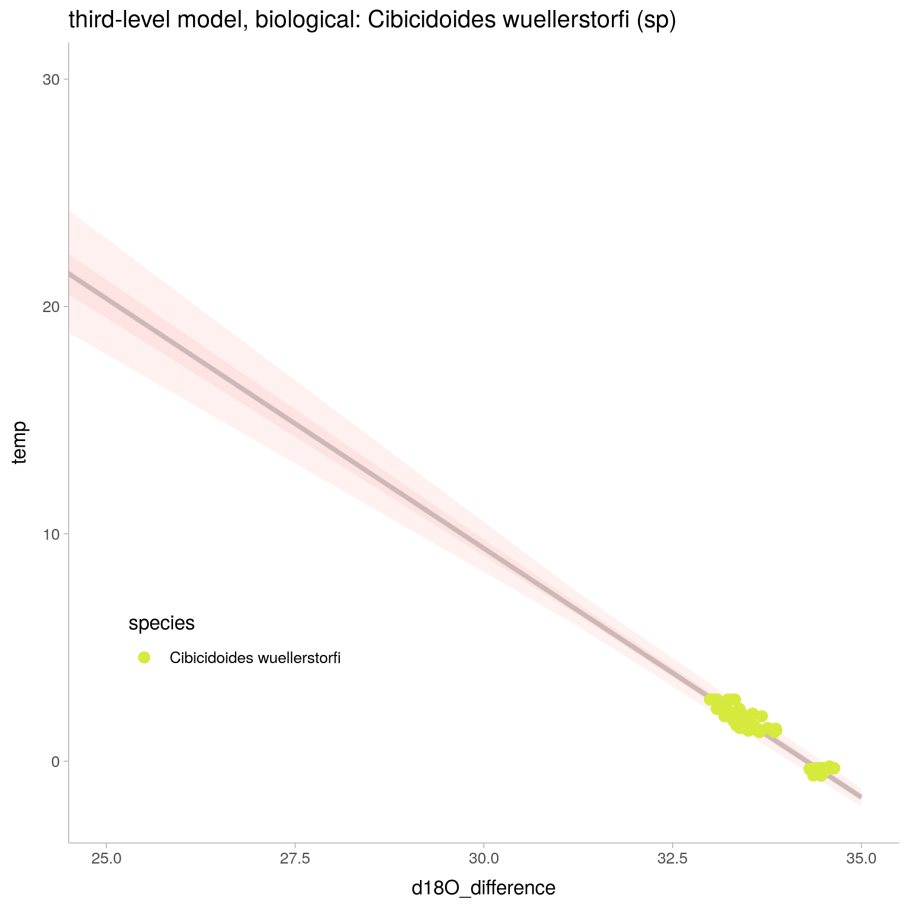

Figure 118: linear regression: three level model (data: bio) for Cibicidoides wuellerstorfi (sp).

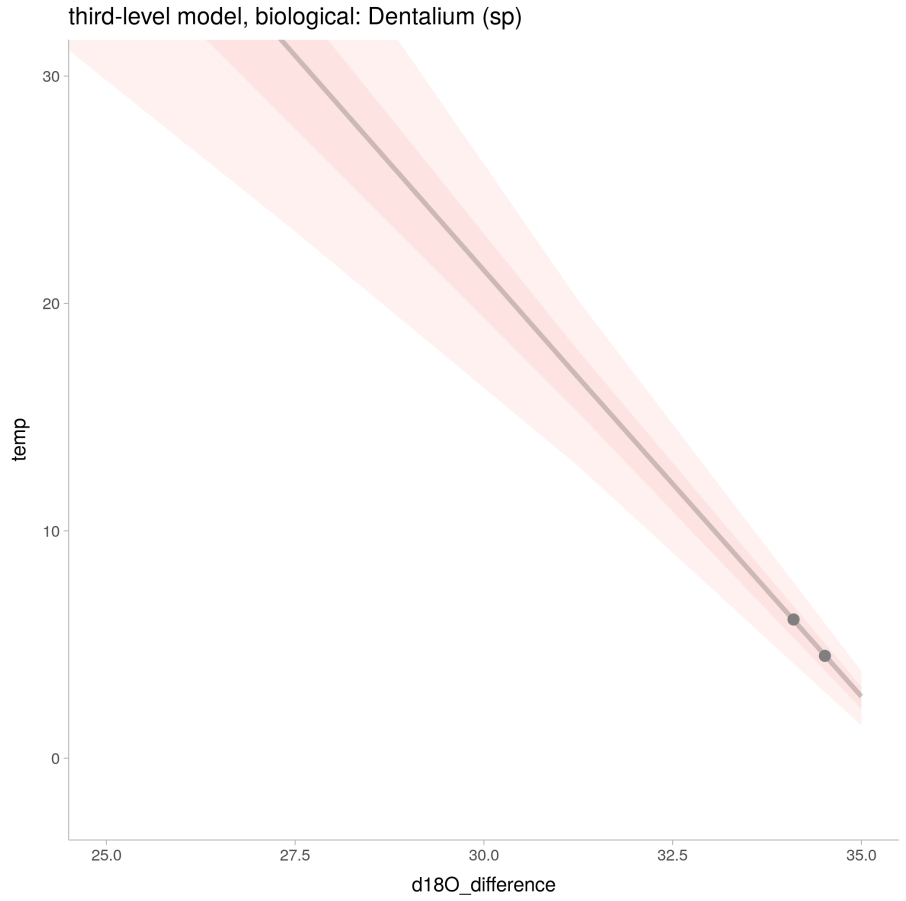

Figure 119: linear regression: three level model (data: bio) for Dentalium (sp).

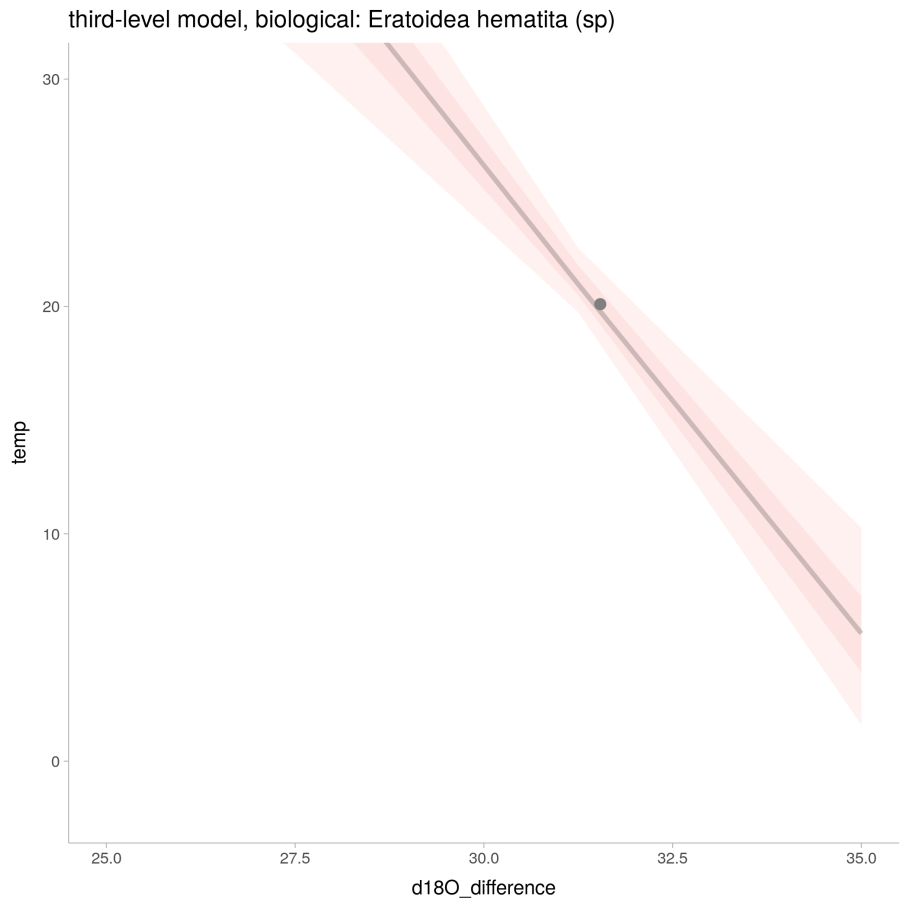

Figure 120: linear regression: three level model (data: bio) for Eratoidea hematita (sp).

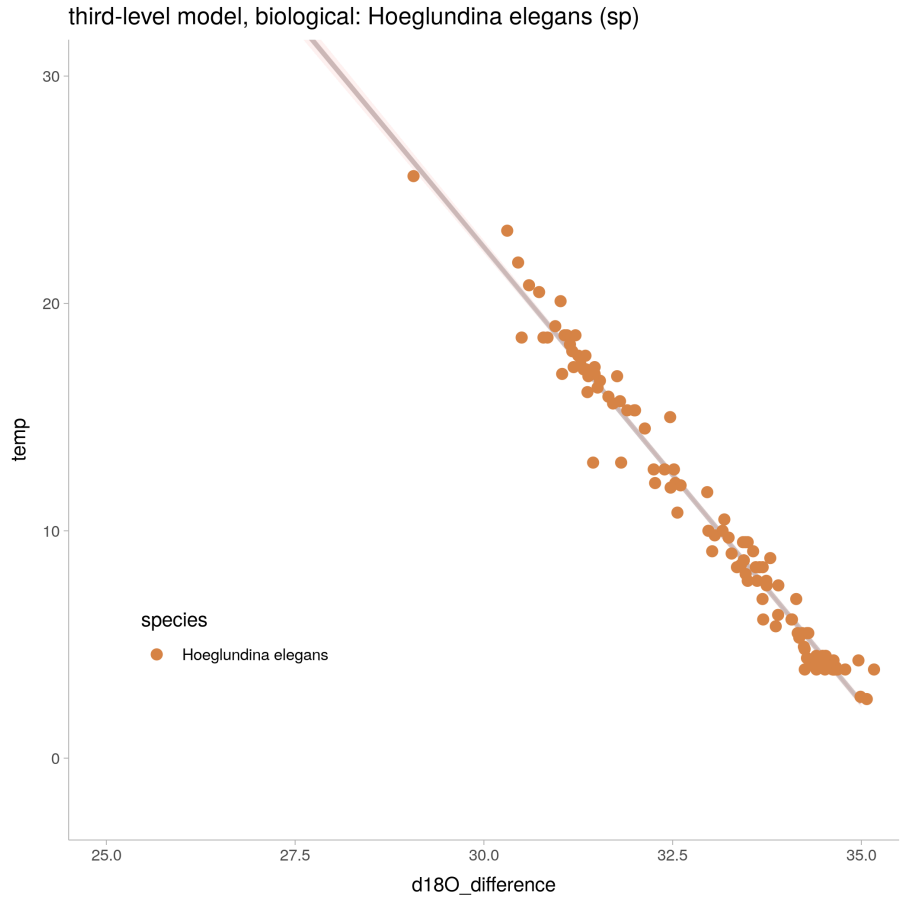

Figure 121: linear regression: three level model (data: bio) for Hoeglundina elegans (sp).

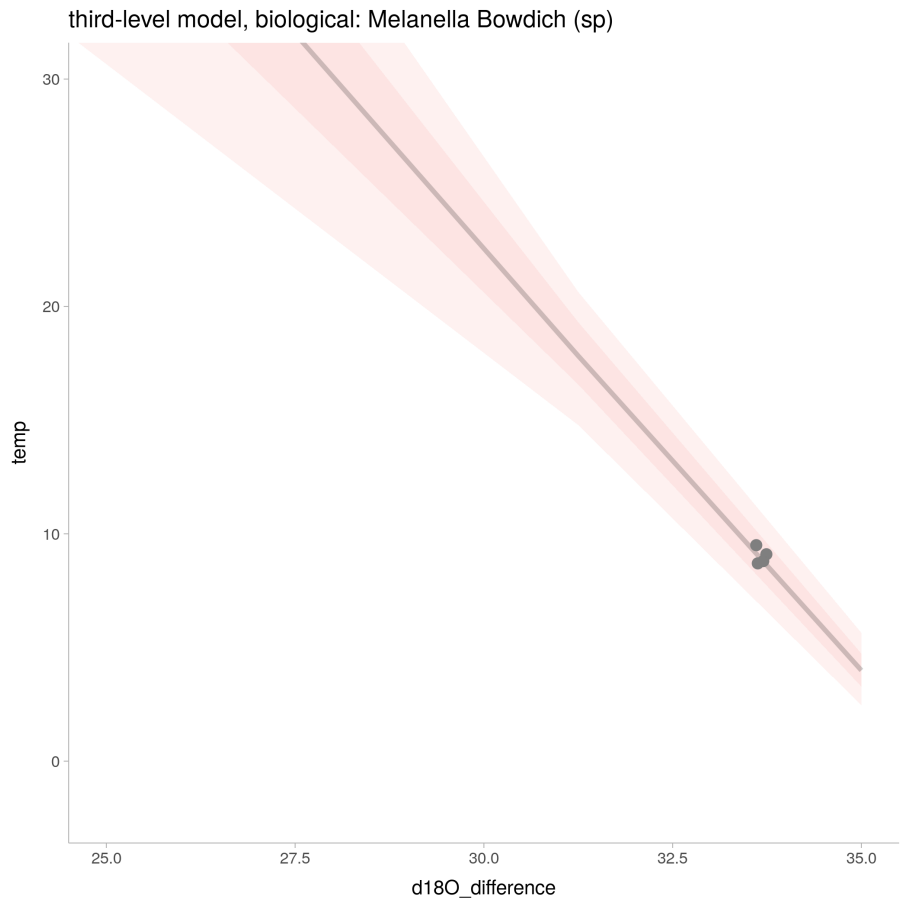

Figure 122: linear regression: three level model (data: bio) for Melanella Bowdich (sp).

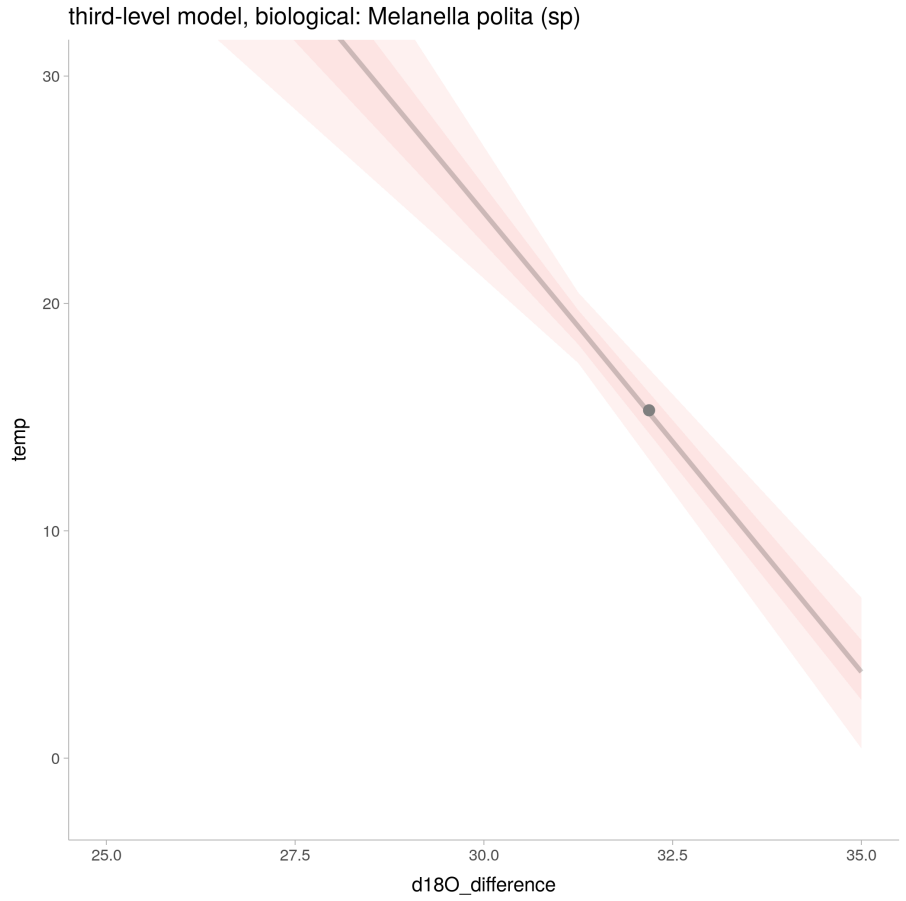

Figure 123: linear regression: three level model (data: bio) for Melanella polita (sp).

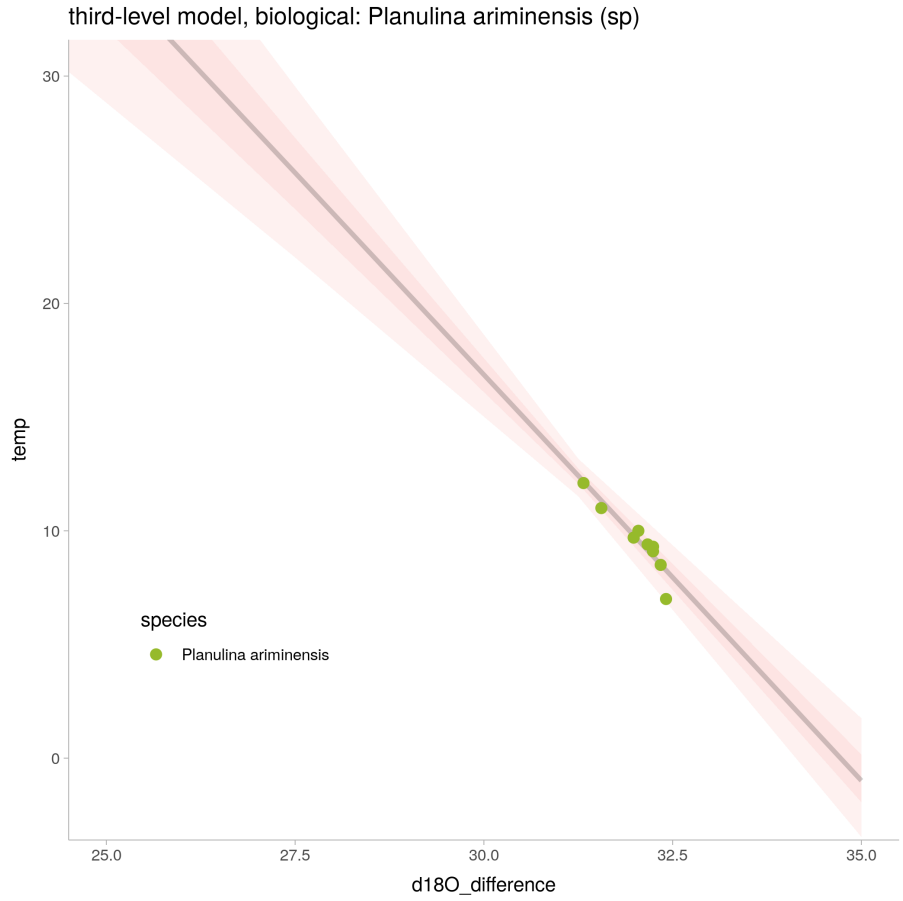

Figure 124: linear regression: three level model (data: bio) for Planulina ariminensis (sp).

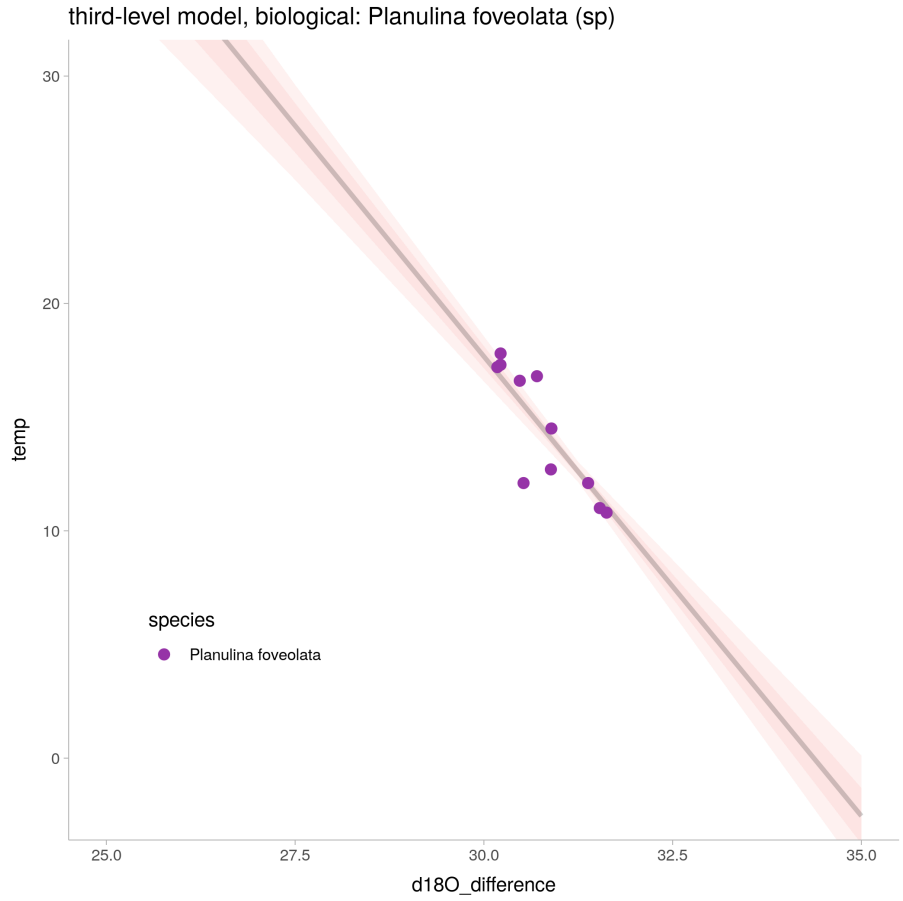

Figure 125: linear regression: three level model (data: bio) for Planulina foveolata (sp).

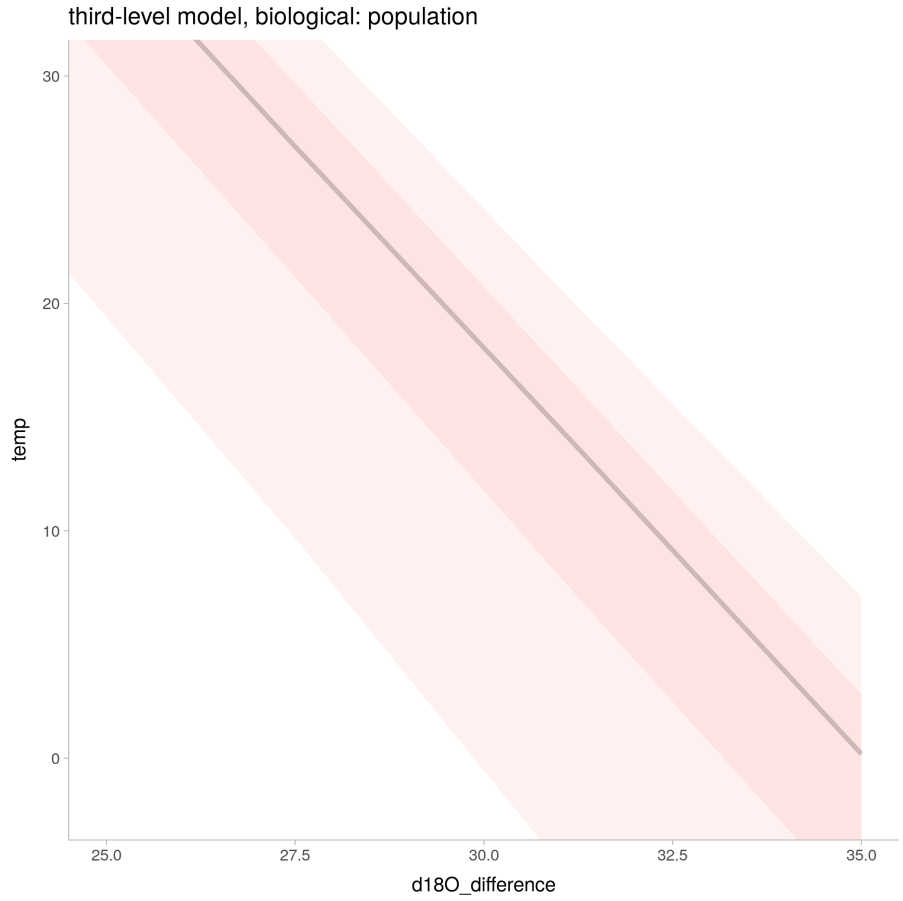

Figure 126: linear regression: three level model (data: bio) for population

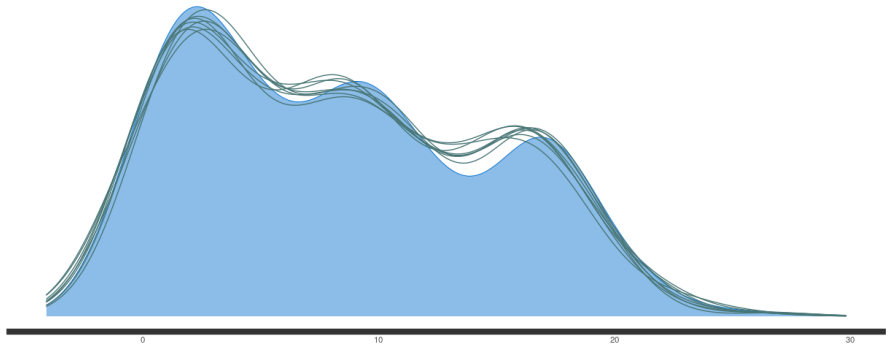

Figure 127: posterior predictive check for the three level model ran on biological data

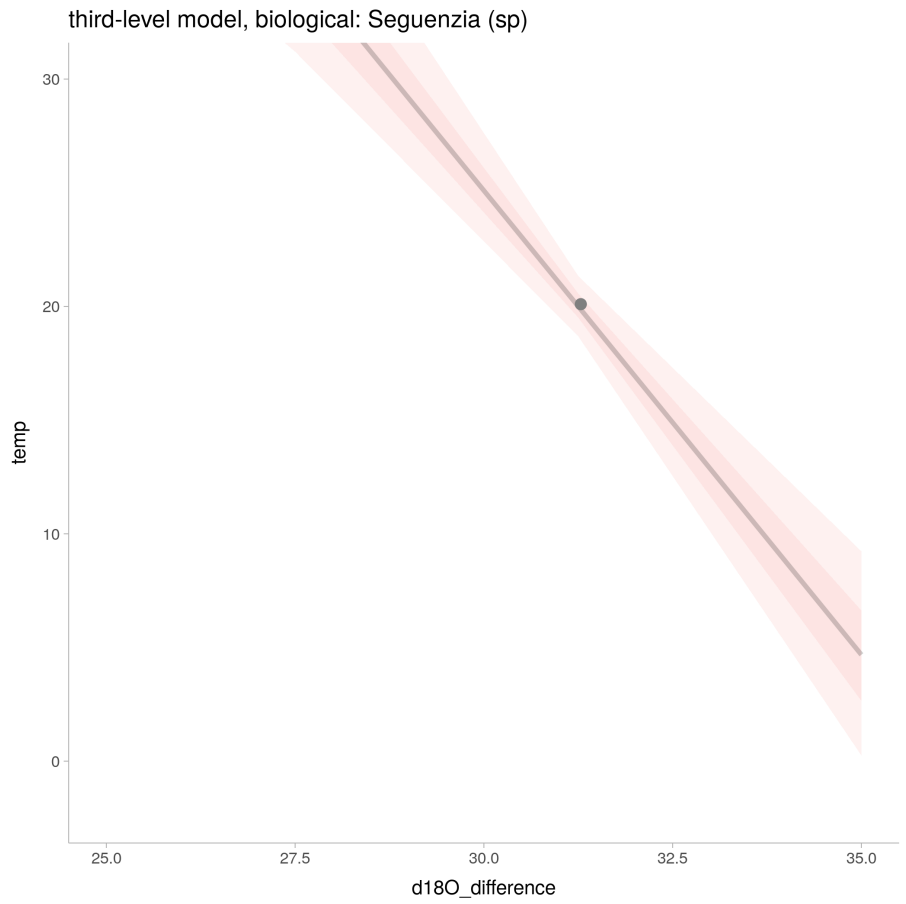

Figure 128: linear regression: three level model (data: bio) for Seguenzia (sp).

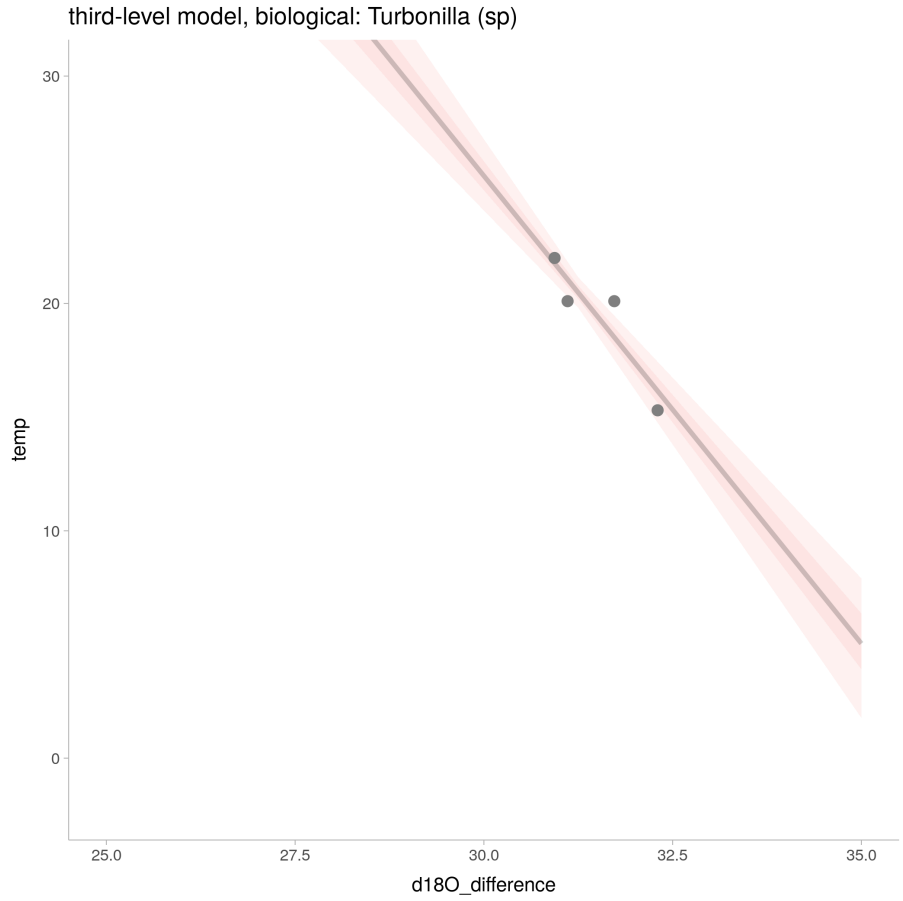

Figure 129: linear regression: three level model (data: bio) for Turbonilla (sp).

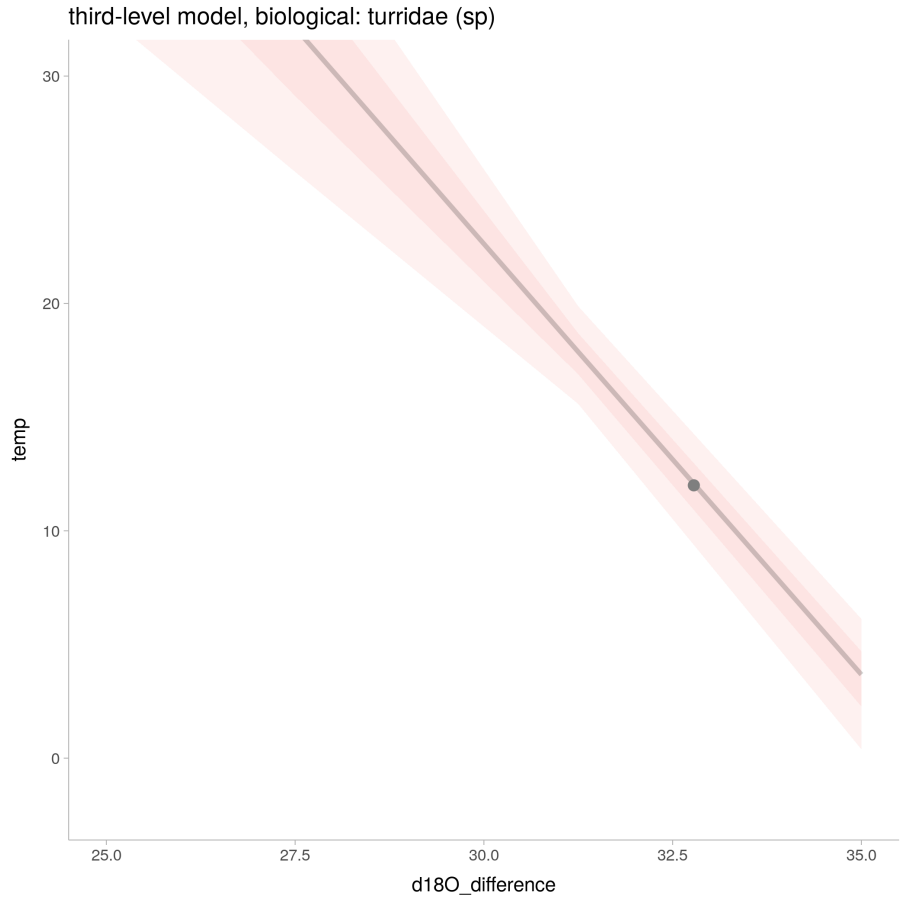

Figure 130: linear regression: three level model (data: bio) for turridae (sp).

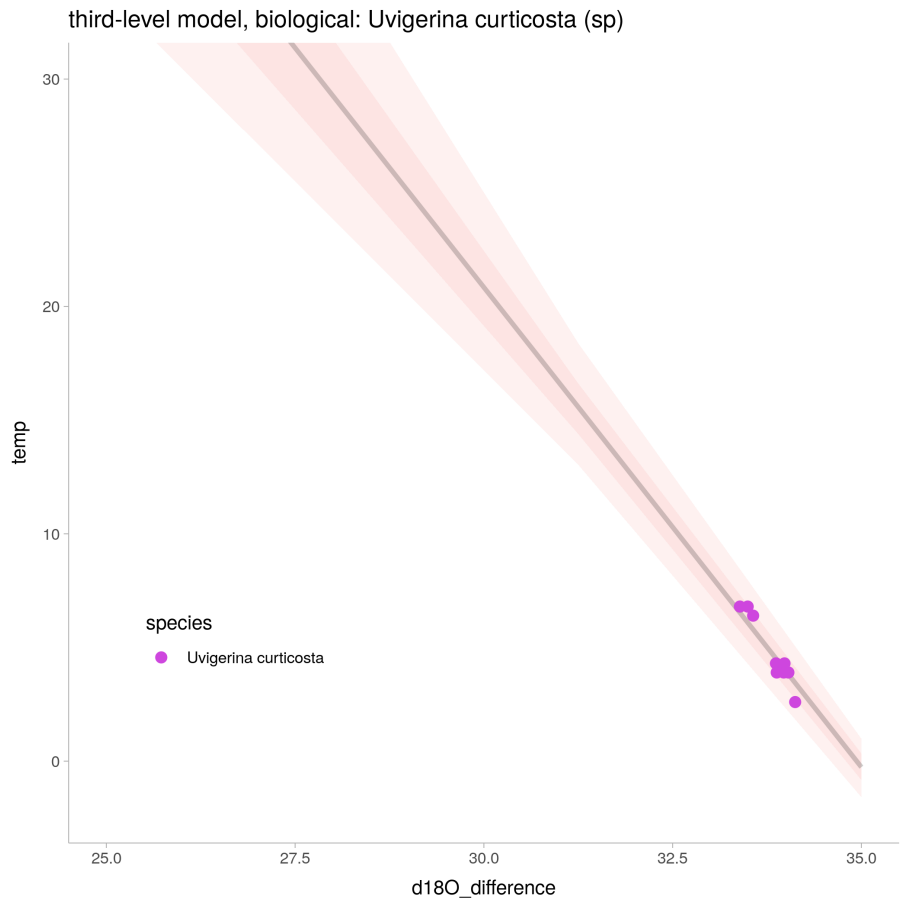

Figure 131: linear regression: three level model (data: bio) for Uvigerina curticosta (sp).

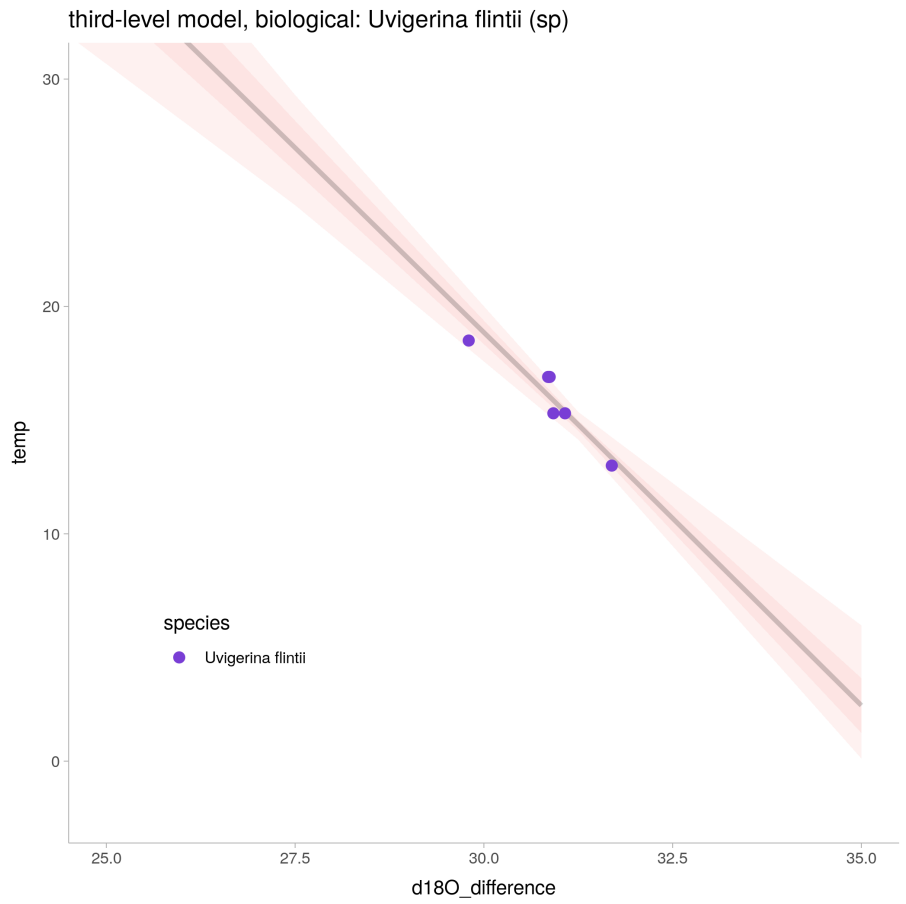

Figure 132: linear regression: three level model (data: bio) for Uvigerina flintii (sp)

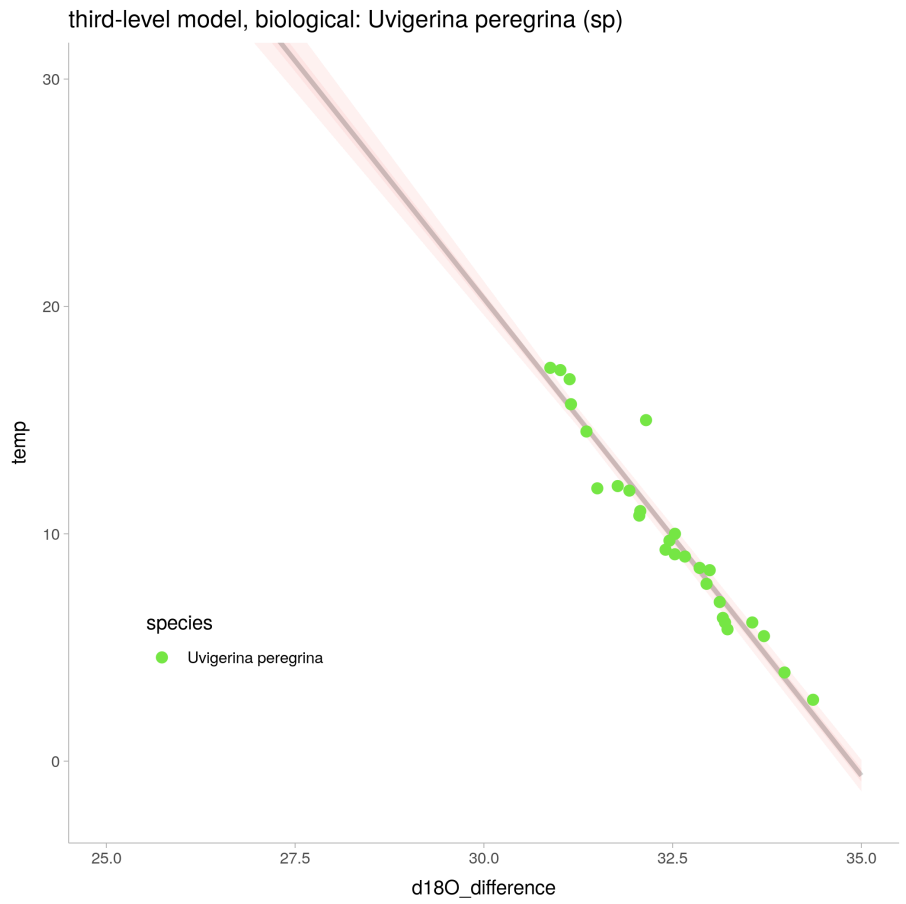

Figure 133: linear regression: three level model (data: bio) for Uvigerina peregrina (sp)

8.13 partial pooling model with a third level of composition: (foraminifera and lab data)

Parameter	Rhat	n_eff	mean	sd	se_mean	2.5%	97.5%
ab_pop[1]	1.00	2837	125.78	4.82	0.09	116.12	134.99
ab_pop[2]	1.00	2185	-3.74	0.18	0.00	-4.11	-3.38
sigma_obs	1.00	5521	1.41	0.05	0.00	1.31	1.52
log-posterior	1.00	961	-337.93	6.25	0.20	-351.14	-326.47
corr_co[1,1]			1.00	0.00		1.00	1.00
corr_co[2,1]	1.00	1620	-0.31	0.59	0.01	-0.99	0.90
corr_co[1,2]			0.00	0.00		0.00	0.00
corr_co[2,2]	1.00	1393	0.70	0.27	0.01	0.13	1.00
sd_sp[1]	1.00	1204	28.06	6.83	0.20	16.16	43.46
sd_sp[2]	1.00	1201	0.82	0.20	0.01	0.46	1.27
ab_co[1,1]	1.00	1658	128.56	6.70	0.16	115.46	142.32
ab_co[2,1]	1.00	1742	-3.69	0.20	0.00	-4.11	-3.29
ab_co[1,2]	1.01	846	129.77	11.07	0.38	113.78	160.08
ab_co[2,2]	1.00	881	-3.84	0.33	0.01	-4.78	-3.36
ab_co[1,3]	1.00	1571	129.99	8.97	0.23	114.39	149.82
ab_co[2,3]	1.00	1678	-3.69	0.27	0.01	-4.26	-3.17
ab_co[1,4]	1.00	793	127.32	9.75	0.35	111.33	151.20
ab_co[2,4]	1.00	777	-3.88	0.29	0.01	-4.62	-3.41
ab_co[1,5]	1.00	793	130.07	10.72	0.38	114.89	160.43
ab_co[2,5]	1.00	781	-3.85	0.32	0.01	-4.73	-3.38
ab_co[1,6]	1.00	1529	125.70	9.09	0.23	109.00	145.69
ab_co[2,6]	1.00	1369	-3.85	0.27	0.01	-4.45	-3.36
ab_co[1,7]	1.00	3899	125.54	11.57	0.19	99.68	148.08
ab_co[2,7]	1.00	3225	-3.74	0.36	0.01	-4.46	-2.98
ab_sp[1,1]	1.00	3906	142.78	3.09	0.05	136.87	148.88
ab_sp[2,1]	1.00	3923	-4.01	0.09	0.00	-4.19	-3.83
ab_sp[1,2]	1.00	3891	127.01	16.88	0.27	91.62	159.12
ab_sp[2,2]	1.00	3883	-3.60	0.55	0.01	-4.63	-2.46
ab_sp[1,3]	1.00	5398	144.98	8.94	0.12	127.85	162.84
ab_sp[2,3]	1.00	5430	-4.16	0.28	0.00	-4.71	-3.63
ab_sp[1,4]	1.00	4780	139.46	25.80	0.37	90.07	194.89
ab_sp[2,4]	1.00	4781	-3.98	0.76	0.01	-5.62	-2.52
ab_sp[1,5]	1.00	3231	79.39	10.16	0.18	59.57	99.23
ab_sp[2,5]	1.00	3233	-2.32	0.30	0.01	-2.91	-1.73
ab_sp[1,6]	1.00	5143	143.74	8.90	0.12	126.30	160.96
ab_sp[2,6]	1.00	5129	-4.18	0.28	0.00	-4.73	-3.63
ab_sp[1,7]	1.00	5268	126.02	20.48	0.28	87.02	167.23
ab_sp[2,7]	1.00	5259	-3.63	0.64	0.01	-4.93	-2.42
ab_sp[1,8]	1.00	4829	135.12	16.05	0.23	105.73	168.62
ab_sp[2,8]	1.00	4830	-3.92	0.52	0.01	-5.01	-2.97
ab_sp[1,9]	1.00	4198	137.50	2.25	0.03	133.01	142.03
ab_sp[2,9]	1.00	4228	-3.84	0.08	0.00	-3.99	-3.68
ab_sp[1,10]	1.00	4376	137.66	3.50	0.05	130.74	144.37
ab_sp[2,10]	1.00	4365	-4.24	0.13	0.00	-4.48	-3.99
ab_sp[1,11]	1.00	4410	153.40	3.64	0.05	146.26	160.66

ab_sp[2,11]	1.00	4430	-4.67	0.13	0.00	-4.93	-4.41
ab_sp[1,12]	1.00	5613	169.52	10.45	0.14	149.26	190.63
ab_sp[2,12]	1.00	5648	-5.00	0.37	0.00	-5.74	-4.28
ab_sp[1,13]	1.00	3539	173.78	1.35	0.02	171.08	176.41
ab_sp[2,13]	1.00	3578	-5.12	0.05	0.00	-5.21	-5.02

Table 18: hierarchical model with correlated parameters

model run time: 262.00

```
Computed from 4000 by 382 log-likelihood matrix

         Estimate   SE
elpd_loo   -682.6 27.2
p_loo        28.1  4.0
looic      1365.1 54.4
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     374   97.9%   582
 (0.5, 0.7]   (ok)         6    1.6%   532
   (0.7, 1]   (bad)        2    0.5%   48
   (1, Inf)   (very bad)   0    0.0%   <NA>
See help('pareto-k-diagnostic') for details.
Warning message:
Some Pareto k diagnostic values are too high.
See help('pareto-k-diagnostic') for details.
```

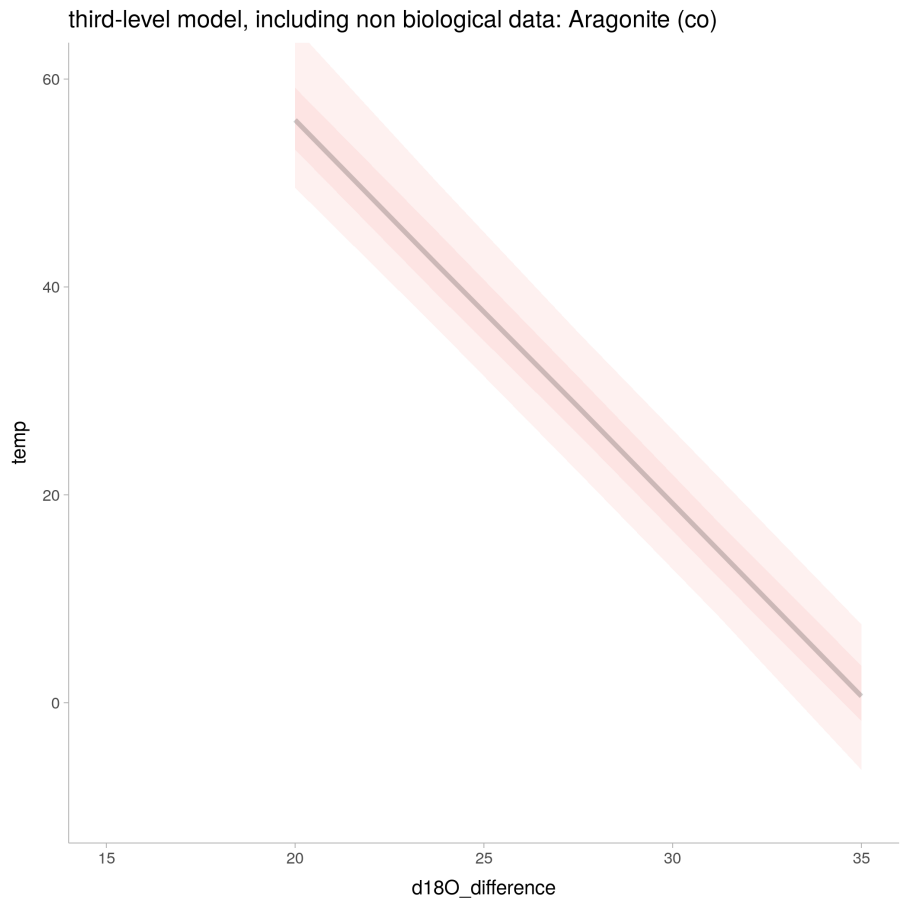

Figure 134: linear regression: three level model (foraminifera and lab data) for Aragonite (co) (datapoints excluded)

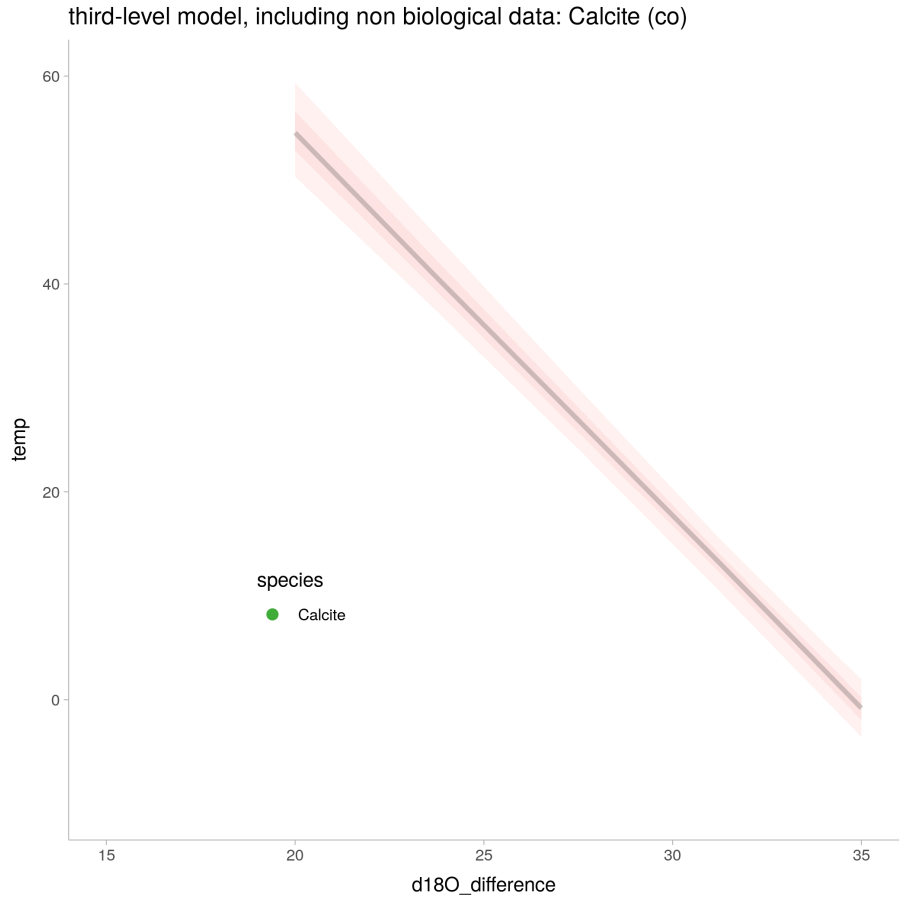

Figure 135: linear regression: three level model (foraminifera and lab data) for Calcite (co)

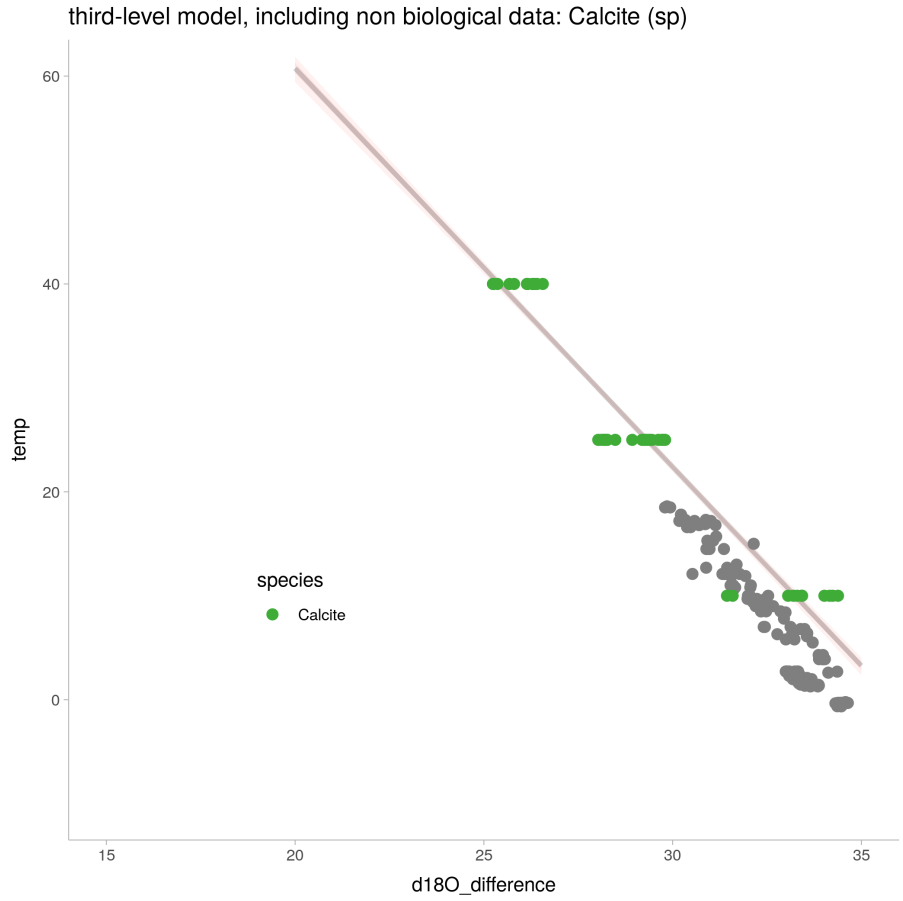

Figure 136: linear regression: three level model (foraminifera and lab data) for Calcite (sp)

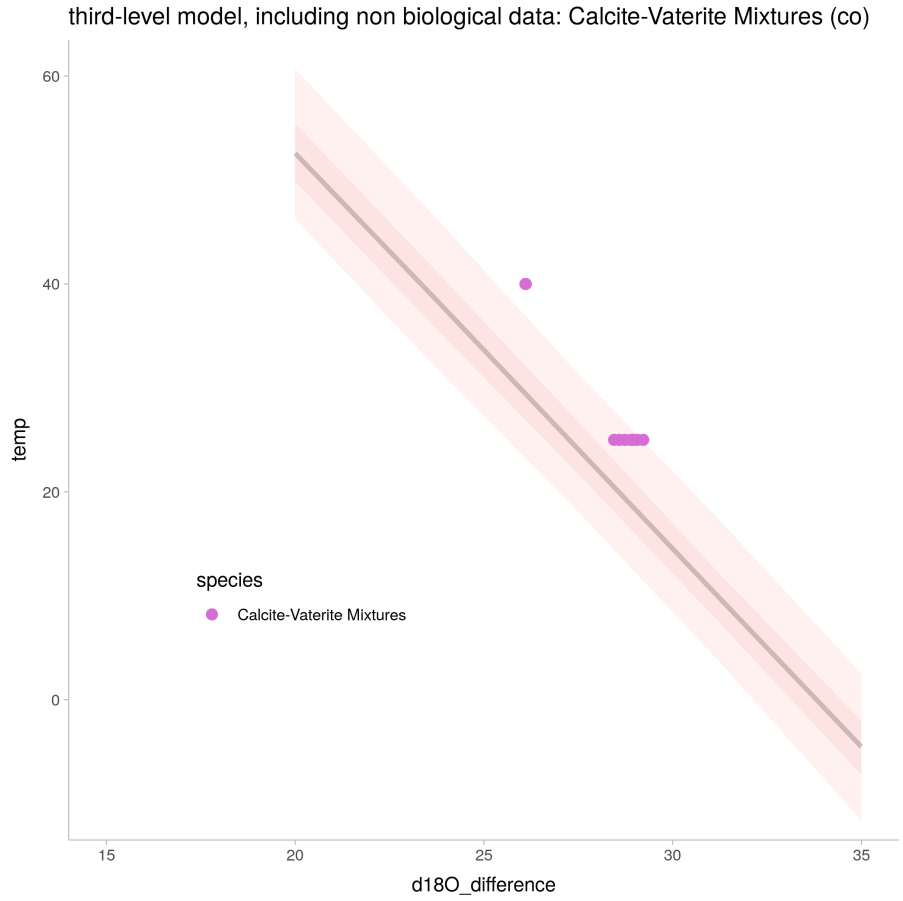

Figure 137: linear regression: three level model (foraminifera and lab data) for Calcite-Vaterite Mixtures (co).

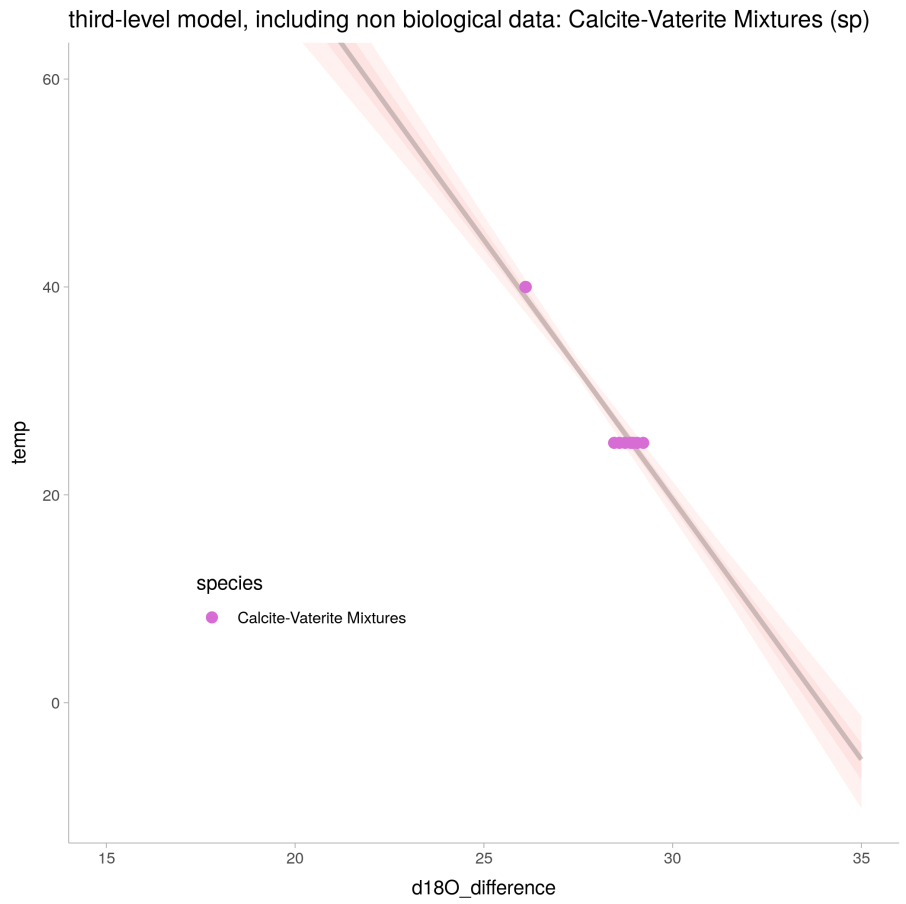

Figure 138: linear regression: three level model (foraminifera and lab data) for Calcite-Vaterite Mixtures (sp)

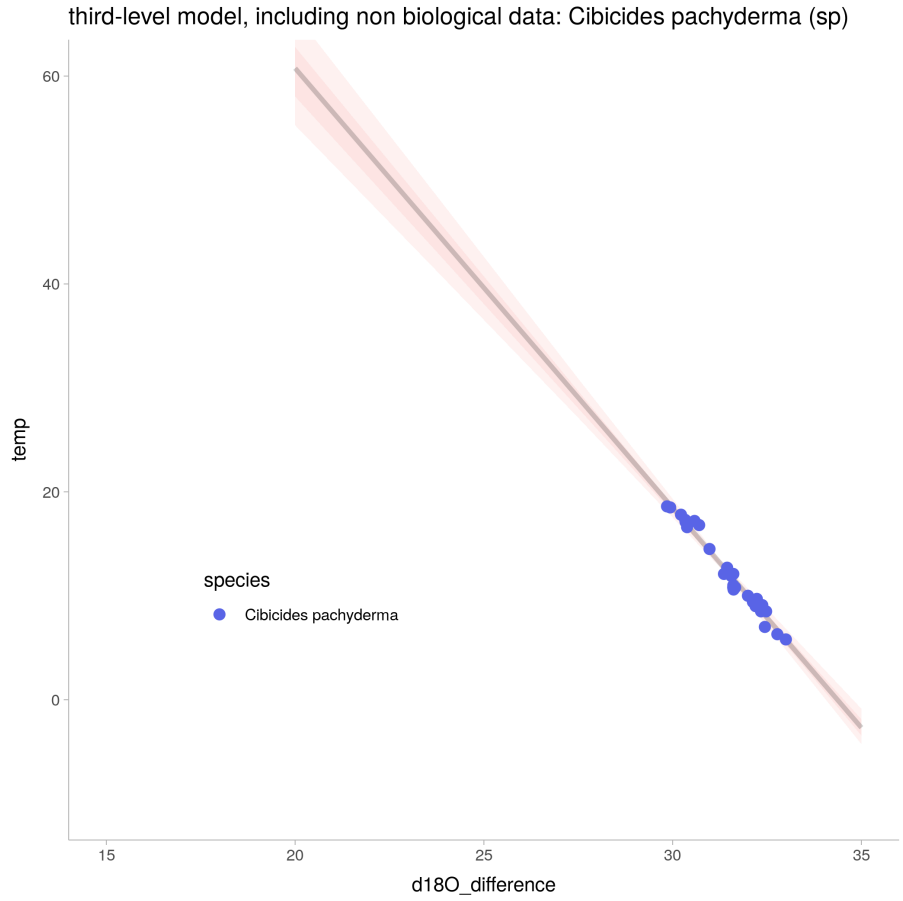

Figure 139: linear regression: three level model (foraminifera and lab data) for Cibicides pachyderma (sp)

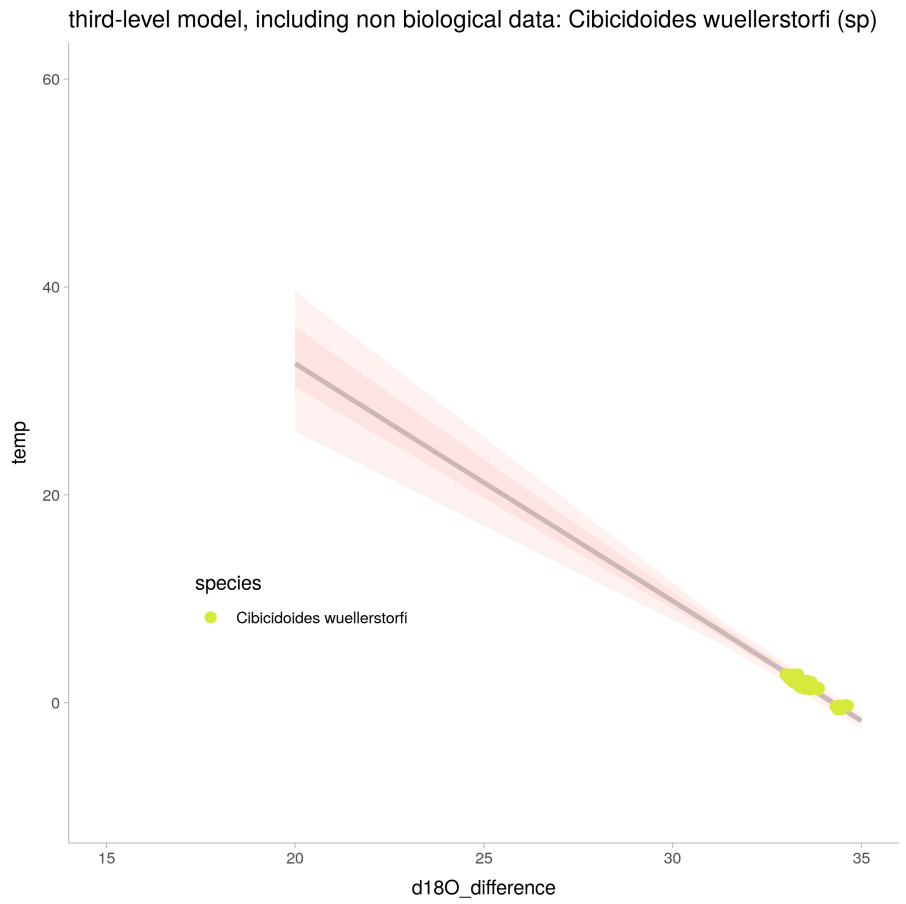

Figure 140: linear regression: three level model (foraminifera and lab data) for Cibicidoides wuellerstorfi (sp)

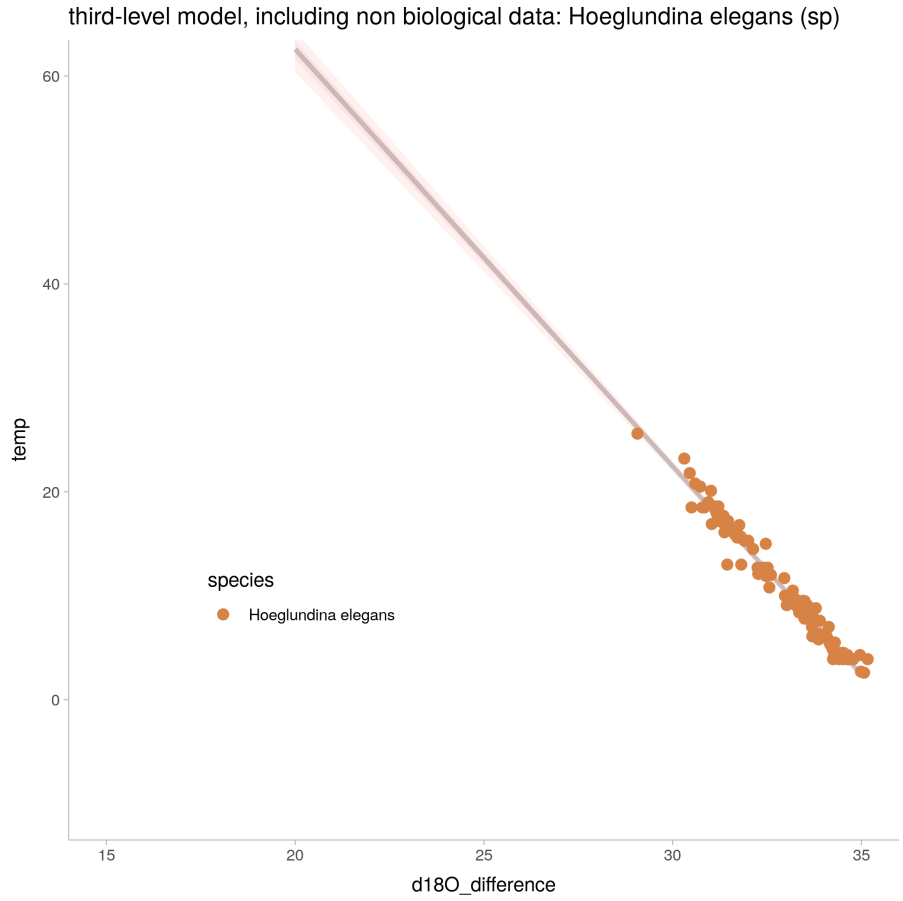

Figure 141: linear regression: three level model (foraminifera and lab data) for Hoeglundina elegans (sp)

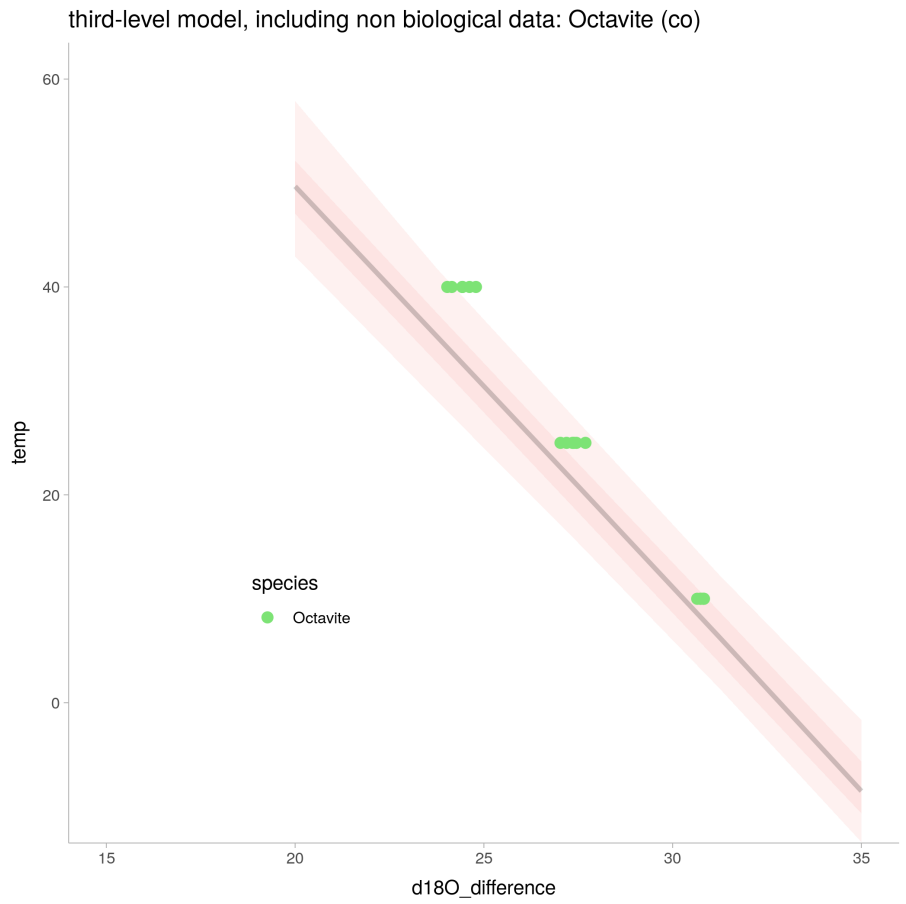

Figure 142: linear regression: three level model (foraminifera and lab data) for Octavite (co) (datapoints excluded)

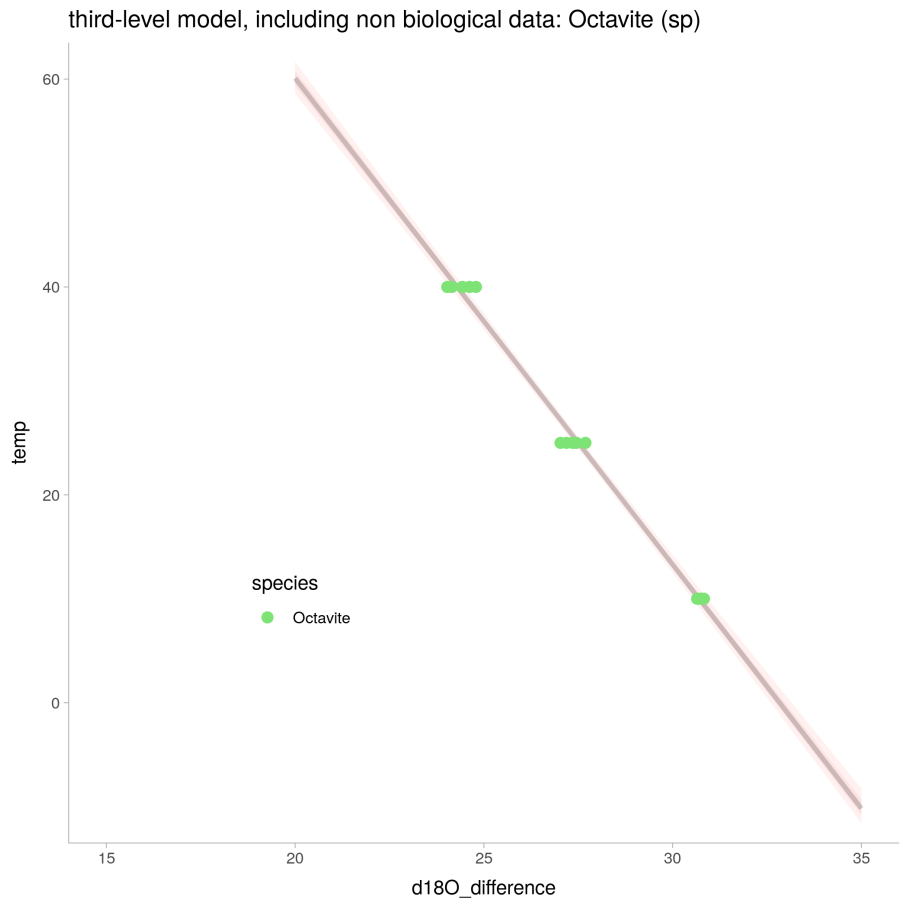

Figure 143: linear regression: three level model (foraminifera and lab data) for Octavite (sp)

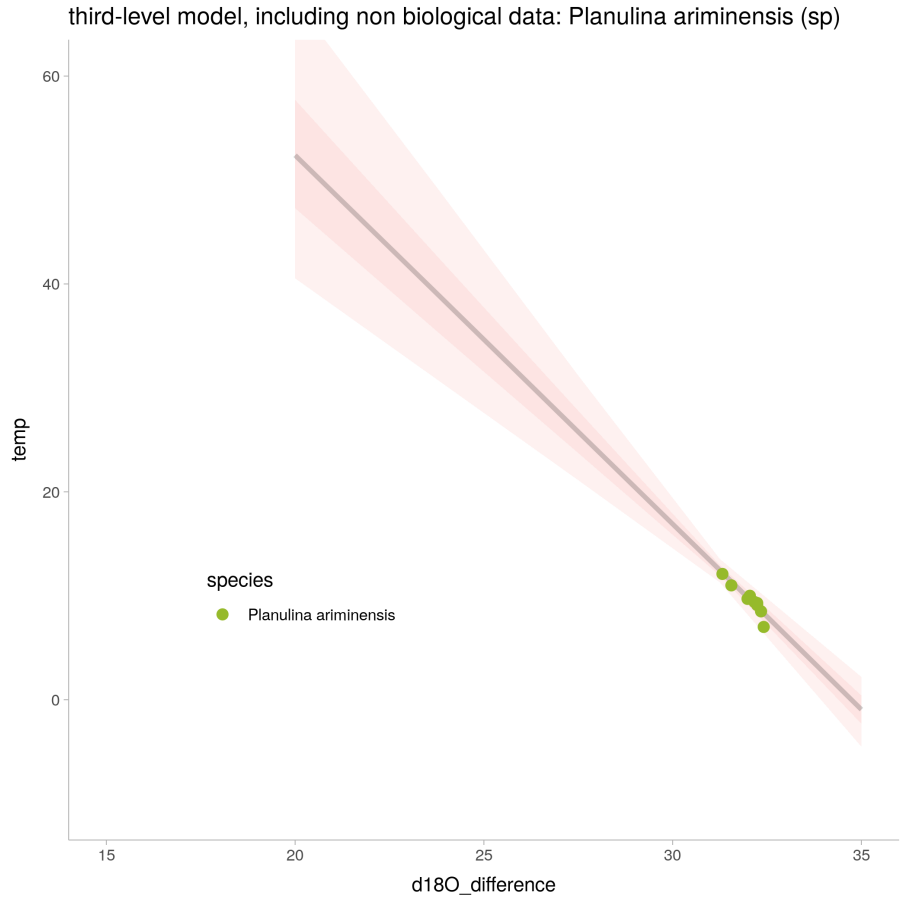

Figure 144: linear regression: three level model (foraminifera and lab data) for Planulina ariminensis (sp)

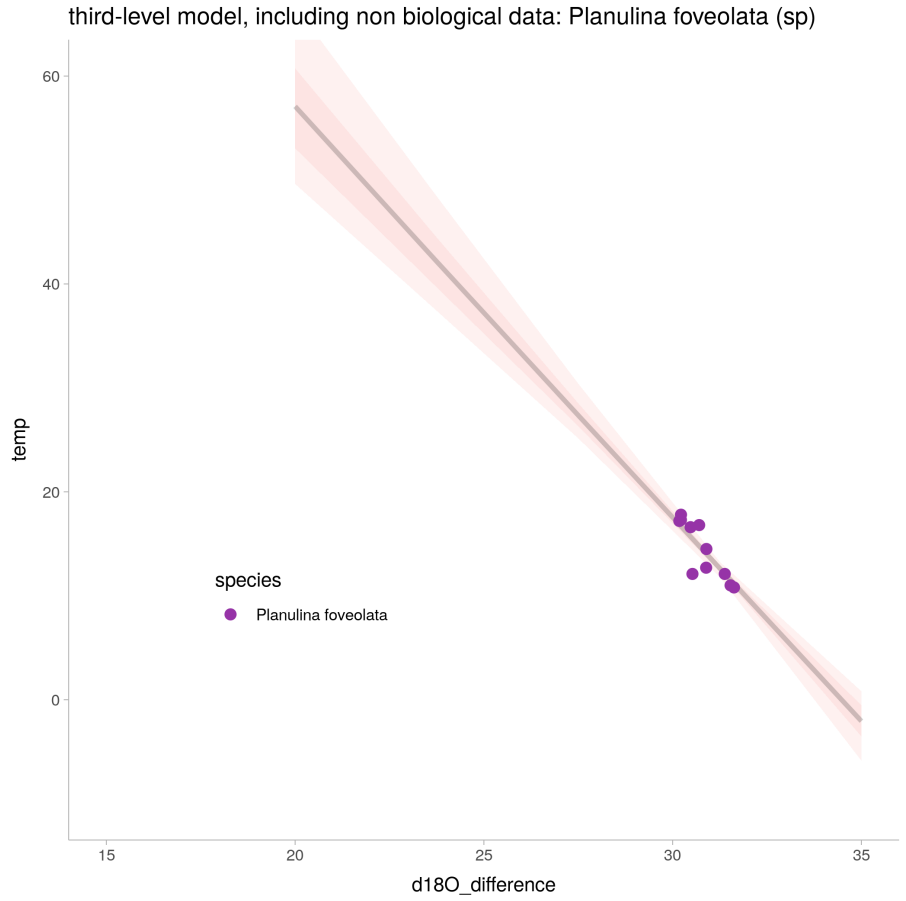

Figure 145: linear regression: three level model (foraminifera and lab data) for Planulina foveolata (sp)

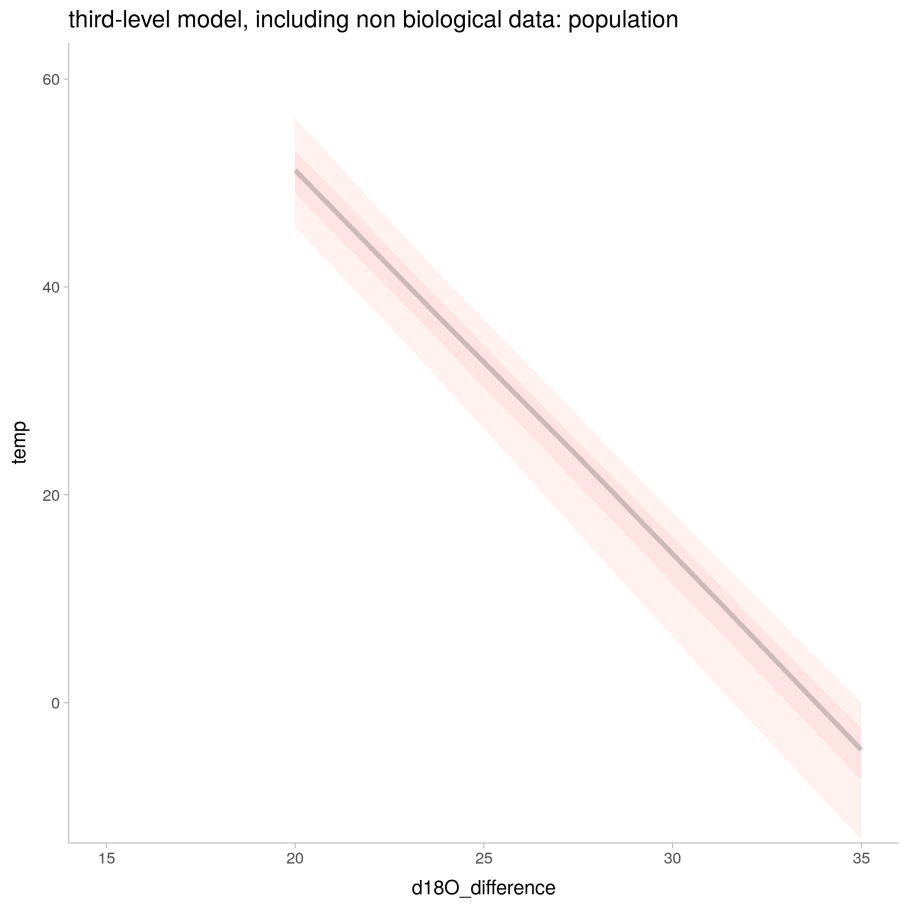

Figure 146: linear regression: three level model (foraminifera and lab data) for population

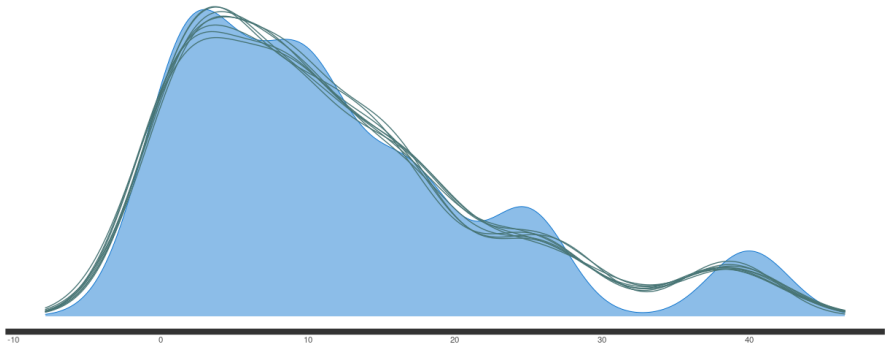

Figure 147: posterior predictive check for the three level model ran over foraminifera and lab data

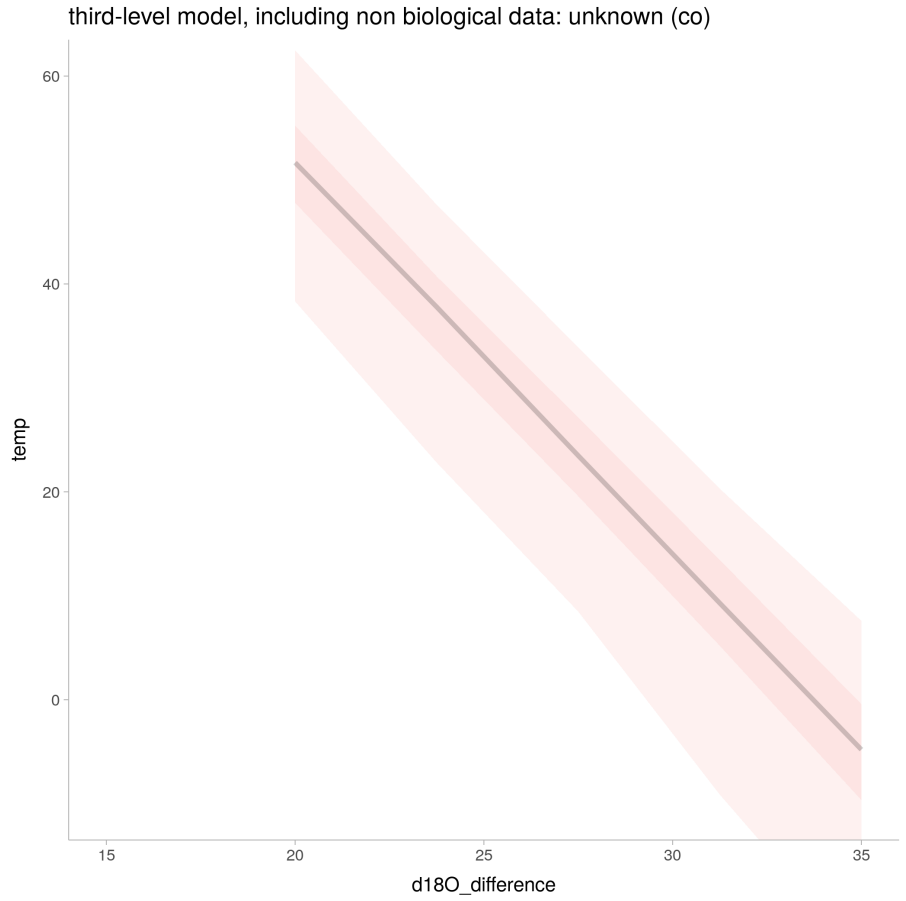

Figure 148: linear regression: three level model (foraminifera and lab data) for unknown (co) (datapoints excluded)

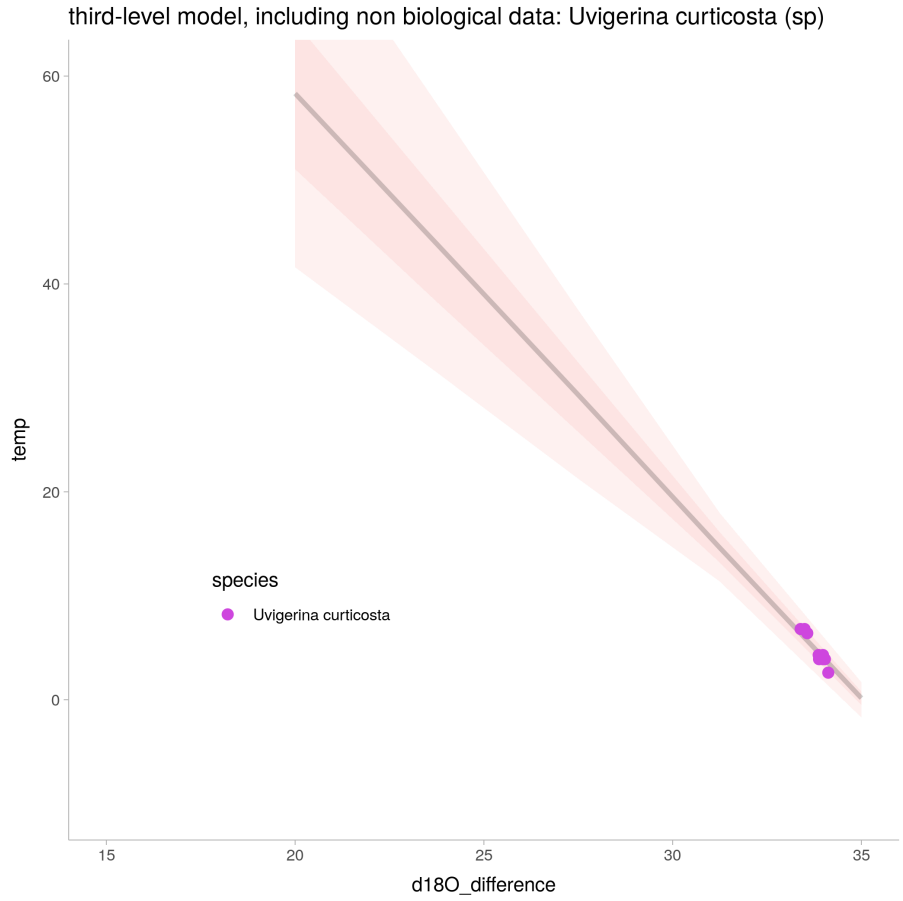

Figure 149: linear regression: three level model (foraminifera and lab data) for Uvigerina curticosta (sp)

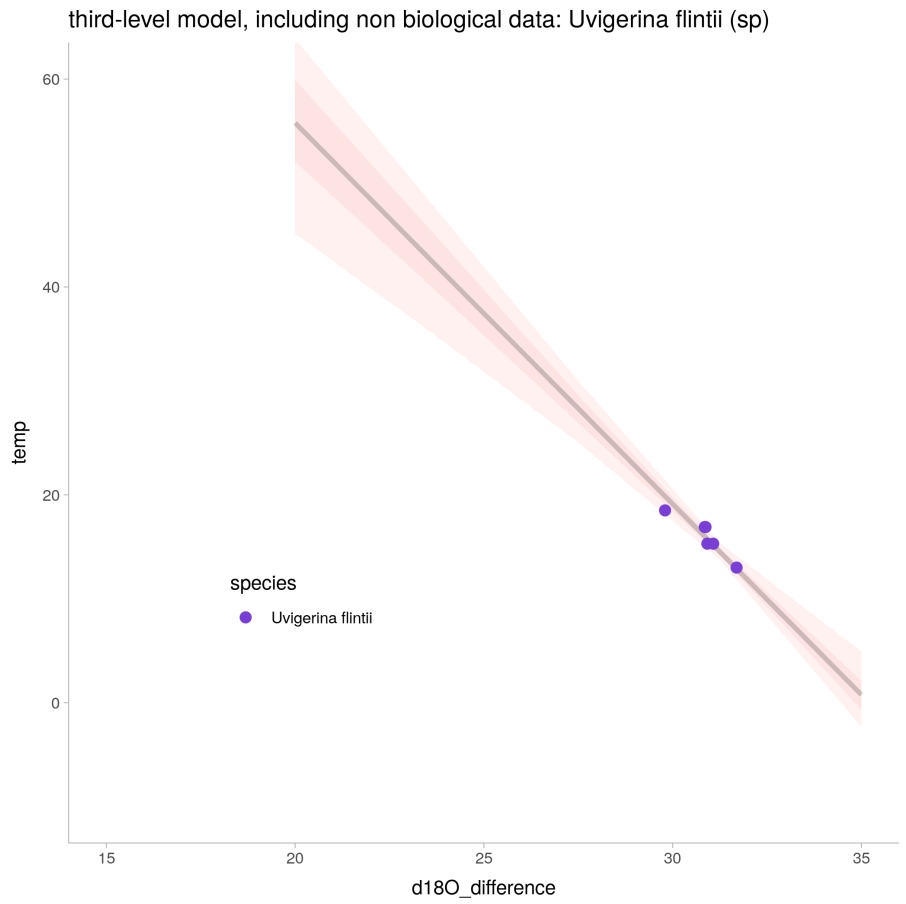

Figure 150: linear regression: three level model (foraminifera and lab data) for Uvigerina flintii (sp)

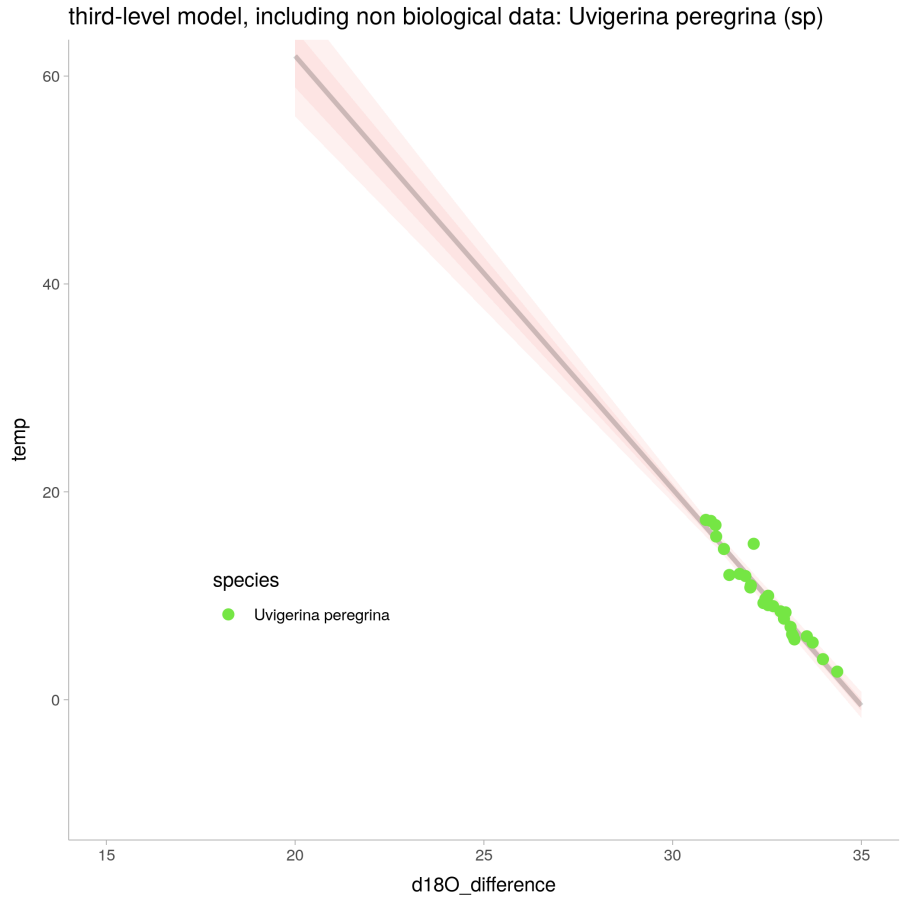

Figure 151: linear regression: three level model (foraminifera and lab data) for Uvigerina peregrina (sp)

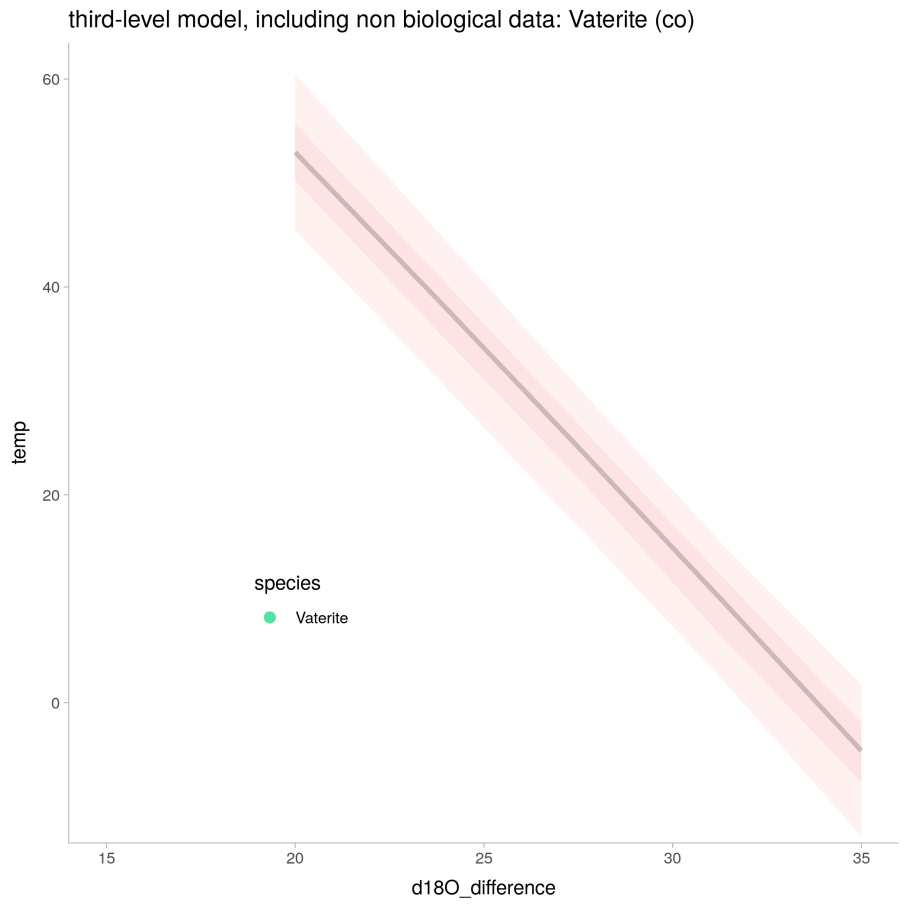

Figure 152: linear regression: three level model (foraminifera and lab data) for Vaterite (co) (datapoints excluded

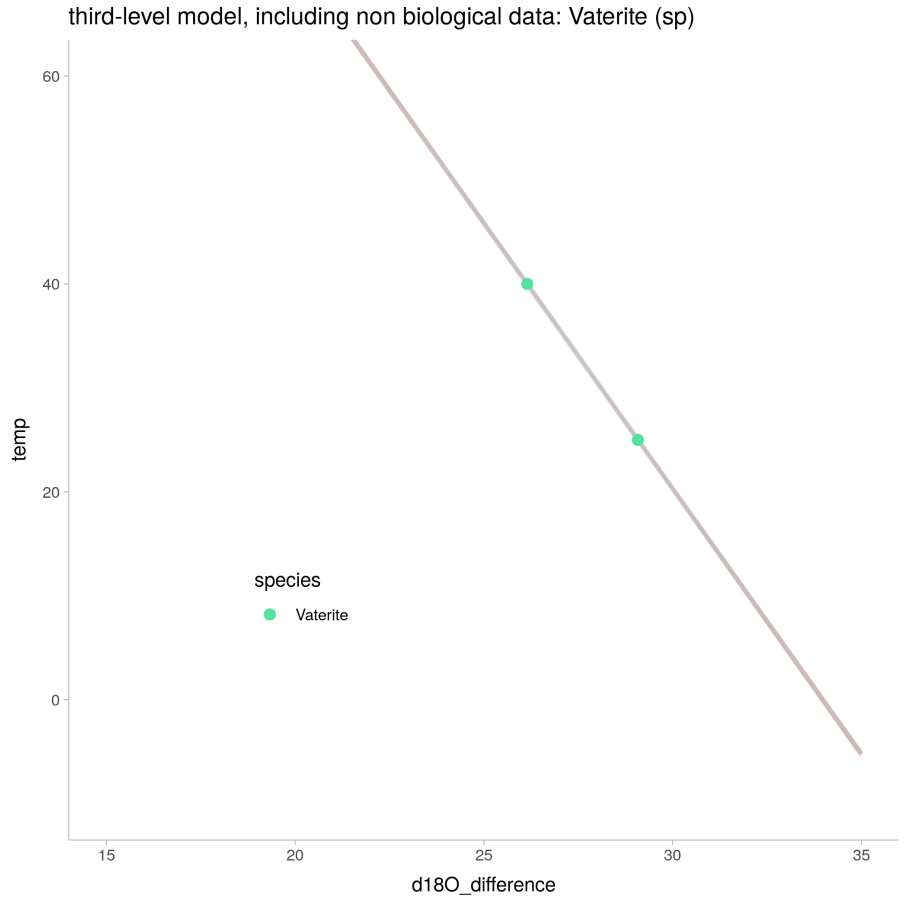

Figure 153: linear regression: three level model (foraminifera and lab data) for Vaterite (sp)

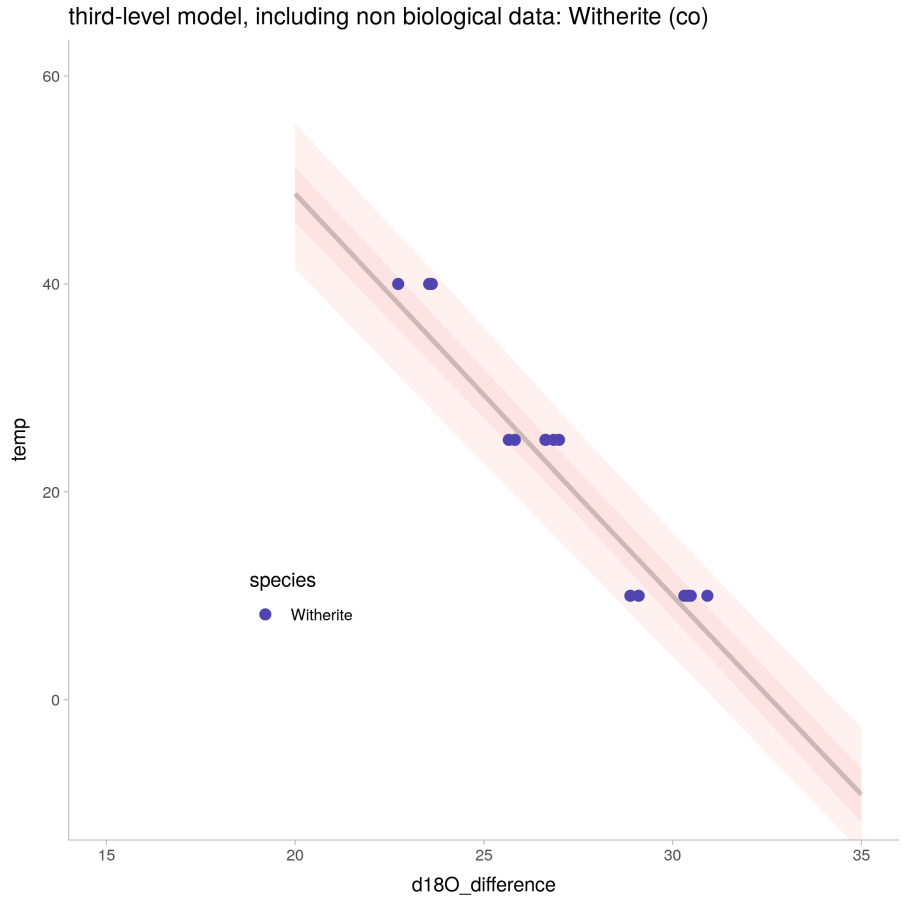

Figure 154: linear regression: three level model (foraminifera and lab data) for Witherite (co)

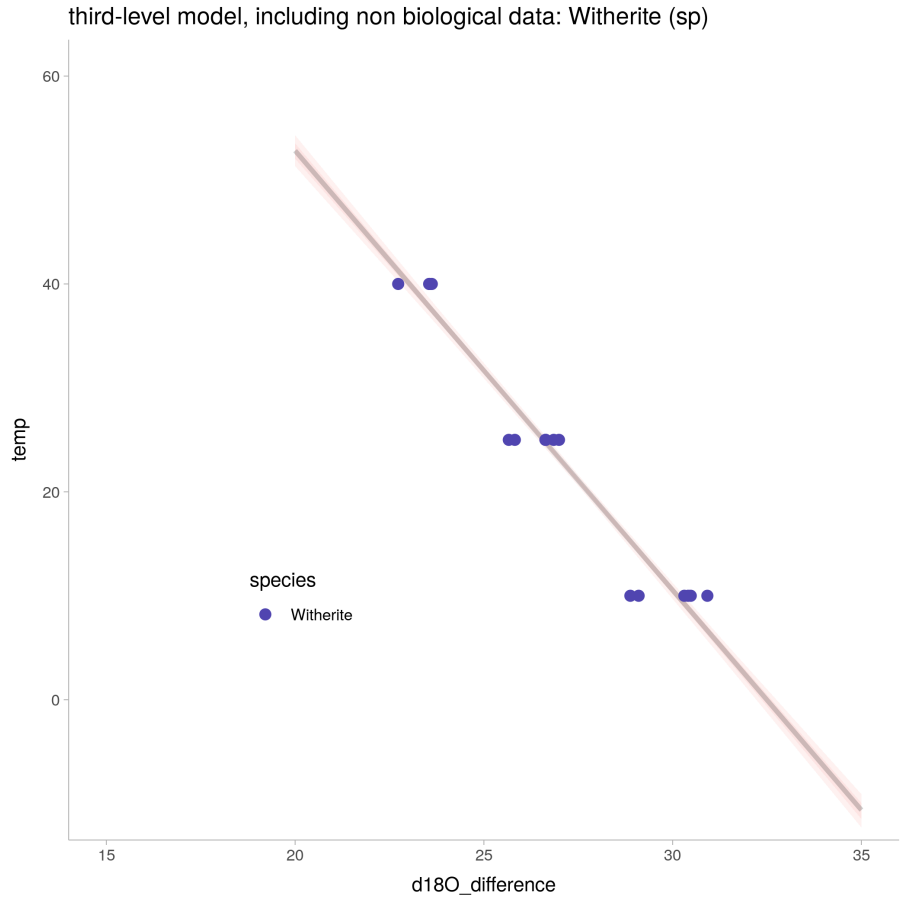

Figure 155: linear regression: three level model (foraminifera and lab data) for Witherite (sp)