

# **Trophic interaction models predict interactions across ecosystems, not food webs.**

Dominique Caron, Ulrich Brose, Miguel Lurgi, F. Guillaume Blanchet, Dominique Gravel, Laura J. Pollock

## **Table of Content:**

**Appendix S1 – Description of the food webs:** Brief description of how the four food web were collated, and how we extracted predator-prey interactions

**Appendix S2 – Description of the structure of the predictive model:** We detail the structure of the predictive model (response variable, predictors, transformations, priors).

**Appendix S3 - Description of species functional roles and food web-level properties:** We describe each functional role and food web-level metrics, how we calculated them, and their relation to functioning.

**Appendix S4 - Environmental distance sensitivity analysis to random sampling:** We test the effect of the random sampling for the environmental distance estimate.

**Appendix S5 - Convergence of the predictive models:** Distribution of rank-normalized potential scale reduction factor on split chains (Rhat) and rank plots of the population-level effects.

**Appendix S6 - Predicting species interactions extended results:** Results of the predicting species interactions with other evaluation metrics, and results of the linear models between model performance and distances estimates.

**Appendix S7 - Predicting functional roles extended results:** Additional results for the prediction of species functional role (linear model R<sup>2</sup> and coefficients).

**Appendix S8 - Predicting food web properties extended results:** Additional results for the prediction of food web results.

**Appendix S9 – References:** References for the supplemental material

## **Appendix 1: Description of the food webs**

*Europe food web:* The food web comprising European tetrapods was assembled by Maiorano et al. (2020). It contains the potential trophic interactions between all terrestrial vertebrates in Europe based on data from atlases, field guides, published papers, and expert knowledge. Maiorano et al. (2020) originally reported interactions between predators and their prey at different life-stage (eggs, larvae/young, or adult). In this study, we focused on interactions between predators and adult prey only.

*Pyrenees food web:* The Pyrenees vertebrate food web was assembled independently from the European food web by Lurgi et al. (2012) who compiled all known interactions between vertebrate species in the alpine ecosystem of the southeastern Pyrenees mountain range. Data was collated from atlases, existing databases, published scientific articles and reports, and expert knowledge. This network is described at the species level and for this study, we excluded fish species.

*Northern Québec and Labrador food web:* The trophic tundra food web of Labrador and northern Quebec (Canada) was compiled by Berteaux et al. (2018). The Northern Québec and Labrador food web documents the presence (absence) of trophic relationships between all birds, mammals, and groups of plants (e.g., moss, ferns, grass) and arthropods (e.g., Lepidoptera, Diptera, Arachnida) whose distribution overlap Labrador and Quebec north of 50°N (Nunavik region). Trophic interactions were collated from a literature review, data from existing databases, and expert knowledge. For our study, we focused on the bird and mammal species (i.e., the vertebrates) of this network.

*Serengeti food web:* The Serengeti food web compiles all interactions between resource categories (e.g., seeds, shrubs, fruits and nectar), invertebrate groups (e.g., thrips, bees, moths, crickets), and vertebrate species in the Serengeti National Park (Tanzania) and was assembled from a literature review (de Visser et al., 2011). The nodes of the original food webs are resource categories, or trophic groups including one or more invertebrate groups and vertebrate species. In this study, we focused on vertebrate species, and assumed that species within a trophic group share the same predator and prey species.

## Appendix 2: Description of the structure of the predictive model

The predictive models we use is a Bayesian hierarchical generalized linear model with 13 trait-based predictor variables with varying intercepts and slopes for the order of the predator. All traits were extracted from Etard et al. (2020). Gaps in the database were imputed using MissForest (Stekhoven, 2013; Stekhoven & Buehlmann, 2012). All traits are described in the original paper, but here, we describe briefly each derived variable used in the predictive models:

- Population-level effects:
  - Trophic level of the predator: Categorical with 3 levels - Herbivore, Omnivore, and Carnivore. Coded as 2 binary variables (Omnivore and Carnivore).
  - Habitat Breadth of the predator: Quantitative - Number of habitats the predator is known to use based on the second level of the IUCN habitat data.
  - Body mass of the predator: Quantitative - Log-transformed body mass of the predator
  - Longevity of the predator: Quantitative - Life span of an individual of the predator species.
  - Clutch Size of the predator: Quantitative - Number of offspring or eggs per clutch/litter by the predator.
  - Trophic level of the prey: Categorical with 3 levels - Herbivore, Omnivore, and Carnivore. Coded as 2 binary variables (Omnivore and Carnivore)
  - Habitat Breadth of the prey: Quantitative - Number of habitats the prey is known to use based on the second level of the IUCN habitat data.
  - Body mass of the prey: Quantitative - Log-transformed body mass of the prey
  - Longevity of the prey: Quantitative - Life span of an individual of the prey species.
  - Clutch Size of the prey: Quantitative - Number of offspring or eggs per clutch/litter by the prey.
  - Activity Time match: Categorical binary – The predator and the prey are active during the same time of the day (1) or not (0). The original activity time data for the prey and the predator is a binary variable with possible values nocturnal and non-nocturnal (diurnal, crepuscular, or catemeral)
  - Habitat match: Quantitative – Jaccard similarity between the habitat use of the predator and the prey. The original habitat data are 12 binary variables: Forest, Savanna, Shrubland, Grassland, Wetland, Rocky areas, Cave and subterranean, Desert, Marine, Marine/Intertidal or coastal/supratidal, Artificial, Introduced vegetation
  - Body mass match: The squared difference between the log-transformed body mass of the predator and the log-transformed body mass of the prey.

For each food web, we centered and scaled the quantitative predictors by subtracting each value by the mean and dividing by twice the standard deviation. The predictors that we scaled are: the log-transformed habitat breadth of the predator and the prey, the log-transformed body mass of the predator and the prey, the log-transformed longevity of the predator and the prey, the log-transformed clutch/litter size of the predator and the prey, the habitat match, and the body mass match.

- Group-level effects (intercept and slopes):
  - Order of the predator
    - Northern Québec and Labrador food web: 21 levels
    - Europe: 36 levels
    - Pyrenees: 23 levels
    - Serengeti: 41 levels

- Model:

$$L_{ij} \sim Bernoulli(p_{ij}),$$

$$\text{logit}(p_{ij}) = \alpha + \alpha_{\text{predator}[j]} + \sum_{k=1}^{13} (\beta_k \times T_k + \beta_{k,\text{predator}[j]} \times T_k),$$

- $L_{ij}$  is the interaction between predator  $j$  and prey  $i$ .
- $P_{ij}$  is the probability of interaction between predator  $j$  and prey  $i$
- $\alpha$  and  $\beta$  are the fixed intercept and slopes
- $\alpha_{\text{predator}[j]}$  and  $\beta_{\text{predator}[j]}$  are the order-specific intercept and slopes for the order of predator  $j$
- $T_k$  are the trait-based predictors.

- Priors:

$$\alpha, \beta \sim Normal(0, 1)$$

$$\alpha_{\text{predator}[j]}, \beta_{k,\text{predator}[j]} \sim Normal(\mu, \sigma)$$

$$\mu \sim Normal(0, 1)$$

$$\sigma \sim HalfCauchy(0, 5)$$

- Number of iterations:
  - Warmup: 2000 iterations
  - Inference: 2000 iterations
- Calibration data:
  - 70% of all interactions in each food web and a equal number of non-interactions (random undersampling): Northern Québec and Labrador food web (769 interactions and non-interactions), Europe (40 422 interactions and non-interactions), Pyrenees (582 interactions and non-interactions), Serengeti (7727 interactions and non-interactions). Random undersampling is a strategy to handle imbalanced dataset where one class of the response variable (non-interactions) is a lot more prevalent than the other class (interactions). Imbalanced dataset can influence and hinder calibration of some predictive models.
- Validation:
  - For within food web predictions (i.e., trophic interaction predictions in the food web used for calibration): We used the remaining 30% of the interactions of each food webs and a number of non-interactions that makes the prevalence of interactions in the validation dataset equal to the prevalence of interactions in the entire food web. Some measures like the area under the precision-recall curve (auprc) are influenced by prevalence. Controlling for prevalence allows comparison of within food web predictions and across food webs predictions for a given food web.
  - For across food web predictions (i.e., trophic interaction predictions in food webs other than the one used for calibration): We used the entire food web (all interactions and non-interactions).

- Model formula:

```
brms_form <- bf(interaction ~ 1 + (Omnivore.predator +
Carnivore.predator + Habitat_breadth.predator + BM.predator +
Longevity.predator + ClutchSize.predator + Omnivore.prey +
Carnivore.prey + Habitat_breadth.prey + BM.prey + Longevity.prey +
ClutchSize.prey + ActivityTime.match + Habitat.match + BM.match) +
(1 + (Omnivore.predator + Carnivore.predator +
Habitat_breadth.predator + BM.predator + Longevity.predator +
ClutchSize.predator + Omnivore.prey + Carnivore.prey +
Habitat_breadth.prey + BM.prey + Longevity.prey + ClutchSize.prey +
ActivityTime.match + Habitat.match + BM.match) || Order.predator),
family = bernoulli())
```

### Appendix 3: Description of species functional roles and food web-level properties

#### Species functional roles:

For more information on what each species role metrics represent and aim to describe, we recommend the reviews made by Cirtwill et al. (2018) and Delmas et al. (2019). Here, we very briefly describe what these metrics are, how they are related to the functioning and the stability of food webs:

- **Number of prey:** We calculated the number of prey of a species as the number of interactions for which the species is the predator (in-degree). It measures how generalized the species is. The loss of species with many prey are more likely to trigger top-down cascades (Curtsdotter et al., 2011). In contrast, specialist predators are more vulnerable to co-extinction. In-degree and out-degree quantify the direct influence of species on other species.
- **Number of predators:** We calculated the number of predators of a species as the number of interactions for which the species is the prey (out-degree). It measures the vulnerability of the prey. The loss of species with many predators are more likely to trigger bottom-up cascades (Curtsdotter et al., 2011). In-degree and out-degree quantify the direct influence of species on other species.
- **Betweenness:** The betweenness centrality of a species is the number of times the shortest path linking any two species in the network goes through the focal species (Freeman, 1977). Betweenness quantify the number of food chains a species takes part in, and thus how it affects energy flows (Cirtwill et al., 2018). It is considered one important measures of species topological importance (Jordán, 2009). We calculated betweenness using the function *centr\_betw* in the R package *igraph* (Csardi & Nepusz, 2006).
- **Closeness:** Closeness centrality measures how far a species is to any other species. It is therefore a more “global” measure of centrality than betweenness and quantify how “efficiently” the change in this species will spread through the food web. We used the function *centr\_clo* in the R package *igraph* (Csardi & Nepusz 2006).
- **Eigenvector centrality:** Eigenvector centrality uses the eigenvector of the largest eigenvalue of the adjacency matrix describing the food web. It is another “global” measure of centrality that aims to describe species importance in the food web. Eigenvector centrality is related to the contribution of species to network stability (Allesina & Pascual, 2009). We used the function *centr\_eigen* of the R package *igraph* (Csardi & Nepusz 2006).
- **Trophic level:** Trophic level describe the “vertical” position (primary consumer, secondary consumer, etc.) of a species in a food web. We calculated trophic level using the function *TrophInd* of the package *NetIndices* package (Kones et al., 2009). The function calculate trophic level as 1 + the weighted average of the trophic levels of its food items. Top predators and basal species are expected to have larger effect on the community through top-down and bottom-up cascades (Dyer & Letourneau, 2003).
- **Omnivory:** Omnivory is defined as the variety of trophic levels a species feed on. Omnivory has the potential to stabilize food webs by mitigating trophic cascades, although this effect is debated (Bascompte et al., 2005). We calculated omnivory using the function *TrophInd* of the package *NetIndices* package (Kones et al. 2009).

- **Within-module degree:** Within-module degree quantify how a species is connected within its module (Guimerà & Amaral, 2005). Modules (groups of species interacting more together than with other species) are first defined using a spin-glass model and simulated annealing using the function *cluster\_springlass* from the package *igraph* (Csardi & Nepusz 2006). We used the function *calc\_topological\_roles* of the R package *multiweb* (Saravia 2022) to calculate within-module degree. Species with high within-module degree are important in the stability and functioning of their module.
- **Participation coefficient:** Participation coefficient quantify how much a species interact with species in other module (Guimerà & Amaral 2005). A participation coefficient equal to 0 defines a species solely interacting within its module, whereas 1 defines a species interacting with species with all other modules equally. Species with high participation coefficient connects module together, and therefore, are important in the cohesion of the entire food web. We used the function *calc\_topological\_roles* of the R package *multiweb* (Saravia, 2022) to calculate participation coefficient.
- **Motif profile:** The motif profile of a species is the frequency of which a species occupies a position within each motif (Milo et al., 2002). We used 3-species motifs, which are the 13 different arrangements 3 species can take. For 3-species motifs, there are 30 different position a species can take. The motif profile has been proposed to be a more complete description of the Eltonian niche of a species (Cirtwill et al., 2018; Stouffer et al., 2012). Some motifs such as “linear chain”, “intraguild predation”, “direct competition”, and “apparent competition” influence the stability of food web (Borrelli, 2015; Stouffer & Bascompte, 2010).

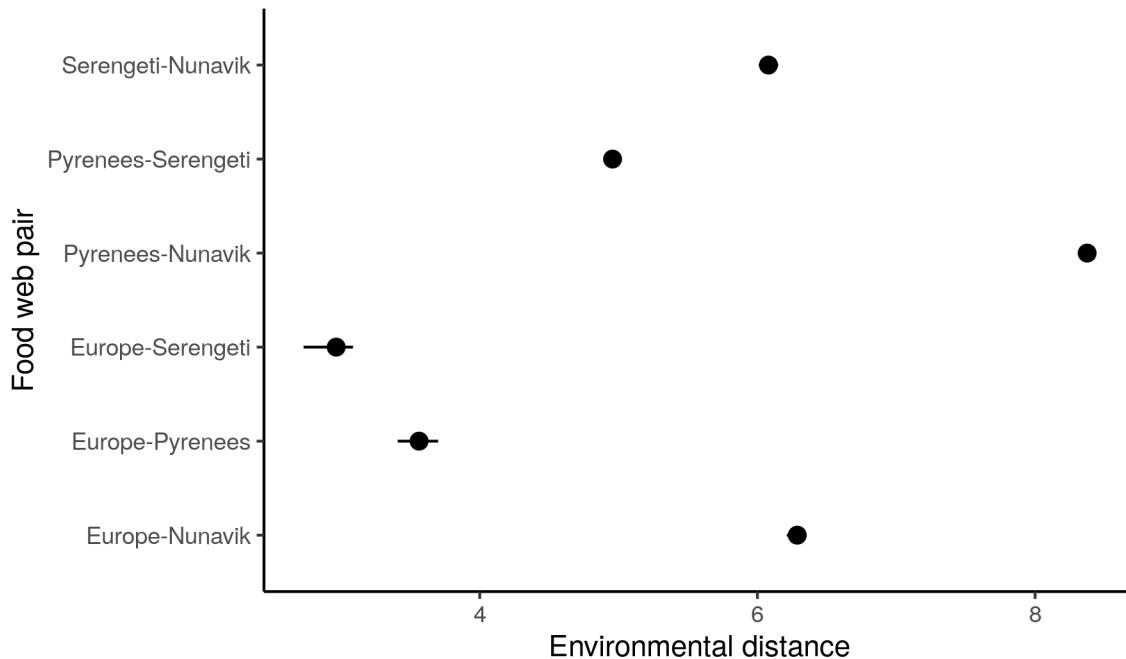
### **Food web properties:**

- **Connectance:** Connectance is defined as the proportion of interactions relative to the total number of possible interactions. Here, since we have a directed network with self-loop allowed, it is calculated as the number of interactions (L) divided by the squared the number of species ( $S^2$ ). Connectance relates to community sensitivity to perturbations and influence its dynamics (Dunne et al., 2002; Vieira & Almeida-Neto, 2015).
- **Mean and maximum trophic level:** The mean and maximum trophic level are calculated as the average and highest trophic levels among species within a food web. These measures “vertical” diversity, describing on how many trophic levels energy flows through in the food web. Vertical diversity interacts with horizontal diversity (within trophic level diversity) to influence community stability (Duffy et al., 2007; Zhao et al., 2019).
- **Motifs distribution:** The motif distribution is the frequency of motifs (Milo et al. 2002). Here, we focused on the 13 possible 3-species motifs. As mentioned in species role section, some motifs such as “linear chain”, “intraguild predation”, “direct competition”, and “apparent competition” influence the stability of food web (Borrelli, 2015; Stouffer & Bascompte, 2010). We used the function *motif* of the package *igraph* (Csardi & Nepusz 2006) to get the motif distribution of food webs.
- **Food web diameter:** The diameter is the longest of all shortest paths between any two species in the food web. The diameter influence how “efficiently” a perturbation spread thought the entire food web. We calculated food web diameter using the function *diameter* of the package *igraph* (Csardi & Nepusz 2006).

- **Number of clusters:** The number of clusters (i.e., modules) are the number of non-overlapping highly connected species. We determined the number of clusters using the function *cluster\_springlass* from the package *igraph* (Csardi & Nepusz 2006). Perturbations spread more efficiently within a cluster than across clusters. So, an higher number of clusters should increase resistance to perturbations.
- **Modularity:** Modularity quantify how food webs are divided into modules (groups of species interacting more together than with other species). Modularity increases food web persistence (Stouffer & Bascompte, 2011). We determined modules/clusters using the function *cluster\_springlass* from the package *igraph* (Csardi & Nepusz 2006).

#### Appendix 4: Environmental distance sensitivity analysis to random sampling

We sampled 10 times 500 points within the spatial domain of each food web. We extracted the 19 bioclimatic variables from WordClim (Hijmans, 2021). We then calculated environmental distance as the Euclidean distance between the food web bioclimatic centroids (mean of each bioclimatic variables) for the 10 iterations.



*Figure S4.1: Environmental distance estimates between the food webs included in our study. We repeated the random sampling of 500 spatial points within the domain of each food web 10 times and calculated the environmental distances between food webs for each sampling sets. The points are the mean environmental distances and the error bar are the range of estimates across sampling sets.*

The within food web pair variation is much lower than the across pairs distances. We conclude that sampling 500 random points is big enough to get an accurate estimate of the environmental distances between food webs.

## Appendix 5: Convergence of the predictive models

We evaluated convergence and adequacy of the four predictive models using rank plots, posterior predictive checks, and the rank-normalized potential scale reduction factor on split chains. Rank plots are similar to trace plots and are used to assess the mixing of the chains (Gelman et al., 2013). Posterior predictive checks compare the distribution of the observed data to the distribution of the posterior predictive distribution (Gelman et al., 2013). A discrepancy in distributions can be caused by a model missfit. Finally, the rank-normalized potential scale reduction factor on split chains (Rhat) is similar to the potential scale reduction factor, and is a summary statistics evaluating model convergence (Vehtari et al., 2021). Rhat close to 1 indicate a correctly converged model.

### *Northern Québec and Labrador model:*

- Rank plots of the population-level effects:

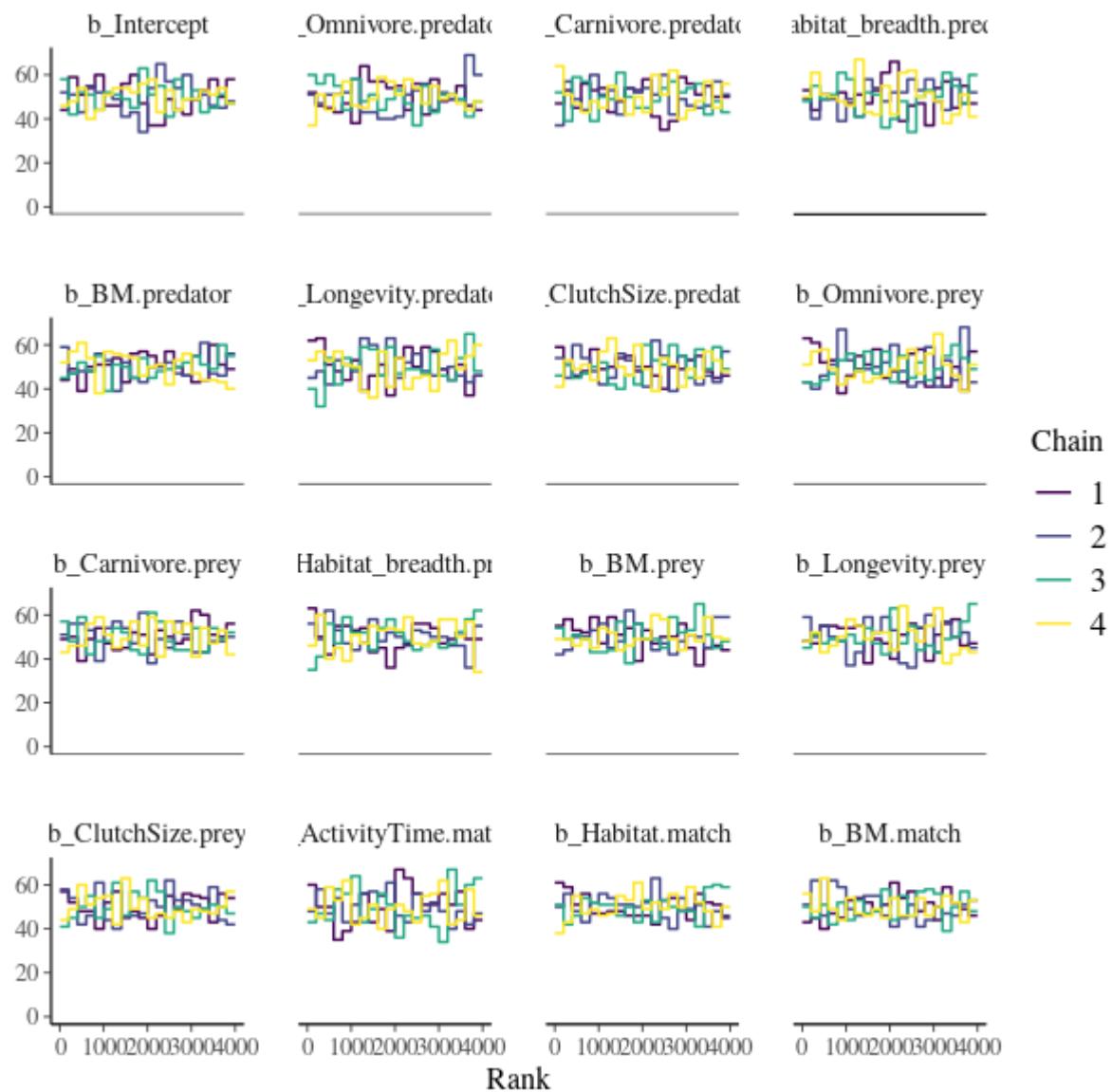


Figure S5.1: Rank plots of the population-level effects for the model calibrated on the Northern Québec and Labrador food web.

- Posterior predictive checks:

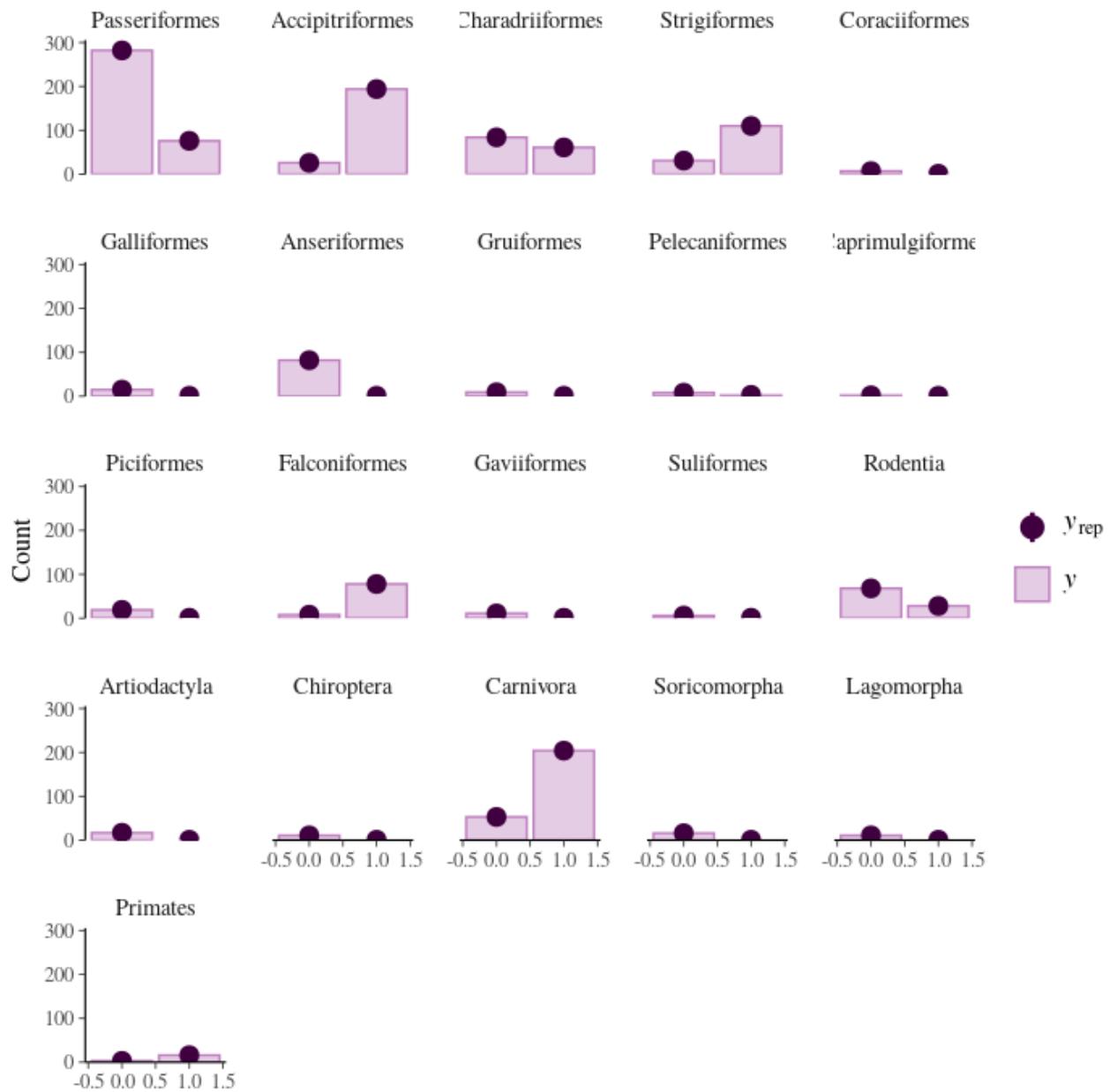
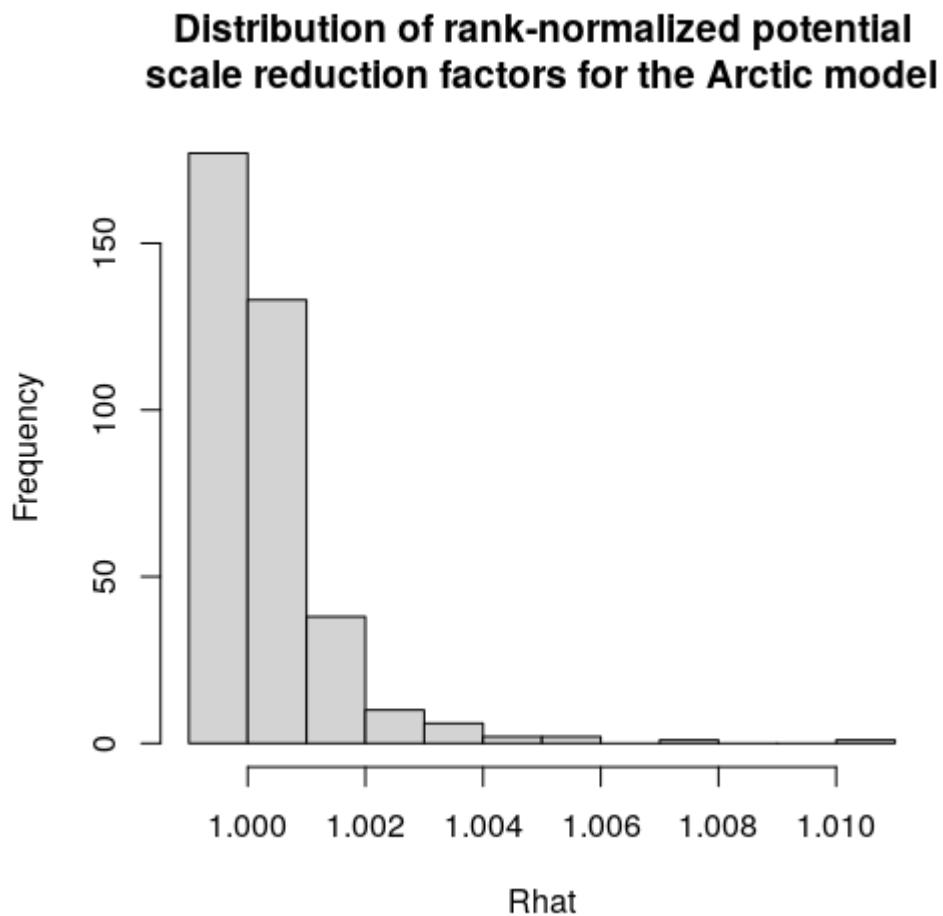


Figure S5.2: Comparison of the posterior predictive distribution ( $y_{rep}$ ; purple points) and the observed distribution ( $y$ ; purple bars) of the model calibrated on the Northern Québec and Labrador food web.

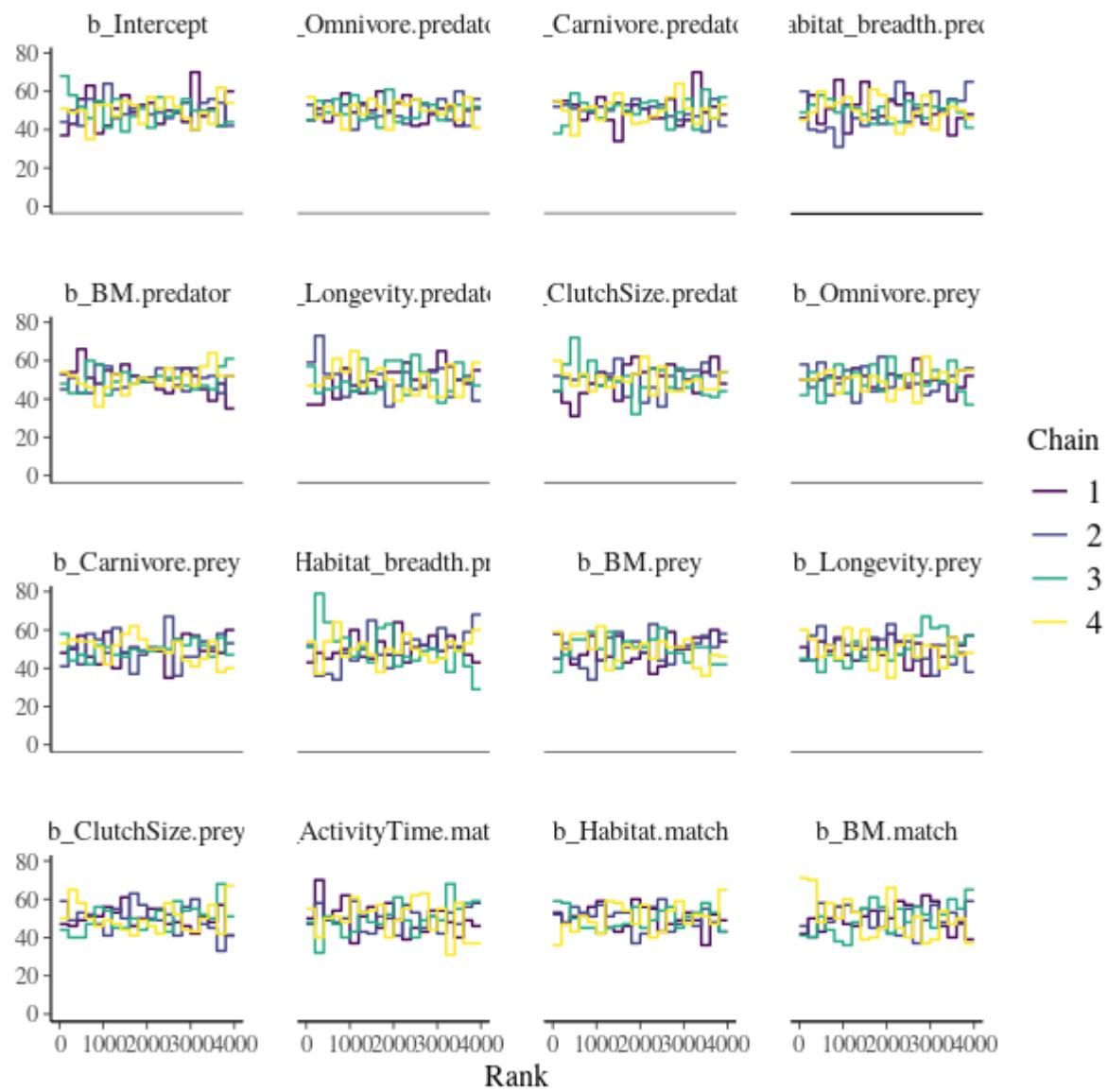
- Rank-normalized potential scale reduction factors on split chains:



*Figure S5.3: Distribution of the rank-normalized potential scale reduction factors on split chains for the Northern Québec and Labrador food web*

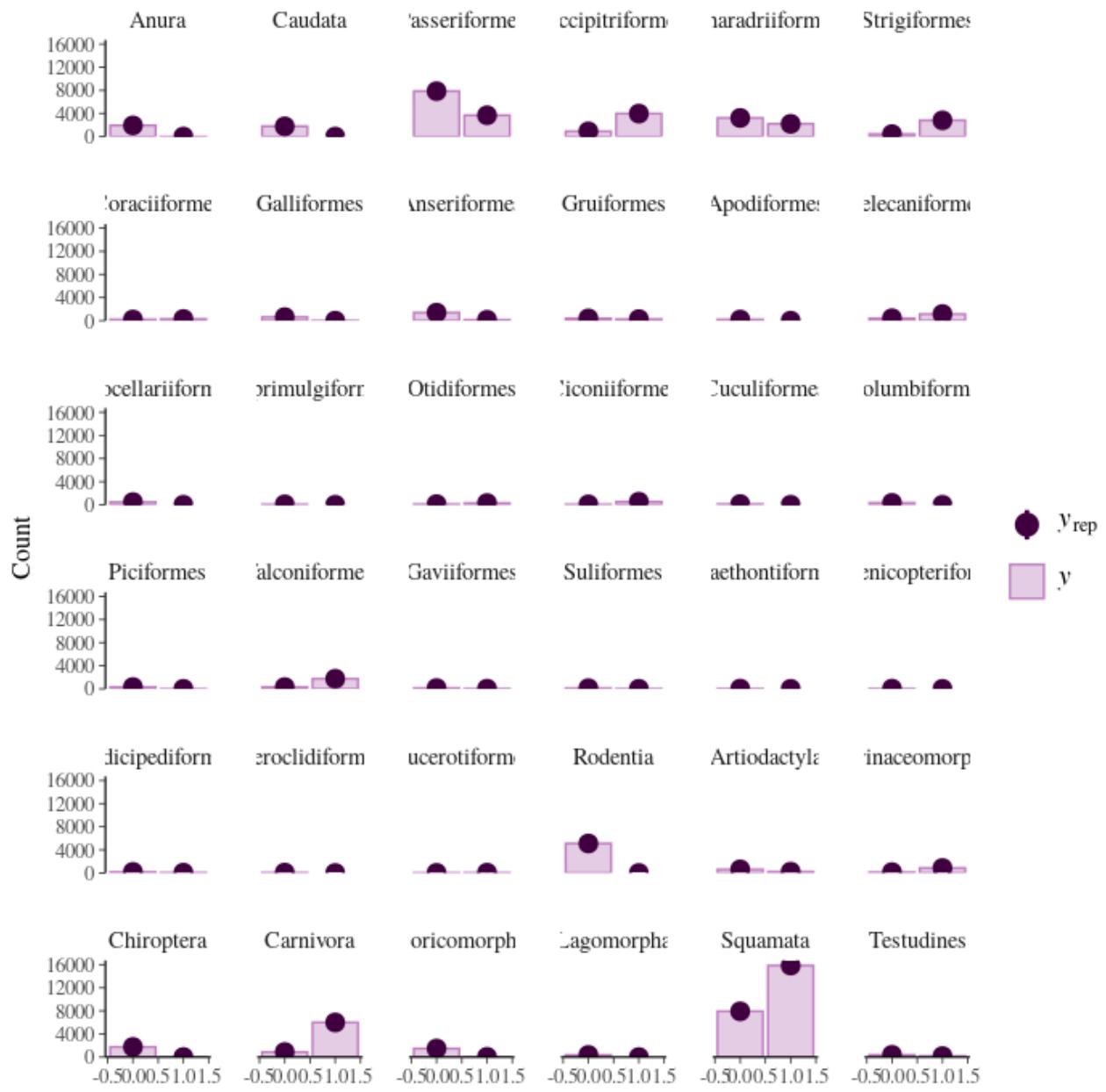
*Europe model:*

- Rank plots of the population-level effects:



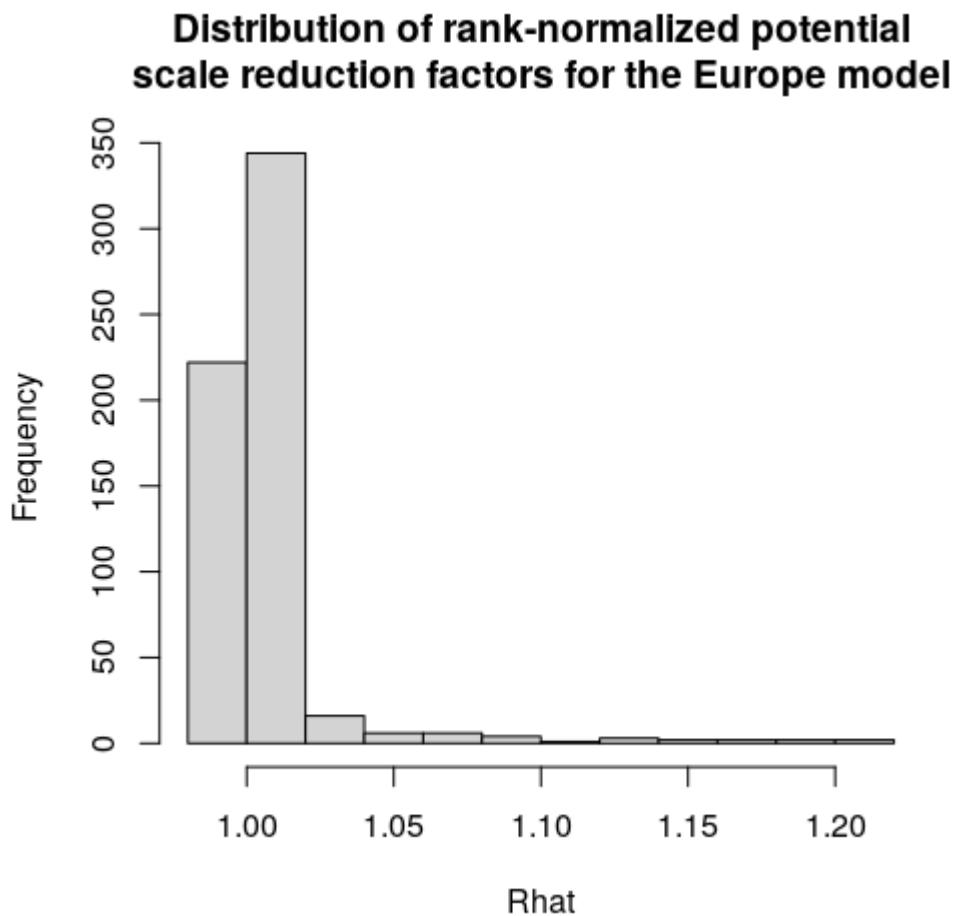
*Figure S5.4: Rank plots of the population-level effects for the model calibrated on the European food web of terrestrial vertebrates.*

- Posterior predictive checks:



*Figure S5.5: Comparison of the posterior predictive distribution ( $y_{rep}$ ; purple points) and the observed distribution ( $y$ ; purple bars) of the model calibrated on the food web of European terrestrial vertebrates.*

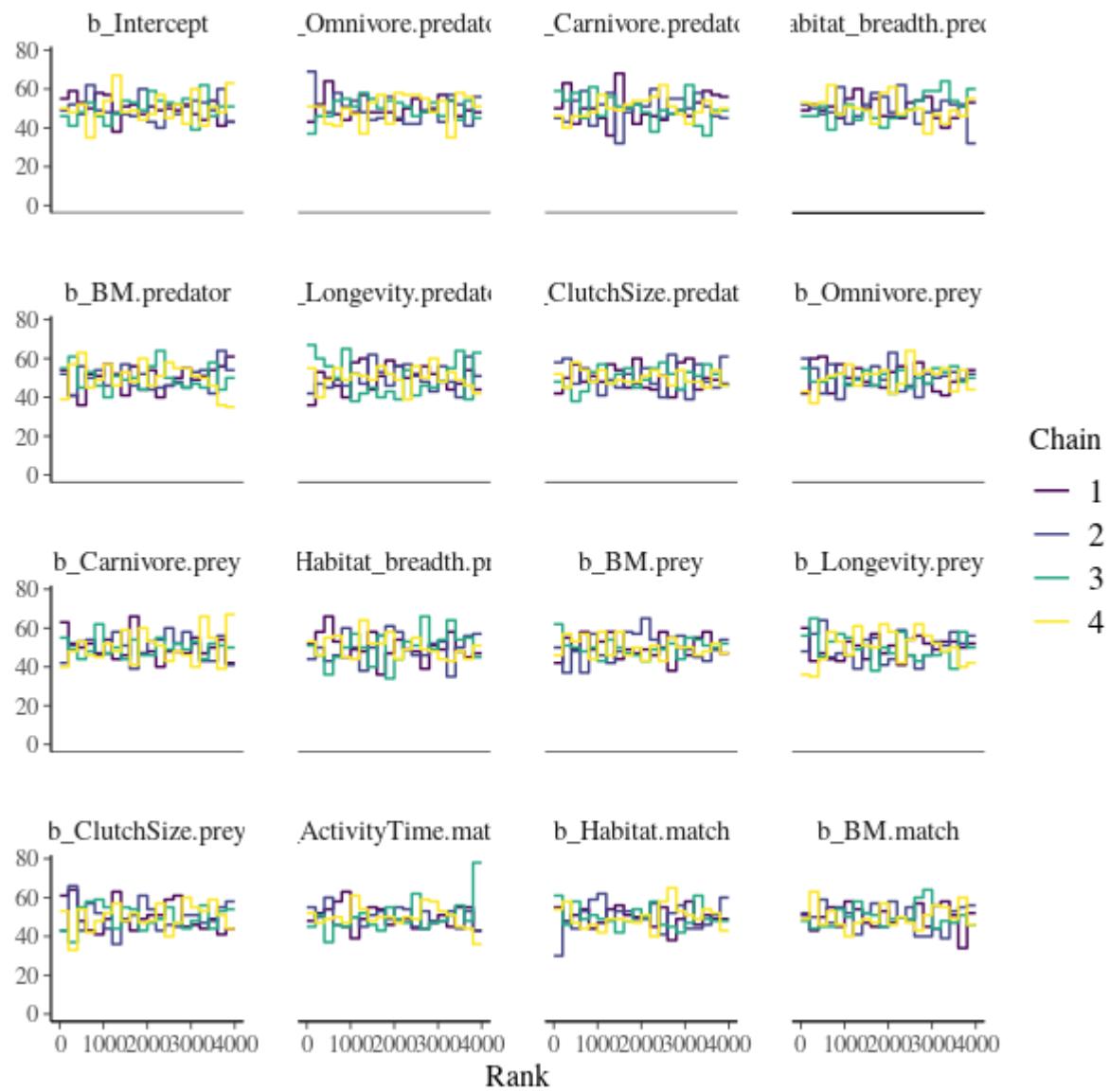
- Potential scale reduction factors:



*Figure S5.6: Distribution of the rank-normalized potential scale reduction factors on split chains for the food web of European terrestrial vertebrates.*

*Pyrenees model:*

- Rank plots of the population-level effects:



*Figure S5.7: Rank plots of the population-level effects for the model calibrated on the Pyrenees food web.*

- Posterior predictive checks:

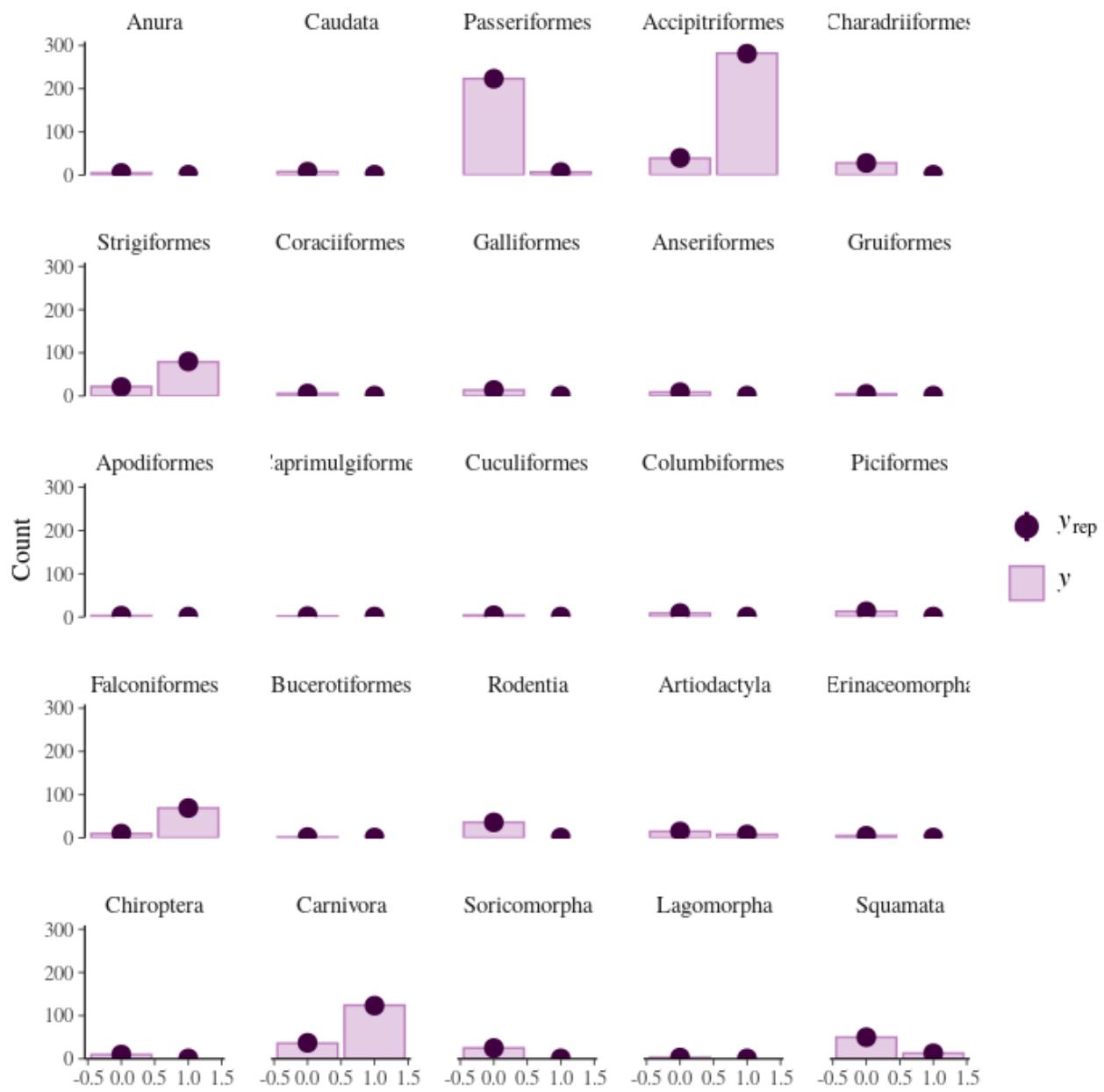


Figure S5.8: Comparison of the posterior predictive distribution ( $y_{rep}$ ; purple points) and the observed distribution ( $y$ ; purple bars) of the model calibrated on the Pyrenees food web.

- Potential scale reduction factors:

### Distribution of rank-normalized potential scale reduction factors for the Pyrenees model

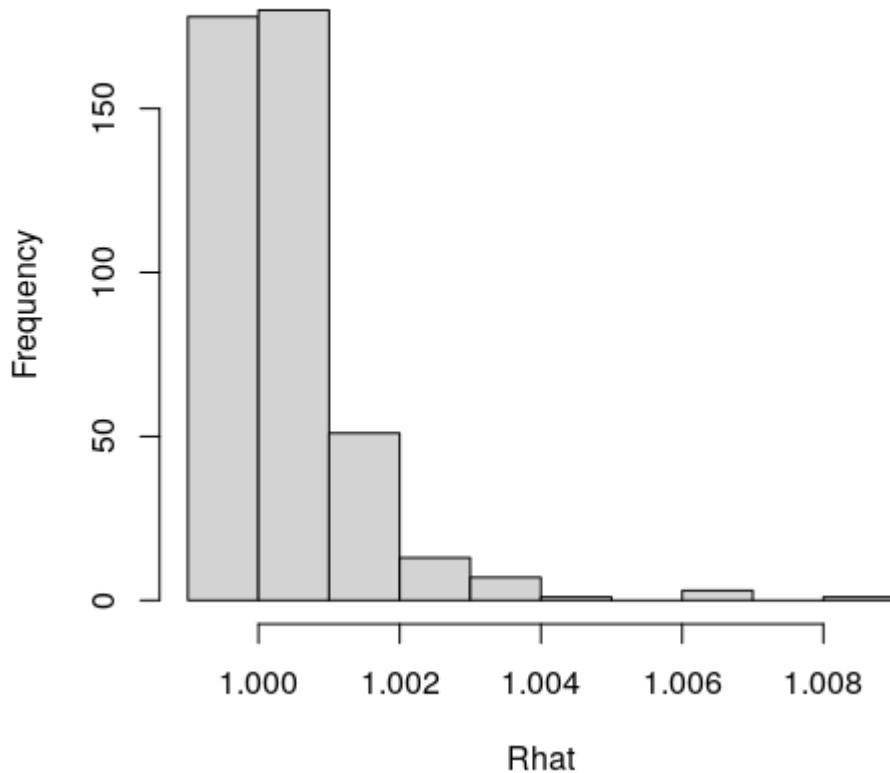
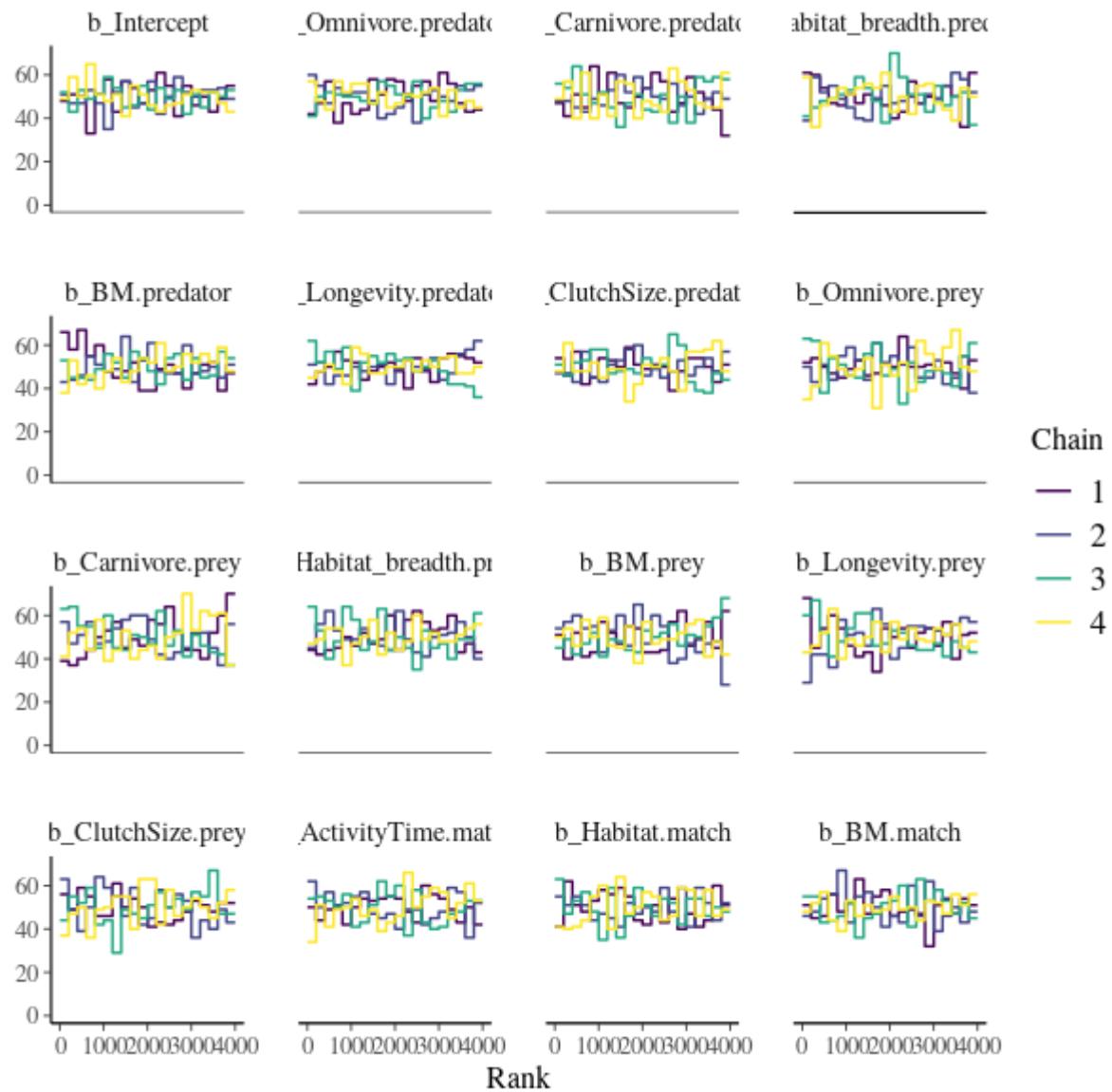


Figure S5.9: Distribution of the rank-normalized potential scale reduction factors on split chains for the Pyrenees food web

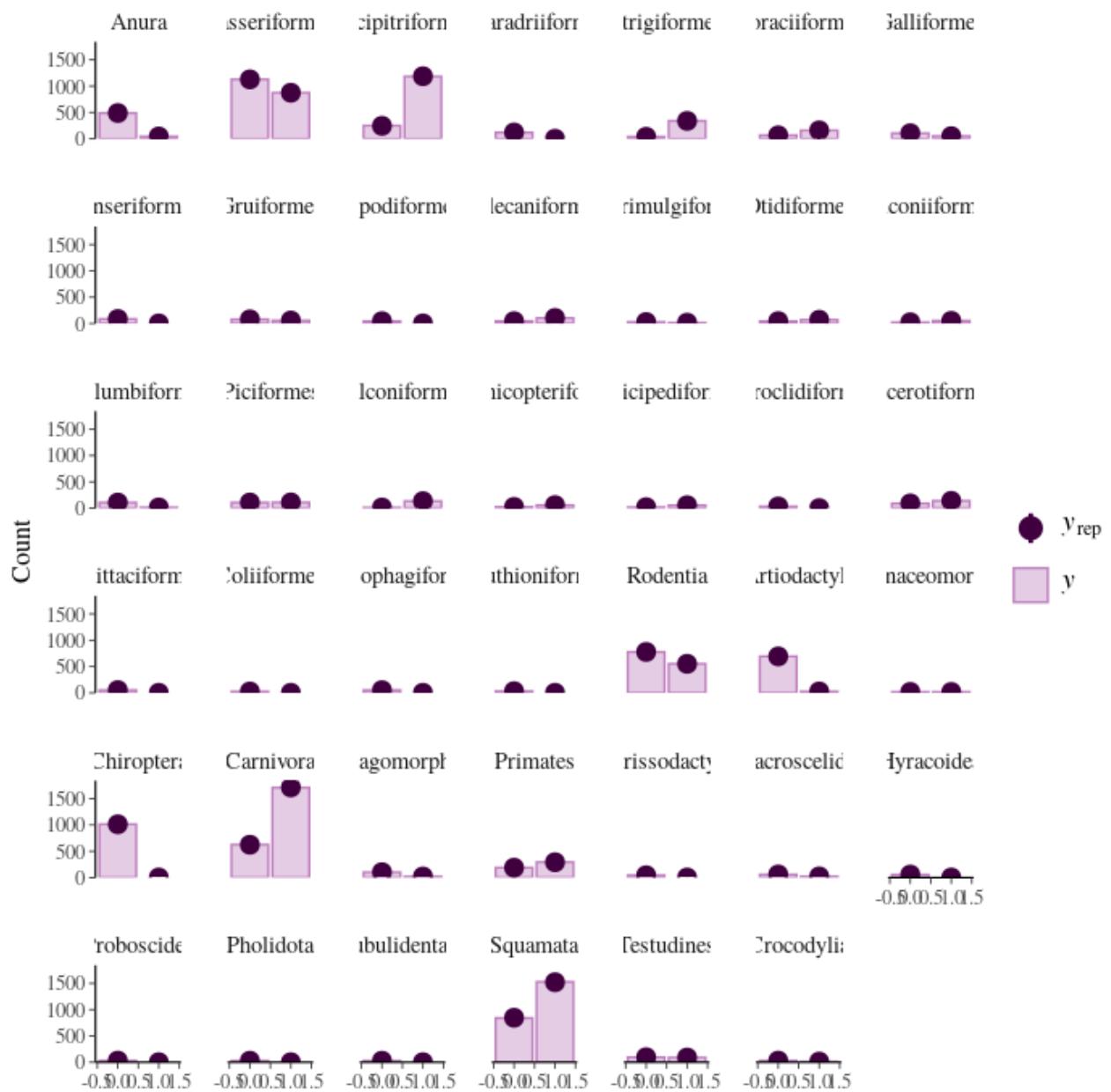
*Serengeti model:*

- Rank plots of the population-level effects:



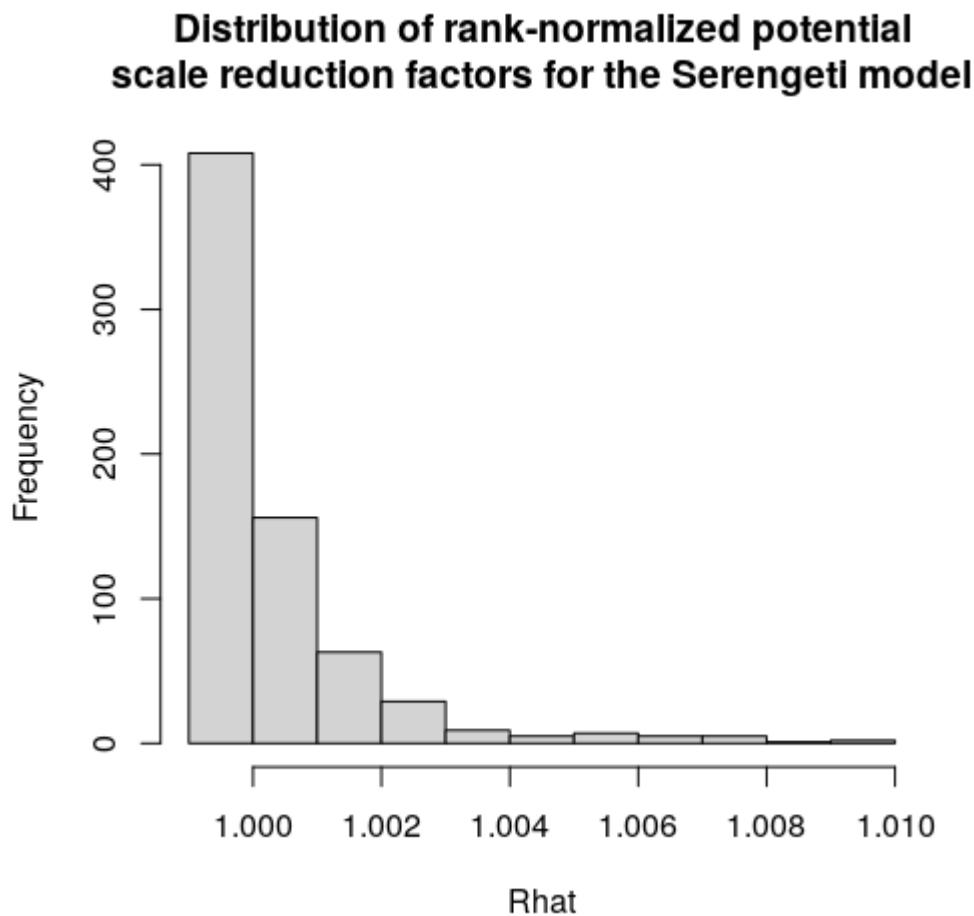
*Figure S5.10: Rank plots of the population-level effects for the model calibrated on the Serengeti food web.*

- Posterior predictive checks:



*Figure S5.11: Comparison of the posterior predictive distribution ( $y_{rep}$ ; purple points) and the observed distribution ( $y$ ; purple bars) of the model calibrated on the Serengeti food web.*

- Potential scale reduction factors:

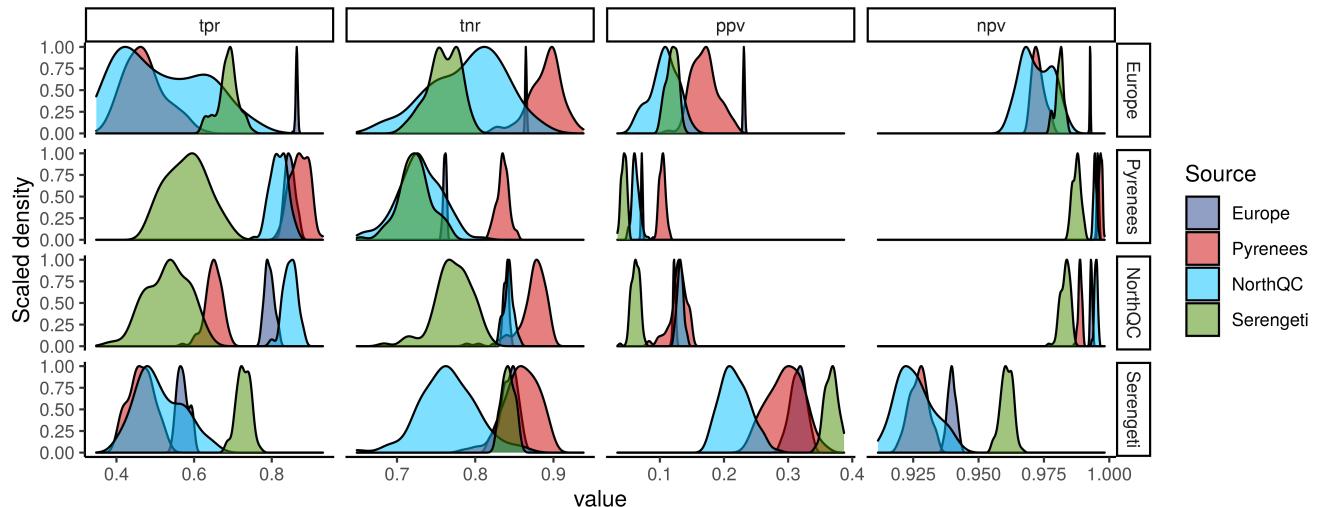


*Figure S5.12: Distribution of the rank-normalized potential scale reduction factors on split chains for the Serengeti food web*

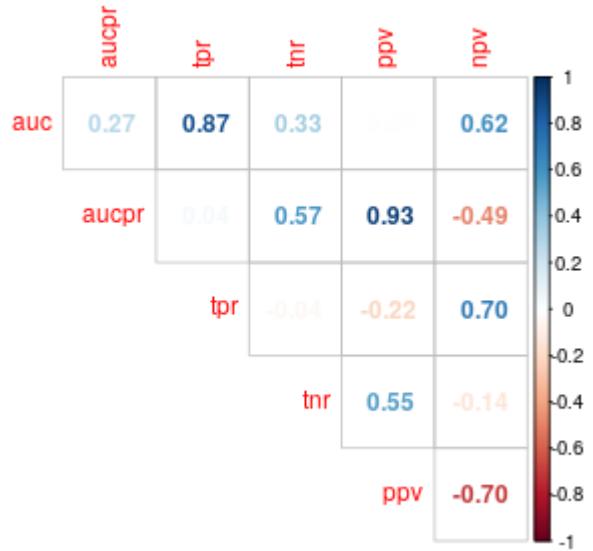
## Appendix 6: Predicting species interactions extended results

*Table S6.1: Area under the precision-recall curve (auprc) of each model predicting every food webs. The baseline for auprc (value for random predictions) depends on the prevalence of interactions in the validation dataset (Europe: 0.05; Pyrenees: 0.02; North Québec and Labrador: 0.03; Serengeti: 0.12). Consequently, model performance can be compared for a given predicted food web (within a column) but not across food webs (across columns).*

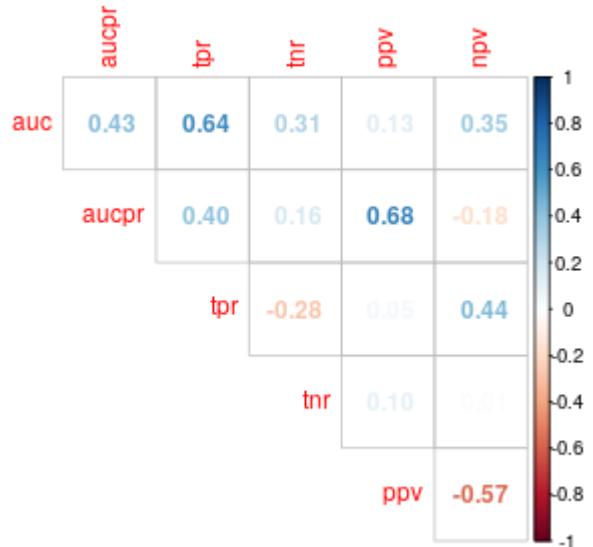
Model	Food web			
	Europe	Pyrenees	North Québec and Labrador	Serengeti
Europe	<b>0.47</b>	0.09	0.19	0.37
Pyrenees	0.23	<b>0.22</b>	0.2	0.37
North Québec and Labrador	0.18	0.15	<b>0.27</b>	0.32
Serengeti	0.18	0.06	0.08	<b>0.56</b>



*Figure S6.1: Scaled density distribution of true positive rate (tpr), true negative rate (tnr), positive predictive value (ppv), and negative predictive value (npv) of each model (Source) predicting every food webs (rows). tpr is the proportion of correctly predicted interactions, tnr is the proportion of correctly predicted non-interactions, ppv is the proportion of predicted interactions that are observed interactions, and npv is the proportion of predicted non-interactions that are observed non-interactions. We calculated tpr, tnr, ppv, and npv with 100 samples of the posterior distribution of the entire food webs.*



*Figure S6.2: Correlation between performance metrics for predicting the interactions of the entire food webs. auc is the area under the receiver operating curve, aucpr is the area under the precision-recall curve, tpr is the true positive rate, tnr is the true negative rate, ppv is the positive predictive value, and npv is the negative predictive value*



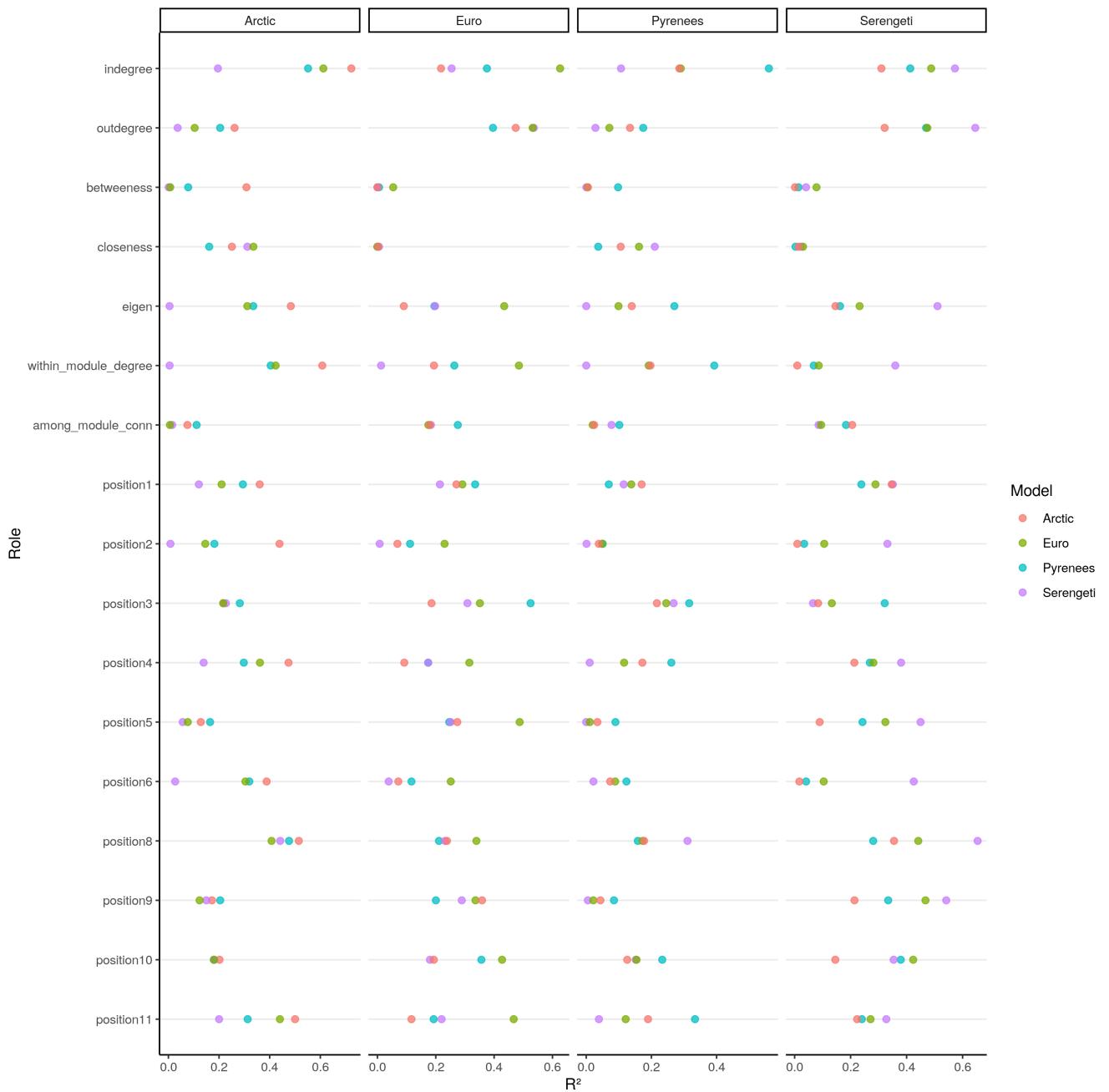
*Figure S6.3: Correlation between performance metrics for predicting the interactions of each species within a food web. See caption of figure S6.2 for abbreviations.*

## Model performance in relationship with distance:

Here, we present the outputs of the model relating model performance to geographic, phylogenetic, and environmental distances:

- Model 1: Total effect of geographic distance on AUC  
 $\text{logitauc} \sim \text{geo.dist\_sc} + (1|\text{Model}) + (1|\text{Food web})$   
fixed effects:  
Intercept: Estimate = 1.73, 95% CrI = [0.71, 2.30]  
 $\text{geo.dist\_sc}$ : Estimate = -0.54, 95% CrI = [-0.79, -0.30]  
random effect:  
 $\text{sd}(1|\text{Model})$ : Estimate = 0.38, 95% CrI = [0.01, 1.67]  
 $\text{sd}(1|\text{Food web})$ : Estimate = 0.41, 95% CrI = [0.01, 1.160]
- Model 2: Partial effect of geographic distance on AUC  
 $\text{logitauc} \sim \text{geo.dist\_sc} + \text{phylo.dist\_sc} + \text{env.dist\_sc} + (1|\text{Model}) + (1|\text{Food web})$   
fixed effects:  
Intercept: Estimate = 1.84, 95% CrI = [1.17, 2.26]  
 $\text{geo.dist\_sc}$ : Estimate = -0.07, 95% CrI = [-0.48, 0.34]  
 $\text{phylo.dist\_sc}$ : Estimate = -0.20, 95% CrI = [-0.63, 0.20]  
 $\text{env.dist\_sc}$ : Estimate = -0.36, 95% CrI = [-0.84, 0.17]  
random effect:  
 $\text{sd}(1|\text{Model})$ : Estimate = 0.25, 95% CrI = [0.01, 0.99]  
 $\text{sd}(1|\text{Food web})$ : Estimate = 0.25, 95% CrI = [0.01, 1.10]
- Model 3: Effect of phylogenetic distance on AUC after controlling for geographic distance  
 $\text{logitauc} \sim \text{phylo.dist\_sc} + \text{geo.dist\_sc} + (1|\text{Source}) + (1|\text{Target})$   
fixed effects:  
Intercept: Estimate = 1.75, 95% CrI = [0.88, 2.29]  
 $\text{geo.dist\_sc}$ : Estimate = -0.26, 95% CrI = [-0.59, 0.08]  
 $\text{phylo.dist\_sc}$ : Estimate = -0.39, 95% CrI = [-0.78, -0.03]  
random effect:  
 $\text{sd}(1|\text{Model})$ : Estimate = 0.40, 95% CrI = [0.02, 1.44]  
 $\text{sd}(1|\text{Food web})$ : Estimate = 0.31, 95% CrI = [0.01, 1.24]
- Model 4: Effect of environmental distance on AUC after controlling for geographic distance  
 $\text{logitauc} \sim \text{env.dist\_sc} + \text{geo.dist\_sc} + (1|\text{Source}) + (1|\text{Target})$   
fixed effects:  
Intercept: Estimate = 1.69, 95% CrI = [0.40, 2.23]  
 $\text{geo.dist\_sc}$ : Estimate = -0.08, 95% CrI = [-0.49, 0.29]  
 $\text{env.dist\_sc}$ : Estimate = -0.52, 95% CrI = [-0.89, -0.11]  
random effect:  
 $\text{sd}(1|\text{Model})$ : Estimate = 0.23, 95% CrI = [0.01, 1.13]  
 $\text{sd}(1|\text{Food web})$ : Estimate = 0.35, 95% CrI = [0.01, 1.51]

## Appendix 7: Predicting species role extended results



**Figure S7.1:** Predictive performance of each model (colors) to predict species role for every food webs (column). Performance is measured with the coefficient of determination ( $R^2$ ). Arctic refers to the Northern Québec and Labrador food web. From top to bottom: Number of prey, number of predator, betweenness, closeness, eigenvector centrality, within-module degree, participation coefficient, frequency of predator position in a linear chain motif, consumer position in a linear chain motif, resource in a linear chain motif, omnivore in a intraguild predation motif, consumer in a intraguild predation motif, resource in a intraguild predation motif, predation in a direct competition motif, prey in a direct competition motif, predator in a apparent competition motif, and prey in a apparent competition motif.

## Number of prey

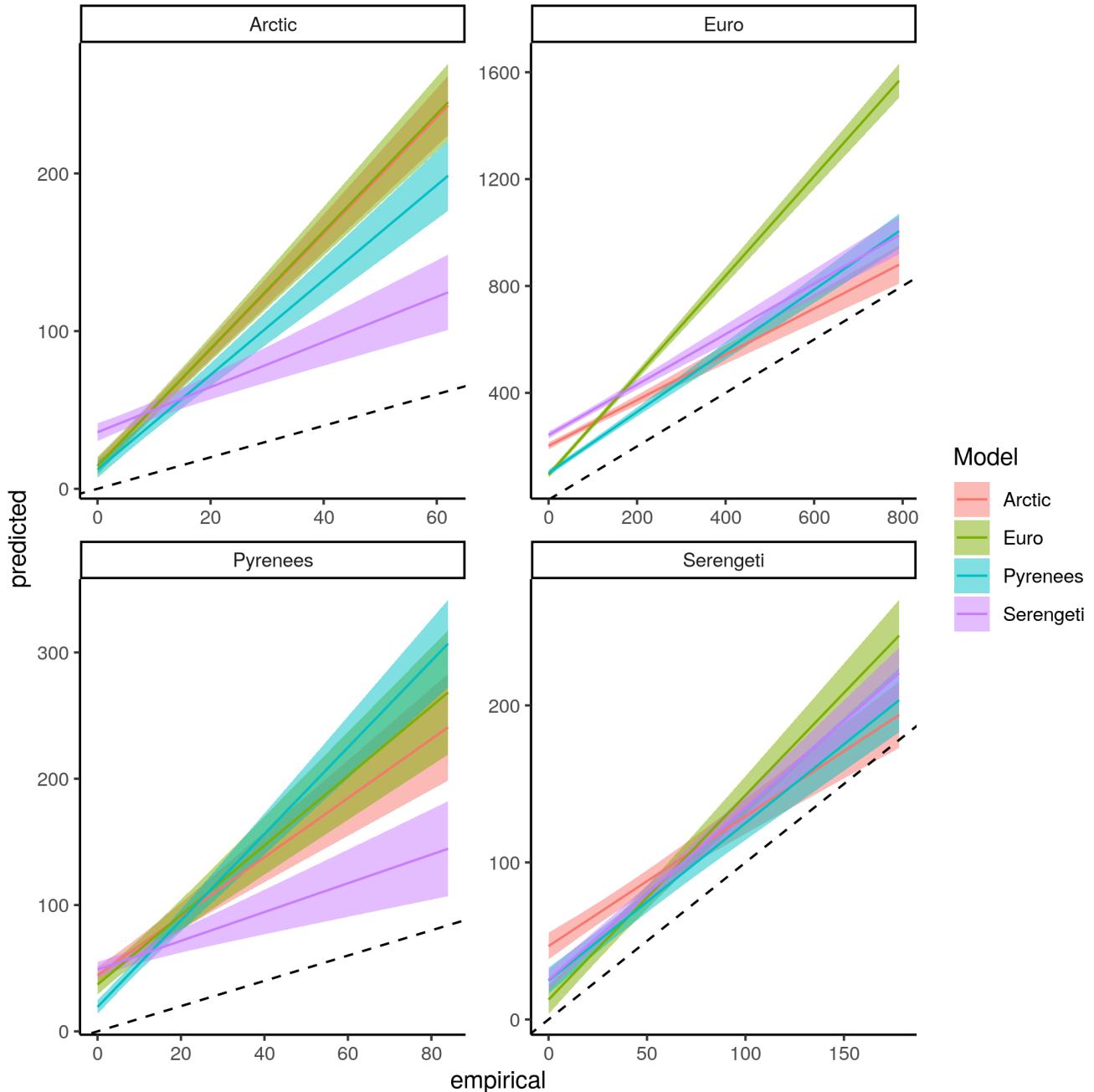


Figure S7.2: Linear regression comparing the predicted number of prey and empirical number of prey of species for each model predicting every food webs.

## Number of predators

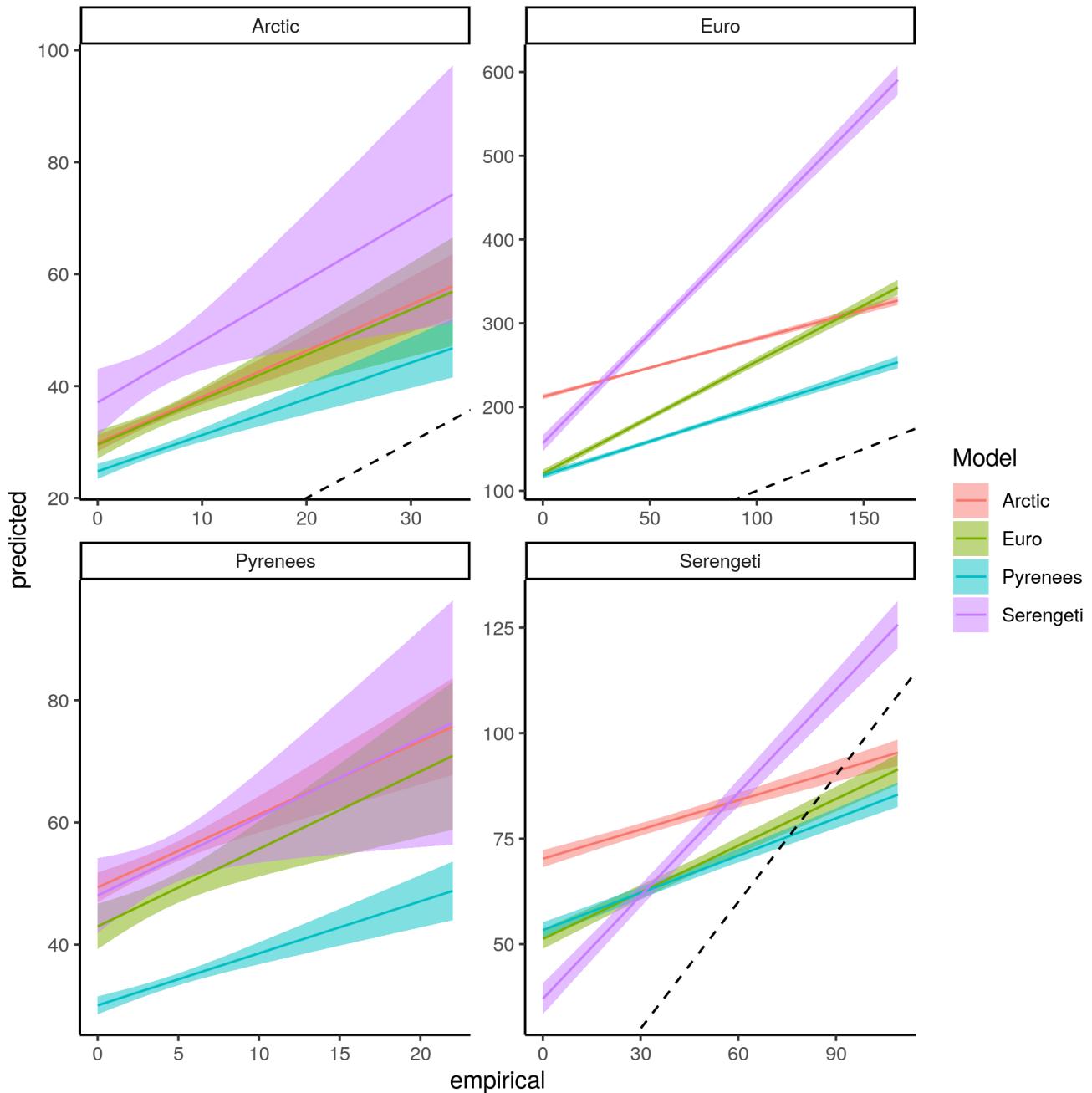


Figure S7.3: Linear regression comparing the predicted number of predators and empirical number of predators of species for each model predicting every food webs.

## Betweenness

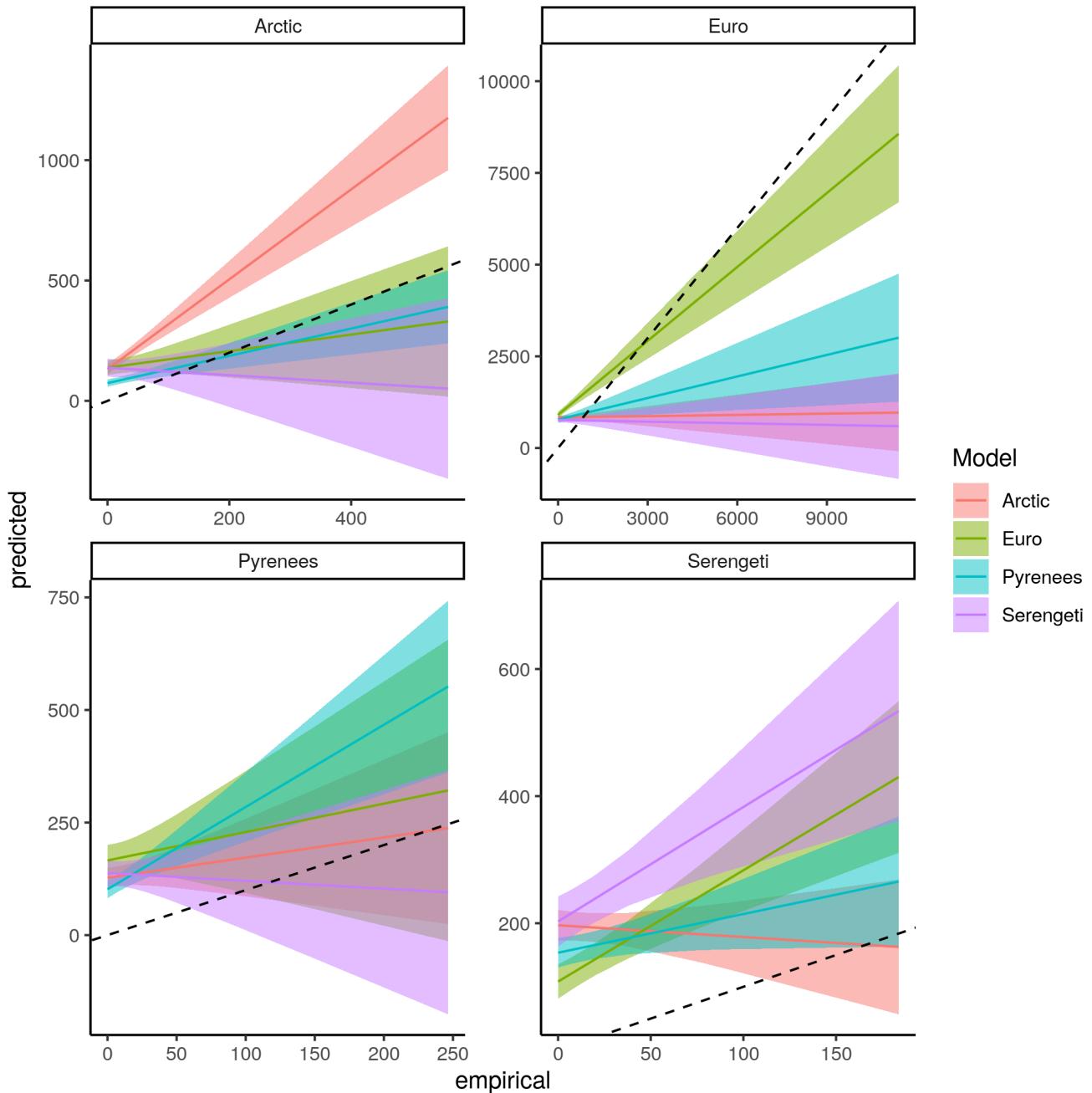


Figure S7.4: Linear regression comparing the predicted betweenness centrality and empirical betweenness centrality of species for each model predicting every food webs.

## Closeness

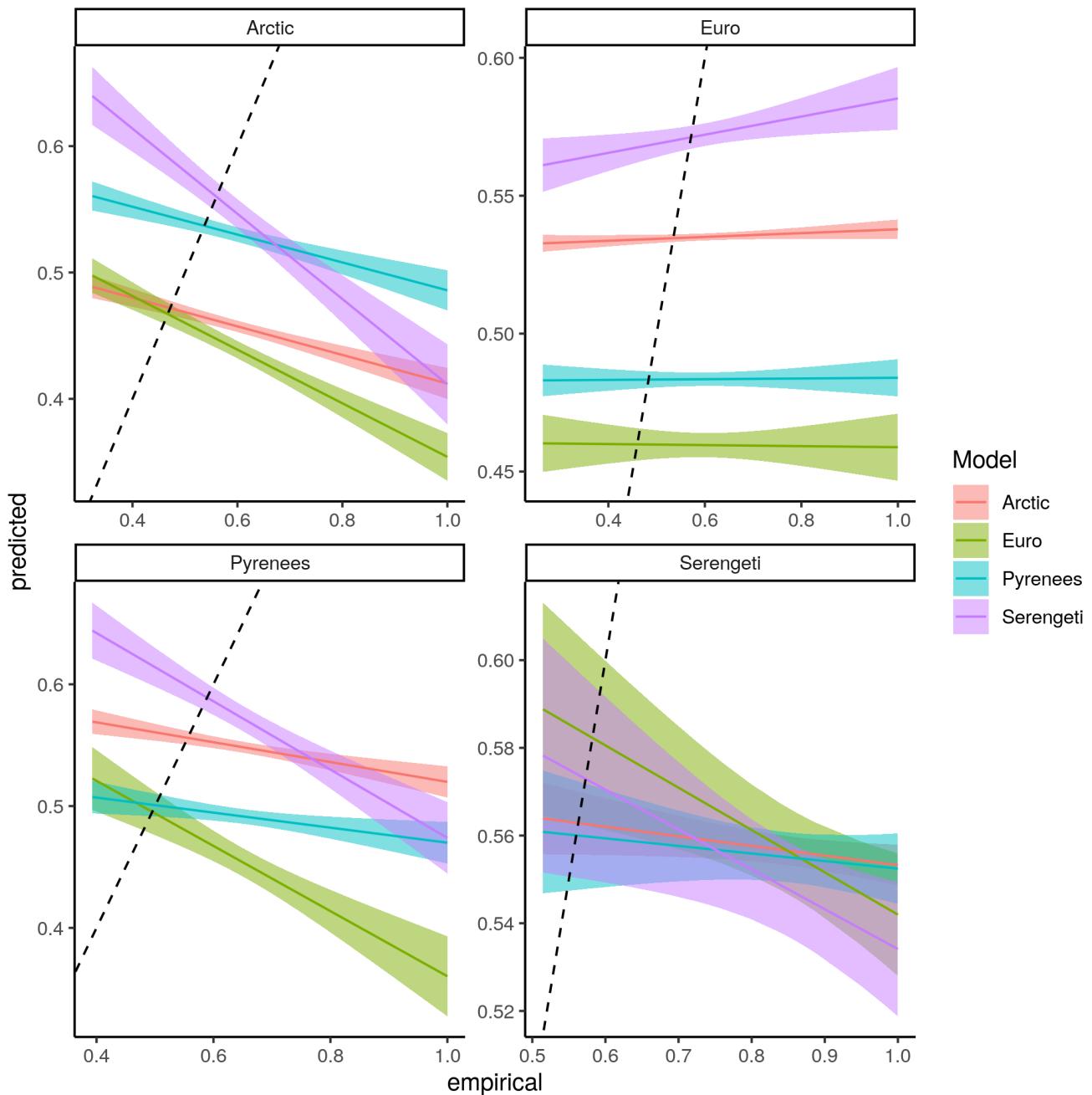


Figure S7.5: Linear regression comparing the predicted closeness centrality and empirical closeness centrality of species for each model predicting every food webs.

## Eigenvector centrality

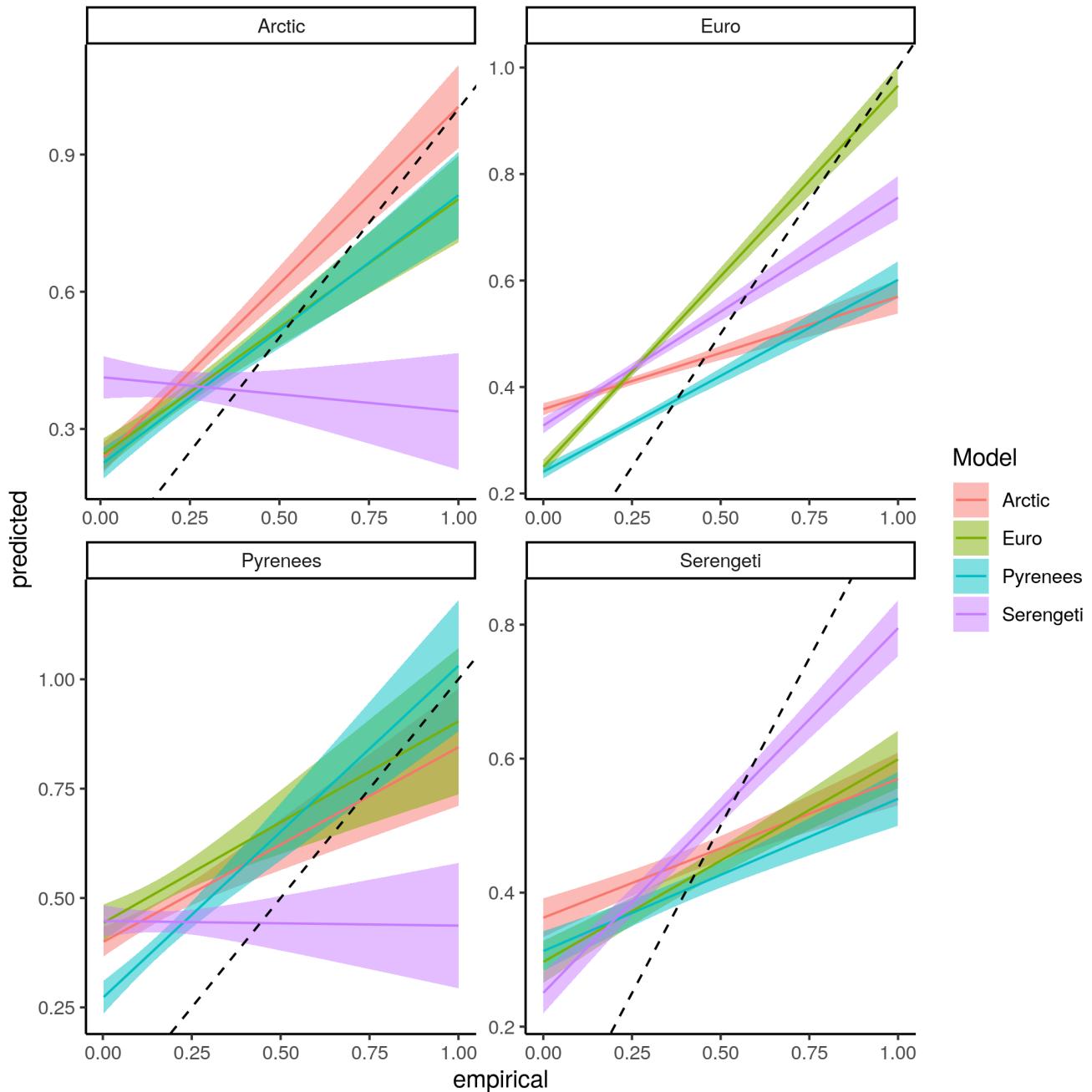


Figure S7.6: Linear regression comparing the predicted eigenvector centrality and empirical eigenvector centrality of species for each model predicting every food webs.

## Trophic Level

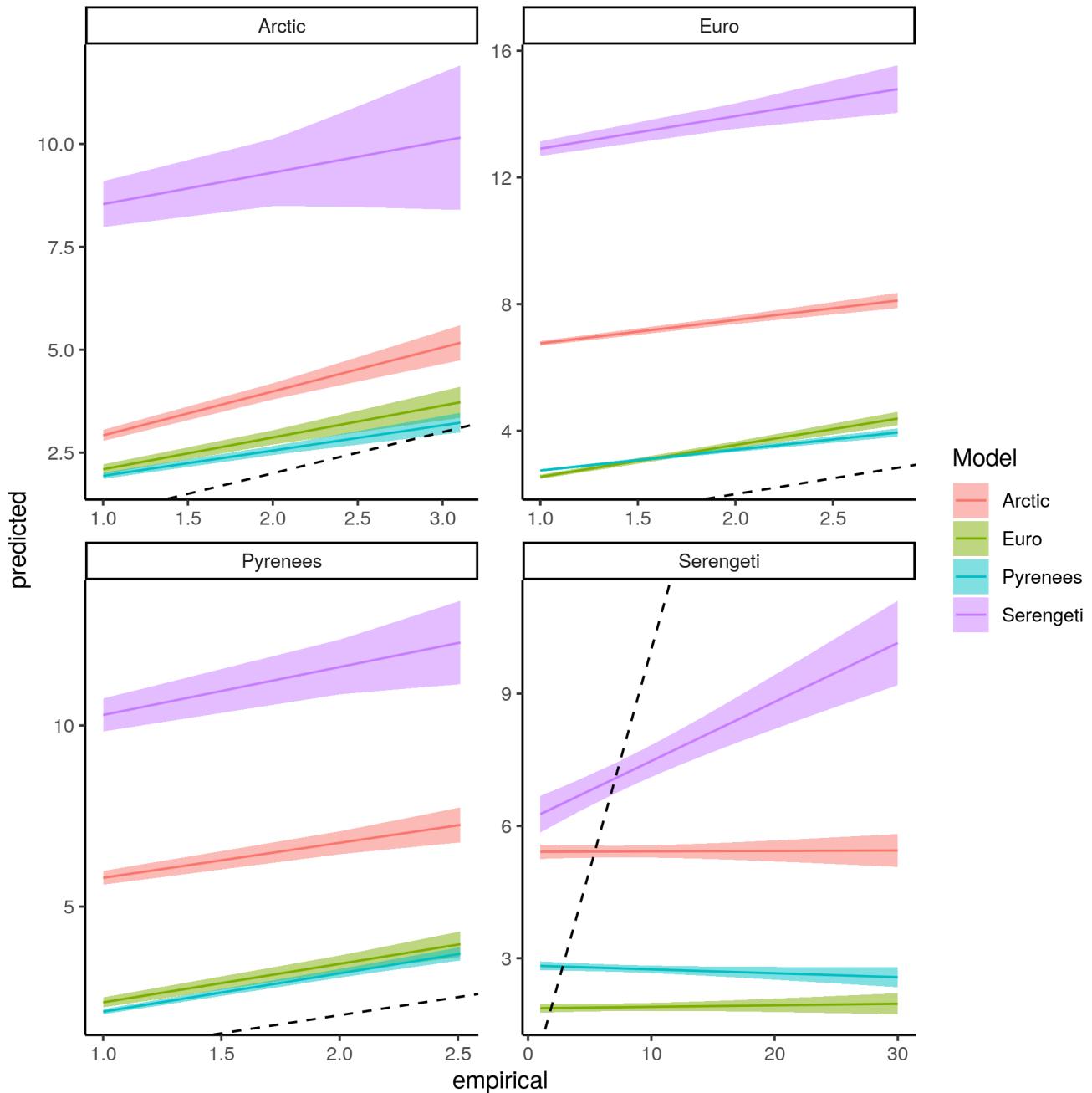


Figure S7.7: Linear regression comparing the predicted trophic level and empirical trophic level of species for each model predicting every food webs.

## Omnivory

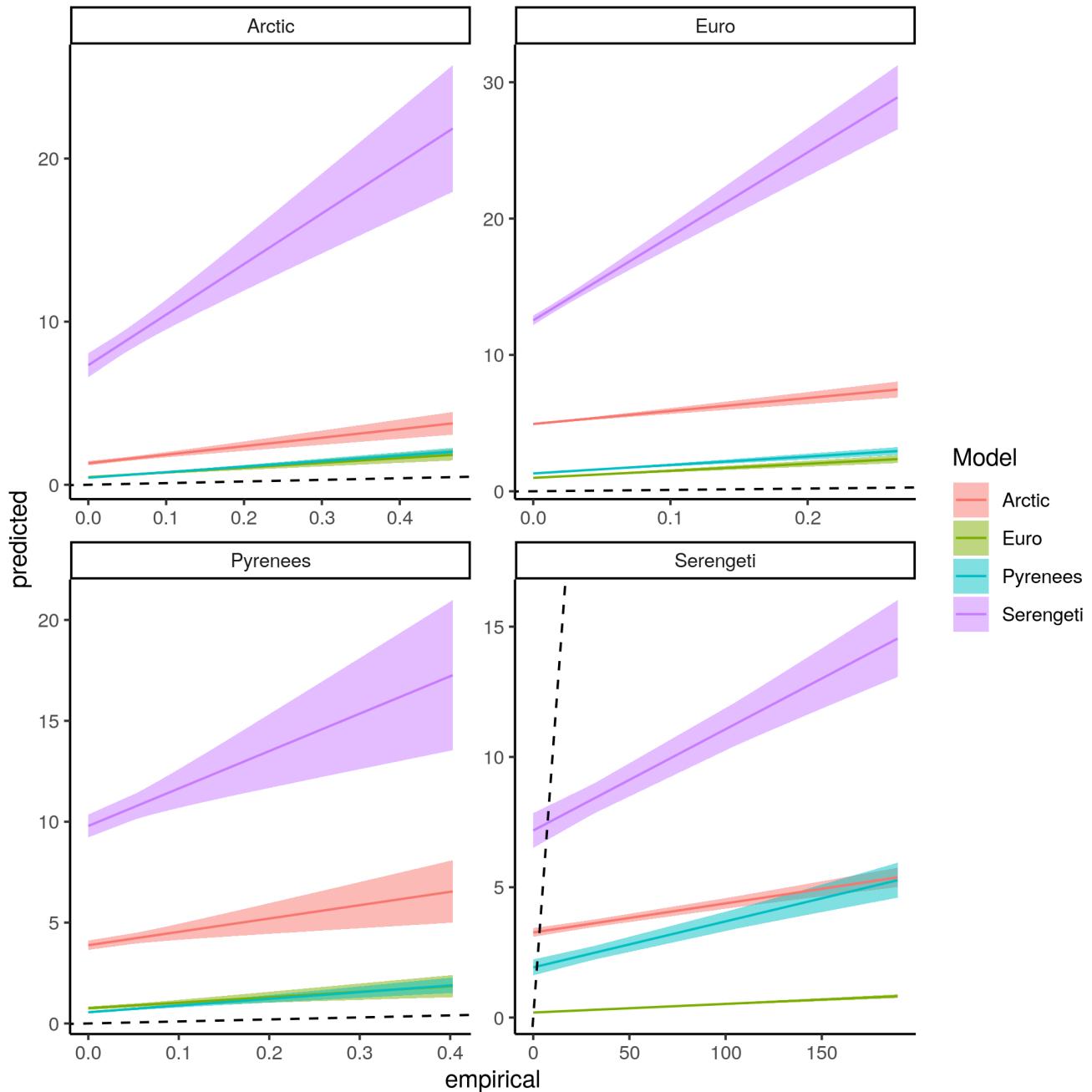


Figure S7.8: Linear regression comparing the predicted omnivory index and empirical omnivory index of species for each model predicting every food webs.

### Within-module degree

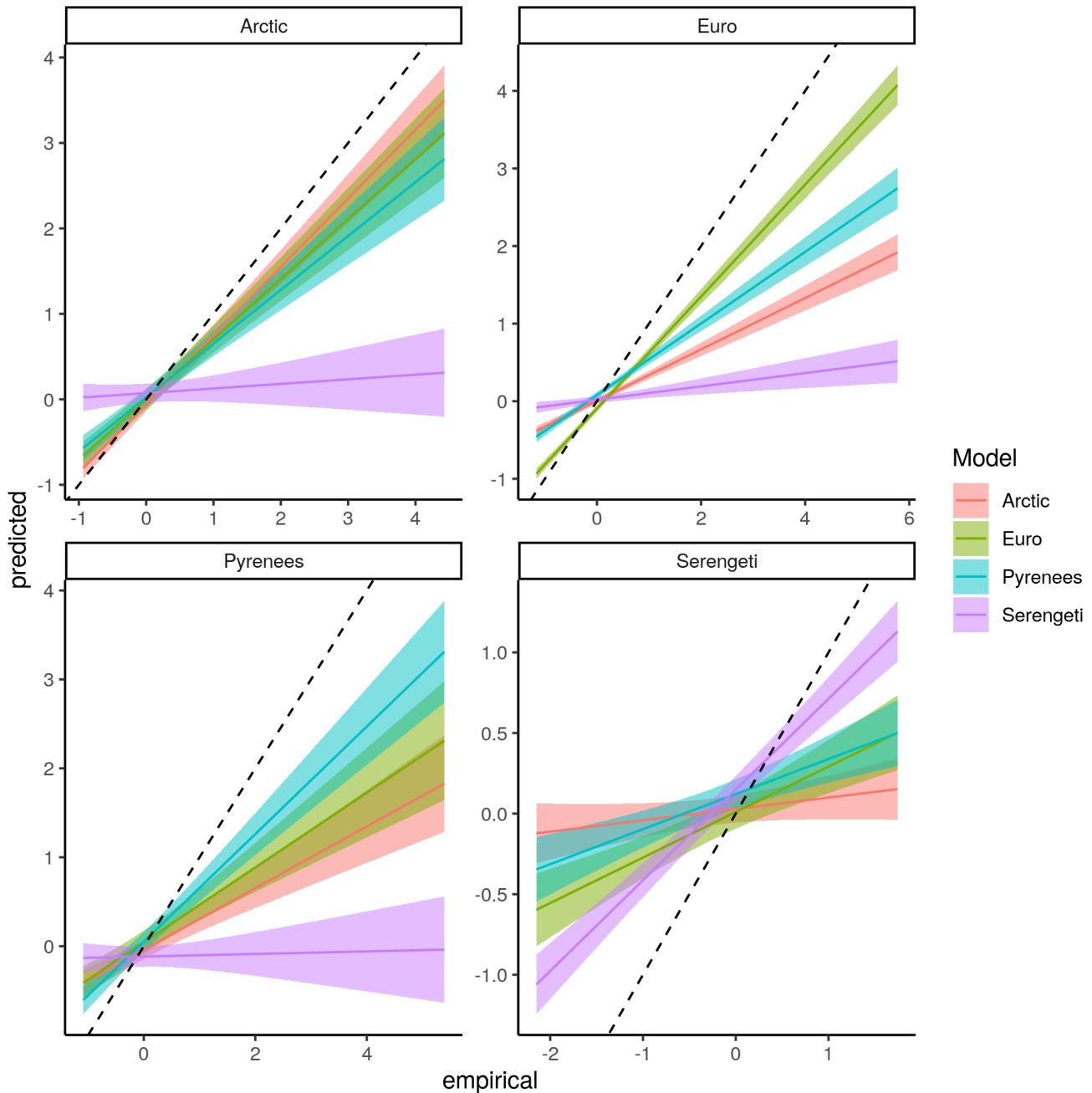


Figure S7.9: Linear regression comparing the predicted within-module degree and empirical within-module degree of species for each model predicting every food webs.

## Participation coefficient

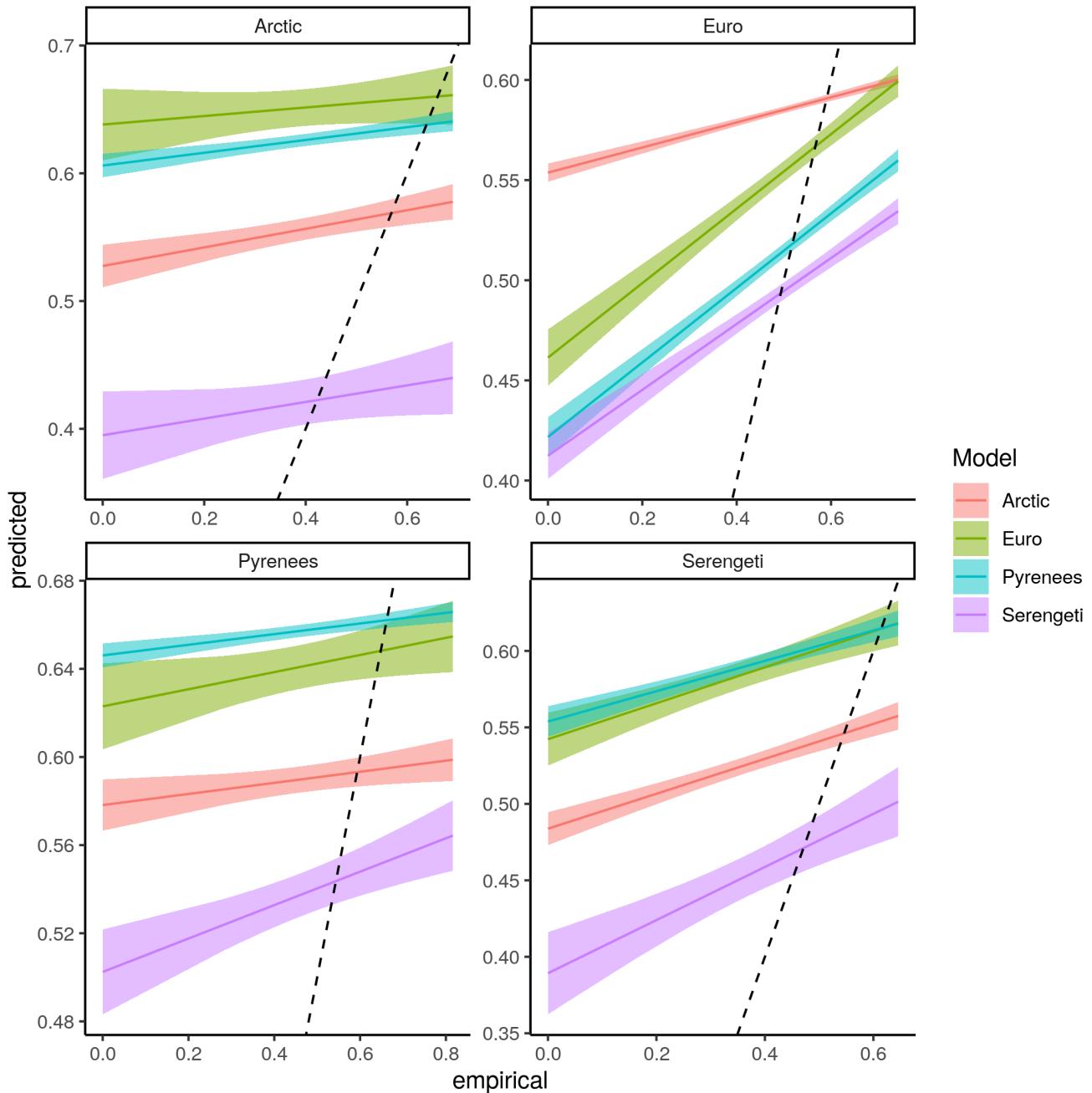


Figure S7.10: Linear regression comparing the predicted participation coefficient and empirical participation coefficient of species for each model predicting every food webs.

### Frequency of the top predator position in the linear food chain motif

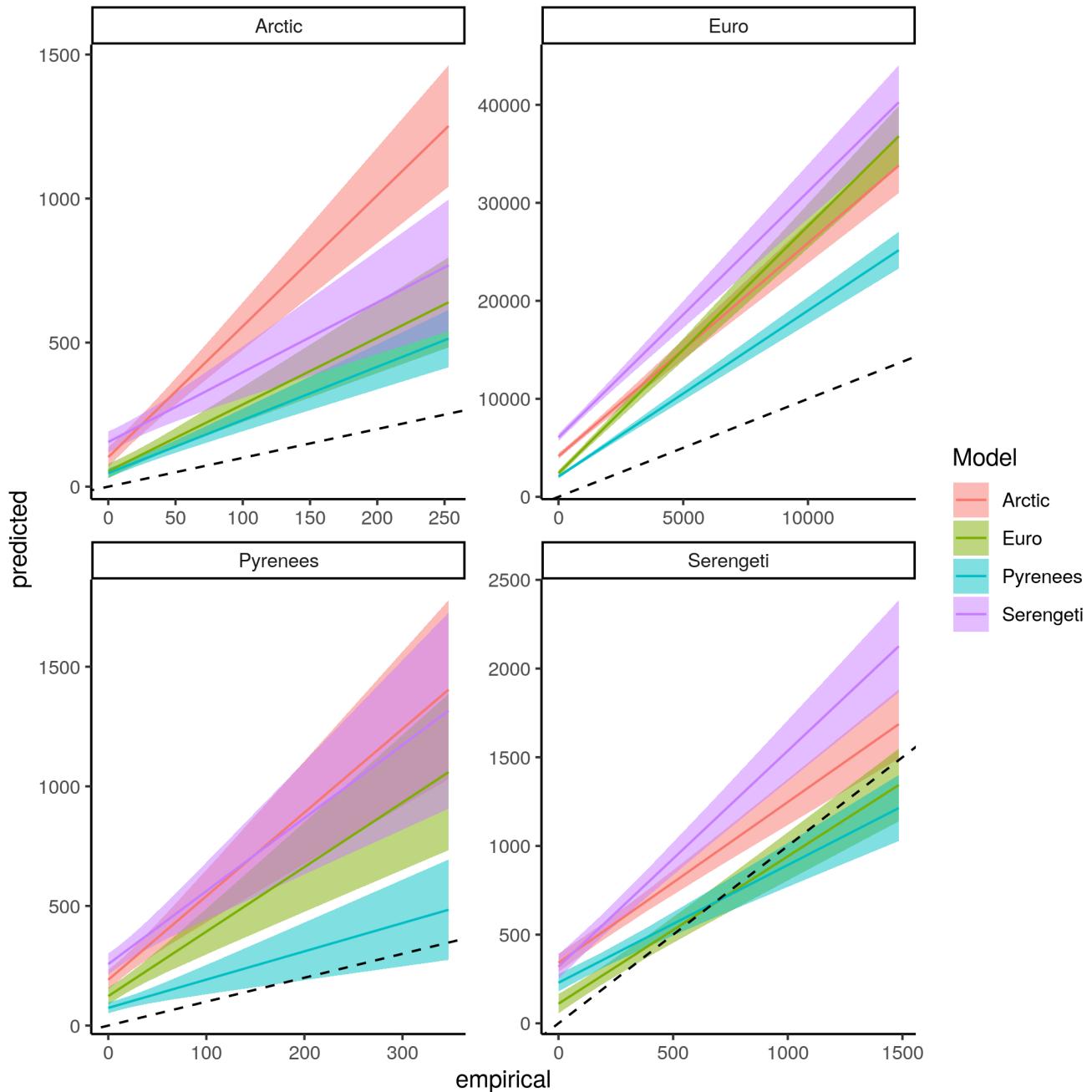
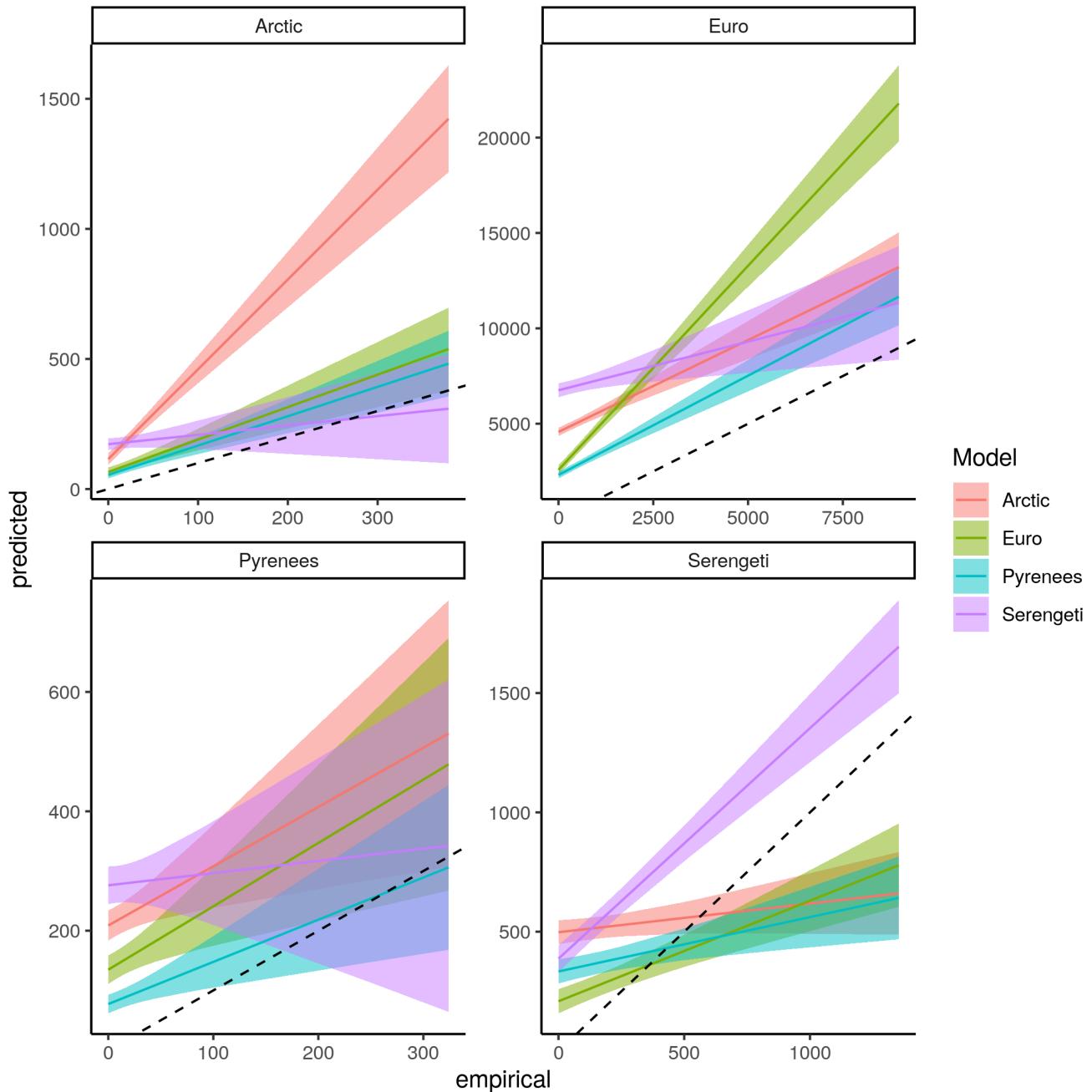


Figure S7.11: Linear regression comparing the predicted frequency of the top predator position in the linear food chain motif and empirical frequency of the top predator position in the linear food chain motif of species for each model predicting every food webs.

## Frequency of the consumer position in the linear food chain motif



*Figure S7.12: Linear regression comparing the predicted frequency of the consumer position in the linear food chain motif and empirical frequency of the consumer position in the linear food chain motif of species for each model predicting every food webs.*

### Frequency of the resource position in the linear food chain motif

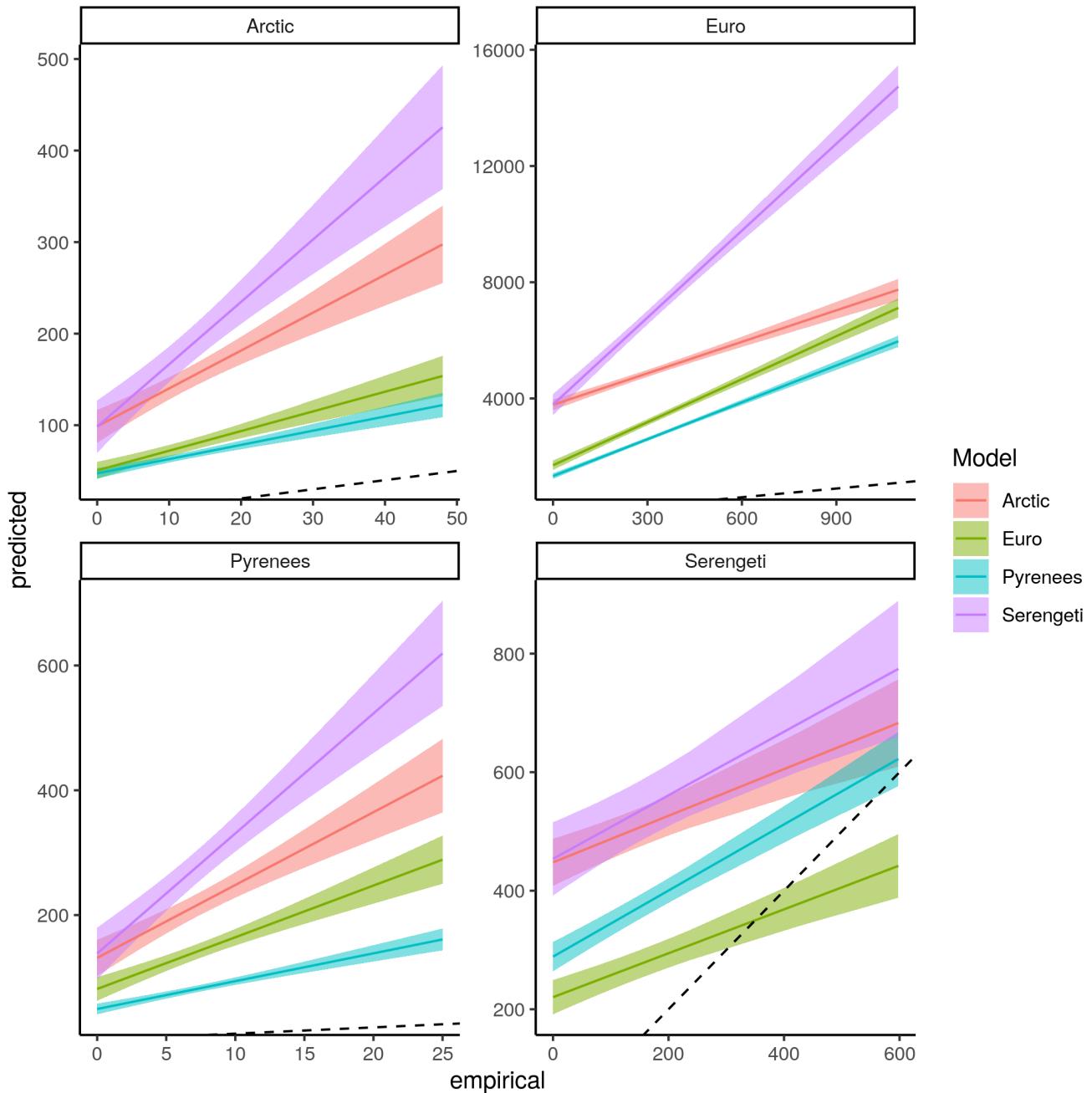


Figure S7.13: Linear regression comparing the predicted frequency of the resource position in the linear food chain motif and empirical frequency of the resource position in the linear food chain motif of species for each model predicting every food webs.

## Frequency of the omnivore position in the intraguild predation motif

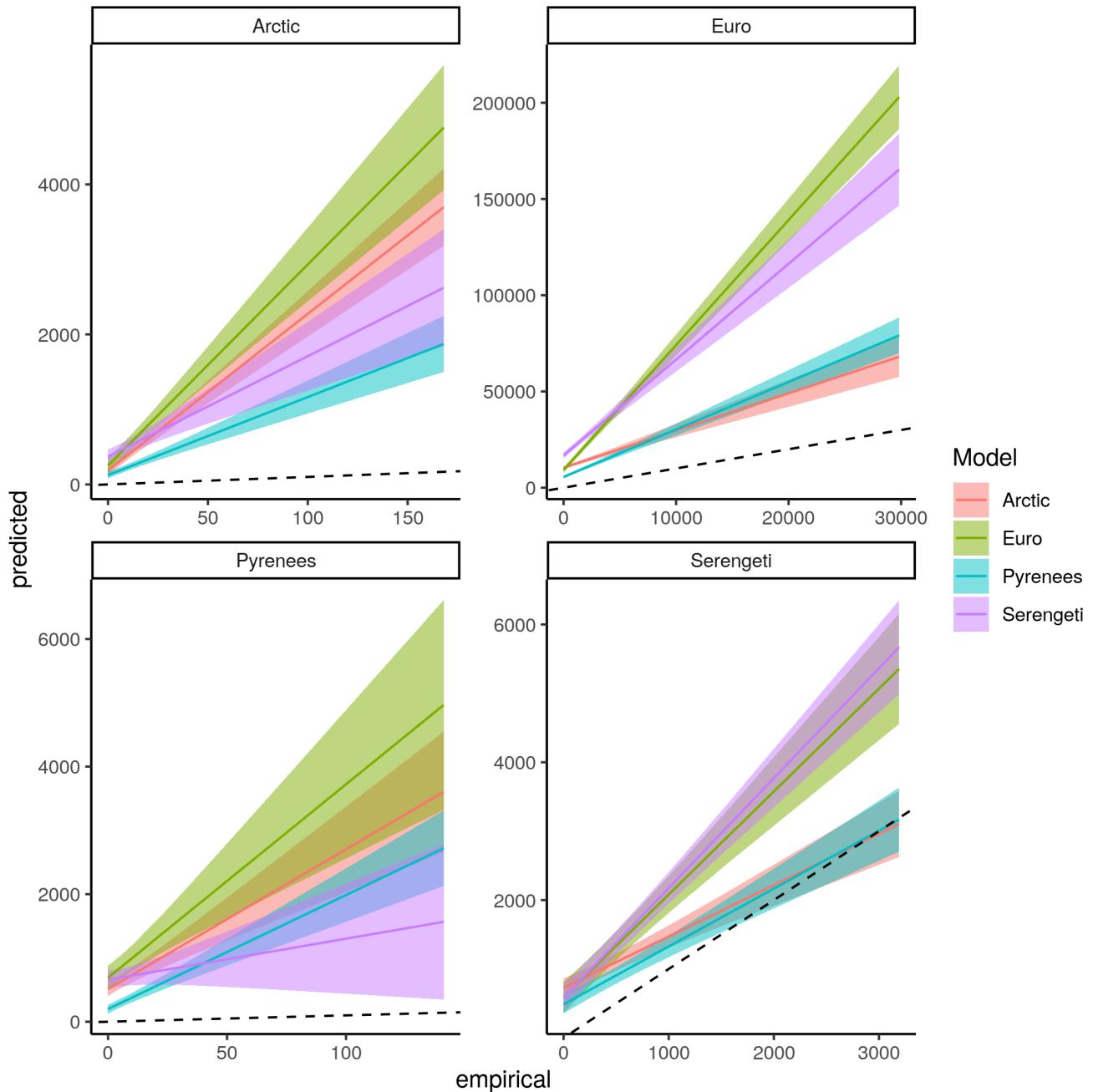


Figure S7.14: Linear regression comparing the predicted frequency of the omnivore position in the intraguild predation motif and empirical frequency of the omnivore position in the intraguild predation motif of species for each model predicting every food webs.

## Frequency of the consumer position in the intraguild predation motif

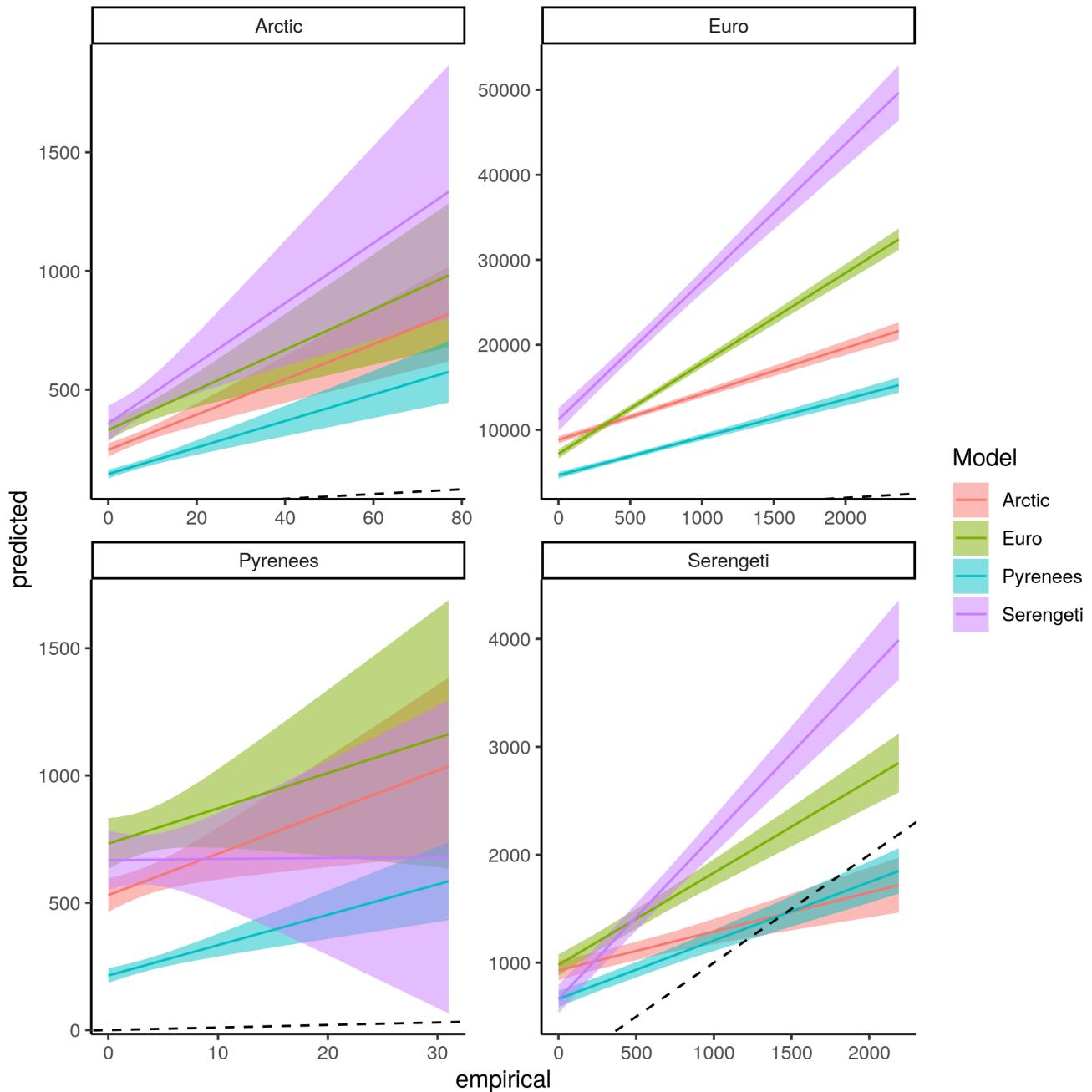
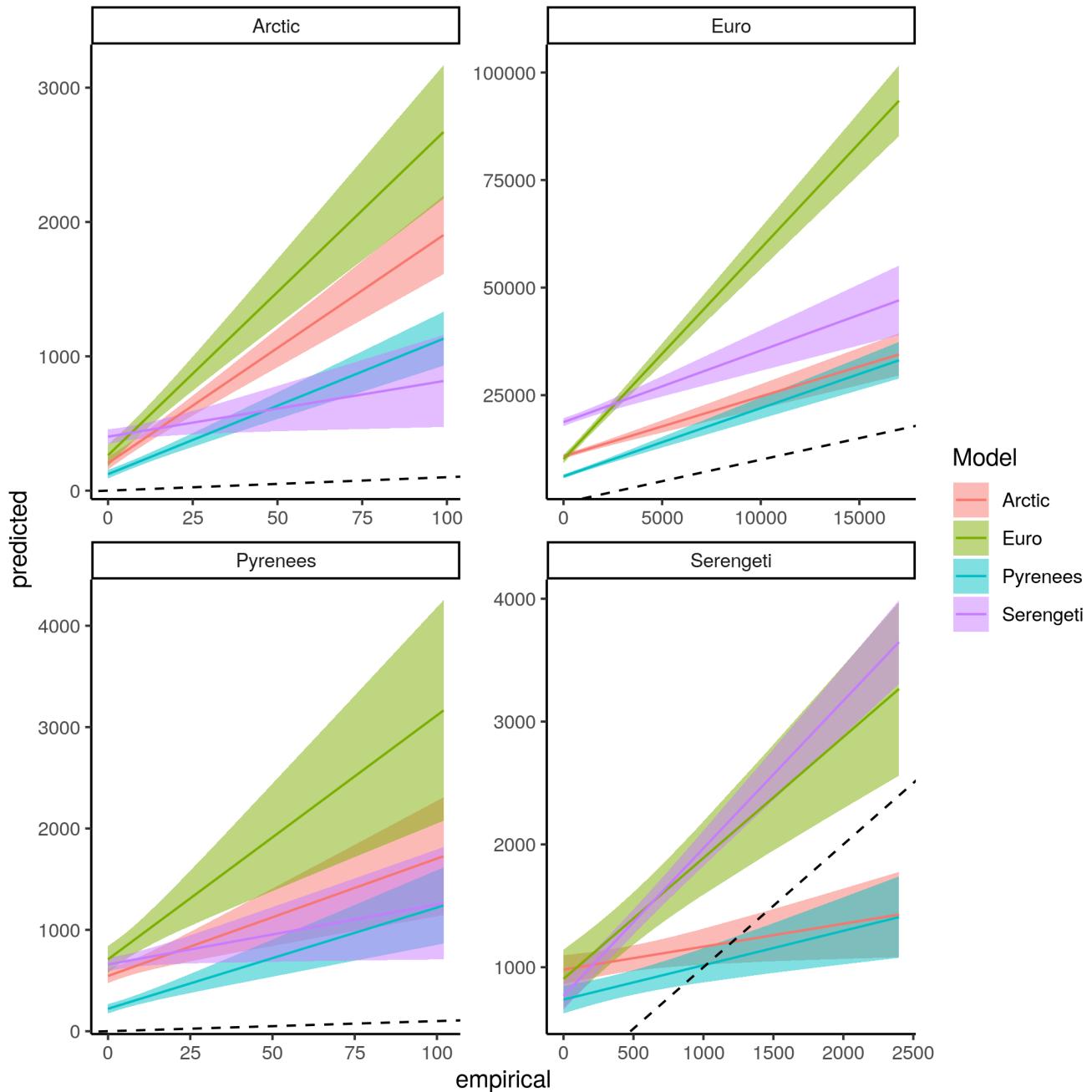


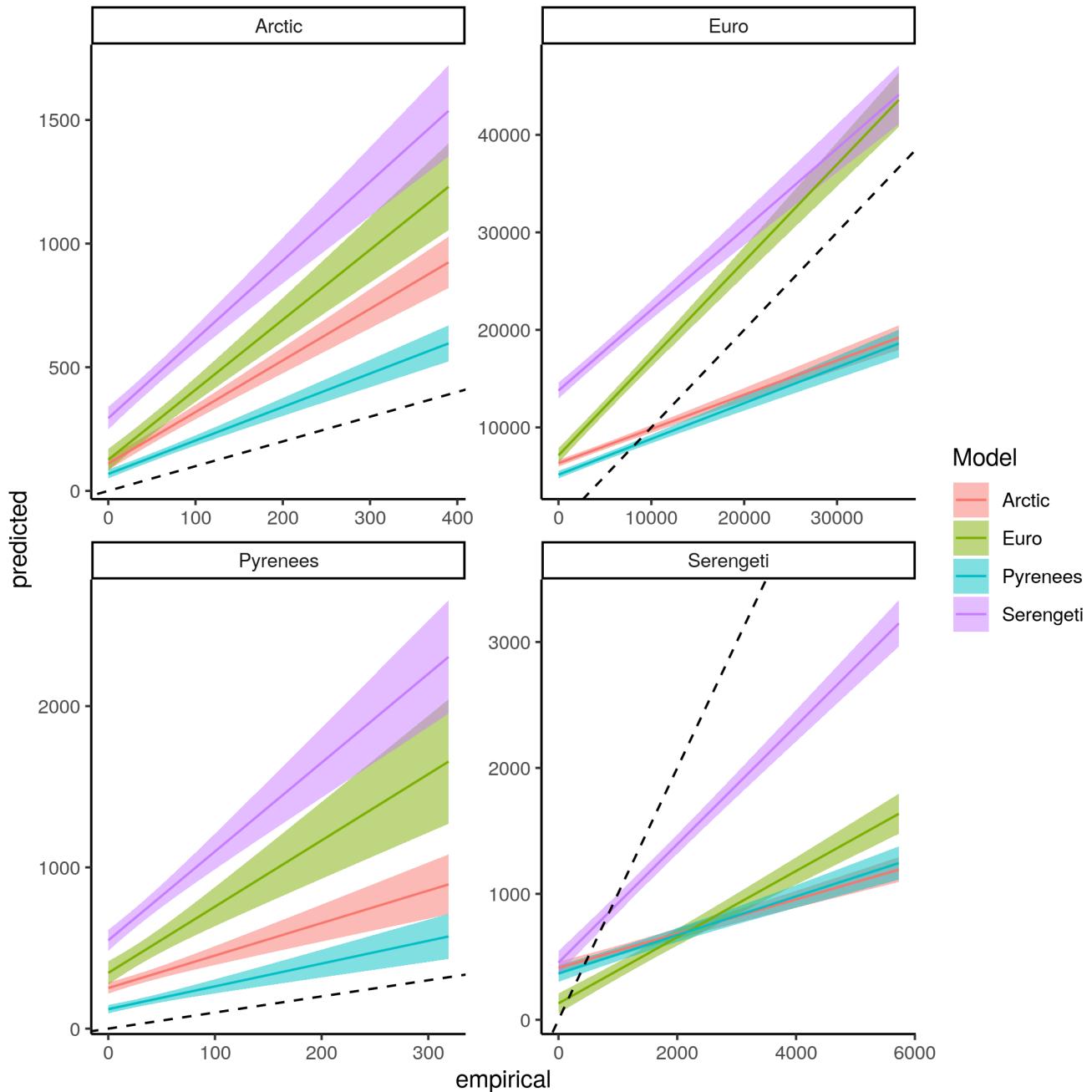
Figure S7.15: Linear regression comparing the predicted frequency of the consumer position in the intraguild predation motif and empirical frequency of the consumer position in the intraguild predation motif of species for each model predicting every food webs.

### Frequency of the resource position in the intraguild predation motif



*Figure S7.16: Linear regression comparing the predicted frequency of the resource position in the intraguild predation motif and empirical frequency of the resource position in the intraguild predation motif of species for each model predicting every food webs.*

### Frequency of the consumer position in the direct competition motif



*Figure S7.17: Linear regression comparing the predicted frequency of the consumer position in the direct competition motif and empirical frequency of the consumer position in the direct competition motif of species for each model predicting every food webs.*

### Frequency of the resource position in the direct competition motif

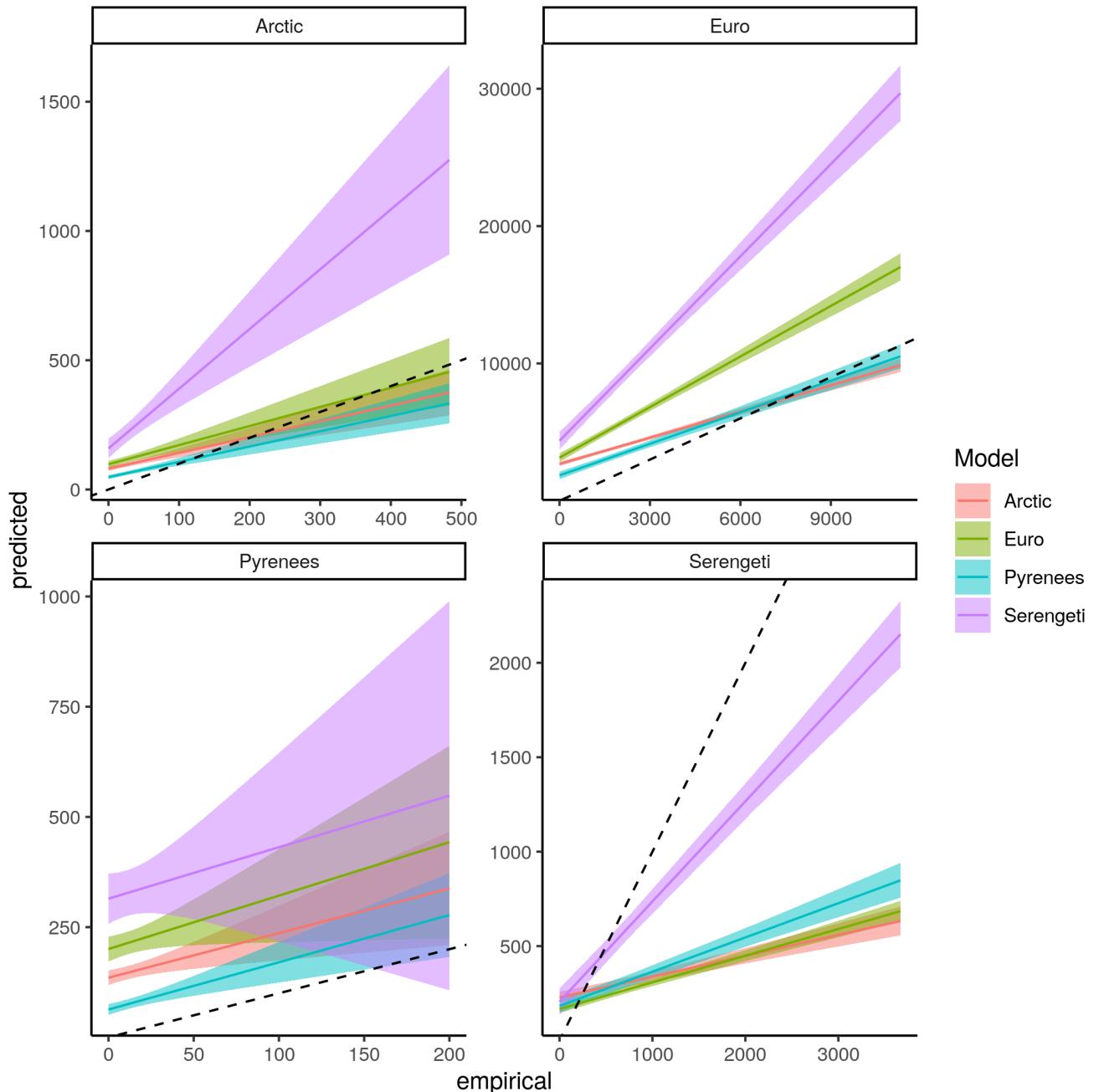


Figure S7.18: Linear regression comparing the predicted frequency of the resource position in the direct competition motif and empirical frequency of the resource position in the direct competition motif of species for each model predicting every food webs.

### Frequency of the consumer position in the apparent competition motif

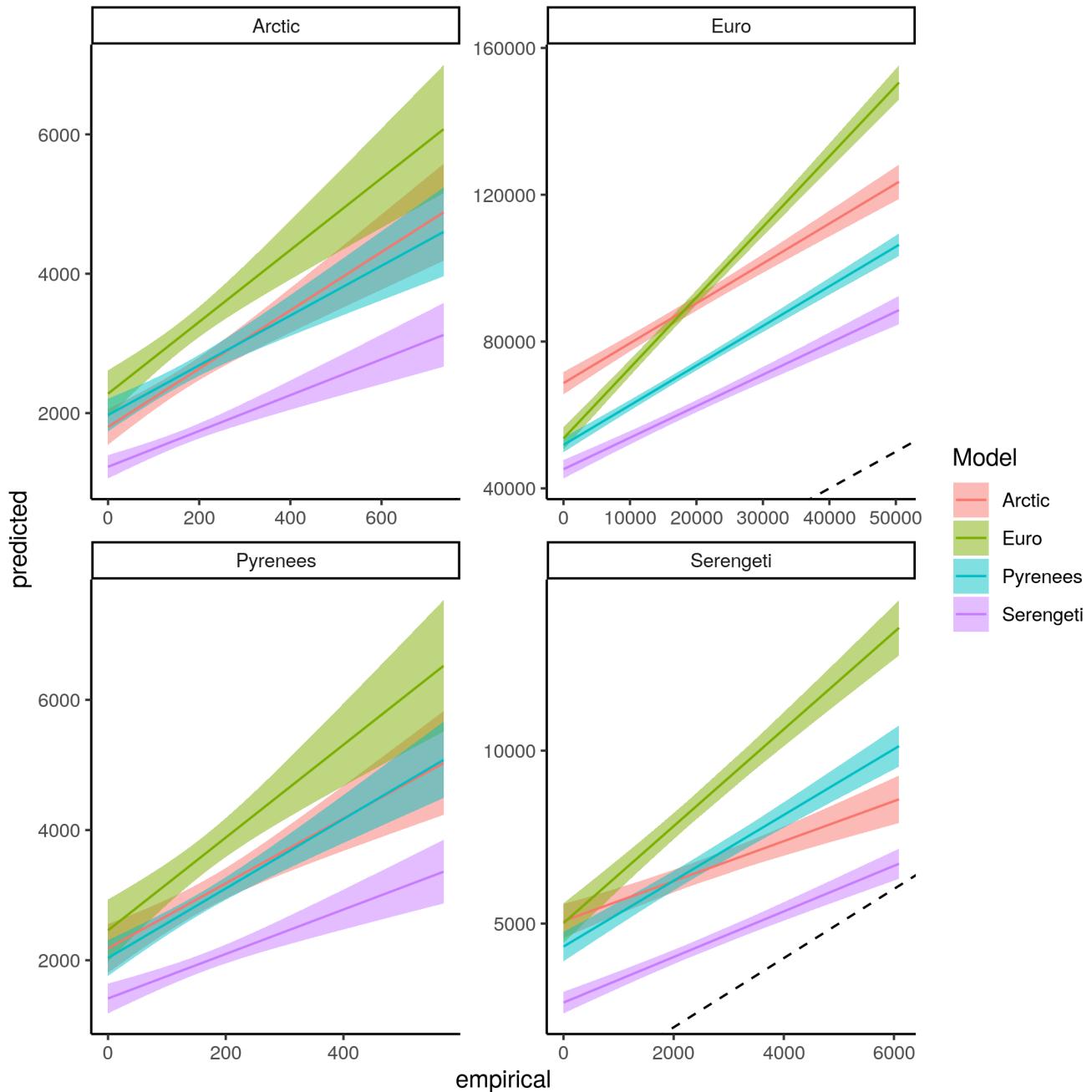


Figure S7.19: Linear regression comparing the predicted frequency of the consumer position in the apparent competition motif and empirical frequency of the consumer position in the apparent competition motif of species for each model predicting every food webs.

## Frequency of the resource position in the apparent competition motif

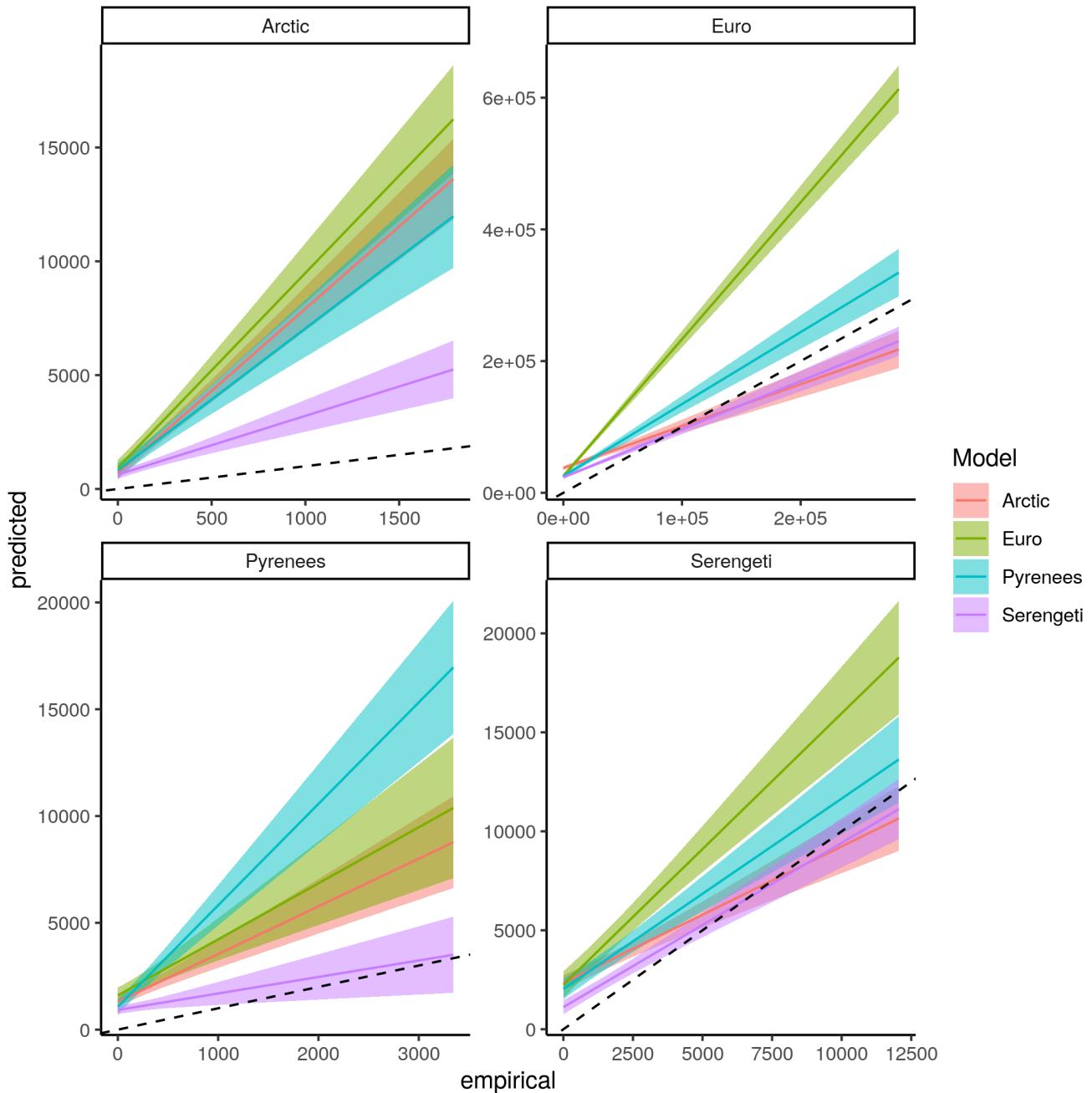
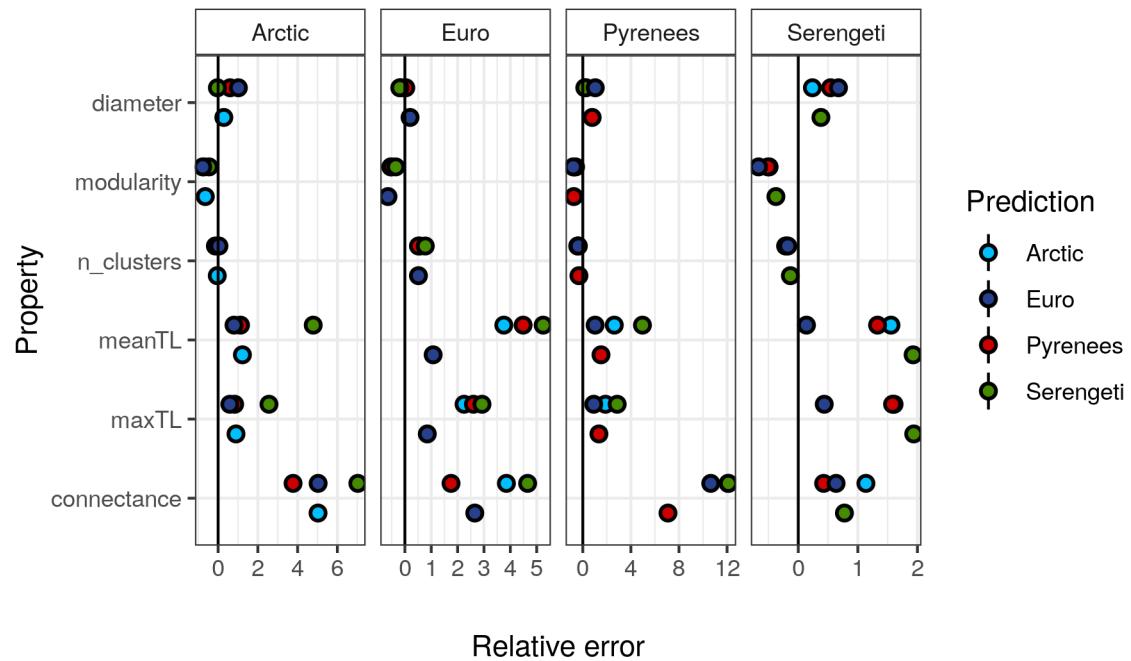


Figure S7.20: Linear regression comparing the predicted frequency of the resource position in the apparent competition motif and empirical frequency of the resource position in the apparent competition motif of species for each model predicting every food webs.

## Appendix 8: Predicting food web properties extended results



*Figure S8.1: The relative error of the predicted food web properties. Relative error is the difference between the predicted and empirical estimates divided by the empirical estimate. From top to bottom food web diameter, modularity, number of clusters, mean trophic level, maximum trophic level, and connectance. Arctic refers to the Northern Québec and Labrador food web*

## **Appendix S9 – References:**

- Allesina, S., & Pascual, M. (2009). Googling Food Webs: Can an Eigenvector Measure Species' Importance for Coextinctions? *PLOS Computational Biology*, 5(9), e1000494.  
<https://doi.org/10.1371/journal.pcbi.1000494>
- Bascompte, J., Melián, C. J., & Sala, E. (2005). Interaction strength combinations and the overfishing of a marine food web. *Proceedings of the National Academy of Sciences*, 102(15), 5443–5447.  
<https://doi.org/10.1073/pnas.0501562102>
- Berteaux, D., Ropars, P., & Casajus, N. (2018). Toundra Nunavik: Matrice des relations trophiques entre espèces du Labrador et du Québec nordique, v. 1.0 (1980-2010). *Nordicana D36*.  
<https://doi.org/10.5885/45555CE-DA1FF11FA4254703>
- Borrelli, J. J. (2015). Selection against instability: Stable subgraphs are most frequent in empirical food webs. *Oikos*, 124(12), 1583–1588. <https://doi.org/10.1111/oik.02176>
- Cirtwill, A. R., Dalla Riva, G. V., Gaiarsa, M. P., Bimler, M. D., Cagua, E. F., Coux, C., & Dehling, D. M. (2018). A review of species role concepts in food webs. *Food Webs*, 16, e00093.  
<https://doi.org/10.1016/j.fooweb.2018.e00093>
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.
- Curtsdotter, A., Binzer, A., Brose, U., de Castro, F., Ebenman, B., Eklöf, A., Riede, J. O., Thierry, A., & Rall, B. C. (2011). Robustness to secondary extinctions: Comparing trait-based sequential deletions in static and dynamic food webs. *Basic and Applied Ecology*, 12(7), 571–580. <https://doi.org/10.1016/j.baae.2011.09.008>
- de Visser, S. N., Freymann, B. P., & Olff, H. (2011). The Serengeti food web: Empirical quantification and analysis of topological changes under increasing human impact. *Journal of Animal Ecology*, 80(2), 484–494. <https://doi.org/10.1111/j.1365-2656.2010.01787.x>
- Delmas, E., Besson, M., Brice, M.-H., Burkle, L. A., Riva, G. V. D., Fortin, M.-J., Gravel, D., Guimarães, P. R., Hembry, D. H., Newman, E. A., Olesen, J. M., Pires, M. M., Yeakel, J. D., & Poisot, T. (2019). Analysing ecological networks of species interactions. *Biological Reviews*, 94(1), 16–36. <https://doi.org/10.1111/brv.12433>

- Duffy, J. E., Cardinale, B. J., France, K. E., McIntyre, P. B., Thébaud, E., & Loreau, M. (2007). The functional role of biodiversity in ecosystems: Incorporating trophic complexity. *Ecology Letters*, 10(6), 522–538. <https://doi.org/10.1111/j.1461-0248.2007.01037.x>
- Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20), 12917–12922. <https://doi.org/10.1073/pnas.192407699>
- Dyer, L. A., & Letourneau, D. (2003). Top-down and bottom-up diversity cascades in detrital vs. Living food webs. *Ecology Letters*, 6(1), 60–68. <https://doi.org/10.1046/j.1461-0248.2003.00398.x>
- Etard, A., Morrill, S., & Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates. *Global Ecology and Biogeography*, 29(12), 2143–2158. <https://doi.org/10.1111/geb.13184>
- Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1), 35–41. <https://doi.org/10.2307/3033543>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (0 ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>
- Guimerà, R., & Amaral, L. A. N. (2005). Cartography of complex networks: Modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02), P02001. <https://doi.org/10.1088/1742-5468/2005/02/P02001>
- Hijmans, R. J. (2021). *raster: Geographic data analysis and modeling* [Manual]. <https://CRAN.R-project.org/package=raster>
- Jordán, F. (2009). Keystone species and food webs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1524), 1733–1741. <https://doi.org/10.1098/rstb.2008.0335>
- Kones, J. K., Soetaert, K., Oevelen, D. van, & Owino, J. (2009). Are network indices robust indicators of food web functioning? A Monte Carlo approach. *Ecological Modelling*, 220, 370–382. <http://dx.doi.org/10.1016/j.ecolmodel.2008.10.012>
- Lurgi, M., López, B. C., & Montoya, J. M. (2012). Climate change impacts on body size and food web structure on mountain ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1605), 3050–3057. <https://doi.org/10.1098/rstb.2012.0239>

- Maiorano, L., Montemaggioli, A., Ficetola, G. F., O'Connor, L., & Thuiller, W. (2020). TETRA EU 1.0: A species level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, geb.13138. <https://doi.org/10.1111/geb.13138>
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594), 824–827. <https://doi.org/10.1126/science.298.5594.824>
- Saravia, L. A. (2022). *multiweb: Ecological network analysis includind multiplex networks*. <https://github.com/lasaravia/multiweb>
- Stekhoven, D. J. (2013). *missForest: Nonparametric missing value imputation using random forest* [Manual].
- Stekhoven, D. J., & Buehlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics (Oxford, England)*, 28(1), 112–118.
- Stouffer, D. B., & Bascompte, J. (2010). Understanding food-web persistence from local to global scales. *Ecology Letters*, 13(2), 154–161. <https://doi.org/10.1111/j.1461-0248.2009.01407.x>
- Stouffer, D. B., & Bascompte, J. (2011). Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9), 3648–3652. <https://doi.org/10.1073/pnas.1014353108>
- Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science*, 335(6075), 1489–1492. <https://doi.org/10.1126/science.1216556>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved  $R^*$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2), 667–718. <https://doi.org/10.1214/20-BA1221>
- Vieira, M. C., & Almeida-Neto, M. (2015). A simple stochastic model for complex coextinctions in mutualistic networks: Robustness decreases with connectance. *Ecology Letters*, 18(2), 144–152. <https://doi.org/10.1111/ele.12394>
- Zhao, Q., Brink, P. J. V. den, Carpentier, C., Wang, Y. X. G., Rodríguez Sánchez, P., Xu, C., Vollbrecht, S., Gillissen, F., Vollebregt, M., Wang, S., & Laender, F. D. (2019). Horizontal and

vertical diversity jointly shape food web stability against small and large perturbations. *Ecology Letters*, 22(7), 1152–1162. <https://doi.org/10.1111/ele.13282>