SIDER: AN R PACKAGE FOR PREDICTING TROPHIC DISCRIMINATION FACTORS OF CONSUMERS BASED ON THEIR ECOLOGY AND PHYLOGENETIC RELATEDNESS

Kevin Healy[1,2], T Guillerme[3], Sean B. A. Kelly[1], Richard Inger[4], Stuart Bearhop[4], Andrew L. Jackson[1]

[1]School of Natural Sciences, Trinity College Dublin, Dublin 2, Republic of Ireland
[2]School of Biology, University of St Andrews, St Andrews, U.K
[3]Imperial College London, Dept of Life Sciences, Ascot, UK
[4]Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Cornwall Campus TR10 9EZ, UK

Corresponding author: Kevin Healy, School of Natural Sciences, Trinity College Dublin, Dublin 2, Republic of Ireland. Email: healyke@tcd.ie

Decision date: 07-Nov-2017

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/ecog.03371].

Summary

Stable isotope mixing models (SIMMs) are an important tool used to study species' trophic ecology. These models are dependent on, and sensitive to, the choice of trophic discrimination factors (TDF) representing the offset in stable isotope delta values between a consumer and their food source when they are at equilibrium. Ideally, controlled feeding trials should be conducted to determine the appropriate TDF for each consumer, tissue type, food source, and isotope combination used in a study. In reality however, this is often not feasible nor practical. In the absence of species-specific information, many researchers either default to an average TDF value for the major taxonomic group of their consumer, or they choose the nearest phylogenetic neighbour for which a TDF is available. Here, we present the SIDER package for R, which uses a phylogenetic regression model based on a compiled dataset to impute (estimate) a TDF of a consumer. We apply information on the tissue type and feeding ecology of the consumer, all of which are known to affect TDFs, using Bayesian inference. Presently, our approach can estimate TDFs for two commonly used isotopes (nitrogen and carbon), for species of mammals and birds with or without previous TDF information. The estimated posterior probability provides both a mean and variance, reflecting the uncertainty of the estimate, and can be subsequently used in the current suite of SIMM software. SIDER allows users to place a greater degree of confidence on their choice of TDF and its associated uncertainty, thereby leading to more robust predictions about trophic relationships in cases where study-specific data from feeding trials is unavailable. The underlying database can be updated readily to incorporate more stable isotope tracers, replicates and taxonomic groups to further increase the confidence in dietary estimates from stable isotope mixing models, as this information becomes available.

Keywords: SIDER, Trophic Discrimination Factors, Comparative analysis, Stable isotope analysis

background

Stable isotopes act as a set of naturally occurring tracers that are altered by the ecological processes of feeding and assimilation such that information on trophic interactions can be inferred through analysis of their relative patterns. One frequent application of inferring trophic interactions using stable isotopes is the reconstruction of consumer diet. This is achieved by parsing the observed isotopic mixture of the consumer into the constituent parts of its food sources using stable isotope mixing models (SIMMs). Since the advent of Bayesian SIMMs, there has been a substantial increase in the use of this technique (Phillips et al. 2014). However, these models require caution with regards to using appropriate data and parameterisation. In addition to knowing the isotopic values of the consumer and its putative sources, it is essential that information on the trophic discrimination factor (TDF) is provided. Essentially, this additive factor represents the difference between the isotopic value of a specific tissue in the consumer compared with the isotopic value of its food, assuming that the consumer tissue is at equilibrium with the food. To quote the originator of this theory: "you are what you eat, plus a few permill" (DeNiro and Epstein 1976).

These isotopic changes between a consumer and its food sources are known to vary predictably with a number of factors. Some of the main drivers relate to the route isotopes take from ingested biomass to consumers' tissue. For example, both the nature of the food source an isotope is derived from and the tissue type it is eventually assimilated into, are strong determinants of TDF values (Caut et al. 2009, McCutchan et al. 2003). Other factors that TDF values may vary with include the physiological condition of the consumer (Pecquerie et al. 2010) and other factors with less clear mechanisms such as the consumers' habitat (Caut et al. 2009). In any case, it is clear there is considerable variation among species (Caut et al. 2009).

The most appropriate method to acquiring TDF estimates is through controlled feeding trials on the consumer in question (Martínez del Rio et al. 2009). Ideally, these experiments should use the same food sources they encounter in the wild as some sources can be assimilated through different metabolic pathways, leading to food source-specific TDF values (Caut et al. 2009). For a variety of reasons, often logistical, this information is not always available, and so researchers often turn to a phylogenetically similar species for which a TDF is available as the next best option. Failing that, researchers will default to a broad approximation derived from averaging TDFs across all species from the major taxonomic grouping to which the consumer belongs: for example, using the oft cited +3.5 per mill for nitrogen or +1.3 per mill for carbon (DeNiro and Epstein 1978, DeNiro and Epstein 1981, Post 2002).

Since mixing models include TDFs as additive offsets, the solutions they generate are inherently dependent on their values. For example, the choice of TDF and the inclusion of uncertainty attributed to this value can affect both the estimates of dietary proportions and the confidence attributed to these proportions (Inger et al. 2010, Inger et al. 2006). Although the main conclusions of SIMMs are often robust to changes in TDF values, such as when sources are widely separated isotopically, the potential sensitivity of SIMMs to TDF choice has been previously highlighted by several authors (Bond and Diamond 2011, Brett et al. 2016, Caut et al. 2009). In some respects, the importance of this problem has been over-stated by focussing too much on the effects of this uncertainty on point estimate summary statistics for the central tendency of the output of the SIMMS. If one follows best practice for applying SIMMs and reports the full posterior distribution, the manifestation of sensitivity to TDFs more often results in overlapping estimates of proportional diet (Phillips et al. 2014). Hence, the ability to incorporate not only more appropriate estimates of the mean TDF, but also the associated variance, is a natural progression within the Bayesian framework of the current SIMMs.

Here we provide an R package that combines these somewhat *ad hoc* approaches under the umbrella of a statistically rigorous phylogenetic regression model fitted to a collated database of TDF values for bird and mammal species. We show how this model can impute (predict or estimate) TDFs for a consumer based on the desired tissue type, their basic ecology and their position in the phylogenetic tree. While previous attempts have been made to estimate TDF values (Caut et al. 2009), these methods are only capable of providing mean TDF values for a given tissue, environment or diet restricted to a predetermined taxon or group. Instead, the inclusion of phylogenetic information as a random term, along with terms for tissue type, repeated measures on the same species and fixed terms for other potential influences including diet and environment type (Caut et al. 2009), means that SIDER uses all the information in the data, weighted accordingly by the estimated correlation structures. This allows SIDER to provide an estimate specific to the species of interest, rather than a generic one for the entire taxon. The regression model is fitted using Bayesian Inference and returns a posterior distribution describing the estimated TDF values that, in turn, can be used to calculate the mean value and the uncertainty associated with it (typically variance or standard deviation). As such, these distributions are also compatible with all the major Bayesian stable isotope mixing models including `MixSIAR` (Stock and Semmens 2013); `IsotopeR` (Ferguson and Hopkins 2016); `SIAR` (Parnell et al. 2010); and `MixSIR` (Moore and Semmens 2008). This approach allows researchers to estimate TDFs for species that are either present or absent from the database and to report, with rigour, both the mean and uncertainty of this estimate. The nature of `SIDER` also allows for flexibility in the predictors included within the regression model, the isotopes used and the taxa in the phylogenetic tree. This flexibility will allow `SIDER` to be updated easily as further data and greater understanding of isotopic routing becomes available.

methods and features

The process for generating predicted TDFs via imputation in `SIDER` is straightforward. A collated database of 409 observed TDFs spanning a range of birds and mammals, tissue types, and variables describing their ecology is provided. A list of values for each of the known explanatory variables associated with the focal species for which we wish to estimate a TDF is appended to the provided dataset, with the phylogenetic tree updated to include the new species where required; i.e. if it is not in the tree then the tree needs to be re-built and relative phylogenetic distances recalculated. The corresponding entries for the focal species' TDF values, which are to be imputed, are encoded as missing values (specifically as NAs) since they are unknown. The phylogenetic regression model is then fitted using the package `MCMCglmm` (Hadfield 2010). This package fits generalised linear mixed models using Markov chain Monte Carlo techniques, with blank NA values in the response variable imputed inherently as part of the MCMC process. The phylogenetic mixed model in `MCMCglmm` includes phylogeny as a random-effect variance structure defined by a distance matrix calculated using the provided phylogenies. This allows for the model to account for the variance associated with the hierarchical structure of phylogenetic relatedness when imputing TDF values. This model produces a simulated posterior distribution for the imputed TDF value which represents an estimate with error of the focal species' TDF. `SIDER` also includes functionality to incorporate the uncertainty associated with building phylogenetic trees by using the `mulTree` package to re-run the analysis across a subsample of a Bayesian distribution of trees (Guillerme and Healy 2014); see methods described in Healy 2015 and Healy et al. 2014. A small distribution of trees are provided in `SIDER`, which were built using the 10k mammal tree (Bininda-Emonds et al. 2007), and the Jetz et al. (2012) bird trees.

The Underlying Data

The data comprise 409 observations of $\Delta^{15}N$ and $\Delta^{13}C$ TDFs derived from controlled feeding trials on 26 bird and 27 mammal species. Explanatory variables that can be included in creating the predictive model are detailed in Table 1. Species taxonomic names follow those used in the Bininda-Emonds et al. (2007) and Jetz et al. (2012) phylogenies. Habitat was defined as either marine or terrestrial based on typical foraging habitat. Sampled tissues were reduced into categories that reflected the metabolic pathway and expected turnover of the tissue. For example, fur and whiskers were all defined under the heading of hair. As diet can affect TDF values, through the specific metabolic pathways a source takes in becoming incorporated into a consumer's tissues (Greer et al. 2015), we defined a consumers diet type as either herbivory, carnivory, omnivory or as a pellet-based diets. Pellet diets were included separately as controlled experiments using pellets often include supplements atypical to the consumer's diet. Our inclusion criterial followed Caut et al. (2009), with $\Delta^{15}N$ and $\Delta^{13}C$ values only included in the dataset if the experiment reported isotopic measurement of both consumers and sources, involved full control of diet, was reported as reaching equilibrium and was conducted on at least four individuals.

The Phylogenetic Regression Model

A generalized linear model of the TDF is fitted using the data that is described in Table 1 and bundled with the package. The response variable is set as being either one of $\Delta^{13}C$ or $\Delta^{15}N$ with explanatory variables of habitat and diet type set as fixed effects. Phylogeny, tissue type and within-species variation (to account for multiple observations reported in some species) are all set as random effects. The model is fitted using the animal model in the MCMCglmm package with typical priors set based on the `MCMCglmm` course notes (Hadfield 2010). Model diagnostics including chain convergence using the Rubin-Gelman diagnostic (Gelman and Rubin 1992) and effective sample size are also automatically tested and displayed for each model run to allow the user to easily assess reliability of the estimated values over large numbers of model runs.

Imputation of an Unknown Observation

In order for `MCMCglmm` to perform imputation, an unknown TDF must be included in the data passed to the fitting algorithm. To do this, the user specifies a new observation for a given species along with the desired tissue type and corresponding information on the habitat, diet type and taxonomic class of the species. Either one of Δ13C or Δ15N are entered as "NA" for this observation, which prompts `MCMCglmm` to impute the missing value automatically during model fitting. The species may be one present in the provided dataset, or it may be a new species, but it must be recognised as present in the phylogeny. In the case of the default scenario, this must match an entry in one of the two phylogenies provided with `SIDER`: either the distribution of mammal trees (Kuhn et al. 2011) or the Jetz *et al.* (2012) bird trees. The function `recipeSider` checks the user-provided species against the underlying phylogeny to make sure it is present in the full phylogeny and prompts the user to import a different phylogeny if the species is not present in either phylogeny. The function `prepareSider` then adds this checked observation to the dataset provided and described in Table 1; prunes the phylogeny to include only the species needed to create the corresponding distance matrix; and adds this observation to the dataset described in Table 1. The cleaned data object from `prepareSider` is then passed to the `imputeSider` function, which calls `MCMCglmm` to run the models. The `imputeSider` function then provides the user with the posterior distribution of the focal TDF estimate and also provides the user with standard diagnostics of MCMC chain convergence and the effective sample size for each model parameter.

model validation

To validate our model, we conducted a series of leave-one-out analyses, where single experimental observations are removed from the dataset and re-estimated using SIDER. This allows TDF estimates to be compared to experimentally observed values. We conducted leave-one-out analyses both for single observations and separately for entire species. In the case of the single observation leave-one-out analyses a single TDF experimental estimate is removed from the dataset and then re-estimated using SIDER in turn. In the case of the species removal analysis, each observation is re-estimated with all experimental observations for that species removed from the dataset. This second analysis hence replicated estimation of a TDF value for a species not present in the dataset.

These analyses generated a series of posterior distributions relating to each TDF imputation (i.e. each row in the dataset) for $\Delta^{13}C$ and $\Delta^{15}N$ isotopes in both mammals and birds. For each of these sets of posteriors we calculated the signed distance of the observed TDF value to each sampled value in its corresponding SIDER posterior to produce a second series of distributions. We then calculate the mode of each of these distributions and plotted them as a density plot as shown in Figure 1a. Across each of the analyses, SIDER was not found to be biased in terms of over- or under-estimating TDF values. In the single observation removal analysis, the modal difference between the SIDER estimations and the experimental values was -0.05 (2.5% CI = -1.95, 97.5% = 1.91) for $\Delta^{13}C$ and 0.01 (2.5% CI = -1.38, 97.5% CI = 1.42) for $\Delta^{15}N$ in birds; and 0.28 (2.5% CI = -2.56, 97.5% CI = 2.38) for $\Delta^{13}C$ and 0.09 (2.5% CI = -1.90, 97.5% CI = 1.87) for $\Delta^{15}N$ in mammals (Figure 1a). In the species removal analysis, the modal difference between the SIDER estimations and the experimental values was 0.04 (2.5% CI = -1.96, 97.5% CI = 2.32) for $\Delta^{13}C$ and 0.18 (2.5% CI = -1.76, 97.5% CI = 2.10) for $\Delta^{15}N$ in birds; and 0.01 (2.5% CI = -4.04, 97.5% CI = 3.64) for $\Delta^{13}C$ and 0.43 (2.5% CI = -3.81, 97.5% CI = 3.29) for $\Delta^{15}N$ in mammals (Figure 1a).

To report the absolute distance SIDER estimates are expected to be from observed values we calculated the root mean squared error between each observed value to the sampled values in its corresponding SIDER posterior. As above we calculated the mode of each of these distributions and plotted them in Figure 1b. The mode of the root mean squared error was found, as expected, to be lower in the single observation removal analysis in comparison to the species removal analysis (Figure 1b). The individual removal analysis for $\Delta^{15}N$ in birds was the lowest at 0.95 (2.5% CI = 0.33, 97.5% CI = 3.12) followed by 1.00 (2.5% CI = 0.69, 97.5% CI = 2.57) for $\Delta^{13}C$. The modal root mean squared error was found to be higher in mammals with 1.29 (2.5% CI = 0.69, 97.5% CI = 3.59) for $\Delta^{13}C$ and 1.03 (2.5% CI = 0.70, 97.5% CI = 2.87) for $\Delta^{15}N$ (Figure 1.b). This was also seen in the species removal analysis with birds found to have 1.45 (2.5% CI = 1.09, 97.5% CI = 2.57) for $\Delta^{13}C$ and 1.44 (2.5% CI = 1.06, 97.5% CI = 2.64) for $\Delta^{15}N$, while mammals modal root mean squared error was found to be 2.03 (2.5% CI = 1.48, 97.5% CI = 3.83) for $\Delta^{13}C$ and 1.60 (2.5% CI = 1.10, 97.5% CI = 2.92) for $\Delta^{15}N$ in (Figure 1.b).

A Worked Example

After loading the SIDER package (https://github.com/healyke/SIDER), we can explore the TDF database, outlined in Table 1, using the `scrumpSider` function." The `scrumpSider` function can search for particular species or can be used to import the entire dataset, which will be required in order to impute TDF values.

```
SIDER_data <- scrumpSider(iso.data = "all")
```

The SIDER package also contains a distribution of Mammalian and Aves phylogenies as outlined above which we can also be upload using `scrumpSider`.

```
SIDER_trees <- scrumpSider(tree = "all")
```

We next define an unknown observation for imputation using `recipeSider`. This function checks that our unobserved data point matches all explanatory variables and is present in the phylogeny. In this case, a grey wolf *Canis lupus* hair sample (which is a terrestrial mammal with a carnivorous diet).

```
new_data_test <- recipeSider(species = "Canis_lupus",
                             habitat = "terrestrial",
                             taxonomic.class = "mammalia",
                             tissue = "hair",
                             diet.type = "carnivore",
                             tree = SIDER_trees)
```

We then add this observation which is to be estimated via imputation during model fitting by defining which of the two isotopes we wish to impute (either `carbon` or `nitrogen` in the current formulation) combining it with the main dataset using `prepareSider`

```
tdf_data_n <- prepareSider(data.estimate = new_data_test,
                           data.isotope = SIDER_data,
                           tree = SIDER_trees,
                           isotope = "nitrogen")
```

The default formulae for the fixed and random parts of the glmm is set within SIDER as

```
formula_n <- delta15N ~ diet.type + habitat
random_terms <- ~ animal + sp.col + tissue
```

The user can directly specify the fixed and random terms, however, any changes will also require re-specifying various parameters associated with running a Bayesian model. These include the prior probability distributions associated with the model parameters and the parameters associated with the length and sampling of the mcmc chains. `SIDER` defaults to use non-informative priors and two chains with 1,200,000 iterations, a burn-in of 200,000, and sampling thinning of 500 (see vignette). These defaults have been optimised on a basic model that runs on the full `SIDER` dataset with no species to be imputed so that all chains converge and have effective sample sizes (ESS) above 1000. The inclusion of a new species to impute may require increased iterations based on whether the chains converge and the ESS is above 1000. In the case of the wolf nitrogen TDF imputation we will use the SIDER defaults, which we do not need to specify in the `imputeSider` function. This model takes approximately seven minutes to run using an intel i5 with 16 GB of RAM.

```
TDF_est_n <-imputeSider(mulTree.data = tdf_data_n,
            formula = formula_n,
            output = "test_n_run")
```

The posterior estimates can then be accessed in `TDF_est.c$tdf_global` and summarized to yield Bayesian credible intervals using the package `hdrcde`. For most users though, the mean and standard deviation or variance of the estimated TDF obtained with the `summary` function, is of primary use for secondary analysis in a mixing model such as `MixSIAR`, `SIAR` or `MixSIR`.

```
Credible_intervals <- hdrcde::hdr(TDF_est_n $tdf_global, prob =
                      c(50, 95, 99))
summary(TDF_est_n$tdf_global)
```

Which returns estimates of a TDF for our wolf (*Canis lupus*) blood tissue with a mean of 3.3 and standard deviation of 1.4.

Incorporating a sider tdf into a mixing model analysis

To demonstrate the full workflow of using a SIDER TDF estimate into a SIMM we used a subset of the wolf analysis from Darimont et al. (2009) which is one of the SIDER-to-MixSIAR-pipeline package vignettes. Using this data, we calculated the proportion of deer (*Odocoileus hemionus*) in the diet of a mainland wolf pack for the following cases;

(1) Using a fox (*Vulpes vulpes*) TDF value with no added uncertainty (Roth and Hobson 2000). We choose this case as a representative of the closest available species with a TDF measured using a controlled feeding experiment and as per the original analysis in Darimont et al. (2009);
(2) Using a recent TDF value estimated for wolves (*Canis lupus*) and its associated uncertainty (Derbridge et al. 2015). This TDF value was not included in our dataset as it does not conform to our data criteria.
(3) Using a SIDER estimate with associated uncertainty.

We found that the SIDER TDF value produced a similar estimate of the proportion of deer compared to using the latest wolf TDF (Figure 2; Derbridge et al. 2015). In contrast, the fox TDF estimated a slightly higher proportion of deer in the wolves' diet (Figure 2). The proportions estimated using the fox TDF also showed a higher degree of confidence in comparison to either of the other analyses (See vignette for full description). This example demonstrates that in the absence of a species specific TDF with realistic associated uncertainty the resulting estimates of dietary proportions from mixing models may be over confident. However, while SIDER can appropriately incorporate the current uncertainty in TDF values based on current knowledge, experimental measures of TDF values, particularly those tailored for a particular system, will remain the gold standard (Martínez del Rio et al. 2009). In particular, the use of TDFs measured from controlled feeding trials tailored to a specific system will remain as the best approach to reduce TDF related uncertainty in the outputs of SIMMs. However, as SIDER's ability to estimate TDF values depends on the range of species with available experimentally controlled measures, feeding trials conducted on new species will not only help reduce uncertainty in SIMMs for those species but also for other species though the use of SIDER.

Acknowledgements

KH acknowledges support from the Marie Curie Research Grants Scheme, grant [749594] and Science Foundation Ireland awarded to Yvonne Buckley. TG acknowledges support from European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement number 311092 awarded to Martin D. Brazeau. AJ was funded by a research scholarship administered by the Fulbright Commission of Ireland and in conjunction with the Marine Institute of Ireland. SB & RI are funded by an ERC consolidators grant (STATEMIG: 310820). We would also like to thank Adam Kane and Mike Wunder for useful comments during the development of the package and the two anonymous reviewers who helped improve both the manuscript and the package.

Author Contributions statement

KH, SK and AJ conceived the original ideas with further contributions from TG, SB and RI. KH and SK collected the data. KH, TG and AJ developed the code. All authors contributed towards writing the manuscript.

Data Accessibility

The data used to fit the regression models, along with the code itself is bundled within the R package and is available on GitHub https://github.com/healyke/SIDER, with the data also on Figshare https://figshare.com/articles/Dataset_for_the_SIDER_R_package/4737481

References

Bininda-Emonds, O.R.P. et al. 2007. The delayed rise of present-day mammals. — Nature 446: 507.

Bond, A.L. and Diamond, A.W. 2011. Recent Bayesian stable-isotope mixing models are highly sensitive to variation in discrimination factors. — *Ecol. Appl*. **21**: 1017-1023.

Brett, M. et al. 2016. Using multiple tracers and directly accounting for trophic modification improves dietary mixing-model performance. — *Ecosphere* **7**(8): e01440.

Caut, S. et al. 2009. Variation in discrimination factors (Δ15N and Δ13C): the effect of diet isotopic values and applications for diet reconstruction. — *J. Appl. Ecol.* **46**: 443-453.

Darimont, C.T. et al. 2009. Landscape heterogeneity and marine subsidy generate extensive intrapopulation niche diversity in a large terrestrial vertebrate. — *J. Anim. Ecol.* **78**: 126-133.

DeNiro, M. and Epstein, S. 1976. You are what you eat (plus a few‰): the carbon isotope cycle in food chains. — *Geol. Soc. Am.* **6**: 834.

DeNiro, M. J. and Epstein, S. 1978. Influence of diet on the distribution of carbon isotopes in animals. — *Geochim. Cosmochim. Acta.* **42**: 495-506.

DeNiro, M. J. and Epstein, S. 1981. Influence of diet on the distribution of nitrogen isotopes in animals. — *Geochim. Cosmochim. Acta.* **45**: 341-351.

Derbridge, J.J. et al. 2015. Experimentally Derived δ13C and δ15N Discrimination Factors for Gray Wolves and the Impact of Prior Information in Bayesian Mixing Models. — *PLoS One* **10**: e0119940.

Ferguson, J. and Hopkins, J. 2016. IsotopeR: Stable Isotope Mixing Model. R package version 0.5.4. https://CRAN.R-project.org/package=IsotopeR.

Gelman, A. and Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences. — *Statist. Sci.* **7**: 457-472.

Greer, A.L. et al. 2015. Simple ways to calculate stable isotope discrimination factors and convert between tissue types. — *Methods Ecol. Evol.* **6**: 1341-1348.

Guillerme, T. and Healy, K. 2014. mulTree: a package for running MCMCglmm analysis on multiple trees. 10.5281/zenodo.12902:

Hadfield, J.D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. — *J. Stat. Softw.* **33**: 1-22.

Healy, K. 2015. Eusociality but not fossoriality drives longevity in small mammals. — *Proc. R. Soc. Lond., B, Biol. Sci.* **282**: 20142917.

Healy, K. et al. 2014. Ecology and mode-of-life explain lifespan variation in birds and mammals. — *Proc. R. Soc. Lond., B, Biol. Sci.* **281**: 20140298.

Inger, R. et al. 2010. Do non-native invasive fish support elevated lamprey populations? — *J. Appl. Ecol.* **47**: 121-129.

Inger, R. et al. 2006. Temporal and intrapopulation variation in prey choice of wintering geese determined by stable isotope analysis. — *J. Anim. Ecol.* **75**: 1190-1200.

Jetz, W. et al. 2012. The global diversity of birds in space and time. — *Nature* **491**: 444-448.

Kuhn, T.S. et al. 2011. A simple polytomy resolver for dated phylogenies. — *Methods Ecol. Evol.* **2**: 427-436.

Martínez del Rio, C. et al. 2009. Isotopic ecology ten years after a call for more laboratory experiments. — *Biological Reviews* **84**: 91-111.

McCutchan, J.H. et al. 2003. Variation in trophic shift for stable isotope ratios of carbon, nitrogen, and sulfur. — *Oikos* **102**: 378-390.

Moore, J.W. and Semmens, B. X. 2008. Incorporating uncertainty and prior information into stable isotope mixing models. — *Ecol. Lett.* **11**: 470-480.

Parnell, A.C. et al. 2010. Source partitioning using stable isotopes: coping with too much variation. — *PloS one* **5**: e9672.

Pecquerie, L. et al. 2010. The impact of metabolism on stable isotope dynamics: a theoretical framework. — *Phil. Trans. R. Soc. B.* **365**: 3455-3468.

Phillips, D.L. et al. 2014. Best practices for use of stable isotope mixing models in food-web studies. — *Can. J. Zoo.* **92**: 823-835.

Post, D.M. 2002. Using stable isotopes to estimate trophic position: models, methods, and assumptions. — Ecology 83: 703-718.

Roth, J.D. and Hobson, K.A. 2000. Stable carbon and nitrogen isotopic fractionation between diet and tissue of captive red fox: implications for dietary reconstruction. — *Can. J. Zoo.* **78**: 848-852.

Stock, B. C. and Semmens, B. X. 2013. MixSIAR GUI User Manual. Version 3.1. https://github.com/brianstock/MixSIAR. doi:10.5281/zenodo.56159

FIGURE LEGENDS

Figure 1: Density histograms of (a) the modal difference and (b) the root mean square error between estimated values using Bayesian imputation and observed values from experimentally controlled dietary studies. The figures in the dark grey boxes on the left give distributions for the individual replacements analysis with the figures in the light grey box on the right giving the distributions for estimates calculated for the species replacements analysis. Each of the outlined boxes give the distributions for Aves $\Delta^{13}C$ and $\Delta^{15}N$ on the top rows and Mammalia $\Delta^{13}C$ and $\Delta^{15}N$ on the bottom row orange bars represent 50%, blue bars 95% and red bars 99% credibility intervals.

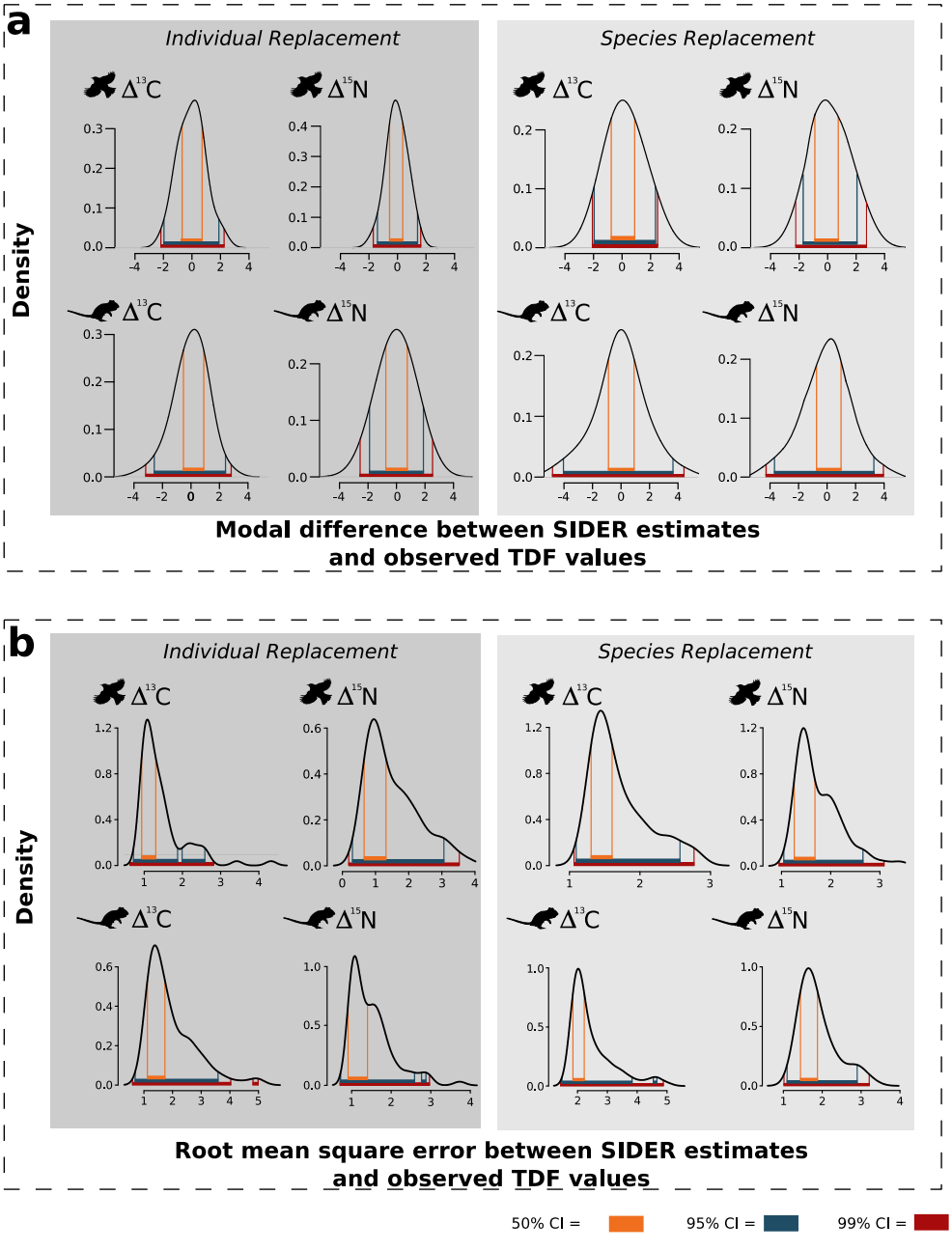

Figure 2: Violin plot of the estimated proportions of Deer (*Odocoileus hemionus*) in the diet of wolves (*Canis lupus*) using three models with alternative TDF estimates: fox TDF values (*Vulpes vulpes*) from Roth and Hobson 2000, wolf TDF values from Derbridge et al. 2015 and TDF values estimated using SIDER. Horizontal lines show the median at the middle, and then the interquartile range and 95% credible intervals (CI).

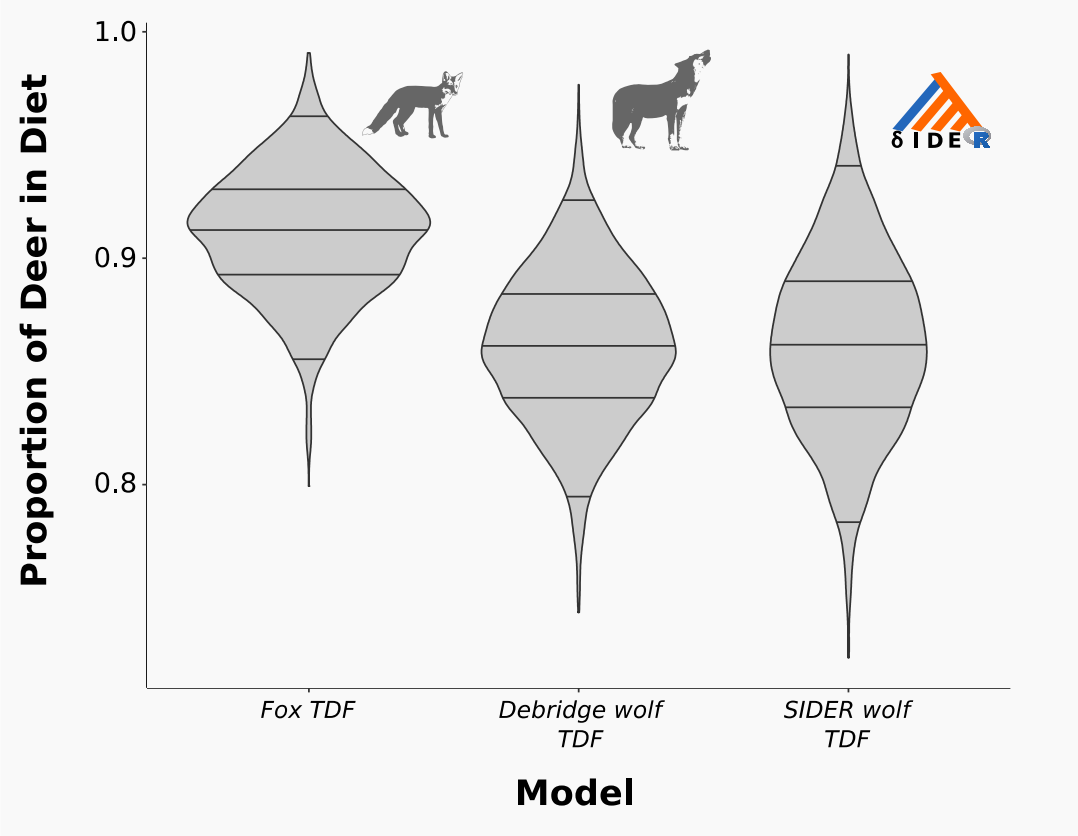

TABLE LEGEND

Table 1: Summary of the dataset used to fit the predictive SIDER model. There are a total of 409 TDF estimates over 53 species in the dataset, comprised of 9 variables.

Variable name	Values	Notes
species	A binomial species name that matches the taxonomy of the corresponding phylogenies.	53 unique species in the present dataset.
habitat	terrestrial / marine	As defined by typical foraging environment
taxonomic.class	mammalia / aves	
sample tissue	liver / blood / kidney / muscle / hair / milk / feather / claws / collagen	Hair and milk are specific to mammals, while feather and claws are specific to birds.
diet.type	herbivore / carnivore / omnivore / pellet	This is a description of the controlled diet they were fed in the experiments. Pellets refer to laboratory food pellets.
source.iso.13C	A numeric value	This is the isotopic δ of the food source. Not currently used in the predictive modelling and present in the dataset for future use.
source.iso.15N	A numeric value	This is the isotopic δ of the food source. Not currently used in the predictive modelling and present in the dataset for future use.
delta13C	A numeric value	The known trophic discrimination factor for $\delta^{13}C$. with 211 observations for 49 species.
delta15N	A numeric value	The known trophic discrimination factor for $\delta^{15}N$ with 198 observations for 49 species.
citation	A character string	The published source of the isotopic data