

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Bruneaux, Matthieu; López-Sepulcre, Andrés

**Title:** isotracer: An R package for the analysis of tracer addition experiments

**Year:** 2022

**Version:** Accepted version (Final draft)

**Copyright:** © 2022 The Authors. Methods in Ecology and Evolution published by John Wiley &

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Bruneaux, M., & López-Sepulcre, A. (2022). isotracer: An R package for the analysis of tracer addition experiments. *Methods in Ecology and Evolution*, 13(5), 1119-1134.

<https://doi.org/10.1111/2041-210x.13822>

## RESEARCH ARTICLE

# isotracer: An R package for the analysis of tracer addition experiments

Matthieu Bruneaux<sup>1</sup>  | Andrés López-Sepulcre<sup>1,2,3</sup> 

<sup>1</sup>Department of Biological and Environmental Sciences, University of Jyväskylä, Jyväskylä, Finland

<sup>2</sup>Department of Biology, Washington University in St. Louis, St. Louis, MO, USA

<sup>3</sup>CNRS UMR 7618, Institute of Ecology and Environmental Sciences of Paris (iEES), Sorbonne Université, Paris, France

**Correspondence**

Matthieu Bruneaux

Email: [matthieu.bruneaux@ens-lyon.org](mailto:matthieu.bruneaux@ens-lyon.org)

**Funding information**

Academy of Finland, Grant/Award

Number: 295941

**Handling Editor:** Jessica Royles

**Abstract**

1. Tracer addition experiments, particularly using isotopic tracers, are becoming increasingly important in a variety of studies aiming at characterizing the flows of molecules or nutrients at different levels of biological organization, from the cellular and tissue levels, to the organismal and ecosystem levels. However, performing rigorous statistical analyses to gain reliable quantitative insights from these experiments often remains challenging.
2. We present an approach based on Hidden Markov Models (HMMs) to estimate nutrient flow parameters across a network, and its implementation in the R package *isotracer*. The *isotracer* package is capable of handling a variety of tracer study designs, including continuous tracer drips, pulse experiments and pulse-chase experiments. It can also take into account tracer decay when radioactive isotopes are used.
3. To illustrate its use, we present three case studies based on published data and spanning different levels of biological organization: a molecular-level study of protein synthesis and degradation in *Arabidopsis thaliana*, an organismal-level study of phosphorus incorporation in the eelgrass *Zostera marina* and an ecosystem-level study of nitrogen dynamics in Trinidadian montane streams. With these case studies, we illustrate how *isotracer* can be used to estimate uptake rates, turnover rates and total flows, as well as their uncertainty. We also show how to perform model selection to compare alternative hypotheses.
4. The *isotracer* package allows researchers from a broad range of disciplines to fully take advantage of their datasets through rigorous statistical analyses. We conclude by discussing *isotracer*'s further applications, limitations and possible future improvements and expansions.

**KEYWORDS**

molecular labelling, nutrient flows, resource allocation, web dynamics

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

## 1 | INTRODUCTION

Tracer addition experiments are an increasingly common tool used to answer a wide variety of biological, ecological and evolutionary questions. These experiments consist in injecting a labelled element into a biological system and tracing its fate throughout the system compartments across time to estimate flows of material across the compartments. Their applications encompass all levels of biological organization, from cells and tissues (Allen & Young, 2020; Yuan et al., 2006) to organisms (Kim et al., 2016; Williams et al., 2016), communities (Freeman et al., 2013; Shik et al., 2018), and ecosystems (Collins et al., 2016; Riis et al., 2012). Tracer additions have a long history. They have been used in seminal experiments revealing the nature of the genetic material (Hershey & Chase, 1952), and more recently in high-throughput metabolomic techniques (e.g. Li et al., 2017; Yuan et al., 2006). They also have a long history in ecosystem science (Crossley & Reichle, 1969) and are becoming a common tool to study the dynamics of nutrients or elements in a variety of ecosystems including: water dynamics in a savanna (Kulmatiski et al., 2010), carbon in coral reefs (de Goeij et al., 2013), nitrogen in deciduous forest soil (Goodale et al., 2015), the effects of reindeer urine in tundra ecosystems (Barthelemy et al., 2018) and nutrient cycling in a variety of aquatic environments (Sánchez-Carrillo & Álvarez Cobelas, 2017). However, gaining reliable quantitative insights and making robust inferences from such data requires appropriate statistical techniques, which can sometimes be challenging to develop and use.

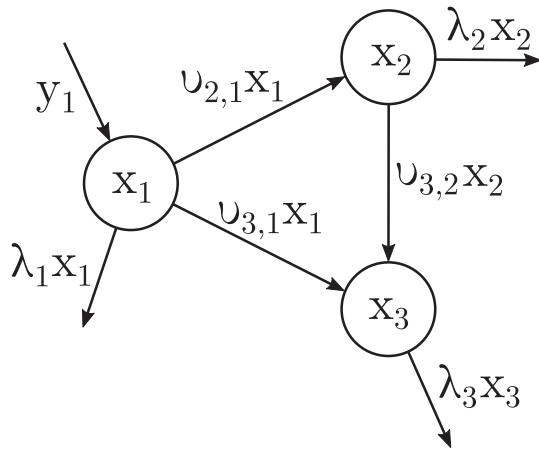
Analysis of tracer addition data typically requires models incorporating a time component via a system of differential equations describing the flows of material across connected compartments, which can be outside the typical statistical expertise of biologists or field geologists. While some sophisticated frameworks exist to analyse pharmacokinetic data (Gelman et al. (1996), Lunn et al. (2002) for PKBugs, Gillespie et al. (2021) for Torsten, currently in development) or metabolomic data (Weindl et al., 2015), other disciplines such as evolutionary biology, ecology or ecosystem science suffer from a lack of a standard analytical tool to make statistically rigorous inferences from tracer additions. For example, analysis of tracer addition experiments to study food web dynamics is usually done by fitting mass balance equations focusing on one trophic compartment at a time, and trophic relationships are assumed to be known a priori, including the relative proportions of diet sources when a consumer feeds on more than one source (Collins et al., 2016; Dodds et al., 2000; Mulholland et al., 2000). Such an approach does not take into account the uncertainty in assumed trophic relationships (Ainsworth et al., 2010; Dodds et al., 2014), and does not allow either the uncertainty in flow estimates at one trophic level to be taken into account when estimating flows in the rest of the trophic network. Statistically reliable estimates of parameter uncertainty are particularly important when performing comparative or experimental studies (e.g. Barneche et al., 2021; Collins et al., 2016; Norman et al., 2017; Tank et al., 2018; Whiles et al., 2013).

We created the `isotracer` R package in an effort to develop a statistical framework for such analyses more easily accessible for researchers. Our package implements the mathematical framework recently described in López-Sepulcre et al. (2020), which is based on a similar logic as the approach used by Wollheim et al. (1999). It considers a whole network system simultaneously and uses a Bayesian approach to incorporate experimental uncertainty into model parameter estimates. For a given set of parameter values, likelihood is calculated by solving the system of differential equations governing the material flows, and by comparing the expected compartment sizes and tracer contents to observations. The current version of `isotracer` implements first-order kinetics for material transfer across compartments, allows replicated units in an experimental design and can take into account categorical covariates to model treatment effects. By using a statistical framework, we can make rigorous inferences and estimate parameter uncertainty. We choose a Bayesian approach to leverage its ability to easily perform marginalization and propagate errors from estimated parameters to derived parameters of interest such as network-wide properties (e.g. total nutrient flow in an ecosystem). Model comparison is possible when several plausible models exist. Additionally, the package also allows researchers to perform power analyses when designing their tracer addition experiments in order to make the most out of such cost- and effort-intensive studies.

We first describe the mathematical framework used in our modelling approach and an overview of its implementation in the package. We then illustrate the use of the package through three case studies based on published datasets: protein turnover in *Arabidopsis thaliana* leaves (Li et al., 2017), phosphate uptake in *Zostera marina* individuals (McRoy & Barsdate, 1970) and nitrogen flows in a Trinidadian mountain stream (Collins et al., 2016). Detailed tutorials to reproduce those case studies are available as an appendix.

## 2 | MATHEMATICAL FRAMEWORK

The system of interest can be represented as a network of connected compartments (Figure 1). As exemplified in our case studies, compartments can be anything from pools of molecules (e.g. intermediates in cellular biosynthesis pathways), to tissues (e.g. blood, liver and muscle; or leaf, stem and root), to species or functional groups (e.g. algae, invertebrate grazer and fish predator). Each compartment represents a distinct pool of matter of any given chemical element of interest, and that matter can flow from one compartment to another when compartments are connected. In a tracer addition experiment, a very small fraction of labelled or marked element of interest (the tracer) is added. The *marked* fraction is followed in time throughout the system's compartments in comparison to the natural *unmarked* fraction. In essence, modelling data from a tracer addition experiment consists in comparing the expected trajectories of tracer with observations, given a set of parameter values. In the case of stable isotope studies, the tracer is often a naturally rare heavy isotopic form of the element of interest (e.g.  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{18}\text{O}$ ,  $^{34}\text{S}$ ) which is used to enrich the



**FIGURE 1** Example of a basic network with three compartments. The  $x_i$  values in each compartment are quantities of matter. The values along arrows are fluxes (quantity per time unit).  $y_1$  is a flux (quantity per time unit);  $v_{ij}$  and  $\lambda_i$  are rate coefficients of first-order kinetics (per time unit), simply called ‘rates’ in this study for convenience.  $x_i$  and  $y_1$  are functions of time  $t$  while  $v_{ij}$  and  $\lambda_i$  are constant rates

pools of the corresponding naturally abundant unmarked form ( $^1\text{H}$ ,  $^{12}\text{C}$ ,  $^{14}\text{N}$ ,  $^{16}\text{O}$ ,  $^{32}\text{S}$ ). Observations usually comprise both the total pool size for each compartment (i.e. the sum of the quantities of heavy and light isotopes) and the proportion of tracer in each pool (i.e. the proportion of heavy isotope to the total amount of heavy and light isotopes, often measured as  $\delta$  enrichment values; Fry, 2006). In the case where only the tracer is being tracked (e.g. a radioactive tracer such as  $^{32}\text{P}$ ), the general framework presented above can be used by considering that the observed sizes represent the tracer alone, and that the observed marked proportions are always one. Radioactive decay is taken into account in the turnover rate of the tracer. It is important to note, however, that the method works for any traceable marker (e.g. immunolabelled molecules), not just rare isotopes.

The mathematical model used to calculate the likelihood of a set of parameter values is detailed in the sections below, and can be divided into two parts. The first part is the calculation of the expected movement of material across the network for those parameter values, which allows to predict the latent state of the network at any point in time. The second part is the likelihood calculation through the incorporation of process and observation error: observations are assumed to be generated via some statistical distribution parameterized with the expected latent states of the network calculated previously.

## 2.1 | General formulation of transfer equations

The mathematical framework described here is similar to the one described in López-Sepulcre et al. (2020). We try to keep mathematical notations consistent with it as much as possible. While the presentation in López-Sepulcre et al. (2020) focuses on discrete-time modelling, we present only the equivalent continuous-time approach here, which allows to specify the model based on a system of differential equations.

Let us consider a network with  $N$  distinct compartments ( $N = 3$  in Figure 1). Each compartment is a pool of the element of interest (e.g. sulphur), and we can define the network state at time  $t$  by a  $N \times 1$  vector  $\mathbf{x}^{(t)} = (x_1^{(t)}, x_2^{(t)}, \dots, x_N^{(t)})$  where  $x_i^{(t)}$  indicates the quantity of material in compartment  $i$  at time  $t$ . Flows between connected compartments are assumed to follow first-order kinetics, and the transfer rate coefficients (later referred to as ‘rates’ for simplicity) are contained in a matrix  $\Upsilon$  (capital upsilon) where each rate  $v_{ij}$  determines the flux from compartment  $j$  to  $i$  as  $v_{ij}x_j^{(t)}$ . Transfer rates can represent a variety of processes depending on the modelled system, including: chemical transformations, resource allocation among tissues, nutrient uptake or consumption among trophic levels. In addition to flowing between compartments, matter can be lost from the system (e.g. through excretion or emigration): each compartment  $i$  has a loss rate  $\lambda_i$  which also defines first-order kinetics for the material exiting the system from this compartment. Finally, some exogenous input can add matter to the system, and is defined for each compartment  $i$  as an input function  $y_i^{(t)}$  giving the input flux for this compartment. Exogenous inputs can represent a variety of scenarios such as nutrients entering an ecosystem from adjacent systems (e.g. upstream in rivers, terrestrial or aerial inputs in aquatic ecosystems), available food or nutrients to be eaten or uptaken by animals or plants, or pollutants in an organism or ecosystem, to name a few examples.

The evolution of such a network system is entirely described by solving for  $\mathbf{x}^{(t)}$  in a linear system of first-order differential equations, for a given  $\mathbf{x}^{(t_0)}$  defining the initial conditions. For the simple network example shown in Figure 1, the system of equations is:

$$\begin{cases} \frac{dx_1}{dt} = -(v_{2,1} + v_{3,1} + \lambda_1)x_1^{(t)} + y_1^{(t)} \\ \frac{dx_2}{dt} = v_{2,1}x_1^{(t)} - (v_{3,2} + \lambda_2)x_2^{(t)} \\ \frac{dx_3}{dt} = v_{3,1}x_1^{(t)} + v_{3,2}x_2^{(t)} - \lambda_3x_3^{(t)} \end{cases} \quad (1)$$

which can be written in a matrix form:

$$\begin{pmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \frac{dx_3}{dt} \end{pmatrix} = \begin{pmatrix} -(v_{2,1} + v_{3,1} + \lambda_1) & 0 & 0 \\ v_{2,1} & -(v_{3,2} + \lambda_2) & 0 \\ v_{3,1} & v_{3,2} & -\lambda_3 \end{pmatrix} \begin{pmatrix} x_1^{(t)} \\ x_2^{(t)} \\ x_3^{(t)} \end{pmatrix} + \begin{pmatrix} y_1^{(t)} \\ 0 \\ 0 \end{pmatrix}. \quad (2)$$

In this form, non-zero  $v_{ij}$  outside the diagonal of the matrix defines a transfer from  $j$  to  $i$ , while the diagonal coefficients are the overall turnover rates for each compartment. The added vector  $\mathbf{y}^{(t)}$  represents the exogenous input fluxes.

In the more general case, the system of differential equations can be written:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}^{(t)} + \mathbf{y}^{(t)} \quad (3)$$

with transfer matrix  $\mathbf{A} = \mathbf{Y} - \mathbf{K} \cdot \mathbf{I}_N$  where  $\mathbf{K}$  is a vector containing the turnover rates for each compartment such that, for compartment  $j$ ,  $k_j = \lambda_j + \sum_{i=1}^N v_{ij}$ , and  $\mathbf{I}_N$  is the identity matrix.

## 2.2 | Likelihood calculation

As mentioned above, tracer addition data usually comprise time series of the compartment sizes as well as their *labelled proportions*. We use interchangeably the terms ‘compartments’ and ‘pools’, and the terms ‘tracer’, ‘labelled’ element and ‘marked’ element to refer to the tracer being added, and ‘unlabelled’ or ‘unmarked’ element for the corresponding common isotope being measured at the same time as the tracer.

To model compartment sizes and labelled proportions, it is necessary to distinguish between two subpopulations of isotopes: the marked population (usually the heavy isotope) and the unmarked population (usually the light population). Following the notations from López-Sepulcre et al. (2020), we define the state of those subpopulations in the network at time  $t$  by two vectors  $\mathbf{m}^{(t)} = (m_1^{(t)}, \dots, m_N^{(t)})$  and  $\mathbf{n}^{(t)} = (n_1^{(t)}, \dots, n_N^{(t)})$  giving the marked and unmarked quantities of atoms, respectively, for each compartment. These vectors are related to the total pool sizes by  $\mathbf{x}^{(t)} = \mathbf{m}^{(t)} + \mathbf{n}^{(t)}$  and the compartment labelled fractions are then defined as  $\mathbf{z}^{(t)} = \mathbf{m}^{(t)} \oslash \mathbf{x}^{(t)}$  where  $\oslash$  is the element-wise division. In a typical tracer addition experiment, the observed time series are for  $\mathbf{x}^{(t)}$  and  $\mathbf{z}^{(t)}$  (not necessarily at the same time points), but for radioactive elements the observed time series can be simply  $\mathbf{m}^{(t)}$ .

The system of differential equations governing  $\mathbf{m}^{(t)}$  and  $\mathbf{n}^{(t)}$  is the same as for  $\mathbf{x}^{(t)}$  above, the only differences being that  $\mathbf{y}^{(t)}$  is split into  $\mathbf{y}_m^{(t)}$  and  $\mathbf{y}_n^{(t)}$  (the vectors defining the input of marked and unmarked matter, respectively, for each compartment) and that initial conditions usually differ for  $\mathbf{m}^{(t)}$  and  $\mathbf{n}^{(t)}$ . Additionally, for radioactive tracers, the  $\lambda$  rates are adjusted to take into account the radioactive decay rate.

For a given set of parameter values, once the system of differential equations is solved and expected trajectories are calculated for  $\mathbf{x}^{(t)}$  and  $\mathbf{z}^{(t)}$ , the observed time series for compartment sizes and labelled proportions are modelled as sampling and measurement errors around the expected trajectories. Several error distributions are implemented in *isotracer*. For example, for the observed size of compartment  $i$  at time  $t$ , one can assume a truncated Normal distribution with the mean being the expected compartment size and a coefficient of variation  $cv$  estimated by the model:

$$x_{\text{obs},i}^{(t)} \sim \text{Truncated Normal}_{\text{lower}=0} \left( \mu = x_i^{(t)}, \sigma = cv \times x_i^{(t)} \right) \quad (4)$$

and for the observed labelled proportions one can assume, for example, a Beta distribution with mean the expected compartment labelled proportion and a precision parameter  $\phi$  estimated by the model:

$$z_{\text{obs},i}^{(t)} \sim \text{Beta} \left( \alpha = \phi \times z_i^{(t)}, \beta = \phi \times (1 - z_i^{(t)}) \right). \quad (5)$$

Other distributions can be a reasonable choice for the generation of observations from the expected trajectories. For example, Normal distributions truncated at 0 or Gamma distributions can be a good approximation if their standard deviations are small and if the labelled proportions are well below 1, and are offered as an option in *isotracer*.

## 2.3 | Modelling experimental design variations

The mathematical model described above can be adjusted to reflect specific features of a given tracer addition setup.

### 2.3.1 | The addition regime

A core component of modelling a tracer addition is the addition regime. The design of the addition regime is flexible, and additions can be done either at discrete time points (pulses, e.g. Barthelemy et al., 2018; McRoy & Barsdate, 1970; Rønnestad et al., 2000) or continuously over a given duration (continuous intervals or drips, e.g. Collins et al., 2016; de Goeij et al., 2013; Williams et al., 2016). Additions can consist of only marked material, or a mix of marked and unmarked material. Finally, additions may be followed or not by an addition of fully unmarked material (chase, e.g. Bacher et al., 2016; Carbone et al., 2007; Simard et al., 1997). All those regimes can be specified by an appropriate definition of the elements of vector  $\mathbf{y}_i^{(t)}$ .

Some source compartments might be considered to be in a steady state on the time-scale of the experiment. This can be the case of dissolved nutrients in a stream which are constantly renewed by the water flow, or when the source compartment is extremely large relative to the amount of material which is transferred to consumer compartments (e.g. ants feeding on a large source of prepared medium). To model compartment  $i$  as being in a steady state, one needs to set all the coefficients of the corresponding  $i$ th row of the  $\mathbf{A}$  matrix to zero and  $y_i(t)$  to zero: this will result in a constant value for  $x_i^{(t)}$  equal to the initial condition  $x_i^{(t_0)}$ .

### 2.3.2 | Over-enrichment

Some sampled compartments can appear over-enriched compared to their source compartments. Over-enrichment happens when a receiving compartment appears more labelled than its source compartment. This occurs notably in ecosystem studies. Over-enrichment should not be possible according to the model, which assumes that the marked material is instantaneously mixed in the receiving compartment pool and that transfer from one compartment does not preferentially affect marked or unmarked material. However, some sampled compartments might actually represent the average of several sub-compartments, for example an active compartment involved in material flows and a refractory one which

behaves as if isolated from the rest of the network on the time-scale of the experiment. This is the case, for example, for a detrital compartment sampled on a stream bed, and which might represent a mix of algae and bacteria (the active portion), assimilating dissolved nutrients and preferentially grazed by invertebrate consumers, and a refractory portion of sediment and slowly decaying matter (e.g. wood and leaves), not involved in significant nutrient cycling in the span of the experiment. Invertebrate consumers can appear over-enriched compared to the average detrital compartment, but they are actually feeding mostly on the rich microbial fraction. For a split compartment  $i$ , comprised of an active and a refractory portions, modelling can be adjusted by adding a  $\pi_i$  parameter representing the active fraction of its pool at  $t_0$ , and by only taking into account this active sub-pool when calculating material flows, but by taking into account the inactive, refractory portion when calculating expected total pool sizes and apparent labelled proportions (López-Sepulcre et al., 2020).

### 3 | PACKAGE IMPLEMENTATION

#### 3.1 | Overview

The model exposed above is implemented in `isotracer` using a Bayesian approach. The Stan program (Carpenter et al., 2017) and its R interface (the `rstan` package, Stan Development Team, 2020) are used for posterior sampling. Stan uses a Hamiltonian Monte Carlo (HMC) algorithm to sample parameter posteriors with MCMC sampling. The HMC sampling allows for efficient and robust sampling of the posterior, and Stan output provides diagnostics to check that the sampled Markov chains behave appropriately. The Stan model calculates the expected compartment trajectories for a set of parameter values using matrix exponentials to solve the system of differential equations, and the likelihood is calculated by comparing observations to those expected trajectories. Matrix exponentials can be used to solve the system of differential equations in Equation 3 because all the addition regimes currently implemented by `isotracer` are equivalent to partitioning the experiment timeline into segments over which  $\mathbf{y}^{(t)} = \mathbf{0}$ .

Note that, while the matrix exponential approach is numerically accurate and is the default solver used by `isotracer`, runtime can become prohibitively long for large models (i.e. with a large number of compartments and/or a large number of unique observation times). A fallback solver is available to the user: it uses a forward Euler scheme with constant step size, but offers no extensive check of numerical accuracy during the numerical solving. This approach allows for greater speed of model evaluation and is expected to perform robustly when an appropriate  $dt$  time step and reasonable parameter priors are chosen, but numerical accuracy is not guaranteed. As such, it should be used only for prototyping purposes when exploring the modelling of large datasets, and the matrix exponential solver should always be used for final runs.

#### 3.2 | Interface overview

The `isotracer` package separates the process of modelling tracer additions into two steps: (a) model definition, and (b) model fitting. This is due to the amount of information required to define a network model, namely: network topology, initial conditions, observations and priors. Figure 2 shows an overview of the typical workflow when using `isotracer` and Table 1 presents the main functions that are used in the model building step.

The package functions have been designed for convenient use with the pipe operator `%>%` provided by the `magrittr` package and familiar to `tidyverse` users, but can also be used without the pipe. A new, empty network model is initialized with the `new_networkModel()` function. This function returns a `tibble` data frame with zero rows and the supplementary class `networkModel`, ready to be populated as the user defines the model in subsequent steps. The `isotracer` package stores its user-facing objects using a tidy table approach (Wickham, 2014) as much as possible to make them easy to inspect. However, the user should use the package functions to modify a model to ensure it results in a valid model. Each row of a `networkModel` object corresponds to one network replicate. It is often necessary to be able to handle several replicates of a similar network system (e.g. modelling replicate sampling locations in a lake, modelling several transect locations along a stream, sampling several plants in a study of nutrient allocation or modelling the removal of an injected drug in the bloodstream of several patients). All rows (i.e. all network replicates) share the same network topology, and each row can be considered either as a simple replication unit with its own initial conditions and observations, but sharing the parameter values that govern flows with all other replicates, or as a level of a treatment where categorical covariates are taken into account when estimating the model parameters, which can differ across treatments in this case. In other words, defining parameters as same or different among replicates defines whether there are different treatment or factor levels.

#### 3.3 | Model definition

##### 3.3.1 | A minimal model

The usage of the functions that define a network model (Table 1) is presented in more details in the case studies and the corresponding appendix below. Here we give an overview of the minimum steps required to define a minimal executable model, and later of the optional steps which allow fine-tuning model definition.

At least five functions have to be used to define a network model (the functions with an asterisk in the last column of Table 1):

- `new_networkModel()` to initialize a new, empty `networkModel` object.
- `set_topo()` to define the network topology of a `networkModel` (i.e. the network compartments and their connections).

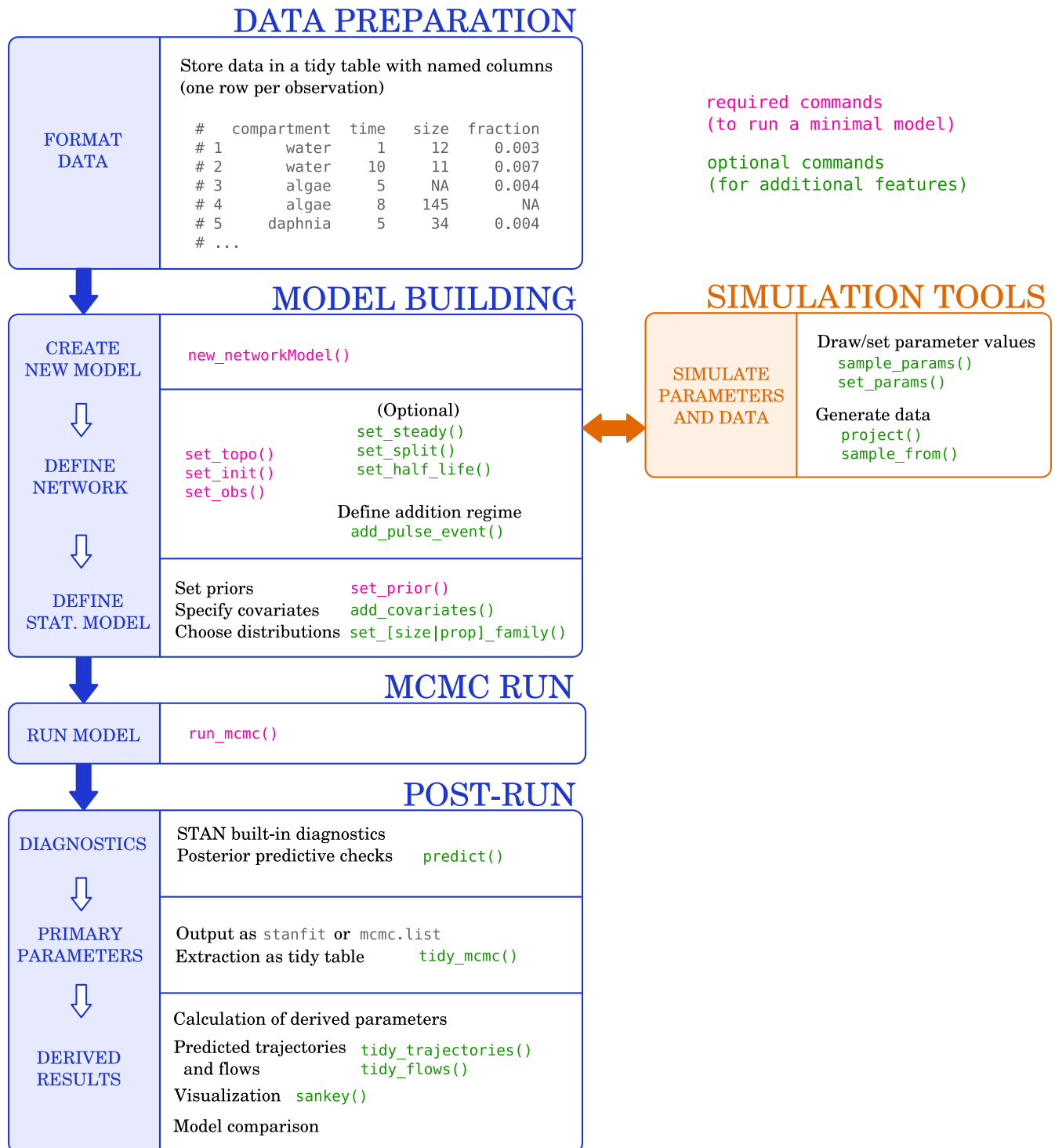


FIGURE 2 Workflow overview for network modelling with *isotracer*

- `set_init()` to provide the initial conditions (total size and labelled proportion) for each compartment. It takes a data frame with columns for compartment identity, size and labelled proportion and accepts supplementary grouping variables used to define replicates.
- `set_obs()` to provide the observations (sampling time, total size, labelled proportion) for each compartment. It accepts a data frame in the same format as `set_init()`, with an extra column for time.
- `set_prior()` to set prior distributions for individual model parameters. Implemented priors are Uniform, Normal, half-Cauchy, Exponential, Gamma and scaled Beta distribution. All priors are truncated to 0. Additionally, a 'constant' prior is available to fix the value of some parameters during MCMC.

TABLE 1 *isotracer* functions used to define a network model (valid for version 1.0)

Function	Role	Example	Notes
Definition Of Network Properties			
<code>new_networkModel()</code>	Create a new, empty network model	<code>m &lt;- new_networkModel()</code>	*
<code>set_topo()</code>	Define a network topology	<code>m &lt;- set_topo(m, "NH4 -&gt; algae -&gt; grazer")</code>	*
<code>set_init()</code>	Define initial conditions	<code>m &lt;- set_init(m, inits, comp = "compartment", size = "biomass", prop = "fraction", group_by = "stream_id")</code>	*, a
<code>set_obs()</code>	Define observations	<code>m &lt;- set_obs(m, obs, time = "time_days")</code>	*, a
<code>set_steady()</code>	Set steady-state compartment	<code>m &lt;- set_steady(m, "NH4")</code>	
<code>set_split()</code>	Set split compartment	<code>m &lt;- set_split(m, "algae")</code>	
<code>set_half_life()</code>	Set a half-life (for radioactive tracer)	<code>m &lt;- set_half_life(m, hl = 34.152)</code>	
<code>add_pulse_event()</code>	Set pulse and drip events	(see Trinidadian stream case study for examples)	b
Definition of statistical properties			
<code>set_prior()</code>	Set prior distributions for parameters	<code>m &lt;- set_prior(m, normal_p(0, 5), "lambda algae")</code>	*, c
<code>add_covariates()</code>	Define categorical covariates	<code>m &lt;- add_covariates(m, upslon_NH4_to_algae ~ stream_id)</code>	
<code>set_size_family()</code>	Set distribution for observed pool sizes	<code>m &lt;- set_size_family(m, "normal_sd")</code>	d
<code>set_prop_family()</code>	Set distribution for observed proportions	<code>m &lt;- set_prop_family(m, "beta_phi")</code>	d

\*Required to build a runnable model.

<sup>a</sup>The example assumes that both `inits` and `obs` have columns named `compartment`, `biomass`, `fraction` and `group_id`, and that `obs` also has a column named `time_days`.

<sup>b</sup>(Steady) drips can be defined by applying pulse events to steady-state compartments.

<sup>c</sup>Note that *isotracer* does not provide any default prior: the user must explicitly set priors for all model parameters. Tutorials in the appendix and package documentation provide guidance for reasonable prior choices.

<sup>d</sup>Default families are normal distributions for sizes and gamma distributions for proportions.

Importantly, while many packages implementing Bayesian methods provide default priors, *isotracer* does not: the user has to explicitly set priors for all parameters. This is a conscious design choice made to encourage users to make reasoned, explicit choices about their model priors, rather than trusting some default priors that might be woefully inappropriate for the system being modelled. The tutorials presented in appendix and the package documentation provide some guidance for choosing slightly informative priors.

### 3.3.2 | Additional network properties

Depending on the network system that is being modelled, additional network properties can be set using the following functions:

- `set_steady()` to define which compartments should be considered at a steady state. The total size and labelled fraction of compartments at steady state do not change during the calculation

of expected compartment trajectories (except if a pulse event is defined for them). This allows modelling, for example, compartments which are constantly replenished (such as dissolved nutrients in flowing stream water) or whose flows are insignificant relative to the time-scale of the experiment.

- `set_split()` to define which compartments should be modelled as comprised of an active portion and a refractory portion.
- `set_half_life()` to define the half-life of radioactive isotopic tracers. This allows to take into account the radioactive decay of the tracer in the estimate of flows. By default, *isotracer* assumes that a stable tracer is used and no half-life is applied in the model.
- `add_pulse_event()` to add pulse events to specific network compartments. A pulse event is defined by an event time and by the quantities of labelled and unlabelled material added to the compartment at that time. In addition to allowing to define versatile 'pulse' designs, this function can also be used to define 'drip' designs when used along `set_steady()`: a pulse event applied to a steady-state compartment will result in a new steady state for this compartment, and 'drip-on' and 'drip-off' phases can thus



be defined using a succession of pulse events to adjust the steady state of a compartment.

### 3.3.3 | Additional statistical model properties

All the functions in the previous section are used to define network properties. The statistical properties of a network model can be defined with the following functions, in addition to the `set_prior()` function required to define a minimal runnable model:

- `add_covariates()` to define categorical covariates with fixed effects on the estimated parameters.
- `set_size_family()` to define the distribution used to model measurement and sampling error when calculating the likelihood of observed compartment sizes compared to expected trajectories. Implemented distributions are:

Normal (mean, coefficient of variation)

Normal (mean, standard deviation)

- `set_prop_family()` to define the distribution used to model measurement and sampling error when calculating the likelihood of observed labelled fractions compared to expected trajectories. Implemented distributions are:

Gamma (mean, coefficient of variation)

Normal (mean, coefficient of variation)

Normal (mean, standard deviation)

Beta (mean, precision)

## 3.4 | MCMC and posterior analysis

Once a network model is defined as explained above, MCMC sampling is performed with the `run_mcmc()` function. This function returns an object of class `mcmc.list` as implemented by the `coda` package, and which can be used directly by many R packages designed for Bayesian analyses (e.g. `bayesplot`). Optionally, `run_mcmc()` can be called with the `stanfit = TRUE` argument. In this case it will return the raw `stanfit` object produced by Stan. This is especially useful when Stan produces errors or warnings during the MCMC sampling: the `stanfit` object can be used for diagnostics (e.g. with the `shinystan` package) as explained in the Stan documentation (Carpenter et al., 2017). Once a model runs without any error or warning from Stan, the user is expected to use the default output from `run_mcmc()` (the `mcmc.list` object). Table 2 provides an overview of the main functions available for post-run processing of `mcmc.list` objects.

The `isotracer` package provides additional methods for the `mcmc.list` class to facilitate the manipulation of such objects. For example, a `plot` method provides a compact way of visualizing MCMC traces and their histograms, and MCMC traces for derived parameters can be easily calculated from the primary parameters returned by `run_mcmc()` via methods implementing the common mathematical operators (+, −, ×, /, exp, log) for `mcmc.list` objects.

To assess model validity, `isotracer` provides functions to perform posterior predictive checks. A `predict()` method is provided for `networkModel` objects, which takes a model fit (the `mcmc.list` output from `run_mcmc()`) and calculates predicted trajectories based on the parameter posteriors. A `tidy_dpp()` function

**TABLE 2** `isotracer` post-run functions typically used on the `mcmc.list` output of `run_mcmc()` (valid for version 1.0). The examples below assume that MCMC sampling was run on a network model `m` with for example: `fit <- run_mcmc(m)`

Function	Role	Example	Notes
Visualization and manipulation of MCMC samples			
<code>plot()</code>	Draw traceplots and histograms	<code>plot(fit)</code>	
<code>summary()</code>	Get summary of parameter estimates	<code>summary(fit)</code>	
<code>as.matrix</code>	Convert MCMC samples to a matrix	<code>as.matrix(fit)</code>	
<code>tidy_mcmc()</code>	Convert MCMC samples to a tidy table	<code>tidy_mcmc(fit)</code>	
<code>[, ...]</code>	Extract parameters	<code>loss &lt;- fit[, "lambda_algae"]</code>	
<code>+, -, *, /, exp(), log()</code>	Calculate derived parameters	<code>turnover &lt;- 1 / loss</code>	
Posterior predictive checks			
<code>predict()</code>	Add posterior predictions to a model (a <code>plot()</code> method is implemented)	<code>p &lt;- predict(m, fit, probs = 0.9)</code> <code>plot(p)</code>	
<code>tidy_dpp()</code>	Provide tidy data and posterior predictions (useful for use with e.g. the <code>bayesplot</code> package)	<code>dpp &lt;- tidy_dpp(m, fit)</code>	
Network properties (calculated for each MCMC iteration, as a tidy table)			
<code>tidy_trajectories()</code>	Calculate compartment trajectories	<code>tidy_trajectories(m, fit)</code>	
<code>tidy_steady_states()</code>	Calculate steady-state compartment sizes	<code>tidy_steady_states(m, fit)</code>	<sup>a</sup>
<code>tidy_flows()</code>	Calculate flows between compartments	<code>tidy_flows(m, fit)</code>	<sup>b</sup>

<sup>a</sup>Steady states can only be calculated for compatible networks (e.g. with at least one steady-state source).

<sup>b</sup>Flows can be averaged over the predicted trajectories or calculated for steady states (if the network admits steady states).

is provided for tidy data and posterior prediction calculation: its output contains the observed data as a `y` list element and the corresponding posterior predictions as a `y_rep` list element, which can be fed directly into the posterior predictive check functions provided by `bayesplot`, such as `ppc_dens_overlay()` or `ppc_intervals()`. The output of `tidy_dpp()` also contains a `vars` list element which can be used to provide the time points to the `bayesplot` plotting functions, and facetting variables when grouping factors were used in the model.

In addition to derived parameters and posterior predictions, `isotracer` can be used to calculate network properties in the particular case of a network admitting equilibrium states. A network can admit an equilibrium state if all  $\lambda$  parameters are set to 0 (no material is lost from the network system), or if at least one compartment is set to a steady state and not all  $\lambda$  are set to zero (a balance can be reached between network in-flows and out-flows). Equilibrium states are calculated using eigenvectors of the transfer matrix. The calculation is performed by the `tidy_steady_states()` function, which will return the compartment size estimates at equilibrium for each MCMC iteration. Similarly, the `tidy_flows()` function can calculate the flows at equilibrium for each iteration, or the average flows over the experiment duration if the network does not admit an equilibrium state. More generally, the `tidy_trajectories()` function will return the full compartment trajectories for each MCMC iteration if they are needed for any downstream analysis.

Finally, since the package can predict compartment trajectories for a given set of parameter values, it can also be used to simulate experiments and generate datasets. The package provides functions to sample parameter values from their priors (`sample_params()`), to calculate trajectories for a set of parameter values (`project()`) and to generate predicted datasets including measurement and sampling error (`sample_from()`). Those functions can be used to optimize experimental designs before running costly and time-consuming real-life experiments, or to check parameter identifiability for a given design.

## 4 | CASE STUDIES

The following case studies demonstrate the use of the `isotracer` package. In the first case study on *Arabidopsis* protein turnover, we introduce basic concepts such as the use of covariates and the calculation of derived parameters. In the second case study on *Zostera* phosphate uptake, we demonstrate how to model radioactive isotopes, and how to compare derived parameters to a reference value. Finally, the study of the nitrogen cycle in a Trinidadian stream illustrates how to model drip experiments, non-homogeneous compartments, and how to perform model comparison. These case studies also illustrate the diversity of addition regimes that can be used in tracer additions (Figure 3). All datasets used in the case studies are available from the `isotracer` package.

### 4.1 | Protein turnover in *Arabidopsis thaliana*

Well-regulated protein synthesis and degradation are crucial for organism development and maintenance. In a study by Li et al. (2017),  $^{15}\text{N}$  labelling was used to measure the degradation rates for 1,228 proteins in *Arabidopsis thaliana* leaves from live plants. This experiment aimed at measuring the in vivo degradation rate of proteins in young plants (21-day old) in three different leaf tissues (third, fifth and seventh leaves). Obtaining turnover estimates for a large number of proteins allowed the authors to examine the determinants of protein degradation rates, such as protein domains and involvement in protein complexes, and to identify proteins for which degradation rates differed between leaves. Here, we demonstrate how to estimate turnover rates for a handful of proteins from the original study using `isotracer`, and how to compare those rates across leaf tissues.

In the original study, seeds were grown on a medium with naturally abundant N isotopes (mostly  $^{14}\text{N}$ ) and switched to a  $^{15}\text{N}$  medium after 21 days. Leaves 3, 5 and 7 were taken from plants at 21 days ( $t_0$ ), 22 days ( $t_1$ ), 24 days ( $t_2$ ) and 26 days ( $t_3$ ) (Figure 3). Each sample consisted of four leaves (from four individual plants) pooled together into one biological replicate. For each leaf type and sampling time, three such replicates were collected. Leaf samples were used for protein separation on gel, digestion and mass spectrometry analysis in order to identify proteins and estimate their labelled fraction. In addition, a similar experiment was performed in which the seeds were grown only on natural N isotope medium (mostly  $^{14}\text{N}$ ) and leaves were sampled at similar time points (without pooling multiple individuals). Those leaves were used in chromatography and mass spectrometry analysis to estimate relative abundance changes for individual proteins.

The data from Li et al. (2017) are available as a Dryad repository (Li et al., 2018) and the `isotracer` package ships several tables derived from this dataset. Here, we only focus on a handful of proteins and estimate their turnover in *Arabidopsis* leaves. The proteins we selected are RBCL (ribulose-bisphosphate carboxylase or 'rubisco', the most abundant protein in leaves in ppm), CPN60A (chaperone involved in rubisco folding), PSBA (protein of the photosystem II), THI1 (thiazole biosynthetic enzyme) and PGK1 (phosphoglycerate kinase 1). Several ways exist in which the biological question at hand can be answered with `isotracer`. For illustrative purposes, we choose to use a network model where the topology is extremely simple and each of the five protein compartments receives nitrogen 'directly' from the growth medium (Figure 4a). This topology is replicated in three instances, one per leaf tissue, and leaf identity is used as a covariate in the model so that leaf-specific parameter values are estimated. This approach assumes that no exchange of nitrogen across proteins or across leaves occurs once nitrogen is incorporated into a protein. Finally, we assume that the N content in the medium is so large that it is constant compared to the protein pools, so we set it to a steady state in the model (Figure 4a). This is equivalent to the zero-order process assumed by Li et al. (2017) for protein synthesis.

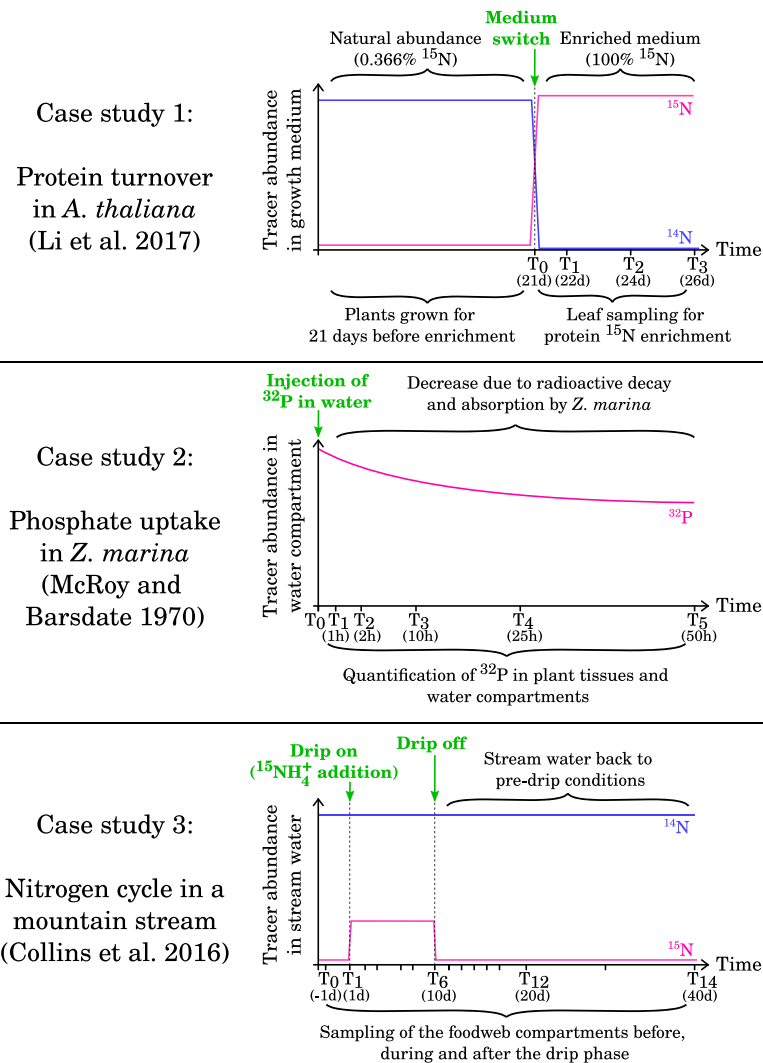


FIGURE 3 Overview of the addition regimes used in the three case studies

From a modelling perspective, the dataset from Li et al. (2017) contains all the data needed for an analysis with *isotracer*: a time series for compartments enrichment (labelled fractions) and a time series for compartment sizes (relative protein abundances). We model the network trajectories from the instant the medium was switched. We assume that the starting labelled fractions for all protein pools are the standard  $^{15}\text{N}$  abundance (0.3663%); for the medium, which is set to steady state in our model, we use a dummy relative abundance of 1 and a labelled fraction of 1 since almost all its N is  $^{15}\text{N}$  during the experiment. Once the MCMC sampling is complete, an important step is to check that the fitted model can correctly predict the original observations using a posterior predictive check (Figure 4b). A posterior predictive check consists in generating predicted data from the parameter posterior distribution, and comparing this predicted data with the actual observations used to fit the model. For example, if we use a 90% probability level, we would expect the predicted trajectory envelopes to contain about 90% of the original observations. Figure 4b shows this check for the labelled proportions of some compartments.

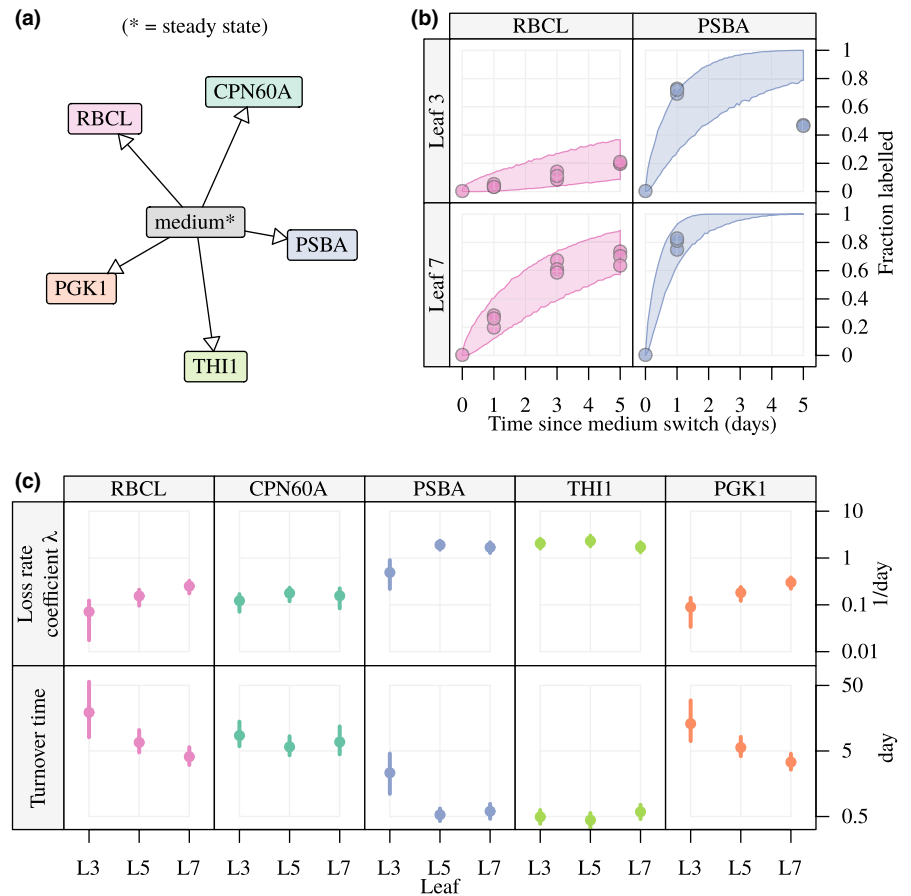
One of the aims of the original paper was to estimate the turnover rate of as many proteins as possible in the three leaf types, and

compare those rates across proteins and leaves. Here, we estimated the turnover rates  $\lambda$  for five proteins, in three leaf types (since each protein pool is not transferring nitrogen to any other compartment, its turnover rate is equal to  $\lambda$ ) (Figure 4c). A powerful aspect of Bayesian MCMC is that we can calculate posterior distributions for derived parameters by doing calculations on the primary parameters sampled during MCMC. For example, turnover times are the inverse of turnover rates, and Figure 4c illustrates posteriors for turnover time which were derived from the turnover rates.

#### 4.2 | Phosphate uptake in *Zostera marina*

Eelgrass can make dense populations in shallow water areas and have an important role in nutrient cycles in the ecosystems they inhabit. McRoy and Barsdate (1970) used controlled laboratory conditions to characterize the uptake of phosphorus by eelgrass plants (*Zostera marina*) from the surrounding water or sediment and its transfer in the plant tissues under light and dark conditions. Their experimental setup consisted in individual eelgrass plants kept in closed jars, with a watertight septum separating the upper and the

**FIGURE 4** Modelling of protein degradation in *Arabidopsis thaliana* leaves based on Li et al. (2017). (a) Network topology used in our model. (b) Examples of posterior predictive checks for compartment enrichment (data points and 90% posterior prediction intervals). (c) Estimates for loss rates and turnover times for proteins across different leaves (posterior means and 95% credible intervals). Since the protein compartments do not transfer nitrogen to other compartments in the network model, their turnover times are simply the inverse of their  $\lambda$  parameters

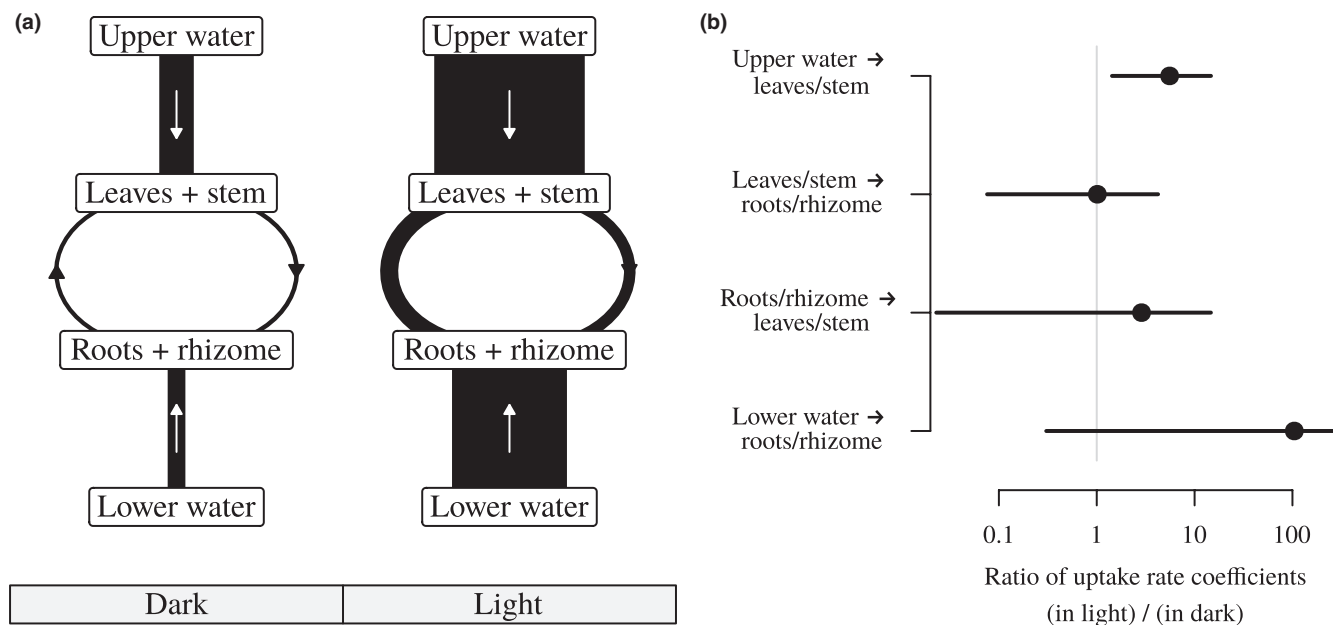


lower compartments. The upper compartment contained the leaves and stems of the plant, while the lower compartment contained the roots and rhizome.  $^{32}\text{P}$  was added as phosphate to either the upper or lower water compartment at the beginning of the experiment, and jars were kept in either light or dark conditions. At several time points over the course of two days, individual plants were taken from the experiment and  $^{32}\text{P}$  abundance in the water and in the plant tissues was quantified by measuring radioactive decay (cpm/mg of dry material; Figure 3). Here, we estimate the phosphate flows in a simplified network with four compartments: sediment water ('lower water'), free water ('upper water'), roots + rhizome ('lower plant') and leaves + stem ('upper plant'). To make the model more amenable, we neglect the release of phosphorus by the plant into the water and only consider four connections in the network model: from upper water to leaves and stem, from lower water to roots and rhizome, and bidirectional flow between leaves and stem (upper plant) and roots and rhizome (lower plant). We examine the light effect by using the light treatment as a covariate in our model, while the addition of  $^{32}\text{P}$  to either the upper or the lower compartment is considered as a replication within light treatment. In other words, we assume that the rates defining the phosphorus fluxes do not depend on the compartment where  $^{32}\text{P}$  was added, but that the light conditions could have an effect on the rates. The radioactive decay of  $^{32}\text{P}$  is automatically taken into account by *isotracer* by adding the appropriate decay rate to the estimated loss rate of each compartment once the half-life has been specified in the model. Note that since

only  $^{32}\text{P}$  (not total P) was measured in these experiments, our model considers only  $^{32}\text{P}$  pools and we fixed the labelled fractions of all observations to 1 (fully labelled).

We compiled the data from McRoy and Barsdate (1970) from the original article figures and tables. We used some approximations to convert the reported data to usable time series. Notably, the reported data for each time point was in cpm/mg of dry weight, and we used a single mean value of each tissue total dry weight per treatment to convert those values into  $^{32}\text{P}$  quantities per compartment for the purpose of this illustrative case study. Given those approximations, the results reported here should be taken as an illustration of the use of *isotracer*, rather than as a biologically precise report on eelgrass physiology.

After fitting the model to determine the transfer rate parameters, we estimated the average flows between compartments over the time of the experiment, in each light treatment (Figure 5a). The Sankey plot shown in this figure is a useful way to visualize flows across a network, but it should be noted that it only includes point estimates of the flows. For a statistically robust comparison between treatments, we can perform calculations on the MCMC values of the primary parameters to build posteriors for derived parameters. For example, to test if the uptake rates differed between light treatments, we calculated the posteriors for their ratios between light and dark conditions (Figure 5b). Based on Figure 5b, only the rate for phosphate uptake from upper water to the leaves and stem increased significantly in light. For the other rates, the 95% credible



**FIGURE 5** Modelling of phosphate absorption in *Zostera marina* tissues under dark and light conditions using data from McRoy and Barsdate (1970). (a) Sankey plot summarizing estimated average phosphate flows between study system compartments. The width of the connectors is proportional to the estimated flows. In the original experiment, upper and lower water compartments are separated by a septum. (b) Mean estimates and 95% credible intervals for the ratios of the uptake rate values between light and dark conditions

intervals were so large that they overlapped with a ratio value of 1. This large uncertainty is due to the variability of the original data and the blunt assumptions that we made when we converted the original data to usable time series. We expect that a dataset where total tissue dry weight would be estimated at the same time as cpm/mg dry weight would allow to reduce the uncertainty a lot. However, this illustrates the ability of our method to estimate uncertainty in derived parameters and to test rigorously for differences between treatments.

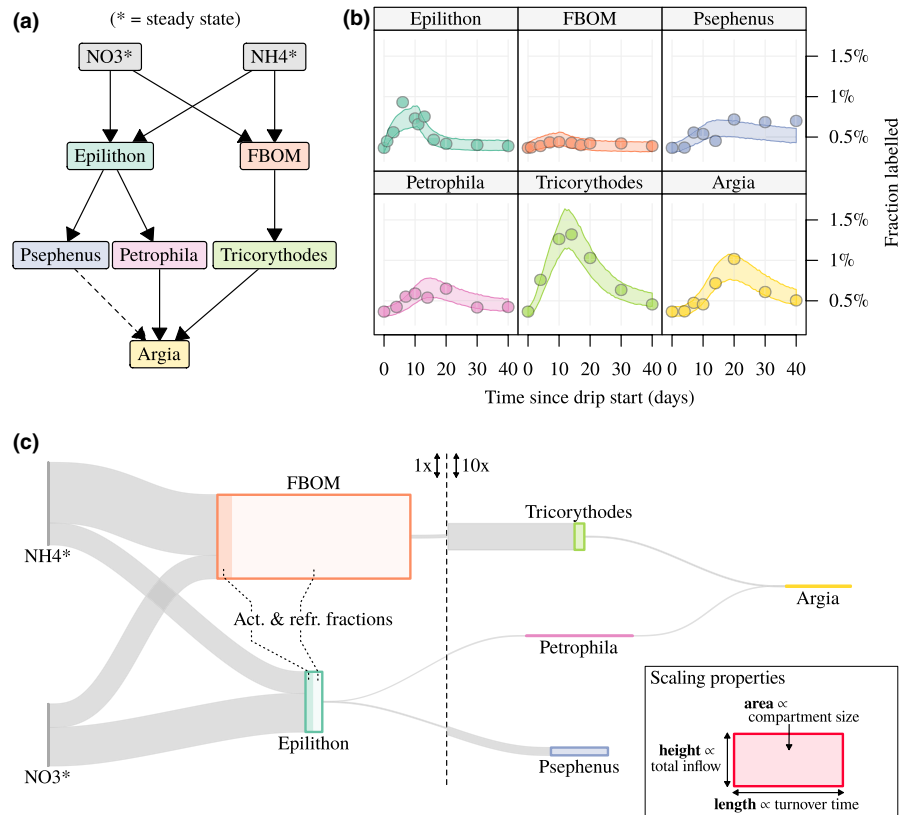
### 4.3 | Nitrogen cycle in a Trinidadian mountain stream

Tracer addition experiments have been used profitably in ecology to identify the strength of trophic connections and to quantify nutrient cycling in ecosystems. In a study by Collins et al. (2016), the authors quantified the nitrogen exchanges in the food webs of mountain streams in Trinidad, from the dissolved nutrients ( $\text{NH}_4^+$  and  $\text{NO}_3^-$ ) to the invertebrate consumers. They examined the effect of light conditions and of the presence/absence of a fish consumer on the nitrogen dynamics in the food web. Data were collected by dripping  $^{15}\text{N}$ -enriched ammonium into two streams in Trinidad, and samples from each food web compartment were taken during the drip and after the drip in several transects in each stream. The transects were located at different locations downstream of each drip. The drip phase lasted 10 days, and the post-drip phase lasted 30 days (Figure 3). Compartment enrichment was measured by determining  $^{15}\text{N}/^{14}\text{N}$  ratios by mass spectrometry, while compartment sizes were estimated in  $\text{mgN}/\text{m}^2$  using various field sampling techniques

depending on the compartment type (dissolved, primary producer, invertebrate). Streams differed in their exposure to sunlight (natural vs. trimmed canopy) and transects within streams differed in the presence/absence of guppies (*Poecilia reticulata*). The dataset used in Collins et al. (2016) is shipped with *isotracer* as the *la-laja* table. In this case study, we will use the data from only one stream and a subset of the original food web for simplicity, but a full analysis including both streams and using the framework used in *isotracer* was presented in López-Sepulcre et al. (2020). We show how *isotracer* can be used to test for the existence of a candidate trophic link by comparing models, to determine active and refractory portions of inhomogeneous compartments, and to estimate the relative proportions of inputs into a given compartment.

We ran two models using the topology presented in Figure 6a. Those models differed only by the inclusion or not of a candidate trophic link between the grazer genus *Psephenus* and the predator genus *Argia* (dashed arrow in Figure 6a). Dissolved nutrients were set to a steady state given that they were renewed by the stream flow, and we allowed the two primary producer compartments, the algal cover on rock substrate (epilithon) and the fine benthic organic matter (FBOM), to be modelled as split (inhomogeneous) compartments given that the field sampling cannot distinguish between active and refractory portions of those compartments. We used data from three transects from one stream (Upper LaLaja, for which the canopy cover was thinned). The transects differed in their effective addition regime due to their different distances from the drip source, but otherwise were used as simple replicates in our model (i.e. they shared the same parameter values). We used DIC to compare the models: the  $\Delta\text{DIC}$  between the models was relatively small (2.66), and suggested that the model without the candidate trophic

**FIGURE 6** Modelling on nitrogen transport across food webs in Trinidadian mountain streams using data from Collins et al. (2016). (a) Foodweb topology used in our case study. The dashed arrow depicts the candidate trophic link that is tested by model comparison. (b) Examples of posterior predictive checks for the labelled fraction data in the first transect of the stream (data points and 90% posterior prediction intervals). (c) Sankey plot showing the estimated nitrogen flows across compartments in the first transect. Connecting ribbon widths are proportional to estimated flows. For epilithon and FBOM, an active portion of the compartment was estimated by our model. Vertical scale is expanded in the right half of the plot for readability



link fitted the data better, with a DIC-derived weight of 80% for this model (López-Sepulcre et al., 2020). Posterior predictive checks were satisfactory (Figure 6b).

As for the *Zostera* case study, we can use a Sankey plot to visualize flows between compartments (Figure 6c). In this case, the area of a given compartment is proportional to the compartment size (a quantity of nitrogen) while the thickness of the connecting ribbons is proportional to the flows (a quantity of nitrogen per time unit). It results from this that the length of a given compartment is proportional to its turnover time (per time unit). The Sankey plot clearly shows the difference in flows between nutrients and primary producers, compared to between primary producers and upper trophic levels. The estimated active portions in epilithon and FBOM were 0.47 (95% CI 0.19–0.73) and 0.08 (95% CI 0.01–0.18) respectively. Being able to estimate an active portion instead of having to set it to a fixed value a priori is important when no prior knowledge is available to make a robust, educated guess about this value. It is also the case for proportions of different inputs: the modelling approach implemented in *isotracer* estimates proportions of different inputs into a compartment from the tracer data, rather than having to rely on estimates from, for example, gut content analysis or other independent studies. For epilithon, the proportion of nitrogen assimilated from  $\text{NH}_4^+$  was 0.37 (95% CI 0.27–0.52) and from  $\text{NO}_3^-$  0.63 (95% CI 0.48–0.73). The high part of  $\text{NO}_3^-$  is explained by the open canopy of this stream, while estimates in the other, closed canopy stream showed that epilithon favoured  $\text{NH}_4^+$  when light was limiting (López-Sepulcre et al., 2020).

## 5 | CURRENT LIMITATIONS, EXTENSIBILITY AND FUTURE DIRECTIONS

The *isotracer* package implements a Bayesian approach to model tracer addition experiments and to estimate parameter uncertainty. Its primary aim is to allow researchers whose typical expertise might not cover statistical modelling of systems governed by ordinary differential equations to perform rigorous and reliable statistical analyses, so that the valuable data produced by tracer addition experiments can be fully taken advantage of. The package is built upon the versatile probabilistic programming language Stan, but provides a user-friendly interface to define network structure and experimental design that does not require the user to write any Stan code and makes quick prototyping of models possible.

*isotracer* allows for modelling of a variety of addition designs (pulse, drip, pulse-chase), can take into account both stable and decaying (e.g. radioactive) tracers, and can be used for model comparison to test hypotheses relating to network structure. The toolkit provided for post-MCMC analyses is rich and can be used to calculate network-wide properties (and their uncertainties) such as flow estimates or compartments steady state sizes when applicable. At the moment, the package does not provide explicit spatial modelling beyond discrete replicate units. However, a continuous handling of the spatial component might be useful to study, for example, stream bed nitrification processes. This type of model would typically involve systems of partial differential equations, which are not implemented for now in the package and for which we recommend the

user to use Stan directly, which might require higher expertise in statistical analysis and coding.

From a technical point of view, the default solver for differential equations used by `isotracer` uses matrix exponentials for numerical accuracy and is reasonably fast (running 2,000 MCMC iterations for the Trinidadian stream case study presented here took less than 20 min on a 3.6 GHz processor with one core per chain). The runtime with this default solver can, however, become problematic for larger models, in which case Stan HMC tuning parameters such as `adapt_delta` and `max_treedepth` can be adjusted directly via `run_mcmc()`. An alternative solver using a forward Euler scheme is also offered for exploratory modelling, but without any guarantee of numerical accuracy.

From a modelling perspective, a future improvement could include the implementation of continuous covariates in addition to discrete ones. This would allow the study of phenomena such as size-specific uptake in consumers, or temperature-specific resource allocation. Isotope fractionation could also be taken into account: the current version of `isotracer` ignores it and assumes that the tracer and the corresponding 'unmarked' material are governed by identical transfer rate coefficients. For ranges of isotope enrichment usually observed in trophic networks (e.g.  $\delta^{15}\text{N}$  typically increasing by about 3.4‰ per trophic level (Cabana & Rasmussen, 1994), but with large variations depending on the study system, for example, Mill et al. (2007)), it is likely that this is a reasonable approximation given the very large enrichment caused by the experimental addition (e.g.  $\delta^{15}\text{N}$  up to about 2,500‰ in some primary consumers in Collins et al. (2016), but as low as 40‰ in higher trophic levels further away from the drip source). However, to accurately take into account isotopic fractionation at higher trophic levels or due to other processes (Robinson, 2001), an option to model it when it is known to occur could be implemented.

Importantly, the current version of `isotracer` only models first-order reactions: the quantity of transported material per unit of time is calculated as the product of a constant rate coefficient and of the source compartment size. Such first-order reactions result in simple systems of differential equations, and offer in most cases a good approximation to the modelled network dynamics on the time-scale of the experiment. However, in some scenarios, second-order reactions, or more complex equations, govern the material flows. In other cases, the rate coefficients are not constant over the time-scale of the experiment (e.g. time-dependent rate of glucose uptake from bloodstream after injection). The `isotracer` package can be extended with more complex models to be able to model such cases, including cases where several sources must react together following stoichiometric constraints (e.g. kinetic flux profiling of known metabolic pathways; Yuan et al., 2006). Such a task would involve extending both the Stan code (to calculate the expected compartment trajectories in the Stan model) and the R code (to provide the interface to those new models and to define the post-run methods used, for example, to perform posterior predictive checks on the Stan output). We hope that feedback from users interested in package extension with new models will help prioritize which models should be implemented in later versions.

In summary, we have presented a package that implements and extends López-Sepulcre et al. (2020)'s method to analyse isotope tracer addition experiments in a diversity of systems and levels of organization, and under a variety of experimental designs. This method allows for the quantification of uncertainty in the estimation of matter and nutrient flows, turnover times and other derived properties. As such, it provides a statistical framework allowing for reliable inferences in comparative or experimental contexts, compared to some previous approaches such as manually solving mass-balance equations iteratively. The package in its current format already fulfils the needs of a majority of short-term experiments at typically studied scales, and it could be extended to accommodate new analyses and scenarios in future versions. We believe it represents a valuable tool to advance the study of the dynamics of chemical elements in tissues, organisms and ecosystems.

## ACKNOWLEDGEMENTS

We thank Sarah Collins, Sebastiano De Bona, Rana El-Sabaawi, Swanne Gordon, Sophia Lambert, Alex Lee and Steven Thomas for discussions. Sarah Collins and Jonathan Benstead helped test the package interface. We are also grateful to an anonymous reviewer and to Prof. Bob Hall for their thorough and constructive feedback on the submitted manuscript. Funding was provided by the Academy of Finland (grant number 295941).

## CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

## AUTHORS' CONTRIBUTIONS

A.L.-S. conceived the original methodology and the initial R code; M.B. implemented the method into an R package, implementing Stan integration and improving efficiency and computability. Both authors participated in the writing of the manuscript.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13822>.

## DATA AVAILABILITY STATEMENT

The `isotracer` package is available from CRAN at <https://cran.r-project.org/package=isotracer>. The CRAN package contains the datasets used in the case studies presented in this manuscript.

## ORCID

Matthieu Bruneaux  <https://orcid.org/0000-0001-6997-192X>

Andrés López-Sepulcre  <https://orcid.org/0000-0001-9708-0788>

## REFERENCES

- Ainsworth, C. H., Kaplan, I. C., Levin, P. S., & Mangel, M. (2010). A statistical approach for estimating fish diet compositions from multiple data sources: Gulf of California case study. *Ecological Applications*, 20(8), 2188–2202. <https://doi.org/10.1890/09-0611.1>
- Allen, D. K., & Young, J. D. (2020). Tracing metabolic flux through time and space with isotope labeling experiments. *Current*

- Opinion in Biotechnology*, 64, 92–100. <https://doi.org/10.1016/j.copbio.2019.11.003>
- Bacher, A., Chen, F., & Eisenreich, W. (2016). Decoding biosynthetic pathways in plants by pulse-chase strategies using  $^{13}\text{C}_2$  as a universal tracer. *Metabolites*, 6(3), 21. <https://doi.org/10.3390/metab6030021>
- Barneche, D. R., Hulatt, C. J., Dossena, M., Padfield, D., Woodward, G., Trimmer, M., & Yvon-Durocher, G. (2021). Warming impairs trophic transfer efficiency in a long-term field experiment. *Nature*, 592(7852), 76–79. <https://doi.org/10.1038/s41586-021-03352-2>
- Barthelemy, H., Stark, S., Michelsen, A., & Olofsson, J. (2018). Urine is an important nitrogen source for plants irrespective of vegetation composition in an Arctic tundra: Insights from a  $^{15}\text{N}$ -enriched urea tracer experiment. *Journal of Ecology*, 106(1), 367–378. <https://doi.org/10.1111/1365-2745.12820>
- Cabana, G., & Rasmussen, J. B. (1994). Modelling food chain structure and contaminant bioaccumulation using stable nitrogen isotopes. *Nature*, 372(6503), 255–257. <https://doi.org/10.1038/372255a0>
- Carbone, M. S., Czimczik, C. I., McDuffee, K. E., & Trumbore, S. E. (2007). Allocation and residence time of photosynthetic products in a boreal forest using a low-level  $^{14}\text{C}$  pulse-chase labeling technique. *Global Change Biology*, 13(2), 466–477. <https://doi.org/10.1111/j.1365-2486.2006.01300.x>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32. <https://doi.org/10.18637/jss.v076.i01>
- Collins, S. M., Thomas, S. A., Heatherly, T., MacNeill, K. L., Leduc, A. O. H. C., López-Sepulcre, A., Lamphere, B. A., El-Sabaawi, R. W., Reznick, D. N., Pringle, C. M., & Flecker, A. S. (2016). Fish introductions and light modulate food web fluxes in tropical streams: A whole-ecosystem experimental approach. *Ecology*, 97, 3154–3166. <https://doi.org/10.1002/ecy.1530>
- Crossley, D. A., & Reichle, D. E. (1969). Analysis of transient behavior of radioisotopes in insect food chains. *Bioscience*, 19(4), 341–343. <https://doi.org/10.2307/1294518>
- de Goeij, J. M., van Oevelen, D., Vermeij, M. J. A., Osinga, R., Middelburg, J. J., de Goeij, A. F. P. M., & Admiraal, W. (2013). Surviving in a marine desert: The sponge loop retains resources within coral reefs. *Science*, 342(6154), 108–110. <https://doi.org/10.1126/science.1241981>
- Dodds, W. K., Collins, S. M., Hamilton, S. K., Tank, J. L., Johnson, S., Webster, J. R., Simon, K. S., Whiles, M. R., Rantala, H. M., McDowell, W. H., Peterson, S. D., Riis, T., Crenshaw, C. L., Thomas, S. A., Kristensen, P. B., Cheever, B. M., Flecker, A. S., Griffiths, N. A., Crowl, T., ... Martí, E. (2014). You are not always what we think you eat: Selective assimilation across multiple whole-stream isotopic tracer studies. *Ecology*, 95(10), 2757–2767. <https://doi.org/10.1890/13-2276.1>
- Dodds, W. K., Evans-White, M. A., Gerlanc, N. M., Gray, L., Gudder, D. A., Kemp, M. J., López, A. L., Stagliano, D., Strauss, E. A., Tank, J. L., Whiles, M. R., & Wollheim, W. M. (2000). Quantification of the nitrogen cycle in a prairie stream. *Ecosystems*, 3(6), 574–589. <https://doi.org/10.1007/s100210000050>
- Freeman, C. J., Thacker, R. W., Baker, D. M., & Fogel, M. L. (2013). Quality or quantity: Is nutrient transfer driven more by symbiont identity and productivity than by symbiont abundance? *The ISME Journal*, 7(6), 1116–1125. <https://doi.org/10.1038/ismej.2013.7>
- Fry, B. (2006). *Stable isotope ecology*. Springer-Verlag. ISBN 978-0-387-30513-4.
- Gelman, A., Bois, F., & Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, 91(436), 1400–1412. <https://doi.org/10.1080/01621459.1996.10476708>
- Gillespie, W. R., Zhang, Y., Margossian, C., & Torsten Development Team. (2021). Torsten, a pharmacokinetics/pharmacodynamics library for Stan (in development). <https://github.com/metrumresearchgroup/torsten>
- Goodale, C. L., Fredriksen, G., Weiss, M. S., McCalley, C. K., Sparks, J. P., & Thomas, S. A. (2015). Soil processes drive seasonal variation in retention of  $^{15}\text{N}$  tracers in a deciduous forest catchment. *Ecology*, 96(10), 2653–2668.
- Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology*, 36(1), 39–56. <https://doi.org/10.1085/jgp.36.1.39>
- Kim, I.-Y., Suh, S.-H., Lee, I.-K., & Wolfe, R. R. (2016). Applications of stable, nonradioactive isotope tracers in in vivo human metabolic research. *Experimental & Molecular Medicine*, 48(1), e203. <https://doi.org/10.1038/emm.2015.97>
- Kulmatiski, A., Beard, K. H., Verweij, R. J. T., & February, E. C. (2010). A depth-controlled tracer technique measures vertical, horizontal and temporal patterns of water use by trees and grasses in a subtropical savanna. *New Phytologist*, 188(1), 199–209. <https://doi.org/10.1111/j.1469-8137.2010.03338.x>
- Li, L., Nelson, C. J., Troesch, J., Castleden, I., Huang, S., & Millar, A. H. (2018). Data from: Protein degradation rate in *Arabidopsis thaliana* leaf growth and development. <https://doi.org/10.5061/dryad.q3h85>
- Li, L., Nelson, C. J., Trösch, J., Castleden, I., Huang, S., & Millar, A. H. (2017). Protein degradation rate in *Arabidopsis thaliana* leaf growth and development. *The Plant Cell*, 29(2), 207–228. <https://doi.org/10.1105/tpc.16.00768>
- López-Sepulcre, A., Bruneaux, M., Collins, S. M., el-Sabaawi, R., Flecker, A. S., & Thomas, S. A. (2020). A new method to reconstruct quantitative food webs and nutrient flows from isotope tracer addition experiments. *The American Naturalist*, 195(6), 964–985. <https://doi.org/10.1086/708546>
- Lunn, D. J., Best, N., Thomas, A., Wakefield, J., & Spiegelhalter, D. (2002). Bayesian analysis of population PK/PD models: General concepts and software. *Journal of Pharmacokinetics and Pharmacodynamics*, 29(3), 271–307. <https://doi.org/10.1023/A:1020206907668>
- McRoy, C. P., & Barsdate, R. J. (1970). Phosphate absorption in eelgrass. *Limnology and Oceanography*, 15(1), 6–13. <https://doi.org/10.4319/lo.1970.15.1.0006>
- Mill, A. C., Pinnegar, J. K., & Polunin, N. V. C. (2007). Explaining isotope trophic-step fractionation: Why herbivorous fish are different. *Functional Ecology*, 21(6), 1137–1145. <https://doi.org/10.1111/j.1365-2435.2007.01330.x>
- Mulholland, P. J., Tank, J. L., Sanzone, D. M., Wollheim, W. M., Peterson, B. J., Webster, J. R., & Meyer, J. L. (2000). Nitrogen cycling in a forest stream determined by a  $^{15}\text{N}$  tracer addition. *Ecological Monographs*, 70(3), 471–493. <https://doi.org/10.2307/2657212>
- Norman, B. C., Whiles, M. R., Collins, S. M., Flecker, A. S., Hamilton, S. K., Johnson, S. L., Rosi, E. J., Ashkenas, L. R., Bowden, W. B., Crenshaw, C. L., Crowl, T., Dodds, W. K., Hall, R. O., el-Sabaawi, R., Griffiths, N. A., Martí, E., McDowell, W. H., Peterson, S. D., Rantala, H. M., ... Webster, J. R. (2017). Drivers of nitrogen transfer in stream food webs across continents. *Ecology*, 98(12), 3044–3055. <https://doi.org/10.1002/ecy.2009>
- Riis, T., Dodds, W. K., Kristensen, P. B., & Baisner, A. J. (2012). Nitrogen cycling and dynamics in a macrophyte-rich stream as determined by a  $^{15}\text{N-NH}_4^+$  release. *Freshwater Biology*, 57(8), 1579–1591. <https://doi.org/10.1111/j.1365-2427.2012.02819.x>
- Robinson, D. (2001).  $\delta^{15}\text{N}$  as an integrator of the nitrogen cycle. *Trends in Ecology & Evolution*, 16(3), 153–162. [https://doi.org/10.1016/S0169-5347\(00\)02098-X](https://doi.org/10.1016/S0169-5347(00)02098-X)
- Rønnestad, I., Dominguez, R. P., & Tanaka, M. (2000). Ontogeny of digestive tract functionality in Japanese flounder, *Paralichthys olivaceus* studied by in vivo microinjection: pH and assimilation of free amino acids. *Fish Physiology and Biochemistry*, 22(3), 225–235. <https://doi.org/10.1023/A:1007801510056>



- Sánchez-Carrillo, S., & Álvarez Cobelas, M. (2017). Stable isotopes as tracers in aquatic ecosystems. *Environmental Reviews*, 26(1), 69–81. <https://doi.org/10.1139/er-2017-0040>
- Shik, J. Z., Rytter, W., Arnan, X., & Michelsen, A. (2018). Disentangling nutritional pathways linking leafcutter ants and their co-evolved fungal symbionts using stable isotopes. *Ecology*, 99(9), 1999–2009. <https://doi.org/10.1002/ecy.2431>
- Simard, S. M., Durall, D., & Jones, M. (1997). Carbon allocation and carbon transfer between *Betula papyrifera* and *Pseudotsuga menziesii* seedlings using a  $^{13}\text{C}$  pulse-labeling method. *Plant and Soil*, 191, 41–55. <https://doi.org/10.1023/A:1004205727882>
- Stan Development Team (2020). RStan: The R interface to Stan. R package version 2.21.2. <https://cran.r-project.org/package=rstan>
- Tank, J. L., Martí, E., Riis, T., Schiller, D., Reisinger, A. J., Dodds, W. K., Whiles, M. R., Ashkenas, L. R., Bowden, W. B., Collins, S. M., Crenshaw, C. L., Crowl, T. A., Griffiths, N. A., Grimm, N. B., Hamilton, S. K., Johnson, S. L., McDowell, W. H., Norman, B. M., Rosi, E. J., ... Webster, J. R. (2018). Partitioning assimilatory nitrogen uptake in streams: An analysis of stable isotope tracer additions across continents. *Ecological Monographs*, 88(1), 120–138. <https://doi.org/10.1002/ecm.1280>
- Weindl, D., Wegner, A., & Hiller, K. (2015). Metabolome-wide analysis of stable isotope labeling—Is it worth the effort? *Frontiers in Physiology*, 6, 1–3. <https://doi.org/10.3389/fphys.2015.00344>
- Whiles, M. R., Hall, R. O., Dodds, W. K., Verburg, P., Huryn, A. D., Pringle, C. M., Lips, K. R., Kilham, S. S., Colón-Gaud, C., Rugenski, A. T., Peterson, S., & Connelly, S. (2013). Disease-driven amphibian declines alter ecosystem processes in a tropical stream. *Ecosystems*, 16(1), 146–157.
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>
- Williams, C. M., McCue, M. D., Sunny, N. E., Szejner-Sigal, A., Morgan, T. J., Allison, D. B., & Hahn, D. A. (2016). Cold adaptation increases rates of nutrient flow and metabolic plasticity during cold exposure in *Drosophila melanogaster*. *Proceedings of the Royal Society B*, 283(1838), 20161317. <https://doi.org/10.1098/rspb.2016.1317>
- Wollheim, W. M., Peterson, B. J., Deegan, L. A., Bahr, M., Hobbie, J. E., Jones, D., Bowden, W. B., Hershey, A. E., Kling, G. W., & Miller, M. C. (1999). A coupled field and modeling approach for the analysis of nitrogen cycling in streams. *Journal of the North American Benthological Society*, 18(2), 199–221. <https://doi.org/10.2307/1468461>
- Yuan, J., Fowler, W. U., Kimball, E., Lu, W., & Rabinowitz, J. D. (2006). Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*. *Nature Chemical Biology*, 2(10), 529–530. <https://doi.org/10.1038/nchembio816>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Bruneaux, M. & López-Sepulcre, A. (2022). *isotracer*: An R package for the analysis of tracer addition experiments. *Methods in Ecology and Evolution*, 13, 1119–1134. <https://doi.org/10.1111/2041-210X.13822>