



1 **PyAges v1.0: an extensible Python toolkit for** 2 **inferring groundwater transit-time** 3 **distributions from environmental tracers**

4 Jean-Raynald de Dreuzy^{1*}, Sarah Leray², Jean Marçais³

5 ¹ Univ Rennes, CNRS, Géosciences Rennes – UMR 6118, F-35000 Rennes, France.

6 ² HydroSciences Montpellier (HSM), Université de Montpellier, CNRS, IRD, F-34093 Montpellier Cedex 5, France.

7 ³ INRAE, UR RiverLy, F-69625 Villeurbanne, France.

8 * Corresponding author: Jean-Raynald.de-Dreuzy@univ-rennes.fr

9 **Abstract**

10 Lumped-parameter models infer groundwater transit-time distributions by combining environmental-tracer
11 input histories with parameterized age distributions. Their practical use requires flexible tracer
12 representation, reliable repeated numerical integration, and explicit treatment of uncertainty when different
13 parameter combinations fit observations similarly. We present PyAges v1.0, an open-source Python toolkit
14 that separates tracer definitions, lumped-parameter models, convolution, and calibration; supports Bayesian
15 inference of parameter uncertainty; and uses an integration scheme based on the cumulative distribution
16 function and partial first moment for repeated evaluation of narrow, shifted, mixed, and shape-free transit-
17 time distributions. Independent reference calculations across 270 cases characterize numerical
18 discrepancies of the PyAges convolution operator across four classical model families. A cross-software
19 benchmark against TracerLPM shows that both tools estimate the prescribed parameters with broadly
20 comparable accuracy. Mean transit time is estimated more robustly under added noise than the model-
21 specific parameter controlling either exponential–piston mixing or dispersion. Two field sites, Holten and
22 Ploemeur, illustrate the framework with real observations. At Holten, PyAges posterior medians reproduce
23 all 28 published four-bin age fractions within 0.02. At Ploemeur, calibrations using the full record or only
24 2014–2015 observations show how repeated measurements can resolve age ambiguity or reveal



25 incompatibility among the assumptions used to interpret the tracers. PyAges therefore provides a
26 reproducible framework for testing tracer and transit-time hypotheses, determining what observations
27 constrain, and integrating tracer inference into broader hydrogeological workflows.

28 **Keywords**

29 Environmental tracers, groundwater dating, lumped-parameter models, transit-time distributions, Bayesian
30 inference, uncertainty quantification

31 **1. Introduction**

32 Environmental and geogenic tracers constrain groundwater-system structure, recharge, surface-water
33 interactions, and transit times (Cook and Herczeg, 2000; Popp et al., 2021). Because they integrate signals
34 along flow paths from recharge to sampling locations, tracers complement hydraulic observations such as
35 piezometric heads (Bethke and Johnson, 2002; Ginn et al., 2009). Their use nevertheless requires rigorous
36 sampling, specialized analytical facilities, and careful separation of transport effects from biogeochemical
37 reactions and water–rock interaction (Appelo and Postma, 1993). Translating tracer observations into
38 quantities relevant to groundwater management therefore depends on an appropriate modeling framework
39 (Clark, 2015; Kazemi et al., 2006; Leibundgut et al., 2009). A central challenge is that tracer records and
40 model structure often admit multiple transit-time distributions that fit the observations. Even with several
41 tracers, correlated input histories and limited age coverage can leave parameters weakly constrained.
42 Interpretation tools should therefore quantify uncertainty, posterior correlations, and trade-offs rather than
43 return only a single optimum (Massoudieh et al., 2012, 2014b; Thiros et al., 2023).

44 Lumped-parameter models (LPMs) represent groundwater flow and mixing with transit-time distributions
45 defined by a small set of parameters (Małoszewski and Zuber, 1982, 1985). Classical formulations range
46 from piston and exponential mixing to dispersion and shifted models; multi-component and shape-free
47 representations provide additional flexibility (Eberts et al., 2012; Engdahl et al., 2013; Engdahl and



48 Maxwell, 2014; Massoudieh et al., 2014b; McCallum et al., 2014). PyAges considers conservative tracers
49 and tracers whose simple production or loss kinetics are specified independently. The same formalism can
50 in principle accommodate non-conventional geochemical proxies, such as dissolved silica, when their
51 evolution during transport can be represented by the simple prescribed production or loss laws supported
52 here (Marçais et al., 2018). The theoretical basis of LPMs is well established, and analytical convolution
53 solutions exist for selected combinations of distributions and input functions (Jódar et al., 2014).
54 Historically, many widely used practical implementations have been delivered as spreadsheets or desktop
55 codes, including TracerLPM, Tracermodell1, TRACER, GTI, Lumpy, and LUMPED (Aeschbach et al.,
56 2006; Bayari, 2002; Böhlke, 2006; Dávila et al., 2013; IAEA, 2006; Jurgens et al., 2012; Ozyurt and Bayari,
57 2003; Suckow, 2012). These tools remain useful for groundwater-age interpretation but are less readily
58 integrated into automated testing, scripted uncertainty analysis, and reproducible Python workflows.

59 Recent Python developments broaden this software ecosystem. `mesas.py` provides a flexible
60 implementation of StorAge Selection functions for solute-transport and transit-time calculations (Harman
61 and Xu Fei, 2024), and RoGeR couples process-based hydrology with travel-time calculations in a Python
62 toolbox (Schwemmle et al., 2024). Related TTD concepts are also implemented in other tracer domains,
63 for example TRACE-Python for transient ocean tracers (Sandborn et al., 2026). PyTracerLab is a recent
64 preprint describing an open-source Python framework for LPM analysis that combines programmatic
65 workflows, a standalone graphical interface, deterministic optimization, and DREAM-based Bayesian
66 sampling (Rudolph et al., 2026). PyAges complements these developments through the combination of four
67 features: scriptable separation of tracer, LPM, convolution, and calibration components; independent
68 numerical characterization of the forward operator; posterior analysis of parameter identifiability; and case-
69 level reproducibility of the analyses reported in the manuscript. PyAges' contribution lies not in a new
70 Bayesian algorithm or LPM family, but in their integration within a tested and reproducible framework.
71 The following sections describe the scientific and software architecture, qualify the forward and inverse



calculations independently, and illustrate the uncertainty-aware workflow with synthetic and field applications.

2. Models and methods

This section presents the conceptual and computational foundations of PyAges. We first introduce the lumped-parameter formulation of tracer transport and the Bayesian framework used for parameter inference and uncertainty quantification, before describing how these elements are implemented in the modular software architecture, including tracer handling, LPMs, convolution, and calibration.

2.1. Theoretical foundations of LPMs and Bayesian calibration

2.1.1. Lumped-parameter transit-time modeling and tracer representation

For a sample collected at location x and time t , the LPM framework expresses the tracer concentration $c(x, t)$ as the convolution of the recharge input history with the transit-time density $p(x, \tau)$, where τ denotes the elapsed time between recharge and arrival at the sampling location:

$$c(x, t) = \int_0^\infty c_T(t - \tau, \tau) p(x, \tau) d\tau \quad (1)$$

PyAges represents tracer transformations during transit with an effective constant zeroth-order production rate α and a first-order loss rate β :

$$c_T(t - \tau, \tau) = \begin{cases} \alpha\tau + c_T^0(t - \tau), & \beta = 0, \\ \frac{\alpha}{\beta}(1 - e^{-\beta\tau}) + c_T^0(t - \tau)e^{-\beta\tau}, & \beta > 0. \end{cases} \quad (2)$$

In Eq. (2), $c_T(t - \tau, \tau)$ is the concentration contributed by water recharged at $t - \tau$ after a transit time τ , $c_T^0(t - \tau)$ is the recharge input function, α is the zeroth-order production rate, and β is the first-order loss rate. Conservative tracers correspond to $\alpha = \beta = 0$; radioactive decay is represented by $\beta = \ln(2)/t_{1/2}$, where $t_{1/2}$ denotes the tracer half-life; and $\beta = 0$ yields linear accumulation at rate α . For $\beta > 0$, the production-loss contribution approaches the asymptotic value α/β . The framework assumes a stationary transit-time distribution and tracer transformations that depend only on transit time, with recharge input



92 histories and tracer-specific kinetic parameters specified externally. Within this framework, different
 93 groundwater flow and mixing organizations can be represented through alternative transit-time distributions
 94 (Amin and Campana, 1996; Leray et al., 2016). PyAges implements classical and extended parametric
 95 formulations, including shifted, mixed, and Weibull models, together with a shape-free representation;
 96 selected built-in models are listed in Table 1.

97 **Table 1.** Selected lumped-parameter transit-time distributions implemented in PyAges.

LPM	Parameters	Expression
Dirac (piston flow)	μ	$p_{\mu}(\tau) = \delta(\tau - \mu)$
Exponential (perfect mixing)	μ	$p_{\mu}(\tau) = \left(\frac{1}{\mu}\right) e^{-\frac{\tau}{\mu}}$
Inverse Gaussian (dispersion)	μ, σ	$p_{\mu,\sigma}(\tau) = \sqrt{\frac{\mu^3}{2\pi\sigma^2\tau^3}} e^{-\mu(\tau-\mu)^2/(2\sigma^2\tau)}$
Shifted exponential (exponential-piston flow)	μ, t_0	$p_{\mu,t_0}(\tau) = \begin{cases} 0 & \text{for } \tau < t_0 \\ \frac{1}{\mu} e^{-\frac{\tau-t_0}{\mu}} & \text{otherwise} \end{cases}$
Uniform (linear piston flow)	$t_{\min}, \Delta t$	$p_{t_{\min},\Delta t}(\tau) = \begin{cases} \frac{1}{\Delta t} & \text{for } t_{\min} \leq \tau < t_{\min} + \Delta t \\ 0 & \text{otherwise} \end{cases}$
Gamma	k, θ	$p(\tau) = \frac{\tau^{k-1} \exp\left[-\frac{\tau}{\theta}\right]}{\Gamma(k)\theta^k}, \tau \geq 0$
Double Dirac (binary mixing)	μ_1, μ_2, r	$p(\tau) = r\delta(\tau - \mu_1) + (1-r)\delta(\tau - \mu_1 - \mu_2), 0 \leq r \leq 1$
Shifted inverse Gaussian	μ, σ, t_0	$p_{\mu,\sigma,t_0}(\tau) = \begin{cases} 0 & \text{for } \tau < t_0 \\ \sqrt{\frac{\mu^3}{2\pi\sigma^2(\tau-t_0)^3}} \exp\left(-\frac{\mu(\tau-t_0-\mu)^2}{2\sigma^2(\tau-t_0)}\right) & \text{otherwise} \end{cases}$
Dirac + shifted exponential	μ_1, μ_2, t_0, r	$p(\tau) = r\delta(\tau - \mu_1) + (1-r)h(\tau)$ $h(\tau) = 0, \tau < \mu_1 + t_0$ $h(\tau) = \frac{1}{\mu_2} \exp\left[-\frac{\tau - \mu_1 - t_0}{\mu_2}\right], \tau \geq \mu_1 + t_0$
Weibull	k, λ	$p(\tau) = \frac{k}{\lambda} \left(\frac{\tau}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{\tau}{\lambda}\right)^k\right], \tau \geq 0$



LPM	Parameters	Expression
Shape-free model	$t_0, \dots, t_n;$ $f_0, \dots, f_{n-1}$	$p(\tau) = \frac{f_i}{t_{i+1} - t_i}, t_i \leq \tau < t_{i+1}, i = 0, \dots, n-1$ $p(\tau) = 0, \tau \notin [t_0, t_n)$ $\sum_{i=0}^{n-1} f_i = 1$

98 The parameterizations in Table 1 are those implemented in PyAges and used consistently throughout the
 99 manuscript. For the shifted exponential distribution, μ denotes the exponential timescale and the mean
 100 transit time is $\mu + t_0$. For the inverse Gaussian, μ and σ denote the mean and standard deviation,
 101 respectively; for the shifted inverse Gaussian, the total mean transit time is $\mu + t_0$. In the double-Dirac
 102 formulation, μ_2 denotes the additional delay between the first and second Dirac components, so that the
 103 second Dirac component occurs at $\mu_1 + \mu_2$. The mixed Dirac-shifted-exponential model contains a discrete
 104 mass r at μ_1 and a normalized continuous component of mass $1 - r$ beginning at $\mu_1 + t_0$. In the mixed
 105 Dirac-shifted-exponential model, μ_2 is the exponential scale and mean measured from the lower support μ_1
 106 $+ t_0$; the mean age of the continuous component is therefore $\mu_1 + t_0 + \mu_2$.

107 The generic tracer representations used or discussed in this study and supported by PyAges are summarized
 108 in Table 2. PyAges requires tracer input histories and observations to be expressed in mutually consistent
 109 units. For atmospheric gas tracers such as CFCs and SF₆, the input histories used here are atmospheric-
 110 equivalent mixing ratios (pptv), which are used directly in the convolution. Any conversion required to
 111 place observations and tracer input histories on a consistent scale is not performed by PyAges and must be
 112 handled externally as a preprocessing step. Foundational applications and tracer behavior are described for
 113 CFCs and SF₆ by Busenberg and Plummer (1992, 2000), for tritium by Solomon and Cook (2000), and in
 114 the review of Chambers et al. (2019). The modern portions of the CFC and SF₆ histories used in this work
 115 are based on NOAA Global Monitoring Laboratory datasets (Dutton et al., 2023, 2025a, b, c); the generic
 116 tritium history is derived from the IAEA/WMO GNIP/WISER database (IAEA and WMO, 2026), and the
 117 ³⁹Ar implementation uses a constant normalized modern atmospheric abundance. ⁸⁵Kr and tritiogenic ³He



are used with site-specific definitions in the Holten benchmark (Sect. 4.1) and are therefore not included among the generic tracer representations in Table 2.

Table 2. Generic tracer representations used or discussed in the present study and supported by PyAges.

Tracer	Origin	Input function	Reactivity / transformation
CFC-11, CFC-12, CFC-113	Anthropogenic atmospheric	Semi-annual atmospheric-equivalent mixing-ratio history	Conservative by default; optional user-specified first-order loss
SF ₆	Anthropogenic atmospheric; possible lithogenic contributions in some settings	Semi-annual atmospheric-equivalent mixing-ratio history	Conservative by default; optional user-specified zeroth-order production
³ H (tritium)	Natural and bomb-produced atmospheric	Semi-annual precipitation record	Radioactive first-order decay ($t_{1/2} = 12.32$ yr)
³⁹ Ar	Cosmogenic atmospheric	Constant normalized modern abundance	Radioactive first-order decay ($t_{1/2} = 267$ yr)

2.1.2. Bayesian formulation and MCMC solution for LPM calibration from tracer data

Calibration of lumped-parameter models (LPMs) from tracer observations is formulated within a Bayesian framework following Tarantola (2005) and Gelman et al. (2013). Bayesian inference is widely used in groundwater hydrology for hydraulic and flow-related inverse problems (Carrera and Neuman, 1986; de Marsily et al., 2000; Kitanidis, 1986; Rojas et al., 2008) and was introduced for groundwater age-distribution modeling using environmental tracers by Massoudieh et al. (2012). Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ denote the vector of LPM parameters and $\mathbf{c}^{\text{obs}} = (c_1^{\text{obs}}, \dots, c_m^{\text{obs}})$ the vector of tracer observations. The posterior probability distribution $p(\boldsymbol{\theta} | \mathbf{c}^{\text{obs}})$ is given by Bayes' theorem:

$$p(\boldsymbol{\theta} | \mathbf{c}^{\text{obs}}) = \frac{p(\mathbf{c}^{\text{obs}} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{c}^{\text{obs}})} \quad (3)$$

where $p(\mathbf{c}^{\text{obs}} | \boldsymbol{\theta})$ is the likelihood function, $p(\boldsymbol{\theta})$ is the prior distribution of the model parameters, and $p(\mathbf{c}^{\text{obs}})$ is the evidence, which acts as a normalization constant. Assuming independent Gaussian observation errors with specified standard deviations σ_j , treated as fixed within the adopted observation-error model, the likelihood is:



$$p(c^{\text{obs}} | \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{m}{2}} \prod_{j=1}^m \sigma_j} \exp \left[-\frac{1}{2} J(\boldsymbol{\theta}) \right] \quad (4)$$

133 where $c_j(\boldsymbol{\theta})$ denotes the modeled tracer value corresponding to observation j . Equation (5) defines $J(\boldsymbol{\theta})$,
 134 the uncertainty-weighted data-misfit function.

$$J(\boldsymbol{\theta}) = \sum_{j=1}^m \left(\frac{c_j(\boldsymbol{\theta}) - c_j^{\text{obs}}}{\sigma_j} \right)^2 \quad (5)$$

135 For comparison across datasets containing different numbers of observations, the normalized root-mean-
 136 square data misfit is reported as

$$137 \quad \text{RMS}_{\text{norm}}(\boldsymbol{\theta}) = \sqrt{\frac{J(\boldsymbol{\theta})}{m}}.$$

138 Prior information on LPM parameters is introduced independently through $p(\boldsymbol{\theta})$. Depending on the
 139 application, priors may be weakly informative, for example uniform distributions over physically
 140 admissible parameter ranges, or informative distributions reflecting independent hydrogeological
 141 knowledge. The posterior therefore combines the information provided by tracer observations through the
 142 likelihood with the information supplied independently through the prior. For uniform priors within
 143 prescribed parameter bounds, as used in the synthetic benchmarks presented in Sect. 3, the prior term is
 144 constant within the admissible domain and zero probability is assigned outside it. Consequently, within
 145 these bounds, variations in the posterior are controlled directly by the data likelihood.

146 The posterior is explored using a Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithm
 147 (Hastings, 1970; Metropolis et al., 1953), with Gaussian proposals and an acceptance probability based on
 148 posterior probability ratios. MCMC sampling characterizes the posterior distribution rather than only its
 149 optimum, allowing uncertainty, correlations, and parameter trade-offs to be quantified. Proposal covariance
 150 may be diagonal or correlated, and burn-in is discarded before posterior summaries are computed (Roberts
 151 et al., 1997; Robert and Casella, 2009).



152 Convergence and sampling efficiency are assessed from multiple chains using split-R-hat and effective
 153 sample size (ESS) diagnostics. Split-R-hat measures the agreement among chains, with values close to 1
 154 indicating that they sample the same posterior distribution. ESS estimates the amount of effectively
 155 independent information contained in the correlated MCMC samples, with larger values indicating more
 156 efficient sampling. The exact ESS implementation is campaign-specific, so diagnostic values are interpreted
 157 within each campaign rather than compared across campaigns. In all analyses, split-R-hat < 1.01 and ESS
 158 ≥ 300 are used as practical convergence and sampling-efficiency thresholds. The general use and
 159 interpretation of these diagnostics follow Vehtari et al. (2021).

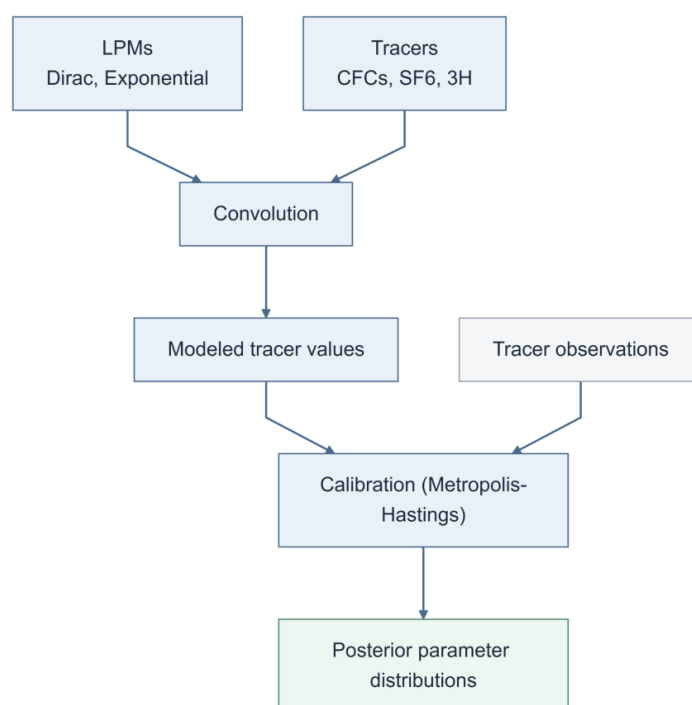
160 **2.2. Extensible software architecture for tracer-based modeling**

161 **2.2.1. Repository structure and modular design**

162 The repository separates the installable package from distributed tracer and LPM data, workflows,
 163 reproducible examples, tests, documentation, and environment specifications, while the package itself is
 164 organized around four components that mirror the scientific workflow (Fig. 1). The Tracer module supplies
 165 the tracer-specific input and transformation term $c_T(t - \tau, \tau)$ in Eq. (1), the LPM module supplies $p(\tau)$,
 166 the Convolution module combines both to compute modeled tracer values, and the Calibration module
 167 compares these values with observations through the Bayesian formulation in Eqs. (3)–(5). These
 168 components communicate through common interfaces, keeping scientific code, configuration, validation,
 169 and user-facing material distinct. This modular design makes extension straightforward: a new tracer can
 170 be introduced through a self-contained tracer definition and input history, while a new transit-time
 171 distribution can be implemented as an LPM with its own parameters and probability law. Once a tracer
 172 definition is placed in the expected data structure, or an LPM is registered through the shared interface, it
 173 becomes available to the same convolution and calibration workflows without requiring changes to the
 174 other components.



175 User-facing workflow configuration files are YAML mappings validated before execution with strict
 176 Pydantic models. Unknown fields are rejected, expected types and admissible ranges are checked, and path
 177 and name constraints are validated so malformed configurations fail early with explicit errors. This schema
 178 validation applies to workflow configuration; tracer histories and observation files retain their documented
 179 tabular formats and are checked separately when loaded.



180

181 **Figure 1.** Core organization of PyAges. Tracer definitions and LPMs provide the tracer-response and
 182 transit-time-distribution components used by the convolution module to calculate modeled tracer values.
 183 Calibration compares these modeled values with tracer observations and samples parameter combinations,
 184 here illustrated with the Metropolis–Hastings workflow, to obtain posterior parameter distributions. The
 185 tracer and LPM names shown are examples rather than an exhaustive list.



186 2.2.2. Tracer framework and extensibility

187 The configurable Tracer class represents the tracer-specific term $c_T(t - \tau, \tau)$ in Eq. (1). Each tracer
188 definition combines (i) an input history specified either as a constant or as a time series, (ii) optional zeroth-
189 order production, (iii) optional first-order loss specified through a half-life or mean lifetime and converted
190 internally to a decay rate, and (iv) the temporal domain and unit metadata. Time-series inputs are linearly
191 interpolated over their available domain and set to zero outside it. For a given sampling date and transit
192 time, the tracer object evaluates the corresponding concentration after transport according to Eq. (2). This
193 interface isolates tracer-specific assumptions from the LPM, convolution, and calibration modules.

194 File-configured tracers distributed with PyAges are each stored in their own directory under
195 `data_core/data_tracer/`. No central registry is required: PyAges identifies a tracer from the directory name
196 and loads the corresponding configuration file `<name>/<name>.yaml`. If the configuration specifies `recharge:`
197 `true`, the associated `recharge.csv` file is loaded from the same directory. New tracer directories can be
198 created automatically with `pyages new tracer <name>` or manually following the same structure. The
199 packaged tracers can be listed with `pyages list tracers`. Site- or region-specific definitions can be used
200 by setting `tracers.data_directory` in the workflow configuration, or the corresponding API parameter,
201 and then become available to the standard convolution and calibration workflows without changes to the
202 core code. For synthetic or programmatic studies, PyAges also provides `SyntheticTracer`, a concrete
203 implementation of `TracerProtocol` whose response is supplied by a Python function and which is accepted
204 by the same `Convolution` interface. PyAges does not perform physical-unit conversion or dimensional
205 analysis. It does, however, validate explicit unit labels, require one exact observation unit per tracer, and
206 check exact equality between observation and modeled-tracer units before calibration. Input histories and
207 observations must therefore still be prepared on mutually consistent scales.



2.2.3. LPM framework and extensibility

Lumped-parameter models share a common interface for parameter management and distribution functions, including probability density functions, cumulative distribution functions (CDFs), inverse CDFs, and moments. For the default continuous-convolution pathway, PyAges also uses the partial first moment

$$M(t; \boldsymbol{\theta}) = \int_0^t \tau p(\tau; \boldsymbol{\theta}) d\tau = \mathbb{E}_{\boldsymbol{\theta}}[T \mathbf{1}_{\{T \leq t\}}].$$

The built-in continuous LPM families provide analytical or vectorized expressions for both the CDF and the partial first moment. If an external model provides a CDF but no partial first moment, PyAges can use a numerical fallback, although this is less efficient for long repeated calibrations. Model-specific configuration files store parameter bounds, prior definitions, initial values, and proposal settings.

Adding a new transit-time distribution therefore requires defining its parameterization and distribution functions and registering the model through the shared LPM interface; no changes to the tracer, convolution, or calibration layers are needed. For efficient repeated use, a vectorized CDF and preferably an analytical or vectorized partial first moment are recommended, and dedicated tests should verify normalization, moments, CDF/quantile consistency, and convolution behavior. Detailed implementation guidance is provided with the project documentation. Versioned user and API documentation for PyAges v1.0 is available at <https://pyage-gw.readthedocs.io/en/stable/>.

2.2.4. CDF–partial-first-moment product integration

PyAges evaluates the convolution in Eq. (1) using a piecewise-linear product-integration approach (Sloan and Smith, 1982). The key idea is to avoid resolving the potentially narrow or strongly peaked LPM density directly. Instead, the tracer response is approximated on an age grid, while the LPM is integrated analytically over each age interval through its cumulative distribution function and partial first moment.



For a fixed sampling time t , let $K(\tau) = c_T(t - \tau, \tau)$ denote the tracer response associated with transit time τ . PyAges represents $K(\tau)$ as piecewise linear over the age grid and integrates this approximation with respect to the LPM distribution.

$$w_i = F(b_i) - F(a_i), \quad \Delta M_i = M(b_i) - M(a_i). \quad (6)$$

For each interval $[a_i, b_i]$, w_i is the exact probability mass of the LPM within the interval, while ΔM_i is its corresponding first-moment increment, with F the CDF and M the partial first moment defined above. The tracer response is approximated linearly within the interval as

$$K(\tau) \simeq K(a_i) + s_i(\tau - a_i), \quad s_i = \frac{K(b_i) - K(a_i)}{b_i - a_i}. \quad (7)$$

The resulting interval contribution is

$$C_i = K(a_i)w_i + s_i[\Delta M_i - a_i w_i], \quad C \simeq \sum_i C_i. \quad (8)$$

Equation (8) is exact whenever $K(\tau)$ is linear within the interval. Consequently, a narrow or shifted LPM does not require the probability density itself to be resolved by a fine age grid: within each interval, its contribution is determined from the exact interval probability mass and partial first-moment increment. The formulation above applies to continuous transit-time distributions. Discrete components are simpler: a Dirac mass at transit time τ_0 is evaluated directly from the tracer response $K(\tau_0)$, without numerical integration. Mixed discrete–continuous models combine this direct contribution with the CDF–partial-first-moment integration of the continuous component.

For continuous LPMs, the tracer-response age grid is constructed from the input-history nodes for each sampling date and is independent of the LPM parameters. For conservative time-series tracers, $K(\tau)$ is affine between successive input nodes, so Eq. (8) is exact on those intervals. When tracer transformations introduce additional nonlinearity, PyAges refines the grid locally; retained intervals use Eq. (8) when the midpoint confirms near-linearity and otherwise use a midpoint contribution. The grid is constructed once for each tracer–sampling-date pair and reused throughout calibration. The adaptive refinement criterion, default tolerances, and fallback are detailed in Appendix A1.



250 **2.2.5. Calibration**

251 The calibration layer provides a common interface to the forward model and objective functions. The
252 default stochastic method is the Metropolis–Hastings sampler described in Sect. 2.1.2, which returns
253 posterior parameter samples and convergence diagnostics. For Metropolis–Hastings calibration, PyAges
254 can estimate a correlated Gaussian proposal from a preliminary pilot chain. After burn-in, the empirical
255 covariance matrix of the sampled parameter vectors is regularized by a small diagonal term and then fixed
256 as the covariance structure of the Gaussian random-walk proposal used in the sampling chains. Multiple
257 independent sampling chains can be run from distinct initial states and combined for posterior summaries
258 after convergence has been assessed using the diagnostics described in Sect. 2.1.2. Deterministic
259 optimization and systematic parameter sampling are also available for complementary tasks and use the
260 same tracer, LPM, and convolution components. This separation allows the inverse strategy to be changed
261 without modifying the forward model.

262 **3. Numerical verification and synthetic calibration benchmarks**

263 The numerical assessment of PyAges is organized in complementary stages. In the terminology used here,
264 verification tests whether the selected equations are solved correctly, whereas evaluation assesses how the
265 complete framework compares with another implementation or with field-based interpretations. The
266 PyAges forward operator is compared with independent reference calculations in 270 cases spanning the
267 piston flow model (PFM), exponential mixing model (EMM), exponential piston-flow model (EPM), and
268 dispersion model (DM). Across these families, the observed absolute RMSE ranges from 2.70×10^{-13} to
269 9.90×10^{-4} and the maximum absolute discrepancy from 5.12×10^{-13} to 6.12×10^{-3} . These absolute metrics
270 are reported in the arbitrary concentration units used for the synthetic inputs. Symmetric relative
271 discrepancies can be substantially larger when reference concentrations approach zero, so absolute and
272 relative diagnostics are reported together. The inverse workflow is additionally cross-checked against
273 TracerLPM using independently generated four-tracer synthetic datasets, while a separate Bayesian



274 benchmark examines posterior uncertainty and parameter identifiability. Detailed qualification criteria are
 275 given in Appendix A2.

276 **3.1. Cross-software inverse benchmark with TracerLPM**

277 To qualify the inverse workflow against an established implementation, PyAges and TracerLPM (Jurgens
 278 et al., 2012) are applied to synthetic datasets composed of CFC-11, CFC-12, CFC-113, and SF₆, expressed
 279 as atmospheric-equivalent mixing ratios in pptv. Synthetic reference values are generated independently of
 280 both software packages from analytical LPM definitions and a separate Gauss–Legendre quadrature over
 281 the tracer input histories. Controls with 0 % added noise cover EMM, EPM, and DM. The 480 paired noisy
 282 inversions cover EPM and DM, mean transit times of 5, 20, and 50 years, several values of the second
 283 model parameter, and added relative-noise levels of 1, 5, 10, and 20 %. For each tracer i , the independently
 284 generated reference value is perturbed multiplicatively as

$$285 \quad y_i^{\text{obs}} = y_i^{\text{true}}(1 + \varepsilon_i), \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_{\text{rel}}^2), \quad \sigma_{\text{rel}} \in \{0, 0.01, 0.05, 0.10, 0.20\}.$$

286 The same predefined random seeds are used for corresponding PyAges and TracerLPM runs, so that both
 287 implementations are fitted to exactly the same noisy four-tracer realization. The two implementations retain
 288 their native inversion frameworks: PyAges uses the uncertainty-weighted quadratic misfit of Eq. (5),
 289 whereas TracerLPM minimizes a sum of absolute relative residuals (L1) through Excel Solver. The exercise
 290 is therefore a cross-software qualification rather than a ranking of optimizers. Parameter recovery across
 291 the full range of added noise is summarized in Table 3.



Table 3. Recovery of prescribed LPM parameters in the PyAges–TracerLPM benchmark as a function of added relative noise. Errors are absolute relative errors for individual configurations and median absolute relative errors where repeated realizations are available; n is the number of cases per implementation. The second parameter is η for EPM and DP for DM; their definitions and the parameter mappings used for TracerLPM are given in Appendix B.

Model	Noise (%)	n	MTT error		Second-parameter error	
			PyAges	TracerLPM	PyAges	TracerLPM
EMM	0	1	numerical precision	0.4 %	—	—
EPM	0	1	numerical precision	1.3 %	numerical precision	1.2 %
DM	0	1	numerical precision	1.6 %	numerical precision	4.0 %
EPM	1	40	0.9 %	0.8 %	3.5 %	1.7 %
DM	1	40	1.0 %	1.4 %	8.2 %	6.9 %
EPM	5	40	4.4 %	4.2 %	10.0 %	13.7 %
DM	5	40	4.7 %	4.2 %	34.0 %	36.2 %
EPM	10	80	9.3 %	8.8 %	18.2 %	10.8 %
DM	10	80	12.7 %	8.3 %	58.1 %	54.0 %
EPM	20	80	19.2 %	19.7 %	32.6 %	28.5 %
DM	20	80	28.2 %	24.2 %	89.7 %	78.6 %

Across the benchmark, PyAges and TracerLPM lead to the same overall interpretation of parameter recoverability. In the controls with 0 % added noise, PyAges reproduces the generating values to numerical precision, whereas small deviations remain in the native TracerLPM workflow. With added noise, the two implementations yield broadly comparable point estimates, and neither is uniformly superior when parameter recovery and common residual metrics are considered together. Mean transit time remains substantially more robust to increasing added noise than the second model parameter, whose recovery deteriorates rapidly between 1 and 5 % noise and becomes increasingly poorly recovered at higher noise levels. Because the native objective functions differ, a smaller residual for a particular noisy realization should not be interpreted as intrinsically better recovery of the generating parameters.



306 The common residual metrics and the parameter mappings used for the cross-software comparison are given
 307 in Appendix B; complete case-level results, parameter bounds, random seeds, software versions, and native
 308 objective values are retained in the reproducibility archive. The cross-software benchmark therefore
 309 supports broad consistency at the level of point estimates. Section 3.2 then uses PyAges' posterior output
 310 to quantify uncertainty and parameter dependence that cannot be inferred from a single optimum.

311 **3.2. Bayesian calibration, uncertainty, and identifiability**

312 Unlike Sect. 3.1, which evaluates parameter recovery after adding random noise to independently generated
 313 synthetic tracer values, the present experiment does not perturb the generating values. Instead, exact
 314 synthetic values are assigned an 8 % relative observational standard deviation in the Bayesian likelihood.
 315 The generating parameter pair therefore remains a maximum-likelihood solution by construction. The
 316 objective is not to measure recovery error for a particular noise realization, but to quantify posterior
 317 uncertainty and parameter identifiability.

318 The benchmark uses CFC-11, CFC-12, CFC-113, and sulfur hexafluoride (SF₆) (Chambers et al., 2019),
 319 treated as conservative during transport, for a sampling date of 1 January 2010. Nineteen shifted-
 320 exponential configurations cover the parameter domain $\mu \leq 40 \text{ yr}, t_0 \leq 40 \text{ yr}, \text{MTT} = \mu + t_0 \leq$
 321 50 yr . Uniform priors are used over $\mu \in [0.1, 70] \text{ yr}, t_0 \in [0, 70] \text{ yr}$. The 8 % relative uncertainty is
 322 intentionally larger than the few-percent analytical intercomparison uncertainty reported for CFCs and SF₆
 323 (Labasque et al., 2014) and is intended to represent combined analytical and sampling uncertainty in a
 324 conservative benchmark.

325 Following the calibration procedure described in Sect. 2.2.5, each case uses a 4,000-iteration pilot chain to
 326 estimate a fixed correlated proposal covariance, followed by five independent 10,000-iteration sampling
 327 chains with 20 % burn-in. All 19 cases satisfy the campaign split-R-hat < 1.01 convergence criterion. For
 328 this workflow, the reported ESS is the sum of autocorrelation-based single-chain ESS values; the minimum
 329 ESS across μ , t_0 , and MTT ranges from 756 to 4,784. Table 4 summarizes the posterior distribution of



330 MTT. For each retained posterior draw s , $MTT^{(s)} = \mu^{(s)} + t_0^{(s)}$, so the MTT quantiles preserve the posterior
 331 dependence between μ and t_0 . Full marginal summaries of μ and t_0 , 2.5–97.5 % MTT intervals, and multi-
 332 chain diagnostics are provided in Appendix C.

333 **Table 4.** Mean-transit-time posterior summaries and minimum ESS across μ , t_0 , and MTT for the 19 shifted-
 334 exponential synthetic configurations. $MTT = \mu + t_0$ is calculated sample by sample from the joint posterior.
 335 Full marginal summaries of μ and t_0 and 2.5–97.5 % MTT intervals are provided in Appendix C. Posterior
 336 MTT values are rounded to 0.1 yr.

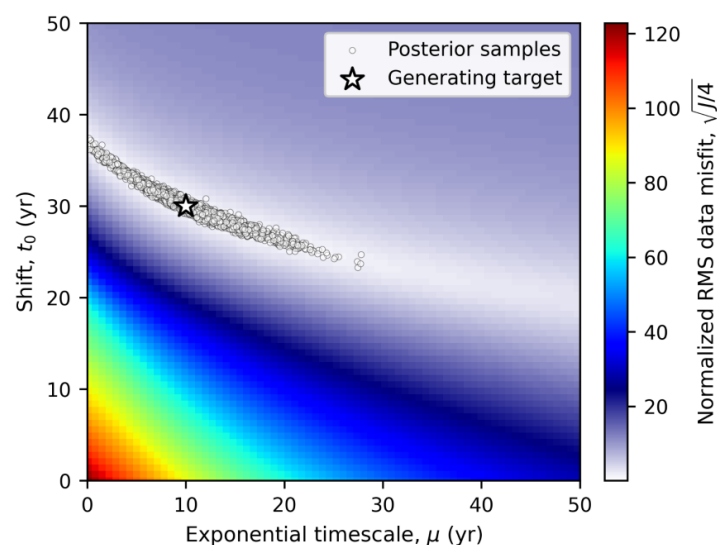
Case	Target (μ , t_0) (yr)	Target MTT (yr)	Posterior MTT median [10–90 %] (yr)	Min ESS
1	(1, 1)	2	3.5 [1.6–5.8]	2,357
2	(1, 10)	11	11.0 [9.0–13.0]	4,421
3	(1, 20)	21	21.1 [20.2–22.0]	3,978
4	(1, 30)	31	31.4 [30.7–32.6]	3,807
5	(1, 40)	41	44.0 [41.1–50.1]	1,571
6	(10, 1)	11	11.2 [8.9–13.4]	3,692
7	(10, 10)	20	20.2 [18.7–21.7]	4,733
8	(10, 20)	30	30.5 [28.1–33.8]	3,321
9	(10, 30)	40	40.3 [37.9–43.7]	2,513
10	(10, 40)	50	55.1 [47.3–73.3]	756
11	(20, 1)	21	21.0 [19.0–23.2]	3,258
12	(20, 10)	30	30.3 [28.0–33.4]	4,531
13	(20, 20)	40	41.1 [37.0–46.8]	4,284
14	(20, 30)	50	50.4 [44.4–57.7]	1,116
15	(30, 1)	31	30.4 [28.1–33.2]	4,305
16	(30, 10)	40	40.6 [36.6–45.4]	4,775
17	(30, 20)	50	51.7 [45.6–60.3]	4,439
18	(40, 1)	41	40.0 [36.7–43.6]	3,799
19	(40, 10)	50	50.7 [45.1–57.5]	4,784

337 The 8 % observational uncertainty is amplified substantially in the individual shifted-exponential
 338 parameters. Across the 19 configurations, the median posterior standard deviations relative to their target
 339 values are 32.1 % for μ , 26.6 % for t_0 , and 8.6 % for MTT. This contrast reflects the strong negative
 340 posterior dependence between μ and t_0 , whose median posterior correlation across the 19 configurations is



341 -0.95 . Their individual contributions can therefore compensate while their sum remains much better
 342 constrained. Fig. 2 illustrates this trade-off for the target $\mu = 10$ yr and $t_0 = 30$ yr, corresponding to an MTT
 343 of 40 yr. The posterior follows a negatively correlated ridge; the resulting MTT posterior has a median of
 344 40.3 yr and a central 10–90 % interval of 37.9–43.7 yr.

345 Within the present 19-case design, identifiability varies across the tested domain, with relatively better joint
 346 constraints mainly for μ between about 20 and 40 years and t_0 around 10–20 years, whereas very small
 347 parameter values and some high-shift configurations are less well resolved. All configurations nevertheless
 348 retain near-zero best-fit normalized residuals relative to the assumed observational uncertainties. The
 349 differences therefore primarily reflect parameter identifiability rather than fit quality.



350

351 **Figure 2.** Bayesian posterior geometry for the shifted-exponential benchmark with target $\mu=10$ years and
 352 $t_0=30$ years. The background shows the normalized root-mean-square data misfit, $\sqrt{J/4}$, calculated from
 353 CFC-11, CFC-12, CFC-113, and SF_6 with an assumed 8 % relative observational standard deviation; the
 354 star marks the generating parameter pair. Points show samples retained after burn-in from the five
 355 converged sampling chains. The x axis is the exponential timescale μ , and the y axis is the shift t_0 .



356 More generally, this experiment distinguishes fit from identifiability: near-zero best-fit normalized
357 residuals relative to the assumed observational uncertainties can coexist with substantial uncertainty in
358 individual LPM parameters, while a derived quantity such as MTT remains considerably better constrained.

359 **4. Field benchmark and demonstration**

360 The field applications illustrate two complementary uses of PyAges. Holten provides an external
361 benchmark against a published multi-tracer interpretation. Ploemur provides a long-term CFC
362 demonstration in which the analysis is deliberately restricted to the shifted-exponential LPM. The purpose
363 is not to revisit the full hydrogeological study or compare alternative LPM parameterizations, but to show
364 how a single PyAges workflow can contrast the information contained in a full tracer record with that
365 available from a restricted sampling window.

366 **4.1. Holten multi-tracer benchmark**

367 We first consider the well-documented Holten well field in the Netherlands (Visser et al., 2013). Production
368 wells with long screens mix groundwater over a broad range of transit times (Henri et al., 2026), making
369 the site a useful external benchmark for groundwater age-distribution modeling. Visser et al. (2013)
370 interpreted seven production wells sampled in April 2010 using four tracer observables – ^3H , corrected
371 tritogenic ^3He , ^{85}Kr , and ^{39}Ar – and reported groundwater fractions in four age classes. These published
372 fractions provide an external reference for the complete PyAges data-preparation, forward-model, and
373 Bayesian-calibration workflow.

374 The final PyAges Holten benchmark uses the same four observables and the input conventions appropriate
375 to the original study: site-specific time-dependent inputs for ^3H and ^{85}Kr , whereas ^{39}Ar is represented by its
376 standard modern atmospheric abundance and radioactive decay. Tritogenic ^3He is derived from ^3H decay
377 rather than prescribed as an independent input. For the ^3H – ^3He pair, a two-year unsaturated-zone transit is
378 imposed before groundwater enters the saturated aquifer. For a parcel of water with saturated-zone age a ,
379 the tritium input is therefore taken two years before saturated recharge and is allowed to decay during this

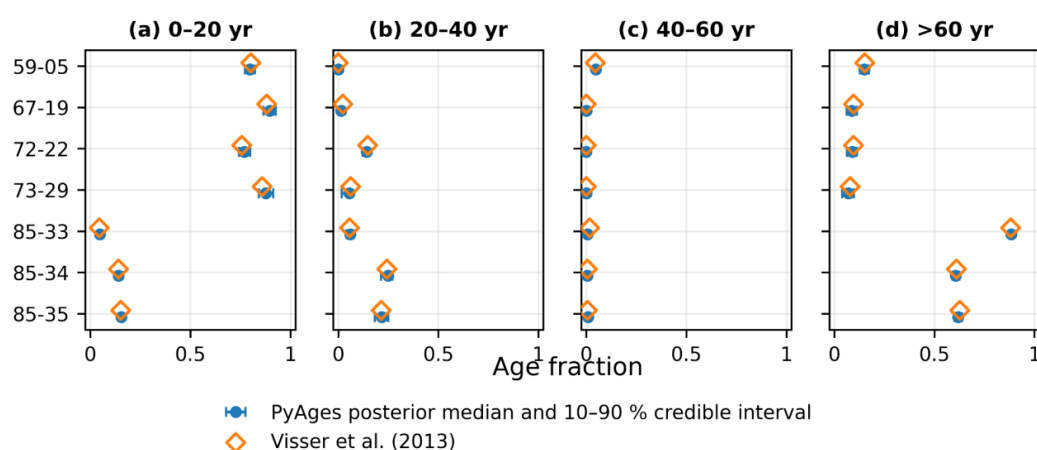


380 delay. The remaining ^3H then decays during saturated transport, while the resulting tritiogenic ^3He is
381 assumed to accumulate in the closed saturated system. The published ^3He values are already corrected for
382 noble-gas, degassing, and radiogenic contributions, so these corrections are not applied again. Tritiogenic
383 ^3He is expressed in tritium-equivalent units (TU), following Visser et al. (2013), who report it as the
384 equivalent tritium concentration. A tritiogenic ^3He uncertainty of 0.5 TU is used for the six wells with
385 reported uncertainties; because the source uncertainty is absent for well 59-05, the same 0.5 TU value (the
386 median of the six reported values) is assigned and treated explicitly as an imputed uncertainty. Visser et al.
387 (2013) also noted possible excess tritiogenic ^3He at well 59-05; the imputed 0.5 TU uncertainty is therefore
388 a benchmark convention rather than a new estimate of measurement uncertainty.

389 The pumped-water age distribution is represented as a four-component mixture, with fractions assigned to
390 the age intervals 0–20, 20–40, 40–60, and >60 years. For each of the first three classes, the tracer response
391 is averaged over 120 uniformly spaced ages within the interval, so each fitted fraction represents the
392 contribution of an age class rather than a single representative age. The >60-year class is treated differently
393 because it has no finite upper bound. Rather than imposing an arbitrary discretization beyond 60 years, it is
394 represented by the fixed old-water end member used in the published benchmark: ^3H and tritiogenic ^3He
395 are evaluated at 310 years, ^{85}Kr is set to zero, and ^{39}Ar to 0.45 fraction modern. The fitted >60-year fraction
396 therefore quantifies mixing with this prescribed old-water tracer signature and should not be interpreted as
397 a fraction of water having exactly 310 years of age. The four fractions are sampled through three stick-
398 breaking variables, which automatically ensures non-negative fractions summing to one and avoids
399 proposing physically invalid age distributions. This parameterization is also well suited to mixtures in which
400 one or more age classes may contribute very little: the bounds $z \in [-8, 8]$ apply to these variables rather
401 than to the fractions themselves, allowing fractions to approach zero or one while keeping the sampled
402 parameters finite and numerically stable. For each well, five independent sampling chains were used and
403 extended when required by the convergence diagnostics. For two wells exhibiting slow mixing near zero-
404 valued fractions, an additional symmetric proposal was used to improve exploration without changing the



405 target posterior; implementation details are retained in the reproducibility archive. All seven final
 406 calibrations satisfy the campaign split-R-hat < 1.01 criterion, with minimum summed autocorrelation-based
 407 ESS values between 313 and 1,373. Fig. 3 compares the resulting PyAges posterior fractions with the values
 408 reported by Visser et al. (2013).



409

410 **Figure 3.** Comparison of the four-bin groundwater age fractions reported by Visser et al. (2013) with the
 411 PyAges posterior distributions for the seven Holten production wells. Panels correspond to the age classes
 412 (a) 0–20, (b) 20–40, (c) 40–60, and (d) >60 years; wells are shown on the shared y axis. Blue circles and
 413 horizontal bars show PyAges posterior medians and central 10–90 % credible intervals, respectively;
 414 orange diamonds show the published fractions. The calibration uses ^3H , corrected tritiogenic ^3He , ^{85}Kr ,
 415 and ^{39}Ar .

416 The final multi-chain PyAges reconstruction closely reproduces the published four-bin distributions. Across
 417 the seven wells and four age classes, the mean absolute difference between the PyAges posterior median
 418 and the Visser et al. (2013) fraction is 0.0055, the median absolute difference is 0.0038, the RMSE is
 419 0.0070, and the maximum absolute difference is 0.018. All 28 reconstructed fractions differ from the
 420 published values by less than 0.02. The largest standardized residual among the four fitted tracer



421 observables is 1.80. Holten therefore provides an external benchmark in which the Bayesian workflow
422 reproduces a published multi-tracer age-distribution interpretation within the same four-bin representation
423 and tracer-input conventions, while also returning posterior uncertainty and explicit convergence
424 diagnostics. Because a uniform prior on the three stick-breaking variables z does not give the same prior
425 distribution to the four age fractions, sensitivity was assessed with a symmetric Dirichlet(1,1,1,1) prior.
426 Although the two prior distributions differ markedly before the observations are used, the largest absolute
427 change in any posterior median age fraction is 0.021 (2.1 % of the total age distribution), and the total
428 fraction redistributed between age classes is at most 0.021 (2.1 %). The global RMS standardized residual
429 changes only from 0.77 to 0.79 (Appendix D).

430 **4.2. Ploemeur long-term CFC demonstration**

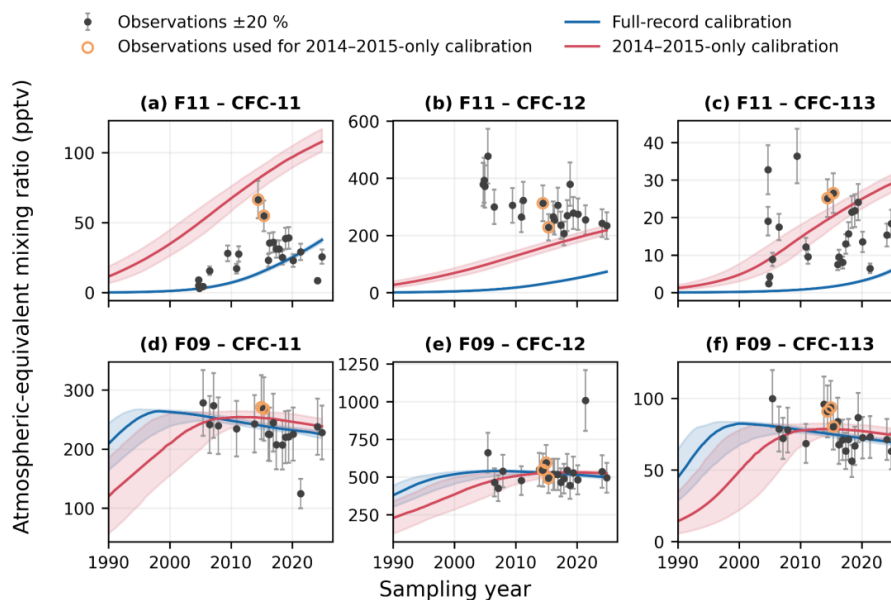
431 The Ploemeur fractured aquifer in Brittany, France, has been extensively characterized hydrogeologically
432 (Le Borgne et al., 2004; Ruelleu et al., 2010) and provides a long-term record of CFC-11, CFC-12, and
433 CFC-113 in groundwater. The CFC observations used here are available in the Ploemeur dataset of the H+
434 national network of hydrogeological observatories (INSU CNRS, 2005); the multisite database architecture
435 underlying H+ is described by de Dreuzay et al. (2006). Here, the purpose is not to revisit the hydrogeological
436 interpretation of the site, but to demonstrate that the same PyAges workflow can be applied consistently
437 either to observations from a restricted sampling window or jointly to a longer tracer record. Two
438 contrasting wells are considered: F09, characterized by relatively young groundwater, and F11,
439 characterized by older and more strongly mixed groundwater. A 20 % relative uncertainty is assigned to all
440 CFC observations as a benchmark choice. This value was selected in consultation with the researchers
441 responsible for the dataset to represent sampling-related variability together with measurement uncertainty,
442 rather than analytical uncertainty alone.

443 For each well, two inference problems are compared using exactly the same shifted-exponential LPM,
444 tracer definitions, likelihood, and uncertainty model: one calibration uses all available observations in the
445 monitoring record, whereas the other uses only the observations from the 2014–2015 sampling window.



446 The final calculations use the CDF–partial-first-moment forward operator described in Sect. 2.2.4. Each
 447 calibration is sampled with five independent Metropolis–Hastings sampling chains using a pilot-estimated
 448 fixed correlated proposal. For the Ploemur demonstration, the inferred transit-time distribution is
 449 illustrated by its median, $t_{50} = F^{-1}(0.5) = t_0 + \mu \ln(2)$.

450 This quantity is evaluated for every posterior sample. All four calibrations satisfy the convergence criteria,
 451 with a maximum split-R-hat of 1.004 and a minimum multi-chain ESS of 1,741 across μ , t_0 , and t_{50} . Fig.
 452 4 shows the observed CFC time series and the posterior trajectories obtained from the two calibration
 453 windows.



454

455 **Figure 4.** Ploemur CFC time series for F11 (top row) and F09 (bottom row) interpreted with the shifted-
 456 exponential LPM. Black points and error bars show atmospheric-equivalent mixing-ratio observations with
 457 the adopted $\pm 20\%$ benchmark uncertainty; orange circles identify the observations used in the 2014–2015-
 458 only calibration. Blue lines and shaded envelopes show the median modeled response and central 10–90
 459 % posterior interval for calibration to the full record, whereas red lines and envelopes show the
 460 corresponding results for the 2014–2015-only calibration. Modeled trajectories are displayed over 1990–



461 *2025 to facilitate comparison of their temporal behavior. The envelopes propagate posterior parameter*
 462 *uncertainty only and do not include an additional realization of the observation-error model.*

463 At F09, the 2014–2015-only observations admit substantially older solutions, with a posterior median t_{50}
 464 of 13.3 years and a 10–90 % interval of 6.2–20.3 years. When all sampling dates are used jointly, the
 465 declining temporal evolution of the CFC record provides an additional constraint and shifts the posterior
 466 toward younger groundwater, with a median t_{50} of 4.2 years [1.5–8.1 years]. The example therefore shows
 467 directly how repeated observations can resolve an age ambiguity that remains when only a limited sampling
 468 window is available.

469 F11 illustrates a different behavior. The 2014–2015-only calibration yields a median t_{50} of 57.4 years
 470 [52.6–61.5 years], whereas calibration to the full record shifts the posterior to 85.4 years [84.7–86.0 years].
 471 The longer record also reveals a clear tracer-specific contrast: the temporal evolution of CFC-11 is
 472 consistent with the full-record interpretation, whereas CFC-12 and CFC-113 are not reproduced
 473 consistently within the adopted stationary tracer–LPM–error framework. Thus, the additional sampling
 474 dates do not simply narrow the posterior; they also reveal which tracers support a consistent transit-time
 475 interpretation and which exhibit persistent discrepancies. Together, the two wells demonstrate that PyAges
 476 responds coherently to the amount and consistency of temporal information provided to the inversion:
 477 repeated observations resolve an ambiguity at F09, while at F11 the longer record distinguishes a temporally
 478 consistent CFC-11 signal from persistent discrepancies in CFC-12 and CFC-113. The persistent
 479 discrepancies at F11 show that a single stationary shifted-exponential TTD, together with the adopted tracer
 480 inputs and error model, cannot reproduce all three CFC time series simultaneously. They therefore reveal
 481 a limitation of at least one assumption in the present interpretation. The data alone do not determine whether
 482 the mismatch results from temporal variation in the TTD, tracer-specific processes affecting individual
 483 CFCs, inaccuracies in the reconstructed input histories, or unrepresented sampling or measurement errors.
 484 A more detailed hydrogeological interpretation of the Ploemeur CFC time series, including the temporal



behavior of the individual tracers, the comparison of alternative transit-time distributions, and the inferred groundwater-age dynamics, is presented in de Dreuz et al. (in review).

5. Discussion

Environmental-tracer inference is most useful when the analysis can be used to test how scientific conclusions respond to alternative assumptions, to determine what the observations actually constrain, and to reuse the resulting components in broader workflows. The discussion therefore focuses on controlled hypothesis testing, uncertainty and identifiability, and workflow integration.

5.1. Exploring tracer information through controlled hypothesis testing

Environmental-tracer interpretation is rarely resolved by a single inversion. It progresses by testing whether conclusions persist when assumptions about tracer representations, observation sets, transit-time distributions, priors, or error models are changed. Because PyAges separates these components, one hypothesis can be modified while the rest of the workflow is held fixed. Differences between analyses can therefore be related more directly to the assumption being tested rather than to changes in implementation.

The Ploemur application illustrates this experimental use of the framework. Long-term, high-resolution multi-tracer records are increasingly used to characterize groundwater recharge and flow regimes in complex aquifers (Kang et al., 2026). The same tracer definitions, shifted-exponential LPM, likelihood, and calibration settings are applied to a restricted sampling window and to the full record, thereby isolating the information added by repeated observations. More generally, information content depends not only on the number of tracers or observations, but also on the independent temporal sensitivities they contribute and on the quantity of interest. Tracers with similar input histories may be partly redundant, and detailed TTD shape may remain non-unique even when moments, age fractions, or predictive quantities are constrained (Green et al., 2014; McCallum et al., 2014; Marçais et al., 2015). Multiple dates can likewise expose structural limitations hidden by a satisfactory single-date fit (Massoudieh et al., 2014a).



PyAges does not remove these intrinsic ambiguities; it provides a practical way to investigate them systematically by comparing alternative hypotheses within the same workflow. In the present Ploemur analysis, multiple dates constrain one stationary TTD. When sufficiently informative time series are available, more flexible transient formulations, including StorAge Selection (SAS) functions, can instead describe how the age composition of discharge changes through time rather than imposing a stationary TTD (Botter et al., 2011; Rinaldo et al., 2015; Harman and Xu Fei, 2024; Türk et al., 2026). Such formulations address a different inference problem and require observations rich enough to constrain the added temporal degrees of freedom (Massoudieh, 2013; Jurgens and Levy, 2026).

5.2. What uncertainty analysis adds to tracer interpretation

Uncertainty analysis changes the question from “Which parameter set fits best?” to “Which conclusions remain supported when all parameter combinations compatible with the observations are considered?” A joint posterior shows which parameters or parameter combinations are constrained by the data and which remain difficult to distinguish because they produce nearly the same modeled tracer values. In the shifted-exponential benchmark, μ and t_0 can compensate strongly, whereas their sum is much better constrained. The Bayesian formulation used here is one practical way of revealing such trade-offs; the objective is to determine what the observations actually constrain and what remains ambiguous (Massoudieh et al., 2012; McCallum et al., 2014).

Inference quality should be assessed separately from fit quality and posterior concentration. A low residual can coexist with weak identifiability, while a narrow posterior only indicates precision within the assumptions of the selected model. Previous studies similarly show that excellent tracer fits need not recover the relevant characteristics of a TTD or provide the best prediction of the quantity of interest (Green et al., 2014; Marçais et al., 2015; Massoudieh et al., 2014a). Information-rich observations can also reduce estimated uncertainty without eliminating predictive bias when the model cannot represent the processes they contain (Knowling et al., 2020).



Bayesian analysis is useful in this context because assumptions about priors and observational errors are explicit and can be tested. At Holten, substantially different priors produce only minor changes in the inferred age fractions, supporting the robustness of the reconstruction for the tested assumptions. This does not remove uncertainty about the model structure itself. A posterior describes uncertainty assuming that the chosen TTD family, tracer representation, and error model are appropriate; a different TTD or tracer-process model may lead to a different interpretation. Such structural uncertainty can be assessed by explicitly fitting and comparing plausible alternatives (Rojas et al., 2008). Where independent hydrogeological information is available, it can reduce parameter equifinality through physically justified parameter bounds or informative prior distributions on relevant LPM parameters (e.g., μ and t_0). Such constraints leave the tracer input histories, transport transformations, and convolution operator unchanged, but their influence on the posterior should be reported explicitly and assessed independently of the tracer likelihood (Corcho Alvarado et al., 2007; Leray et al., 2012; Thiros et al., 2023). The contribution is therefore not Bayesian inference in itself, but the ability to expose parameter trade-offs, test assumptions, and distinguish robust conclusions from those that remain model dependent.

5.3. Integration into broader tracer-analysis workflows

PyAges is designed so that tracer preparation, transit-time distributions, forward simulation, and inference can be reused as components of larger scientific workflows. Their explicit inputs and outputs allow analyses to be applied in batch across sites or sampling dates, connected to upstream data-processing steps, or followed by other modeling and prediction tasks. Retaining observations in tracer space is important in this respect: conversion to apparent groundwater ages is itself model dependent and may discard information contained in tracer-specific input histories and uncertainties (Suckow, 2014; Turnadge and Smerdon, 2014; McCallum et al., 2015). Direct use of tracer observations can also provide stronger constraints when incorporated into groundwater-model calibration (Thiros et al., 2021).



555 Tracer-based LPM inference and distributed groundwater modeling are therefore complementary rather
556 than competing approaches. Environmental tracers can constrain or test process-based models (Sanford,
557 2011), while physically based flow models can provide information on plausible transit-time structures and
558 parameter ranges (Leray et al., 2019; Gauvain et al., 2021). Once inferred, a TTD can also be reused as a
559 temporal transfer function with other recharge or solute-input histories, provided that the corresponding
560 transport and reaction assumptions are specified. Tracer-constrained LPMs have, for example, been used
561 to assess well vulnerability and to reproduce or project nitrate trends (Eberts et al., 2012; Jurgens et al.,
562 2016; Alikhani et al., 2016). These applications reinforce the broader point that the usefulness of an inferred
563 TTD ultimately depends on the prediction or interpretation for which it is used.

564 The same framework can also support questions about what should be measured next. Candidate tracers,
565 additional sampling dates, or improved measurement precision can in principle be compared according to
566 how much they reduce uncertainty in a selected parameter or prediction (Green et al., 2014; Zell et al.,
567 2018). Formal data-worth or experimental-design tools are not part of PyAges v1.0, but the forward and
568 inference components provide the basis for such extensions. Reuse in broader workflows also requires
569 reproducibility: versioned code, configurations, tracer inputs, random seeds, and post-processing steps
570 connect the scientific interpretation to an identifiable computational experiment (Hutton et al., 2016).

571 **6. Conclusions and perspectives**

572 PyAges v1.0 provides an open, scriptable framework for inferring groundwater transit-time distributions
573 directly from environmental-tracer observations. Its contribution is the integration of modular tracer and
574 LPM definitions, a numerically characterized convolution operator, interchangeable calibration
575 components, and reproducible case-level workflows. The CDF–partial-first-moment formulation resolves
576 variation in the tracer response rather than requiring the potentially narrow LPM density itself to be densely
577 sampled, while discrete components are treated directly.



578 The benchmarks characterize discrepancies relative to independent reference calculations, reproduce a
579 published multi-tracer interpretation, and consistently exploit differences in the temporal information
580 supplied by field observations. Additional information may either reduce ambiguity or reveal
581 inconsistencies that are not apparent from a restricted dataset. More generally, fit quality alone does not
582 determine inference quality: different parameter combinations can reproduce the observations similarly,
583 while some derived quantities remain much better constrained than the individual parameters. PyAges
584 makes these distinctions accessible through joint uncertainty analysis and tracer-level diagnostics, without
585 making Bayesian inference itself the scientific objective.

586 Because each scientific component can be exchanged independently, PyAges can support controlled tests
587 of alternative tracer assumptions, transit-time distributions, priors, sampling designs, and inverse methods,
588 and can serve as a building block in broader hydrogeological calibration or water-quality workflows. The
589 present release is limited to stationary LPMs and the tracer transformations explicitly implemented or
590 supplied by the user; transient and more complex reactive formulations remain extensions rather than
591 implicit capabilities. The public v1.0 tag, validated reproducibility archive, machine-readable results, and
592 case-level check, post-processing, and rerun commands make both the software and the analyses reported
593 here directly inspectable and reusable.



594 **Code and data availability**

595 PyAges is openly developed at <https://github.com/dreuzy/PyAges> and distributed under the CeCILL 2.1
596 license. The public annotated tag 1.0 (<https://github.com/dreuzy/PyAges/tree/1.0>), which points to commit
597 bc97d05f608f983b08df6d870cca72caa088d5e2, identifies the exact source-code version used for this
598 study. The CI workflow associated with this tag completed successfully, including tests on Python 3.12,
599 3.13, and 3.14, dependency checks, documentation and packaging checks, and TracerLPM validation. A
600 validated reproducibility archive has been assembled for the manuscript and includes the redistributable
601 benchmark input data and tracer histories, configuration files, run manifests and random seeds, scripts used
602 to generate the reported figures and tables, machine-readable results, and environment and provenance
603 metadata. The tagged source and reproducibility package are archived together on Zenodo at
604 <https://doi.org/10.5281/zenodo.22150863> (de Dreuzy, 2026). The exact Ploemur observations used in this
605 study are available through the H+ data record (INSU CNRS, 2005), and their use and redistribution for
606 this study are authorized by the data providers. GitHub remains the development repository, while the
607 Zenodo record provides the immutable software-and-data archive corresponding to this study. The mapping
608 between manuscript results and the common case-level reproduction interface is summarized in Appendix
609 E.



610 **Appendix A. Numerical implementation and verification of the forward** 611 **operator**

612 **A1. Adaptive tracer-response discretization and interval integration**

613 This section provides the numerical details of the continuous convolution scheme introduced in Sect. 2.2.4.
 614 After the tracer response has been defined for a given sampling date, the main numerical question is how
 615 finely it must be discretized over water age before the interval contributions are evaluated.

616 For the continuous pathway, the remaining numerical choice is the age discretization of the tracer response
 617 $K(\tau)$. The age grid is determined by the tracer and the sampling date, independently of the LPM parameters.
 618 For a tracer defined by a time series, all input-history dates within the convolution window are converted
 619 to water ages and used as initial grid nodes, together with zero age and the oldest age represented by the
 620 available input record.

621 For conservative tracers, the time-series input is linearly interpolated between successive chronicle nodes
 622 (Sect. 2.2.2), so $K(\tau)$ is affine within each interval and Eq. (8) is exact there. No additional subdivision
 623 below the temporal resolution of the input chronicle is therefore required. For example, the CFC and SF₆
 624 histories used in Sect. 3.2 are tabulated every 0.5 years; for the 2010 sampling date, their chronicle nodes
 625 directly define 140 age intervals between 0 and 70 years.

626 When the complete tracer response $K(\tau)$ is not sufficiently resolved by these initial nodes—for example,
 627 because radioactive decay makes the response nonlinear between chronicle dates—PyAges refines the grid
 628 only where the tracer response changes too much. For each interval $[a, b]$, K is evaluated at the left endpoint,
 629 the midpoint m , and the right endpoint. The variation of K within the interval is defined as

$$630 \quad \Delta K = \max[K(a), K(m), K(b)] - \min[K(a), K(m), K(b)]$$

631 Thus, ΔK is simply the range of tracer-response values within the interval. It is compared with a tolerance
 632 T ,

$$633 \quad T = f_a \max(K_{\text{global}}, \varepsilon) + f_r K_{\text{local}}$$



634 This tolerance defines the maximum variation in K that is accepted within an interval before the grid is
 635 considered too coarse. The first term provides a small global tolerance based on the overall amplitude of
 636 the tracer response and prevents excessive refinement where K is close to zero. The second term scales the
 637 tolerance with the response amplitude in the current interval. Here, K_{global} is the largest absolute response
 638 over the whole grid, K_{local} is the largest absolute response in the current interval, and ε is a small positive
 639 numerical floor that prevents the tolerance from becoming zero. The interval is divided into two equal parts
 640 when

$$641 \quad \Delta K > T$$

642 Otherwise, it is retained. With the default values $f_a = 5 \times 10^{-4}$ and $f_r = 2 \times 10^{-2}$, an interval is therefore
 643 retained only if the sampled tracer response varies by no more than 0.05 % of the global response scale plus
 644 2 % of the local response scale. If this limit is exceeded, the interval is subdivided and the same test is
 645 applied again.

646 Once an interval has been retained, PyAges determines how its contribution to the convolution should be
 647 calculated. The previous refinement criterion limits the total variation of K within the interval, but it does
 648 not ensure that K varies linearly between the two endpoints. This distinction matters because Eq. (8) is
 649 exact when K is linear within the interval. PyAges therefore compares the actual midpoint response with
 650 the value expected from a straight line between the two endpoints. The deviation from linearity is

$$651 \quad D = \left| K(m) - \frac{K(a) + K(b)}{2} \right|$$

652 The response is considered sufficiently linear when

$$653 \quad D \leq 0.1 T$$

654 In that case, Eq. (8) is used. If this condition is not satisfied, the straight-line approximation is no longer
 655 considered reliable. Rather than extrapolating a linear variation that is contradicted by the evaluated
 656 midpoint, PyAges uses the midpoint response as a representative value for the interval. The contribution is



then approximated as $K(m) \times w_i$, where w_i is the exact LPM probability mass within the interval. This provides a more robust fallback for locally curved tracer responses, while the preceding grid-refinement criterion ensures that the total variation of K within the interval remains small. The tracer grid is constructed once for each tracer–sampling-date pair and then reused for all subsequent LPM evaluations during calibration.

A2. Independent verification and numerical qualification

The resulting forward operator was then verified against an independently implemented reference calculation in a 270-case campaign, without any tracer-parameter inference. The campaign contains 45 piston flow model (PFM), 45 exponential mixing model (EMM), 90 exponential-piston flow model (EPM), and 90 dispersion model (DM) cases. Reference concentrations are generated without calling PyAges LPM classes: PFM is evaluated directly, whereas the continuous models use analytical LPM density functions and composite eight-point Gauss–Legendre quadrature, with the integration split at every monthly interpolation knot of the synthetic input histories. The complete reference matrix and per-case results are retained with provenance metadata in the reproducibility package. Numerical agreement is assessed with two complementary measures: the absolute concentration difference and the symmetric relative difference in Eq. (A1). The symmetric form treats PyAges and the independent reference calculation equally, without selecting either result as the denominator.

$$d_{\text{abs}} = |C_{\text{PyAges}} - C_{\text{ref}}|, \quad d_{\text{sym}} = \frac{2|C_{\text{PyAges}} - C_{\text{ref}}|}{\max(|C_{\text{PyAges}}| + |C_{\text{ref}}|, 10^{-12})}. \quad (\text{A1})$$

Relative differences become unstable near zero. We therefore define, for each tracer, the reference scale C_{scale} as the maximum concentration in its input history:

$$C_{\text{scale}} = \max_t |C_{\text{input}}(t)|. \quad (\text{A2})$$

When at least one calculated concentration reaches 0.1 % of C_{scale} , agreement is assessed with the symmetric relative criterion:

$$\max(|C_{\text{PyAges}}|, |C_{\text{ref}}|) \geq 10^{-3} C_{\text{scale}}, \quad d_{\text{sym}} \leq 5 \times 10^{-4} \text{ (0.05 \%)}. \quad (\text{A3})$$

Otherwise, the near-zero case is assessed with the absolute criterion:

$$d_{\text{abs}} \leq 2 \times 10^{-5} C_{\text{scale}}. \quad (\text{A4})$$



Thus, relative errors are used for appreciable concentrations and absolute errors near zero. Because the synthetic input histories are non-negative and the tested tracer response is conservative, both calculated concentrations must also be finite and lie between 0 and C_{scale} , allowing only a $10^{-12} C_{\text{scale}}$ tolerance for floating-point round-off. All 270 cases must satisfy these conditions individually.

Table A1. Observed forward discrepancies by LPM family in the 270-case campaign. Absolute RMSE and maximum discrepancies are reported in the arbitrary concentration units used for the synthetic inputs; symmetric relative discrepancies are dimensionless.

LPM	Cases	RMSE (a.u.)	Max. abs. discrepancy (a.u.)
PFM	45	2.70×10^{-13}	5.12×10^{-13}
EMM	45	7.25×10^{-4}	2.92×10^{-3}
EPM	90	9.90×10^{-4}	6.12×10^{-3}
DM	90	9.71×10^{-4}	6.11×10^{-3}

Of the 270 verification cases, 237 are evaluated with the relative criterion and 33 with the near-zero absolute criterion. The largest symmetric relative difference among the 237 relative-criterion cases is 2.15×10^{-4} , below the 5×10^{-4} limit; among the 33 near-zero cases, the largest absolute difference normalized by C_{scale} is 1.09×10^{-5} , below the 2×10^{-5} limit. All 270 cases therefore pass the qualification contract. Table A1 additionally reports family-level RMSE and maximum absolute discrepancy: RMSE ranges from 2.70×10^{-13} to 9.90×10^{-4} , and the maximum absolute discrepancy from 5.12×10^{-13} to 6.12×10^{-3} .

Appendix B. Common metrics and parameter mappings for the PyAges–TracerLPM benchmark

Because PyAges and TracerLPM use different native objective functions, the modeled concentrations from both codes were reevaluated with the same two residual measures before the final solutions were compared:



$$L_1^{\text{rel}} = \sum_{i=1}^m \left| \frac{C_{\text{mod},i} - C_{\text{obs},i}}{C_{\text{obs},i}} \right|. \quad (\text{B1})$$

$$L_2^{\text{rel}} = \sum_{i=1}^m \left(\frac{C_{\text{mod},i} - C_{\text{obs},i}}{C_{\text{obs},i}} \right)^2. \quad (\text{B2})$$

695 These are sums over the four tracer observations; $L_{2,\text{rel}}$ is not square-rooted. TracerLPM parameters were
 696 also converted to the parameterization used throughout this manuscript before parameter-recovery errors
 697 were calculated. For EPM, TracerLPM uses the piston-to-exponential ratio r , whereas the manuscript uses
 698 $\eta = 1 + r$ and the shifted-exponential timescale μ and shift t_0 :

$$\eta = 1 + r, \quad \mu = \frac{\tau}{\eta}, \quad t_0 = \tau \left(1 - \frac{1}{\eta} \right). \quad (\text{B3})$$

699 For DM, the corresponding inverse-Gaussian parameters are obtained from:

$$\mu = \tau, \quad \sigma = \tau \sqrt{2 \text{DP}}. \quad (\text{B4})$$

700 The common search bounds used for the benchmark are given below. Table B1 reports parameter recovery
 701 for the 0 % added-noise controls, and Table B2 summarizes the 480 noisy paired inversions using the
 702 common residual measures and parameter-recovery criteria.

$$\tau \in [0.1, 200] \text{ yr}, \quad \eta \in [1.01, 11] \quad (r \in [0.01, 10]). \quad (\text{B5})$$

$$\tau \in [0.1, 200] \text{ yr}, \quad \text{DP} \in [0.001, 3]. \quad (\text{B6})$$

703 **Table B1.** Parameter-recovery errors for the 0 % added-noise controls. “Numerical precision” denotes
 704 discrepancies negligible at the reported precision.

Model	Parameter	Target	PyAges relative error	TracerLPM relative error
EMM	MTT	20	Numerical precision	0.4 %
EPM	MTT	20	Numerical precision	1.3 %
EPM	η	3	Numerical precision	1.2 %
DM	MTT	20	Numerical precision	1.6 %
DM	DP	0.2	Numerical precision	4.0 %



Table B2. Aggregate comparison of the 480 noisy paired inversions using common residual measures and parameter-recovery criteria.

Comparison	PyAges	TracerLPM
Median $L_{1,rel}$	0.179	0.181
Median $L_{2,rel}$	0.0135	0.0164
Lower $L_{2,rel}$	461 / 480	19 / 480
Lower $L_{1,rel}$	111 / 480	369 / 480
Closer to generating MTT	208 / 480	272 / 480
Closer to generating second parameter	204 / 480	233 / 480 (43 ties)

The contrasting rankings under $L_{1,rel}$, $L_{2,rel}$, and parameter-recovery criteria show that neither software can be ranked independently of the chosen metric. A closer fit to a particular noisy realization does not necessarily imply parameters closer to the values used to generate the observations. All 480 paired inversions completed successfully. Complete case-level results, native objective values, parameter bounds, software-version metadata, random seeds, and input files are provided in the reproducibility archive.

Appendix C. Detailed posterior results for the shifted-exponential benchmark

This appendix provides the complete posterior summaries supporting the shifted-exponential benchmark presented in Sect. 3.2. The main results and interpretation are reported in the main text and Table 4; the tables below provide the corresponding parameter-level posterior summaries and multi-chain diagnostics for all 19 synthetic configurations.

The benchmark uses exact synthetic tracer values with an 8 % relative observational standard deviation in the likelihood. Five converged sampling chains are combined after burn-in. MTT is calculated for every paired posterior sample before its summaries are computed, as defined in Sect. 3.2; its quantiles are therefore not obtained by adding marginal quantiles. Relative posterior standard deviations and relative 10–90 % half-widths summarize different aspects of posterior spread and are reported separately. Table C1 reports the marginal posterior summaries of μ and t_0 , while Table C2 reports the corresponding MTT summaries and multi-chain diagnostics.



724 **Table C1.** Posterior summaries of the two shifted-exponential parameters. Values are median [10–90 %
 725 credible interval], in years, rounded to 0.1 yr.

Case	Target μ	Posterior μ	Target t_0	Posterior t_0
1	1	1.6 [0.4–4.1]	1	1.4 [0.2–3.6]
2	1	3.6 [0.8–7.5]	10	7.2 [3.2–10.6]
3	1	3.5 [0.8–7.3]	20	17.5 [14.0–20.2]
4	1	3.0 [0.7–6.7]	30	28.3 [25.7–30.3]
5	1	9.2 [1.8–19.0]	40	34.8 [31.2–39.3]
6	10	8.1 [3.6–11.2]	1	2.9 [0.6–7.0]
7	10	10.4 [6.4–14.6]	10	9.7 [6.1–13.6]
8	10	11.1 [5.2–17.6]	20	19.4 [16.0–23.0]
9	10	10.5 [5.2–16.3]	30	29.8 [27.4–32.7]
10	10	17.9 [4.9–41.0]	40	37.2 [32.3–42.4]
11	20	18.6 [14.9–21.7]	1	2.3 [0.4–5.5]
12	20	20.7 [15.3–26.8]	10	9.6 [6.0–13.2]
13	20	21.7 [15.1–30.2]	20	19.4 [16.6–22.0]
14	20	20.6 [11.3–30.4]	30	29.8 [27.3–33.2]
15	30	28.1 [23.6–31.9]	1	2.3 [0.4–5.4]
16	30	31.1 [23.8–38.9]	10	9.6 [6.1–13.0]
17	30	32.2 [23.7–43.6]	20	19.4 [16.6–21.9]
18	40	37.7 [32.1–42.4]	1	2.3 [0.5–5.4]
19	40	41.2 [32.3–51.1]	10	9.7 [6.1–13.0]



Table C2. Posterior summaries of MTT and multi-chain diagnostics. MTT intervals are calculated directly from the joint posterior. MTT values are in years and are rounded to 0.1 yr; split-R-hat is reported to three decimals.

Case	Target MTT	MTT median [10–90 %]	MTT [2.5–97.5 %]	Max split- $\hat{R}$	Min ESS
1	2	3.5 [1.6–5.8]	[0.9–7.3]	1.002	2,357
2	11	11.0 [9.0–13.0]	[7.9–14.1]	1.002	4,421
3	21	21.1 [20.2–22.0]	[19.7–22.6]	1.001	3,978
4	31	31.4 [30.7–32.6]	[30.4–33.5]	1.003	3,807
5	41	44.0 [41.1–50.1]	[40.8–53.9]	1.004	1,571
6	11	11.2 [8.9–13.4]	[7.7–14.6]	1.001	3,692
7	20	20.2 [18.7–21.7]	[18.0–22.6]	1.001	4,733
8	30	30.5 [28.1–33.8]	[27.3–35.9]	1.000	3,321
9	40	40.3 [37.9–43.7]	[37.1–46.1]	1.002	2,513
10	50	55.1 [47.3–73.3]	[46.3–84.3]	1.010	756
11	21	21.0 [19.0–23.2]	[18.0–24.5]	1.001	3,258
12	30	30.3 [28.0–33.4]	[26.9–35.3]	1.001	4,531
13	40	41.1 [37.0–46.8]	[35.4–50.3]	1.001	4,284
14	50	50.4 [44.4–57.7]	[41.7–62.2]	1.002	1,116
15	31	30.4 [28.1–33.2]	[26.9–34.9]	1.001	4,305
16	40	40.6 [36.6–45.4]	[34.7–48.2]	1.002	4,775
17	50	51.7 [45.6–60.3]	[43.2–66.2]	1.000	4,439
18	41	40.0 [36.7–43.6]	[35.1–45.5]	1.002	3,799
19	50	50.7 [45.1–57.5]	[42.5–61.6]	1.001	4,784

Across the 19 configurations, the median posterior standard deviation relative to the target is 32.1 % for μ , 26.6 % for t_0 , and 8.6 % for MTT. The corresponding median relative half-widths of the central 10–90 % intervals are 40.8 %, 34.2 %, and 11.0 %. Across the 19 configurations, the median pooled posterior correlation between μ and t_0 is -0.95 . The shortest case, $(\mu, t_0) = (1, 1)$ years, is a clear exception in relative resolution; the case (10, 40) years has the lowest ESS and the broadest MTT posterior among the remaining configurations.



735 **Appendix D. Prior sensitivity of the Holten age-fraction benchmark**

736 **D1. Prior distributions**

737 This appendix tests whether the Holten age-fraction results reported in Sect. 4.1 depend materially on the
 738 prior implied by the stick-breaking parameterization. The baseline analysis assigns a uniform prior to each
 739 of the three stick-breaking variables $z \in [-8, 8]$. Because the four age fractions are constructed sequentially
 740 from these variables, this choice does not give the same prior distribution to each age class. As an
 741 alternative, we use a symmetric Dirichlet(1,1,1,1) prior defined directly on the four age fractions; under this
 742 prior, the four fractions have the same prior distribution, are non-negative, and sum to one. The
 743 implementation of the change of variables required for sampling was verified numerically. Because
 744 sampling is restricted to $z \in [-8, 8]$, the Dirichlet comparison is restricted to this represented range; 99.76
 745 % of independent Dirichlet draws fall within it. Table D1 compares the resulting prior-only distributions.

746 **Table D1.** Prior-only distributions of the four age fractions. The Dirichlet prior is restricted to the range
 747 represented by $z \in [-8, 8]$.

Fraction	Uniform- z mean	Uniform- z median [10–90 %]	Dirichlet mean	Dirichlet median [10–90 %]
0–20 yr	0.50	0.50 [0.00–1.00]	0.25	0.21 [0.04–0.54]
20–40 yr	0.25	0.02 [0.00–0.94]	0.25	0.21 [0.04–0.54]
40–60 yr	0.12	0.00 [0.00–0.57]	0.25	0.21 [0.04–0.54]
>60 yr	0.12	0.00 [0.00–0.57]	0.25	0.21 [0.04–0.54]

748 **D2. Effect of the prior on inferred age distributions**

749 The alternative prior was used in seven independent five-chain calibrations with the same observations,
 750 uncertainties, tracer transformations, four age-class end members, likelihood, and sampling-chain lengths
 751 as the baseline analysis. All calculations met the convergence criterion used in the main benchmark; the
 752 maximum split-R-hat was 1.0087 and the minimum ESS was 910. To assess the practical effect of changing
 753 the prior, posterior median age fractions were compared well by well. For each well, Table D2 reports the
 754 largest change in any one age class and the total fraction redistributed among the four age classes. The latter
 755 is defined as half the sum of the absolute changes across the four fractions, and can therefore be read as the

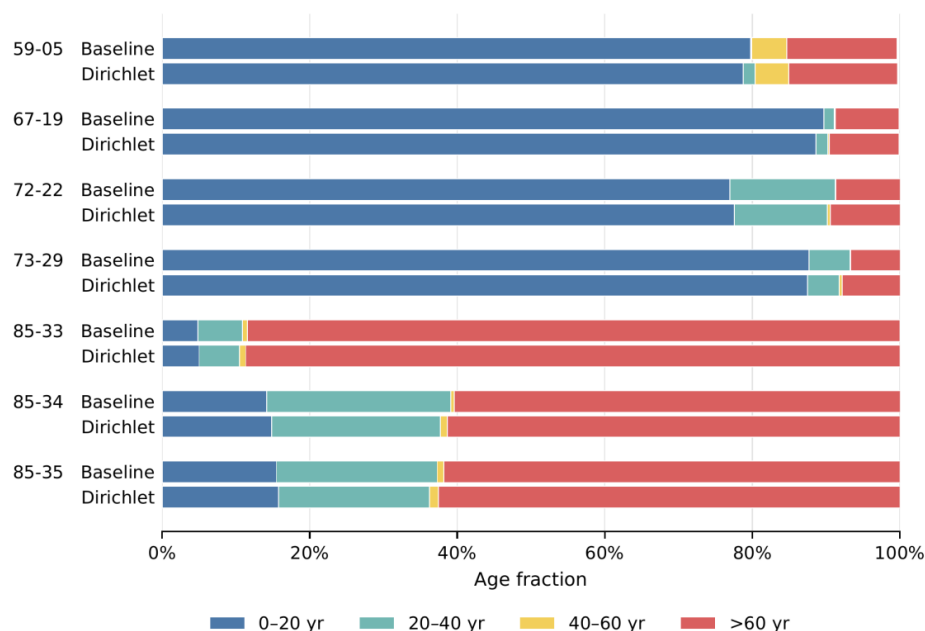


756 fraction of the inferred age distribution that would have to move between classes to transform one median
 757 distribution into the other.

758 **Table D2.** *Effect of replacing the baseline uniform-z prior with the Dirichlet(1,1,1,1) prior. Changes are*
 759 *reported as percentages of the total age distribution; RMS residuals are standardized by the observational*
 760 *standard deviations.*

Well	Max. absolute age-fraction change (%)	Age class	Total redistributed (%)	RMS baseline	RMS Dirichlet
59-05	1.5	20–40 yr	1.5	0.78	0.92
67-19	1.1	0–20 yr	1.1	1.03	1.01
72-22	1.7	20–40 yr	1.7	0.39	0.59
73-29	1.3	20–40 yr	1.4	0.67	0.73
85-33	0.5	20–40 yr	0.5	0.21	0.03
85-34	2.1	20–40 yr	2.1	1.00	0.93
85-35	1.4	20–40 yr	1.4	0.92	0.88

761 The effect of the prior is small for all seven wells. The largest absolute change in any single age class is
 762 0.021 (2.1 % of the total age distribution), for the 20–40 yr fraction in well 85-34. The total fraction
 763 redistributed between age classes is also at most 0.021 (2.1 %) in well 85-34, with a median of 0.014 (1.4
 764 %) across the seven wells. The global RMS standardized residual is 0.77 for the baseline prior and 0.79 for
 765 the Dirichlet prior, indicating essentially unchanged fit quality. Fig. C1 compares the two posterior median
 766 age distributions directly for each well. Thus, despite markedly different prior-only distributions, the
 767 inferred Holten age distributions are only weakly affected by this change of prior.



768

769 **Figure D1.** Sensitivity of the Holten posterior median age distributions to prior choice. For each well,
 770 paired stacked bars compare the four age-class fractions obtained with the baseline uniform-z prior and
 771 with the symmetric Dirichlet(1,1,1,1) prior. The close agreement between the two bars for each well
 772 illustrates the small effect of the prior change.

773 Appendix E. Reproducibility of the manuscript analyses

774 The calculations reported in this manuscript are organized as six reproducibility cases, each corresponding
 775 to a specific benchmark, field application, or sensitivity analysis. For each case, the archived material
 776 identifies the input data and configuration, the scripts used for the calculations and post-processing, the
 777 retained numerical outputs, and the associated provenance information. Table E1 provides the
 778 correspondence between manuscript results and case identifiers.



Table E1. Mapping between manuscript results and reproducibility cases.

Manuscript result	Case identifier	Main archived output
Sect. 3 / Appendix A	s3_forward_verification	Forward-verification results
Sect. 3.1 / Appendix B	s3_1_tracerlpm	Table 3 and paired PyAges–TracerLPM benchmark results
Sect. 3.2 / Appendix C	s3_2_shifted_exponential	Figure 2, Table 4, posterior samples and diagnostics
Sect. 4.1	s4_1_holten	Figure 3, posterior age fractions and diagnostics
Sect. 4.2	s4_2_ploemur	Figure 4, posterior time series and diagnostics
Appendix D	holten_prior_dirichlet1	Prior-sensitivity results, figure and tables

The Zenodo archive associated with this manuscript contains the frozen PyAges v1.0 source code, the input data and configurations used for these cases, retained numerical outputs including the MCMC samples required to reproduce posterior summaries, publication figures and tables, and machine-readable provenance and integrity information. These retained outputs allow the reported results to be inspected and post-processed without rerunning the computationally intensive calibration campaigns.

For users who wish to interact with the archived source, the common command-line interface is summarized in Table E2.

Table E2. Common command-line interface for the archived reproducibility cases. All commands are invoked as `python -m scripts.article.run_case <command>`. A complete rerun of `s3_1_tracerlpm` additionally requires the external TracerLPM/Excel calculation.

Command	Purpose
<code>list</code>	Lists manuscript cases.
<code>check <case_id></code>	Verifies archived files, checksums, and provenance; no calculation.
<code>postprocess <case_id></code>	Regenerates summaries and figures from archived outputs.
<code>run <case_id></code>	Runs a new calculation in a separate directory; archived results are unchanged.

The archived calculations correspond to the public PyAges 1.0 tag, which identifies Git revision `bc97d05f608f983b08df6d870cca72caa088d5e2`. The corresponding Zenodo record provides the immutable software-and-data snapshot used for this manuscript, with file inventories and SHA-256 checksums allowing its integrity to be verified: <https://doi.org/10.5281/zenodo.22150863>.



794 **Author contribution**

795 J.-R. de Dreuzy developed and implemented PyAges, performed the analyses, prepared the figures and
796 tables, and drafted the manuscript. J.-R. de Dreuzy, S. Leray, and J. Marçais jointly developed the scientific
797 concept, designed the benchmark and application examples, interpreted the results, and reviewed and edited
798 the manuscript.

799 **Competing interests**

800 The authors declare that they have no competing interests.

801 **Acknowledgements**

802 The authors acknowledge the long-term research infrastructure provided by OZCAR (Observatoire de la
803 Zone Critique) and the SNO H+ network of hydrogeological observatories.

804 **Financial support**

805 This work was supported by the France 2030 investment plan through targeted project 22-PEXO-0005
806 under the PEPR OneWater – Bien Commun programme.

807 **References**

- 808 Aeschbach, W., Beyerle, U., Zoellmann, K. P., and Kinzelbach, W.: Use of spreadsheet Boxmodel for age
809 determination with tritium and CFCs, in: Use of Chlorofluorocarbons in Hydrology: A Guidebook,
810 International Atomic Energy Agency, Vienna, 243–246, 2006.
- 811 Alikhani, J., Deinhart, A. L., Visser, A., Bibby, R. K., Purtschert, R., Moran, J. E., Massoudieh, A., and
812 Esser, B. K.: Nitrate vulnerability projections from Bayesian inference of multiple groundwater age
813 tracers, J. Hydrol., 543, 167–181, <https://doi.org/10.1016/j.jhydrol.2016.04.028>, 2016.



- 814 Amin, I. E. and Campana, M. E.: A general lumped parameter model for the interpretation of tracer data
 815 and transit time calculation in hydrologic systems, *J. Hydrol.*, 179, 1–21,
 816 [https://doi.org/10.1016/0022-1694\(95\)02880-3](https://doi.org/10.1016/0022-1694(95)02880-3), 1996.
- 817 Appelo, C. A. J. and Postma, D.: *Geochemistry, Groundwater and Pollution*, Balkema, Rotterdam, 1993.
- 818 Bayari, S.: TRACER: an EXCEL workbook to calculate mean residence time in groundwater by use of
 819 tracers CFC-11, CFC-12 and tritium, *Comput. Geosci.*, 28, 621–630,
 820 [https://doi.org/10.1016/S0098-3004\(01\)00094-2](https://doi.org/10.1016/S0098-3004(01)00094-2), 2002.
- 821 Bethke, C. M. and Johnson, T. M.: Ground water age, *Ground Water*, 40, 337–339,
 822 <https://doi.org/10.1111/j.1745-6584.2002.tb02510.x>, 2002.
- 823 Böhlke, J. K.: Tracermodel1: Excel workbook for calculation and presentation of environmental tracer data
 824 for simple groundwater mixtures, in: *Use of Chlorofluorocarbons in Hydrology: A Guidebook*,
 825 International Atomic Energy Agency, Vienna, 239–243, 2006.
- 826 Botter, G., Bertuzzo, E., and Rinaldo, A.: Catchment residence and travel time distributions: The master
 827 equation, *Geophys. Res. Lett.*, 38, L11403, <https://doi.org/10.1029/2011GL047666>, 2011.
- 828 Busenberg, E. and Plummer, L. N.: Use of chlorofluorocarbons (CCl_3F and CCl_2F_2) as hydrologic tracers
 829 and age-dating tools: the alluvium and terrace system of central Oklahoma, *Water Resour. Res.*,
 830 28, 2257–2283, <https://doi.org/10.1029/92WR01263>, 1992.
- 831 Busenberg, E. and Plummer, L. N.: Dating young groundwater with sulfur hexafluoride: natural and
 832 anthropogenic sources of sulfur hexafluoride, *Water Resour. Res.*, 36, 3011–3030,
 833 <https://doi.org/10.1029/2000WR900151>, 2000.



- 834 Carrera, J. and Neuman, S. P.: Estimation of aquifer parameters under transient and steady state conditions:
 835 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, 22, 199–210,
 836 <https://doi.org/10.1029/WR022i002p00199>, 1986.
- 837 Chambers, L. A., Gooddy, D. C., and Binley, A. M.: Use and application of CFC-11, CFC-12, CFC-113
 838 and SF₆ as environmental tracers of groundwater residence time: A review, *Geosci. Front.*, 10,
 839 1643–1652, <https://doi.org/10.1016/j.gsf.2018.02.017>, 2019.
- 840 Clark, I.: *Groundwater Geochemistry and Isotopes*, CRC Press, Boca Raton,
 841 <https://doi.org/10.1201/b18347>, 2015.
- 842 Cook, P. G. and Herczeg, A. L. (Eds.): *Environmental Tracers in Subsurface Hydrology*, Kluwer Academic
 843 Publishers, Boston, Massachusetts, USA, <https://doi.org/10.1007/978-1-4615-4557-6>, 2000.
- 844 Corcho Alvarado, J. A., Purtschert, R., Barbecot, F., Chabault, C., Rueedi, J., Schneider, V., Aeschbach-
 845 Hertig, W., Kipfer, R., and Loosli, H. H.: Constraining the age distribution of highly mixed
 846 groundwater using ³⁹Ar: A multiple environmental tracer (³H/³He, ⁸⁵Kr, ³⁹Ar, and ¹⁴C) study in the
 847 semiconfined Fontainebleau Sands Aquifer (France), *Water Resour. Res.*, 43, W03427,
 848 <https://doi.org/10.1029/2006WR005096>, 2007.
- 849 Dávila, P. F., Külls, C., and Weiler, M.: A toolkit for groundwater mean residence time interpretation with
 850 gaseous tracers, *Comput. Geosci.*, 61, 116–125, <https://doi.org/10.1016/j.cageo.2013.08.006>, 2013.
- 851 de Dreuzy, J.-R.: *PyAges: Groundwater Age Dating Toolkit*, version 1.0 [software], Zenodo,
 852 <https://doi.org/10.5281/zenodo.22150863>, 2026.
- 853 de Dreuzy, J.-R., Bodin, J., Le Grand, H., Davy, P., Boulanger, D., Battais, A., Bour, O., Gouze, P., and
 854 Porel, G.: General database for ground water site information, *Ground Water*, 44, 743–748,
 855 <https://doi.org/10.1111/j.1745-6584.2006.00220.x>, 2006.



- 856 de Dreuzy, J.-R., Bouchez, C., Marçais, J., Yvard, B., Sedel, A., Vergnaud, V., Chatton, E., Labasque, T.,
 857 and Aquilina, L.: Long-term CFC monitoring advances groundwater age interpretation and reveals
 858 transit-time dynamics in a fractured crystalline aquifer, *Hydrol. Process.*, in review.
- 859 de Marsily, G., Delhomme, J. P., Coudrain-Ribstein, A., and Lavenue, A. M.: Four decades of inverse
 860 problems in hydrogeology, in: *Theory, Modeling, and Field Investigation in Hydrogeology: A*
 861 *Special Volume in Honor of Shlomo P. Neuman's 60th Birthday*, edited by: Zhang, D. and Winter,
 862 C. L., Geological Society of America, Boulder, Colorado, USA, 1–17, [https://doi.org/10.1130/0-](https://doi.org/10.1130/0-8137-2348-5.1)
 863 [8137-2348-5.1](https://doi.org/10.1130/0-8137-2348-5.1), 2000.
- 864 Dutton, G. S., Hall, B. D., Dlugokencky, E. J., Lan, X., Nance, J. D., and Madronich, M.: Combined
 865 Atmospheric Sulfur Hexafluoride Dry Air Mole Fractions from the NOAA GML Halocarbons
 866 Sampling Network, 1995–2023, NOAA Global Monitoring Laboratory [data set],
 867 <https://doi.org/10.15138/TQ02-ZX42>, 2023.
- 868 Dutton, G. S., Hall, B. D., Montzka, S. A., Nance, J. D., Clingan, S. D., and Petersen, K. M.: Combined
 869 Atmospheric Chlorofluorocarbon-11 Dry Air Mole Fractions from the NOAA GML Halocarbons
 870 Sampling Network, 1977–2025, NOAA Global Monitoring Laboratory [data set],
 871 <https://doi.org/10.15138/BVQ6-2S69>, 2025a.
- 872 Dutton, G. S., Hall, B. D., Montzka, S. A., Nance, J. D., Clingan, S. D., and Petersen, K. M.: Combined
 873 Atmospheric Chlorofluorocarbon-12 Dry Air Mole Fractions from the NOAA GML Halocarbons
 874 Sampling Network, 1977–2025, NOAA Global Monitoring Laboratory [data set],
 875 <https://doi.org/10.15138/PJ63-H440>, 2025b.
- 876 Dutton, G. S., Hall, B. D., Montzka, S. A., Nance, J. D., Clingan, S. D., and Petersen, K. M.: Combined
 877 Atmospheric Chlorofluorocarbon-113 Dry Air Mole Fractions from the NOAA GML Halocarbons
 878 Sampling Network, 1992–2025, NOAA Global Monitoring Laboratory [data set],
 879 <https://doi.org/10.15138/4N0D-4M07>, 2025c.



- 880 Eberts, S. M., Böhlke, J. K., Kauffman, L. J., and Jurgens, B. C.: Comparison of particle-tracking and
 881 lumped-parameter age-distribution models for evaluating vulnerability of production wells to
 882 contamination, *Hydrogeol. J.*, 20, 263–282, <https://doi.org/10.1007/s10040-011-0810-6>, 2012.
- 883 Engdahl, N. B. and Maxwell, R. M.: Approximating groundwater age distributions using simple streamtube
 884 models and multiple tracers, *Adv. Water Resour.*, 66, 19–31,
 885 <https://doi.org/10.1016/j.advwatres.2014.02.001>, 2014.
- 886 Engdahl, N. B., Ginn, T. R., and Fogg, G. E.: Using groundwater age distributions to estimate the effective
 887 parameters of Fickian and non-Fickian models of solute transport, *Adv. Water Resour.*, 54, 11–21,
 888 <https://doi.org/10.1016/j.advwatres.2012.12.008>, 2013.
- 889 Gauvain, A., Leray, S., Marçais, J., Roques, C., Vautier, C., Gresselin, F., Aquilina, L., and de Dreuz, J.-
 890 R.: Geomorphological controls on groundwater transit times: A synthetic analysis at the hillslope
 891 scale, *Water Resour. Res.*, 57, e2020WR029463, <https://doi.org/10.1029/2020WR029463>, 2021.
- 892 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: *Bayesian Data*
 893 *Analysis*, 3rd edn., Chapman and Hall/CRC, Boca Raton, Florida, USA,
 894 <https://doi.org/10.1201/b16018>, 2013.
- 895 Ginn, T. R., Haeri, H., Massoudieh, A., and Foglia, L.: Notes on groundwater age in forward and inverse
 896 modeling, *Transp. Porous Media*, 79, 117–134, <https://doi.org/10.1007/s11242-009-9406-1>, 2009.
- 897 Green, C. T., Zhang, Y., Jurgens, B. C., Starn, J. J., and Landon, M. K.: Accuracy of travel-time-distribution
 898 models as affected by TTD complexity, observation errors, and model and tracer selection, *Water*
 899 *Resour. Res.*, 50, 6191–6213, <https://doi.org/10.1002/2014WR015625>, 2014.
- 900 Harman, C. J. and Xu Fei, E.: *mesas.py v1.0: a flexible Python package for modeling solute transport and*
 901 *transit times using StorAge Selection functions*, *Geosci. Model Dev.*, 17, 477–495,
 902 <https://doi.org/10.5194/gmd-17-477-2024>, 2024.



- 903 Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*,
 904 57, 97–109, <https://doi.org/10.1093/biomet/57.1.97>, 1970.
- 905 Henri, C. V., Fogg, G. E., and Harter, T.: Groundwater age and nonpoint source pollutant mixing in alluvial
 906 aquifer wells: Comparing the role of diffusion, dispersion, aquifer heterogeneity, and well screen
 907 length, *Water Resour. Res.*, 62, e2025WR040063, <https://doi.org/10.1029/2025WR040063>, 2026.
- 908 Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., and Arheimer, B.: Most computational hydrology is
 909 not reproducible, so is it really science?, *Water Resour. Res.*, 52, 7548–7555,
 910 <https://doi.org/10.1002/2016WR019285>, 2016.
- 911 IAEA: Use of Chlorofluorocarbons in Hydrology: A Guidebook, International Atomic Energy Agency,
 912 Vienna, 2006.
- 913 IAEA and WMO: Global Network of Isotopes in Precipitation (GNIP) database, Water Isotope System for
 914 Data Analysis, Visualization and Electronic Retrieval (WISER), International Atomic Energy
 915 Agency and World Meteorological Organization [data set],
 916 <https://www.iaea.org/services/networks/gnip>, last access: 19 August 2026, 2026.
- 917 INSU CNRS: Database of the H⁺ network of hydrogeological observatories, SNO H⁺ [data set],
 918 <https://doi.org/10.26169/HPLUS.ALL>, 2005.
- 919 Jódar, J., Lambán, L. J., Medina, A., and Custodio, E.: Exact analytical solution of the convolution integral
 920 for classical hydrogeological lumped-parameter models and typical input tracer functions in natural
 921 gradient systems, *J. Hydrol.*, 519, 3275–3289, <https://doi.org/10.1016/j.jhydrol.2014.10.027>, 2014.
- 922 Jurgens, B. C. and Levy, Z. F.: Decadal shifts in groundwater age detected by environmental tracers across
 923 California, USA, *Geophys. Res. Lett.*, 53, e2025GL119794,
 924 <https://doi.org/10.1029/2025GL119794>, 2026.



- 925 Jurgens, B. C., Böhlke, J. K., and Eberts, S. M.: TracerLPM (Version 1): An Excel® workbook for
 926 interpreting groundwater age distributions from environmental tracer data, U.S. Geological Survey
 927 Techniques and Methods 4-F3, <https://doi.org/10.3133/tm4F3>, 2012.
- 928 Jurgens, B. C., Böhlke, J. K., Kauffman, L. J., Belitz, K., and Esser, B. K.: A partial exponential lumped
 929 parameter model to evaluate groundwater age distributions and nitrate trends in long-screened
 930 wells, *J. Hydrol.*, 543, 109–126, <https://doi.org/10.1016/j.jhydrol.2016.05.011>, 2016.
- 931 Kang, H.-J., Moon, S., Moon, S., Yoo, H. J., Kang, K., Yu, S., and Yun, S.-T.: Long-term (> 5 years), high
 932 resolution multi-tracer-based analysis of groundwater recharge-flow regimes in a complex volcanic
 933 aquifer, *J. Hydrol.*, 664, 134559, <https://doi.org/10.1016/j.jhydrol.2025.134559>, 2026.
- 934 Kazemi, G. A., Lehr, J. H., and Perrochet, P.: Groundwater Age, Wiley Interscience, Hoboken, New Jersey,
 935 USA, <https://doi.org/10.1002/0471929514>, 2006.
- 936 Kitanidis, P. K.: Parameter uncertainty in estimation of spatial functions – Bayesian analysis, *Water Resour.*
 937 *Res.*, 22, 499–507, <https://doi.org/10.1029/WR022i004p00499>, 1986.
- 938 Knowling, M. J., White, J. T., Moore, C. R., Rakowski, P., and Hayley, K.: On the assimilation of
 939 environmental tracer observations for model-based decision support, *Hydrol. Earth Syst. Sci.*, 24,
 940 1677–1689, <https://doi.org/10.5194/hess-24-1677-2020>, 2020.
- 941 Labasque, T., Aquilina, L., Vergnaud, V., and Barbecot, F.: Inter-laboratory comparison of the analyses of
 942 sulphur hexafluoride (SF₆) and three chlorofluorocarbons (CFC-11, -12 and -113) in groundwater
 943 and an air standard, *Appl. Geochem.*, 50, 118–129,
 944 <https://doi.org/10.1016/j.apgeochem.2014.03.009>, 2014.
- 945 Le Borgne, T., Bour, O., de Dreuz, J.-R., Davy, P., and Touchard, F.: Equivalent mean flow models for
 946 fractured aquifers: Insights from a pumping tests scaling interpretation, *Water Resour. Res.*, 40,
 947 W03512, <https://doi.org/10.1029/2003WR002436>, 2004.



- 948 Leibundgut, C., Małoszewski, P., and Külls, C.: Tracers in Hydrology, Wiley,
 949 <https://doi.org/10.1002/9780470747148>, 2009.
- 950 Leray, S., de Dreuzy, J.-R., Bour, O., Labasque, T., and Aquilina, L.: Contribution of age data to the
 951 characterization of complex aquifers, J. Hydrol., 464–465, 54–68,
 952 <https://doi.org/10.1016/j.jhydrol.2012.06.052>, 2012.
- 953 Leray, S., Engdahl, N. B., Massoudieh, A., Bresciani, E., and McCallum, J.: Residence time distributions
 954 for hydrologic systems: Mechanistic foundations and steady-state analytical solutions, J. Hydrol.,
 955 543, 67–87, <https://doi.org/10.1016/j.jhydrol.2016.01.068>, 2016.
- 956 Leray, S., Gauvain, A., and de Dreuzy, J.-R.: Residence time distributions in non-uniform aquifer recharge
 957 and thickness conditions – An analytical approach based on the assumption of Dupuit-Forchheimer,
 958 J. Hydrol., 574, 110–128, <https://doi.org/10.1016/j.jhydrol.2019.04.032>, 2019.
- 959 Małoszewski, P. and Zuber, A.: Determining the turnover time of groundwater systems with the aid of
 960 environmental tracers, 1. Models and their applicability, J. Hydrol., 57, 207–231,
 961 [https://doi.org/10.1016/0022-1694\(82\)90147-0](https://doi.org/10.1016/0022-1694(82)90147-0), 1982.
- 962 Małoszewski, P. and Zuber, A.: On the theory of tracer experiments in fissured rocks with a porous matrix,
 963 J. Hydrol., 79, 333–358, [https://doi.org/10.1016/0022-1694\(85\)90064-2](https://doi.org/10.1016/0022-1694(85)90064-2), 1985.
- 964 Marçais, J., Gauvain, A., Labasque, T., Abbott, B. W., Pinay, G., Aquilina, L., Chabaux, F., Viville, D.,
 965 and de Dreuzy, J.-R.: Dating groundwater with dissolved silica and CFC concentrations in
 966 crystalline aquifers, Sci. Total Environ., 636, 260–272,
 967 <https://doi.org/10.1016/j.scitotenv.2018.04.196>, 2018.
- 968 Marçais, J., de Dreuzy, J.-R., Ginn, T. R., Rousseau-Gueutin, P., and Leray, S.: Inferring transit time
 969 distributions from atmospheric tracer data: Assessment of the predictive capacities of lumped



- 970 parameter models on a 3D crystalline aquifer model, *J. Hydrol.*, 525, 619–631,
 971 <https://doi.org/10.1016/j.jhydrol.2015.03.055>, 2015.
- 972 Massoudieh, A.: Inference of long-term groundwater flow transience using environmental tracers: A
 973 theoretical approach, *Water Resour. Res.*, 49, 8039–8052,
 974 <https://doi.org/10.1002/2013WR014548>, 2013.
- 975 Massoudieh, A., Sharifi, S., and Solomon, D. K.: Bayesian evaluation of groundwater age distribution using
 976 radioactive tracers and anthropogenic chemicals, *Water Resour. Res.*, 48, W09529,
 977 <https://doi.org/10.1029/2012WR011815>, 2012.
- 978 Massoudieh, A., Leray, S., and de Dreuz, J.-R.: Assessment of the value of groundwater age time-series
 979 for characterizing complex steady-state flow systems using a Bayesian approach, *Appl. Geochem.*,
 980 50, 240–251, <https://doi.org/10.1016/j.apgeochem.2013.10.006>, 2014a.
- 981 Massoudieh, A., Visser, A., Sharifi, S., and Broers, H. P.: A Bayesian modeling approach for estimation of
 982 a shape-free groundwater age distribution using multiple tracers, *Appl. Geochem.*, 50, 252–264,
 983 <https://doi.org/10.1016/j.apgeochem.2013.10.004>, 2014b.
- 984 McCallum, J. L., Engdahl, N. B., Ginn, T. R., and Cook, P. G.: Nonparametric estimation of groundwater
 985 residence-time distributions: what can environmental tracer data tell us about groundwater
 986 residence time?, *Water Resour. Res.*, 50, 2022–2038, <https://doi.org/10.1002/2013WR014974>,
 987 2014.
- 988 McCallum, J. L., Cook, P. G., and Simmons, C. T.: Limitations of the use of environmental tracers to infer
 989 groundwater age, *Groundwater*, 53, 56–70, <https://doi.org/10.1111/gwat.12237>, 2015.
- 990 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state
 991 calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087–1092,
 992 <https://doi.org/10.1063/1.1699114>, 1953.



- 993 Ozyurt, N. N. and Bayari, C. S.: LUMPED: a Visual Basic code of lumped-parameter models for mean
 994 residence-time analyses of groundwater systems, *Comput. Geosci.*, 29, 79–90,
 995 [https://doi.org/10.1016/S0098-3004\(02\)00075-4](https://doi.org/10.1016/S0098-3004(02)00075-4), 2003.
- 996 Popp, A. L., Pardo-Álvarez, Á., Schilling, O. S., Scheidegger, A., Musy, S., Peel, M., Brunner, P.,
 997 Purtschert, R., Hunkeler, D., and Kipfer, R.: A framework for untangling transient groundwater
 998 mixing and travel times, *Water Resour. Res.*, 57, e2020WR028362,
 999 <https://doi.org/10.1029/2020WR028362>, 2021.
- 1000 Rinaldo, A., Benettin, P., Harman, C. J., Hrachowitz, M., McGuire, K. J., van der Velde, Y., Bertuzzo, E.,
 1001 and Botter, G.: Storage selection functions: A coherent framework for quantifying how catchments
 1002 store and release water and solutes, *Water Resour. Res.*, 51, 4840–4847,
 1003 <https://doi.org/10.1002/2015WR017273>, 2015.
- 1004 Robert, C. P. and Casella, G.: Metropolis–Hastings algorithms, in: *Introducing Monte Carlo Methods with*
 1005 *R*, Springer, New York, 167–197, https://doi.org/10.1007/978-1-4419-1576-4_6, 2009.
- 1006 Roberts, G. O., Gelman, A., and Gilks, W. R.: Weak convergence and optimal scaling of random walk
 1007 Metropolis algorithms, *Ann. Appl. Probab.*, 7, 110–120, <https://doi.org/10.1214/aoap/1034625254>,
 1008 1997.
- 1009 Rojas, R., Feyen, L., and Dassargues, A.: Conceptual model uncertainty in groundwater modeling:
 1010 Combining generalized likelihood uncertainty estimation and Bayesian model averaging, *Water*
 1011 *Resour. Res.*, 44, W12418, <https://doi.org/10.1029/2008WR006908>, 2008.
- 1012 Rudolph, M. G., Burghardt, D., Hartmann, A., and Wöhling, T.: PyTracerLab: An interactive tool for
 1013 lumped parameter modelling of environmental tracers, *SSRN [preprint]*,
 1014 <https://doi.org/10.2139/ssrn.7180989>, 2026.



- 1015 Ruelleu, S., Moreau, F., Bour, O., Gapais, D., and Martelet, G.: Impact of gently dipping discontinuities on
 1016 basement aquifer recharge: An example from Ploemeur (Brittany, France), *J. Appl. Geophys.*, 70,
 1017 161–168, <https://doi.org/10.1016/j.jappgeo.2009.12.007>, 2010.
- 1018 Sandborn, D. E., Carter, B. R., Warner, M. J., and Dias, L. M.: TRACE-Python: tracer-based rapid
 1019 anthropogenic carbon estimation implemented in Python (version 1.0), *Geosci. Model Dev.*, 19,
 1020 5961–5978, <https://doi.org/10.5194/gmd-19-5961-2026>, 2026.
- 1021 Sanford, W. E.: Calibration of models using groundwater age, *Hydrogeol. J.*, 19, 13–16,
 1022 <https://doi.org/10.1007/s10040-010-0637-6>, 2011.
- 1023 Schwemmler, R., Leistert, H., Steinbrich, A., and Weiler, M.: RoGeR v3.0.5 – a process-based hydrological
 1024 toolbox model in Python, *Geosci. Model Dev.*, 17, 5249–5262, [https://doi.org/10.5194/gmd-17-](https://doi.org/10.5194/gmd-17-5249-2024)
 1025 5249-2024, 2024.
- 1026 Sloan, I. H. and Smith, W. E.: Properties of interpolatory product integration rules, *SIAM J. Numer. Anal.*,
 1027 19, 427–442, <https://doi.org/10.1137/0719027>, 1982.
- 1028 Solomon, D. K. and Cook, P. G.: ^3H and ^3He , in: *Environmental Tracers in Subsurface Hydrology*, edited
 1029 by: Cook, P. G. and Herczeg, A. L., Springer, Boston, Massachusetts, USA, 397–424,
 1030 https://doi.org/10.1007/978-1-4615-4557-6_13, 2000.
- 1031 Suckow, A.: Lumpy – an interactive Lumped Parameter Modeling code based on MS Access and MS Excel,
 1032 *Geophys. Res. Abstr.*, 14, EGU2012-2763, 2012.
- 1033 Suckow, A.: The age of groundwater – Definitions, models and why we do not need this term, *Appl.*
 1034 *Geochem.*, 50, 222–230, <https://doi.org/10.1016/j.apgeochem.2014.04.016>, 2014.
- 1035 Tarantola, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia,
 1036 Pennsylvania, USA, <https://doi.org/10.1137/1.9780898717921>, 2005.



- 1037 Thiros, N. E., Gardner, W. P., and Kuhlman, K. L.: Utilizing environmental tracers to reduce groundwater
 1038 flow and transport model parameter uncertainties, *Water Resour. Res.*, 57, e2020WR028235,
 1039 <https://doi.org/10.1029/2020WR028235>, 2021.
- 1040 Thiros, N. E., Siirila-Woodburn, E. R., Dennedy-Frank, P. J., Williams, K. H., and Gardner, W. P.:
 1041 Constraining bedrock groundwater residence times in a mountain system with environmental tracer
 1042 observations and Bayesian uncertainty quantification, *Water Resour. Res.*, 59, e2022WR033282,
 1043 <https://doi.org/10.1029/2022WR033282>, 2023.
- 1044 Türk, H., Stumpp, C., Hrachowitz, M., Strauss, P., Blöschl, G., and Stockinger, M.: Catchment transit time
 1045 variability with different SAS function parameterizations for the unsaturated zone and
 1046 groundwater, *Hydrol. Earth Syst. Sci.*, 30, 1053–1076, <https://doi.org/10.5194/hess-30-1053-2026>,
 1047 2026.
- 1048 Turnadge, C. and Smerdon, B. D.: A review of methods for modelling environmental tracers in
 1049 groundwater: Advantages of tracer concentration simulation, *J. Hydrol.*, 519, 3674–3689,
 1050 <https://doi.org/10.1016/j.jhydrol.2014.10.056>, 2014.
- 1051 Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C.: Rank-normalization, folding, and
 1052 localization: An improved R-hat for assessing convergence of MCMC, *Bayesian Anal.*, 16, 667–
 1053 718, <https://doi.org/10.1214/20-BA1221>, 2021.
- 1054 Visser, A., Broers, H. P., Purtschert, R., Sültenfuß, J., and de Jonge, M.: Groundwater age distributions at
 1055 a public drinking water supply well field derived from multiple age tracers (^{85}Kr , $^3\text{H}/^3\text{He}$, and ^{39}Ar),
 1056 *Water Resour. Res.*, 49, 7778–7796, <https://doi.org/10.1002/2013WR014012>, 2013.
- 1057 Zell, W. O., Culver, T. B., and Sanford, W. E.: Prediction uncertainty and data worth assessment for
 1058 groundwater transport times in an agricultural catchment, *J. Hydrol.*, 561, 1019–1036,
 1059 <https://doi.org/10.1016/j.jhydrol.2018.02.006>, 2018.