Bayesian fitting in positronium lifetime imaging

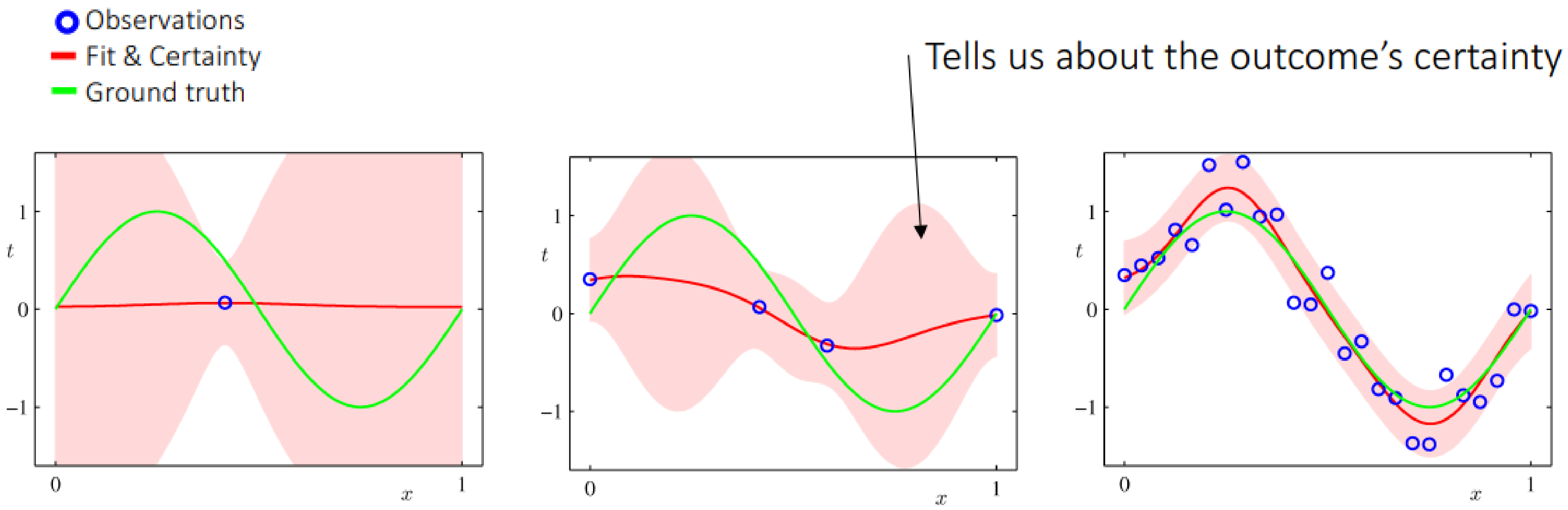
Roman Y. Shopa
16-06-2025

Outline

- Bayesian fitting
  - Uncertainty & probability
  - Knowledge update
  - General Bayesian Inference
- Linear and non-linear fitting
  - Update new data
  - Posterior sampling
- Ps decay and Ps lifetime spectrum
- Examples

Uncertainty & probability

Use probability as a vehicle to quantify degrees of belief on uncertainty in unobserved events ⟶ Bayesian statistics



https://shapemodelling.cs.unibas.ch/pmm2017/slides/bayes.pdf

Uncertainty & probability

Why does the distribution change when we have more data? Shouldn't there be a real distribution $P(\theta)$?

Bayesian probabilities rely on a subjective perspective:

probability is used to express our *current knowledge*. It can change when we learn or see more: With more data, we are more certain about our result.

Subjectivity: There is no single, real underlying distribution. A probability distribution expresses our knowledge – It is different in different situations and for different observers since they have different knowledge.

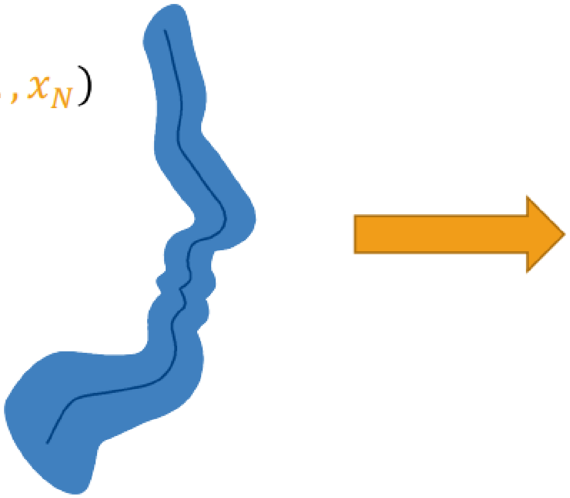
https://shapemodelling.cs.unibas.ch/pmm2017/slides/bayes.pdf

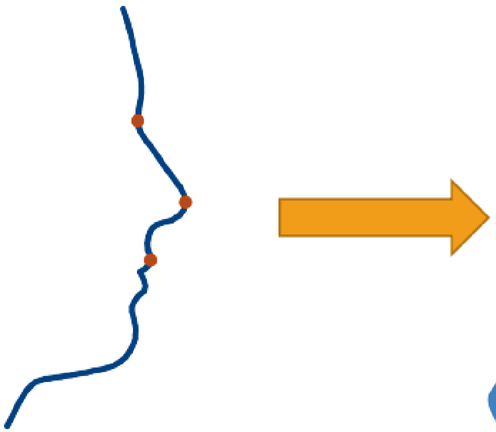
Knowledge update

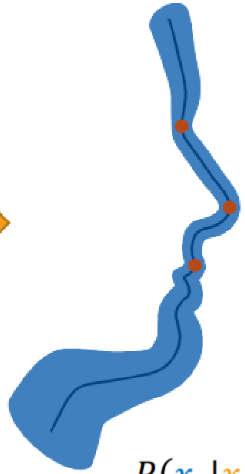
Update belief about x_1 by *observing* $x_2, \dots, x_N$

$$P(x_1) \to P(x_1 | x_2 \dots x_N)$$

$P(x_2, \dots, x_N)$







$$P(x_1 | x_2 \dots x_N) = \frac{P(x_1, x_2, \dots, x_N)}{P(x_2, \dots, x_N)}$$

Model	Observation	Posterior
Face distribution	Concrete points	Face distribution
	Possibly uncertain	*consistent with observation*
Prior belief	More knowledge	Posterior belief

https://shapemodelling.cs.unibas.ch/pmm2017/slides/bayes.pdf

General Bayesian Inference

General Bayesian inference case:

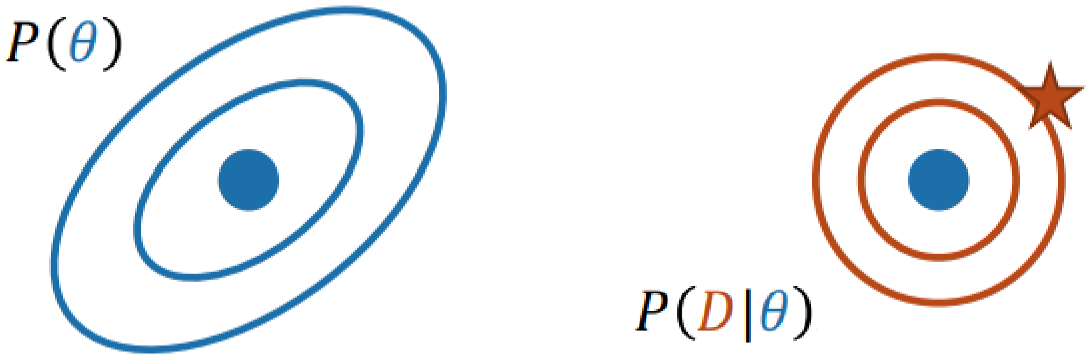
- Distribution of data $\boldsymbol{D} \equiv \boldsymbol{y}$ (or Evidence)
- Parameters θ (or Query)
- Update method (similar to MLEM): $P(\theta|D) \propto P(D|\theta)P(\theta)$

likelihood prior

posterior

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

marginal likelihood
(constant)



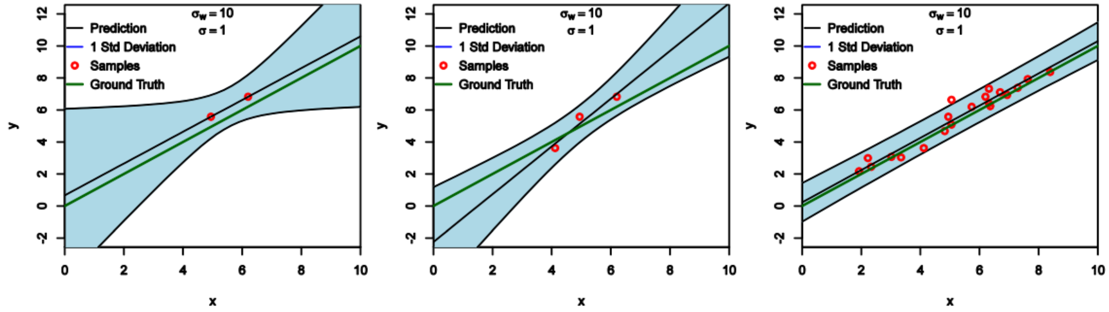
https://shapemodelling.cs.unibas.ch/pmm2017/slides/bayes.pdf

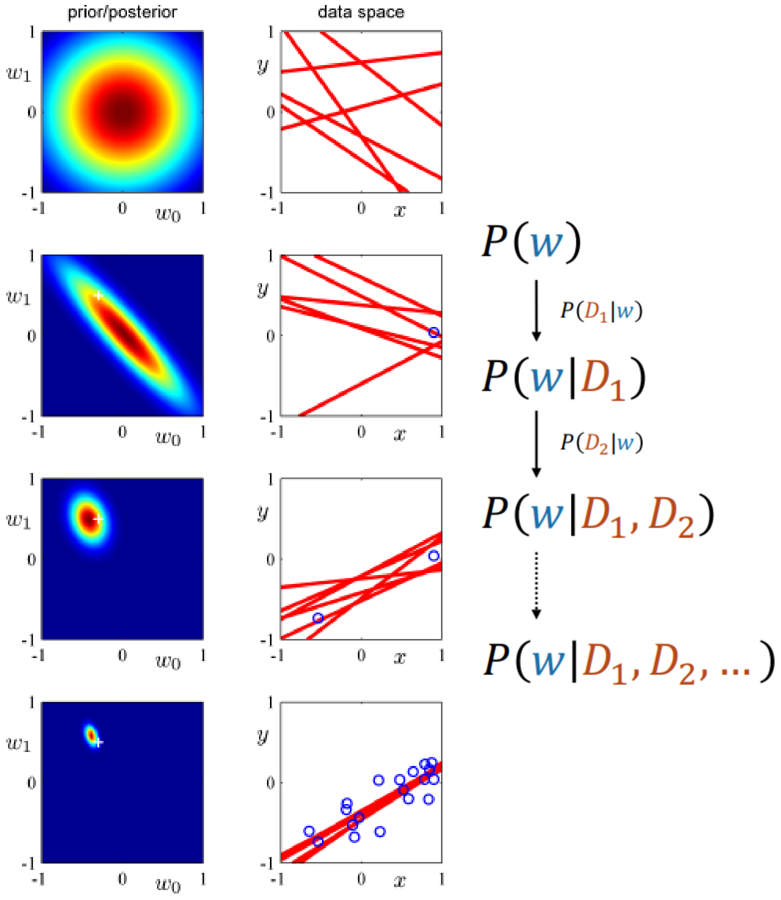
Update new data

Inference case:

- beliefs evolve with observation
- Recursive: Posterior becomes prior for the next inference step (new data arrives)

Example: curve fitting





Bishop PRML, 2006

https://shapemodelling.cs.unibas.ch/pmm2017/slides/bayes.pdf

Linear and non-linear fitting

Linear regression: many predictors (***X*** as a matrix), errors ε are independent and identically normally distributed

$$y = \beta^T X + \varepsilon$$

Non-linear, multilevel models (MLMs): accounts for the population-level and grouped-level coefficients β and u, with the corresponding design matrices ***X*** and ***Z***.
Also, custom distribution ("family" D and link function $f(\eta)$)

$$y_i \sim D(f(\eta_i), \theta)$$

$$\eta = \mathbf{X}\beta + \mathbf{Z}u$$

Bürkner, brms: An R Package... (2017)

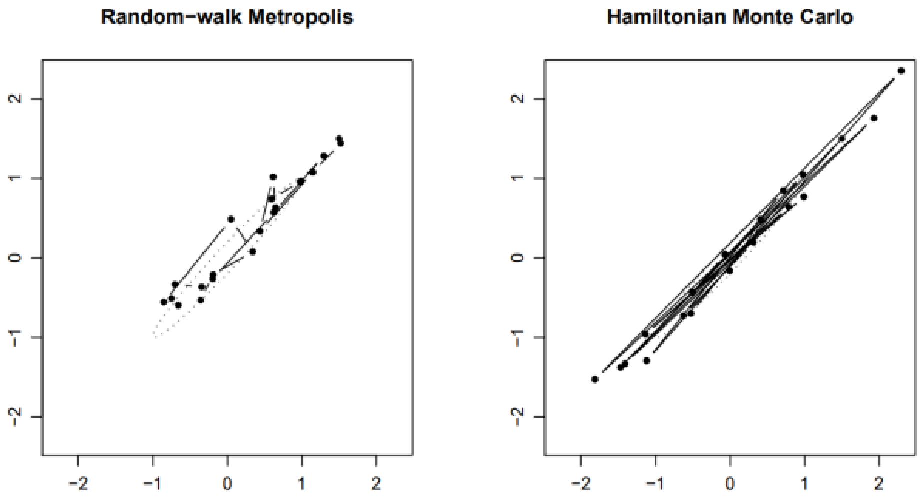
What if we don't have new data?

Posterior sampling

What if we don't have new data? – sample many candidates and refine:

– start with initial parameters θ as random guesses (rough distribution)
– generate a new candidate in the parameter space
– accept or reject the new value based on likelihood
– repeat for thousands of iterations

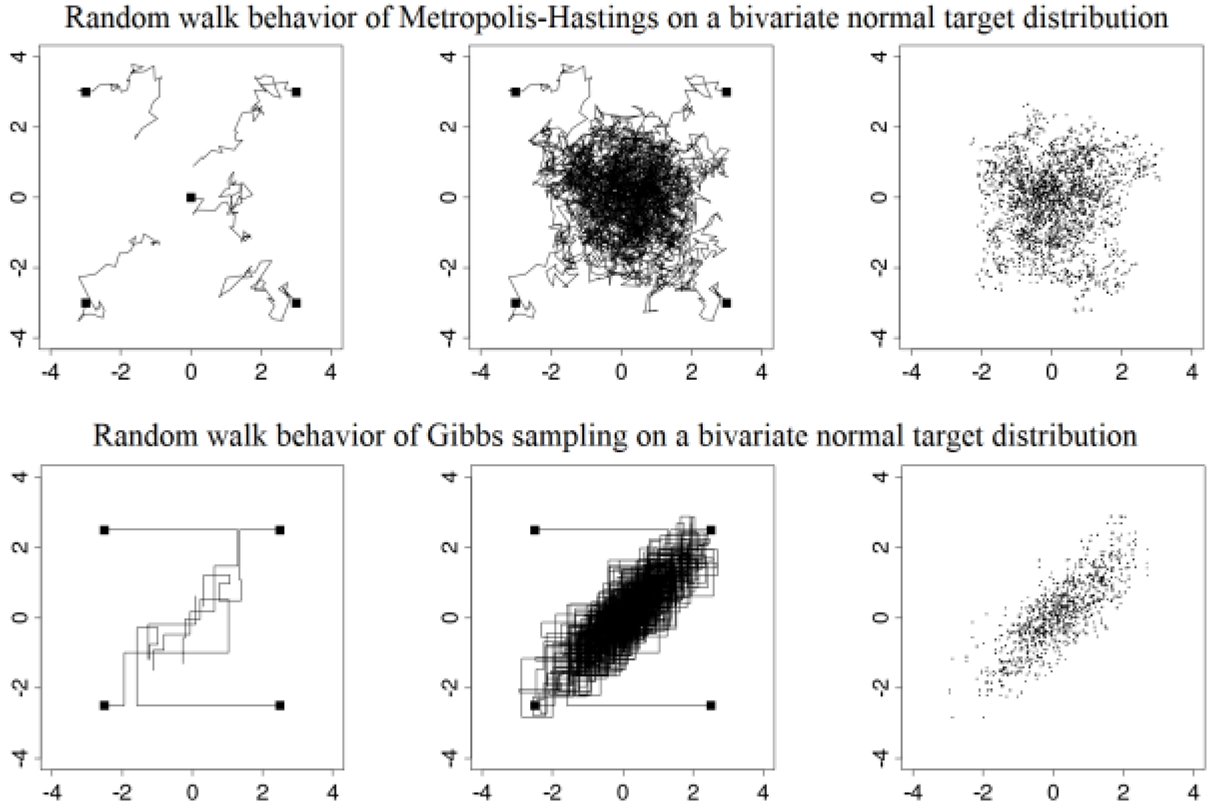
How to generate samples:
Random walk (least efficient)
Gibbs,
Markov Chain Monte Carlo (MCMC):
Metropolis-Hastings,
Hamiltonian Monte Carlo,
No-U-turn sampler (NUTS)



(figure from Neal (2011))

Posterior sampling

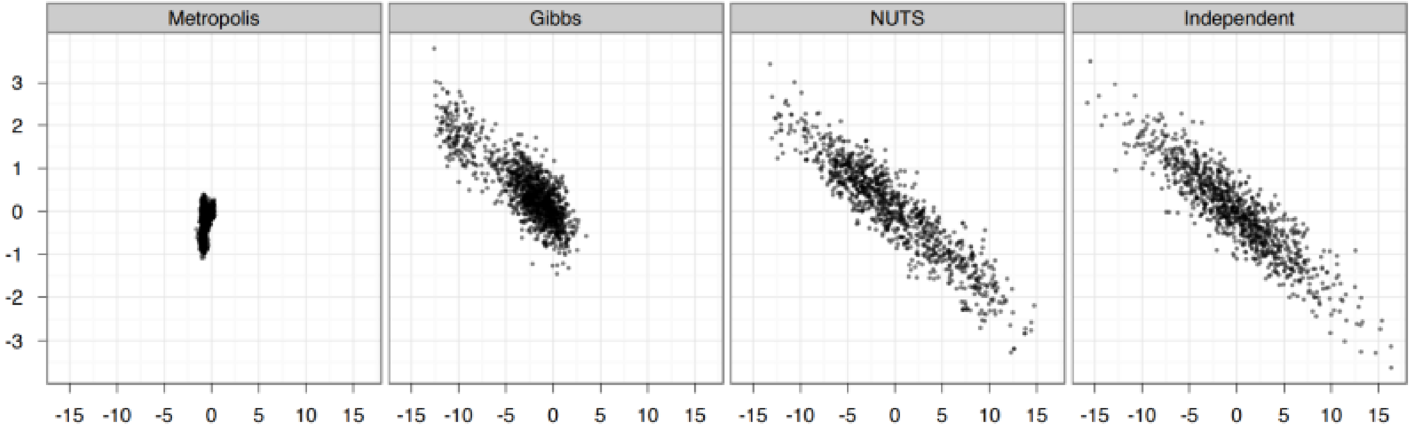
figure from Gelman et al. (2013), BDA3, Chapter 11



Posterior sampling

figure from Hoffman & Gelman (2014)

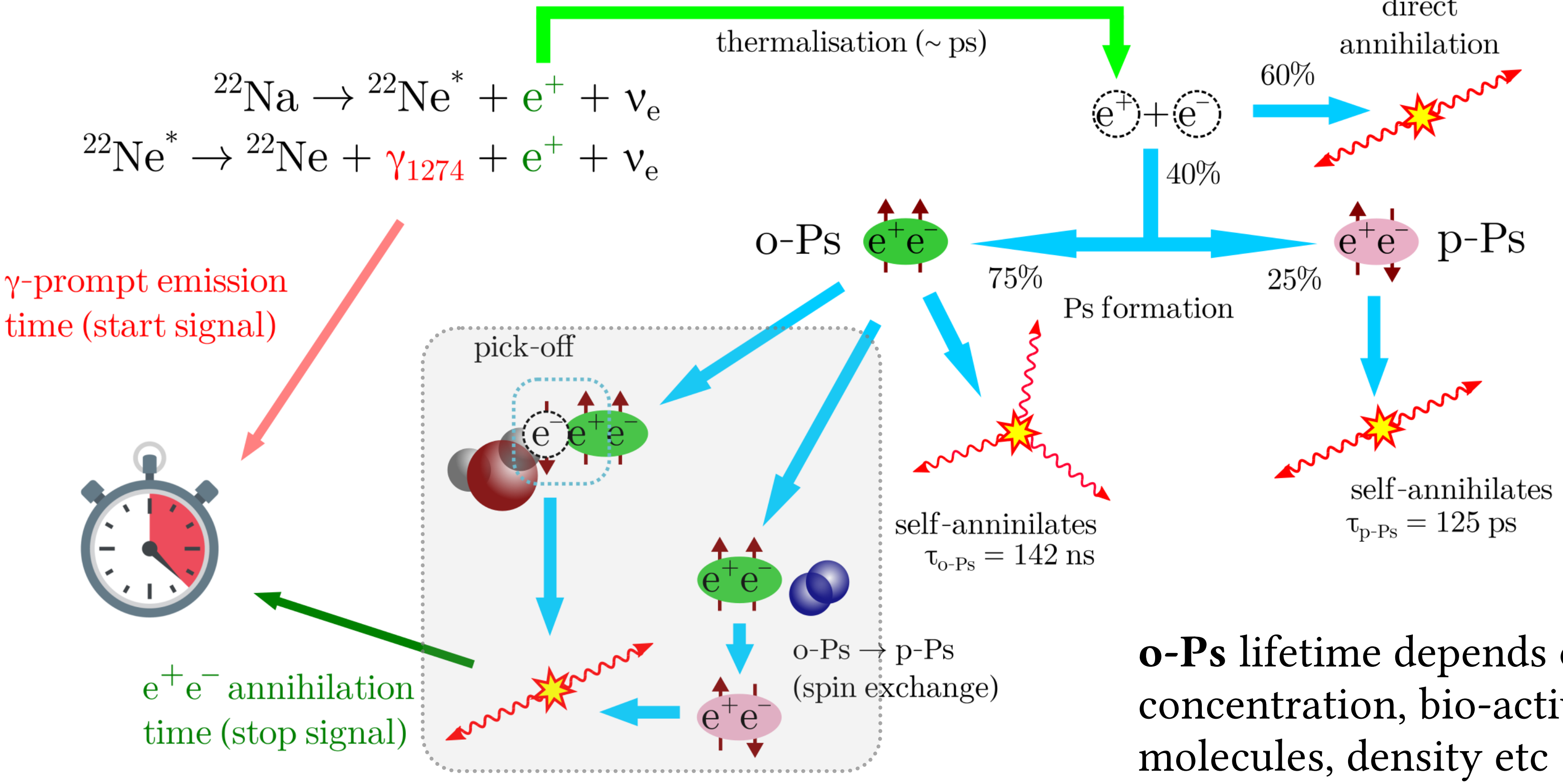
Advanced sample generation: Hamiltonian Monte Carlo → no-U-turn sampler (NUTS).



Simulating a "*physical system*" via multiple MCMC chains where parameter changes are guided by *gradient-based forces* derived from the likelihood function. Accepted values if increased $P(y\,|\theta)$

Convergence & Diagnostics: chains should mix well, R-hat stats ~1.0 indicate convergence, Effective Sample Size (ESS) should be sufficiently large

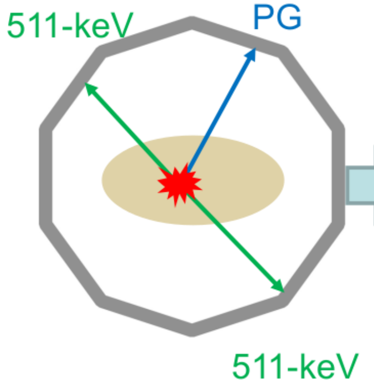
Positronium (Ps) decay



o-Ps lifetime depends on O_2 concentration, bio-active molecules, density etc

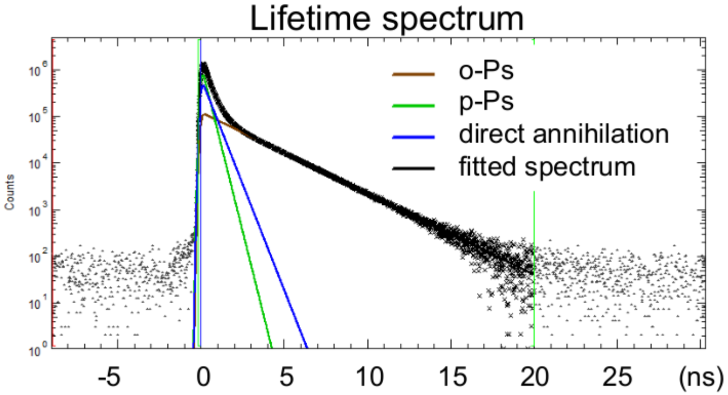
Ps lifetime spectrum

From J.Qi presentation (Jagiellonian symposiun 2024)



Lifetime measurement
(after travel-time correction)

$$\tau = \frac{(t_{1st\ 511} + t_{2nd\ 511})}{2} - t_{PG}$$



List-mode event parameters

- LOR i_k, including TOF information
- Time delay τ_k

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \qquad P(\theta|y) \propto P(y|\theta)P(\theta)$$

Likelihood model of time measurement

$$p(\tau|\boldsymbol{\lambda}, \boldsymbol{A}) = \sum_l g(\tau) * [A_l \lambda_l \exp(-\lambda_l \tau) u(\tau)] + B$$

- λ_l: 1/lifetime of the l^{th} pathway
- A_l: Fraction of the l^{th} pathway
- $g(\tau)$: Detector timing response
- $u(\tau)$: Heaviside function
- B: random background events

Examples


Nuclear Inst. and Methods in Physics Research, A 928 (2019) 37–42

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima

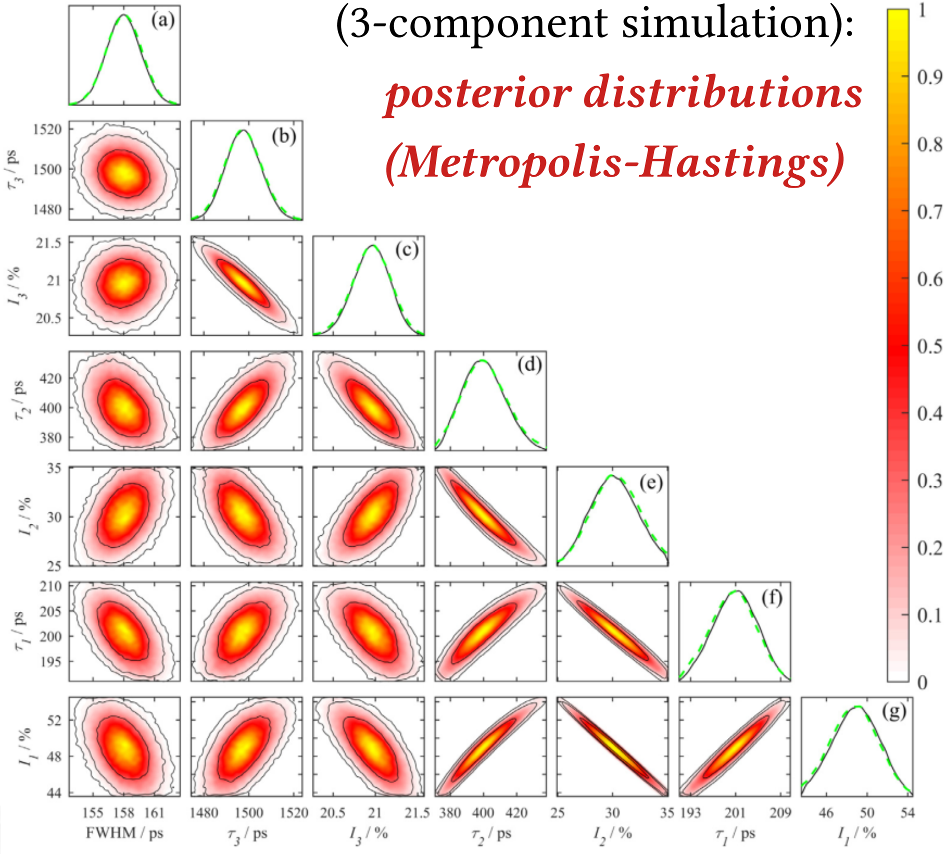
Accurate and informative analysis of positron annihilation lifetime spectra by using Markov Chain Monte-Carlo Bayesian inference method

B.C. Gu [a,b], W.S. Zhang [c,*], J.D. Liu [a,b], H.J. Zhang [a,b], B.J. Ye [a,b,**]

[a] State Key Laboratory of Particle Detection and Electronics, University of Science and Technology of China, Hefei 230026, China
[b] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China
[c] Network Information Center, Supercomputing Center, University of Science and Technology of China, Hefei, 230026, China

(3-component simulation):

posterior distributions (Metropolis-Hastings)



$$N(t) = \sum_{j=1}^{k_0} [A_j \exp(-t/\tau_j)] \otimes R(t) + B$$

likelihood:

$$p(\boldsymbol{D}|\boldsymbol{\theta}) = \frac{\exp(-\frac{1}{2}\chi^2)}{\prod_k \sqrt{2\pi N(k)}}$$

$$\chi^2 = \sum_k \frac{[N^\theta(k) - N(k)]^2}{N^\theta(k)}$$

time delay t is quantised by “channels” k

Examples

Software:

Julia, Turing.jl, ArviZ, Stan, brms…

30. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: a fresh approach to numerical computing. SIAM Rev. 2017;59:65–98.
31. Ge H, Xu K, Ghahramani Z. Turing: a language for flexible probabilistic inference. In International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9–11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain, 2018;1682–1690.
32. Kumar R, Carroll C, Hartikainen A, Martin O. ArviZ a unified library for exploratory analysis of Bayesian models in python. J Open Source Softw. 2019;4:1143. https://doi.org/10.21105/joss.01143.

Another example:

Steinberger *et al. EJNMMI Physics* (2024) 11:76
https://doi.org/10.1186/s40658-024-00678-4

EJNMMI Physics

ORIGINAL RESEARCH Open Access

Positronium lifetime validation measurements using a long-axial field-of-view positron emission tomography scanner

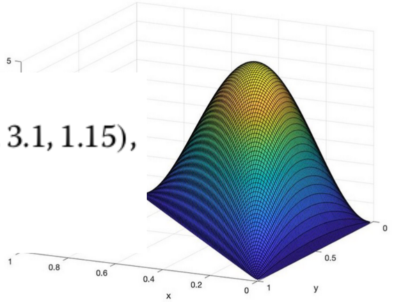
William M. Steinberger[1*], Lorenzo Mercolli[2], Johannes Breuer[3], Hasan Sari[4], Szymon Parzych[5], Szymon Niedzwiecki[5], Gabriela Lapkiewicz[5], Pawel Moskal[5], Ewa Stepien[5], Axel Rominger[2], Kuangyu Shi[2] and Maurizio Conti[1]

$$F(t) = b + N \cdot \sum_{i=1}^{3} \frac{BR_i}{2\tau_i} e^{(\sigma^2 - 2t\tau_i + 2\Delta\,\tau_i)/(2\tau_i^2)} \cdot \mathrm{erfc}\left(\frac{\sigma}{\sqrt{2}\tau_i} + \frac{\Delta - t}{\sqrt{2}\sigma}\right)$$

$$\tau_1 = 0.125\,\text{ns}$$
$$\tau_2 = 0.388\,\text{ns}$$

priors:

$$\tau_3 \sim \mathcal{N}(1.5\,\text{ns}, 0.5\,\text{ns}),$$
$$BR_{1,2,3} \sim \text{Dirichlet}(0.75, 3.1, 1.15),$$
$$\sigma \sim \mathcal{N}(0.1\,\text{ns}, 0.05\,\text{ns}),$$
$$\Delta \sim \mathcal{N}(0.0\,\text{ns}, 0.5\,\text{ns}).$$



Examples

Steinberger *et al. EJNMMI Physics* (2024) 11:76
https://doi.org/10.1186/s40658-024-00678-4

EJNMMI Physics

ORIGINAL RESEARCH Open Access

Positronium lifetime validation measurements using a long-axial field-of-view positron emission tomography scanner

William M. Steinberger[1*], Lorenzo Mercolli[2], Johannes Breuer[3], Hasan Sari[4], Szymon Parzych[5], Szymon Niedzwiecki[5], Gabriela Lapkiewicz[5], Pawel Moskal[5], Ewa Stepien[5], Axel Rominger[2], Kuangyu Shi[2] and Maurizio Conti[1]

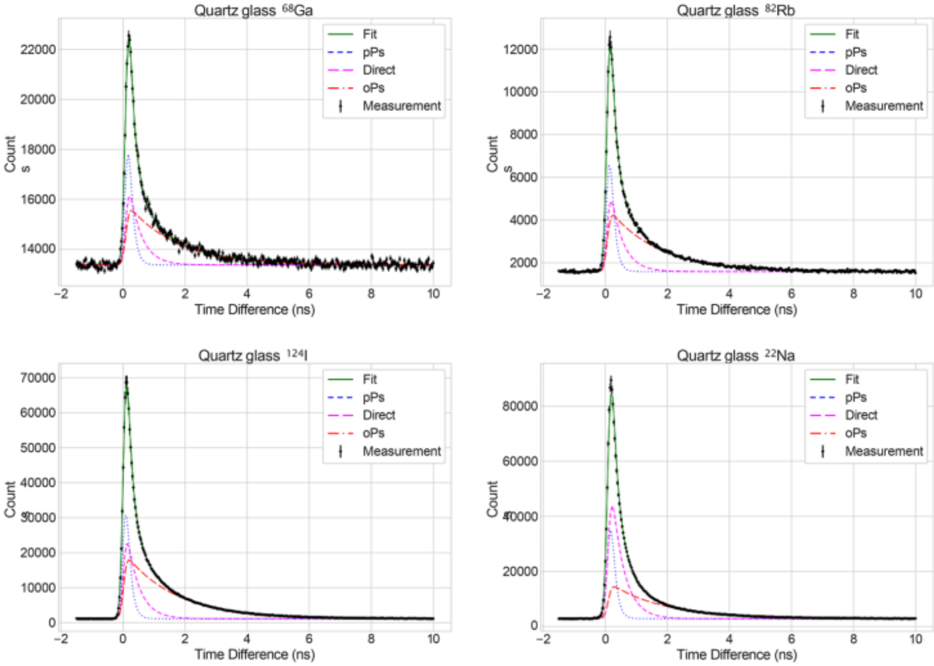


Fig. 14 Bayes fit using Eq. (7) to quartz glass data with the contributions of the Ps lifetime components

$$\tau_3 \sim \mathcal{N}(1.5\,\text{ns}, 0.5\,\text{ns}),$$
$$BR_{1,2,3} \sim \text{Dirichlet}(0.75, 3.1, 1.15),$$
$$\sigma \sim \mathcal{N}(0.1\,\text{ns}, 0.05\,\text{ns}),$$
$$\Delta \sim \mathcal{N}(0.0\,\text{ns}, 0.5\,\text{ns}).$$

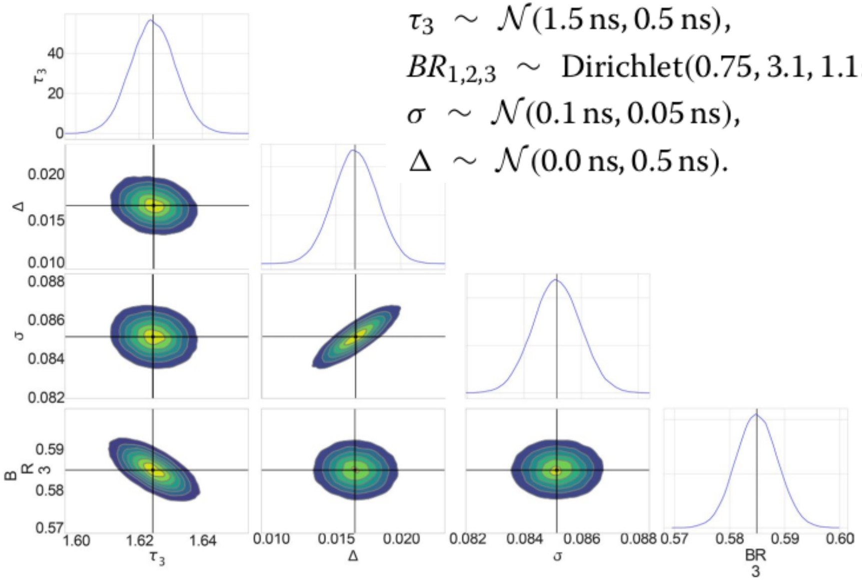


Fig. 15 Pair plot of the posterior distributions for the parameters τ_3, BR_3, σ and Δ for the ^{124}I quartz glass data

Examples

Example with hierarchical models:

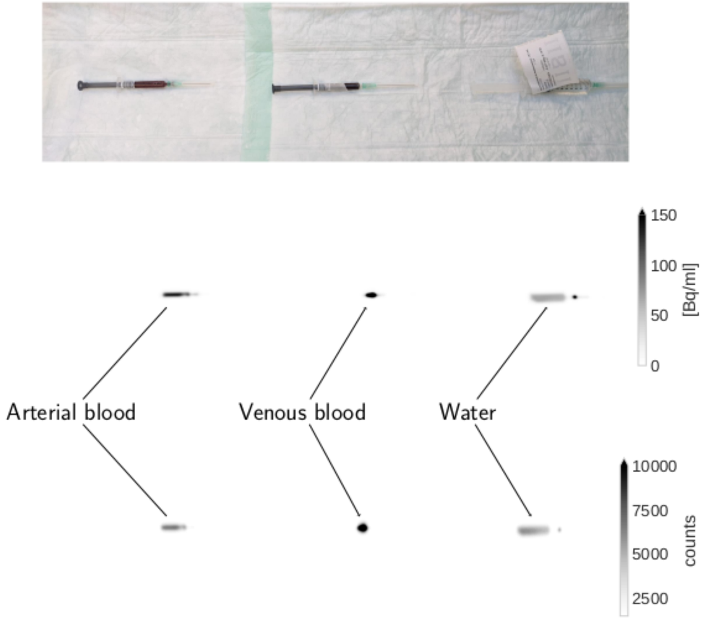


Fig. 11: Picture of the ^{124}I blood samples and water lying on Quadra's patient bed (top) together with the MIP of the coincidence PET image (middle) and 3γE histoimage (bottom).

In Vivo Positronium Lifetime Measurements with a Long Axial Field-of-View PET/CT

Lorenzo Mercolli[1,2*], William M. Steinberger[3], Hasan Sari[1,2,4], Ali Afshar-Oromieh[1], Federico Caobelli[1], Maurizio Conti[3], Ângelo R. Felgosa Cardoso[1], Clemens Mingels[1], Paweł Moskal[5,6], Thomas Pyka[1], Narendra Rathod[1,2], Robin Schepers[1], Robert Seifert[1], Kuangyu Shi[1,2], Ewa Ł. Stępień[5,6], Marco Viscione[1], Axel O. Rominger[1]

[1]Department of Nuclear Medicine, Inselspital, Bern University Hostpital, University of Bern, Bern, Switzerland.
[2]ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland.
[3]Siemens Medical Solutions USA, Inc., Knoxville TN, USA.
[4]Siemens Healthineers International AG, Zürich, Switzerland.
[5]Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krakow, Poland.
[6]Centre for Theranostics, Jagiellonian University, Krakow, Poland.

Examples

Example with hierarchical models
(some parameters are shared across samples):

In Vivo Positronium Lifetime Measurements with
a Long Axial Field-of-View PET/CT

Lorenzo Mercolli[1,2*], William M. Steinberger[3], Hasan Sari[1,2,4],
Ali Afshar-Oromieh[1], Federico Caobelli[1], Maurizio Conti[3],
Ângelo R. Felgosa Cardoso[1], Clemens Mingels[1], Paweł Moskal[5,6],

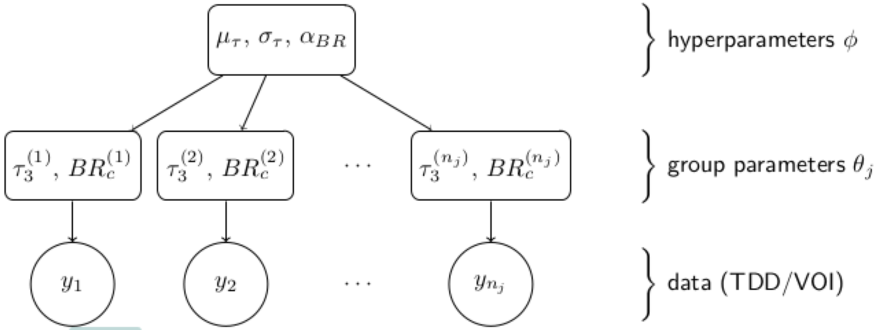


Fig. 2: Hierarchical model for the combined analysis of data from different VC

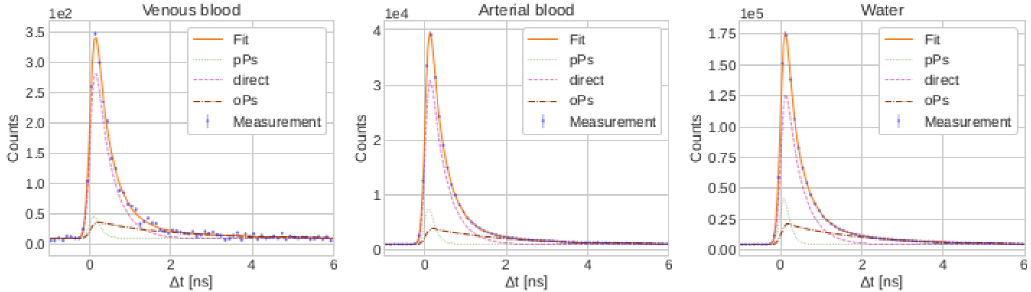


Fig. 12: TDD and fit prediction for the three ^{124}I samples.

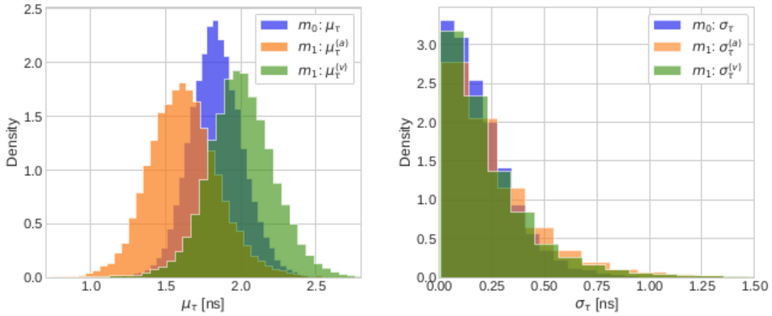


Fig. 13: Comparison between the sampled posterior distributions of the hyperparameters of the models m_0 and m_1 (see Fig. 3).

Examples

PLI using Bayesian fitting:

$$F(t) = b + N \cdot \sum_{i=1}^{3} \frac{BR_i}{2\tau_i} e^{(\sigma^2 - 2t\tau_i + 2\Delta\,\tau_i)/(2\tau_i^2)} \cdot \mathrm{erfc}\left(\frac{\sigma}{\sqrt{2}\tau_i} + \frac{\Delta - t}{\sqrt{2}\sigma} \right)$$

$$\tau_2 = 0.388\,\mathrm{ns}$$
$$\tau_1 = 0.125\,\mathrm{ns}$$

$$\tau_3 \sim \mathcal{N}(1.78\,\mathrm{ns}, 0.8\,\mathrm{ns})\ ,$$
$$BR_{1,2,3} \sim \mathrm{Dirichlet}(0.75, 3.1, 1.15)$$
$$\sigma \sim \mathcal{N}(0.1\,\mathrm{ns}, 0.05\,\mathrm{ns})\,,$$
$$\Delta \sim \mathcal{N}(0\,\mathrm{ns}, 0.5\,\mathrm{ns})\ ,$$

Positronium Lifetime Imaging with the Biograph Vision Quadra using ^{124}I

Lorenzo Mercolli[1,2*], William M. Steinberger[3], Narendra Rathod[1,2], Maurizio Conti[3], Paweł Moskal[4,5], Axel Rominger[1], Robert Seifert[1], Kuangyu Shi[1,2], Ewa Ł. Stępień[4,5], Hasan Sari[1,2,6]

[1*]Department of Nuclear Medicine, Inselspital, Bern University Hostpital, University of Bern, Bern, Switzerland.
[2]ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland.
[3]Siemens Medical Solutions USA, Inc., Knoxville TN, USA.
[4]Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krakow, Poland.
[5]Centre for Theranostics, Jagiellonian University, Krakow, Poland.

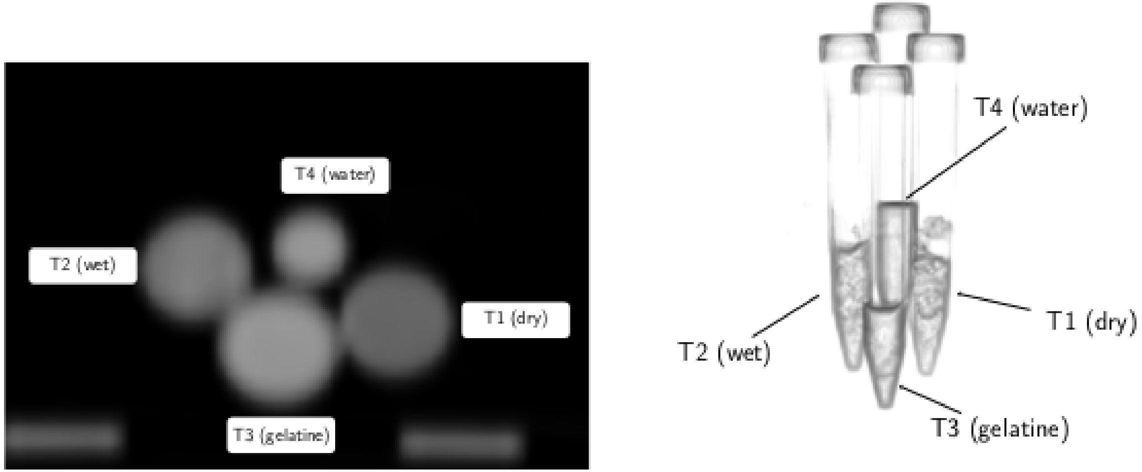
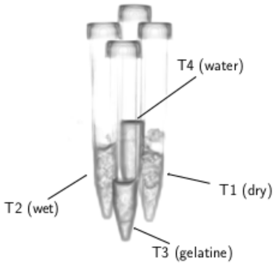


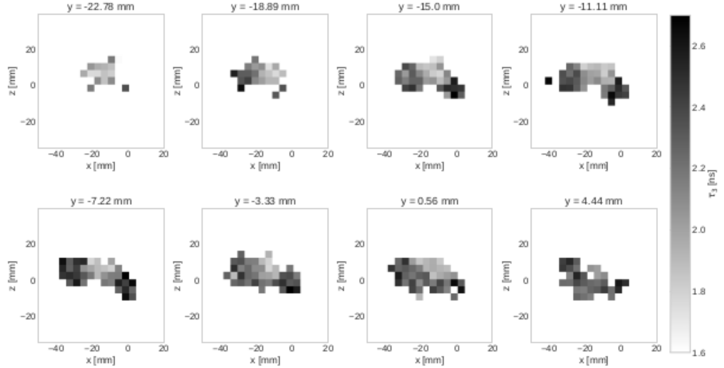
Fig. 3: Top view of a CT slice (right) and 3D rendering of the CT (left) of the setup with the four tubes taped together.

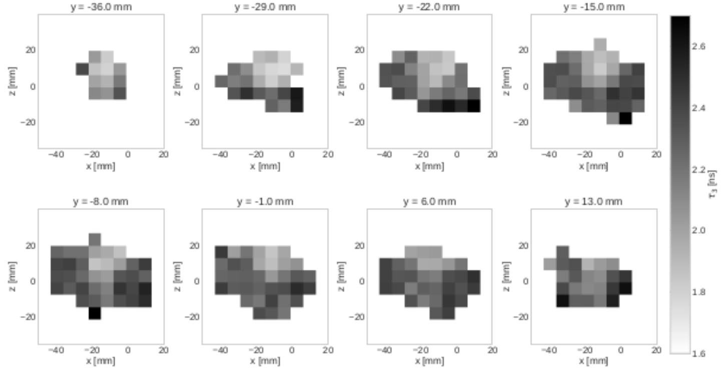
Examples

PLI using Bayesian fitting:



) rendering of the CT (left) of the setup





Positronium Lifetime Imaging with the Biograph Vision Quadra using ^{124}I

Lorenzo Mercolli[1,2*], William M. Steinberger[3], Narendra Rathod[1,2], Maurizio Conti[3], Paweł Moskal[4,5], Axel Rominger[1], Robert Seifert[1], Kuangyu Shi[1,2], Ewa Ł. Stępień[4,5], Hasan Sari[1,2,6]

[1*]Department of Nuclear Medicine, Inselspital, Bern University Hostpital, University of Bern, Bern, Switzerland.
[2]ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland.

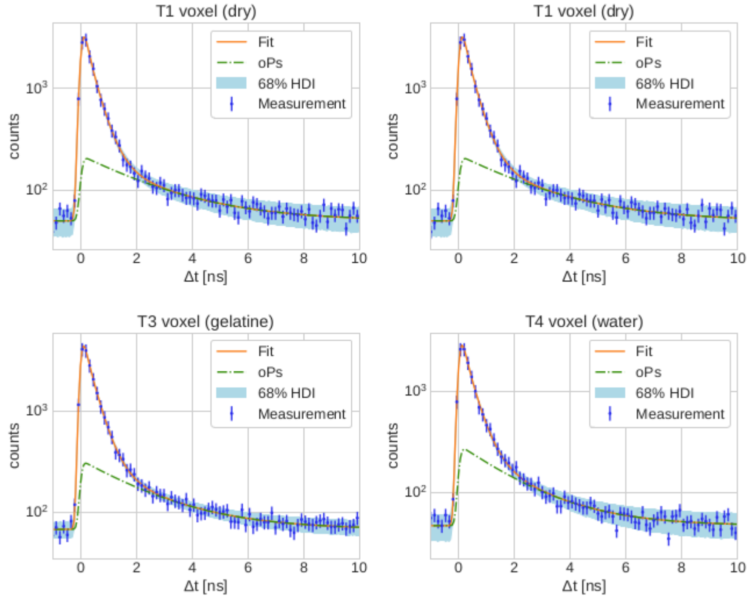


Fig. 4: Single-voxel TDD together with the fit prediction, posterior 68% HDI and oPs component in logarithmic scale ($4 \times 4 \times 4\,\mathrm{mm}^3$ voxel size). The fit results are reported in Tab. 2.

My example (NEMA IEC spheres)

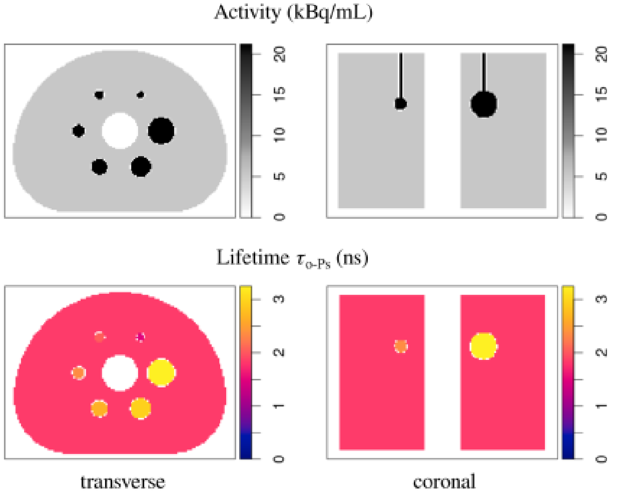
R package **'brms' / Stan** programming language (cf., Stan Development Team 2017b), provides a wide range of *non-linear distributional multilevel models*

$$f(\Delta t) = BG + \sum_{l \in \mathcal{L}} \frac{I_l}{2\tau_l} \exp\left[\frac{\mu - \Delta t}{\tau_l} + \frac{\sigma^2}{2\tau_l^2}\right] \cdot \mathrm{erfc}\left[\frac{\mu - \Delta t}{\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}\tau_l}\right]$$

$$\mathcal{L} \equiv \{\mathrm{dir}, \text{p-Ps}, \text{o-Ps}\}$$

$$\tau_{\text{p-Ps}} = 0.125 \ \text{ns}$$

$$\sigma \sim \mathcal{N}(0.15\,\text{ns},\ 0.05\,\text{ns}),$$
$$\mu \sim \mathcal{N}(0.0\,\text{ns},\ 0.5\,\text{ns}),$$
$$N_{BR} \sim \mathcal{N}(0.30,\ 0.7),$$
$$BR \sim \mathrm{Beta}(3,\ 2),$$
$$\tau_{\mathrm{dir}} \sim \mathcal{N}(0.5\,\text{ns},\ 0.1\,\text{ns}),$$
$$\tau_{\text{o-Ps}} \sim \mathcal{N}(2.0\,\text{ns},\ 0.8\,\text{ns}),$$

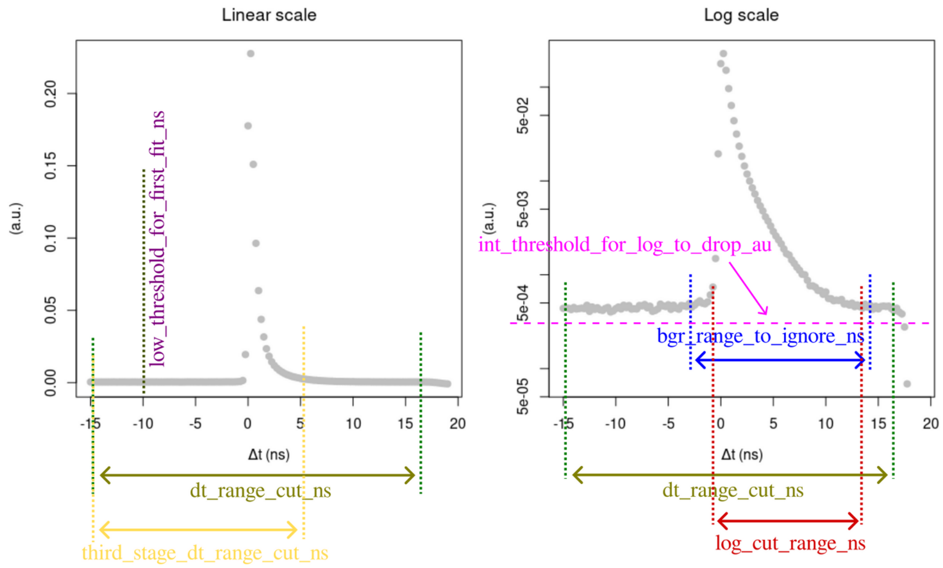


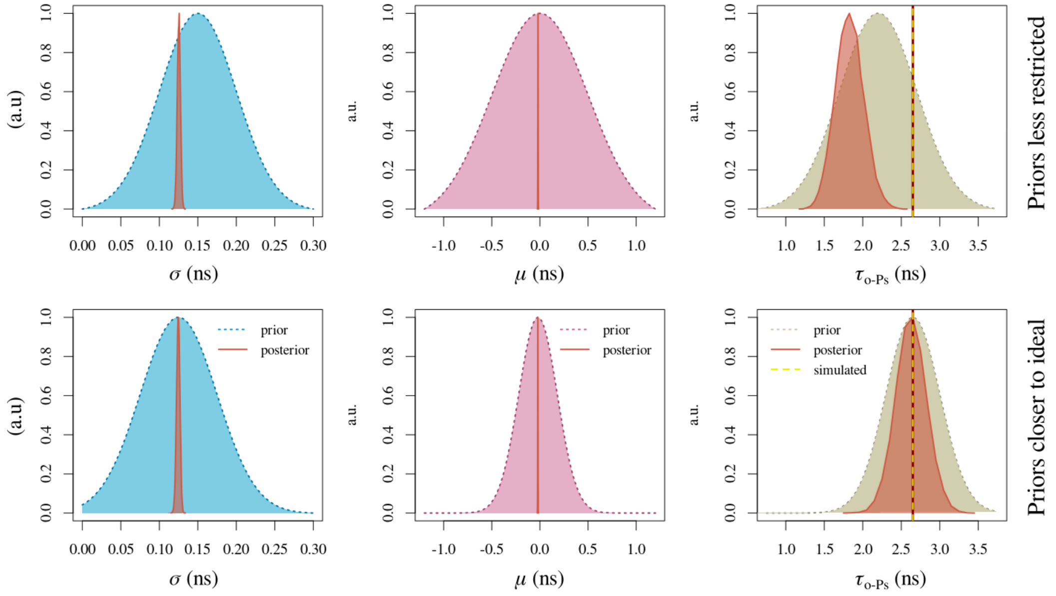
where N_{BR} is the normalisation constant and the branching ratio $BR = I_{\mathrm{dir}}/(I_{\mathrm{dir}} + I_{\text{p-Ps}} + I_{\text{o-Ps}})$, so that $I_{\mathrm{dir}} = N_{BR} \cdot BR$, $I_{\text{p-Ps}} = 0.25 \cdot N_{BR}(1 - BR)$ and $I_{\text{o-Ps}} = 0.75 \cdot N_{BR}(1 - BR)$ (see [28]). All normal distributions were truncated at zero, except for μ.

My example

$$f(\Delta t) = BG + \sum_{l\in\mathcal{L}} \frac{I_l}{2\tau_l} \exp\left[\frac{\mu-\Delta t}{\tau_l} + \frac{\sigma^2}{2\tau_l^2}\right] \cdot \mathrm{erfc}\left[\frac{\mu-\Delta t}{\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}\tau_l}\right]$$

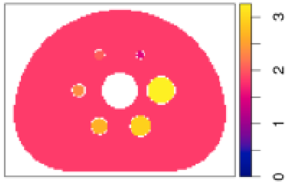
Strong dependence on priors and fitting range, in particular in log scale $f(\Delta t) \longrightarrow \log f(\Delta t)$



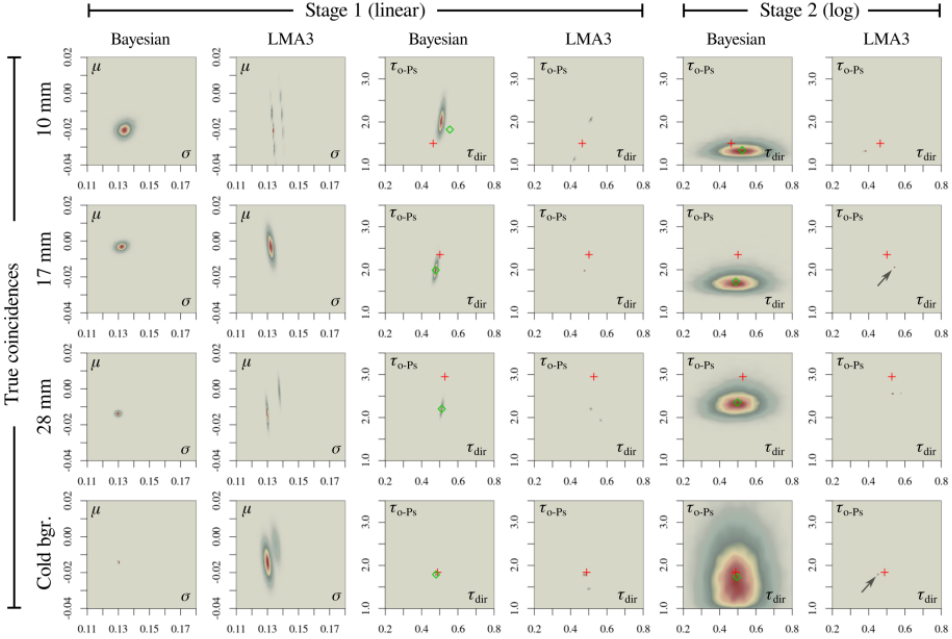


22-mm sphere

My example



$f(\Delta t)$ $\log f(\Delta t)$



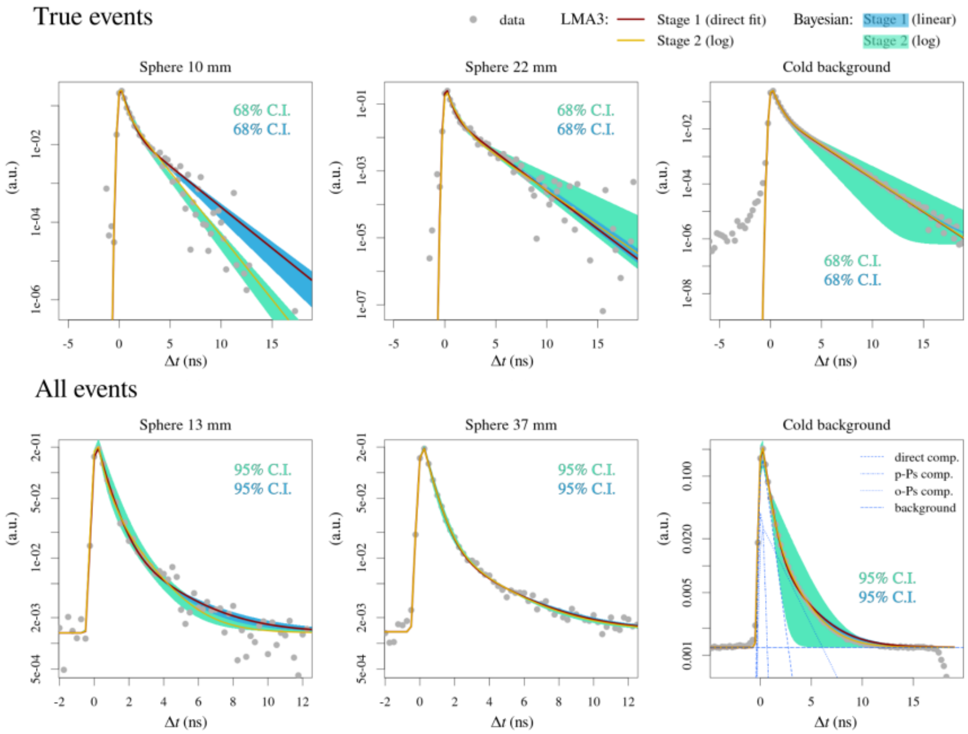


Figure 7: Time delay histograms, built from the MLEM weights (5th iteration) and fitted by various methods for the VOIs of the NEMA IEC phantom.

Thank You for Your attention!