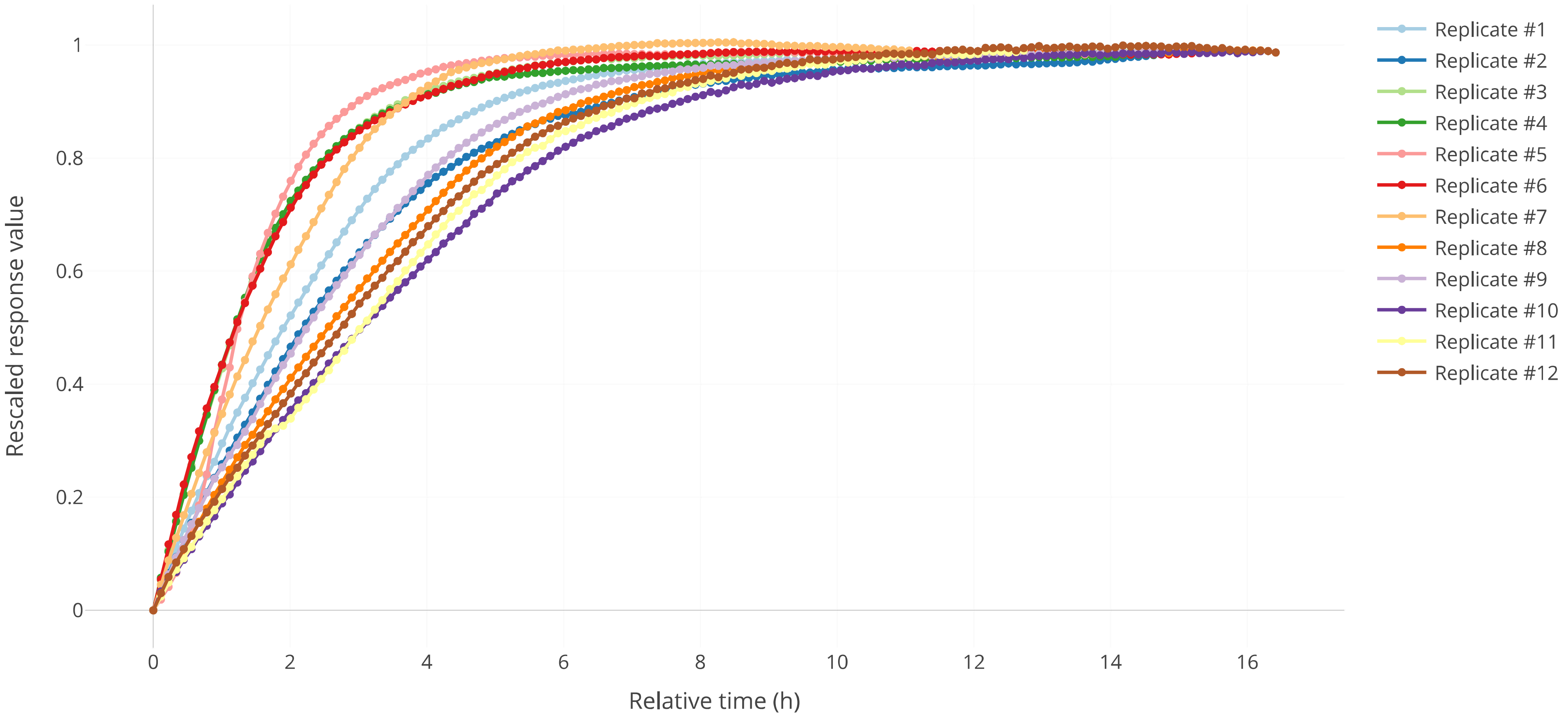
A generative modeling approach for trend outlier detection during reaction process transfer

NCS 2022, Louvain-la-Neuve

Joris Chau – Open Analytics
Jan Dijkmans – Janssen R&D
Tatsiana Khamiakova – Janssen R&D
Niels Vandervoort – Janssen R&D
Tor Maes – Janssen R&D

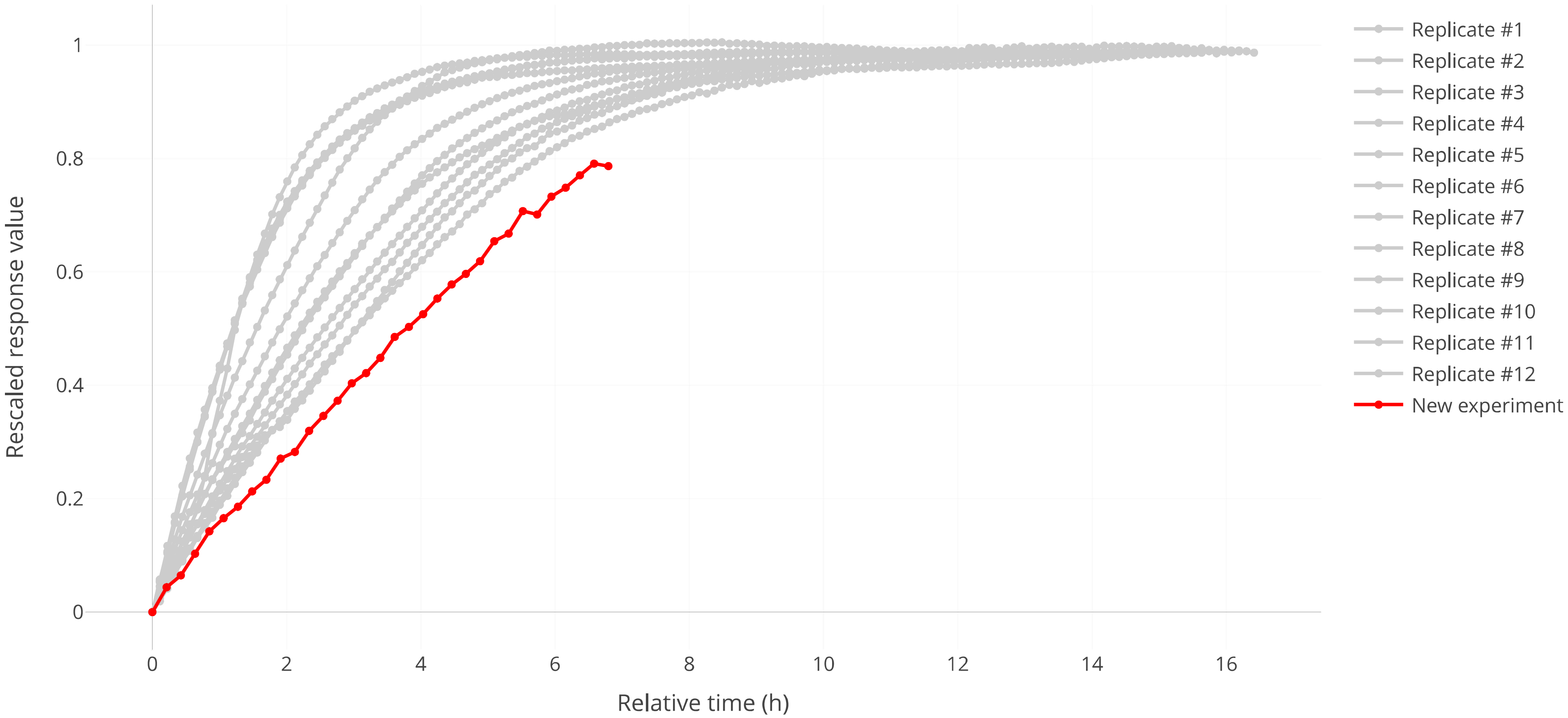
Example data

- **Reference reaction experiments**: IR spectral data
- **Trend data**: univariate (single wavelength) baseline–corrected peak heights over time



Example data

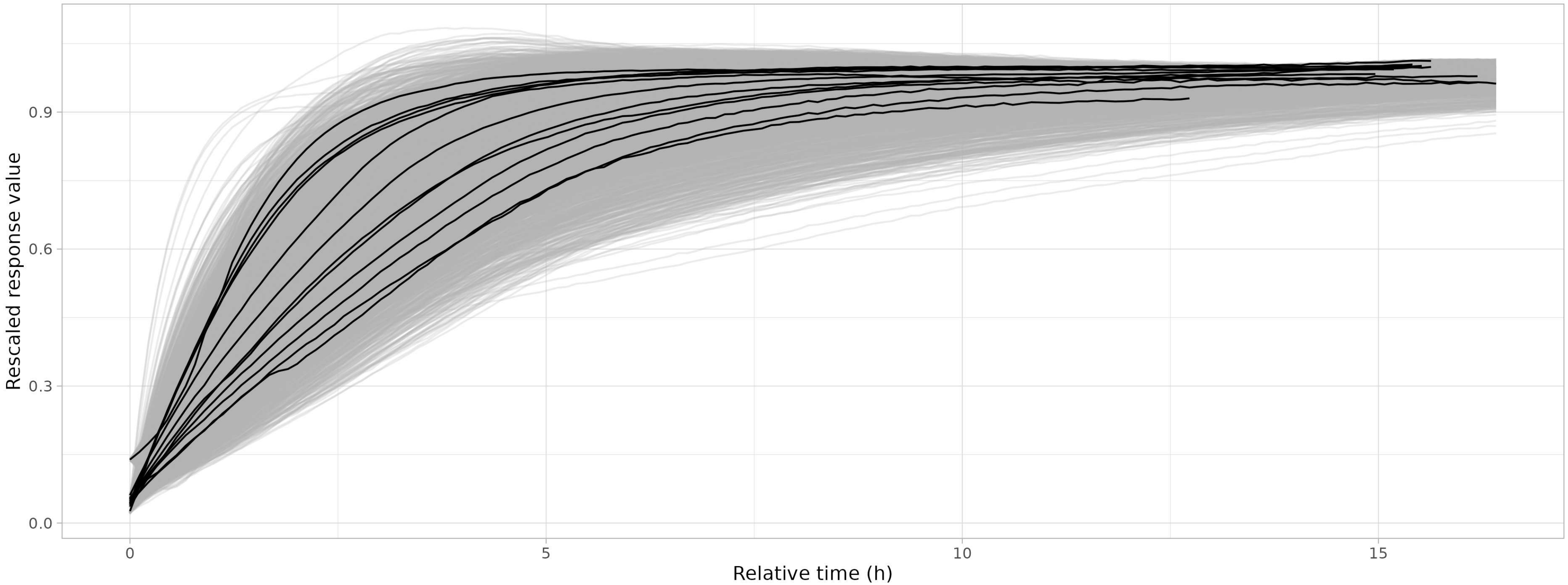
- **Reference reaction experiments**: IR spectral data
- **Trend data**: univariate (single wavelength) baseline-corrected peak heights over time



Step 1: Kinetic model fitting

Workflow

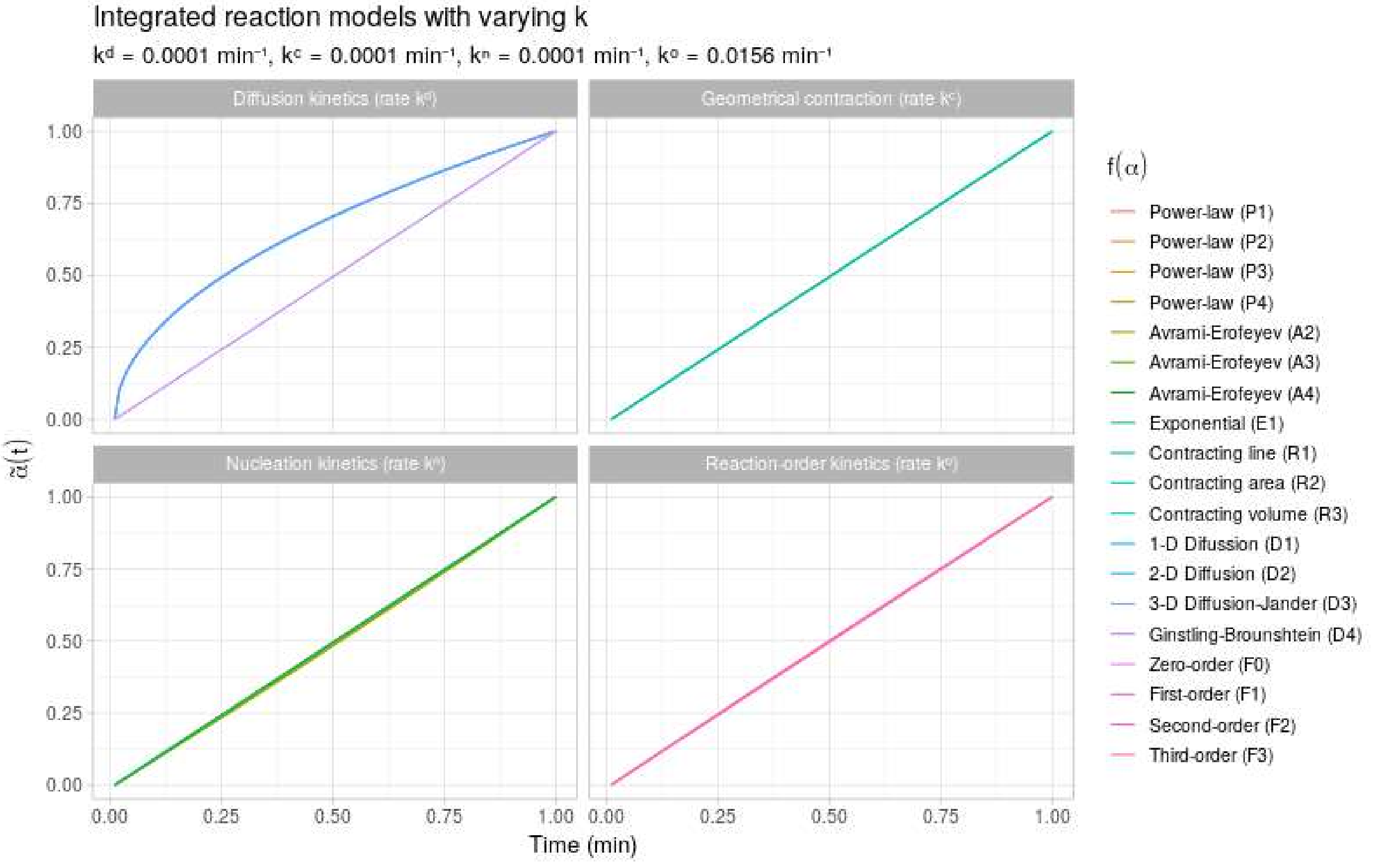
1. Select parametric kinetic model based on goodness-of-fit criteria for **individual** trend model fits;
2. (Optional) normalize trends based on estimated *intercept* and *asymptote* parameters of individual model fits;
3. Conditional on selected kinetic model, fit a single Bayesian hierarchical (random effects) model to all replicated trends;
4. Sample posterior predicted trends from fitted Bayesian model to form a distribution of functional trends.



Kinetic model fitting

Nonlinear random effects reaction models

$$\frac{d\alpha_s}{dt} = k_s \cdot f(\alpha_s), \qquad \log(k_s) \sim N(\log(k), \sigma_u^2 \boldsymbol{I})$$



Step 2: Functional outlier detection

Workflow

1. Evaluate functional band depths (e.g. Sun and Genton, 2011 JCGA) and epigraph indices for all sampled trends (with respect to the sampling distribution);

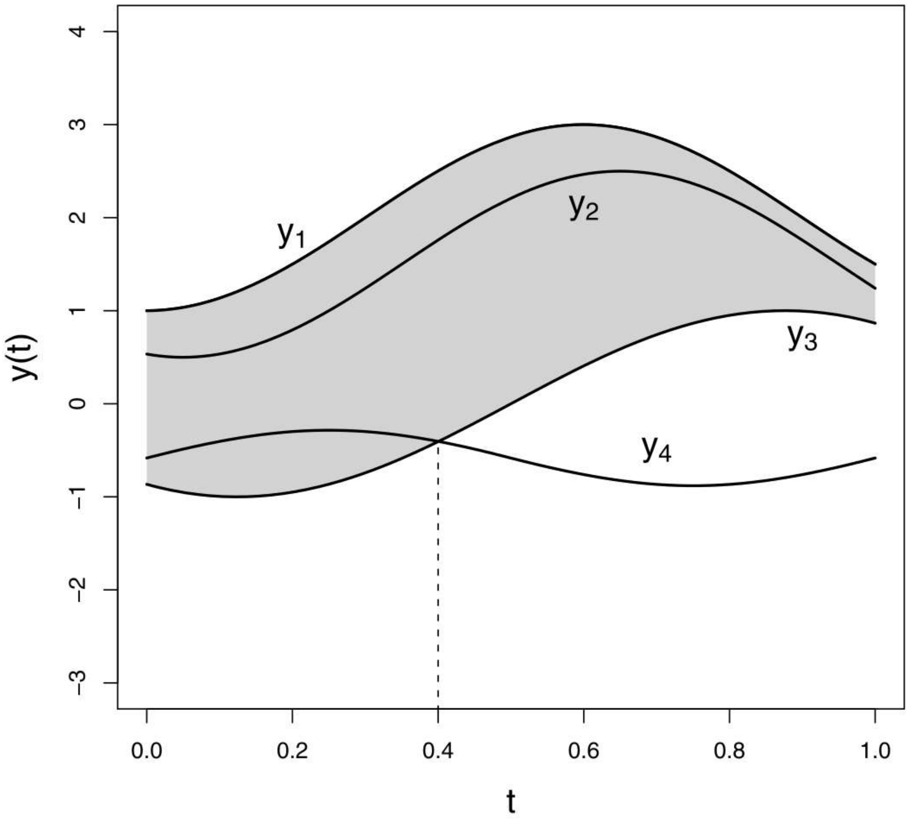


Figure 1 from [Sun and Genton, 2011 JCGA]

Step 2: Functional outlier detection

Workflow

1. Evaluate functional band depths (e.g. Sun and Genton, 2011 JCGA) and epigraph indices for all sampled trends (with respect to the sampling distribution);
2. Construct outliergram visualization (Arribas-Gil and Romo, 2014 Biostatistics) by plotting band depth against epigraph index.

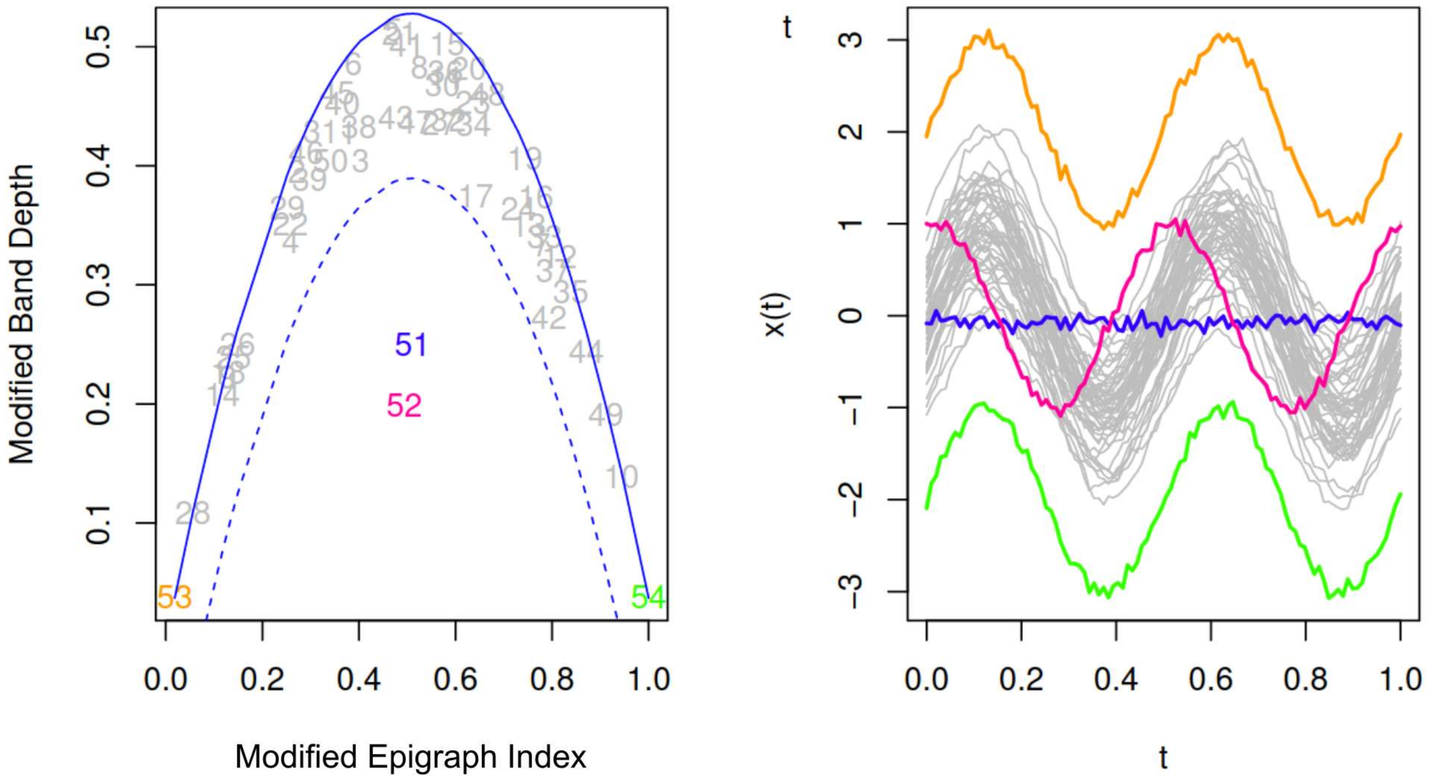
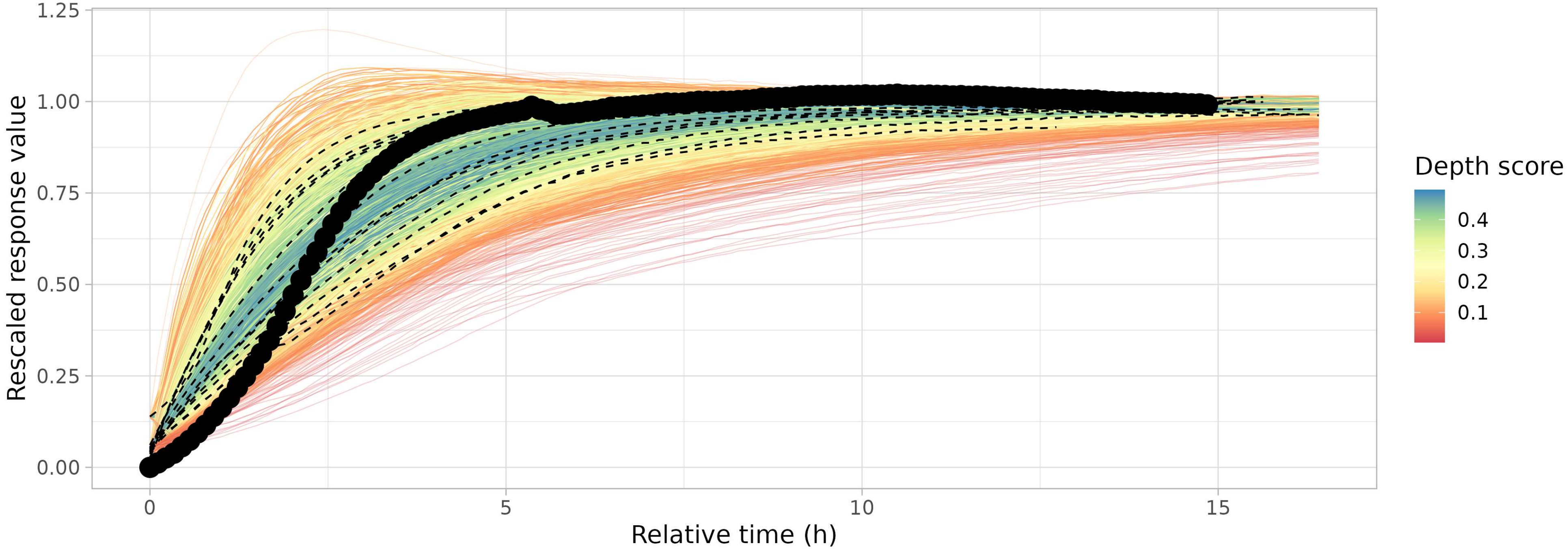


Figure 3 from [Arribas-Gil and Romo, 2014 Biostatistics]

Step 2: Functional outlier detection

Workflow

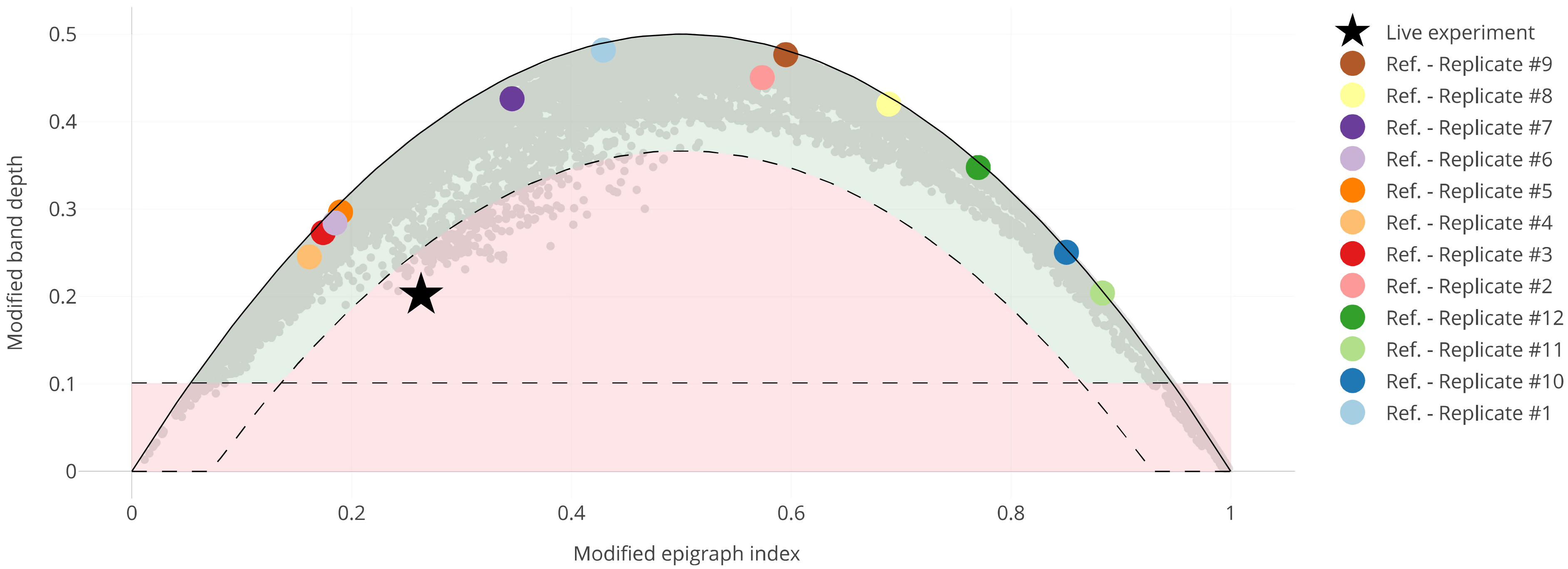
1. Evaluate functional band depths (e.g. Sun and Genton, 2011 JCGA) and epigraph indices for all sampled trends (with respect to the sampling distribution);
2. Construct outliergram visualization (Arribas-Gil and Romo, 2014 Biostatistics) by plotting band depth against epigraph index.



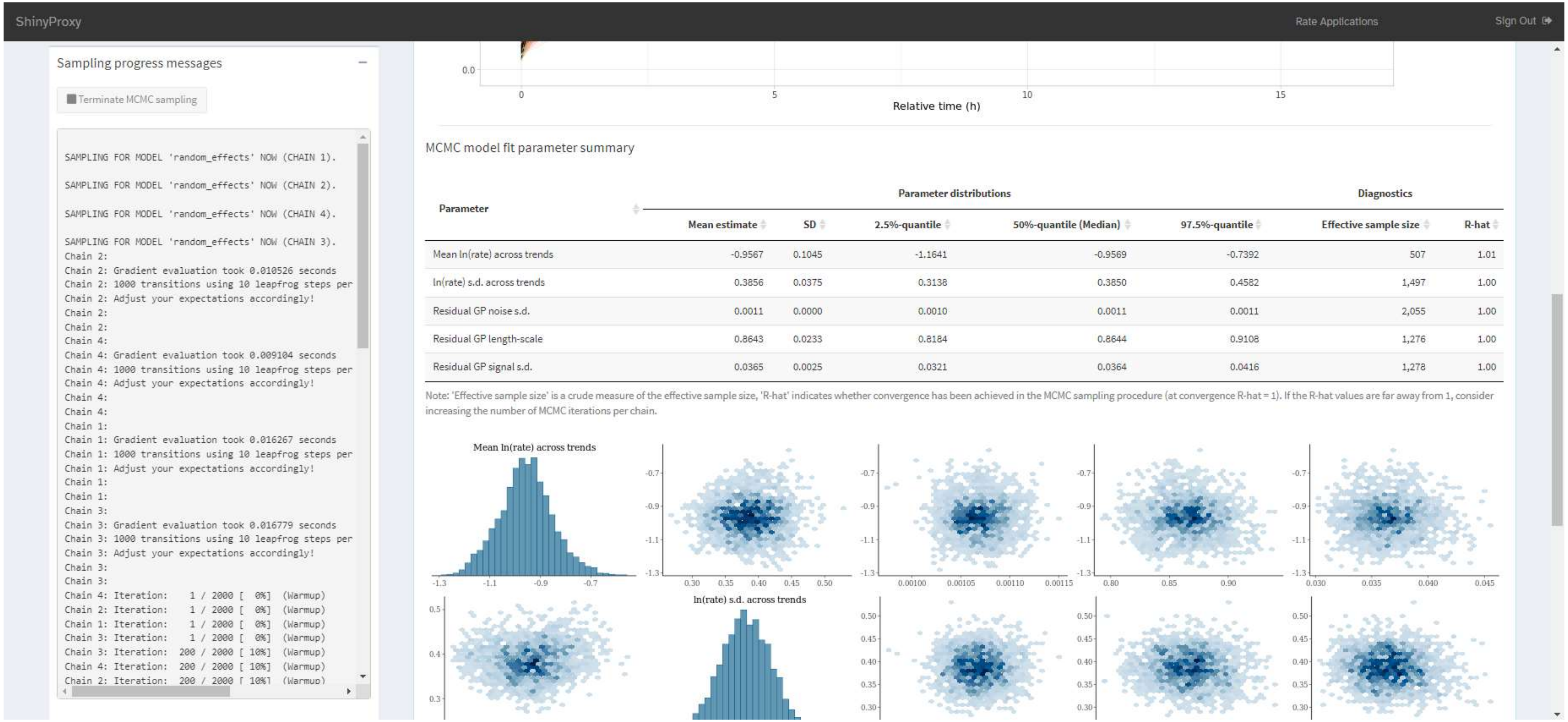
Step 2: Functional outlier detection

Workflow

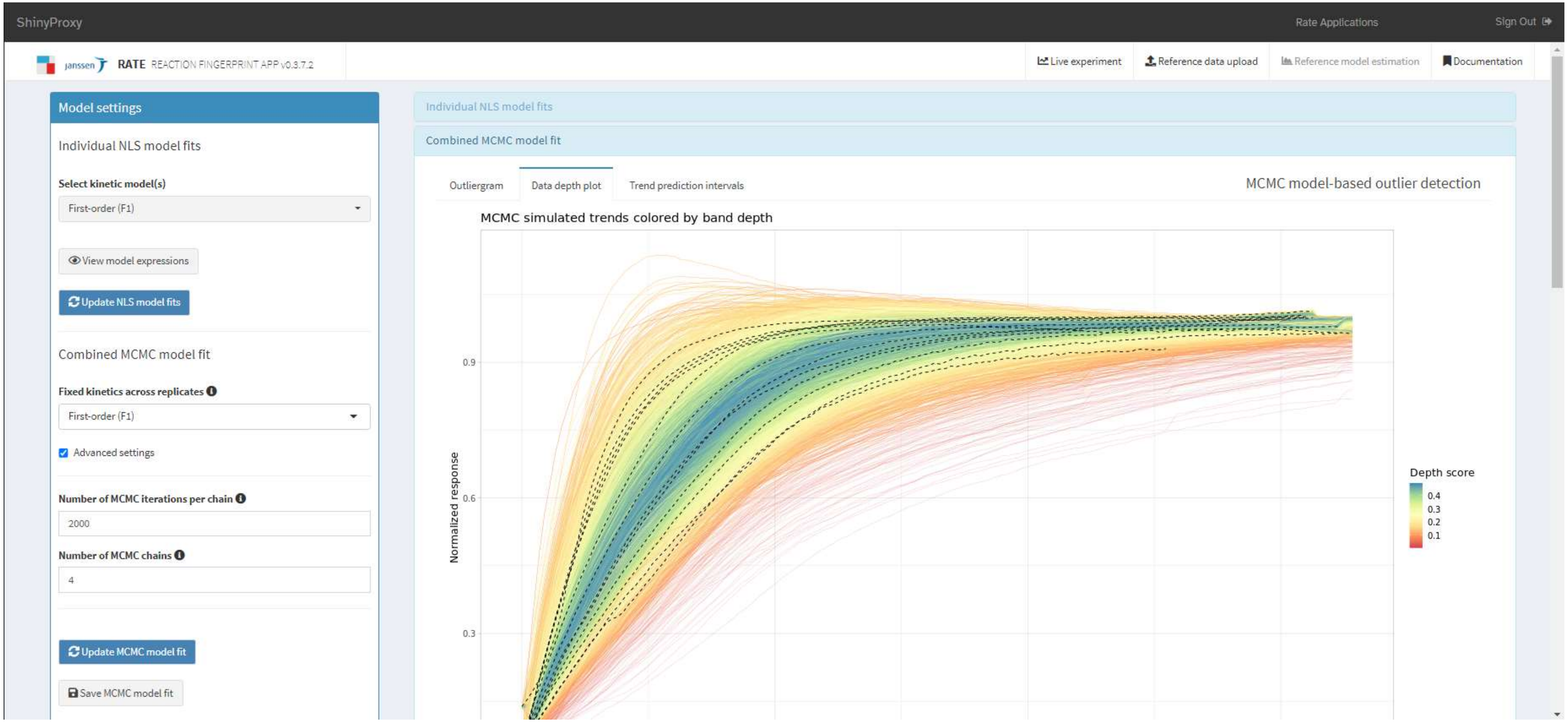
1. Evaluate functional band depths (e.g. Sun and Genton, 2011 JCGA) and epigraph indices for all sampled trends (with respect to the sampling distribution);
2. Construct outliergram visualization (Arribas-Gil and Romo, 2014 Biostatistics) by plotting band depth against epigraph index.



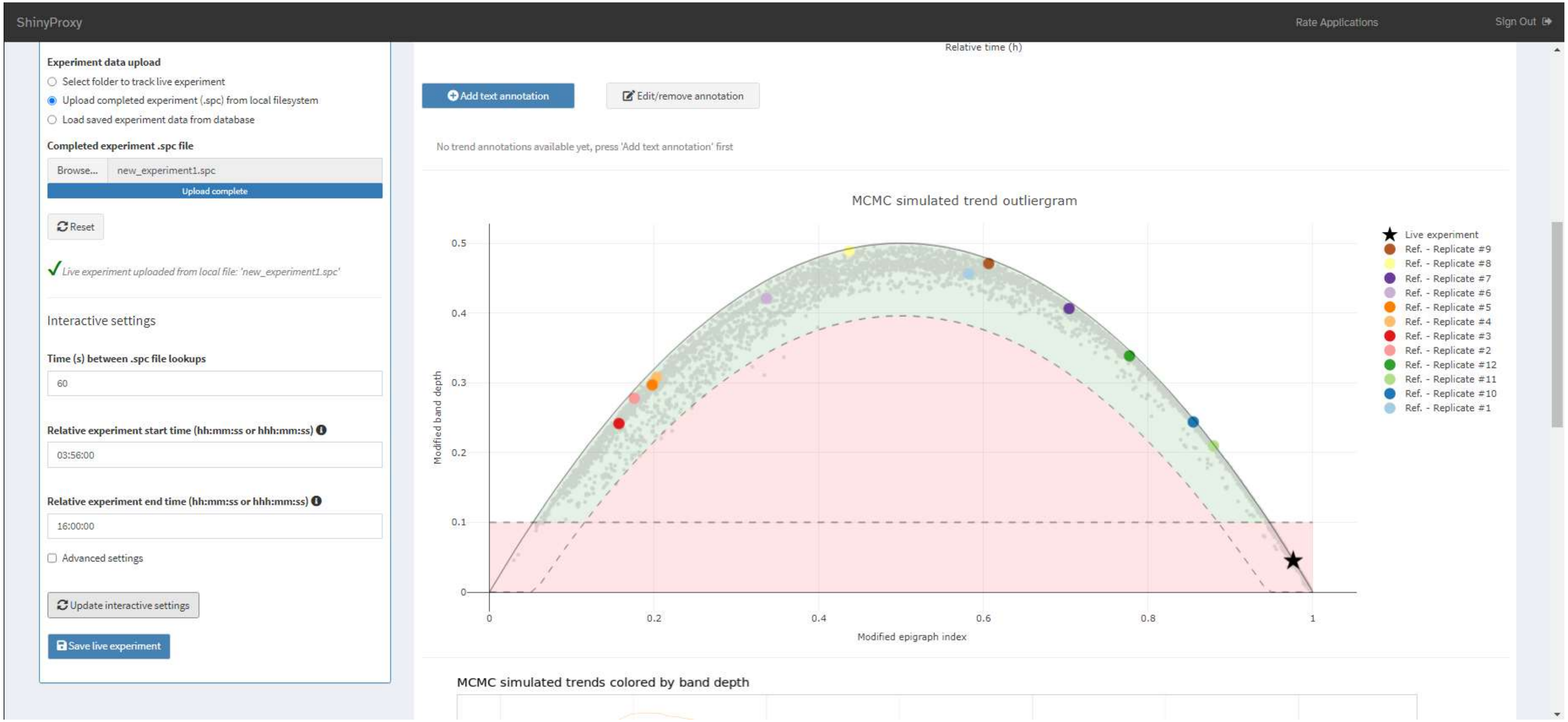
Live fingerprinting application (R-Shiny)



Live fingerprinting application (R-Shiny)



Live fingerprinting application (R-Shiny)



Useful references

- Sun, Y., and M. G. Genton. “Functional boxplots.” Journal of Computational and Graphical Statistics 20.2 (2011): 316–334.
- Arribas-Gil, A., and J. Romo. “Shape outlier detection and visualization for functional data: the outliergram.” Biostatistics 15.4 (2014): 603–619.