

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

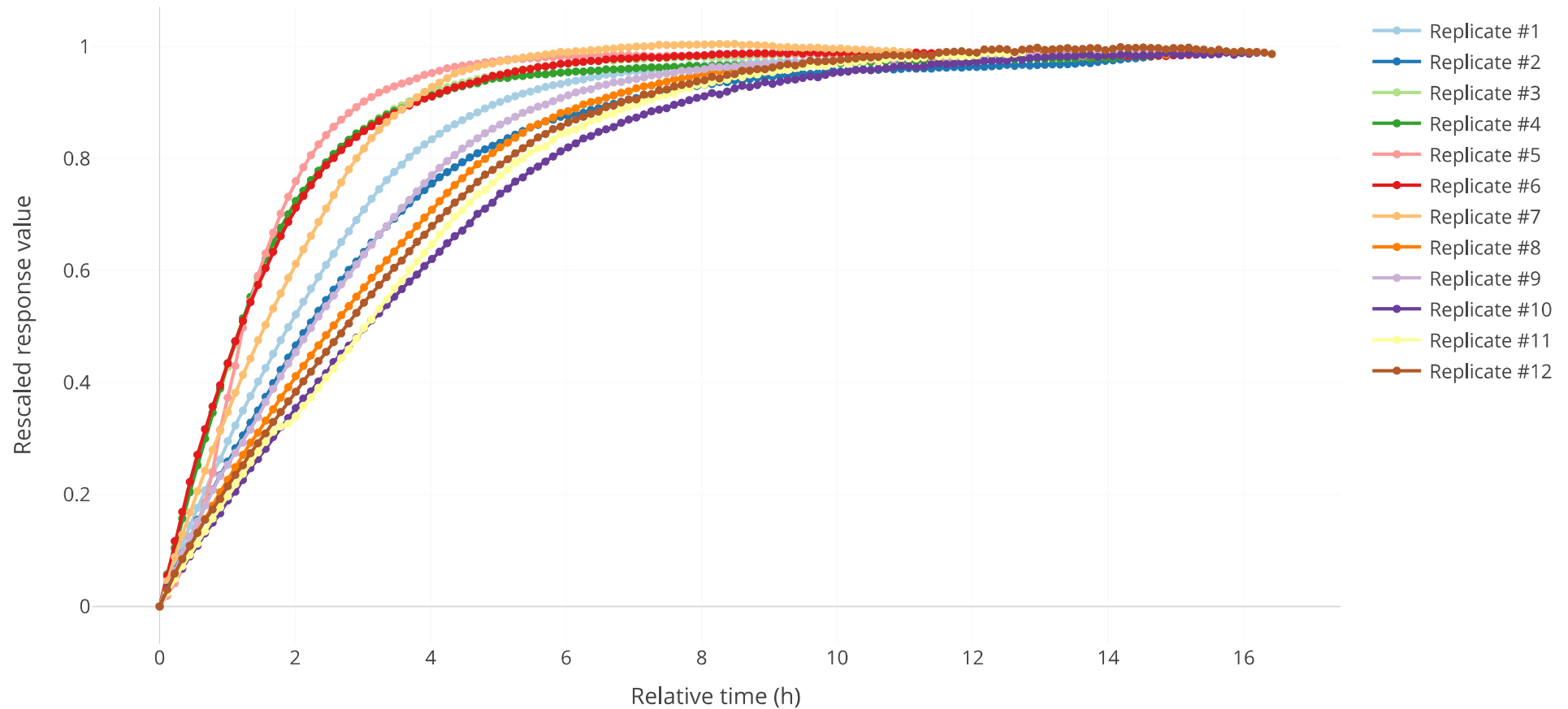
Tatsiana Khamiakova – Janssen R&D

Niels Vandervoort – Janssen R&D

Tor Maes – Janssen R&D

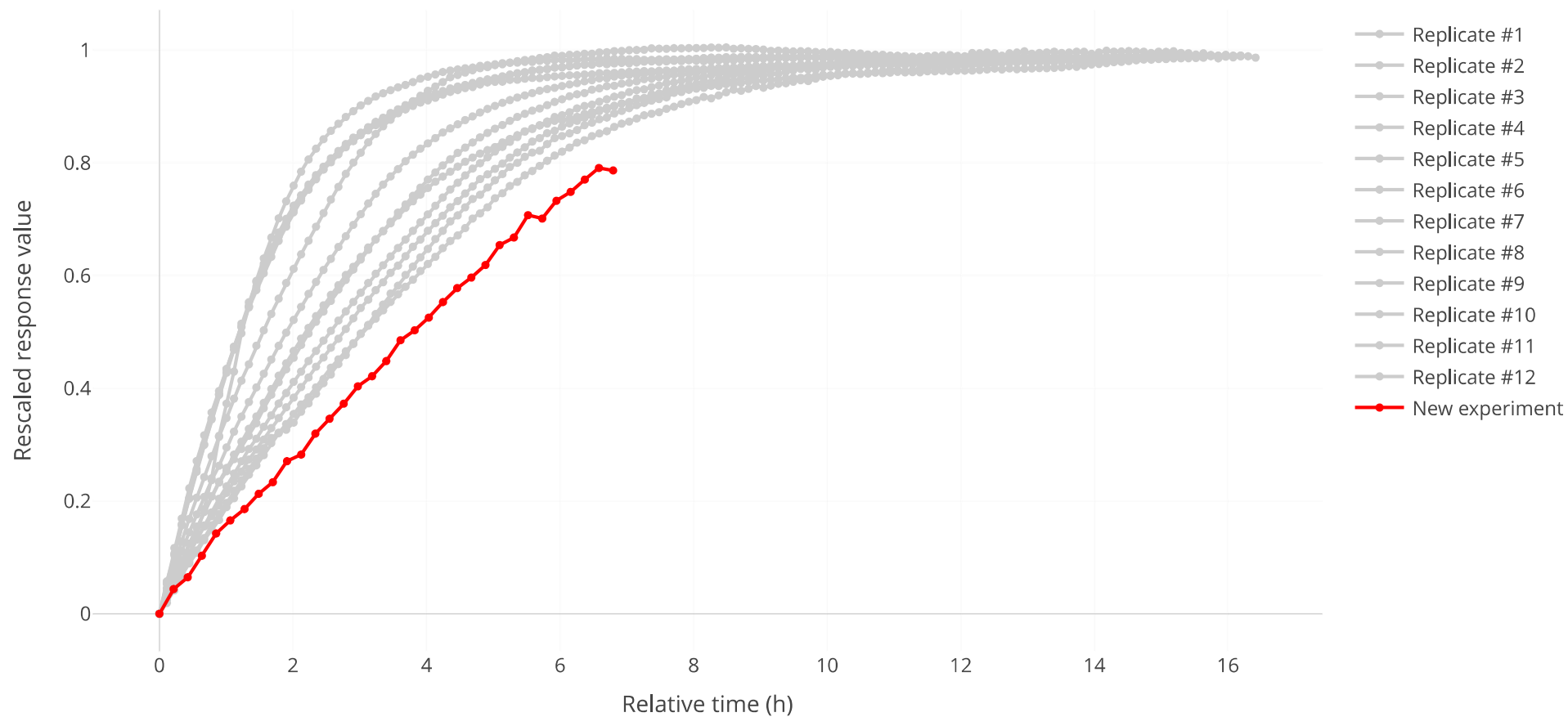
Example data

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



Example data

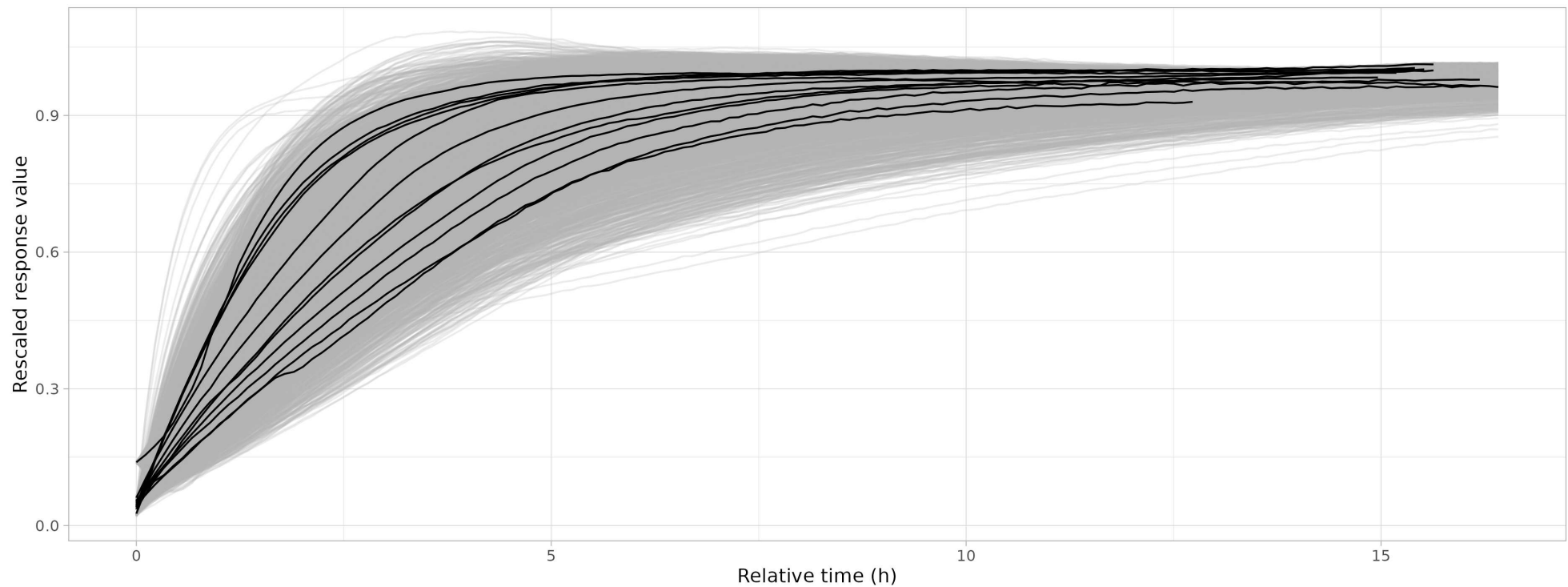
- Reference reaction experiments: IR spectral data
- Trend data: univariate (single wavelength) baseline-corrected peak heights over time



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.



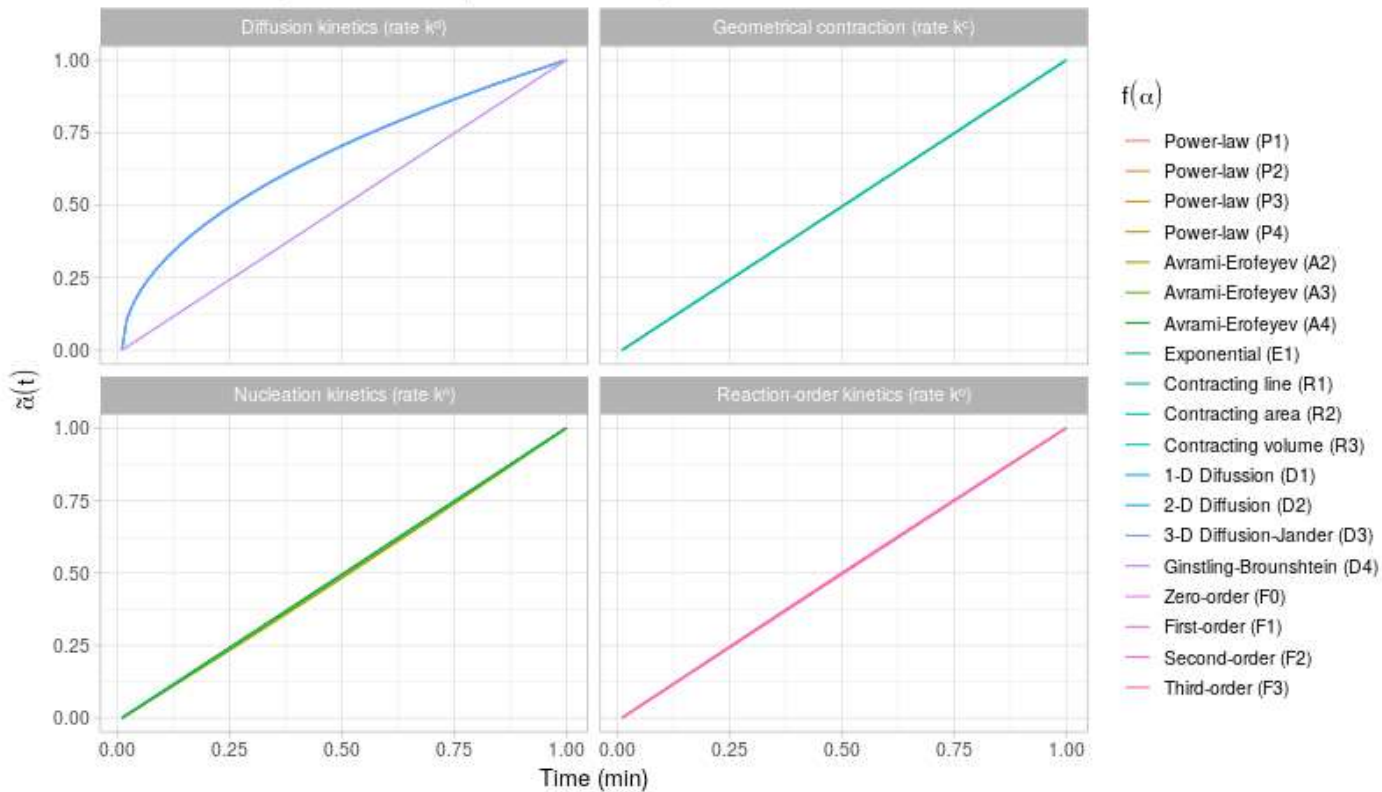
Kinetic model fitting

Nonlinear random effects reaction models

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

Integrated reaction models with varying k

$k^d = 0.0001 \text{ min}^{-1}$, $k^c = 0.0001 \text{ min}^{-1}$, $k^n = 0.0001 \text{ min}^{-1}$, $k^o = 0.0156 \text{ min}^{-1}$



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);

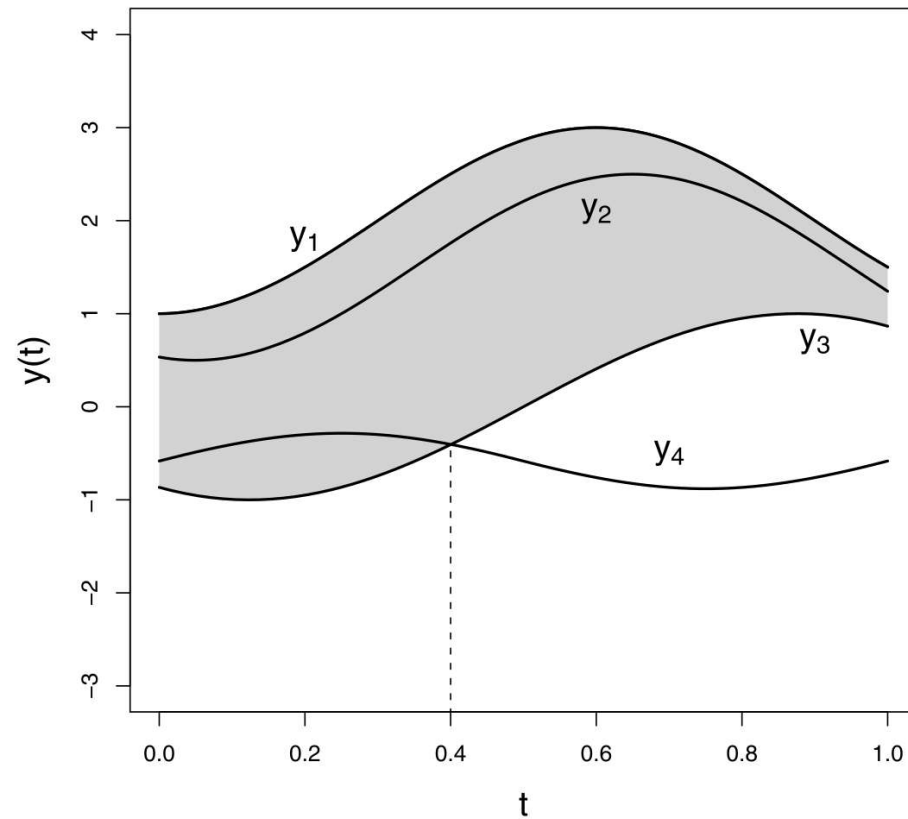


Figure 1 from [Sun and Genton, 2011 JCGA]

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.

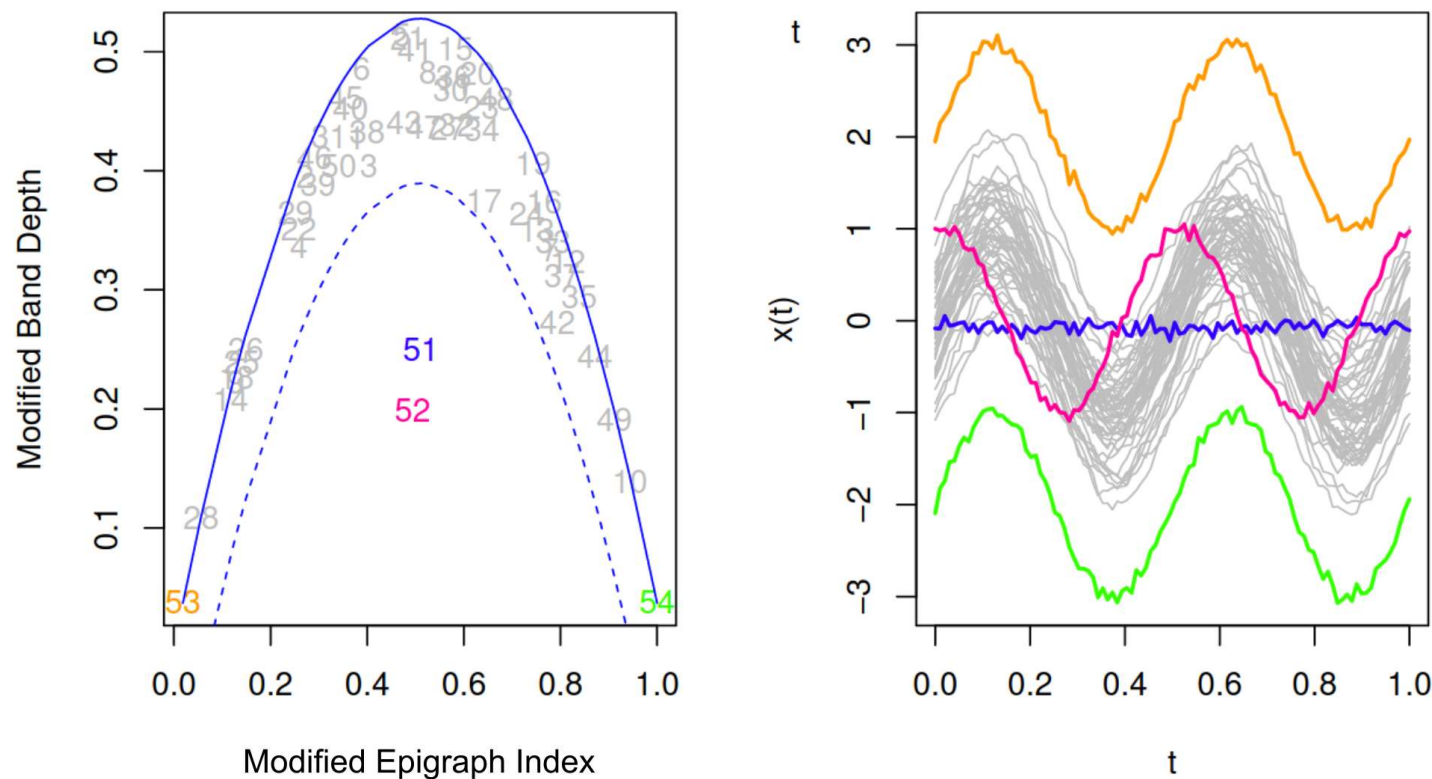
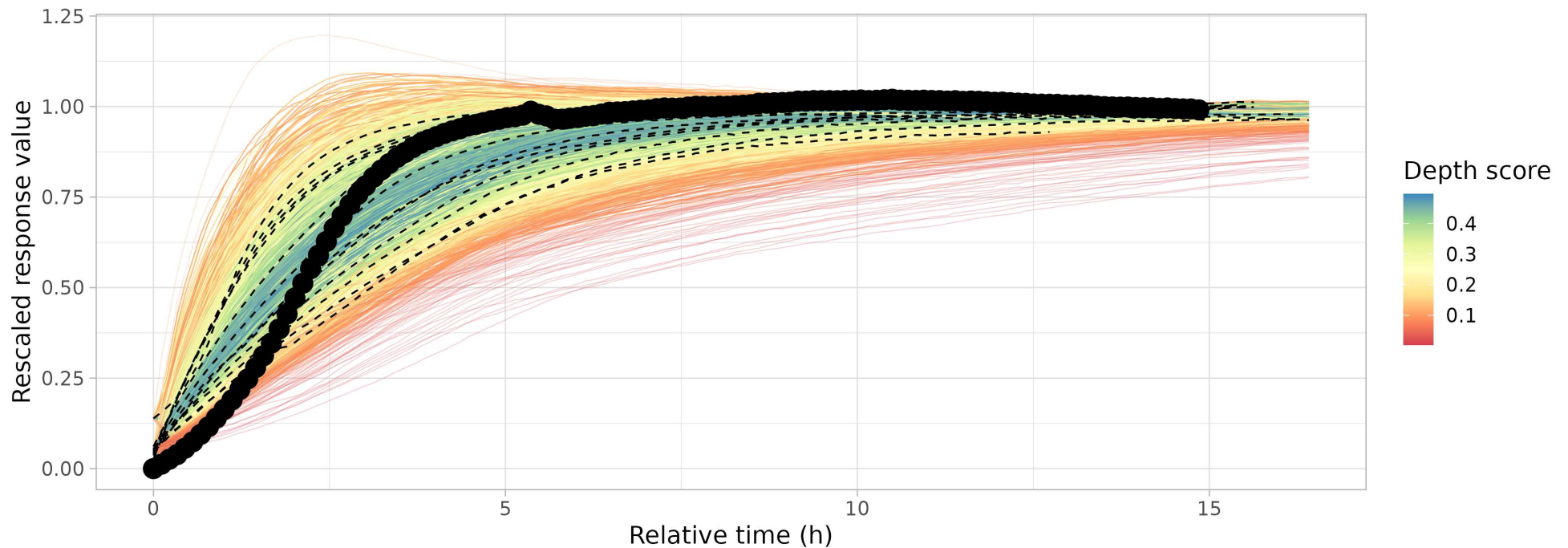


Figure 3 from [Arribas-Gil and Romo, 2014 Biostatistics]

Step 2: Functional outlier detection

Workflow

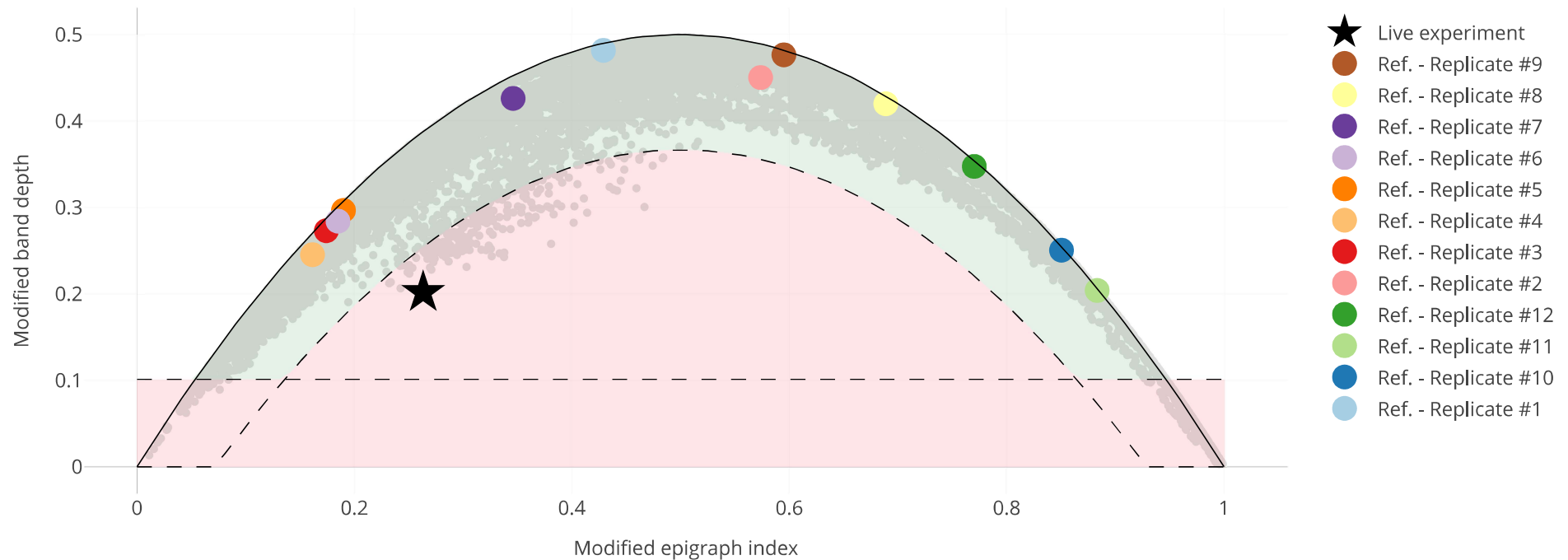
1. Evaluate functional band depths (e.g. [Sun and Genton, 2011 JCGA](#)) and epigraph indices for all sampled trends (with respect to the sampling distribution);
2. Construct outliergram visualization ([Arribas-Gil and Romo, 2014 Biostatistics](#)) by plotting band depth against epigraph index.



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)


ShinyProxy
Rate Applications
Sign Out

Sampling progress messages

Terminate MCMC sampling

```

SAMPLING FOR MODEL 'random_effects' NOW (CHAIN 1).
SAMPLING FOR MODEL 'random_effects' NOW (CHAIN 2).
SAMPLING FOR MODEL 'random_effects' NOW (CHAIN 4).
SAMPLING FOR MODEL 'random_effects' NOW (CHAIN 3).
Chain 2:
Chain 2: Gradient evaluation took 0.010526 seconds
Chain 2: 1000 transitions using 10 leapfrog steps per chain
Chain 2: Adjust your expectations accordingly!
Chain 2:
Chain 4:
Chain 4: Gradient evaluation took 0.009104 seconds
Chain 4: 1000 transitions using 10 leapfrog steps per chain
Chain 4: Adjust your expectations accordingly!
Chain 4:
Chain 1:
Chain 1: Gradient evaluation took 0.016267 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per chain
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 3:
Chain 3: Gradient evaluation took 0.016779 seconds
Chain 3: 1000 transitions using 10 leapfrog steps per chain
Chain 3: Adjust your expectations accordingly!
Chain 3:
Chain 3:
Chain 4: Iteration: 1 / 2000 [ 0%] (Warmup)
Chain 2: Iteration: 1 / 2000 [ 0%] (Warmup)
Chain 1: Iteration: 1 / 2000 [ 0%] (Warmup)
Chain 3: Iteration: 1 / 2000 [ 0%] (Warmup)
Chain 3: Iteration: 200 / 2000 [ 10%] (Warmup)
Chain 4: Iteration: 200 / 2000 [ 10%] (Warmup)
Chain 2: Iteration: 200 / 2000 [ 10%] (Warmup)
                    
```

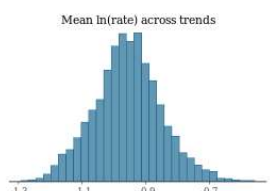


MCMC model fit parameter summary

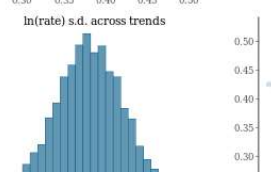
Parameter	Parameter distributions					Diagnostics	
	Mean estimate	SD	2.5%-quantile	50%-quantile (Median)	97.5%-quantile	Effective sample size	R-hat
Mean ln(rate) across trends	-0.9567	0.1045	-1.1641	-0.9569	-0.7392	507	1.01
ln(rate) s.d. across trends	0.3856	0.0375	0.3138	0.3850	0.4582	1,497	1.00
Residual GP noise s.d.	0.0011	0.0000	0.0010	0.0011	0.0011	2,055	1.00
Residual GP length-scale	0.8643	0.0233	0.8184	0.8644	0.9108	1,276	1.00
Residual GP signal s.d.	0.0365	0.0025	0.0321	0.0364	0.0416	1,278	1.00

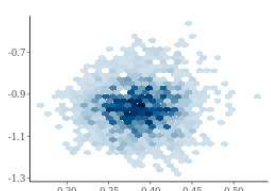
Note: 'Effective sample size' is a crude measure of the effective sample size, 'R-hat' indicates whether convergence has been achieved in the MCMC sampling procedure (at convergence R-hat = 1). If the R-hat values are far away from 1, consider increasing the number of MCMC iterations per chain.

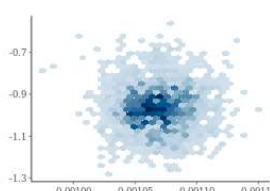
Mean ln(rate) across trends

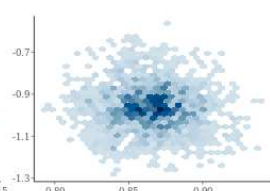


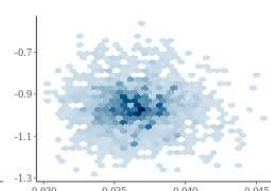
ln(rate) s.d. across trends

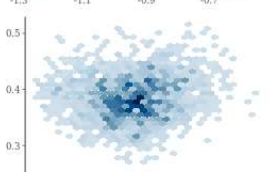


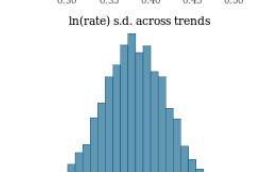


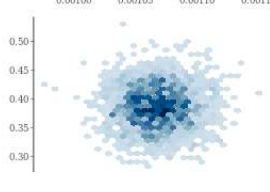


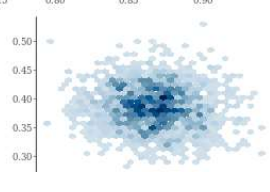


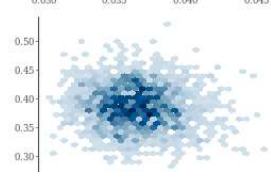




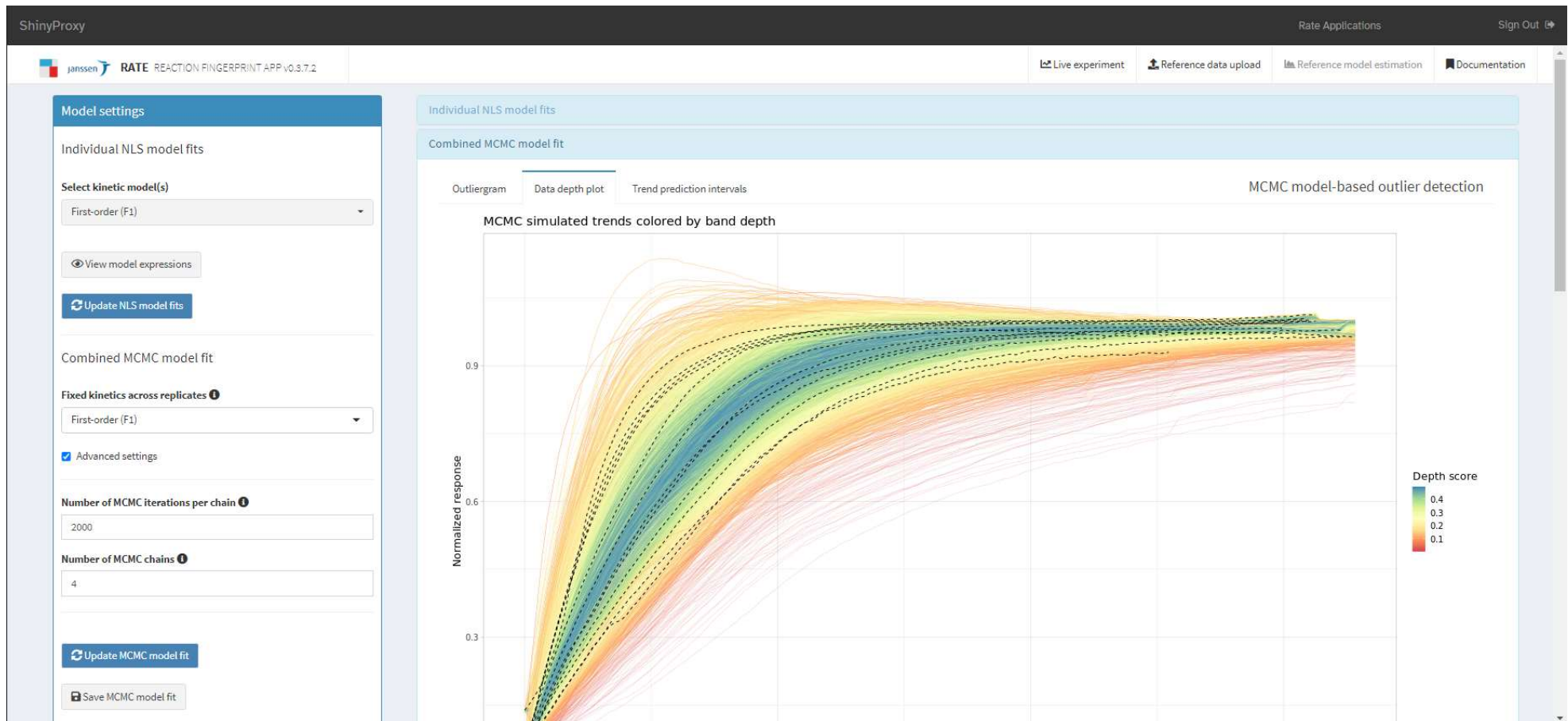








Live fingerprinting application (R-Shiny)



Live fingerprinting application (R-Shiny)

ShinyProxy Rate Applications [Sign Out](#)

Experiment data upload

Select folder to track live experiment

Upload completed experiment (.spc) from local filesystem

Load saved experiment data from database

Completed experiment .spc file

Browse... new_experiment1.spc

Upload complete

Reset

✓ Live experiment uploaded from local file: 'new_experiment1.spc'

Interactive settings

Time (s) between .spc file lookups

60

Relative experiment start time (hh:mm:ss or hhh:mm:ss)

03:56:00

Relative experiment end time (hh:mm:ss or hhh:mm:ss)

16:00:00

Advanced settings

Update interactive settings

Save live experiment

Relative time (h)

+ Add text annotation Edit/remove annotation

No trend annotations available yet, press 'Add text annotation' first

MCMC simulated trend outliergram

MCMC simulated trends colored by band depth

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.