blotIt2: Optimal Alignment
of Biological Replicate Data by Latent Variable Models

Daniel Kaschek, Physikalisches Institut Freiburg

Abteilungsseminar, 9.5.2017

Albert-Ludwigs-Universität Freiburg

UNI
FREIBURG

Introduction

What is dynamics?

> **Definition (Dynamics)**
>
> Let $x : t \mapsto x(t)$ and $y : t \mapsto y(t)$ be two curves.
> Then x and y share the same dynamics if there exists $s \in \mathbb{R}$ such that
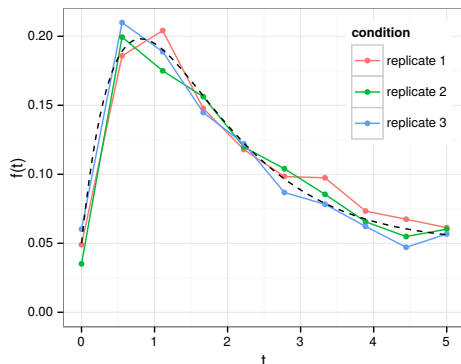>
> $$y(t) = sx(t)$$

Estimation of the dynamics:

- time points t_i

- replicates j

- data S_{ij}

$$\text{RSS} = \sum_{ij} \left(S_{ij} - \frac{y_i}{s_j} \right)^2$$

$\rightarrow$ dynamics y_i (pars of interest)

$\rightarrow$ scales s_j (latent pars)

Latent Variable Modeling

The objective function

Example: simulated data

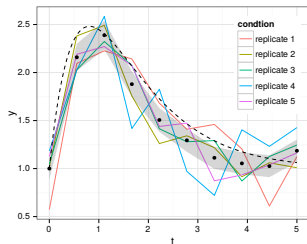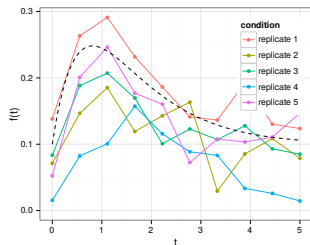$$\text{RSS} = \sum_{ij} \left(S_{ij} - \frac{y_i}{s_j} \right)^2$$

Scale invariance:

- $(\hat{y}, \hat{s})$ minimizes RSS

- transformation $(\hat{y}', \hat{s}') = (\alpha \hat{y}, \alpha \hat{s})$
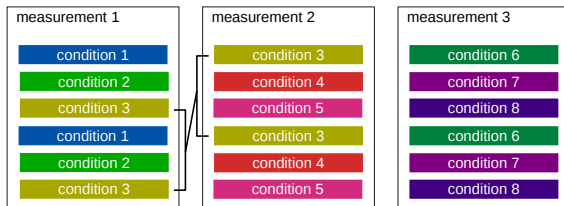
$\Rightarrow \text{RSS} = \text{RSS}'$

Fold change scale:

- choose $\alpha = {}^1/\hat{y}_1$

$\rightarrow$ $\left(\frac{\hat{y}_1}{\hat{y}_1}, \ldots, \frac{\hat{y}_n}{\hat{y}_1} \right)$ is the dynamics estimate

Experimental view

Matrix representation:

- Rows correspond to **conditions**
- Columns correspond to **measurements**
- Block-diagonal form obtained by **pertumation algorithm**

$\rightarrow$ Define common scale for list of conditions

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Key assumptions for blotIt2

1. Biological conditions are different → fixed effects parameters
   - Target & Timepoint
   - Treatment
   - Stimulation

2. Measurement conditions are different → latent effects parameters
   - Gel-to-gel differences
   - Loading or signal amplification

3. Biological and measurement noise → error model parameters

Model equations

Alignment model

$$S_{ij} = f(y_i, s_j) + \varepsilon_{ij} = \frac{y_i}{s_j} + \varepsilon_{ij}$$

- Model response $S_{ij} \sim$ fixed effects y_i and latent effects s_j
- Noise $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ and σ_{ij} governed by error model

Error model

$$\sigma_{ij} = h(y_i, s_j, \theta_k) = \frac{\sigma_{\mathrm{rel}} \cdot y_i + \sigma_{\mathrm{abs}}}{s_j}$$

- Error model $\sim$ fixed/latent effects and error model parameters

- $i \sim$ (target, time-point, treatment, stimulus, etc.)
- $j \sim$ (gel, slide, sample, measurement)
- $k \sim$ (global or gel, slide, sample)

Parameter estimation

Log-likelihood function

$$-2\log L\big(\{y_i\},\{s_j\},\{\theta_k\}\big) = \sum_{ij}\left(\frac{f(y_i,s_j) - S_{ij}}{h(y_i,s_j,\theta_k)}\right)^2 + \quad \text{(least squares)}$$

$$+ \sum_{ij}\log h(y_i,s_j,\{\theta_k\})^2 + \quad \text{(variance parameters)}$$

$$+ \left(\frac{1 - \sum_i y_i}{10^{-3}}\right)^2 \qquad \text{(normalization constraint)}$$

- ▶ **Trust region** optimization for parameters y_i, s_j, $\theta_k \rightarrow$ value, gradient and $J^{\mathrm{T}}J$
- ▶ **Symbolic derivatives** computed expressions for f and h
- ▶ **Independent estimation** for each unconnected target $\rightarrow$ can be parallelized

Uncertainty estimation and error propagation

Computation of uncertainties:

- **Generalized** inverse of the hessian (FIM)
  ↳ remove impact of non-identifiable parameters

- Bessel factor $\gamma = \sqrt{\frac{\#\text{data}}{\#\text{data} - \#\text{pars}}}$
  ↳ adjust parameter uncertainty: $\sigma_{y_i}^{(\gamma)} = \gamma\sigma_{y_i}$, $\sigma_{s_j}^{(\gamma)} = \gamma\sigma_{s_j}$
  ↳ adjust error model prediction: $\sigma_{ij}^{(\gamma)} = \gamma h(y_i, s_j, \theta_k) \neq h(y_i^{(\gamma)}, s_j^{(\gamma)}, \theta_k^{(\gamma)})$

Outputs generated:

	data	$\sigma(\text{data})$	prediction	$\sigma(\text{prediction})$
original scale	S_{ij}	$\sigma_{ij}^{(\gamma)}$ (*)	$f(y_i, s_j)$	$\sigma_{ij}^{(\gamma)}$
common scale	$f_j^{-1}(S_{ij})$	$\left\|(f_j^{-1})'(S_{ij})\right\|\sigma_{ij}^{(\gamma)}$ (**)	y_i	$\sigma_{y_i}^{(\gamma)}$ (***)

(*) Estimated measurement uncertainty $\hat{=}$ standard deviation

(**) Propagated measurement uncertainty $\hat{=}$ propagated standard deviation

(***) Uncertainty of the fixed effects parameters $\hat{=}$ standard error of the mean

Applications

Activation of the IκB kinase (IKK)

```
out0 <- alignME(
  data = data,
  model = "yi/sj",
  errmodel = "sigmaR*yi/sj",
  fixed = yi~stimulus+compound+dose,
  latent = sj~experiment,
  error = sigmaR~1,
  log = TRUE
)

plot1(out0)
```

- ▶ Show prediction of the alignment model on original scale
- ▶ Error model prediction shown as ribbon around prediction → estimate for data uncertainty
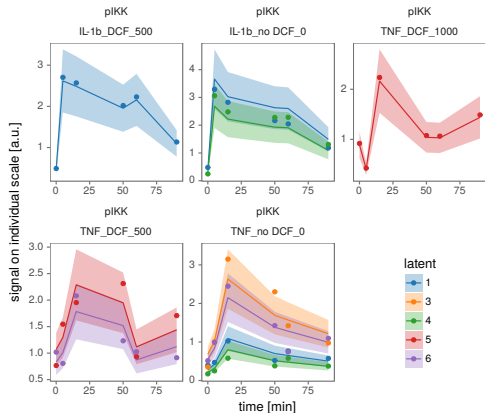
Activation of the IκB kinase (IKK)

```
out0 <- alignME(
  data = data,
  model = "yi/sj",
  errmodel = "sigmaR*yi/sj",
  fixed = yi~stimulus+compound+dose,
  latent = sj~experiment,
  error = sigmaR~1,
  log = TRUE
)

plot2(out0)
```

- ▶ Scaled datapoints plotted on the same scale
- ▶ Estimated data uncertainty translated to common scale (error bars)
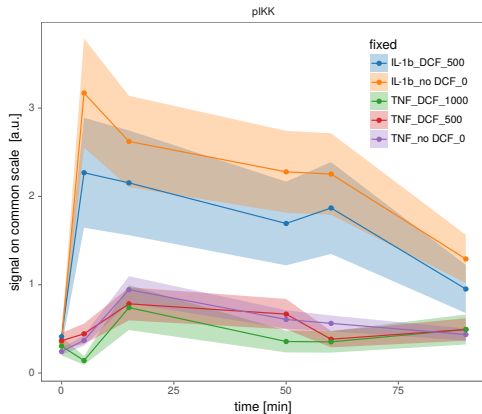- ▶ $y_i \pm \sigma_{y_i}$, result from parameter estimation, in gray with ribbon
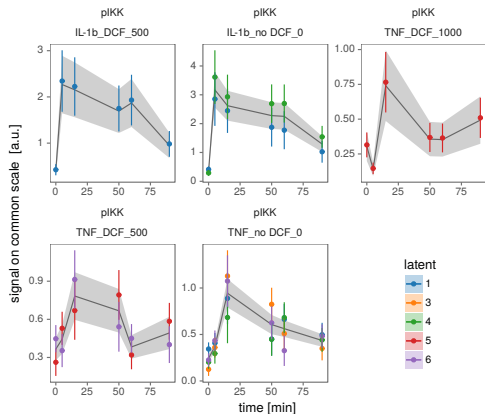
Activation of the IκB kinase (IKK)

```
out0 <- alignME(
  data = data,
  model = "yi/sj",
  errmodel = "sigmaR*yi/sj",
  fixed = yi~stimulus+compound+dose,
  latent = sj~experiment,
  error = sigmaR~1,
  log = TRUE
)


plot3(out0)
```

- $y_i \pm \sigma_{y_i}$, result from parameter estimation
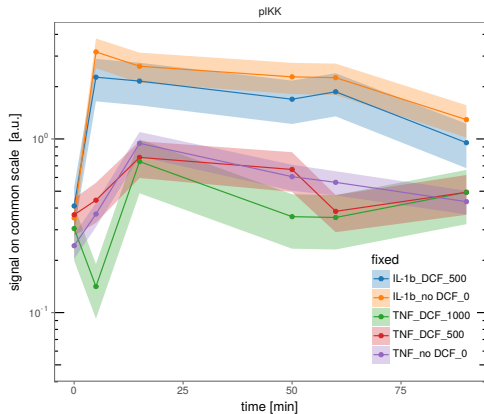- Comparison of different fixed effects in one plot

Activation of the IκB kinase (IKK)

```
out0 <- alignME(
  data = data,
  model = "yi/sj",
  errmodel = "sigmaR*yi/sj",
  fixed = yi~stimulus+compound+dose,
  latent = sj~experiment,
  error = sigmaR~1,
  log = TRUE
)


plot3(out0)
```

- $y_i \pm \sigma_{y_i}$, result from parameter estimation
- Comparison of different fixed effects in one plot (log-scale)

pIKK

signal on common scale [a.u.]

time [min]

fixed
- IL-1b_DCF_500
- IL-1b_no DCF_0
- TNF_DCF_1000
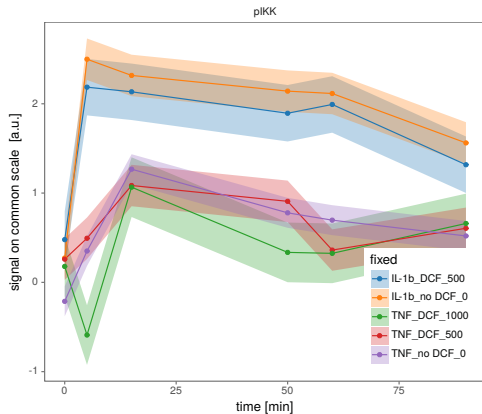- TNF_DCF_500
- TNF_no DCF_0

Explicitly deal with log-data

```
logdata <- data
logdata$value <- log(logdata$value)


out0.log <- alignME(
  data = logdata,
  model = "yi - sj",
  errmodel = "sqrt(sigmaR^2)",
  fixed = yi~stimulus+compound+dose,
  latent = sj~experiment,
  error = sigmaR~1,
  log = FALSE
)
```

- ▶ Use log-data
- ▶ Scaling model turns into offset model
- ▶ Negative parameter values allowed

pIKK

Saturating alignment model

Background:

- Western: enzyme-linked AB
- bound $E \propto$ prot. conc.
- (invisible) $S \xrightarrow{E} P$ (visible)

Alignment model:

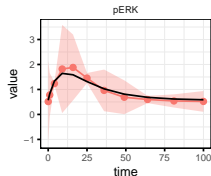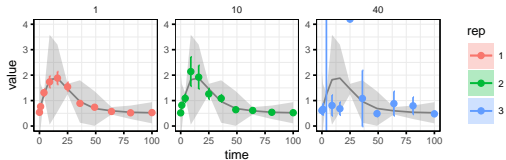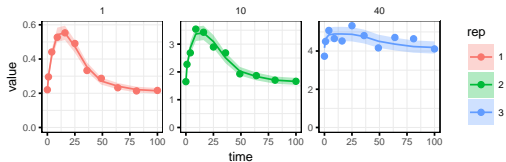$$x_{ij} = \bar{x}_j \left(1 - \mathrm{e}^{-\frac{\alpha_j \, y_i}{\bar{x}_j} T_j}\right)$$

$\bar{x}_j$ ∝ substrate amount

α_j aplification

T_j development time

Simulation:

	$\bar{x}$	α	T
rep1	10	1	1
rep2	5	1	10
rep3	5	1	40

Hypothesis testing

Likelihood-ratio testing:

- ▶ Formulate overarching model (H1)
- ▶ Formulate specialized model (H0)
- ▶ Test log-likelihood difference against χ^2 distribution

Example:

- ▶ IL-1β stimulation

```
$llr
[1] 5.865884

$statistic
[1] "chisquare with 6 degrees of freedom."

$p.value
[1] 0.4383805
```

```
data <- subset(data, name == "pIKK" &
                       stimulus == "IL-1b")

out_H1 <- alignME(
  data = data,
  model = "yi/sj",
  errmodel = "sigmaR*yi/sj",
  fixed = yi~stimulus+compound+dose,
  latent = sj~experiment,
  error = sigmaR~1,
  log = TRUE
)

out_H0 <- alignME(
  data = data,
  model = "yi/sj",
  errmodel = "sigmaR*yi/sj",
  fixed = yi~stimulus,
  latent = sj~experiment,
  error = sigmaR~1,
  log = TRUE
)

llrtest(out_H0, out_H1)
```
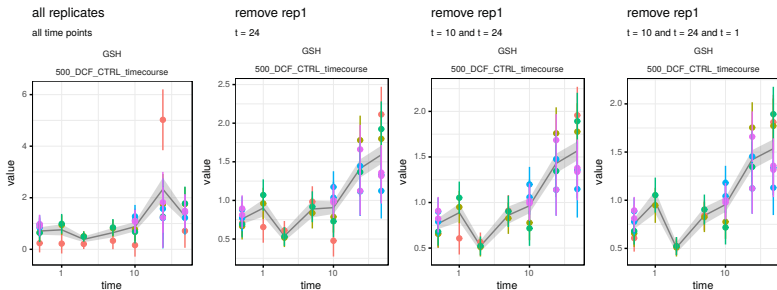
Outlier testing

H_0 Usual alignment model

H_1 Treat tested data point as additional parameters $\rightarrow$ zero resiual

Deviance $(-2\log L(H_0) + 2\log L(H_1))$	72.60	7.64	4.53
p-value	$1.6 \cdot 10^{-17}$	$5.7 \cdot 10^{-3}$	$3.3 \cdot 10^{-2}$

$\rightarrow$ Multiple-testing problem: significance threshold needs to be increased

Conclusion

Conclusion

blotIt2 provides . . .

► flexible alignment models linking dynamics and observed signal

► flexible error models to cover different error sources

► diagnostic plots to inspect qualitiy of the alignment result

► likelihood-ratio testing for hypothesis and outlier tests