

An error model for protein quantification: Supplementary information

C. Kreutz^{a,b,†,*}, M. M. Bartolome Rodriguez^{c,†}, T. Maiwald^{a,b}, M. Seidl^c, H. E. Blum^c, L. Mohr^c, J. Timmer^{a,b}

^aFreiburg Center for Data Analysis and Modeling FDM, Eckerstr. 1, D-79104 Freiburg, Germany,

^bDepartment for Physics, University of Freiburg, Hermann-Herder-Str. 3, D-79104 Freiburg,

Germany ^cDepartment of Medicine II, University Hospital of Freiburg, Hugstetter Str. 55, D-79106 Freiburg, Germany

† both authors contributed equally

6 SUPPLEMENT

6.1 Experimental data

Time courses of phosphorylated insulin receptor (IR^{*}) and insulin receptor substrate (IRS-1^{*}) as well as binding of PI-3 kinase to IRS-1 and phosphorylation of extracellular regulated kinase (ERK-1^{*} and ERK-2^{*}) as functional outcome are measured for different insulin stimulations. In addition, total concentrations of IR, IRS-1, ERK-1 and ERK-2 and of housekeeping proteins gp96, hsc70, β -actin are measured. Further, positive and negative controls are performed within every stimulation with insulin. Total protein concentrations and concentrations of housekeeping proteins are not affected by insulin stimulation and are therefore considered to be constant.

Cells used for measurements are primary hepatocytes of mice origin isolated as previously described (Klingmueller *et al.*, 2006). Isolation procedure is done with cells from two mice livers. Cells are incubated for different times with insulin 24 hours after isolation. After incubation with insulin, cells are harvested and cell pellets are incubated in lysis buffer (10mM Tris-HCl PH 7.4, 150 mM NaCl, 5 mM EGTA, 5 mM EDTA, 1 mM Na₃VO₄, 20 mM NaF, 20 mM Na₄P₂O₇, 1% Triton x-100, protease inhibitors). For immunoprecipitation, 500 μ g protein is mixed with protein A sepharose (Sigma, Deisenhofen) previous coated with antibodies against IRS-1 or IR. Immunoprecipitated proteins as well as total cell lysates separated by SDS/PAGE and transferred to Immobilon P membranes (Millipore, Schwalbach, Germany). It could be shown that randomized gel loading enables correction of within gel inhomogeneities (Schilling *et al.*, 2005b). Following this approach gels are loaded in non-chronological time order.

Membranes are incubated with antibodies against Tyrosine phosphorylated sites (PY-20, BD Heidelberg, Germany), IRS-1 (Upstate biotechnology) IR (Santa Cruz), gp96 (9G10.F8.2, NeoMarkers, Asbach, Germany), hsc 70 (B-6, Santa Cruz), ppERK (V8031, Promega), ERK (V1141, Promega) and β -actin (AC-15, S) and stained with ECL-plus (Amersham, Freiburg, Germany) following

manufacturer's instructions. For quantitative measurements of luminescence, LAS-3000 from Fujifilm (Raytest, Straubenherdt, Germany) was used. Luminescence was quantified using 2D facilities of the AIDA[®] imaging software.

6.2 Additive and multiplicative noise

In Section 3.1 an additive error model

$$y = \alpha + \beta x + \varepsilon, \varepsilon \sim N(0, \sigma_\varepsilon^2) \quad (1)$$

as well as a model

$$\tilde{y} = \beta_0 + \beta_1 x^{\beta_2} \eta, \eta \sim e^{N(0, \sigma_\eta)} \quad (2)$$

with multiplicative noise were introduced. A more general model with additive and multiplicative noise

$$y' = \beta_0 + \beta_1 x^{\beta_2} \eta + \varepsilon \quad (3)$$

is dominated by additive noise ε for small concentrations x and by multiplicative noise η for large concentrations. Unfortunately, an exact transformation leading to Gaussian noise does not exist for model (3) and intensity dependency of measurement errors cannot be removed completely. The most appropriate transformation is a generalized log-transformation (Durbin *et al.*, 2004), (Rocke *et al.*, 2003)

$$y = \log \left(y' - \beta_0 + \sqrt{(y' - \beta_0)^2 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2}} \right) \quad (4)$$

which leads to symmetric errors with first order signal independent variance. Transformed data points can be analysed equivalently to (1) (see Geller *et al.* (2003); Huber *et al.* (2002, 2003)).

Generalized log-transformation requires knowledge about parameters β_0 , σ_ε , σ_η which can be estimated using a general maximum likelihood approach (Honerkamp, 1993; Rocke *et al.*, 1995) or by estimating σ_ε and β_0 from data with low and σ_η from data with high intensities (Geller *et al.*, 2003).

Note, that for large signal intensities $y' - \beta_0 \gg \sigma_\varepsilon^2/\sigma_\eta^2$, e.g. small additive noise $\sigma_\varepsilon \ll \sigma_\eta$, generalized log-transformation corresponds to a simple log-transformation of $y' - \beta_0$ and model (3) becomes equivalent to model (1). For small signals $y' - \beta_0 \ll \sigma_\varepsilon^2/\sigma_\eta^2$, e.g. $\sigma_\varepsilon \ll \sigma_\eta$, measurements consist primarily of observational noise. In this case, transformation (4) reduces to $y \approx \sigma_\varepsilon/\sigma_\eta$.

*to whom correspondence should be addressed

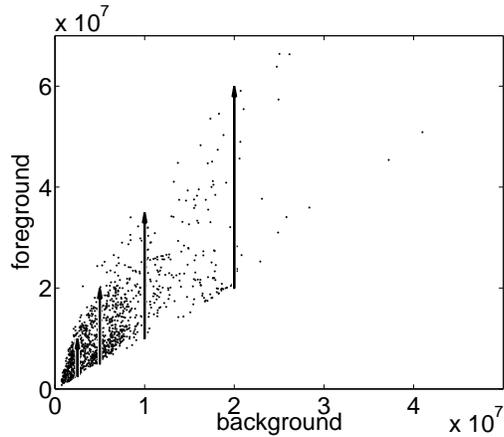


Fig. 1. Foreground intensities are correlated with background intensities. These correlations cannot be removed by background subtraction. This is indicated by the arrows.

6.3 Potential response variables

Background constitutes a systematic bias of measurements. In immunoblotting, signals S are usually calculated by background subtraction

$$S = F - B \quad (5)$$

to eliminate this bias. However, if measurements of background are very noisy, this step may introduce additional variability. In this case, it could be superior to abandon from background correction.

If foreground and background intensities are strongly correlated, e.g. by a common multiplicative error, it could be superior to apply a background correction by division. Ratios

$$R = \frac{F}{B} \quad (6)$$

are more correlated with underlying true protein concentrations if there is a strong common multiplicative effect in foreground as well as in background intensities. This circumstance is illustrated by a simulation study in Section 6.4. Actually, Figure 1 shows that measured foreground and background intensities are strongly correlated. Background subtraction does not remove these correlations completely. This fact is indicated by the arrows in Figure 1.

To check the hypothesis that background correction by division is feasible, repeated measurements of housekeeping proteins on the same gel are used to assess reproducibility. Figure 2 shows relative variability for raw foreground intensities, for background subtracted intensities as well as for ratios foreground over background. The latter show the smallest variability of around 14% for measurement on the same gel and 19% within the same preparation. In contrast, raw foreground intensities have a variability of around 27% and 40%, respectively. Propagation of errors in foreground and background intensities leads to a variability of signals obtained by background subtraction of around 38% within gels and 57% within preparations.

Decreasing variability by multiplicative background correction is a consequence of inhomogeneity of gels, which seems to be a multiplicative influence on foreground as well as on background intensities.

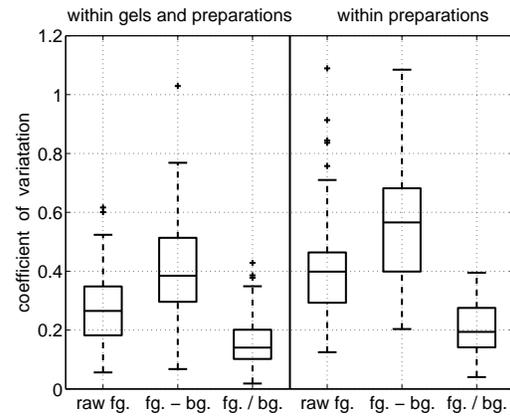


Fig. 2. Ratios foreground / background leads to clearly smaller relative variability. In fact background subtraction increases variability in comparison with uncorrected raw foreground intensities. Boxes around the medians indicates 50% quantiles of observed relative residuals.

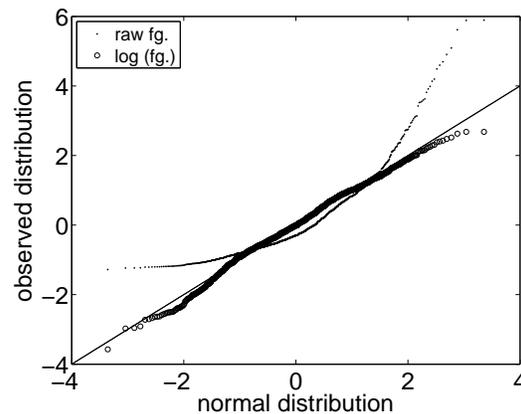


Fig. 3. QQ-plot of foreground intensities without and after log-transformation against normal distribution shows the benefit of the transformation.

For measurements of housekeeping proteins, the underlying protein concentration is constant, e.g. independent from time and stimulation. Therefore repeated observations can be used to determine the distribution of technical and biological noise.

A possibility to assess assumptions about distribution of noise is a plot of theoretical quantiles against observed quantiles. If distribution of observations coincide with theoretical distribution, such a “quantile-quantile plot” or “qq-plot” result in a straight line with slope 1. A deviation from a straight line is a qualitative criterion to assess the assumption of normal distributed noise.

QQ-Plots of raw and log-transformed foreground intensities against normal distribution confirms that log-transformation is required to obtain normally distributed measurements (see Figure 3).

Additionally, log-transformed intensities agrees better with the assumption of normally distributed noise. This is indicated by orders

Table 1. Comparison of repeated measurements of housekeeping proteins with normal distribution.

| Response variable | Abbreviation | Kolmogorov-Smirnov test |
|-------------------|--------------|-------------------------|
| Foreground | F | $p < 1e - 19$ |
| Signals | S | $p = 2.2e - 19$ |
| Signal ratios | R | $p = 1.8e - 8$ |
| log-foreground | $\log(F)$ | $p = 0.224$ |
| log-signals | $\log(S)$ | $p = 0.0014$ |
| log-ratios | $\log(R)$ | $p = 0.00027$ |

A Kolmogorov-Smirnov test shows that log-transformation is required to obtain normally distributed noise.

of magnitude larger p -values obtained by Kolmogorov-Smirnov test (see Table 1).

6.4 Simulation study

In this section, it will be illustrated by a simulation study that multiplicative background correction is superior if strong multiplicative noise affects both, foreground as well as background intensities.

Scanned foreground intensity is composed of background and concentration dependent signal intensity. An evident background correction procedure for estimation of signal S would be a background subtraction

$$\hat{S} = F - B . \quad (7)$$

However, immunoblotting suffers from several systematic multiplicative errors. These errors affect foreground as well as background intensities commonly. For such correlated multiplicative noise, signal ratios

$$R = F/B \quad (8)$$

foreground over background lead to more accurate estimation of underlying concentrations than background subtraction. To illustrate this fact, foreground is simulated according to

$$F = rS + B \quad (9)$$

with a signal to background ratio parameter r . Signals

$$S = x \gamma e^{\eta_1} + x(1 - \gamma)e^{\eta_2} + \varepsilon_1 , \quad (10)$$

$$\eta_i \sim N(0, \sigma_{\eta_i}) , i = 1, 2 , \varepsilon_1 \sim N(0, \sigma_{\varepsilon_1})$$

consist on multiplicative lognormal distributed noise e^{η_1} and e^{η_2} and additive Gaussian noise ε_1 . It is assumed that η_1 is the common error which affects background, too. Parameter $\gamma \in [0, 1]$ is used to vary the proportion of this common error. True underlying protein concentration is denoted by x . Background intensities are simulated according to

$$B = \gamma e^{\eta_1} + (1 - \gamma)e^{\eta_3} + \varepsilon_2 . \quad (11)$$

$$\eta_3 \sim N(0, \sigma_{\eta_3}) , \varepsilon_2 \sim N(0, \sigma_{\varepsilon_2})$$

where η_3 denotes the uncorrelated fraction of multiplicative noise corresponding to η_2 for signals.

The intention of a background correction procedure is to improve the determination of underlying protein dynamics. Correlations

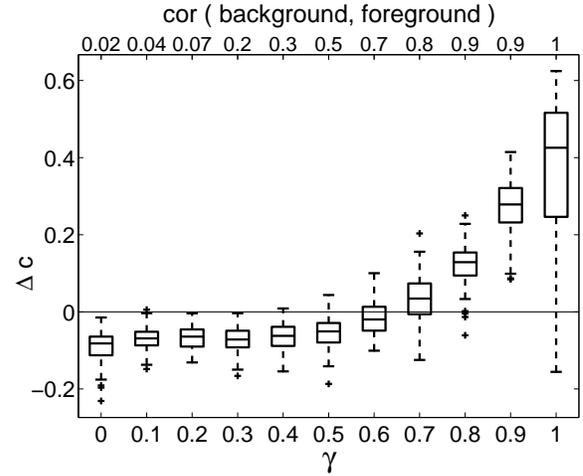


Fig. 4. For $\gamma > 0.65$, corresponding to a correlation of foreground and background of about 0.7, signal ratios are better correlated to underlying protein concentrations than signals obtained after background subtraction.

Table 2. Chosen parameters for simulation study.

| Parameter | Description | Parameter value |
|-----------------|------------------------------|------------------------|
| r | Signal to background ratio | 10 |
| n | Number of simulations | 100 |
| N | Number of data points | 1000 |
| γ | Fraction of common error | $\{0, 0.1, \dots, 1\}$ |
| x | True concentrations | $\sim N(0, 1)$ |
| ε_1 | Additive noise in signals | $\sim N(0, 0.1)$ |
| ε_2 | Additive noise in background | $\sim N(0, 0.1)$ |
| η_1 | Common multiplicative noise | $\sim N(0, 1)$ |
| η_2 | Multiplicative noise of S | $\sim N(0, 1)$ |
| η_3 | Multiplicative noise of B | $\sim N(0, 1)$ |

with underlying true protein concentrations are used to compare background subtraction with signal ratios. For positive differences

$$\Delta c = \text{cor}(R, x) - \text{cor}(F - B, x) \quad (12)$$

of correlations, signal ratios are more accurate estimators of the underlying protein concentration dynamics.

Figure 4 shows Δc obtained by simulations for different values of parameter γ . For small γ , foreground and background intensities are only weakly correlated and signals $F - B$ should be used for background correction. However, for large γ background correction by division is superior. For our simulation, above a threshold of $\gamma = 0.65$ corresponding to a correlation between foreground and background of around 0.7, signal ratios become more accurate than background subtraction.

Realistic values are assumed for parameters (see Table 2). The obtained result depends only weakly on chosen parameter values. Since ratios R diverge for background intensities near zero, lower

Table 3. Model selection of superior models.

| Tested effect | | p -value for model 26 | p -value for model 26' |
|--------------------------|-----------------------|----------------------------|-----------------------------|
| Time and stimulation | T | $< 2.2e - 16$ | $< 2.2e - 16$ |
| Background | B | $1.58e - 15$ | $6.67e - 16$ |
| Preparation specific B | $\epsilon_p^{(1)}$ | 0.095 | $1.67e - 3$ |
| Preparation effects | $\epsilon_p^{(2)}$ | $4.13e - 5$ | $4.01e - 3$ |
| Gel specific B | $\epsilon_{pg}^{(1)}$ | $< 2.2e - 16$ | $< 2.2e - 16$ |
| Gel effects | $\epsilon_{pg}^{(2)}$ | $< 2.2e - 16$ | $< 2.2e - 16$ |

p -values obtained by likelihood-ratio tests of the superior error model against submodels.

1% quantile of background intensities are excluded. This assumption is not in contradiction to our experiments because the magnitude of smallest intensities is sufficiently far away from zero and no background intensities larger than foreground are observed in practice.

6.5 Model selection for superior error model

Submodels obtained by omitting an effect are tested by likelihood-ratio tests against full model 26 and 26', respectively. For this purpose, parameters are fitted by maximum likelihood estimation (Pineiro *et al.*, 2000) instead of restricted maximum likelihood method. Resulting p -values are displayed in Table 3. Only the effect accounting for preparation specific background correction $\epsilon_p^{(1)}$ is not clearly significant. We decided to include $\epsilon_p^{(1)}$, because the full model has superior AIC . Additionally, elimination of weak but true effects can cause a so called *omission bias* (Miller *et al.*, 1984). This means that estimated time courses could be biased if true but weakly significant effects are excluded from a the model. On the basis of our data, it is not possible to decide finally if there is a preparation specific background variation.

For the full data set, a regression coefficient of $b = 0.98 \pm 0.04$ was estimated for background correction. The confidence interval is in agreement with ratios $\log(F)/\log(B)$ corresponding to $b = 1$. Nevertheless a likelihood-ratio test is significant with $p < 2.2e - 16$.

A qq-plot (Figure 5) for the superior error model No. 26' shows agreement of residuals with normal distribution. Additionally a scatter plot of residuals against predicted values shows no pattern (see Fig. 6).

6.6 Matrix notation

The superior mixed effects model 26'

$$\begin{aligned} \log(F_{ostpg}) &= O_o + T_{ost} + \left(b + \epsilon_p^{(1)} + \epsilon_{pg}^{(1)} \right) \log(B_*) \\ &\quad + \epsilon_p^{(2)} + \epsilon_{pg}^{(2)} + \varepsilon_{ostpg} \\ \epsilon_p^{(1)} &\sim N(0, \sigma_1^{(1)}), \quad \epsilon_{pg}^{(1)} \sim N(0, \sigma_2^{(1)}), \quad \epsilon_p^{(2)} \sim N(0, \sigma_1^{(2)}), \\ \epsilon_{pg}^{(2)} &\sim N(0, \sigma_2^{(2)}), \quad \varepsilon_* \sim N(0, \sigma), \end{aligned} \quad (13)$$

can be written in matrix notation as

$$\vec{F}_{pg} = \mathbf{X}_{pg} \vec{\beta} + \mathbf{Z}_{pg} \vec{\gamma}_{pg} + \vec{\varepsilon}_{pg} \quad (14)$$

where

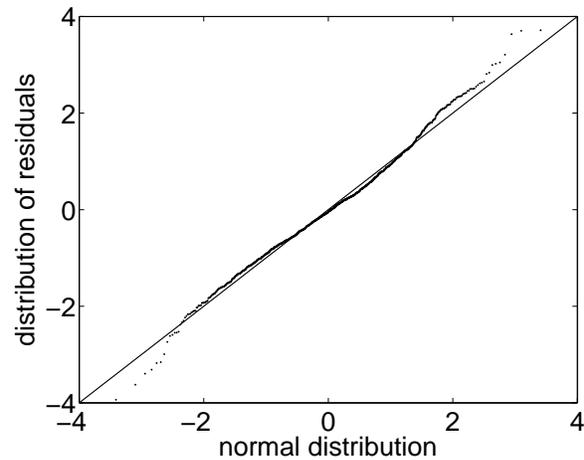


Fig. 5. QQ-plot of residuals of the superior model against normal distribution.

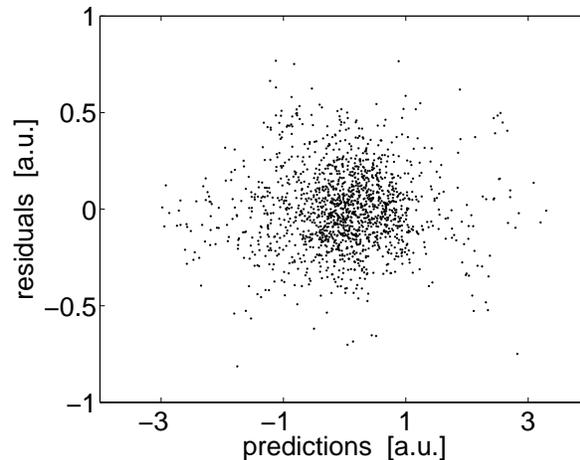


Fig. 6. Residuals of the superior error model show no dependency on predictions of the model.

- $\vec{F}_{pg} \in M(n_{pg} \times 1)$ are arrays containing all n_{pg} log-foreground intensities of preparation p on gel g .
- $\mathbf{X}_{pg} \in M(n_{pg} \times n_{\text{fixed}})$ are model matrices. Without loss of generality \mathbf{X}_{pg} can be written as

$$\mathbf{X}_{pg} = \begin{pmatrix} \vec{B}_{pg} & \mathbf{X}_{pg}^O & \mathbf{X}_{pg}^T \end{pmatrix} \quad (15)$$

where $(X_{ij}^O)_{pg} \in \{0, 1\}$ indicates whether observable j has an influence on data point i of preparation p on gel g . Similarly, $(X_{ij}^T)_{pg} \in \{0, 1\}$ indicates whether time effect j has an influence on data point i . \vec{B}_{pg} are log-background intensities of considered preparation and gel.

- $\vec{\beta} \in M(n_{\text{fixed}} \times 1)$ is an array containing all fixed effects. Model 26' has

$$n_{\text{fixed}} = 1 + n_o + \sum_{o,s} n_t(o, s) \quad (16)$$

Table 4. Measured time dependent observables of insulin signaling pathway.

| Observable | Description | IP | Antibody | Insulin treatment |
|------------|-----------------|-------|----------|-------------------|
| O_1 | Activated IR | IR | Py20 | 1e-7 nm |
| O_2 | Activated IR | IR | Py20 | 1e-5 nm |
| O_3 | Activated IRS-1 | IRS-1 | Py20 | 1e-7 nm |
| O_4 | Activated IRS-1 | IRS-1 | Py20 | 1e-5 nm |
| O_5 | PI3K binding | IRS-1 | PI3K | 1e-7 nm |
| O_6 | PI3K binding | IRS-1 | PI3K | 1e-5 nm |
| O_7 | Activated ERK-1 | - | ppERK-1 | 1e-7 nm |
| O_8 | Activated ERK-1 | - | ppERK-1 | 1e-5 nm |
| O_9 | Activated ERK-2 | - | ppERK-2 | 1e-7 nm |
| O_{10} | Activated ERK-2 | - | ppERK-2 | 1e-5 nm |

fixed parameters. If model matrices \mathbf{X}_{pg} are chosen as described above, it holds

$$\vec{\beta} = \begin{pmatrix} b \\ \vec{O} \\ \vec{T} \end{pmatrix}. \quad (17)$$

The first entry of $\vec{\beta}$ is the regression coefficient b of background correction followed by observable and time effects.

- $\mathbf{Z}_{pg} \in M(n_{pg} \times 4)$ are model matrices for random effects. In our model, every data point is influenced by all four random effects. This yields to

$$\mathbf{Z}_{pg} = \begin{pmatrix} \log((B_1)_{pg}) & \log((B_1)_{pg}) & 1 & 1 \\ \dots & \dots & \dots & \dots \\ \log((B_{n_{pg}})_{pg}) & \log((B_{n_{pg}})_{pg}) & 1 & 1 \end{pmatrix}. \quad (18)$$

n_{pg} is the number of data points of preparation p on gel g .

- $\vec{\gamma}_{pg} \in M(4 \times 1)$, $\vec{\gamma}_{pg} \sim N(\vec{0}, \Psi^2)$ denote arrays containing random effects coefficients of preparation p and gel g . The correlation matrix is

$$\Psi = \begin{pmatrix} \sigma_{pg}^{(1)} & 0 & 0 & 0 \\ 0 & \sigma_g^{(1)} & 0 & 0 \\ 0 & 0 & \sigma_{pg}^{(2)} & 0 \\ 0 & 0 & 0 & \sigma_g^{(2)} \end{pmatrix}. \quad (19)$$

- $\vec{\epsilon}_{pg} \sim N(\vec{0}, \sigma^2 \mathbf{I})$ denotes uncorrelated observational noise of variance σ^2 .

6.7 Time courses for all data and error models

Error models for constant normalizer proteins contain only effects corresponding to background, biological preparation variability, gel-to-gel differences and observed proteins. For time dependent signaling proteins, the models have to be extended to estimate time effects after insulin stimulation. Within the applied parameter estimation process, observed total variability is split to all considered effects in a model. Both, incomplete models and models containing effects that are not required show increased confidence intervals for time effects. Moreover, model assessment criteria are worse because of large residuals and/or large number of parameters. These criteria

Table 6. Model assessment criteria for testing the assumption of time and treatment independence of housekeeping proteins.

| | n_{par} | AIC | BIC | p_{ks} | SNR |
|------------------------------------|-----------|------|------|----------|-----|
| model 26 | 13 | -98 | -33 | 0.008 | 3.7 |
| model 26 + treatment effects | 34 | -9.1 | 162 | 0.0019 | 3.7 |
| model 26 + time effects | 92 | 268 | 735 | 0.18 | 3.9 |
| model 26 + time and treat. effects | 245 | 705 | 1981 | 0.0056 | 4.2 |

AIC and BIC prefer a model without time and treatment effects. We concluded that the selection of housekeeping proteins based on biological prior knowledge is appropriate.

and p -values of Kolmogorov-Smirnov test checking for normally distributed residuals are displayed in Table 5.

In accordance to our result obtained from housekeeping proteins, model 26' is superior in 2 out of 5 criteria. Observed advantage of log-transformation and background subtraction on log-scale is again confirmed.

Obtained time courses for all considered models are displayed in Figures 7 to 9. Rows correspond to different observables which are displayed in Table 4. A zigzag shape of some time courses emerge because neighboring time points are mostly not on the same gel. This causes a badly identifiable parameter which determines how time effects of even and uneven time points have to be merged. This is no problem if gel effects are modeled by a random variable, because only one parameter has to be estimated for all gels.

Error models without application of log-transformation show large error bars (Figures 7). If log-transformation is applied and systematic errors are treated multiplicatively, error bars are decreased (Figures 8). Qualitatively similar results can be seen in Figures 9 where a regression parameter is estimated for background correction. This step improves model assessment criteria and leads to the overall best model No. 26'. For this model log-transformed foreground intensities are used as response variable, preparation and gel effects are modeled as random variables and a gel specific random regression parameter is estimated for background correction.

Although, all models yield qualitatively similar shapes for time dependency, the estimated dynamic behavior depends on the chosen model. One possibility to avoid this dependency would be a *model averaging* procedure. Here, a weighted average of all estimated time courses would be calculated. Weights w_M are given by the posterior probability of considered model M . This posterior probability can be approximated up to first order by the exponential of Bayes' Information criterion BIC_M of model M (Kass et al., 1994):

$$w_M = \frac{\exp(-\frac{BIC_M}{2})}{\sum_m \exp(-\frac{BIC_m}{2})}. \quad (20)$$

Because model 26' has a clearly superior BIC, this model would contribute mainly in a model averaging process ($w_{26'} \approx 1$).

Recapitulating, all appropriate error models lead to qualitatively similar time courses. Nevertheless, the obtained time effects depend on applied models. This emphasizes the need of proper error models for the analysis of immunoblotting and immunoprecipitation measurements.

Table 5. Comparison of error models for time course measurements.

| Model No. | model | n_{par} | AIC | BIC | p_{ks} | SNR | TR |
|-----------|---|------------------|------------|-------------|-----------------|------------|------------|
| 1' | $F_* = O_o + T_{ost} + \varepsilon_*$ | 74 | 56500 | 56900 | 1.6e-58 | 0.28 | 2.1 |
| 2' | $F_* = O_o + T_{ost} + P_p + G_{pg} + \varepsilon_*$ | 185 | 54400 | 55400 | 8.6e-26 | 2 | 7.4 |
| 3' | $F_* = O_o + T_{ost} + \epsilon_p + \epsilon_{pg} + \varepsilon_*$ | 76 | 52300 | 52700 | 4.6e-28 | 1.9 | 6.5 |
| 4' | $S_* = O_o + T_{ost} + \varepsilon_*$ | 74 | 55300 | 55700 | 3.4e-83 | 0.3 | 2.4 |
| 5' | $S_* = O_o + T_{ost} + P_p + G_{pg} + \varepsilon_*$ | 185 | 53700 | 54700 | 2.4e-38 | 1.6 | 5.1 |
| 6' | $S_* = O_o + T_{ost} + \epsilon_p + \epsilon_{pg} + \varepsilon_*$ | 76 | 51700 | 52100 | 9.5e-39 | 1.6 | 5.8 |
| 7' | $R_* = O_o + T_{ost} + \varepsilon_*$ | 74 | 3130 | 3530 | 3.2e-13 | 0.59 | 4.3 |
| 8' | $R_* = O_o + T_{ost} + P_p + G_{pg} + \varepsilon_*$ | 185 | 2740 | 3730 | 7.4e-5 | 1 | 6.8 |
| 9' | $R_* = O_o + T_{ost} + \epsilon_p + \epsilon_{pg} + \varepsilon_*$ | 76 | 2980 | 3390 | 1.3e-5 | 0.92 | 8.2 |
| 10' | $\log(F_*) = O_o + T_{ost} + \varepsilon_*$ | 74 | 3900 | 4290 | 0.00014 | 0.47 | 3.5 |
| 11' | $\log(F_*) = O_o + T_{ost} + P_p + G_{pg} + \varepsilon_*$ | 185 | 2630 | 3620 | 0.02 | 1.5 | 8.2 |
| 12' | $\log(F_*) = O_o + T_{ost} + \epsilon_p + \epsilon_{pg} + \varepsilon_*$ | 76 | 2870 | 3280 | 0.12 | 1.4 | 8.6 |
| 13' | $\log(S_*) = O_o + T_{ost} + \varepsilon_*$ | 74 | 4860 | 5260 | 0.0019 | 0.59 | 5.4 |
| 14' | $\log(S_*) = O_o + T_{ost} + P_p + G_{pg} + \varepsilon_*$ | 185 | 4100 | 5080 | 0.0013 | 1.3 | 8.8 |
| 15' | $\log(S_*) = O_o + T_{ost} + \epsilon_p + \epsilon_{pg} + \varepsilon_*$ | 76 | 4260 | 4670 | 0.00011 | 1.2 | 9.5 |
| 16' | $\log(R_*) = O_o + T_{ost} + \varepsilon_*$ | 74 | 807 | 1200 | 0.031 | 0.62 | 5.1 |
| 17' | $\log(R_*) = O_o + T_{ost} + P_p + G_{pg} + \varepsilon_*$ | 185 | 430 | 1420 | 0.77 | 1 | 8.0 |
| 18' | $\log(R_*) = O_o + T_{ost} + \epsilon_p + \epsilon_{pg} + \varepsilon_*$ | 76 | 773 | 1180 | 0.68 | 0.93 | 8.8 |
| 19' | $F_* = O_o + T_{ost} + b B_* + \varepsilon_*$ | 75 | 54700 | 55100 | 3.6e-60 | 1.6 | 2.6 |
| 20' | $F_* = O_o + T_{ost} + b B_* + P_p + G_{pg} + \varepsilon_*$ | 186 | 53700 | 54700 | 2.8e-40 | 2.6 | 4.2 |
| 21' | $F_* = O_o + T_{ost} + b B_* + \epsilon_p + \epsilon_{pg} + \varepsilon_*$ | 77 | 51600 | 52100 | 6e-41 | 2.5 | 3.5 |
| 22' | $F_* = O_o + T_{ost} + (b + \epsilon_p^{(1)} + \epsilon_{pg}^{(1)}) B_* + \epsilon_p^{(2)} + \epsilon_{pg}^{(2)} + \varepsilon_*$ | 81 | 51200 | 51600 | 4.8e-45 | 2.9 | 2.6 |
| 23' | $\log(F_*) = O_o + T_{ost} + b \log(B_*) + \varepsilon_*$ | 75 | 799 | 1200 | 0.024 | 2.9 | 5.1 |
| 24' | $\log(F_*) = O_o + T_{ost} + b \log(B_*) + P_p + G_{pg} + \varepsilon_*$ | 186 | 382 | 1370 | 0.5 | 3.6 | 9.0 |
| 25' | $\log(F_*) = O_o + T_{ost} + b \log(B_*) + \epsilon_p + \epsilon_{pg} + \varepsilon_*$ | 77 | 737 | 1150 | 0.63 | 3.5 | 7.9 |
| 26' | $\log(F_*) = O_o + T_{ost} + (b + \epsilon_p^{(1)} + \epsilon_{pg}^{(1)}) \log(B_*) + \epsilon_p^{(2)} + \epsilon_{pg}^{(2)} + \varepsilon_*$ | 81 | 490 | 923 | 0.0052 | 4.2 | 8.0 |

Abbreviation * is used instead of all occurring indices in a model, e.g. indices of all predictor variables and an index for replicate measurements. Best values are underlined and the 5 superior values of each model assessment criterion are highlighted in bold face. In accordance with results obtained from housekeeping proteins, log-transformation improves performance. Model 26' is superior for 2 out of 6 criteria.

6.8 Housekeeping proteins

β -actin, gp96 and hsc70 are widely used housekeeping proteins (Li *et al.*, 2002; Picard, 2002; Suzuki *et al.*, 2000; Schilling *et al.*, 2005b). In addition, we considered total insulin receptor and total insulin receptor substrate concentrations as constant because there are no biological indications that both molecule concentrations are changed in mouse hepatocytes after insulin stimulation within the first hour.

Based on this biological prior knowledge we used β -actin, gp96, hsc70, IR_{total} and IRS-1_{total} to determine an error model for constant proteins in Section 4.2.

To validate the assumption that housekeeping proteins are indeed independent on stimulation and constant over time, the obtained superior error model for housekeeping proteins is enlarged by time and treatment effects.

Table 6 shows that *AIC* and *BIC* are clearly superior for the model without time and treatment effects. P-values obtained by a Kolmogorov-Smirnov test indicate that violations from the assumption of normally distributed residuals are similar for the considered four models. Because variance of residuals is always decreased by an enlargement of model 26, signal to noise ratio *SNR* is not very meaningful for comparison of considered models with submodel 26.

REFERENCES

- Durbin, B., Rocke, D. (2004) Variance-stabilizing transformations for two-color microarrays, *Bioinformatics*, **19**, 1360-1367.
- Geller, S., Gregg, J., Hagerman, P., Rocke, D. (2003) Transformation and normalization of oligonucleotide microarray data, *Bioinformatics*, **19**, 1817-1823.
- Honerkamp, J. (1993) Stochastic Dynamical Systems, *VCH*.
- Huber, W., Heydebreck, A., Suetmann, H., Poustka, A., Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression, *Bioinformatics*, **18**, 96-104.
- Huber, W., Heydebreck, A., Suetmann, H., Poustka, A., Vingron, M. (2003) Parameter estimation for the calibration and variance stabilization of microarray data, *Stat Appl Genet Mol Biol*, **2**, 3.
- Kass, R., Raftery, A. (1994) Bayes Factors, *Dep. of Statistics, University of Washington*, **254**.
- Li, Z., Dai, J., Zheng, H., Liu, B., Caudill, M. (2002) An integrated view of the roles and mechanisms of heat shock protein gp96-peptide complex in eliciting immune response., *Front Biosci*, **7**, d731-d751.
- Miller, A.J. (1984) Selection of Subsets of Regression Variables, *Journal of the Royal Statistical Society A*, **147(3)**,389-425.
- Picard, D. (2002) Heat-shock protein 90, a chaperone for folding and regulation, *Cell Mol Life Sci*, **59**, 1640-1648.

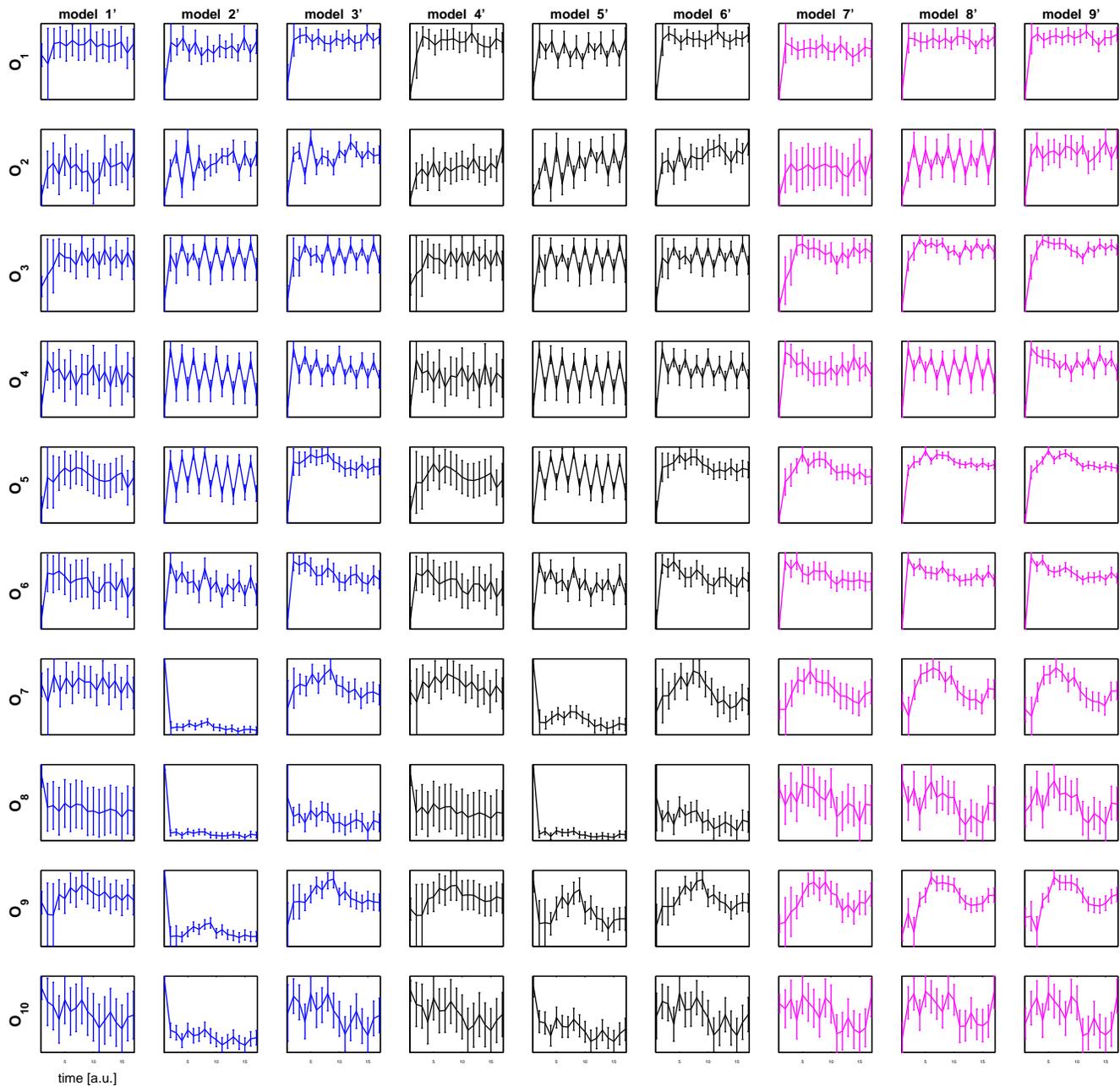


Fig. 7. Estimated time effects of 10 measured observables for error models 1'-9'. For models 1'-3' (blue color) raw foreground intensities are used as response variable. Background subtraction is applied for models 4'-6' (black) and ratios foreground over background are used for models 7'-9' (magenta). Models with fixed gel effect show a zigzag shape of time courses due to a badly identifiable parameter.

Pinheiro, C., Bates, M. (2000) Mixed-Effects models in S and S-Plus, *Springer*.
 Klingmueller, U., Bauer, A., Bohl, S., Nickel, P., Dooley, K., Zellmer, S., Kern, C., Merfort, I., Donauer, T., Walz, G., Geyer, M., Kreutz, C., Goetschel, M., Hecht, A., Walter, D., Egger, L., Borner, K., Brulport, M., Schormann, W., Baumann, C., Preiss, R., MacNelly, S., Godoy, P., Ciuculan, E., Edelmann, J., Zeilinger, K., Zanger, M., Gebhardt, R., Maiwald, T., Timmer, J., Weizsaecker,

F., Hengstler, J. (2006) Primary mouse hepatocytes for Systems Biology approaches: A standardized in vitro system for modeling of signal transduction pathways, *IEE Proceedings - Systems Biology*, **153**, 433-447.

Rocke, D., Lorenzato, S. (1995) A Two-Component Model for Measurement Error in Analytical Chemistry, *Technometrics*, **37**, 176-185.

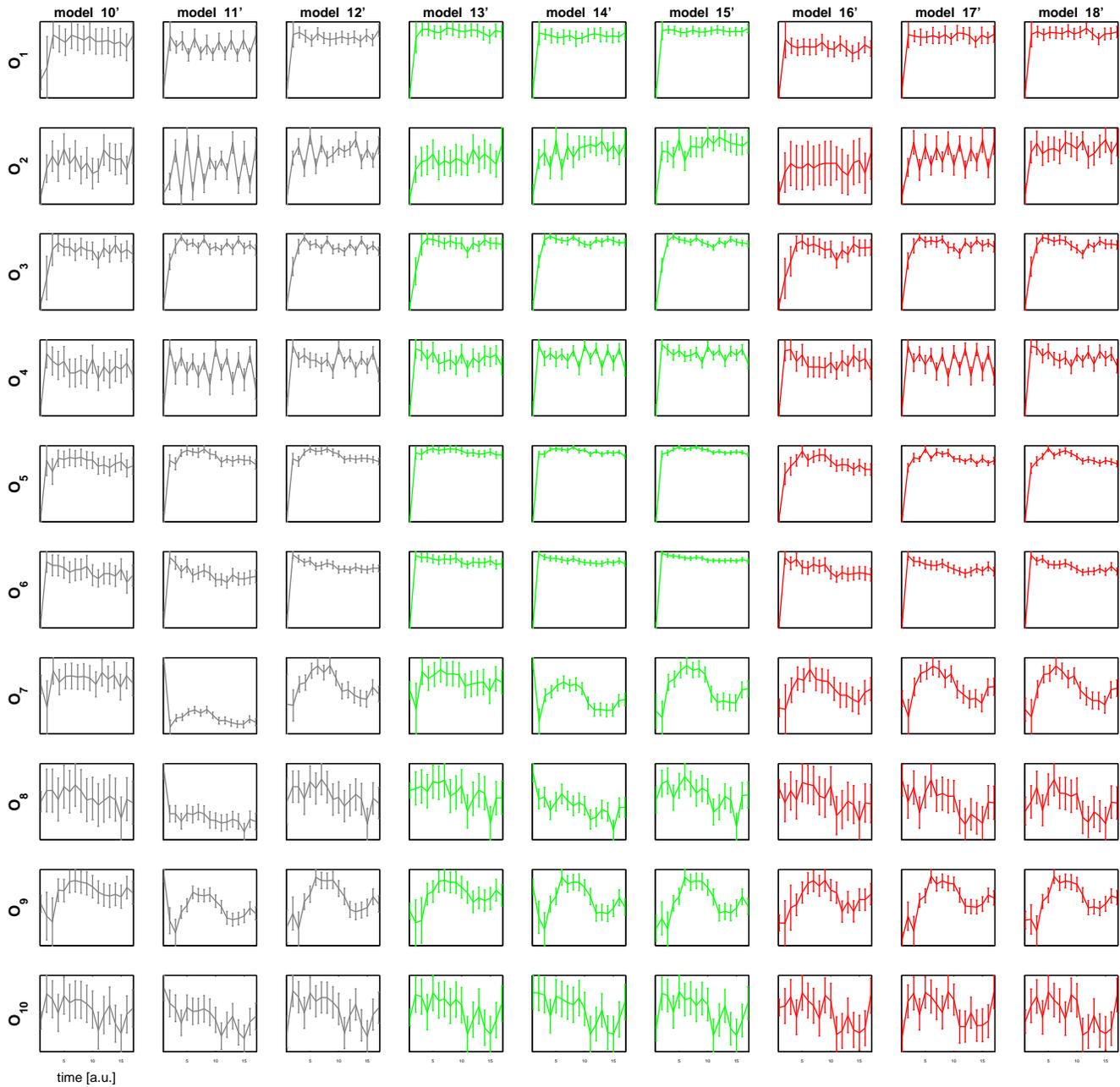


Fig. 8. Estimated time effects of regarded observables after log-transformation. Log-foreground intensities are used as response variable for models 10'-12' (gray color). Models using log-signals (13'-16') are display in green color and models for log-ratios are plotted in red.

Rocke, D., Durbin, B. (2003) Approximate variance-stabilizing transformations for gene-expression microarray data, *Bioinformatics*, **19**, 966-972.

Schilling, M., Maiwald, T., Bohl, S., Kollmann, M., Timmer, J., Klingmueller, U. (2005b) Quantitative data generation for

Systems Biology: The impact of randomisation, calibrators and normalisers, *IEE Proceedings - Systems Biology*, **152**, 193-200.

Suzuki, T., Higgins, P. J., Crawford, D.R. (2000) Control selection for RNA quantitation, *Biotechniques*, **29**, 332-337.

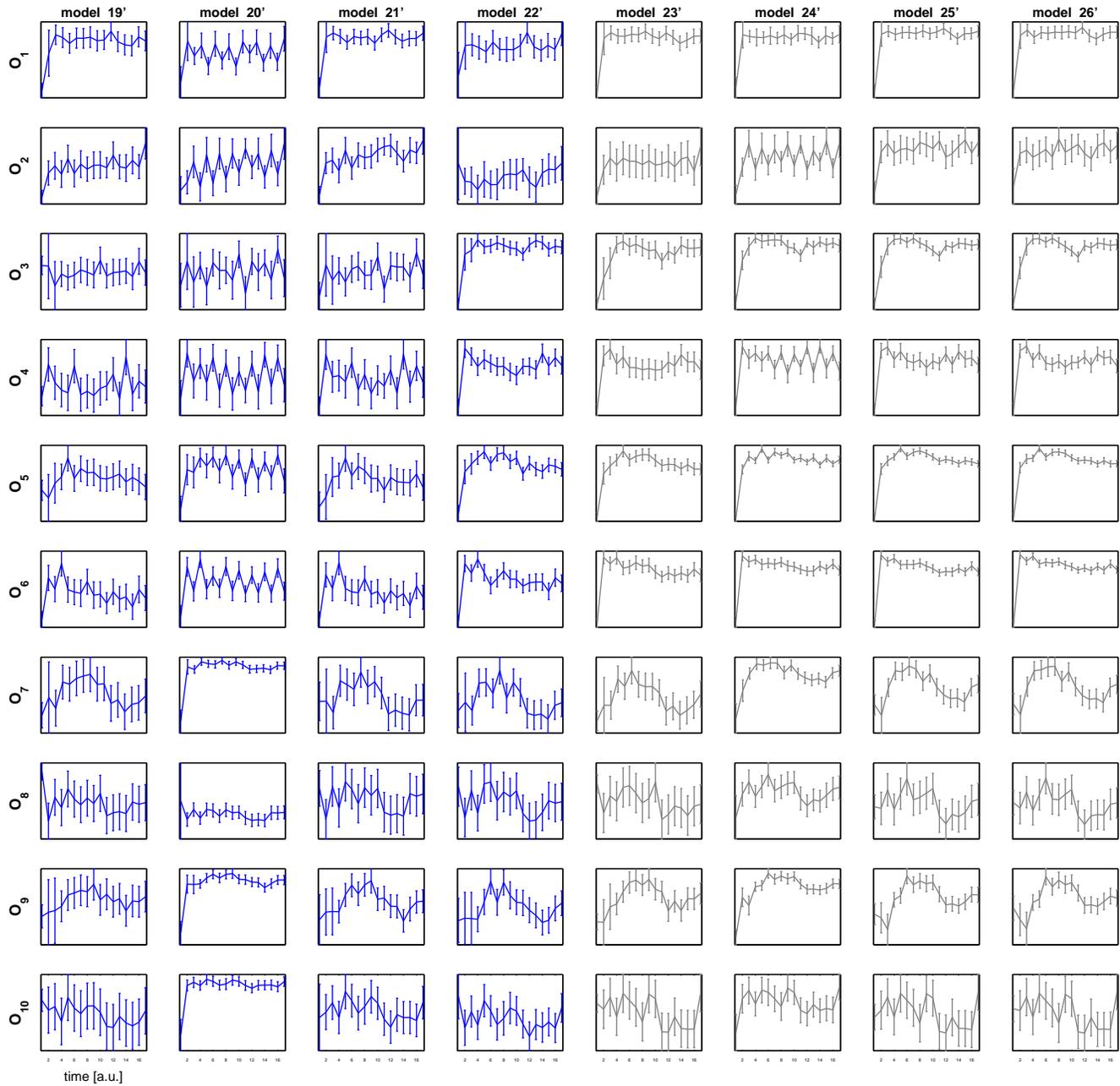


Fig. 9. Estimated time effects for error models with fitted regression parameter for background correction. Untransformed foreground intensities are displayed in blue whereas log-transformed intensities are plotted in gray color. Model 26' is overall best model selected by our model assessment criteria.