Statistical Applications in Genetics and Molecular Biology

Volume 8, *Issue* 1 2009 *Article* 5

A Nonlinear Mixed-Effects Model for Estimating Calibration Intervals for Unknown Concentrations in Two-Color Microarray Data with Spike-Ins

Pushpike J. Thilakarathne, *Catholic University of Leuven and University of Hasselt*
Geert Verbeke, *Catholic University of Leuven and University of Hasselt*
Kristof Engelen, *Catholic University of Leuven*
Kathleen Marchal, *Catholic University of Leuven*

Recommended Citation:
Thilakarathne, Pushpike J.; Verbeke, Geert; Engelen, Kristof; and Marchal, Kathleen (2009) "A Nonlinear Mixed-Effects Model for Estimating Calibration Intervals for Unknown Concentrations in Two-Color Microarray Data with Spike-Ins," *Statistical Applications in Genetics and Molecular Biology*: Vol. 8: Iss. 1, Article 5.
DOI: 10.2202/1544-6115.1401

A Nonlinear Mixed-Effects Model for Estimating Calibration Intervals for Unknown Concentrations in Two-Color Microarray Data with Spike-Ins

Pushpike J. Thilakarathne, Geert Verbeke, Kristof Engelen, and Kathleen Marchal

Abstract

In this study, we propose a calibration method for preprocessing spiked-in microarray experiments based on nonlinear mixed-effects models. This method uses a spike-in calibration curve to estimate normalized absolute expression values. Moreover, using the asymptotic properties of the calibration estimate, 100(1-α)% confidence intervals for the estimated expression values can be constructed. Simulations are used to show that the approximations on which the construction of the confidence intervals are based are sufficiently accurate to reach the desired coverage probabilities. We illustrate applicability of our method, by estimating the normalized absolute expression values together with the corresponding confidence intervals for two publicly available cDNA microarray experiments (Hilson et al., 2004; Smets et al., 2008). This method can easily be adapted to preprocess one-color oligonucleotide microarray data with a slight adjustment to the mixed model.

KEYWORDS: calibration, nonlinear model, mixed model, cDNA microarray, oligonucleotide

Author Notes: We thank the referees and associate editor for the valuable suggestions. This work was financially supported by SymBioSys, the K.U.Leuven Center of Excellence in System Biology.

1 Introduction

Microarray experiments provide indirect measurements of gene expression (mRNA abundance in a biological sample) by measuring intensities of labeled RNA target bound to corresponding probes on the array. These measurements are not only indirect, but are also influenced by several experimental sources of systematic and random variation, making effective preprocessing a crucial factor in any analysis. For preprocessing of spotted microarrays, different methods have been described. Overviews are given, e.g., by Leung and Cavalieri (2002), Quackenbush (2002), and Bilban *et al.* (2002). In general, preprocessing of spotted microarrays largely depends on the calculation of the log-ratios of the measured intensities. However for some analyses having access to absolute expression levels is more suitable (Kerr *et al.*, 2000). ANOVA models for absolute expression levels have been proposed, e.g., by Wolfinger *et al.* (2001).

We propose the use of external reference RNAs (also known as spike-in controls or spikes) to preprocess cDNA microarray data. Spike RNAs have no sequence similarity to the genome of the studied species and they are added in defined amounts to experimental RNA samples before labeling. The use of spikes allows not only data preprocessing but also the evaluation of several parameters of the platform quality, including the sensitivity and specificity of the microarray experiments, the accuracy and reproducibility of the measurements and the assessment of technical variability introduced by labeling procedure, hybridization and image scanning (Badiee *et al.*, 2003; van Bakel *et al.*, 2004).

In this paper we obtain estimates of the actual RNA abundances by quantifying the relationship between RNA abundance and measured intensity using data from the external spikes. This problem can thus be considered as a calibration issue where linear or non-linear models can be employed. In the literature, Dudley *et al.* (2002) applied a linear regression method on data acquired from the same microarray slide under the several photomultiplier tube gains to extend the linear range of a scanner. Shi *et al.* (2005) assumed that the relationship between log fluorescence intensity and log dye RNA concentration can be reasonably described by a Sigmoid function. However, they did not incorporate the hybridization reaction in constructing the relationship between intensity measurements and concentrations. The normalization procedure proposed by Engelen *et al.* (2006), however, is based on a calibration method for spiked-in RNAs and integrates the hybridization reaction and the dye-saturation function in building the calibration model (see Section 3 for more details). This method exploits the spikes (transcripts with known RNA concentration in pg/ml), which are added to the hybridization solution, in order to build a calibration curve. The underlying model compensates for the spot effect and the non-linear behavior of both the red and the green dye and hence normalizes the

data. The calibration curve is subsequently used to estimate the absolute expression level of the transcripts (gene expression levels) from the measured intensities. Such an approach poses advantages over classical normalization methods: it does not make any assumption on the distribution of the data (such as classical normalization approaches based on locally weighted least squares do) and it avoids calculating ratio's. The disadvantage of this model proposed by Engelen *et al.* (2006) is that it lacks a clear statistical framework that would allow correct assessment of the various sources of uncertainty, needed for the calculation of confidence intervals for the predicted expression values. We therefore propose to combine the models Engelen *et al.* (2006) used in the various steps of their estimation procedure, into one non-linear mixed model, on which the statistical calibration for spike based normalization of cDNA arrays can be based, also yielding an estimate for the uncertainty in the predictions. In Section 2, the type of experiments considered is described. In Section 3, the hybridization and dye saturation models of Engelen *et al.* (2006) are presented. Our proposed combined non-linear mixed model, together with some implementation issues, will be introduced in Section 4. The procedures for calibration and for the construction of confidence intervals are described in Sections 5 and 6, respectively, and will be illustrated extensively in the analysis of a publicly available dataset in Section 7. Finally, a simulation study will be presented in Section 8, to study the finite sample behaviour of the proposed confidence intervals. Concluding remarks are given in Section 9.

2 Spike in experiments

Our proposed model makes use of spike in experiments. These are arrays that contain spikes: control spots for which the RNA transcripts are added to the hybridization solution in known concentrations prior to the labeling. These spikes enable us to model a relationship between the concentration of a transcript in the hybridization solution and its measured absolute intensity measurements (estimate the intensity measurement as a function of the concentration). The purpose of the analysis is to estimate the unknown concentrations based on a model fitted to the measured intensities for the concentrations of the spikes.

The most basic designs that are commonly applied for cDNA microarray experiments are Color flip designs (see Figure 1), Reference designs, and Loop designs (Churchill, 2002). The simplest microarray experiments compare expression in two distinct conditions. A test condition (e.g. a cell line triggered with a drug compound) is compared to a reference condition (e.g. a cell line triggered with a placebo). Usually the test is labeled with Cy5 (red dye) while the reference is labeled with Cy3 (green dye). Performing replicate experiments is mandatory to

DOI: 10.2202/1544-6115.1401

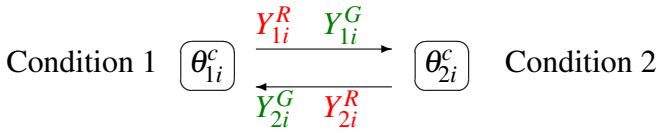

Figure 1: Color flip experiment (self-self experiment): two microarrays (Array 1:⟶ and Array 2:⟵) and two biological conditions; θ^c_{1i} and θ^c_{2i} are the unknown concentrations for the condition 1 *and* 2 *for a given gene i; (Y^R_{1i},Y^G_{1i}), and (Y^R_{2i},Y^G_{2i}) pair of intensity measurements on array* 1 *and* 2 *respectively.*

infer relevant information on a statistically sound basis. However, instead of just repeating the experiments exactly, a more reliable approach is to perform colour flip experiments (also called dye swap experiments). In colour flip experiments, the same test and reference conditions are measured once more as a repeat on a second array but the dyes are swapped, i.e., on the second array, the test condition is labeled with Cy3 (green dye) while the corresponding reference condition is labeled with Cy5 (red dye). This allows better compensating for dye specific biases, to the extent that these biases are repeatable across slides. Generally, color flipped pairs are recommended whenever possible (Yang and Speed, 2002).

3 The hybridization and dye saturation models

Engelen *et al.* (2006) proposed an analysis in two stages, based on the hybridization model and the dye saturation model. The hybridization model reflects the binding of fluorescently labeled target mRNA (i.e. the quantities we want to measure) to its corresponding probes on the microarray. In general, the more target mRNA present, the more will hybridize to the microarray. This reaction however, is heavily dependent on several experimental factors, such as hybridization efficiency of the target and quality of the spotted probes; the hybridization model attempts to captures these dependencies.

The hybridization model includes spot-specific errors, used to explain the large observed variations of absolute intensities for a given spike concentration (Engelen *et al.*, 2006). The relation between the amount of hybridized target (x_s) and the concentration of the corresponding transcript in the hybridization solution (x_0) is modeled by the steady state of the reaction

$$x_0 + s \overset{K}{\Longleftrightarrow} x_s,$$

in which the unknown hybridization constant K is assumed to be equal for all spots on a single microarray. It is also assumed that the hybridization is a first order reaction, and that x_0 is in excess (i.e. x_0 is constant). This ensures that the amount of hybridized target at the end of the reaction only depends on the initial concentration in

the hybridization solution. The amount of probe of a spot(s) available for hybridization will decrease with an increasing amount of hybridized target $x_s (s = s_0 - x_s, s_0$ being the spot size or maximal amount of available probe). Hence, at thermodynamic equilibrium, we have that

$$x_0 = \frac{x_s}{K(s_0 - x_s)}. \tag{1}$$

The spot capacity s_0 is assumed to follow a certain distribution around an average spot capacity μ. More specifically, it is assumed that $s_0 = \mu e^{b_s}$ with $b_s \sim N(0, \sigma_s^2)$. The parameters μ and σ_s are assumed to be equal for all measurements of a single array. Finally, it is assumed that the presence of distinct labels (green: Cy3 and red: Cy5) does not influence the hybridization efficiency of the differentially labeled target transcripts, i.e.,

$$\frac{x_{0,R}}{x_{0,G}} = \frac{x_{s,R}}{x_{s,G}}, \tag{2}$$

where $x_0 = x_{0,R} + x_{0,G}$, and $x_s = x_{s,R} + x_{s,G}$. It follows from (1) and (2) that the expression for the hybridized target $x_{s,G}$ for a certain label (eg. green) is given by

$$x_{s,G} = \frac{x_{0,G} s_0}{(x_{0,G} + x_{0,R} + \frac{1}{K})},$$

and for ratio controls where $x_{0,G}/x_{0,R} = 1:1$ this expression reduces to

$$x_s = \frac{K x_0 \mu e^{b_s}}{(K 2 x_0 + 1)}.$$

The dye saturation model describes the relation between the measured intensity y and the amount of labeled target x_s, hybridized to a single spot on the microarray. It is a simple linear equation incorporating an additive and multiplicative intensity error. This type of function stems from analytical chemistry (Rocke and Durbin, 2001) and has already been used in other normalization strategies (Durbin and Rocke, 2003; Durbin *et al*., 2002; Durbin and Rocke, 2004; Huber *et al*., 2002; Rocke and Durbin, 2003).

More specifically, it is assumed that

$$y = p_1 x_s e^{b_m} + p_2 + \varepsilon,$$

in which ε and b_m are additive and multiplicative intensity errors, assumed to be normally distributed, with means zero, and with variances σ^2 and σ_m^2, respectively.

It follows from the combination of the hybridization and the dye saturation models that the relation of the intensities y_G and y_R for the green and red label, measured

DOI: 10.2202/1544-6115.1401

on a single spot s_0 of a given array, and the amount of corresponding target x_0, is given by

$$y_R = p_{1,R}\left[\frac{x_{0,R}\mu e^{b_s}}{x_{0,G}+x_{0,R}+\frac{1}{K}}\right]e^{b_m}+p_{2,R}+\varepsilon_R, \quad (3)$$

$$y_G = p_{1,G}\left[\frac{x_{0,G}\mu e^{b_s}}{x_{0,G}+x_{0,R}+\frac{1}{K}}\right]e^{b_m}+p_{2,G}+\varepsilon_G, \quad (4)$$

Note that it is implicitly assumed that the spot-specific errors b_s and b_m are common for both dyes, which is not assumed for the errors ε_G and ε_R. Strictly speaking, this implies that models (3) and (4) are not identified which is why, later on, the models will be formulated in terms of $b_{sm} = b_s + b_m$, for new random terms b_{sm}, normally distributed with mean zero, and variance denoted by σ^2_{sm}. Similarly, $p_{1,R}$, $p_{1,G}$, and μ are not identified separately, while the products $\mu p_{1,R}$ and $\mu p_{1,G}$ are. Therefore, from now on, μ will be restricted to $\mu = 10^5$.

Figure 2 shows the graphical representation of the model. The Intercept $p_{2,R}$ is the lower saturation limit, $p_{1,R}$ is the steepness parameter and upper saturation limit is given by $p_{1,R}\left(\frac{\mu e^{b_s}e^{b_m}}{1+\frac{1}{K}}\right)+p_{2,R}$. The vertical downward arrow represents the concentration corresponding to the inflection point. The inflection point is obtained as half of the difference between upper and lower saturation above the lower saturation. The slope at the inflection point is $\frac{p_{1,R}}{2}$.

4 A Non-linear mixed model for parameter estimation

Engelen *et al.* (2006) fitted the models using separate estimation procedures. First, the standard deviations of ε_R and ε_G are estimated by the sample standard deviations of the intensities at the zero concentration level, while standard errors for multiplicative errors (σ_m) are estimated by performing orthogonal regression on a selected set of data points. Estimates for all other parameters are subsequently derived optimizing an objective function, in which the above mentioned standard deviations are replaced by their estimates. Such a two-stage approach does not account for the uncertainty associated with the a priori estimation of the standard deviations, and therefore cannot be used for the calculation of standard errors for the various parameter estimates nor for the construction of confidence intervals for calibrated concentrations.

We therefore propose to directly fit the models (3) and (4), within the framework of non-linear mixed models. Within this framework we can easily estimate the model parameters along with their standard errors, needed to construct the required calibration intervals. Moreover, the proposed non-linear mixed effects model

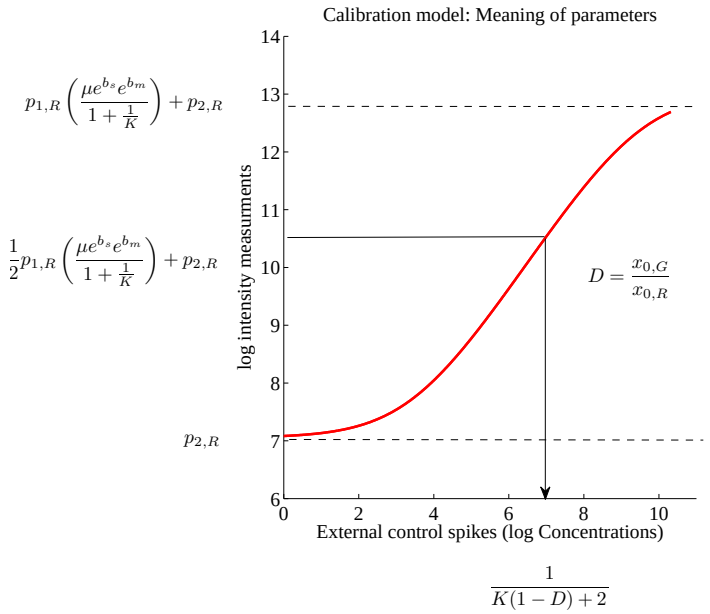

Figure 2: Calibration model: illustration of model shape (one dye only). The solid red curve represents the relationship between concentration and intensity if all errors were zero. The horizontal dashed lines indicate the upper and lower saturation levels. The vertical downward arrow represents the concentration corresponding to the inflection point. The inflection point is obtained as half of the difference between upper and lower saturation above the lower saturation. The slope at the inflection point is $\frac{p_{1,R}}{2}$.

combines all the steps in the previous approach (Engelen *et al.*, 2006) into one single model. A key feature of these models is that, by introducing random effects in addition to fixed effects, they allow us to correctly account for multiple source of variation. For instance in this study it allows to account both within- and between-spot variation. Mixed models are also a powerful class of models used for the analysis of correlated data. Here, we incorporate the correlation between the two intensity measurements which were measured on the same spot by introducing the spot-specific random effect.

Let y_{ij} denote the intensity for spot i of dye j, $i = 1,...,N$, $j \in \{R,G\}$. Models (3) and (4) then become

$$y_{ij} = p_{1,j}\left[\frac{x_{0,ij}\mu e^{b_{sm,i}}}{x_{0,iG}+x_{0,iR}+\frac{1}{K}}\right]+p_{2,j}+\varepsilon_{ij},. \quad (5)$$

DOI: 10.2202/1544-6115.1401

which is a special case of the general linear mixed model, defined as any model of the form

$$y_{ij} = f(X_{ij}, \beta, b_i) + \varepsilon_{ij}, \tag{6}$$

with X_{ij} a design matrix of known within-spot covariates, b_i a q-dimensional vector of spot-specific parameters, β an r-dimensional vector of unknown array-specific parameters, and with $f(\cdot)$ any known function. The b_i are assumed normally distributed with mean vector 0 and covariance D, independently of the errors ε_{ij} which are assumed independent normal, with mean 0 and variance σ^2.

As will be illustrated in Section 7, the standard deviation of the errors is in our context often proportional to the mean, in contrast to the homoscedasticity assumption made by model (6). We therefore adapt model (6) as suggested by (Gilberg and Urfer, 1999), i.e., by assuming

$$y_{ij} = f(x_{ij}, \beta, b_i) + f(x_{ij}, \beta, b_i)\varepsilon_{ij}. \tag{7}$$

Applying log transformation on both sides, model (7) can be rewritten as

$$y^*_{ij} = f^*(x_{ij}, \beta, b_i) + \varepsilon^*_{ij}, \tag{8}$$

in which $y^*_{ij} = \ln(y_{ij})$, $f^*(\cdot) = \ln[f(\cdot)]$, and $\varepsilon^*_{ij} = \ln(1+\varepsilon_{ij})$. The final models then become

$$\ln(y_{ij}) \quad = \quad \ln\left(p_{1,j}\left[\frac{x_{0,ij}\mu e^{b_{sm,i}}}{x_{0,iR} + x_{0,iG} + \frac{1}{K}} \right] + p_{2,j} \right) + \varepsilon^*_{ij} \tag{9}$$

for which we assume $b_{sm,i} \sim N(0, \sigma^2_{sm})$, $\varepsilon^*_{ij} \sim N(0, \sigma^2_a)$, $j \in \{R, G\}$. Note that this model implicitly replaces the normality assumption for the original error components ε_{ij} by the assumption of log-normality, while the mean of the original errors ε_{ij} now becomes $\exp(\sigma^2_a/2) - 1$ and their variance σ^2 is given by $\exp(\sigma^2_a)[\exp(\sigma^2_a) - 1]$.

Standard maximum likelihood principles can be used to estimate the parameters in the above models, and to calculate associated standard errors. After suitable reparameterizations for numerical stability, the models can easily be implemented within the SAS procedure NLMIXED. As suggested by (Davidian and Giltinan, 2002), adaptive Gaussian quadrature is used for the approximation of the likelihood, while the optimization is based on the Newton-Raphson algorithm.

5 The calibration procedure

The main purpose of our analysis is to draw inferences about the concentration levels $x_{0,ij}$ corresponding to measured intensities y_{ij}. Inversion of (9) cannot be used

as this would require knowledge of the spot-specific error $b_{sm,i}$, which is not known for the spots without known concentrations. Therefore, calibration will have to be based on average calibration curves, rather than spot-specific calibration curves.

The spot-specific calibration curve, implied by model (9) is given by

$$E[Y_{ij}|b_{sm,i}] = \left\{ p_{1,j} \left[\frac{x_{0,ij}\mu e^{b_{sm,i}}}{x_{0,iR} + x_{0,iG} + \frac{1}{K}} \right] + p_{2,j} \right\} \exp(\sigma_a^2/2), \tag{10}$$

in which $p_{1,j}$, K, $p_{2,j}$ are replaced by their maximum likelihood estimates, and $b_{sm,i}$ by the empirical Bayes estimate (Molenberghs and Verbeke, 2005 Chapter 19). For spots without known concentrations, $b_{sm,i}$ cannot be estimated, and calibration will be based on the average curve, given by

$$\begin{aligned} E[Y_{ij}] &= E\left[E[Y_{ij}|b_{sm,i}]\right] = \left\{ E\left[p_{1,j} \left[\frac{x_{0,ij}\mu e^{b_{sm,i}}}{x_{0,iR} + x_{0,iG} + \frac{1}{K}} \right] + p_{2,j} \right] \right\} \exp(\sigma_a^2/2) \\ &= \left\{ p_{1,j} \left[\frac{x_{0,ij}\mu e^{\frac{\sigma_{sm}^2}{2}}}{x_{0,iR} + x_{0,iG} + \frac{1}{K}} \right] + p_{2,j} \right\} \exp(\sigma_a^2/2). \end{aligned} \tag{11}$$

As described by Davidian and Giltinan (2002), the concentration corresponding to a intensity can now be estimated by inverting the average curve (11). Let β_j denote the parameters in (11), i.e., $p_{1,j}$, $p_{2,j}$, K, σ_a^2, and σ_{sm}^2. We then have

$$x_{0,ij} = h(y_{ij}, \beta_j) \equiv \frac{1}{K} \left[\frac{y_{ij} e^{-\frac{\sigma_a^2}{2}} - p_{2,j}}{p_{1,j}\mu e^{\frac{\sigma_{sm}^2}{2}} - 2(y_{ij} e^{-\frac{\sigma_a^2}{2}} - p_{2,j})} \right] \tag{12}$$

In practice, an estimate for $x_{0,ij}$ will be obtained from replacing β_j in (12) by the maximum likelihood estimate $\widehat{\beta}_j$, i.e., $\widehat{x}_{0,ij} = h(y_{ij}, \widehat{\beta}_j)$. Observations outside the calibration limits need to be handled with care. If the new observation y_{ij} is below the lower saturation limit of $p_{2j}\exp(\sigma_a^2/2)$ then $\widehat{x}_{0,ij}$ is set equal to the observed minimum concentration. In case y_{ij} is above the upper saturation limit of $[p_{1j}\mu\exp(\sigma_{sm}^2/2) + p_{2j}]\exp(\sigma_a^2/2)$, then $\widehat{x}_{0,ij}$ is set equal to the observed maximum concentration. In case multiple observations y_{ij} are available to estimate $x_{0,ij}$, y_{ij} in (12) is replaced by their average $\overline{y}_{ij}$.

6 The construction of confidence intervals

In order to reflect the uncertainty in $\widehat{x}_{0,ij}$, due to replacing the parameter β_j in (12) by the maximum likelihood estimate $\widehat{\beta}_j$, an approximate $(1-\alpha)100\%$ confidence

DOI: 10.2202/1544-6115.1401

Table 1: Color flip experiment: pooled concentration

Conditions	Array 1	Array 2	Pooled Estimate
Condition 1	Red	Green	
	$\widehat{x}_{0,iR,1} = h(\bar{y}_{ij,1}, \widehat{\beta}_{R,1})$	$\widehat{x}_{0,iG,2} = h(\bar{y}_{ij,2}, \widehat{\beta}_{G,2})$	$\widehat{\theta}^c_{1i} = \frac{\widehat{x}_{0,iG,2} + \widehat{x}_{0,iR,1}}{2}$
Condition 2	Green	Red	
	$\widehat{x}_{0,iG,1} = h(\bar{y}_{ij,1}, \widehat{\beta}_{G,1})$	$\widehat{x}_{0,iR,2} = h(\bar{y}_{ij,2}, \widehat{\beta}_{R,2})$	$\widehat{\theta}^c_{2i} = \frac{\widehat{x}_{0,iG,1} + \widehat{x}_{0,iR,2}}{2}$

interval can be calculated, using a first order Taylor series expansion of the inverse regression function $h(y_{ij}, \beta_j)$. Let $\bar{y}_{ij}$ represent the average of n replicates for the j^{th} dye of a given array, estimating the average $\mu_{ij} = \mu(x_{0,ij}, \beta_j) = E[Y_{ij}]$. We then have

$$\widehat{x}_{0,ij} \equiv h(\bar{y}_{ij}, \widehat{\beta}_j) \approx h(\mu_{ij}, \beta_j) + h_y(\mu_{ij}, \beta_j)(\bar{y}_{ij} - \mu_{ij}) + h^T_\beta(\mu_{ij}, \beta_j)(\widehat{\beta}_j - \beta_j), \quad (13)$$

in which $h_\beta(\mu_{ij}, \beta_j)$ is the gradient vector of $h(y, \beta)$ with respect to β and where $h_y(\mu_{ij}, \beta_j)$ is the first order derivative of $h(y, \beta)$ with respect to y, both evaluated at $y = \mu_{ij}$ and $\beta = \beta_j$. Denoting the covariance matrix of $\widehat{\beta}_j$ by $\Sigma(\widehat{\beta}_j)$, it immediately follows from (13) that an approximation to the standard error of $\widehat{x}_{0,ij}$ is given by

$$\text{s.e.}(\widehat{x}_{0,ij}) \approx \sqrt{h^2_y(\mu_{ij}, \beta_j)\text{Var}(\bar{y}_{ij}) + h^T_\beta(\mu_{ij}, \beta_j)\Sigma(\widehat{\beta}_j)h_\beta(\mu_{ij}, \beta_j)}. \quad (14)$$

Note that an estimate for $\Sigma(\widehat{\beta}_j)$ follows from the maximum likelihood estimation procedure used to fit the non-linear mixed model. Further, we have that $\text{Var}(\bar{y}_{ij}) = \text{Var}(y_{ij})/n$ in which $\text{Var}(y_{ij})$ immediately follows from (5) to be equal to

$$\text{Var}(y_{ij}) = (1 + \sigma^2)Q^2[e^{\sigma^2_{sm}} - 1] + \sigma^2[Q + p_{2j}]^2$$

with

$$Q = \frac{p_{1j}x_{0,ij}\mu e^{\frac{\sigma^2_{sm}}{2}}}{\left(x_{0,iR} + x_{0,iG} + \frac{1}{K}\right)},$$

and with, as before, $\sigma^2 = \exp(\sigma^2_a)[\exp(\sigma^2_a) - 1]$. An estimate $\widehat{\text{s.e.}}(\widehat{x}_{0,ij})$ for $\text{s.e.}(\widehat{x}_{0,ij})$ is obtained from replacing β_j by $\widehat{\beta}_j$, σ^2_a by $\widehat{\sigma}^2_a$, and μ_{ij} by $\bar{y}_{ij}$ in (14).

Assuming $\widehat{x}_{0,ij}$ to be approximately normally distributed, an approximate Wald-type $(1-\alpha)100\%$ confidence interval for $x_{0,ij}$ is obtained by $\widehat{x}_{0,ij} \pm z_{\alpha/2}\,\widehat{\text{s.e.}}(\widehat{x}_{0,ij})$,

in which $z_{\alpha/2}$ is the appropriate quantile of the standard normal distribution. Often, concentrations in biological experiments are positively skewed (Davidian and Giltinan, 2002). We therefore propose to construct the confidence interval assuming normality for $\ln(\widehat{x}_{0,ij})$ instead of $\widehat{x}_{0,ij}$. An estimate for the standard error of $\ln(\widehat{x}_{0,ij})$ then easily follows from the transformation theorem (Serfling 1980, Chapter 3). In cases where $\bar{y}_{ij}$ is outside the calibration limits, no confidence interval can be calculated.

As explained in Section 1 the major objective of this study is to estimate the absolute expression levels (i.e., unknown concentrations) for every gene in each of the tested biological conditions in a color flip experiment. As depicted in Table 1, such experiments yield one concentration for each condition, but each is estimated twice, once from the Green channel, and once from the Red channel. As different arrays are assumed independent, the two estimates can easily be combined by averaging (also shown in Table 1), and the standard error of the combined estimate can easily be obtained from the standard errors of the original individual components.

7 Application

7.1 Analysis of dye-swap experiment

To illustrate the applicability of our method we used a two publicly available datasets. First, we consider a dye-swap experiment based on data from (Hilson *et al.*, 2004). This experiment contains the necessary spots for measuring external control spikes, which are required for estimating the parameters of our model. A series of external controls (Lucidea Universal Scorecard; Amersham Biosciences) consisted of 10 calibration spikes (added to the hybridization solution in a ratio 1:1 and spanning up to 4.5 orders of magnitude), eight ratio spikes provided at both low and high concentration and two negative controls, each of them spotted once per pin group, resulting in a total of 24 repeats of each spike probe per array. The experimental design included only a single biological condition (self-self experiment; all hybridizations were conducted with the same RNA sample, extracted from aerial parts of germinating *Arabidopsis thaliana* seedlings).

Measured intensities and observed concentrations for red and green channels for this experiment are given in Figure 3. The black dots shown in the left panel of Figure 3 are the spots where the concentration ratios between green and red are 1:1. The right panel of Figure 3 depicts the nonlinear relationship between measured intensities and concentrations. However, in our analysis, both dynamic (X:Y) and ratio (1:1) control spikes will be used to estimate the model parameters. Figure 4 shows the variance of the intensities at unique concentration levels versus the true concentrations in original and log scale for arrays 1 and 2. This reveals that the

DOI: 10.2202/1544-6115.1401

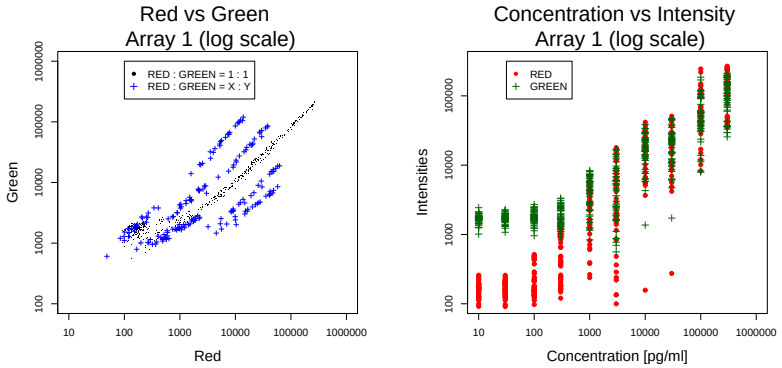

Figure 3: External control spikes (Dye-swap experiment): Genes with known concentration (left panel) and Intensity versus concentration for ratio controls (R:G =1:1) concentrations (pg/ml) of target transcripts in the hybridization solution for all external control spikes (right panel), on log scale. However both even (R:G =1:1) and uneven (R:G =X:Y) spike controls are used in the analysis.

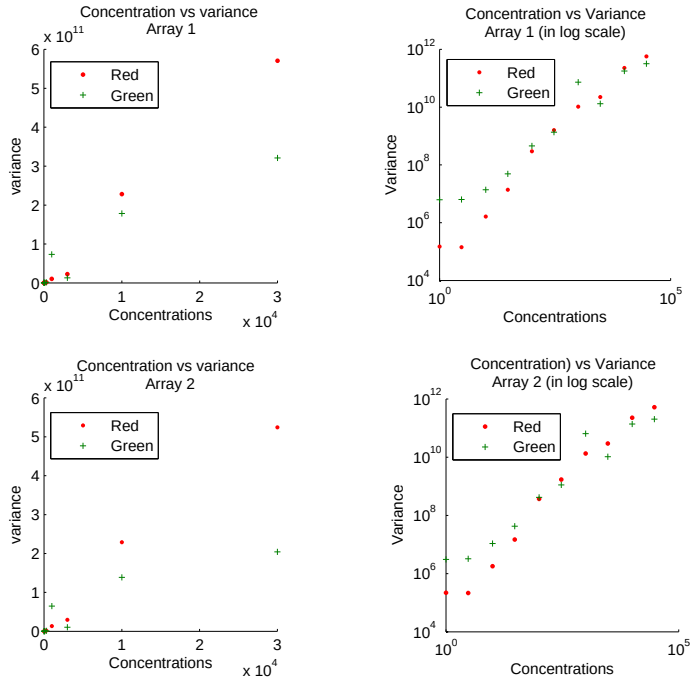

Figure 4: Dye-swap experiment: Variance at unique concentration levels for external control spike data for which the concentration for Array 1 and 2 were known: (left panels in original scale and right panels in log scale)

Table 2: *Results of dye-swap experiment. Maximum Likelihood estimators for Original parameters and Standard errors via Delta Method : Array 1 and 2*

	Array 1		Array 2	
Parameter	Estimate	(s.e.)	Estimate	(s.e.)
K	0.030	(0.009)	0.047	(0.012)
$p_{1,R}$	0.978	(0.076)	0.998	(0.082)
$p_{2,R}$	0.927	(0.019)	0.877	(0.021)
$p_{1,G}$	0.835	(0.013)	0.717	(0.013)
$p_{2,G}$	0.930	(0.011)	0.727	(0.010)
σ^2_{sm}	1.363	(0.108)	1.506	(0.119)
σ^2_a	0.028	(0.002)	0.036	(0.002)
$-2\times$ log-likelihood	811.3		1020.6	

variance changes with the mean leading to heterogeneous errors, motivating the use of the heteroscedastic model (9).

7.2 Model fitting and informal model assessment

Models (9) were fitted using the SAS procedure NLMIXED, using numerical quadrature methods for the approximation of the likelihood (Molenberghs and Verbeke, 2005 Chapter 14). It is wellknown that, in the context of nonlinear mixed-effects models, optimization procedures can be sensitive to the parameter values from which the iteration is initiated. Therefore, the models were fitted repeatedly, for a variety of starting values. Also, various numbers of quadrature points were used (3, 5, 11, 21, 25, 31 and 35), and likelihoods and parameter estimates were found to be stable. Table 2 summarizes the maximum likelihood estimates and their associated standard errors, for both arrays separately.

Based on these results, the implied fitted marginal (average) relations (11) were calculated and are shown in Figure 5, for both arrays. The original observations are also added. Note the sigmoid shape of the estimated curves reflecting the nonlinear relationship between concentrations and the intensity measurements. Further, empirical Bayes estimates $\widehat{b}_{sm,i}$ of $b_{sm,i}$ can be used to calculate spot-specific residuals

$$\widehat{\varepsilon}^*_{ij} = y^*_{ij} - f^*(x_{ij}, \widehat{\beta}_j, \widehat{b}_{sm,i}),$$

for which graphical exploration did not reveal any deviations from the normality assumption made in our model (Figure not shown).

DOI: 10.2202/1544-6115.1401

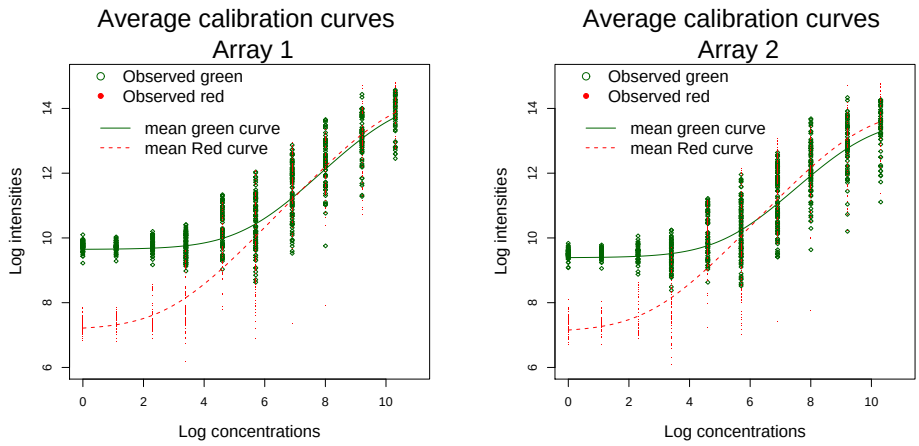

Figure 5: Dye-swap experiment: Solid line indicates estimated marginal curve for green dye, the broken lines indicate marginal curve for the red dye, [0] and [.] show original intensities in log scale for green and red dyes respectively. (Left panel: Array 1 and Right panel: Array 2)

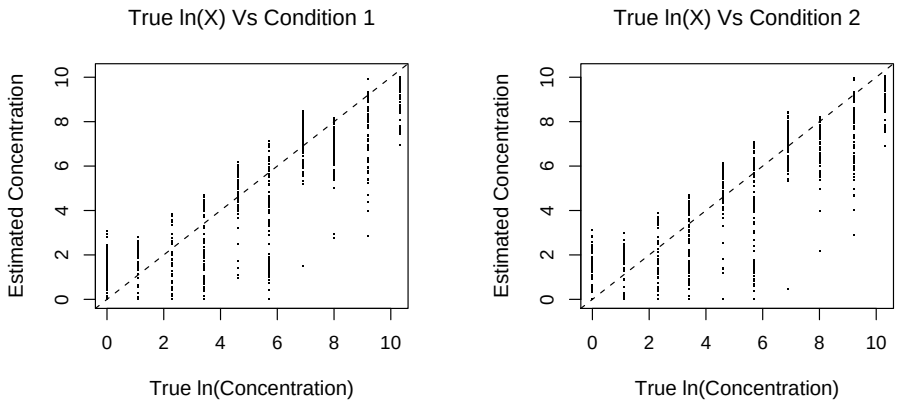

Figure 6: Dye-swap experiment: Estimated versus true concentration for conditions C1 and C2, on log-scale.

7.3 Calibration on external control spike data

In order to investigate the accuracy of the proposed calibration method, we first apply the procedure on spike data where we know the true concentration values. In this example, the two conditions C1 and C2 represent the same biological mRNA

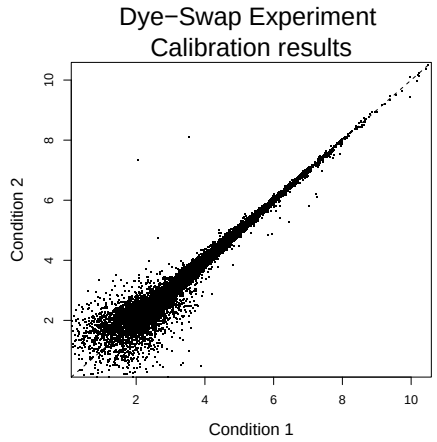

Figure 7: Dye-swap experiment: Calibration estimates by nonlinear mixed-effects model: Estimated concentration on log scale: biological conditions C1 vs C2. The dashed line indicates the bisector.

samples. Figure 6 shows the estimated versus true concentrations, on log-scale, for both conditions separately. Note that very small true concentrations tend to be overestimated, while the opposite is true for very large true concentrations. This bias results from intensity levels below and above the lower and upper saturation limits respectively (Section 5).

Further, we used the estimated model parameters to obtain absolute expression levels (the unknown concentrations) for every gene in each of the tested biological conditions in a color flip experiment. Figure 7 shows the estimated concentration on log scale for biological conditions 1 and 2. Because C1 and C2 represent the same biological condition, all estimates are centered along the first bisector. Figure 8 shows estimated 95% confidence intervals for each calibrated concentrations, for both conditions separately. Similar to the bias observed before for the estimates, we find wide confidence intervals for extreme estimated concentrations, due to observed intensities outside the calibration limits.

7.4 Analysis of loop design experiment

So far we have applied our method to a simple dye-swap experiment. The method in itself however, is capable of tackling real world designs of any complexity. To illustrate, we analyzed a more complex spiked-in cDNA microarray dataset published by Smets *et al.* (2008). A total of 5928 genes were measured for three different yeast strains (Wild-Type, Rim-mutant, and Sch9-mutant) across five consecutive

DOI: 10.2202/1544-6115.1401

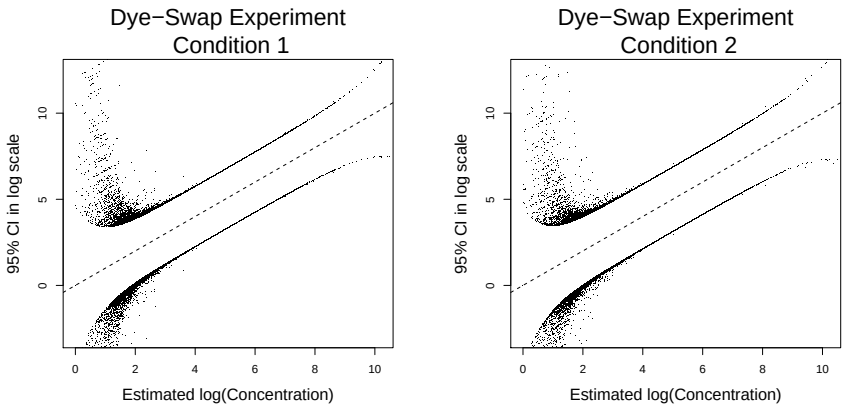

Figure 8: Dye-swap experiment: Confidence intervals versus calibration estimates (in log scale) for biological condition 1 *(left panel) and* 2 *(right panel) for genes of which the concentration in the hybridization solution was unknown.*

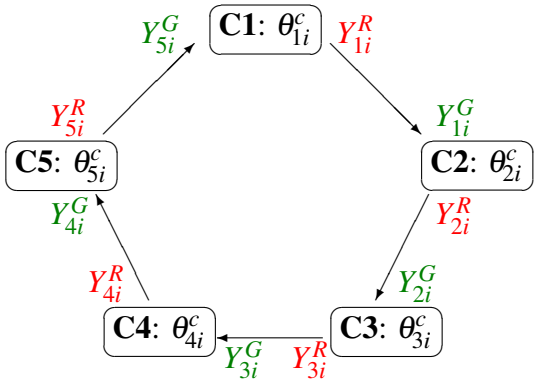

Figure 9: Loop design (Yeast data: Wild-Type Strain). 5 microarrays (Array 1: $\searrow$*, Array 2:* $\swarrow$*, Array 3:* $\longleftarrow$*, Array 4:* $\nwarrow$*, and Array 5:* $\nearrow$*) and 5 biological conditions;* θ^c_{1i}*,* θ^c_{2i}*,* θ^c_{3i}*,* θ^c_{4i}*, and* θ^c_{5i} *are the unknown concentrations for the condition* $1,2,3,4$*, and* 5*. For a given gene i (*Y^R_{1i}*,* Y^G_{1i}*), (*Y^R_{2i}*,* Y^G_{2i}*), (*Y^R_{3i}*,* Y^G_{3i}*),(*Y^R_{4i}*,* Y^G_{4i}*), and (*Y^R_{5i}*,* Y^G_{5i}*) are pair of intensity measurements on array* $1,2,3,4$*, and* 5 *respectively. Each array shares two conditions and each condition shares two arrays.*

time points after adding rapamycin (0, 15, 30, 60, and 120 min), resulting in a total of 15 different biological conditions. A loop design was applied, with one loop for the Wild-Type (see Figure 9), one for Rim samples and one for the Sch9 samples, resulting in two measurements for each of the 15 RNA samples. All the arrays in this experiment were outfitted with the Lucidea^{TM} Universal Scorecard^{TM} (Amer-

sham Biosciences), generating a series of external control spikes. For more detailed information regarding the experimental protocol we refer to Smets *et al.* (2008).

Using the data from the external control spikes, our proposed calibration model was fitted for each array. The maximum likelihood estimators for model parameters along with their standard errors are tabulated in Table 3. Based on those parameter estimates absolute value expression levels have been estimated along with their standard errors for all genes under all 15 conditions surveyed in the experiment. We give an overview of our main findings here. Figure 10 shows estimated expression profiles along with their 95% confidence bands for randomly selected four genes. Interestingly, these expression profiles and their confidence bands can be used to answer relevant biological questions. For instance, downstream analysis such as selecting differentially expressed genes and identifying corresponding biologically important conditions and strains can benefit from these results.

8 Simulation study

The confidence intervals discussed in Section 6 are based on Taylor series expansions which are asymptotic in nature and are therefore only guaranteed to perform well in very large (infinite) samples. The only way to check their accuracy in more realistic settings is in a simulation study, where the following steps are followed. First, data are simulated from a model that is sufficiently realistic to cover real data situations. In our paper, we used the model that was found to be a reasonable description of the data analysed in Section 7. Second, the model is fitted to the simulated data and the approximate confidence intervals are computed. When confidence intervals have the correct 95% coverage, we expect that 5% of them would not contain the correct values (which are known in the simulated data). In order to estimate the realized coverage of the intervals, the simulation is repeated many times and it is counted how often the constructed intervals contain the true values. In Section 8.1, we will do this in the context of the dye-swap experiment from Section 7. Afterwards, in Section 8.2, we will repeat the simulation for a more complex design, i.e. a loop design.

8.1 Simulation for the dye-swap experiment

We generated 1000 data sets from model (9), with all parameters replaced by the estimates obtained in Section 7 for the original data in array 1. As concentration levels, we also used the original concentrations from the spike data, hence all generated data sets are of the same structure as the original data. For each data set, model (9) was fitted, and the calibration procedure was applied. All calculations were performed using the SAS software version 9.1 (SAS Institute, North Carolina, US).

DOI: 10.2202/1544-6115.1401

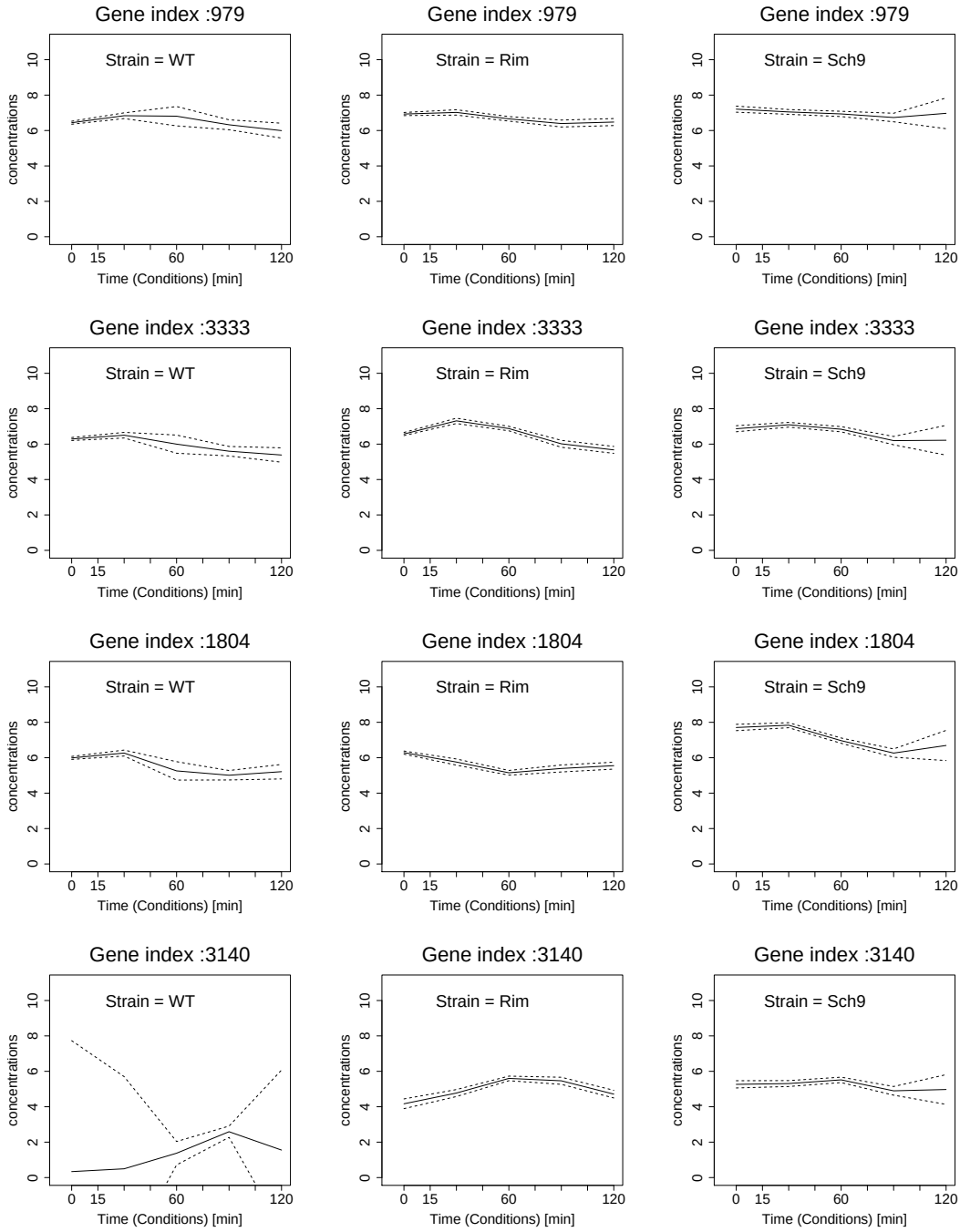

Figure 10: Analysis of yeast data. Estimated expression profiles along with their 95% confidence bands for four randomly selected genes. WT: Wild-Type, Rim: Rim-mutant, and Sch9: Sch9-mutant

Table 3: *Results of yeast analysis: Maximum likelihood estimators (standard errors)*

Strain	Parameter	Microarray number									
		1		2		3		4		5	
		Estimate	(s.e.)	Estimate	(s.e.)	Estimate	(s.e.)	Estimate	(s.e.)	Estimate	(s.e.)
WT	K	0.050	(0.007)	0.036	(0.006)	0.058	(0.007)	0.056	(0.008)	0.042	(0.006)
	$p_{1,R}$	5.304	(0.320)	4.931	(0.296)	5.762	(0.301)	7.966	(0.457)	7.580	(0.454)
	$p_{2,R}$	5.062	(0.216)	4.459	(0.171)	3.810	(0.132)	6.012	(0.221)	8.236	(0.371)
	$p_{1,G}$	1.128	(0.068)	0.919	(0.048)	1.144	(0.050)	0.990	(0.046)	1.263	(0.079)
	$p_{2,G}$	6.703	(0.273)	4.702	(0.173)	4.926	(0.164)	3.733	(0.142)	5.960	(0.282)
	σ^2_{sm}	1.213	(0.397)	2.979	(0.530)	2.290	(0.359)	2.748	(0.363)	0.540	(0.317)
	σ^2_a	0.509	(0.023)	0.394	(0.019)	0.310	(0.015)	0.346	(0.016)	0.594	(0.026)
		6		7		8		9		10	
	K	0.020	(0.005)	0.024	(0.004)	0.028	(0.005)	0.029	(0.005)	0.023	(0.004)
	$p_{1,R}$	3.717	(0.258)	2.837	(0.173)	4.051	(0.230)	3.180	(0.178)	4.553	(0.253)
	$p_{2,R}$	4.750	(0.240)	3.512	(0.153)	3.369	(0.142)	3.173	(0.121)	5.019	(0.224)
Rim	$p_{1,G}$	1.004	(0.076)	1.087	(0.070)	0.936	(0.054)	0.850	(0.045)	0.936	(0.059)
	$p_{2,G}$	4.497	(0.227)	3.482	(0.151)	3.909	(0.158)	2.699	(0.100)	3.212	(0.148)
	σ^2_{sm}	1.310	(0.530)	1.185	(0.407)	1.824	(0.372)	1.922	(0.348)	0.433	(0.308)
	σ^2_a	0.791	(0.035)	0.596	(0.025)	0.496	(0.021)	0.397	(0.017)	0.653	(0.026)
		11		12		13		14		15	
	K	0.024	(0.005)	0.025	(0.005)	0.029	(0.005)	0.027	(0.005)	0.019	(0.004)
	$p_{1,R}$	3.765	(0.250)	3.288	(0.213)	3.326	(0.194)	6.742	(0.401)	2.439	(0.155)
	$p_{2,R}$	5.541	(0.251)	4.919	(0.211)	3.235	(0.132)	8.720	(0.294)	2.543	(0.124)
Sch9	$p_{1,G}$	0.735	(0.052)	0.788	(0.053)	0.809	(0.046)	0.731	(0.036)	0.973	(0.069)
	$p_{2,G}$	3.896	(0.175)	3.504	(0.150)	2.686	(0.107)	5.718	(0.192)	3.208	(0.151)
	σ^2_{sm}	1.236	(0.475)	1.358	(0.431)	1.906	(0.385)	3.521	(0.671)	1.521	(0.520)
	σ^2_a	0.649	(0.028)	0.585	(0.025)	0.466	(0.020)	0.355	(0.019)	0.741	(0.032)

DOI: 10.2202/1544-6115.1401

Table 4: Simulation study results: Summary of parameter estimates

	K	$p_{1,R}$	$p_{2,R}$	$p_{1,G}$	$p_{2,G}$	σ^2_{sm}	σ^2_a
True values ψ	0.0303	0.9783	0.9269	0.8348	0.9301	1.3628	0.0282
$\overline{\widehat{\psi}} = \frac{1}{1000}\sum_s \widehat{\psi}_s$	0.0339	1.0130	0.8504	0.8780	0.8863	2.0576	0.0127
Bias$= \overline{\widehat{\psi}} - \psi$	-0.0035	-0.0347	0.0765	-0.0432	0.0437	-0.6948	0.0155
RMSE	0.0104	0.0825	0.0190	0.0722	0.0092	0.1446	0.0013

The models (9) were fitted with the procedure NLMIXED, using Newton-Raphson optimization, adaptive Gaussian Quadrature with 21 quadrature points, and with 10^{-15} as convergence criterion for absolute parameter values.

Table 4 summarizes the results from fitting the models (9) to the 1000 generated data sets, for the various parameters in the model. For each parameter, let ψ denote the true value, as reported in Table 2 (Array 1). Then each generated data set yields an estimate $\widehat{\psi}_s$ for ψ, $s = 1,\ldots,1000$. The average estimate $\overline{\widehat{\psi}}$ is obtained from averaging the estimates $\widehat{\psi}_s$ over all simulations. The bias $\overline{\widehat{\psi}} - \psi$ reflects the systematic error in the estimation of ψ, while the root mean squared error RMSE $= \sum_s (\widehat{\psi}_s - \psi)^2/1000$ reflects the average quadratic difference between the estimated and true paramater value. We observe that the bias and mean squared error is small for all parameters, but somewhat larger for the random-effects variance σ^2_{sm}, where a slight overestimation seems to occur. This overestimation of variability may affect the coverage probabilities of the confidence intervals.

Figure 11 shows the estimated coverage probabilities of the confidence intervals at each level of the true concentrations. The actual coverage was calculated as the proportion of intervals (out of 1000) that included the true concentration. For all concentrations, the coverage probability is very close to the nominal 95%, indicated that our calibration method performs well in samples of size similar to our application analysed in Section 7.

8.2 Simulation for the loop design

The methodology developed can also be applied in settings far more general than the dye-swap experiments considered so far. As an illustration, we consider the so-called loop design, represented graphically in Figure 12. The loop design contains tests on four biological conditions on four arrays. The simulation study from Section 8.1 was repeated for the loop design. We now generated data for 1000 loop designs resulting in 4000 data sets (each loop design is composed of four microar-

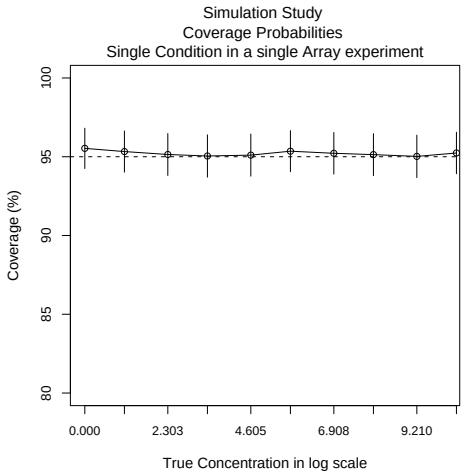

Figure 11: Simulation study based on single condition and single microarray: True Concentration vs Coverage results from 1000 simulations for the nominal level of 95%: Confidence intervals are estimated by linearizing the inverse regression function. Vertical small bars indicate the 95% confidence intervals for coverage probabilities.

rays). As in Section 8.1, calibration intervals are estimated and coverage probabilities for each of the experimental conditions are summarized and shown in Figure 13. As before, all coverage probabilities are around the nominal level of 95%, confirming that the calibration interval estimation method can easily be extended to more complex designs.

9 Discussion

We have developed a spike based normalization procedure based on a nonlinear mixed-effects model. This approach not only allows accurate estimation of absolute expression levels from cDNA arrays based on a spike derived calibration curve, but it also makes it possible to derive confidence bands for each estimated concentration level. This model incorporates parameters and error distributions representing both the hybridization of labeled target to complementary probes and the subsequent measurement of fluorescence intensities. External control spikes were used to estimate the model parameters. The estimated model parameter values are then used to obtain absolute levels of expression for the remaining genes. For each combination of a gene and a tested biological condition, a single absolute target level is estimated, while considering the specificities of the design.

The simulations conducted have shown that our proposed calibration method

DOI: 10.2202/1544-6115.1401

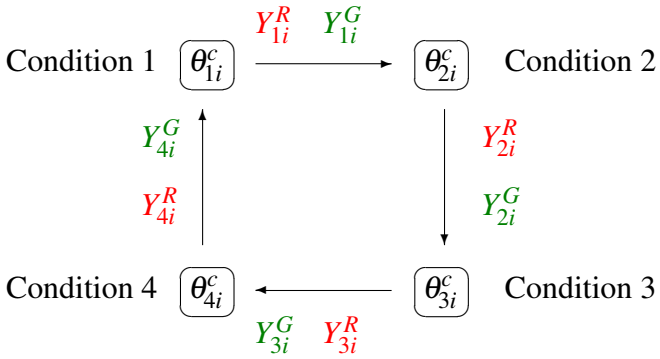

Figure 12: Loop design : four microarrays (Array 1: →, Array 2: ↓, Array 3: ← and Array 4: ↑) and four biological conditions; for a given gene i, θ^c_{1i}, θ^c_{2i}, θ^c_{3i} and θ^c_{4i} are unknown concentrations for the condition 1,2,3, and 4 respectively; (Y^R_{1i},Y^G_{1i}), (Y^R_{2i},Y^G_{2i}), (Y^R_{3i},Y^G_{3i}), and (Y^R_{4i},Y^G_{4i}) pair of intensity measurements on array 1,2,3, and 4 respectively.

yields valid estimates for unknown concentrations, with associated confidence limits that reach the requested nominal coverage levels. Although the simulation results are promising, one has to keep in mind that no good estimation of concentration levels is possible whenever the observed intensity levels are outside the calibration limits. It is expected that this problem will be more prominent as the residual variability σ_a^2 gets larger. However, if multiple intensity measures are available, calibration will be based on their average. Since such averages have smaller variability than single measurements, we then expect this bias problem to be less severe.

Our approach computes absolute expression levels together with their 95% confidence intervals, avoiding the use of intensity ratios. Moreover, for the described experiment, the estimated absolute expression levels (calibration estimates) approximate the actual concentrations fairly well. Although the calibration method was developed in the context of simple dye-swap experiments, we also illustrated that it can fairly easily be extended to more complex designs. This was done for loop designs, but others can be considered as well.

Finally, our method can also be adapted to preprocess one-color oligonucleotide microarray data (such as Affymetrix, Draghici, 2003) with a slight adjustment to our model. For cDNA arrays we have kept the hybridization constant K the same for all measurements of the same array as probe sequences for spotted microarrays are often very long specifically selected to have properties that obviate large differences in transcript specific hybridization effects. On the other hand, average spot capacities were allowed to vary across probes, because two-color cDNA arrays show large differences in intensities for probes that are effectively measuring

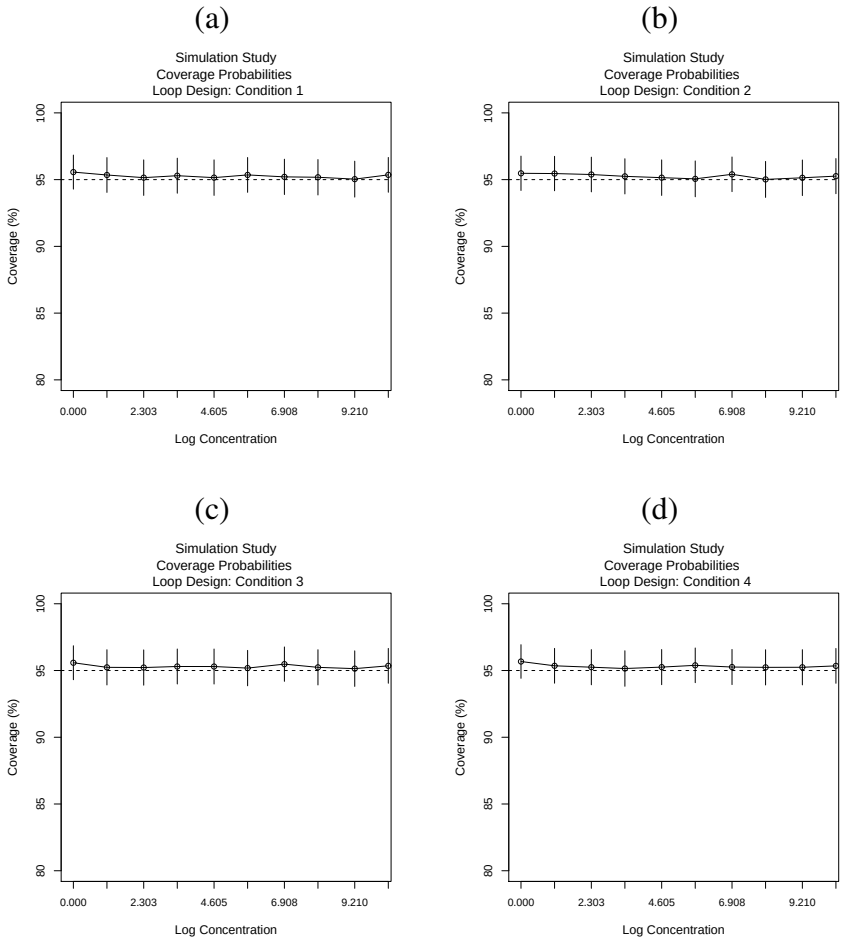

Figure 13: Simulation study based on ***Loop design****: True concentration versus coverage results from 1000 simulations for the nominal level of* 95%*. Confidence intervals are estimated by linearizing the inverse regression function (see Section 6). The broken horizontal line indicates the nominal level of* 95%*. (a) Condition* 1 *(b) Condition* 2 *(c) Condition* 3 *(d) Condition* 4

the same transcript, which are due to the large variation in spot capacities (Engelen *et al.*, 2006; Rocke and Durbin, 2001). For oligonucleotide arrays (e.g. Affymetrix), this is not much of an issue: spot capacities are a lot more homogeneous because the array production process is generally a lot more controlled than with spotted cDNA microarrays. Measured intensities however, also show a large variation for Affymetrix arrays, but due to an entirely different issue: because the small size of the probes ($\sim 25nt$), the hybridization efficiencies (reflected in the parameter K)

DOI: 10.2202/1544-6115.1401

show large differences. So for Affymetrix arrays we would keep the average spot capacity the same for all probes, but allow individual probe values for hybridization constant K to differ. This is contrary to our model for two-color arrays, where we could assume K the same for all probes in a given array. Because of the short probe length and resulting effects on hybridization efficiencies, Affymetrix arrays use on average around 15-20 probes (all with different K's) per gene, i.e. they all measure the same transcript. When we allow the parameter K to differ across the probes for a given gene, our model will be changed and will have a random effect for parameter K. Further, the original spot specific random effect $b_{sm,i}$ will be treated as a fixed effect and can be merged with parameter μ. From the statistical point of view this is still a random-effects model and hence, our general methodology for estimating calibration intervals will follow the same procedure, but will require slightly different computations. Hence absolute expression levels and confidence intervals can be obtained under the one-color microarray settings as well.

References

Badiee, A. Eiken, H. G. Steen, V. M. and Lovlie, R. (2003). Evaluation of five different cDNA labeling methods for microarrays using spike controls. *BMC Biotechnol. 3*, 23.

Bilban, M. (2002). Normalizing DNA microarray data. *Curr. Issues Mol. Biology 4*, 57-64.

Churchill, G. A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genetics 32*, 490–5.

Carroll, R. and Ruppert, D. (1988). *Transformation and Weighting in Regression (Monographs on Statistics and Applied Probability).* Chapman & Hall: New York, 10119.

Davidian, M. and Giltinan, D. M. (2002). *Nonlinear models for repeated measurement data.* Chapman & Hall: London.

Draghici, S. (2003). *Data analysis tools for DNA microarrays.* Chapman & Hall: London.

Durbin, B. P. Hardin, J. S. Hawkins, D. M. and Rocke, D. M. (2002). A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics 18*, S105–S110.

Durbin, B. P. and Rocke, D. M. (2003). Estimation of transformation parameters for microarray data. *Bioinformatics 19*, 1360–1367.

Durbin, B. P. and Rocke, D. M. (2004). Variance-stabilizing transformations for two-color microarrays. *Bioinformatics 20*, 660–667.

Dudley, A. M. Aach, J. Steffen, M. A. and Church, G. M. (2002). Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc Natl Acad Sci U S A 99(11)*, 7554–9.

Engelen, K. Naudts, B. Moor, B. D. and Marchal, K. (2006). A calibration method for estimating absolute expression levels from microarray data. *Bioinformatics 32*, 490–5.

Engelen, K. (2003). Normalizing microarray data. *Bioinformatics 19(7)*, 893–4.

Gilberg, F. and W. Urfer (1999). Heteroscedastic nonlinear regression models with random efects and their application to enzyme kinetic data. *Biometrical Journal 41 Suppl:5.*, 543–557.

Hilson, P. *et al.* (2004). Versatile gene-specific sequence tags for arabidopsis functional genomics: 145 bibliography transcript profiling and reverse genetics applications. *Genome Research 14*, 2176–2189.

Kerr, M. K. (2000). Analysis of variance for gene expression microarray data. *Journal of Computational Biology 7*, 819-837.

Leung, Y. F. and Cavalieri, D. (2003). Fundamentals of cDNA microarray data analysis. *Trends Genetics 19*, 649-659.

Molenberghs, G. and G. Verbeke (2005). *Models for discrete longitudinal data.* Springer-Verlag: New York.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics 32* (Suppl.), 496-501.

Rocke, D. M. and Durbin, B. (2001). A Model for Measurement Error for Gene Expression Arrays. *Journal of Computational Biology 8(6)*, 557–569.

Rocke, D. M. and Lorenzato, S. (1995). A 2-Component Model for Measurement Error in Analytical-Chemistry. *Technometrics 37*, 176–184.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* John Wiley & Sons, Inc..

DOI: 10.2202/1544-6115.1401

Shi, L. *et al.* (2005). Microarray scanner calibration curves: characteristics and implications. *Bioinformatics 6* (Suppl.), 2:S11.

Smets, B. De Snijder, P. Engelen, K. Joossens, E. Ghillebert, R. Thevissen, K. Marchal, K. Winderickx, J. (2008). Genome-wide expression analysis reveals TORC1-dependent and -independent functions of Sch9. *FEMS Yeast Research. 8*, 1276–1288.

Yang, Y. H. and T. Speed (2002). Design issues for cDNA microarray experiments. *Nature Reviews Genetics 3*, 579–588.

van Bakel, H. and Holstege, F. C. (2004). In control:systematic assessment of microarray performance. *EMBO Rep. 5*, 964–969.

Wolfinger, R. D. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology 8*, 625-637.