Analytical Biochemistry 449 (2014) 76–82

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Comparing real-time quantitative polymerase chain reaction analysis methods for precision, linearity, and accuracy of estimating amplification efficiency

Joel Tellinghuisen [a,*], Andrej-Nikolai Spiess [b]

[a] Department of Chemistry, Vanderbilt University, Nashville, TN 37235, USA
[b] Department of Andrology, University Hospital Hamburg-Eppendorf, Hamburg, Germany

ARTICLE INFO

Article history:
Received 28 October 2013
Received in revised form 6 December 2013
Accepted 15 December 2013
Available online 21 December 2013

Keywords:
qPCR
Weighted least squares
Statistical errors
Chi-square
Calibration
Amplification efficiency

ABSTRACT

New methods are used to compare seven qPCR analysis methods for their performance in estimating the quantification cycle (C_q) and amplification efficiency (E) for a large test data set (94 samples for each of 4 dilutions) from a recent study. Precision and linearity are assessed using chi-square (χ^2), which is the minimized quantity in least-squares (LS) fitting, equivalent to the variance in unweighted LS, and commonly used to define statistical efficiency. All methods yield C_qs that vary strongly in precision with the starting concentration N_0, requiring weighted LS for proper calibration fitting of C_q vs $\log(N_0)$. Then χ^2 for cubic calibration fits compares the inherent precision of the C_qs, while increases in χ^2 for quadratic and linear fits show the significance of nonlinearity. Nonlinearity is further manifested in unphysical estimates of E from the same C_q data, results which also challenge a tenet of all qPCR analysis methods — that E is constant throughout the baseline region. Constant-threshold (C_t) methods underperform the other methods when the data vary considerably in scale, as these data do.

The development of real-time quantitative polymerase chain reaction (qPCR)[1] methods has greatly facilitated the quantification of small amounts of genetic material [1]. The target substance is amplified through a cyclical heating/cooling process, during which the amount of the target roughly doubles in early cycles. Unfortunately this early amplification is not directly observable in most procedures, because the optical fluorescence that is commonly used to monitor the reaction progress is dominated by background contributions. Eventually the product fluorescence rises above the background, in the growth phase; but within a few cycles thereafter the process begins to saturate in the approach to the final plateau stage. Typical data are illustrated in Fig. 1, which includes profiles for four initial concentrations of the target gene, giving curves similar in shape but shifted along the cycle axis. Data for different known starting amounts of the target can be used to determine the amount of an unknown, through calibration procedures based on the exponential growth equation,

$$y = y_0 E^x, \tag{1}$$

where E is the amplification efficiency, ranging from $E = 1$ (no amplification) to $E = 2$ (perfect doubling), x is the cycle number, and y represents the fluorescence signal from the target gene, which is assumed to be proportional to the number of target molecules N. Calibration can be accomplished by associating certain cycle location indices with fixed amounts of the amplified target material. For these location benchmarks, which are labeled collectively as C_q (quantification cycle, see below), Eq. (1) implies:

$$y_q = y_0 E^{C_q}, \tag{2}$$

and a plot of C_q vs $\log(N_0)$ provides the desired calibration relation, where N_0 is the number of target molecules before amplification.

In the two decades that real-time fluorescence monitoring of the PCR reaction has been in use [2], many procedures have been described for analyzing the resulting data [3–53]. Some of these are directed toward a better determination of C_q. Others attempt to estimate the starting amount y_0 through Eq. (2) or variants thereof. This requires estimation of E, since initial hopes that E could be taken as 2 for cycles in the baseline region [11] have

* Corresponding author. Fax: +1 615 343 1234.
E-mail address: joel.tellinghuisen@vanderbilt.edu (J. Tellinghuisen).

[1] *Abbreviations and symbols used:* qPCR, quantitative polymerase chain reaction; y and y_0, fluorescence signal above baseline at cycle x and at cycle 0; E, amplification efficiency; C_q, quantification cycle; y_q, signal at $x = C_q$; N_0, initial amount of target nucleic acid in sample; C_t, threshold cycle, where $y = y_q$; FDM and SDM, cycles where y reaches its maximal first and second derivatives, respectively; C_{y0}, intersection of a straight line tangent to the curve at the FDM with the baseline-corrected x-axis; LS, least squares; χ^2, chi-square; w_i, statistical weight for ith data point; σ^2 and σ, variance and standard deviation; S, sum of weighted, squared residuals (="Chisq" in KaleidaGraph fit results, = χ^2 when $w_i = 1/\sigma_i^2$); ν, statistical degrees of freedom, = # of data points − # of adjustable parameters; SE, parameter standard error; FPLM, DART, Miner, Cy0, 5PSM, LinReg, and FPK, published qPCR analysis algorithms.

0003-2697/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ab.2013.12.020

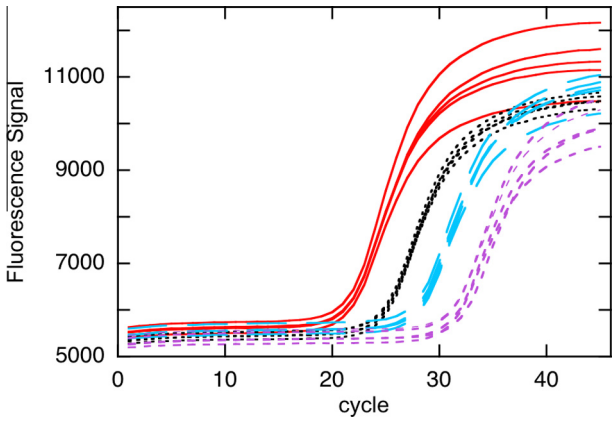

Fig.1. Representative qPCR fluorescence curves, from the 94 × 4 Reps technical dataset in Ref. [1]. Shown are 5 of the 94 replicates at each of the 4 dilutions, spanning a starting concentration range of 1000.

not been borne out [23,38]. The ultimate goal is reliable estimation of y_0 from the data for a single experiment. This goal has proved elusive and will likely remain so, because there is no direct experimental information about y from the early cycles, where the signal is buried in the background. Thus any attempt to extrapolate back to cycle 0 requires assumptions about E in this region. While it may be possible to bolster such assumptions by accumulated experience for specific genetic targets, estimating y_0 purely from single-run data will likely have to retain the assumption of constant E in this region until methods can be devised to permit its direct estimation here. Even if E can be determined, most y_0 methods also require reliable estimates of C_q. Accordingly, in this paper we emphasize the estimation of C_q and its subsequent use in calibration when data are available for multiple dilutions.

Fig. 2 shows several commonly used C_q benchmarks for idealized qPCR trajectories. The first- and second-derivative maxima (FDM and SDM) are the (noninteger) cycle values where those derivatives reach their maxima. C_t is the cycle where the fluorescence reaches a specified threshold level y_q above the baseline. C_{y0} is the intersection of a line tangent to the curve at the FDM with the baseline-subtracted signal level [37]. If the curves are all of the same shape, shifted along the cycle axis, then all of these markers are equally valid for calibration [38]; and the optimal choice is that

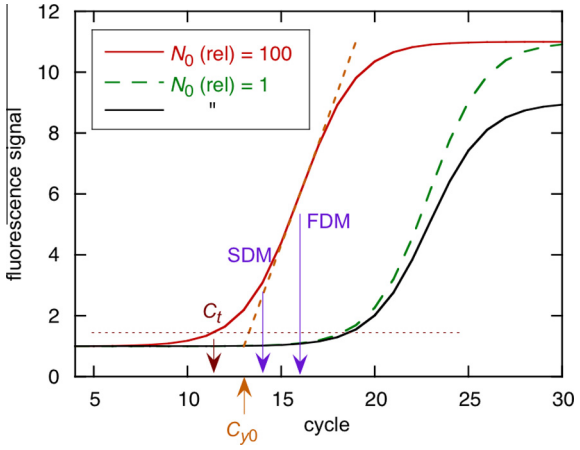

Fig.2. Synthetic qPCR curves, showing for the first (highest-N_0) curve 4 common location indices used as C_q: C_t = 11.4 (for y_q = 0.44, horizontal dashed line), C_{y0} = 13.0, SDM = 14.0, FDM = 16.0. Curves were generated with the logistic equation, $y = b + a/[1 + \exp(c(x_0 - x))]$, with baseline b = 1 and amplitude a = 10 for the first two, with a reduced to 8 for the third. The second and third curves (N_0 = 1) share a common x_0 (=FDM) of 22.89, consistent with E = 1.95.

which can be determined most precisely. However, for the purpose of estimating y_0 with Eq. (2), C_q must be taken within the range where the amplification is thought to follow Eq. (1); this exponential growth phase is commonly taken to end at the SDM or earlier. Note that the FDM, SDM, and C_{y0} are all insensitive to scale changes of the sort shown for N_0 = 1 in Fig. 2. However, C_t is sensitive to scale, as it is based on an absolute threshold level. For this reason the data are often scaled to a common plateau level ("normalized") by the instrumental software prior to analysis. With data like those in Fig. 1, where most curves do not achieve a clear plateau level, such scaling is difficult to implement without fitting to a functional form containing a plateau parameter, like the logistic function used to produce Fig. 2.

A major deficiency in the development of new qPCR analysis methods has been the lack of suitable data sets on which new methods can be compared objectively with existing methods. Recently Ruijter et al. [1] have taken a big step toward remedying that deficiency, in a comprehensive examination of some of the more popular qPCR analysis methods in medical diagnosis applications and in more purely mathematical tests. Their study employed large data sets, analyzed by 9 methods to obtain estimates of C_q and E, and they have made all these data and results available for further such work. Here we illustrate how the χ^2 statistic can be used to assess precision and linearity in the C_q estimates when data are available for multiple dilutions of the target gene. Although some of our results are specific to the data set used for this illustration, the methods will be straightforwardly applicable to results for other test data sets when they become available.

χ^2 is the minimization target in weighted and unweighted regression, equivalent to the estimated variance for the latter. Accordingly, it has a simple physical significance, leading to its use in defining the *statistical efficiency*: Increases in χ^2 are equivalent to proportionate increases in the experimental effort (number of data values) needed to maintain a stated precision. In calibration fitting, its dependence on the choice of calibration function (cubic, quadratic, linear) is thus a simple quantitative measure of the importance of nonlinearity. By contrast, the widely used R (R^2) has no such simple interpretation, though it is mathematically related to χ^2 (see below).

In the following sections, we show that the C_q values for the 94 × 4 Reps technical data set from Ref. [1] (Fig. 1) require weighted least squares for proper calibration analysis — a result that is likely to be generally applicable to C_q calibration fitting but appears not to have been noted before. We find that the methods examined in Ref. [1] vary by more than a factor of 3 in statistical efficiency but typically show efficiency losses <20% from nonlinearity. We also address the estimation of PCR amplification efficiency, and we argue that estimates of E from C_q for multiple N_0 *do* pertain to the early cycles of amplification. The resulting E estimates challenge a basic assumption of most y_0-estimation methods: that E is constant over the baseline region. Finally, we identify the source of the poorer performance by some of the qPCR analysis methods as an experimental flaw that produces variability in the scale of the data profiles, leading to the aforementioned systematic errors in C_t when it is taken as C_q.

Materials and methods

Weighted regression

In ordinary least squares (LS), estimates of the parameters in the fit model function are obtained through procedures that minimize the sum of squared residuals [54–56],

$$S = \sum \delta_i^2 = \sum [f(a, b, c, \ldots; x_i) - y_i]^2, \quad (3)$$

where $a,b,c,\ldots$ are the adjustable parameters in the model function f, and the sum runs over the n data points. For example, in a fit to a quadratic function, the ith residual $\delta_i = a + bx_i + cx_i^2 - y_i$ is calculable for all data points (x_i,y_i) after the equations have been solved for a, b, and c. The assumptions here are that x is error free and the random error σ_i in y is constant. Then the variance σ^2 in y can be estimated from $\sigma^2 = S/\nu$, where ν is the number of degrees of freedom, equal to the number of data points minus the number of adjustable parameters, $\nu = n - p$. (This is a simple extension of the familiar expression for estimating variance for a mean, where $p = 1$.)

If the y_i vary in precision, the correct minimization target is the sum of *weighted* squared residuals,

$$S = \sum w_i \delta_i^2, \tag{4}$$

where w_i is the statistical weight for the ith point. [Note that Eq. (3) is Eq. (4) with all $w_i = 1$.] For fit models that are *linear* in the adjustable parameters (which, e.g., includes polynomials of all orders), the LS procedure will yield *minimum-variance estimates* of the parameters *if and only if* the weights are taken proportional to the inverse of the data variance [57],

$$w_i \propto \sigma_i^{-2}. \tag{5}$$

For example, this is important here in the fitting of C_q data to achieve a calibration function of $\log(N_0)$, because the variance of C_q varies with N_0, increasing sharply for high dilutions.

In our calibration fitting, we use the KaleidaGraph program (Synergy Software), which has an easy-to-use routine that can handle variable weighting and nonlinear response functions, up to 9 adjustable parameters [58]. The user specifies the weighting mode by clicking a box; then when the fit is called, a data column containing the σ_i values is selected, and the w_i are computed using Eq. (5), with the proportionality constant = 1. Weighted fitting is implemented similarly in several other data analysis programs (e.g., Origin, SigmaPlot, Igor, and the qpcR package, version 1.3-7 [http://cran.r-project.org/web/packages/qpcR/index.html]).

Statistical properties of χ^2

The following properties of χ^2 are important in the present work [59]:

- When the weights are $w_i = \sigma_i^{-2}$, each term in the sum S in Eq. (4) becomes $(\delta_i/\sigma_i)^2$, which is about 1 on average. S then becomes an estimate of χ^2, with statistical average ν.
- χ^2 is related to R^2 through

$$R^2 = 1 - S/S_y, \tag{6}$$

  where S_y is defined like S in Eq. (4), but with $\delta_i = y_i - \bar{y}_w$, where $\bar{y}_w$ is the weighted average of y ($=\Sigma\, w_i y_i/\Sigma w_i$). Thus χ^2 and R^2 contain similar information, but χ^2 is easier to "read," and free of certain flaws in R^2 [60,61].

- The output of an LS fit includes estimates of the parameter standard errors (SE). In *ad hoc* LS fitting, where parameters are tested in a model in trial-and-error fashion, any parameter whose SE exceeds its magnitude is statistically insignificant; removal of such a parameter reduces the fit variance σ^2 in unweighted fitting and lowers the reduced $\chi^2(\chi^2/\nu)$ in weighted fitting.
- Changes in variance and χ^2 are equivalent to changes in the amount of data ($n \approx \nu$) needed to achieve a stated precision — hence the definition of statistical efficiency in terms of variance. For example, if $\sigma_2^2 = 2\sigma_1^2$ for methods 1 and 2, method 2 requires twice as much data to match method 1 in performance.

Table 1
Estimated variance of C_q estimates, as a function of initial concentration of the MYCN gene in the 94 × 4 Reps data.

Method	N_0			
	15000	1500	150	15
LinReg	0.0172	0.0295	0.0287	0.1955
FPK	0.0170	0.0325	0.0294	0.1942
Cy0	0.0211	0.0086	0.0215	0.1680
DART	0.0221	0.0303	0.0324	0.1847
FPLM	0.0369	0.0581	0.0428	0.2325
Miner	0.0054	0.0055	0.0178	0.1739
5PSM	0.0084	0.0072	0.0187	0.1917
Mean	0.0183	0.0245	0.0274	0.1915

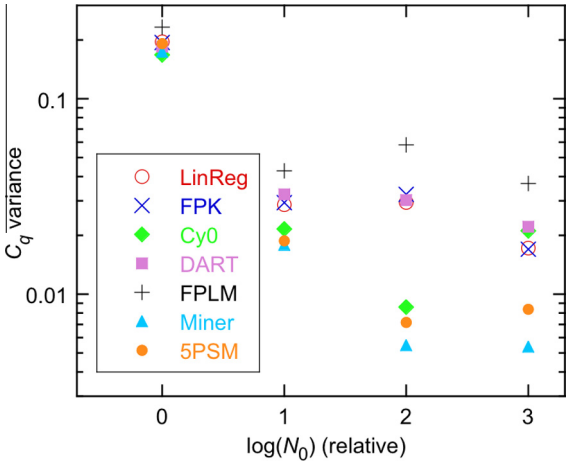

Fig.3. Variance of C_q as estimated by several qPCR analysis methods for the 94 × 4 dataset, displayed for the four initial concentrations of the target gene.

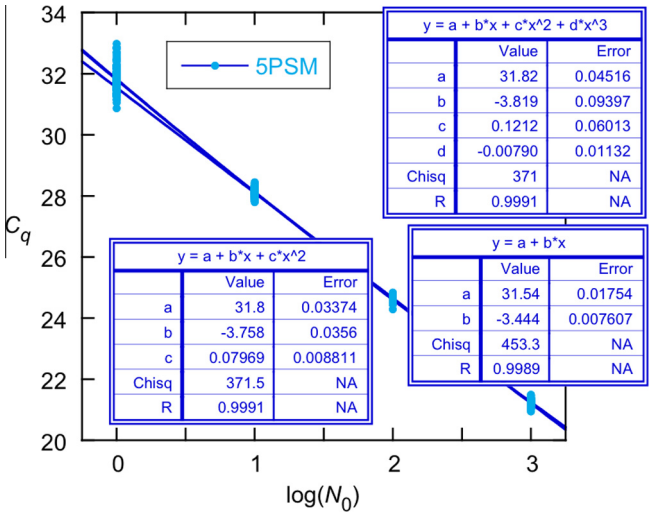

Fig.4. Results from weighted LS calibration fits of C_q for the 94 × 4 dataset, as estimated using the 5PSM results from Ref. [1], with weights taken as σ^{-2} for this method from Table 1.

Data and analysis methods

All of the data and results from Ref. [1] that are used here were obtained from the Excel files contained in the supplementary qPCR-DataMethods.zip file accompanying that paper. For the 94 × 4 replicates dataset, C_q values were provided for the following methods: FPLM [17], DART [18], PCR-Miner [25], Cy0 [37], 5PSM [40], LinRegPCR [44], and FPK-PCR [52]. The results for one run, 150_28, were omitted for all methods, leaving 375 C_q values for the calibra-

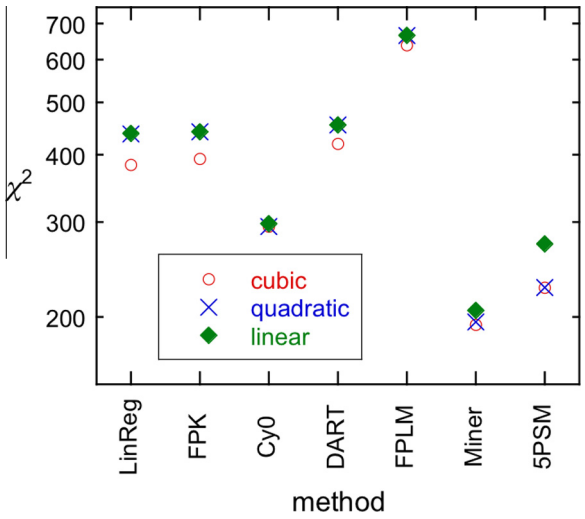

Fig.5. χ^2 values obtained for linear, quadratic, and cubic calibration fits of the C_q values, as estimated for the 94 × 4 Reps data by the indicated methods. All results were obtained using weighted LS, with weights computed using the average σ^2 values in Table 1.

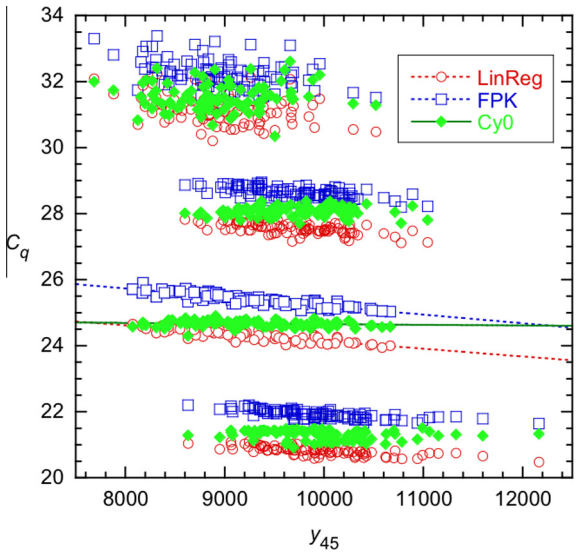

Fig.6. C_q displayed as a function of the signal in the last channel for the 94 × 4 Reps data. Results of linear LS fits are included for the second concentration. The slopes are −0.00024, −0.00026, and −0.00002, for LinReg, FPK, and Cy0, respectively. In the same order, the S values are 0.826, 0.615, and 0.777; these rise to 2.72, 2.99, and 0.79, respectively, when the slope is set = 0 (i.e., when the C_q are simply averaged).

tion fitting. Additional values were either missing or deemed outliers and omitted for two of the other methods, leaving 373 total points for Cy0 and 368 for FPLM. The Ref. [1] supplement also supplied single-run E estimates for all of these methods except Cy0.

The logic and implementation of the various analysis algorithms are described in some detail in Ref. [1], with a convenient summary in Table 1 of that work. Most attempt to estimate E, C_q, and y_0 by fitting selected data ranges to Eqs. (1) and (2) or their logarithmic versions. Since the baseline signal is always significant, it, too, must be estimated, and this is done differently in the different methods. Of the 7 methods considered here, FPLM, DART, Miner, and LinReg fall in this "exponential fitting" category, with different schemes for deciding which cycles are to be considered baseline and which belong to the growth phase, and different ways of estimating the baseline (or "ground phase"), the growth parameters, and C_q. The Cy0 method estimates just C_q, in the manner illustrated in Fig. 2. The FPK method takes E to be a 6-parameter bilinear function of y and fits the data for most cycles in the transition region between baseline and plateau, with the key fit parameter being E_{max}, the E for $y \approx 0$; the fit also yields y_0 from the variable-E version of Eq. (1), where E_x is replaced by the product of all E_i from $i = 1$ to x. 5PSM (five-parameter sigmoidal) fits all cycles to a 5-parameter log-logistic model having a constant baseline, takes C_q as the SDM from the fitted curve, and estimates E from the incremental change in the fit function from x = SDM to x = SDM + 1.

Results and discussion

Calibration testing

Most of the qPCR analysis methods considered by Ruijter et al. [1] estimate y_0 using Eq. (2), which requires that both E and C_q be estimated first. It can be shown that y_0 can never be better than C_q, even if E is perfectly determined (work in preparation).[2] Thus it is appropriate to examine the reliability with which C_q is estimated in the various methods. To this end, we show in Fig. 3 and in Table 1 the calculated variances from the C_q values obtained for each of the four initial target concentrations. From these results, the FPLM method is everywhere the poorest (highest variance) and Miner best. Further, the variance estimates are closely bunched for the lowest N_0 but are much more disperse at higher concentrations; and they span a range of over 30 for some methods, implying that in calibration fitting, the weights should vary by this much.

In the calibration fitting, we examine linear, quadratic, and cubic response functions for C_q vs $\log(N_0)$. If the C_q values for a given method are weighted as the reciprocal variances for that method from Table 1, then the cubic fit should return $\chi^2 = \nu$, because the 4 adjustable parameters permit the response function to seek its optimal value for each of the 4 concentrations. (Higher polynomial orders would be required to achieve this result with more dilutions.) Then as the order of the calibration polynomial is reduced, increases in χ^2 show the importance of nonlinearity. Fig. 4 illustrates results for the 5PSM C_q values from Ref. [1]. For the cubic fit, χ^2 (Chisq) = 371 (=375 points − 4 parameters). For the quadratic fit there is only slight increase in χ^2; this reflects the statistical insignificance of d in the cubic fit (i.e., "Error" > |d|). However, c in the quadratic fit is highly significant, and when the quadratic term is dropped from the response function, χ^2 rises by 22% — a clear marker for the significance of nonlinearity.

The results for each method similarly yield $\chi^2 = \nu$ for the cubic fit when weighted by the reciprocal variances for that method. To permit method comparisons, we use a common set of weights, taken as the reciprocals of the average variances in Table 1. Fig. 5 summarizes the results from this exercise. The best performer is Miner, followed by 5PSM and Cy0. The poorest performer is FPLM: its χ^2 is ~3 times that of Miner, meaning it requires three times as much data (replicates) to match Miner's performance. The Cy0 results are notable for their high linearity, 5PSM for strong nonlinearity, as already noted. The results for LinReg, FPK, DART, and FPLM *seem* highly linear if just linear and quadratic response functions are considered; but this is misleading, because χ^2 is significantly smaller for the cubic fits for these methods.

These results cannot be compared quantitatively with those given in Ref. [1], because all fitting done there was unweighted (all $w_i = 1$). If the present fits are repeated unweighted, the results in

[2] To understand this, consider the logarithmic form of Eq. (1), which is linear in x, with intercept $\log(y_0)$ and slope $\log(E)$. The range of fitted cycles is usually centered around x = 15–30, so the intercept is a lengthy extrapolation removed and hence is subject to variability exceeding that in the fitted $\log(y)$ values, from uncertainty in the slope. Only if the slope is error-free is the scatter in $\log(y_0)$ comparable to that in the fitted $\log(y)$ values. On the other hand, C_q is in the range of fitted cycles, so is determined with optimal precision.

Fig. 5 are reproduced qualitatively: Miner gives the lowest S and FPLM the highest (but larger by a factor of only 1.7); and Cy0 is most linear, 5PSM least (but the increase in S is only 13% for 5PSM, *cf* 22% for weighted LS). Both with and without weighting, the increases in S from nonlinearity are much smaller than the range over methods.

In seeking to explain the noticeably higher χ^2 values for 4 of the methods in Fig. 5, we note that several of these take C_q as C_t for an absolute threshold level. When the data vary in scale, an absolute threshold y_q will lead to larger C_t for curves with lower plateau levels, and smaller C_t for higher plateaus (see Fig. 2). This explanation is supported by examination of the dependence of C_q on plateau signal level (Fig. 6). We see that the systematic dependence on y_{45} accounts for almost 80% of the total S for FPK but less than 2% for Cy0, which along with 5PSM and Miner, defines C_q to be scale-insensitive (e.g., the SDM in 5PSM).

Amplification efficiency

We can estimate E from the slopes in calibration data, using a well-known expression obtainable from Eq. (2) by noting that y_q is constant by definition:[3]

$$\log(y_{0,2}/y_{0,1}) = (C_{q,1} - C_{q,2})\log(E) \equiv \Delta C_q \log(E), \tag{7}$$

where the y_0's now represent the known relative values for any two concentrations, and E the average efficiency over that range. Fig. 7 shows results for E obtained using the C_q values for adjacent concentrations. The middle estimates exceed 2.00 by amounts that are statistically significant at the 2σ level for 4 of the 7 methods; these are the same methods that show significant increases in χ^2 in Fig. 5 on going from the cubic to the quadratic fit, so these unphysical efficiency values can be seen as another manifestation of nonlinearity.

Six of the methods considered in Ref. [1] yielded E estimates for individual runs. In order of increasing E, the averages and SEs (in parentheses, in terms of final digits) of these estimates over all runs are: 1.7987(24) (5PSM), 1.8687(32) (LinReg), 1.9085(24) (DART), 1.9836(20) (FPLM), 1.9918(37) (Miner), and 2.3111(19) (E_{max}, FPK). In Fig. 8, values obtained for two of the methods at each concentration are compared with their counterparts from Fig. 7. The agreement is not good. Furthermore, the N_0-dependence is much greater than the uncertainties in the individual estimates, and their averaging is thus not supported statistically, as is seen in the large χ^2 values obtained for the weighted averages. [Note that when comparing values for consistency, the appropriate dispersion index is the SE, and thus only FPLM and Miner are marginally consistent. The sampling distributions shown in Fig. 6D of Ref. [1] are properly interpreted as applying for single determinations and thus appear to show much greater mutual consistency.]

Since the ΔC_q-based estimates of E mostly differ significantly from the single-curve estimates, we may ask about differences in their physical meaning. The single-curve estimates are intended to represent just the earliest cycles in the growth phase, where E is supposed to be at the same constant level as throughout the baseline region, before its decline to 1 in the plateau region. (Some methods, like LRE [39] and FPK, assume that E is already declining for the analyzed data, but still assume constancy for the baseline region.) On the other hand, the ΔC_q-based estimates arguably apply for the earliest cycles, through the following logic: If we assume that the qPCR data for two different starting concentrations share a common course after the more dilute sample has achieved

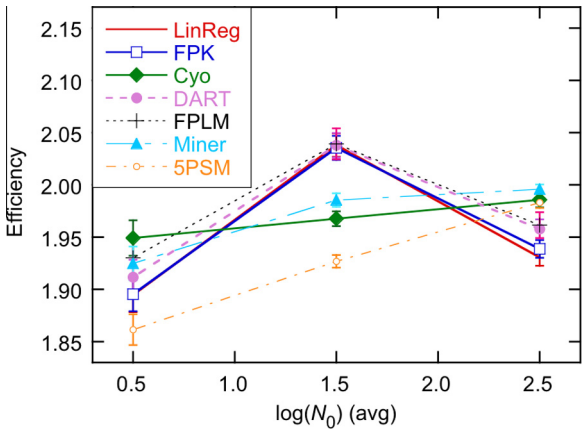

Fig.7. Amplification efficiency as a function of average concentration for the 94 × 4 Reps data, from ΔC_q values returned by the indicated methods for adjacent concentrations.

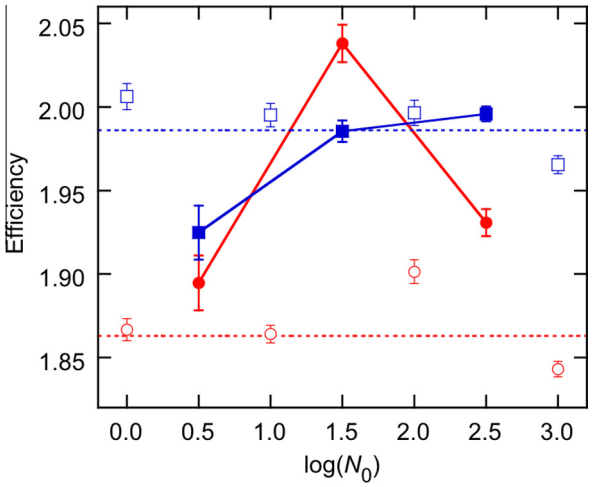

Fig.8. Efficiency estimates for the LinReg (circles) and Miner (squares) methods, as obtained from the analyses of individual runs in Ref. [1] (open) and from Fig. 7 (solid). Error bars are 1 σ; for the open points these are the SEs obtained from the values at each N_0. Weighted fits of these to a constant (broken lines) yield $\chi^2 = 47.5$ (LinReg) and 24.8 (Miner); both values easily exceed the 0.1% probability limits for χ^2, indicating that the assumption of constancy is not supported.

the N_0 of its counterpart, then ΔC_q for these two is the number of cycles needed for this initial amplification in the more dilute sample; and the Eq. (7)-based E is the average efficiency for the first ΔC_q cycles in this sample.

Which estimates are "correct"? The ΔC_q-based estimates are a byproduct of, hence fully consistent with, calibration. To the extent that the single-curve estimates differ from these, they must produce bias in the estimation of y_0.[4] Is E constant throughout the baseline region? From the results in Fig. 7, no. The consensus estimate of efficiency in the early cycles of the most dilute samples is smaller than in later cycles — a result that has been noted before [62]. The single-curve methods may be correctly estimating E in the growth onset regions, and these values may in fact vary with initial N_0 (as suggested in Fig. 8). However, none of the methods agree overall with the C_q-based estimates better than Miner, and only Miner and FPLM give single-run Es that are roughly consistent with these.

[3] y_q is not exactly constant for the FDM, SDM, and Cy0 if the curve shape varies with N_0.

[4] For example, if the true $E = 1.97$ and it is estimated as 1.87 from single-curve analysis, then for $C_q = 25$, y_0 will be overestimated by a factor of $(1.97/1.87)^{25} = 3.7$.

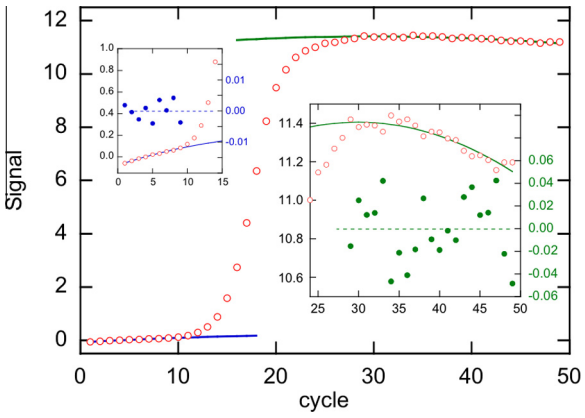

Fig.9. PCR fluorescence data for run F1.1 in the Reps dataset from Ref. [41], with quadratic functions fitted to the baseline and plateau regions. Insets show these regions at expanded scale and include the LS residuals (calculated − observed, ordinate scales to right) for the fitted points (first 9 and last 21). The estimated data variances in the two regions are 1.20×10^{-5} (baseline) and 8.9×10^{-4} (plateau).

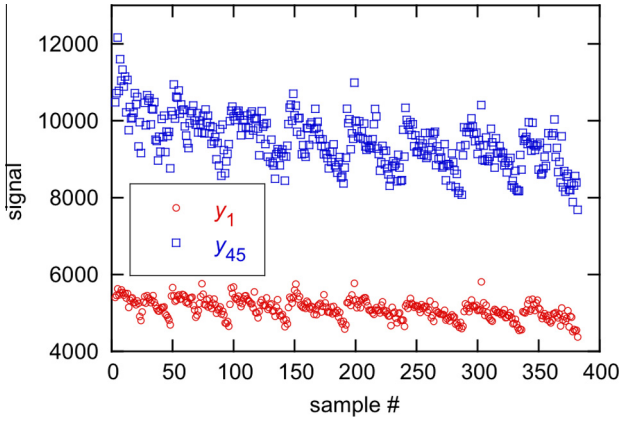

Fig.10. Signal in the first and last channels in the 94 × 4 Reps data, displayed as a function of sample number.

Conclusion

We have re-examined results from a recent study of qPCR analysis methods [1], using the χ^2 statistic to judge precision and linearity in the C_q estimates for multiple dilutions of the target gene. Because the precision of estimation of C_q depends strongly on starting concentration (N_0), weighted least squares is required for optimal results in the calibration fitting; this need appears not to have been recognized in previous work. We find better performance by three methods — PCR-Miner [25], Cy0 [37], and 5PSM [40] — manifested as better precision of estimation of the C_q values, better linearity in their subsequent use in calibration, and better estimation of E from their N_0 dependence. The better methods are insensitive to scale changes in the data, and that helps them outperform the constant-threshold methods on these test data, where significant scale variability is evident.

For the six analysis methods that gave estimates of E from single-run data, the results for the 94 × 4 Reps data are largely mutually inconsistent, and for only two methods — PCR-Miner and FPLM — are the Es close to those obtained from the calibration fitting. The premise behind these efforts — that there exists a region early in the growth phase where E is constant and equal to that in the baseline region — has been challenged in a number of studies [38,42,52]. Our estimates of E from the concentration dependence of C_q show that E is significantly smaller in the earliest cycles for the most dilute samples than in subsequent cycles having larger N. Thus, even if single-curve analysis can correctly estimate E at the start of the growth phase, there may be no guarantee that the same value will apply for all earlier cycles.

Our finding of poorer performance for constant-threshold methods must hold generally when the data vary in scale, and is the reason some instrumental software packages attempt to rescale the data to constant amplitude, as noted earlier. Some of the other results found here apply strictly only for the MYCN gene and the specific amplification chemistry and procedures used to produce the test data. For a full and fair assessment of analysis methods, there is a real need for further such comparison studies on other large data sets, to which end an agreed-upon collection of test data is highly desirable. Then proposed new methods can be tested against existing methods with a common basis for comparison. The lack of such reference data has meant that every newly proposed qPCR analysis method has seemed to be superior to all pre-existing methods, leaving potential users inadequately informed about the real capabilities of the methods.

We know of no other multiple-dilution test data sets as large as the 94 × 4 Reps data used here, but there are a number of smaller replicate data sets that might be similarly tested and are readily available from the web site of one of us (www.dr-spiess.de). These include 20 replicates at each of 6 dilutions from Rutledge and Cote [20], 12 × 6 from Sisti et al. [48], and 18 × 5 from Lievens et al. [52]. In preliminary work on these with the 5PSM method, we have encountered a problem not important for the data used in the present study. As Fig. 9 shows, the data noise can vary strongly from baseline to plateau. Such *heteroscedasticity* appears to be a general property of most qPCR data and it means that in methods that fit data over this full range, weighted LS is required for optimal analysis, requiring, e.g., downweighting by a factor of 75 in the plateau region for the data in Fig. 9. Methods that fit data in just the early growth region should be less affected, but those that transform the data logarithmically may perform less well for lack of attention to weighting resulting from this transformation [63].

Since variation in the scale of the profiles can lead to reduced precision in the estimation of C_q for the C_t methods (Fig. 6), it is relevant to look for the source of such variation. For the 94 × 4 Reps data, we find that the signal amplitude depends upon the sample number for all cycles, including those in the baseline region (Fig. 10). Further, there is a periodic variation superimposed upon the overall trend. These observations imply that the problem originates in the charging of the wells with the target, reagents, and dyes, which is done with a robotic pipetting system. When such effects are present in studies having far fewer replications, they will be completely unrecognizable while compromising the quality of the results. Accordingly, this problem warrants attention, and especially so with new qPCR instrumentation coming online with replication capabilities exceeding those of the current generation of instruments by an order of magnitude or more.

Finally, we acknowledge again that the goal of single-curve analysis methods is estimation of y_0, and performance in estimating C_q and E may not translate into similar rankings for estimating y_0. Further, some methods (like LRE, FPK, MAK2 [47], and Carr-Moore [53]) do not need C_q to estimate y_0. In fact this statement also holds for most methods based on Eqs. (1) and (2), for the following reason: If the point (C_q,y_q) lies on the fitted curve, then the value of y_0 calculated from Eq. (2) must be identically that obtained from the fit of the data to Eq. (1). In preliminary work on y_0 estimation, we find that, as explained in note 2, no methods for estimating y_0 can match C_q-based calibration unless all uncertainty and bias in E is removed — a limitation that should be borne in mind when using these methods.

Acknowledgment

This work was supported by the German Research Foundation under Grant Sp721/1-4 to A.N.S.

References

[1] J.M. Ruijter, M.W. Pfaffl, S. Zhao, A.N. Spiess, G. Boggy, J. Blom, R.G. Rutledge, D. Sisti, A. Lievens, K. De Preter, S. Derveaux, J. Hellemans, J. Vandesompele, Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications, Methods 59 (2013) 32–46.
[2] R. Higuchi, C. Fockler, G. Dollinger, R. Watson, Kinetic PCR analysis: realtime monitoring of DNA amplification reactions, Biotechnology (NY) 11 (1993) 1026–1030.
[3] U.E. Gibson, C.A. Heid, P.M. Williams, A novel method for real time quantitative RT-PCR, Genome Res. 6 (1996) 995–1001.
[4] F. Souazé, A. Ntodou-Thomé, C.Y. Tran, W. Rostene, P. Forgez, Quantitative RT-PCR: limits and accuracy, Biotechniques 21 (1996) 280–285.
[5] J. Peccoud, C. Jacob, Theoretical uncertainty of measurements using quantitative polymerase chain reaction, Biophys. J. 71 (1996) 101–108.
[6] T. Morrison, J.J. Weis, C.T. Wittwer, Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification, Biotechniques 24 (1998) 954–962.
[7] W.M. Freeman, S.J. Walker, K.E. Vrana, Quantitative RT-PCR: pitfalls and potential, Biotechniques 26 (1999) 112–122.
[8] S.A. Bustin, Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, J. Mol. Endocrinol. 25 (2000) 169–193.
[9] M.W. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR, Nucleic Acids Res. 29 (2001) e45.
[10] M.W. Pfaffl, M. Hageleit, Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR, Biotechnol. Lett. 23 (2001) 275–282.
[11] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using realtime quantitative PCR and the $2^{-\Delta\Delta CT}$ method, Methods 25 (2001) 402–408.
[12] J. Meijerink, C. Mandigers, L. van de Locht, E. Tonnissen, F. Goodsaid, J. Raemaekers, A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time pcr, J. Mol. Diagn. 3 (2001) 55–61.
[13] W. Liu, D.A. Saint, A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics, Anal. Biochem. 302 (2002) 52–59.
[14] W. Liu, D.A. Saint, Validation of a quantitative method for real time PCR kinetics, Biochem. Biophys. Res. Commun. 294 (2002) 347–353.
[15] M.W. Pfaffl, G.W. Horgan, L. Dempfle, Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR, Nucleic Acids Res. 30 (2002) e36.
[16] C. Ramakers, J.M. Ruijter, R.H. Deprez, A.F. Moorman, Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data, Neurosci. Lett. 339 (2003) 62–66.
[17] A. Tichopad, M. Dilger, G. Schwarz, M.W. Pfaffl, Standardized determination of real-time PCR efficiency from a single reaction set-up, Nucleic Acids Res. 31 (2003) e122.
[18] S.N. Peirson, J.N. Butler, R.G. Foster, Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis, Nucleic Acids Res. 31 (2003) e73.
[19] J. Wilhelm, A. Pingoud, M. Hahn, Validation of an algorithm for automatic quantification of nucleic acid copy numbers by real-time polymerase chain reaction, Anal. Biochem. 317 (2003) 218–225.
[20] R.G. Rutledge, C. Cote, Mathematics of quantitative kinetic PCR and the application of standard curves, Nucleic Acids Res. 31 (2003) e93.
[21] S. Wilkening, A. Bader, Quantitative real-time polymerase chain reaction: methodical analysis and mathematical model, J. Biomol. Tech. 15 (2004) 107–111.
[22] S. Swillens, J.C. Goffard, Y. Marechal, A.D. d'Exaerde, H. El Housni, Instant evaluation of the absolute initial number of cDNA copies from a single realtime PCR curve, Nucleic Acids Res. 32 (2004) e56.
[23] N. Lalam, Modelling the PCR amplification process by a size-dependent branching process, Adv. Appl. Probab. 36 (2004) 602–615.
[24] V. Luu-The, N. Paquet, E. Calvo, J. Cumps, Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction, Biotechniques 38 (2005) 287–293.
[25] S. Zhao, R.D. Fernald, Comprehensive algorithm for quantitative real-time polymerase chain reaction, J. Comput. Biol. 12 (2005) 1047–1064.
[26] J.L. Gevertz, S.M. Dunn, C.M. Roth, Mathematical model of real-time PCR kinetics, Biotechnol. Bioeng. 92 (2005) 346–355.
[27] A. Larionov, A. Krause, W. Miller, A standard curve based method for relative real time PCR data processing, BMC Bioinf. 6 (2005) 62.
[28] R. Goll, T. Olsen, G. Cui, J. Florholmen, Evaluation of absolute quantitation by nonlinear regression in probe-based real-time PCR, BMC Bioinf. 7 (2006) 107.
[29] N. Lalam, Estimation of the reaction efficiency in polymerase chain reaction, J. Theor. Biol. 242 (2006) 947–953.
[30] J.H. Schefe, K.E. Lehmann, I.R. Buschmann, T. Unger, H. Funke-Kaiser, Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's C (T) difference" formula, J. Mol. Med. 84 (2006) 901–910.
[31] S. Cikos, A. Bukovska, J. Koppel, Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis, BMC Mol. Biol. 8 (2007) 113.
[32] Y. Karlen, A. McNair, S. Perseguers, C. Mazza, N. Mermod, Statistical significance of quantitative PCR, BMC Bioinf. 8 (2007) 131.
[33] M.V. Smith, C.R. Miller, M. Kohn, N.J. Walker, C.J. Portier, Absolute estimation of initial concentrations of amplicon in a real-time RT-PCR process, BMC Bioinf. 8 (2007) 409.
[34] I. Chervoneva, Y.Y. Li, B. Iglewicz, S. Waldman, T. Hyslop, Relative quantification based on logistic models for individual polymerase chain reactions, Stat. Med. 26 (2007) 5596–5611.
[35] M.J. Alvarez, G.J. Vila-Ortiz, M.C. Salibe, O.L. Podhajcer, F.J. Pitossi, Model based analysis of real-time PCR data from DNA binding dye protocols, BMC Bioinf. 8 (2007) 85.
[36] J. Feng, R. Zeng, J. Chen, Accurate and efficient data processing for quantitative real-time PCR using a tripartite plant virus as a model, Biotechniques 44 (2008) 901–912.
[37] M. Guescini, D. Sisti, M.B. Rocchi, L. Stocchi, V. Stocchi, A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition, BMC Bioinf. 9 (2008) 326.
[38] R. Rutledge, D. Stewart, Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR, BMC Mol. Biol. 9 (2008) 96.
[39] R. Rutledge, D. Stewart, A kinetic-based sigmoidal model for the polymerase chain reaction and its application to high-capacity quantitative real-time PCR, BMC Biotechnol. 8 (2008) 47.
[40] A.N. Spiess, C. Feig, C. Ritz, Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry, BMC Bioinf. 9 (2008) 221.
[41] C. Ritz, A.N. Spiess, QpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis, Bioinformatics 24 (2008) 1549–1551.
[42] S. Swillens, B. Dessars, H.E. Housni, Revisiting the sigmoidal curve fitting applied to quantitative real-time PCR data, Anal. Biochem. 373 (2008) 370–376.
[43] W. OConnor, E.A. Runquist, Error minimization algorithm for comparative quantitative PCR analysis: Q-Anal, Anal. Biochem. 378 (2008) 96–98.
[44] J.M. Ruijter, C. Ramakers, W.M.H. Hoogaars, Y. Karlen, O. Bakker, M.J.B. van den Hoff, A.F.M. Moorman, Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data, Nucleic Acids Res. 37 (2009) e45.
[45] S. Cikos, J. Koppel, Transformation of real-time PCR fluorescence data to target gene quantity, Anal. Biochem. 384 (2009) 1–10.
[46] M. Garlick, J. Powell, D. Eyre, T. Robbins, Mathematically modeling pcr: an asymptotic approximation with potential for optimization, Math. Biosci. Eng. 7 (2010) 363–384.
[47] G.J. Boggy, P.J. Woolf, A mechanistic model of PCR for accurate quantification of quantitative PCR data, PLoS ONE 5 (2010) e12355.
[48] D. Sisti, M. Guescini, M.B.L. Rocchi, P. Tibollo, M. D'Atri, V. Stocchi, Shape based kinetic outlier detection in real-time PCR, BMC Bioinf. 11 (2010) 186.
[49] R.B. Page, A.J. Stromberg, Linear methods for analysis and quality control of relative expression ratios from quantitative real-time polymerase chain reaction experiments, Sci. World J. 11 (2011) 1383–1393.
[50] S.L.R. Ellison, K.R. Emslie, Z. Kassir, A standard additions method reduces inhibitor-induced bias in quantitative real-time PCR, Anal. Bioanal. Chem. 401 (2011) 3221–3227.
[51] T. Bar, M. Kubista, A. Tichopad, Validation of kinetics similarity in qPCR, Nucleic Acids Res. 40 (2012) 1395–1406.
[52] A. Lievens, S. Van Aelst, M. Van den Bulcke, E. Goetghebeur, Enhanced analysis of real-time PCR data by using a variable efficiency model: FPK-PCR, Nucleic Acids Res. 40 (2012) e10.
[53] A.C. Carr, S.D. Moore, Robust quantification of polymerase chain reactions using global fitting, PLoS ONE 7 (2012) e37640.
[54] A.M. Mood, F.A. Graybill, Introduction to the Theory of Statistics, second ed., McGraw-Hill, New York, 1963.
[55] W.C. Hamilton, Statistics in Physical Science: Estimation, Hypothesis Testing, and Least Squares, The Ronald Press Co., New York, 1964.
[56] W.E. Deming, Statistical Adjustment of Data, Dover, New York, 1964.
[57] J. Tellinghuisen, Weighted least-squares in calibration: what difference does it make?, Analyst 132 (2007) 536–543.
[58] J. Tellinghuisen, Nonlinear least-squares using microcomputer data analysis programs: KaleidaGraph™ in the physical chemistry teaching laboratory, J. Chem. Educ. 77 (2000) 1233–1239.
[59] J. Tellinghuisen, Understanding least squares through Monte Carlo calculations, J. Chem. Educ. 82 (2005) 157–166.
[60] A.N. Spiess, N. Neumeyer, An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical rsearch: a Monte Carlo approach, BMC Pharmacol. 10 (2010) 6. biomedcentral.com/1471-2210/10/6.
[61] J. Tellinghuisen, C.H. Bolster, Using R^2 to compare least-squares fit models: when it must fail, Chemom. Intell. Lab. Syst. 105 (2011) 220–222.
[62] H.K. Nogva, K. Rudi, Potential influence of the first PCR cycles in real-time comparative gene quantifications, Biotechniques 37 (2004) 246–253.
[63] J. Tellinghuisen, Bias and inconsistency in linear regression, J. Phys. Chem. A 50 (2000) 11829–11835.