Theor Appl Genet (2010) 120:475–490
DOI 10.1007/s00122-009-1210-3

ORIGINAL PAPER

Background correction of two-colour cDNA microarray data using spatial smoothing methods

André Schützenmeister · Hans-Peter Piepho

Received: 13 August 2009 / Accepted: 22 October 2009 / Published online: 15 November 2009

Abstract The analysis of two-colour cDNA microarray data usually involves subtracting background values from foreground values prior to normalization and further analysis. This approach has the advantage of reducing bias and the disadvantage of blowing up the variance of lower abundant spots. Whenever background subtraction is considered, it implicitly assumes locally constant background values. In practice, this assumption is often not met, which casts doubts on the usefulness of simple background subtraction. In order to improve background correction, we propose local background smoothing within the pre-processing pipeline of cDNA microarray data prior to background correction. For this purpose, we employ a geostatistical framework with ordinary kriging using both isotropic and anisotropic models of spatial correlation and 2-D locally weighted regression. We show that application of local background smoothing prior to background correction is beneficial in comparison to using raw background estimates. This is done using data of a self-versus-self experiment in Arabidopsis where subsets of differentially expressed genes were simulated. Using locally smoothed background values in conjunction with existing background correction methods increases the power, increases the accuracy and decreases the number of false positive results.

Communicated by M. Frisch.

Contribution to the special issue "Heterosis in Plants".

The R-code that we used is available from the authors upon request.

A. Schützenmeister · H.-P. Piepho (✉)
Bioinformatics Unit, Institute for Crop Production and Grassland Research, University of Hohenheim, Fruwirthstrasse 23, 70599 Stuttgart, Germany
e-mail: piepho@uni-hohenheim.de

A. Schützenmeister
e-mail: schuemei@uni-hohenheim.de

Introduction

The cDNA microarray technology has been widely used throughout all fields of scientific research that make use of gene expression data, including projects aimed at unravelling the causes of heterosis and studying heterosis itself (Keller et al. 2005; Uzarowska et al. 2007, 2009; Höcker et al. 2008; Paschold et al. 2009; Frisch et al. 2009, Thiemann et al. 2009, Jahnke et al. 2009). Two-colour cDNA microarrays quantify the response of thousands of genes to a specific stimulus at once. Such stimuli could be treatments with specific reagents, different developmental stages, different genotypes, different tissues of the same organism and so on. In general, the experimenter expects differences in expression of a subset of genes due to the different stimuli. On each two-colour cDNA microarray, two mRNA-probes are competitively hybridized after they have been reverse-transcribed into cDNA and marked with two different fluorophores such as Cy-3 (green) and Cy-5 (red). They are expected to bind to their complementary sequences which are immobilized onto the surface of the chip at specific known positions. This allows quantifying the amount of fluorescence emitted when a laser excites these fluorophores. A scanner captures these fluorescence signals, which are subsequently transformed into real numbers (Mary-Huard et al. 2004; Schena 2003).

Several sources of (non-biological) variation have been identified, which directly influence gene expression measurements. For example, Schuchhardt et al. (2000) identified the following sources of non-biological variation: mRNA-preparation, reverse transcription, labelling of the

cDNAs, PCR-amplification, systematic variation due to different groups of pin-tips, hybridization efficiency, slide inhomogeneities, non-specific hybridization, non-specific background and image analysis. Different methods have been proposed, which all aim at adjusting for effects that arise from non-biological sources rather than from biologically caused differences (Fujita et al. 2006; Haldermans et al. 2007; Huber et al. 2002; Irizarry et al. 2003; Piepho et al. 2006; Smyth and Speed 2003; Yang et al. 2002).

Besides estimates of the spot intensities, microarray scanning software provides estimates of background (BG) fluorescence. Local BG fluorescence emerges from labelled cDNA, which binds to the glass surface and is usually assumed to contribute additively to the foreground spot intensity (FG). This BG is estimated from the area or specific parts of the area surrounding a spot. There is no standard procedure or definition how BG should be estimated and different methods coexist. Yin et al. (2005) review different BG estimation procedures and propose their own method.

Prior to normalizing, the data are often corrected for BG fluorescence. Although it is not clear whether one should perform this step or not, background subtraction (BS) has become the "standard" procedure (Kooperberg et al. 2002), and some authors recommend to avoid BS completely (Yang et al. 2001; Tran et al. 2002). BS is known to increase the variance of expression ratios (Scharpf et al. 2006; Kooperberg et al. 2002). Scharpf et al. (2006) pointed out that BS reduces bias but increases variance in the estimates of expression ratios. They used data simulation to study this bias-variance trade-off and developed recommendations to decide whether to perform BS or not. Kooperberg et al. (2002) proposed a Bayesian approach to BG correction, which reduces the variance of low-intensity ratios while leaving intensities for higher ratios nearly unchanged. Yuan and Irizarry (2006) propose a method that requires technical replication on the array. Ritchie et al. (2007) compare several BG correction methods for two-colour microarrays and recommend a normal plus exponential convolution model with offset.

Several groups reported spatially systematic artefacts (Colantuoni et al. 2002; Arteaga-Salas et al. 2008; Mary-Huard et al. 2004; Neuvial et al. 2006). Colantuoni et al. (2002) observed a spatially dependent accumulation of significant log-ratios although they expected a completely random distribution. They proposed their local mean normalization, which consists of fitting a two-dimensional, locally weighted regression (LOWESS) of signal intensities without BG correction for an experimental dataset and the control dataset. Normalization is done by dividing signal intensities of both datasets by the locally estimated mean response. Mary-Huard et al. (2004) describe a so-called spotting effect, which is believed to be caused by printing procedures. They make use of the semivariogram, a geostatistical tool that estimates spatial correlation, in order to analyse spatial dissimilarities. Arteaga-Salas et al. (2008) report spatial flaws in oligonucleotide microarrays. They compare the vicinity of a spot and search for spatially clustered values that are extreme compared to a reference or replicated data. Neuvial et al. (2006) report two types of spatial effects which are not accounted for by other normalization procedures. They identified spatial gradients of centred log-ratios influencing the entire microarray and local spatial bias, which cannot be explained by the microarray spotting design. They suggested a spatial normalization method combining spatial segmentation and spatial trend estimation that accounts for spatial gradients via two-dimensional LOWESS regression. Although their method was originally designed for array comparative genomic hybridization (array-CGH) data (Beatty et al. 2009), it can be applied to any microarray experiment.

In our collaborative work in a research network on heterosis in plants (https://www.uni-hohenheim.de/plantbreeding/350a/dfg/indexd.html), we also observed spatial correlation of the BG intensities. Figure 1 depicts two heatmaps of log-transformed BG intensity estimates of two different cDNA microarrays. The spatial structure extends over a larger area than one would expect for this kind of data. An implicit assumption of BG correction methods is that BG values of cDNA microarrays are locally constant (Kooperberg et al. 2002). We adopt this assumption and investigate whether geostatistical smoothing methods or 2-D locally weighted regression (LOWESS) can improve estimation of BG values, and therefore, BG correction (Cleveland et al. 1988, 1991). We combine these methods with existing BG correction methods which avoid negative corrected signals (Ritchie et al. 2007). Such methods are beneficial because negative signals cannot be used in further analysis whenever log-transformation of BG corrected signals is considered. Log-transformation of negative values is undefined, which would lead to loss of information.

To account for local differences and for computational reasons, we apply our procedures for each block of a microarray separately. Specifically, we consider three approaches to BG smoothing. The first one is a complex procedure which incorporates directional information. Before applying this method, we first check whether there is a sufficient amount of spatial correlation by performing an approximate hypothesis test. In case this test is significant, we use empirical semivariograms of BG values to fit a theoretical model of spatial correlation, which is then used to perform ordinary kriging (OK) to smooth BG values. Second, we use this geostatistical approach for all blocks of a microarray without testing the existence of spatial correlation and do not incorporate directional information. Third, we use 2-D LOWESS on BG values. With all approaches, we obtain new smoothed BG estimates that incorporate the information of nearby BG values. These

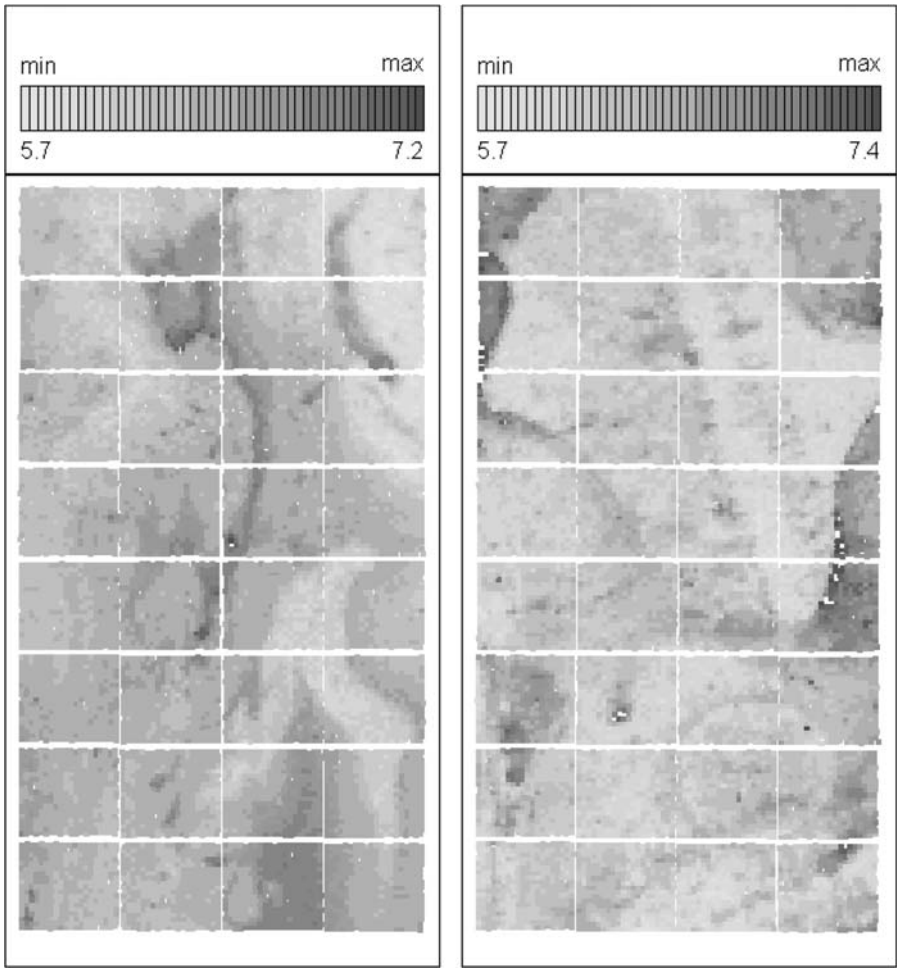

Fig. 1 Two heatmaps of log-transformed background (BG) intensities showing spatial correlation of BG values. Some blocks show clear spatial patterns among BG values, others do not. The data shown is part of a larger dataset used to uncover the molecular causes of heterosis in maize (Keller et al. 2005; Piepho et al. 2006)

values are then used for BG correction. Subsequently, we check if this methodology has an overall positive effect on estimation of genotypic differences. For this purpose, we use a self-versus-self (SVS) dataset where differentially expressed genes were simulated.

In the next section, we introduce the general geostatistical framework and introduce our three approaches. In "Ordinary Kriging" we apply this algorithm to the SVS dataset where we simulated subsets of differentially expressed (DE) genes. To assess the performance in cases where no or just a small fraction of genes is expected to exhibit significant differences, we applied all approaches to the original SVS dataset, i.e. without simulating DE genes. The paper ends with a brief "Discussion".

Materials and methods

Semivariograms

A semivariogram is a function which describes the degree of spatial dependencies for a stochastic process or a random field. For a stationary process, it is defined as:

$$\gamma(h) = E\left\{ \left(Z(x_i) - Z(x_j)\right)^2 \right\} \Big/ 2,$$

where $Z(x_i)$ denotes an observation from the process at location $x_i = (x_{1i}, x_{2i})'$, measured in absolute Cartesian coordinates or in row and column numbers in case of a regular grid, and h is the distance between locations x_i and x_j. The observed quantities are assumed to be the sum of a deterministic trend $m(x)$ and a Gaussian stationary random process $G(x)$, $Z(x) = m(x) + G(x)$. Stationarity means the independence of location, i.e., the random process $G(x)$ only depends on the separation distance. A simple consequence is that $Z(x)$ is stationary if and only if $m(x)$ is constant.

We use the robust semivariogram estimator, proposed by Cressie and Hawkins (1980), defined as:

$$\hat{\gamma}_{\text{Cressie-Hawkins}}(h, \delta d) = \frac{\left(\frac{1}{|N(h,\delta d)|} \sum_{N(h)} \left| Z(x_i) - Z(x_j) \right|^{1/2} \right)^4}{2[0.457 + 0.494/N(h, \delta d)]},$$

where $N(h, \delta d) = \{(i,j) : |x_i - x_j| \in [h - \delta d, h + \delta d[\}$ is the set of location pairs (x_i, x_j) separated by distance h within a tolerance of $\pm\delta d$, where d specifies the lag distance, which is the distance between two consecutive lag

classes, with $h \in kd, k \in 1,\ldots,N_{\text{lag}}, N_{\text{lag}}$ being the number of lag classes. The lag tolerance δd is usually chosen to be equal to half of the lag distance d.

There are three theoretical models of $\gamma(h)$ considered here, i.e., the exponential model with correlation function $\rho(h) = \exp(-h/\phi)$, the spherical model with $\rho(h) = 1 - 1.5(h/\phi) + 0.5(h/\phi)^3$ if $h < \phi$ (0 otherwise), and the Gaussian model with $\rho(h) = \exp[-(h/\phi)^2]$. In each of these models, ϕ denotes the range parameter, which determines the rate at which the correlation decays with distance h. The semivariogram $\gamma(h)$ can be linked to its correlation function via $\gamma(h) = \tau^2 + \sigma^2\{1 - \rho(h)\}$. The intercept τ^2 corresponds to the nugget variance and has much influence on prediction results. A larger value will lead to a smoother predicted surface with a smaller fraction of the original structure retained. The smaller this value, the less smooth the predicted surface, and the more of the original structure of the data is retained. The value σ^2 corresponds to the signal variance. The correlation function $\rho(h)$, which has extra parameters, determines the way the asymptote $\tau^2 + \sigma^2$ of the semivariogram (sill) is reached.

Often, the assumption that values are correlated in a way that only depends on the separation distance of data locations (isotropy) does not reflect reality. Frequently, the correlation also depends on direction. Specifically, there may be a major axis of correlation, which is defined by the correlation model with the largest range, i.e., pairs in this direction are correlated over a larger distance than pairs located in other directions. Models that take direction into account are called anisotropic. In this paper, we restrict attention to so-called geometric anisotropy as explained next.

Anisotropy

If N_{angle} directional semivariograms $\gamma_i(h), i \in 1,\ldots,N_{\text{angle}}$ are to be computed, pairs are located within the area:

$$[\theta_i - \sigma_{\text{angle}}, \theta_i + \sigma_{\text{angle}}[, \text{with angle } \theta_i \in \left\{ j\frac{180^\circ}{N_{\text{angle}}}, j \in 0,\ldots,(N_{\text{angle}} - 1) \right\}, i \in 1,\ldots,N_{\text{angle}}.$$

In case the angle tolerance value σ_{angle} is chosen to be equal to $180^\circ/2N_{\text{angle}}$, it is ensured that each pair can be assigned to a specific direction. Thus, a directional semivariogram has a subscript to indicate the direction (angle) it refers to.

Geometric anisotropy assumes the nugget variance τ^2 as well as the signal variance σ^2 to be constant for each of the N_{angle} semivariograms. The direction with the maximum range parameter $\phi_{\max}$ represents the main axis $\theta_{\max}$ of anisotropy, the perpendicular angle is the minor axis $\theta_{\min}$ with range $\phi_{\min}$. The ratio of both range parameters $R = \phi_{\max}/\phi_{\min}$ defines the shape of an ellipse of equal correlation with the two axes representing lag/Euclidean distances in two dimensions, which can be seen as the geometrical interpretation of geometric anisotropy. N_{angle} directional semivariograms allow to compute its parameters in case $N_{\text{angle}} \bmod 2 = 0$. This type of anisotropy is used here because of the simplicity and availability of prediction methods in the statistical software we use (Cressie 1993; Ribeiro and Diggle 2001).

Ordinary Kriging

Ordinary kriging (OK) estimates values at locations using the data available at nearby locations. These estimates are weighted linear combinations of the observed data. Kriging estimates are unbiased in the sense that they have mean residual error equal to zero. It also aims at minimizing the variance of errors which, all combined, makes OK a best linear unbiased estimator, BLUE according to Isaaks and Srivastava (1989). This definition of OK complies with the definition of the best linear unbiased predictions (BLUP) in mixed model theory (Robinson 1991; Schabenberger and Pierce 2002), which appears to be the more natural definition since there are no fixed effects estimated in OK (apart from a general mean). OK assumes that observed values $z_i = Z(x_i)$ at locations $x_i, i \in 1,\ldots,N$ are the result of a stationary random process $G(x)$, i.e. expected values do not depend on location. For a pair of values located apart from each other, first order and second order probability laws only depend on the separation, not on the location, and, therefore, have the same expected value (mean) regardless location, which corresponds to $m(x) = \text{const}$ within $Z(x) = m(x) + G(x)$.

OK allows to estimate a value z_0 at location x_0 using the information from nearby locations z_i ($i = 1,\ldots, n$):

$$\hat{z}_0 = \sum_{i=1}^{n} w_i z_i,$$

where optimal weights w_i have to be estimated, which are constrained to:

$$\sum_{i=1}^{n} w_i = 1.$$

The isotropic or anisotropic semivariograms represent functions, which can be used to calculate covariances needed to determine the weights w_i for OK (Isaaks and Srivastava 1989).

Usually, OK is used to estimate unsampled locations. In our case, we only require smoothed estimates for an observed regular grid of locations. To prevent OK from honouring the data, i.e., OK estimates representing the data

values, we slightly shift the grid used for prediction relative to the observed grid. Thus, any OK estimate uses the information of the original spot as well as all nearby spots representing a weighted linear combination of these data points depending on the correlation function that was fitted to the empirical semivariogram $\hat{\gamma}(h)$.

Global trend

Global or spatial trend is the simplest form of departure from stationarity. Second-order stationarity holds, if the expected value m of a random process, as well as the covariance of two locations separated by distance h, are independent of location. With global trend the mean response $m = m(x)$ will depend on location x, i.e. $m(x_i) \neq m(x_j), \exists(i,j) : i \neq j; i,j \in 1,\ldots,n$.

In practice, spatial trend is often modelled as polynomial regression using powers and cross products of Cartesian coordinates (Diggle and Ribeiro 2007). If this kind of explanatory variables is used, such models are called trend surface models. Of course, other kinds of explanatory variables might be used to model the mean function $m(x)$ instead. Often, there are additional spatially referenced quantities available besides the variable of primary interest. These values can be used for modelling the mean response. We use only the spatial coordinates and do not consider other covariables.

We considered only trend surface models for $m(x)$ of maximum degree two. We believe, that many types of global trend can be explained with this type of trend surface model, particularly when analysing expression data block by block, where a block consists of about 400–500 spots only.

The trend models we consider are all well-formed according to Nelder (2000). This means that all higher order terms have to be accompanied by their marginal terms. Thus, e.g., it is not reasonable to include x_1^2 if x_1 is not included, where x_1 (x_2) denotes the first (second) of the two coordinates $x_i = (x_{1i}, x_{2i})$ with index i dropped for simplicity.

We consider the following trend surface models:

$$\begin{aligned}&1, x_1, x_2, x_1 + x_2, x_2 + x_2^2, x_1 + x_1^2, x_1 + x_2 + x_1x_2, x_1 + x_2\\ &\quad + x_1^2, x_1 + x_2 + x_2^2, x_1 + x_2 + x_1x_2 + x_1^2, x_1 + x_2 + x_1x_2\\ &\quad + x_2^2, x_1 + x_2 + x_1^2 + x_2^2, x_1 + x_2 + x_1x_2 + x_1^2 + x_2^2.\end{aligned}$$

The first model represents the constant mean model, which corresponds to second-order stationarity of the original data set as defined previously.

Once a global trend model is chosen, all downstream analyses are performed on the residuals of this fitted model, i.e. all analyses are performed on values $G(x_i)$:

$$G(x_i) = Z(x_i) - m(x_i), i \in 1,\ldots,N.$$

This can be seen as transforming a non-stationary random field to a stationary one, thus fulfilling the prerequisites of OK. The procedure is also known as universal kriging.

Model selection and estimation

We select models based on F tests to compare nested models, while BIC is used to compare non-nested models. It would be ideal to fit all models by restricted maximum likelihood (REML), because this would allow accounting for spatial correlation at all stages of the analysis. Unfortunately, we encountered frequent convergence problems with REML, so this approach was not feasible for analysing a large number of genes. Instead, we used a weighted least squares approach, using the number of observations in a lag class as weight.

At first, we check whether the data provide sufficient information to fit a spatial model required to perform OK. For this purpose we use the constant mean global trend model. An approximate F test is performed comparing the best fitting theoretical model of spatial continuity to a pure nugget model (constant semivariance/correlation i.e. $\gamma(h) = \tau^2, \rho(h) = k$). We use an isotropic model here because of its simplicity. After computing the isotropic empirical semivariogram as described above, the best fitting theoretical model (exponential, spherical, Gaussian) is computed using non-linear weighted least squares, i.e. each lag class is assigned a weight equal to the number of observations. The best theoretical model is the one that has the smallest residual sum of squares:

$$\mathrm{RSS} = \sum_{i=1}^{N} [\hat{\gamma}_i(h_i) - \gamma_F(h_i)]^2 w_i$$

where $\gamma_F(h_i)$ denotes the fitted semivariogram model, i ($i \in 1,\ldots,N$) indexes all lag classes, and w_i are weights representing the number of pairs within a lag class i. In order to obtain reliable estimates of the semivariance, we constrain lag classes to have at least 100 pairs.

Each theoretical model considered has three parameters (nugget τ^2, sill σ^2, range ϕ), while the pure nugget model has only one parameter (nugget τ^2). To compare the pure nugget model to the best fitting theoretical model we use the following F statistic:

$$F_1 = \frac{(\mathrm{RSS}_{\mathrm{nugget}} - \mathrm{RSS}_{\mathrm{theo}})/(\mathrm{DF}_{\mathrm{nugget}} - \mathrm{DF}_{\mathrm{theo}})}{\mathrm{RSS}_{\mathrm{theo}}/\mathrm{DF}_{\mathrm{theo}}}$$

with $\mathrm{RSS}_{\mathrm{nugget}}$ and $\mathrm{RSS}_{\mathrm{theo}}$ the residual sum of squares of the pure nugget model and the best fitting theoretical model, respectively, and $\mathrm{DF}_{\mathrm{nugget}}$ and $\mathrm{DF}_{\mathrm{theo}}$ the respective residual degrees of freedom. If $\mathrm{N_{SV}}$ is the number of

semivariances (lag classes) of the empirical semivariogram, we have $DF_{nugget} = N_{SV} - 1$ and $DF_{theo} = N_{SV} - 3$. The value F_1 is compared to a $F(2; DF_{theo})$ distribution, and the P value is obtained, which we require to be ≤ 0.001 to infer that there is enough information to assume spatial dependence.

If this F test is significant, all trend surface models as described in the previous section are fitted to the raw data. The values of the Bayesian information criterion (BIC) for each model are obtained, and the model with the smallest BIC value is chosen as trend surface model. BIC-values have the usual form and were computed as:

$$\mathrm{BIC} = -2\log(\text{likelihood}) + \log(n)k,$$

where n is the number of residuals, and k is the number of parameters in the model. Using maximum likelihood estimates under the assumption of normally distributed errors:

$$-2\log(\text{likelihood}) = n\log(2\pi) + n\log\left(\hat{\sigma}_k^2\right) + n,$$

with $\hat{\sigma}_k^2 = \mathrm{RSS}/n$ (McQuarrie and Tsai 1998). Again, the use of BIC is based on the simplifying assumption of independence and normality of empirical semivariance estimates.

The final step is to select a spatial model. As mentioned above, we consider three such models, namely the exponential model, the spherical model, and the Gaussian model. Furthermore, the decision has to be made whether an isotropic model is preferable or an anisotropic model. Each fit of a theoretical model to an isotropic or anisotropic (directional) empirical semivariogram assumes the usage of a trend surface model as described above.

We only consider geometrical anisotropy. Thus, two parameters have equal values for all directions, namely the nugget-variance τ^2 and the signal variance σ^2, whereas the range parameters ϕ_i $(i \in 1, \ldots, N_{\text{angle}})$ might change. Figure 2 depicts a directional semivariogram computed for a single block of a cDNA two-colour microarray.

A directional empirical semivariogram $\hat{\gamma}_i(h)$ consists of lag classes $N_i(h)$ of direction i $(i \in 1, \ldots, N_{\text{angle}})$. Each lag class comprises all pairs that are separated by distance h for a direction/angle i. We require all lag classes of the directional semivariogram to have at least 100 pairs for each direction, lag classes with lower frequencies are discarded.

At first we obtain the best fitting anisotropic model. For this purpose, we fit all three anisotropic models of spatial continuity numerically. There are $2 + N_{\text{angle}}$ parameters that have to be estimated, $\tau^2, \sigma^2, \phi_1, \ldots, \phi_{N_{\text{angle}}}$. The objective function to be minimized for all three theoretical models $\gamma(h_{ij})$ is the residual sum of squares (RSS) accumulated over all directions:

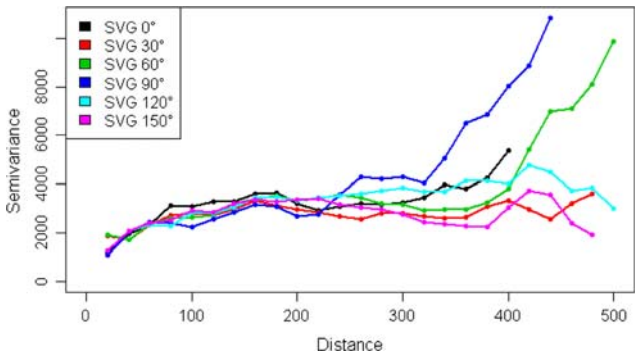

Fig. 2 Plot of a multidirectional empirical semivariogram where directional semivariograms (SVG) correspond to six different directions. All directional semivariograms show a similar form up to a distance where they start to depart. Data are part of a larger data set described in Piepho et al. (2006)

$$\mathrm{RSS} = \sum_{i=1}^{N_{\text{angle}}} \sum_{j=1}^{N_i} \left[\hat{\gamma}_i(h_{ij}) - \gamma_F(h_{ij})\right]^2 w_{ij},$$

where $\gamma_F(h_{ij})$ denotes the fitted semivariogram, i ($i \in 1,\ldots,$ N_{angle}) indexes a particular direction, j ($j \in 1,\ldots, N_i$) indexes all lag classes of a direction, which correspond to a specific lag distance, and w_{ij} are weights representing the number of pairs within lag class j of one particular direction i. If these weights are set equal to one, this results in OLS estimation instead of weighted least squares estimation (WLS). The anisotropic theoretical model with the smallest residual sum of squares is chosen.

The best fitting anisotropic model is compared to the corresponding isotropic model. We obtain the correlation model (exponential, spherical, Gaussian) from the best fitting anisotropic model, and fit it to the directional empirical semivariogram using a single theoretical model (isotropic model). Thus, at each lag distance there are $1,\ldots, N_{\text{angle}}$ values possible. All lag classes with more than one value provide replicated data.

Fitting the isotropic model to the directional semivariogram is done using the same algorithm as for the anisotropic case. Only the respective objective functions differ slightly. Specifically, there is no subscript for the range parameter ϕ. We use an F statistic to decide whether to use the isotropic or anisotropic model:

$$F_2 = \frac{(\mathrm{RSS}_{\text{iso}} - \mathrm{RSS}_{\text{aniso}})/(\mathrm{DF}_{\text{iso}} - \mathrm{DF}_{\text{aniso}})}{\mathrm{RSS}_{\text{aniso}}/\mathrm{DF}_{\text{aniso}}}.$$

The isotropic model has three parameters, namely nugget variance τ^2, signal variance σ^2, and range ϕ. In contrast, the anisotropic model has $N_{\text{angle}} - 1$ additional parameters, since there is a range parameter for each direction, hence $DF_{\text{iso}} = N_{\text{sv}} - 3$ and $\mathrm{DF}_{\text{aniso}} = N_{\text{sv}} - N_{\text{angle}} - 2$, where N_{SV} is the number of all semivariances of the directional empirical semivariogram, and N_{angle} is the number of

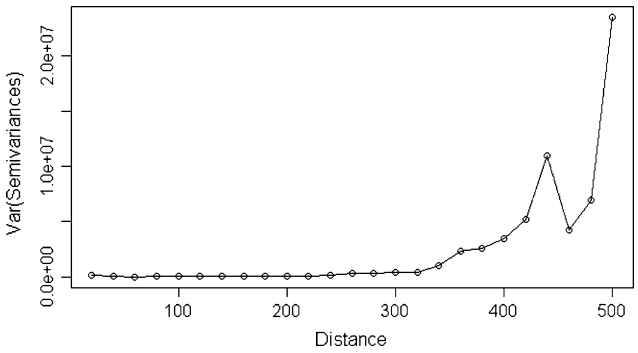

Fig. 3 Plot of the variances of semivariances of 6 directional empirical semivariograms. Up to a distance of 220 distance units, this function is almost constant. Variances start to increase slightly at a distance of 220 distance units, and increase rapidly at a distance of 320 distance units. The same data as in Fig. 2 was used

directions/angles. Subsequently, F_2 is compared to an $F(N_{\text{angle}} - 1; \text{DF}_{\text{aniso}})$ distribution and the P value is obtained. If the P value falls below a threshold value the anisotropic model is chosen. In our investigation, threshold P values of 10^{-5}(WLS) and 10^{-3} (OLS) proved to produce reasonable results, leading to decisions that coincided with the visual inspection. As pointed out previously, the test is approximate in that it is based on the assumption of independence and normality of the empirical semivariogram estimates.

An important point in using the F statistic as described above is the choice of a cutpoint, i.e. a specific distance at which the empirical semivariogram is cut, and all lag classes of distances greater than the cutpoint distance are discarded. As one can see in Fig. 2 (directional SVG), semivariances of different directions lie close together for smaller distances up to a point at which they start to scatter, which is a characteristic feature of isotropic models. This particular example depicts data for which an isotropic model was chosen. If the whole range of distances would have been used, the F test would have been significant, thus preferring an anisotropic model.

A useful tool in choosing this cut-off value is a plot of the variances of semivariances, the rationale being that we do not need to fit the model beyond a point where the sill has been reached for all directions. An example, which shows this type of plot referring to the data of Fig. 2, is depicted in Fig. 3. At a distance of 220, variances of semivariances (VSV) start to increase, whereas at distances smaller than 220 they are approximately equal. At a distance of 320 this is even more obvious for this example.

For anisotropic models the multidirectional semivariogram as well as the VSV-plot do not look similar compared to isotropic cases. An example is depicted in Fig. 4.

In the VSV-plot there is an initially increasing variance, which decreases after a local maximum to a local minimum. In our experience this is a common feature of anisotropic models fitted to cDNA microarray data, if one would classify them as anisotropic by visually inspecting the directional semivariogram. In our procedure this local minimum is chosen to trim the multidirectional semivariogram. Thus, only distances smaller than or equal to this cutpoint are used to fit the anisotropic and isotropic models. Choosing a cutpoint is restricted to distances greater than the minimum distance to avoid having too few degrees of freedom performing the F test described above.

To obtain a reasonable cutpoint we first compute all local minima of the empirical VSV-function $\min_{\text{loc}}(\text{VSV})$. Subsequently, all minima are removed that are located at distances smaller or equal to a minimum distance as described. If there are local minima left, the smallest of these minima is chosen. In case of a single minimum, this is chosen. Otherwise, ratios

$$r_i = \frac{v_i}{v_{i-1}}, i \in 2, \ldots, N$$

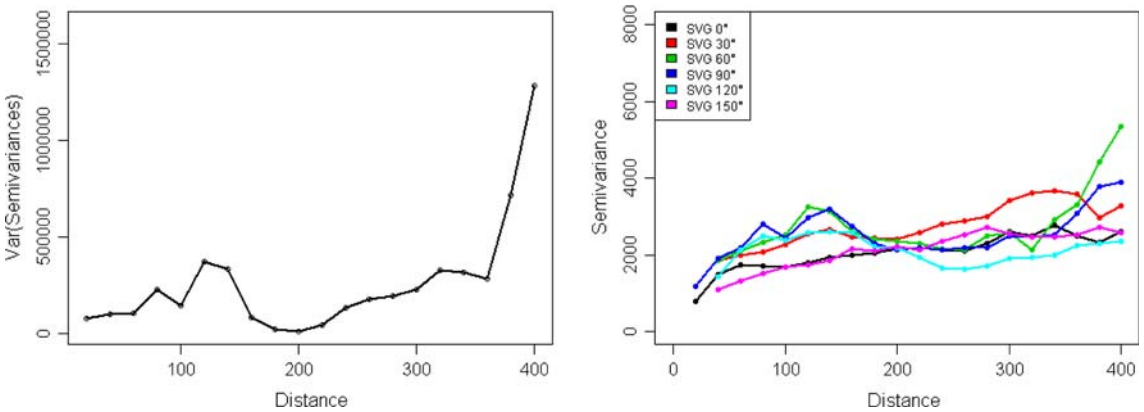

Fig. 4 Plot of variances of semivariances (*left panel*) of six directional empirical semivariograms (SVG, shown in the *right panel*). In contrast to Fig. 3, variances are not constant for shorter distances, and this function has its minimum at a mid-range distance. Data are part of a larger data set described in Piepho et al. (2006)

are computed with v_i being an element of all VSV-values greater than the minimum distance and N being the largest index of these values. The distance with the largest of these ratios is used as cutpoint.

Three approaches on BG smoothing

As stated in the introductory section we used three different algorithms for locally smoothing the BG values in a block-wise manner.

1. The most complex procedure comprises each step as described in the previous section on model selection. The obvious complexity of this algorithm is increased by the fact that in case a specific global trend model resulted in a singular Kriging system, where no smoothed BG values could be obtained, the second best global trend model is chosen. If this fails the next one is used. We will refer to this approach by OK.
2. This approach is a simplification of the first one. We left out the step on checking whether there is sufficient information for fitting a spatial model and used the constant global trend model. Kriging estimates were obtained using the best fitting isotropic model among the three theoretical models (Gaussian, exponential, spherical). Thus, we did not have to perform any F test since anisotropy was not considered. We will refer to this approach by OKiso emphasizing that we focus on isotropic models.
3. This approach consists of 2-D locally weighted regression on BG values. The concept of 2D LOWESS is quite simple. For a point x a LOWESS estimate is required. Let f be a number between 0 and 1 and let $q = fn$, q is truncated to an integer value. Among n points a neighbourhood of point x consisting of points x_i ($i = 1,\ldots, q$, $q \leq n$) is selected. A specific weight is assigned to each point within this neighbourhood:

$$w_i(x) = W\left[\frac{\rho(x_i - x)}{d(x)}\right],$$

where W [] is the tri-cube weight function, ρ is a distance function, d(x) is the distance of x to the qth nearest x_i. A quadratic function using weights $w_i(x)$ is fit to y_i in order to obtain an estimate for x (Cleveland et al. 1988, 1991). LOWESS estimates were obtained for BG values using a span-value $f = 0.35$ with four iterations to obtain a robust fit. We will refer to this method by Loess, which is also the name of the R-function that we used.

These three algorithms were used in conjunction with two BG correction methods, one of which has been found to be superior in a comparison study on different background correction methods (Ritchie et al. 2007). Ritchie et al. recommend the use of *normexp+offset*. The *normexp* algorithm is performed on BG corrected signal intensities. The observed intensities X are the result of a convolution of the true signal S and a noise component Y:

$$X = S + Y, S \sim \exp\left(\frac{1}{\alpha}\right), Y \sim N(\mu, \sigma^2)$$

The normal distribution for the noise component is truncated at zero, which models non-negative signal values. A conditional expectation estimates the signal value S (McGee and Chen 2006). This model is fitted to each dye-channel separately. Kernel density parameter estimation is done by using maximum likelihood (ML). For the *normexp+offset* method a small value is added to the corrected intensities which moves the corrected intensities away from zero and stabilizes the variance for small log-ratios (Ritchie et al. 2007). The typical fishtail effect in M-vs-A plots $[M = \log_2(R) - \log_2(G), A = 0.5\log_2(\mathrm{R}) + 0.5\log_2(\mathrm{G}), \mathrm{R} = \text{red signal}, \mathrm{G} = \text{green signal}]$, for small A-values is mitigated (see Fig. 7 below or Fig. 1 in Ritchie et al. 2007). In this study, we used an offset-value of 50 as suggested by Ritchie et al. (2007). The *normexp* method is similar to the background adjustment within the RMA algorithm for Affymetrix oligonucleotide microarrays (Irizarry et al. 2003).

For comparison, we also applied the traditional BG subtraction FG-BG (*subtract*), and the *Edwards* BG correction method. The latter method avoids negative BG corrected intensities by using a smoothing function of FG–BG values and a threshold δ. If S represents the BG corrected value (Edwards 2003), then:

$$S = \begin{cases} \mathrm{FG} - \mathrm{BG} & \text{if } \mathrm{FG} - \mathrm{BG} > \delta \\ \delta \exp[1 - (\mathrm{BG} + \delta)/\mathrm{FG}] & \text{otherwise} \end{cases}$$

We also applied the *normexp* and *normexp+offset* algorithms without BG smoothing to assess the effect of BG-smoothing prior to BG correction. Thus, there were ten BG correction methods overall: *subtract*, *Edwards*, *normexp*, *normexp+offset*, *OK & normexp*, *OKiso & normexp*, *Loess & normexp*, *OK & normexp+offset*, *OKiso & normexp+offset*, and *Loess & normexp+offset*.

Data

To simulate differentially expressed genes and to assess the behaviour of different BG correction methods under the null-hypothesis, we used a self-versus-self (SVS) dataset consisting of six custom-made microarrays. In total, 25,392 spots were assigned to 48 blocks, each one consisting of 23 × 23 spots (rows × columns). Each microarray represents a replicate of an Arabidopsis mRNA pool of plants, which were grown for 6 weeks in short days (10 h light, 100 microE, 20°C, 60% RH). Rosette leaves from 12 plants

were harvested, pooled and total RNA extracted. 100 mg of RNA were labelled with Cy3- or Cy5-dCTP and hybridized overnight. The microarrays contain gene-specific DNA fragments that were amplified by PCR. Scanning was performed with a ScanArray4000 (PerkinElmer). Scanning parameters were set so that not more than one to four spots were saturated in each subgrid. Spot intensities were obtained by using *Imagene* software with automated spot finding followed by manual checking and adjustment (Hilson et al. 2004; Little et al. 2007).

We used the SVS-data to simulate specific fractions of differentially expressed (DE) genes. At first, we allocated two genotypes randomly to each channel of an individual microarray. Then, we simulated 10% of all 25,392 genes as differentially expressed. We randomly selected four subsets, each consisting of 2.5% (634) of all genes. One half of the raw signal intensities were multiplied with an artificial fold change (FC) in genotype one, the other half of genes were multiplied with the same FC in genotype two. This corresponds to up- and down-regulation of genes in the subsets. We used FC = 1.5, FC = 2, FC = 3, and FC = 5 to simulate different classes of DE genes. To obtain more robust results, we repeated this simulation set up, consisting of genotype allocation and DE simulation, four times and averaged the results. For comparing the accuracy of BG correction methods, we repeated the previous steps with FCs of 2, 3, 5, and 10 to cover a wider range of DE. To adhere to the assumption that only a small number of genes are usually expected to be DE, we additionally created a dataset that consisted of 5% DE genes with FC = 3. Therefore, we obtained three additional datasets from the SVS-data, which were used to classify all ten BG correction methods.

After applying the three BG smoothing algorithms and/or BG correction, the data were normalized prior to estimating genotypic differences. A single chip was normalized with Loess-within-chip-normalization. The log-ratios of microarrays were then scaled to have the same absolute median deviation (MAD), which is sometimes called between-chip-normalization (Smyth and Speed 2003). Since Loess-normalization operates on log-ratios, any spot with higher BG than signal intensity represented a missing value for the traditional BG subtraction. All other BG correction methods avoid negative BG corrected values.

Computing pair-wise linear contrasts

To assess the performance of a specific background correction procedure, we estimated pairwise linear contrasts of two genotypes using a simple linear model. The SVS-dataset consists of data where no truly significant results were expected since there were no genotypic differences. Therefore, we allocated two genotypes (A, B) randomly to the six SVS-chips in such a way that each chip contained both genotypes, and each genotype was present with the same number of replicates in each channel. So, we had six replicates of genotype A, and six replicates of genotype B, each one present three times in the Cy3-channel, and three times in the Cy5-channel. DE genes were simulated as described earlier.

For all data, we fitted the following linear model:

$$y_{ijk} = \alpha_i + \beta_j + \gamma_k + e_{ijk},$$

where y_{ijk} represents the spot intensity for genotype $i \in \{1,\ldots, N_{\text{geno}}\}$ on array $j \in \{1,\ldots, N_{\text{array}}\}$ coming from dye-channel $k \in \{1, 2\}$, while α_i, β_j, and γ_k are fixed effects for genotype, array and dye, respectively. The term e_{ijk} represents the independent and identically distributed residual error with $e_{ijk} \sim N(0, \sigma^2)$. Subsequently, the difference in expression of genotypes A and B were subjected to an ordinary t test using the appropriate linear contrast based on a least-squares fit of the linear model (Searle 1971).

We also used the moderated t statistic to identify DE genes. This is an empirical Bayes approach which makes use of all estimated sample variances; these are shrunk to a pooled estimate. Using the moderated t statistic makes inference far more stable, especially in small microarray experiments (Smyth 2004). We used the same model as shown above. The moderated t statistic is defined as

$$\tilde{t}_{gj} = \frac{\hat{\beta}_{gj}}{\tilde{s}_g \sqrt{\upsilon_{gj}}},$$

where $\hat{\beta}_{gj}$ is the jth estimated contrast of gene g, and υ_{gj} is the jth diagonal element of the estimated covariance matrix for the contrast matrix of gene g, $\text{var}\left(\hat{\beta}_g\right) = C' V_g C s_g^2$. Matrix C defines the contrasts to be estimated for the gth model fit, and $V_g s_g^2$ is the estimated covariance matrix of the coefficient estimator $\hat{\alpha}_g$. The posterior mean of the sample variance $\tilde{s}_g^2$ of gene g is defined as

$$\tilde{s}_g^2 = \frac{d_0 s_0^2 + d_g s_g^2}{d_0 + d_g},$$

where s_g^2 is the residual variance estimator for gene g, with d_g residual degrees of freedom. Prior information is assumed on the residual variances for each gene by using a prior estimator $\tilde{s}_0$ with d_0 degrees of freedom, which are estimated from the data.

The relation between the residual variance of gene g and these prior values is expressed by

$$\frac{1}{\sigma_g^2} \sim \frac{1}{d_0 s_0^2} \chi_{d_0}^2$$

See Smyth (2004) for a thorough development of this empirical Bayes methodology.

We adjusted *P* values for multiplicity by using Storey's *q* value (Storey 2002). This type of adjustment was developed for very large numbers of comparisons and is therefore a natural choice for microarray data comprising of 25,392 genes.

Implementation

All of the previously described steps were implemented using the freely available statistical language R (http://cran.R-project.org). Variogram calculations, non-linear OLS/WLS fitting of theoretical models to isotropic semivariograms, and ordinary kriging were performed using the geoR-package (Ribeiro and Diggle 2001). Non-linear OLS/WLS estimation of isotropic/anisotropic models for multidirectional semivariograms was done with the *optim* function of the R stats-package. Specifically, we used the "L-BGFS-B" algorithm, which allows box-constraints (Byrd et al. 1995, citing the R-help page of function *optim*) to set lower and upper bounds. BG correction, normalization and identifying differentially expressed genes using the moderated *t* statistic were performed using the limma-package of Smyth (citing the R-help page of the limma package; http://www.bioconductor.org). Loess-smoothing was done using the R-function loess. Fitting the linear model was done with the *lm* function and pairwise linear contrasts were computed with the *glht* function of the *multcomp* R-package. FDR-adjustment was done with the *qvalue* R-package.

Results

When we applied the geostatistical framework (OK, OKiso), we used a lag distance of 20 distance units, and 25 lag classes, which resulted in a maximum lag distance of 500 distance units for computation of the empirical semivariograms. The spots of the SVS-data were arranged in a grid with separation distance of approximately 20 distance units in both directions, which led to this choice of parameters.

Table 1 outlines the results obtained for all ten BG correction approaches for the SVS-data with four classes of simulated DE genes, at simulated FC of 1.5, 2, 3 and 5, using the linear model with linear contrasts (ordinary-t) averaged over four simulation cycles. Raw *P* values were adjusted for multiplicity by Storey's *q* value, which resulted in observed false discovery rates (FDR) ranging from 4.78% (*Loess & normexp+offset*) to 6.04% (*normexp*) at the 5% significance level and 9.08% (*normexp*, *OK & normexp*) to 9.84% (*Edwards*) at the 10% nominal alpha-level. The subtract method performed worst. It detected by far fewer simulated DE genes and had by far the highest observed FDR. The best performing method is *Loess & normexp+offset*. For each class corresponding to a specific simulated FC, this method identified most DE genes. Smoothing BG values prior to BG correction clearly increased power (Fig. 6), and lowered the proportion of false discoveries (Fig. 5). Both, *normexp* and *normexp+offset* benefit from smoothing BG values using either OKiso or Loess. The complex OK approach (*OK & normexp, OK & normexp+offset*) did not improve the performance of *normexp* and *normexp+offset* methods. The results were consistent for two different significance thresholds used to identify DE genes ($\alpha = 5\%$, $\alpha = 10\%$).

A similar outcome was observed when Smyth's moderated *t* tests were used to identify DE genes (Table 2). Observed false-positive rates were smaller than for the ordinary *t* statistic (*OK & normexp* 2.72–4.2% for *Edwards* at the 5% significance level; *OK & normexp* 6.74–8.72%

Table 1 Number of true discoveries for ten different background correction methods using ordinary *t* tests

	$\alpha = 5\%$				$\alpha = 10\%$			
	FC = 1.5	FC = 2	FC = 3	FC = 5	FC = 1.5	FC = 2	FC = 3	FC = 5
subtract	27.75	35.50	33.50	37.25	43.25	51.75	47.75	54.00
Edwards	85.25	222.75	378.00	506.00	169.50	349.50	501.00	584.00
normexp	19.50	49.25	99.75	153.25	115.75	256.75	442.00	567.50
normexp+offset	34.50	93.75	187.25	297.75	130.50	290.50	478.75	590.75
OK & normexp	18.75	46.25	92.00	135.00	106.75	241.75	414.25	547.75
OKiso & normexp	69.25	202.50	387.75	536.75	158.00	341.50	527.25	616.75
Loess & normexp	86.25	226.75	428.50	572.25	176.25	372.75	547.25	619.50
OK & normexp+offset	33.00	79.00	167.00	252.75	124.00	281.00	457.50	584.25
OKiso & normexp+offset	85.75	225.50	431.25	569.25	171.00	366.50	544.50	622.00
Loess & normexp+offset	**100.00**	**259.50**	**465.00**	**590.75**	**189.75**	**397.00**	**560.75**	**624.25**
Simulated DE genes	634	634	634	634	634	634	634	634

Columns correspond to specific simulated fold changes at different significance levels. Results were averaged over four independent simulations, and adjusted for multiplicity using Storey's *q* value

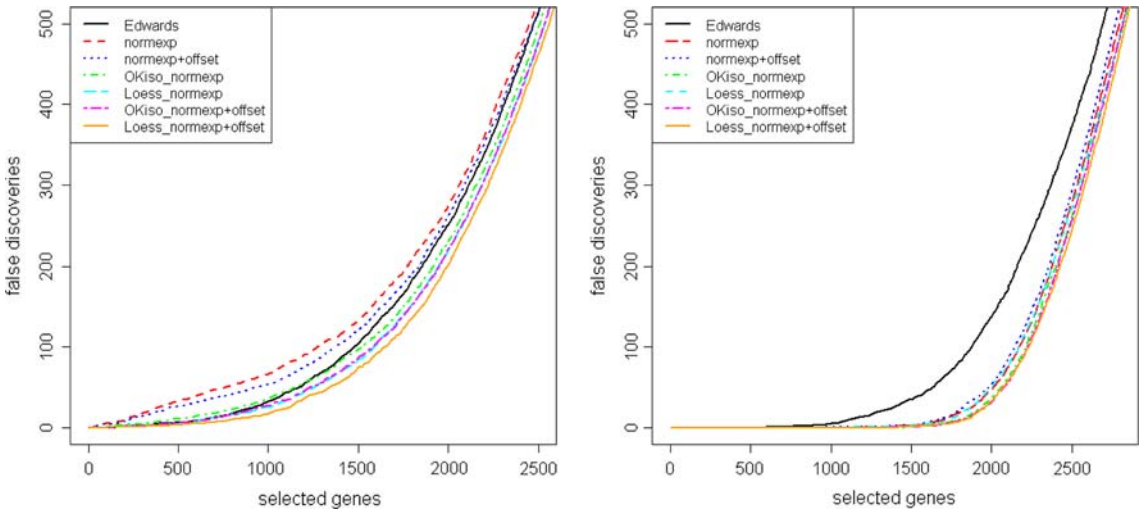

Fig. 5 Trade-off between the number of selected genes and the number of false discoveries for the SVS-data with four classes of DE genes. Results were averaged over four independent simulations, and adjusted for multiplicity using Storey's q value approach. *Left* DE genes were identified using ordinary t tests. *Right* DE genes were identified using moderated t tests

Table 2 Number of true discoveries for ten different background correction methods using moderated t tests

	$\alpha = 5\%$				$\alpha = 10\%$			
	FC = 1.5	FXC = 2	FC = 3	FC = 5	FC = 1.5	FC = 2	FC = 3	FC = 5
subtract	49.25	45.25	57.50	60.00	58.50	54.25	68.00	68.25
Edwards	122.50	400.25	558.00	613.00	223.75	492.00	596.75	625.25
normexp	231.75	500.00	620.50	633.50	321.00	566.00	629.00	633.50
normexp+offset	230.50	488.75	617.50	633.75	317.25	558.00	628.75	633.75
OK & normexp	221.75	506.00	622.75	632.75	309.50	566.00	631.00	633.25
OKiso & normexp	244.25	550.50	628.00	633.75	330.75	585.75	631.50	633.75
Loess & *normexp*	228.00	542.50	624.25	633.75	334.75	586.00	629.25	**634.00**
OK & normexp+offset	226.00	495.00	620.25	632.50	309.75	559.50	630.50	633.25
OKiso & normexp+offset	260.75	557.00	**628.50**	**634.00**	346.75	590.50	**632.00**	**634.00**
Loess & normexp+offset	**270.75**	**566.75**	628.00	633.75	**369.00**	**596.25**	630.50	**634.00**
Simulated DE genes	634	634	634	634	634	634	634	634

Columns correspond to specific simulated fold changes at different significance levels. Results were averaged over four independent simulations, and adjusted for multiplicity using Storey's q value

for *Loess & normexp* at the 10% significance level). Again, the subtract method had by far the highest observed false-positive rates (88.39, 88.59%). An obvious feature of the moderated t statistic was the increased number of identified DE genes. This is the result of borrowing information across genes, and therefore, having more degrees of freedom available. Application of OKiso and Loess increased the power (Fig. 6) and decreased the number of false discoveries (Fig. 5).

To circumvent misleading inference due to a too large proportion of DE genes, we also applied all ten methods to the SVS-data with a single class of DE-genes. There were 5% of DE genes at a simulated FC = 3. Results for the ordinary t tests as well as those of the moderated t tests are shown in Table 3.

These results confirm our findings for the SVS-data with four classes of 10% simulated DE genes. The complex and time-consuming OK approach (*OK & normexp*, *OK & normexp+offset*) does not perform as good as the other two BG-smoothing algorithms (*OKiso*, *Loess*). We conclude that applying this complex framework as described in the "Material and methods" section is not justified by our findings. We do not include the results of this BG smoothing algorithm in Figs. 5 and 6. As one can see in Tables 1, 2 and 3, performing the traditional BG subtraction is the worst BG correction method overall.

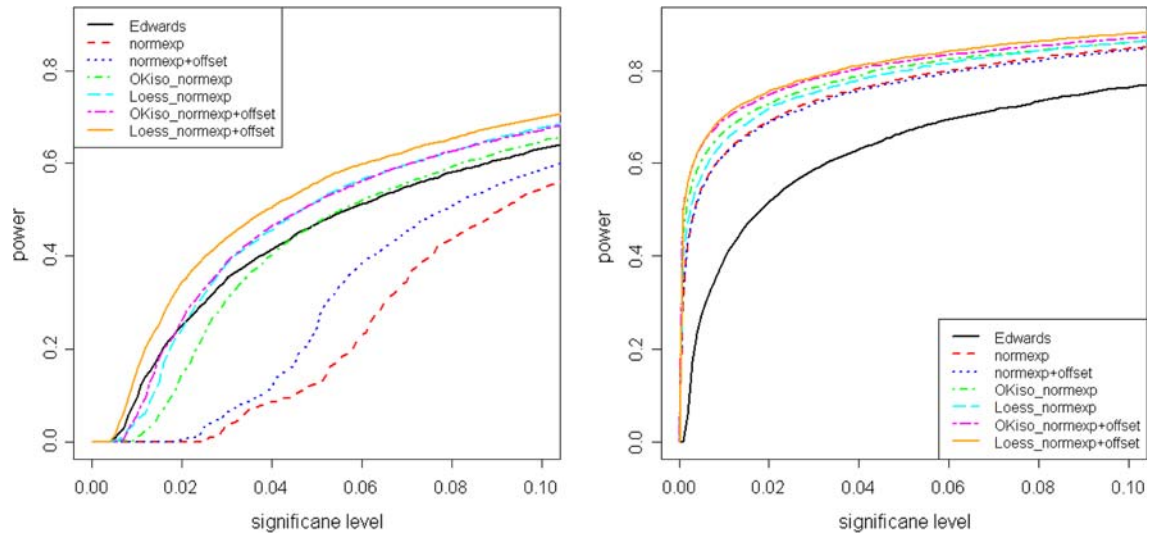

Fig. 6 Power for seven background correction methods up to a significance level of 10%. Results were obtained as average over four independent simulations. *Left plot* shows results for ordinary *t* tests, *right plot* shows results of moderated *t* tests

Table 3 Number of true discoveries for ten different background correction methods using ordinary *t* tests and moderated *t* tests

	Fold Change = 3					
	Ordinary *t*			Moderated *t*		
	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 0.1\%$	$\alpha = 1\%$	$\alpha = 5\%$
subtract	0	43.50	61.50	0	40.50	74.25
Edwards	5.25	627.50	881.25	0	618.25	1067.50
normexp	0	41.00	344.50	343.75	1,070.50	1,232.25
normexp+offset	0.25	59.75	604.00	329.25	1,047.25	1,227.25
OK & normexp	0	26.00	243.25	266.25	1,082.25	1,235.50
OKiso & normexp	0	452.50	878.50	788.00	1,185.75	1,251.25
Loess & normexp	4.00	636.25	966.25	611.50	1,163.25	1,246.50
OK & normexp+offset	0	50.25	488.50	287.50	1,059.50	1,229.50
OKiso & normexp+offset	21.25	632.75	953.50	916.25	1,204.75	1,253.00
Loess & normexp+offset	**24.75**	**765.25**	**1,023.25**	**968.50**	**1,218.00**	**1,258.25**
DE genes	1,269	1,269	1,269	1,269	1,269	1,269

Columns correspond to different significance levels at a single simulated Fold Change. Results were averaged over four independent simulations, and adjusted for multiplicity using Storey's *q* value approach

To further assess the different BG correction approaches, we investigated the trade-off between the number of genes that were termed significant and the number of false discoveries among them. These findings are depicted in Fig. 5. The *subtract* method is not shown because the very large numbers of false discoveries would distort this plot. The corresponding line would run way left of the others. Extending established BG correction methods (*normexp*, *normexp+offset*) by BG smoothing (Loess, OKiso) lowered the proportion of false discoveries (Fig. 5) and increased the power of finding DE genes (Fig. 6).

The raw SVS-data served as means to check whether different methods perform as expected under the null-hypothesis (H_0) or if the proportion of false significances exceeds the nominal threshold. In case there are no significant differences between both genotypes one would expect observed *P* values to show a uniform $U(0, 1)$ distribution. This corresponds to a straight line in the plot of *P* values of the linear contrasts (genotypic differences) versus quantiles of *P* values. None of the six methods exhibit peculiar departure from the expected proportion of 5% false significant results (data not shown). This was also confirmed by our observed false discovery rates for the SVS-data as mentioned earlier.

The set-up of this simulation study allowed to compare the accuracy of DE classification for different background

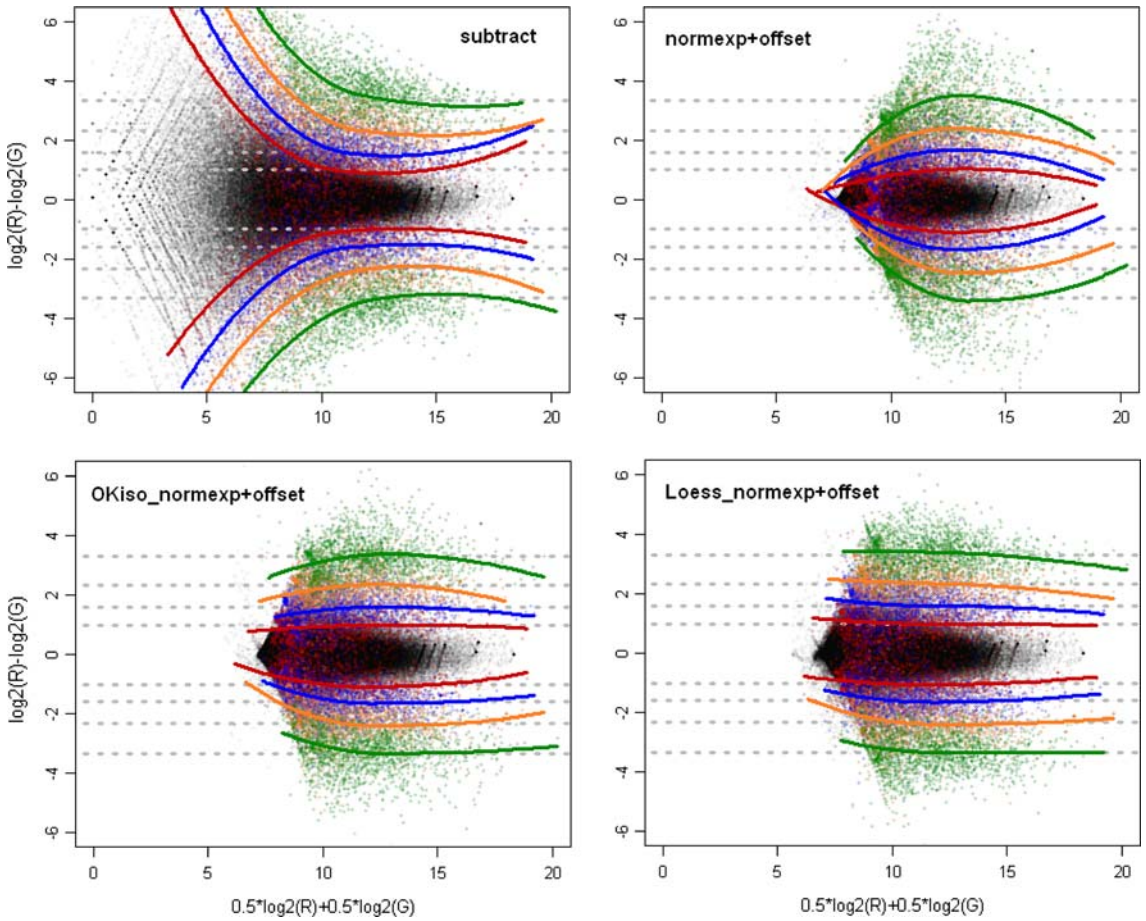

Fig. 7 M versus A plot averaged over microarrays of the self-versus-self data. Classes of simulated DE genes are coloured (*red* FC = 2, *blue* FC = 3, *orange* FC = 5, *green* FC = 10). *Dotted grey lines* correspond to the expected value of each DE-class ($\log_2$-scale), *solid coloured lines* correspond to LOWESS-lines with span = 1 per DE-class

correction methods. Since we simulated DE genes with specific FCs, we expected the four classes of DE genes to be located around distinct horizontal lines in the M versus A plots. Specifically, for simulated FCs of 2, 3, 5 and 10, the different classes were expected to scatter around $-\log_2(10)$, $-\log_2(5)$,..., $\log_2(5)$, $\log_2(10)$ lines. We averaged the results of background corrected and normalized data over microarrays to compare different background correction methods.

Figure 7 depicts the results for *subtract*, *normexp+offset*, *OKiso & normexp+offset, and Loess & normexp+offset*. For each class of DE, we added a scatterplot LOESS-smoother (span = 1) which summarizes the outcome for a specific background correction method. The most accurate method is *Loess & normexp+offset*. The LOESS-fits for this method reflect the expectation to the greatest extent and showed the lowest degree of curvature over the range of expression intensities (*X* axis). Results obtained without using the offset were similar (not shown), although the *OKiso & normexp* method performed better in comparison to *Loess & normexp*, which in turn outperformed *normexp*. Using BG smoothing prior to BG correction resulted in more stable patterns of DE, i.e. the simulated DE genes were less dependent on the mean expression level and the estimated FC better reflected the simulated FC. The upper left plot of Fig. 7 (subtract) reveals the most severe departure from the expectation.

Discussion

We have investigated the general applicability of geostatistical methods represented by a complex *ordinary kriging* approach accounting for anisotropic correlation models (OK) and a simple *ordinary kriging* approach focussing on isotropic models (OKiso), and additionally, a 2D-LOWESS (Loess) smoother for cDNA microarray BG correction. We combined these BG smoothing methods with two BG correction procedures that avoid negative BG corrected intensities (*normexp*, *normexp+offset*). These BG correction methods were found to be superior compared to the standard method, and in case of the latter one, the best

overall (Ritchie et al. 2007). Our focus was on comparing these to the standard method and three established BG correction methods (*Edwards*, *normexp*, *normexp+offset*). After simulating DE genes for a self-versus-self dataset, we used two different methods to detect DE genes: ordinary *t* tests and more sophisticated moderated *t* tests (Smyth 2004). We did not consider more such approaches to estimating these genotypic differences as for example the random variance model proposed by Wright and Simon (2003) or the empirical Bayes approach based on mixed models by Sarholz and Piepho (2008). We expect that the results we obtained for moderated *t* tests would be similar for other methods since our approach is restricted to the BG values, and only aims at improving BG values in terms of adhering to the implicit assumptions of BG correction, a locally constant BG as stated in Kooperberg et al. (2002). Our proposed BG smoothing is an additional step, performed prior to BG correction, and can be combined with any BG correction method. Our findings show that BG smoothing prior to BG correction is beneficial compared to leaving this step out. The OKiso procedure shows that methods which exploit the local correlation structure of BG values might be a useful means within the cDNA microarray pre-processing pipeline.

Although we used pairwise linear contrasts in our study, our methodology is not restricted to this kind of comparison. In heterosis research, for example, two types of comparison are of primary interest, mid-parent heterosis (MPH) and best-parent heterosis (BPH) (Keller et al. 2005; Uzarowska et al. 2007, 2009; Höcker et al. 2008; Paschold et al. 2009; Frisch et al. 2009; Thiemann et al. 2009; Jahnke et al. 2009). BPH, the hybrid being above the better parent, is estimated using a pairwise linear contrast, whereas MPH, the hybrid being significantly different from the average of both parents, consists of three terms. BPH, MPH, and any other comparison (contrast) are expected to benefit from the higher accuracy of the estimated differences between probes on single microarrays (see Fig. 7).

At this developmental stage, we would propose the 2D-LOWESS smoother (Loess) in combination with *normexp+offset* (Ritchie et al. 2007). Using this combination produced the best results regarding the number of false discoveries, power and accuracy. If *normexp* is the BG correction method of choice, we suggest the OKiso method. This approach showed to be less sensitive regarding the choice of an offset. It is fast and improved the outcome in comparison to using *normexp* alone in terms of false discoveries, power and accuracy. Both geostatistical approaches (OKiso, OK) were outperformed by Loess, with OKiso being close to Loess. We believe that this is due to the fact that only the Loess approach smoothes BG values of each block of a microarray. The other two approaches use the original values if any errors occur during estimation or, in case of OK, when the algorithm does not find a model of spatial correlation that is distinguishable from a pure nugget model. Using a simple smoothing strategy in these situations could help the OK and OKiso approaches to better compete with Loess. Combining background smoothing techniques with established background correction methods works equally well using ordinary *t* tests and the empirical Bayes (moderated *t* test) method (see Figs. 5, 6). The *normexp* and *normexp+offset* background correction methods without BG smoothing did way better when using moderated *t* tests. They were even outperformed by the *Edwards* method when ordinary *t* tests were used to find DE genes. We were surprised that the complex OK method could not improve the performance of *normexp* and *normexp+offset*. We conclude that anisotropy and global trend do not have to be accounted for when smoothing BG values. The former is also supported by the fact that OKiso and Loess combined with either *normexp* or *normexp+offset* do not use directional information.

One might argue that restricting smoothing of BG values to only some regions of the microarray, thus treating blocks of a microarray differently, introduces further bias. We believe, however, that confinement to blocks is, in fact, a particular strength of our methodology because the spatial correlation structure of BG values is often restricted to specific parts of the chip surface. This can be seen for example in the two heatmaps in Fig. 1. Application of smoothing methods in a blockwise manner allows accounting for these local features.

Using maximum likelihood (ML) or restricted maximum likelihood (REML) estimation would be a natural approach of finding the theoretical model of spatial correlation to be used for the OK or OKiso methods. We did not use ML/REML, however, because of the high rate of models which did not converge. Furthermore, ML/REML estimation is way more time-consuming than weighted least squares (WLS). We implemented a pipeline that allows performing each step automatically with an appropriate handling of any errors that could occur, e.g. a singular kriging system. This particular exception was handled by applying the next best fitting global trend model. We considered 13 trend surface models, and the most extreme case was that only the ninth best fitting trend surface model resulted in a solvable kriging system (OK).

Based on our results, we would propose to use Loess or OKiso for smoothing BG values prior to applying a BG correction algorithm. Using our time-consuming implementation of the OK approach is not justified by our findings, although we still believe that it is worth further investigation. In case there are departures from homogeneous local BG with severe anisotropic character, we think that this should be accounted for.

Acknowledgments We thank Caroline Gouhier-Darimont and Philippe Reymond (Department of Plant Molecular Biology, University of Lausanne, 1015 Lausanne, Switzerland) for providing the self-vs-self dataset. This work was funded by the *Deutsche Forschungsgemeinschaft* (DFG) within the priority program *SPP1149-Heterosis in Plants* (grant-no. PI 377/7-3). Prof. Uwe Jensen and two anonymous reviewers are thanked for helpful and constructive comments on earlier versions of the paper.

Conflict of interest statement The authors declare that they have no conflict of interest.

References

Arteaga-Salas JM, Harrison AP, Upton GJG (2008) Reducing spatial flaws in oligonucleotide arrays by using neighborhood information. Stat Appl Genet Mol Biol 7 (1), Article 29

Beatty M, Hondred D, Fengler K, Li B, Rafalski A, Beló A (2009) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet (this issue) (in press)

Byrd RH, Lu P, Nocedal J, Zhu C (1995) A limited memory algorithm for bound constrained optimization. SIAM J Sci Comput 16:1190–1208

Cleveland WS, Grosse E (1991) Computational methods for local regression. Stat Comput 1:47–62

Cleveland WS, Devlin SJ, Grosse E (1988) Regression by local fitting—methods, properties, and computational algorithms. J Econom 37:87–114

Colantuoni C, Henry G, Zeger S, Pevsner J (2002) Local mean normalization of microarray element signal intensities across an array surface: quality control and correction of spatially systematic artifacts. BioTech 32:1316–1320

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

Cressie NAC, Hawkins DM (1980) Robust estimation of the variogram: I. Math Geol 12:115–125

Diggle PJ, Ribeiro PJ Jr (2007) Model-based geostatistics. Springer, New York

Edwards D (2003) Non-linear normalization and background correction in one-channel cDNA microarrays. Bioinformatics 19:825–833

Frisch M, Thiemann A, Fu J, Schrag TA, Scholten S, Melchinger AE (2009) Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor Appl Genet (this issue) (in press)

Fujita A, Sato JR, de Oliveira Rodrigues L, Ferreira CE, Sogayar MC (2006) Evaluating different methods of microarray data normalization. BMC Bioinformatics 7, Article 469

Haldermans P, Shkedy Z, Van Sanden S, Burzykowski T, Aerts M (2007) Using linear mixed models for normalization of cDNA microarrays. Stat Appl Genet Mol Biol 6 (1), Article 19

Hilson P, Allemeersch J, Altmann T, Aubourg S, Avon A, Beynon J, Bhalerao RP, Bitton F, Caboche M, Cannoot B, Chardakov V, Cognet-Holliger C, Colot V, Crowe M, Darimont C, Durinck S, Eickhoff H, Falcon de Longuevialle A, Farmer EE, Grant M, Kuiper MTR, Lehrach H, Léon C, Leyva A, Lundeberg J, Lurin C, Moreau Y, Nietfeld W, Paz-Ares J, Reymond P, Rouzé P, Sandberg G, Segura MD, Serizet C, Tabrett A, Taconnat L, Thareau V, Van Hummelen P, Vercruysse S, Vuylsteke M, Weingartner M, Weisbeek PJ, Wirta V, Wittink FRA, Zabeau M, Small I (2004) Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. Genome Res 14:2176–2189

Höcker N, Keller B, Chollet D, Descombes P, Piepho HP, Hochholdinger F (2008) Comparison of maize (*Zea mays*) hybrid and parental inbred line primary root transcriptomes suggests organ specific patterns of non-additive gene expression and conserved expression trends between different hybrids in a subset of genes. Genetics 179:1275–1283

Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18(Suppl.1):96–104

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264

Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics. Oxford University Press, New York

Jahnke S, Sarholz B, Thiemann A, Kühr V, Gutierrez-Marcos JF, Geiger HH, Piepho HP, Scholten S (2009) Heterosis in early seed development: A comparative study of F1 embryo and endosperm tissues six days after fertilization. Theoret Appl Genet (this issue) (in press)

Keller B, Emrich K, Hoecker N, Sauer M, Hochholdinger F, Piepho H-P (2005) Designing a microarray experiment to estimate dominance in maize (*Zea mays* L.). Theoret Appl Genet 111:57–64

Kooperberg C, Fazzio TG, Delrow JJ (2002) Improved background correction for spotted DNA microarrays. J Comput Biol 9:55–66

Little D, Gouhier-Darimont C, Bruessow F, Reymond P (2007) Oviposition by pierid butterflies triggers defense gene expression in Arabidopsis. Plant Physiol 143:784–800

Mary-Huard T, Daudin J-J, Robin S, Bitton F, Cabannes E, Hilson P (2004) Spotting effect in microarray experiments. BMC Bioinformatics 5, Article 63

McGee M, Chen Z (2006) Parameter estimation for the exponential-normal convolution model for background correction of Affymetrix GeneChip data. Stat Appl Genet Mol Biol 5 (1), Article 24

McQuarrie ADR, Tsai CL (1998) Regression and time series model selection. World Scientific, Singapore

Nelder JA (2000) Functional marginality and response-surface fitting. J Appl Stat 27:109–112

Neuvial P, Hupe P, Brito I, Liva S, Manie E, Brennetot C, Radvanyi F, Aurias A, Barillot E (2006) Spatial normalization of array-CGH data. BMC Bioinformatics 7, Article 264

Paschold A, Marcon C, Hoecker N, Hochholdinger F (2009) Molecular dissection of heterosis manifestation during early maize root development. Theor Appl Genet (this issue) (in press)

Piepho HP, Keller B, Hoecker N, Hochholdinger F (2006) Combining signals from spotted cDNA microarrays obtained at different scanning intensities. Bioinformatics 22:802–807

Ribeiro PJ Jr, Diggle PJ (2001) geoR: a package for geostatistical analysis. R-NEWS 1(2):15–18

Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK (2007) A comparison of background correction methods for two-color microarrays. Bioinformatics 23:2700–2707

Robinson GK (1991) That BLUP is a good thing: the estimation of random effects. Stat Sci 6:15–51

Sarholz B, Piepho H-P (2008) Variance component estimation for mixed model analysis of cDNA microarray data. Biom J 50:927–939

Schabenberger O, Pierce FJ (2002) Contemporary statistical models for the plant and soil sciences. CRC Press, Boca Raton

Scharpf RB, Iacobuzio-Donahue CA, Sneddon JB, Parmigiani G (2006) When should one subtract background fluorescence in two color microarrays? Biostatistics 8:695–707

Schena M (2003) Microarray analysis. Wiley, Hoboken
Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzel H (2000) Normalization strategies for cDNA microarrays. Nucleic Acids Res 28, Article 47
Searle SR (1971) Linear Models. Wiley, New York
Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3, Article 3
Smyth GK, Speed TP (2003) Normalization of cDNA microarray data. Methods 31:265–273
Storey JD (2002) A direct approach to false discovery rates. J R Stat Soc B 64:479–498
Thiemann A, Fu J, Schrag TA, Melchinger AE, Frisch M, Scholten S (2009) Correlation between parental transcriptome and field data for the characterization of heterosis in *Zea mays* L. Theoret Appl Genet (this issue) (in press)
Tran PH, Peiffer DA, Shin Y, Meek LM, Brody JP, Cho KWY (2002) Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals. Nucleic Acids Res 30, Article 54
Uzarowska A, Keller B, Piepho HP, Schwarz G, Ingvardsen C, Wenzel G, Lübberstedt T (2007) Comparative expression profiling in meristems of inbred–hybrid triplets of maize. Plant Mol Biol 63:21–34
Uzarowska A, Dionisio G, Sarholz B, Piepho HP, Xu M, Ingvardsen C, Wenzel G, and Lübberstedt T (2009) Validation of candidate genes putatively associated with resistance to SCMV and MDMV in maize (*Zea mays* L.) by expression profiling. BMC Plant Biology 9, Article 15
Wright GW, Simon RM (2003) A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics 19:2448–2455
Yang YH, Buckley MJ, Speed TP (2001) Analysis of cDNA microarray images. Brief Bioinform 2:341–349
Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30, Article 15
Yin W, Chen T, Zhou XS, Chakraborty A (2005) Background correction for cDNA microarray images using the TV+L1 model. Bioinformatics 21:2410–2416
Yuan DS, Irizarry RA (2006) High-resolution spatial normalization for microarrays containing embedded technical replicates. Bioinformatics 22:3054–3060