Johns Hopkins University, Dept. of Biostatistics Working Papers

9-29-2003

Stochastic Models Based on Molecular Hybridization Theory for Short Oligonucleotide Microarrays

Zhijin Wu
Johns Hopkins Bloomberg School of Public Health, Zhijin_Wu@brown.edu

Richard LeBlanc
none

Rafael A. Irizarry
Johns Hopkins Bloomberg School of Public Health, rafa@jhu.edu

Stochastic Models Based on Molecular Hybridization Theory for Short Oligonucleotide Microarrays

Zhijin Wu, Richard Le Blanc, and Rafael A. Irizarry[1]

September 19, 2003

[1]To whom correspondance should be addressed: rafa@jhu.edu

Abstract

High density oligonucleotide expression arrays are a widely used tool for the measurement of gene expression on a large scale. Affymetrix GeneChip arrays appear to dominate this market. These arrays use short oligonucleotides to probe for genes in an RNA sample. Due to optical noise, non-specific hybridization, probe-specific effects, and measurement error, ad-hoc measures of expression, that summarize probe intensities, can lead to imprecise and inaccurate results. Various researchers have demonstrated that expression measures based on simple statistical models can provide great improvements over the ad-hoc procedure offered by Affymetrix. Recently, physical models based on molecular hybridization theory, have been proposed as useful tools for prediction of, for example, non-specific hybridization. These physical models show great potential in terms of improving existing expression measures. In this paper we demonstrate that the system producing the measured intensities is too complex to be fully described with these relatively simple physical models and we propose empirically motivated stochastic models that compliment the above mentioned molecular hybridization theory to provide a comprehensive description of the data. We discuss how the proposed model can be used to obtain improved measures of expression useful for the data analysts.

1 Introduction

In the Affymetrix system, a fair amount of further pre-processing and data reduction occurs following the image processing step to obtain measures of gene expression. Background adjustments, normalization, and summarization of the probe level data are three typical steps. The model proposed in this paper is especially useful for the background adjustment step thus we will focus our discussion on this aspect. However, in Section 6 we briefly discuss how it can be useful for normalization and summarization as well.

Affymetrix GeneChip arrays use short oligonucleotides to probe for genes in an RNA sample. Each gene will be represented by 11-20 pairs of oligonucleotide probes. The first component of these pairs is referred to as a *perfect match* (PM) probe and is designed to be specific to transcripts from the intended gene. However, non-specific hybridization and optical noise are unavoidable. Therefore, the observed intensities need to be adjusted to give accurate measurements of specific hybridization. Affymetrix's approach to adjusting is to pair each perfect match probe with a *mismatch* (MM) probe, that is designed by changing the middle (13th) base, with the intention of measuring only optical background noise and non-specific hybridization (NSB). The default adjustment, provided as part of the Affymetrix system, is based on the difference between perfect match and mismatch probe intensities($PM - MM$).

A final step in the pre-processing of these arrays is to combine the 11-20 probe pair intensities, after background adjustment and normalization, for a given gene to define a measure of expression that represents the amount of the corresponding mRNA species. Affymetrix's default algorithm, MAS 5.0, is based on a robust average of the log $PM - MM$ values (some tweaking is performed to avoid logs of negatives). Various researchers have developed alternative algorithms, motivated by statistical models, that outperform the default algorithm in many applications. For example, Li and Wong (2001) notice a strong probe effect in both PM and $PM - MM$ and describe it via a simple multiplicative model. By analyzing various arrays at once they are able to estimate probe effects and use this to improve outliers detection. Li and Wong also propose a non-linear normalization procedure that improves precision of the default re-scaling approach. Irizarry et al. (2003a) demonstrate that the $PM - MM$ transformation results in gene expression estimates with exaggerated variance. As a practical solution, they propose a global background adjustment step that ignores the MM intensities. This approach sacrifices some accuracy for large gains in precision. After the global background adjustment, arrays are quantile normalized (Bolstad et al., 2003) and a log-scale expression effect plus probe effect model is fitted robustly to define the robust multi-array analysis (RMA) expression measure. Irizarry et al. (2003b) (Irizarry et al., 2003c) and Cope et al. (2003) (Cope et al., 2003) demonstrate that RMA outperforms MAS 5.0 and the Li and Wong procedure in various practical tasks. RMA has been implemented in the Bioconductor project (`http://www.bioconductor.org`) *affy* package (Irizarry et al., 2003b), Iobion's Genetraffic (`http://www.iobion.com/`, and Insightful's S+ArrayAnalyzer (`http://www.insightful.com/products/s-plus_arrayanalyzer/`) and has become a popular alternative to the default algorithm provided by Affymetrix. Various other algorithms have been proposed (Holder et al., 2001; Workman et al., 2002; Naef et al., 2001; Chu et al., 2002; Zhang et al., 2002). have been proposed. In Section 6 we will argue that the model described in this paper can be used to improve these methods, especially in terms of their accuracy.

A simple version of our model can be written as $PM = O + N + S$ with PM the measured intensity of particular PM probe, O representing optical background noise for this probe, N representing NSB and S represents specific signal. Similar models have been proposed by, for example, Hekstra et al. (2003) and Zhang, Miles and Aldape (2003) . A deterministic model that motivates Affymetrix's approach to background adjustment would be $MM = O + N$ which would imply that $PM - MM = S$. However, in Section 2 we demonstrate that a stochastic model is more appropriate. I this case, $PM = O^{(PM)} + N^{(PM)} + S$ and $MM = O^{(MM)} + N^{(MM)}$ where $O^{(PM)}$ and $O^{(MM)}$ have the same expectation but are not perfectly correlated. Similarly, $N^{(PM)}$ and $N^{(MM)}$ have the same expectation but are not perfectly correlated. In this case the

difference $PM - MM$ is unbiased, $\mathrm{E}(PM - MM) = S$, but may have a large variance var (PM - MM).

In Section 2 we demonstrate that the $O + N$ component of the PM and MM are not perfectly correlated, thus var$(PM) <$ var (PM - MM). In part this explains why PM-only measures, such as RMA, are more precise than measures based on $PM - MM$, such as MAS 5.0. Irizarry et al. (2003b) empirically show that for low intensities PM the variance of the difference $PM - MM$ can be considerably larger than that of PM. Furthermore, in general, $MM > PM$ for roughly 40% of all probes. This is problematic because we know S is strictly positive. These facts have led some researchers to consider $PM - only$ measures. However, because O and N are strictly positive, not correcting for optical noise and NSB can lead to biased results: $\mathrm{E}(PM) > S$. To see the negative effect this can have in a practical application of, say, estimating expression fold-change in two samples being compared, consider a simple example: Say that the true expression for a particular gene of interest in two samples being compared are μ_1 and μ_2 picoMolar. Ideally we should observe a fold change of μ_1/μ_2. In practice, we observe intensities $PM_1 = O_1^{(PM)} + N_1^{(PM)} + k\mu_1$ and $PM_2 = O_2^{(PM)} + N_2^{(PM)} + k\mu_2$ and an observed fold change $(O_1^{(PM)} + N_1^{(PM)} + k\mu_1)/(O_2^{(PM)} + N_2^{(PM)} + k\mu_2)$. Thus, as the $k\mu_1$ and $k\mu_2$ become smaller, as compared to the the strictly positive mean of the background components O and N, the estimated fold change converges to 1. This results in attenuated fold change estimates. RMA performs a global background adjustment that improves accuracy over non- background adjusted methods. However, as we will discuss later, different probes have different propensities to NSB which implies RMA does not fully account for NSB. In this paper we develop a model that predicts the behavior of optical noise, NSB, and specific binding very well. We use hybridization theory from molecular biology together along with data from carefully designed experiments to motivate the model. We also propose a model for the distribution of the specific signal S intensities within an array. This model can be used to improve existing expression measures and provides theoretical explanations for various facts observed in practice, for example: 1) $MM > PM$ for roughly 40% of all probes, 2) $PM - MM$ has much larger variance than PM when S is small, and 3) $PM - MM$ is more accurate then PM when S is small.

2 Empirically Motivated Stochastic Models

In the Section we use publicly available data and data from our own experiments to motivate some of the components of our stochastic models The first of these data sets is the Affymetrix spike-in experiments. This experiment is described in detail, for example, by Irizarry et al. (2003b) and Cope et al. (2003). For this experiment, human cRNA fragments matching 16 probe-sets on one of the Affymetrix human chips were added to a hybridization mixture at concentrations ranging from 0 to 1024 picoMolar in a design similar to a Latin square. Apart from the spiked-in probe-sets, the same RNA mixture was hybridized to 59 arrays. Because we know the spike-in concentrations, it is possible to identify statistical features of the data for which the expected outcome is known in advance. The second data come from what we call the *empty chip* experiment. For this experiment, sample RNA control from human embryonic kidney derived cells was not labeled, but hybridized following the Affymetrix protocol. Because the RNA was not labeled, the observed intensities for this hybridization will represent optical noise in the presence of biological sample. Finally, the third data come from what we call the *NSB* experiment. For this experiment, yeast control RNA was hybridized to an array probing for human genes. This hybridization will represent the full component of the noise, NSB and optical noise. These two experiments are described in more detail in Wu et al. (2003).

2.1 Optical noise

A figure of a kernel density estimates (not-shown) of the empty chip probe level data (both PM and MM). This density looks very much like a normal distribution with mean of roughly 30 and standard deviation

(SD) of roughly 2. This plots motivates the first component of our model, the optical noise component, which we will model as normally distributed.

By using a log-scale transformation before analyzing microarray data, a great number of investigators have, implicitly or explicitly, propose a multiplicative measurement error model (Dudoit et al., 2002; Newton et al., 2001; Kerr et al., 2000; Wolfinger et al., 2001) for microarray data. A slightly more complicated additive-background-multiplicative-measurement-error model has been proposed by, for example, Durbin et al. (2002), Huber et al. (2002), Cui, Kerr, and Churchill (2003), and Irizarry et al. (2003a). In Figure 1a we see observed *PM* log (base 2) intensities from the spike-in data plotted against their nominal log (base 2) concentration. The line shows the median value for each concentration. Notice that this line looks very much like the shape of the function $f(x) = \log_2(x+k)$ with k about 60. This confirms that optical noise is additive as opposed to multiplicative.

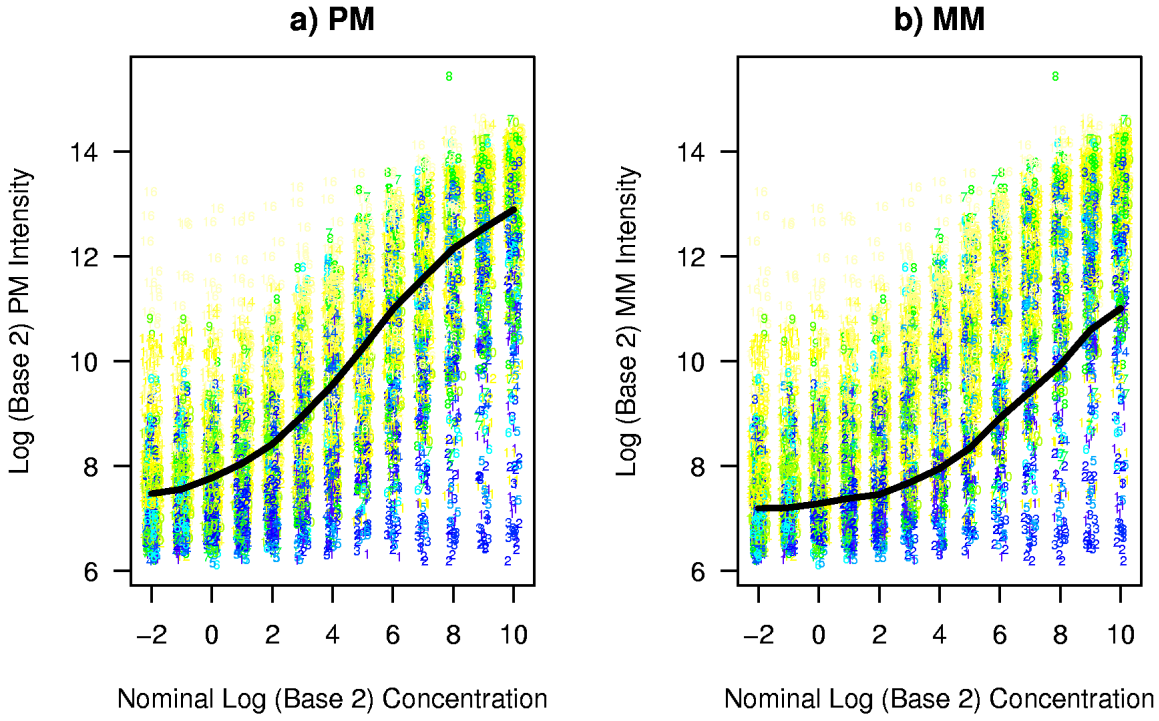

Figure 1: Signal detection by PM and MM probes in the Latin-square spike-in experiment. a). $\log_2(PM)$ intensities of spike-in genes plotted against concentration. Number indicate the order of each probe within probesets, each number is associated with a color for probesets. The line shows median $\log_2(PM)$ for each concentration. b). Same as a) but for MM intensities.

Figure 2a shows SD of probe intensities, computed across 28 replicate arrays, plotted against the respective average intensity. Figure 2b shows the same plot for log intensities. The mean-variance dependence that is removed by applying a log transformation is a strong argument for a multiplicative error model. We therefore propose using an additive-background-multiplicative-measurement-error

In Section 4 we describe a simple procedure for adjusting for optical background noise, similar to the one proposed by Irizarry et al. (2003a). In Figure 3a we see the median intensities for each nominal concentration, as in Figure 1, for the *PM* and the *PM* adjusted for background (along with other adjustments described later). We expect the curves in Figure 3a to be lines with slope 1, since every time the nominal

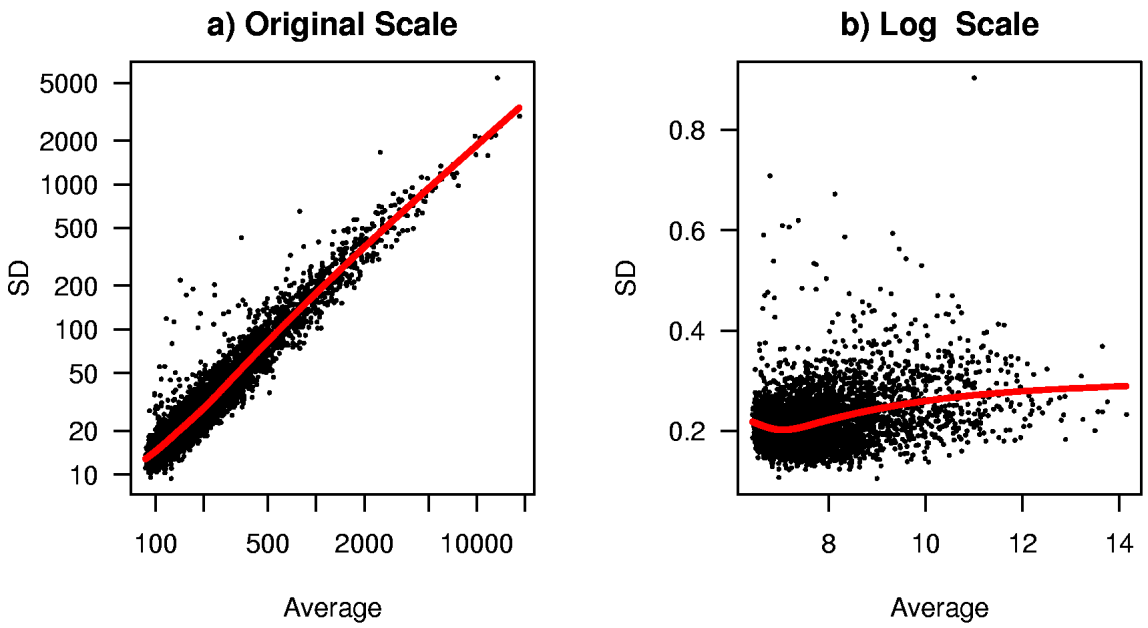

Figure 2: Standard deviation of probe intensity in original scale (a) and log scale (b) plotted a

concentration doubles observed concentration should double. We fit a line to the curve in this Figure and the slope for the *PM* intensities is 0.51. For the adjusted *PM* we have a slope of 0.59. The background adjustment slightly improves accuracy.

2.2 Non-specific binding

Molecular hybridization theory predicts that short oligonucleotides will hybridized to non-complementary transcripts. Using data from the NSB experiment we can see this. Figure 4a demonstrate a log-scale scatter plot of optical noise adjusted *PM*s versus optical noise adjusted *MM*s. This plot demonstrates intensities due to NSB are larger (by orders of magnitude) than those obtained just from optical noise.

Because in this data there is no specific signal, if in fact the *MM* are an exact measure of the NSB captured by the *PM* then the predictive power of the *MM* should be 1 and this plot should have no scatter. However, as expected, we do see scatter. The relative predictive power or R^2 for this scatter plot is 0.70. Although not perfect, the large R^2 suggest that there is information on NSB to be extracted from the *MM*. Notice also that Figure 4a seems to suggest that after adjustment for optical noise the NSB component of the PM, MM pairs follow a bivariate normal distribution.

To see that NSB is an additive effect more than it is a multiplicative effect, we adjusted the *PM* by subtracting and by dividing the *MM*. The resulting median intensity of $PM - MM$ is shown in Figure 3a. The estimated slope is 0.79 which is a good improvement over the non-adjusted *PM*. The PM/MM adjustment is very inaccurate (not shown in Figure 3a). The slope is only about 0.14. This suggest that NSB is an additive effect more than it is a multiplicative effect.

In Figure 3b we show a smooth curve demonstrating the over-all log-scale SD, across 28 replicate arrays, as a function of average log intensity, for the different adjustments. Notice that the $PM - MM$ adjustment is very noisy, especially at the low end. Notice that the loss of precision from subtracting *MM* is quite

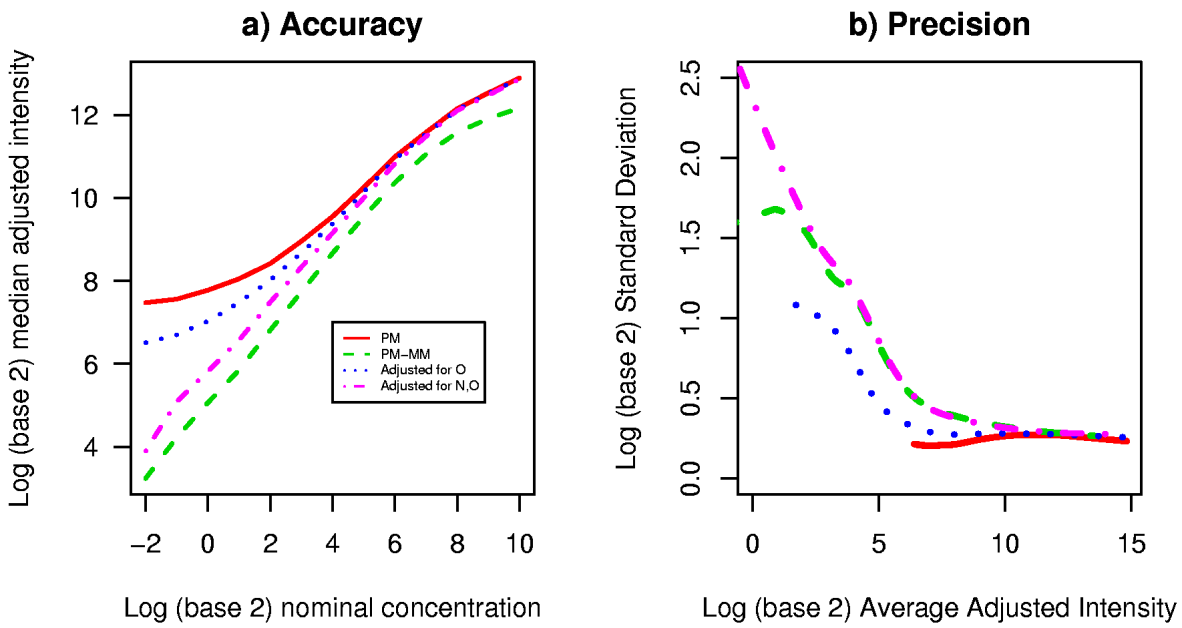

Figure 3: Accuracy and precision of various background adjustments. a).Log median adjusted intensity plotted against log concentration. b). standard deviation of log adjusted intensity plotted against log concentration.

significant. The median SD grows from 0.20 for non-adjusted *PM*, to 0.36 for optical noise adjusted *PM*, to 0.90 for $PM - MM$. The loss in accuracy of ignoring the *MM* is not as drastic.

2.3 Specific Signal

Li and Wong (2001) demonstrate that, even after subtracting *MM*, there is a strong probe effect. Notice in Figure 1 that the range of probe intensities measuring the same nominal amount of RNA cover various orders of magnitude. In Figure 1 we use color and numbers to denote the same probes. The probes that have, on average bigger effects, are shown in yellowish colors, those with lower values in blue colors. The fact that the blue are always at the bottom, the yellow at the top demonstrate the strong and consistent probe effects. The fact that Figure 1 is a log-scale plot, suggests that this probe effect is multiplicative as well as the measurement error.

3 Physical Models

Zhang et al. (2003) propose a stacking energy, positional-dependent-nearest-neighbor (PDNN),model for RNA/DNA duplex formed on microarrays. Their energy model takes into account the sequence of nearest-neighbors (adjacent two bases) and the position of these nucleotide pairs. It has been suggest that the effect of nearest-neighbor nucleotide pairs is the most important factor in determining RNA/DNA duplex stability. Zhang et al. add a positional weight factor to reflect the different contributions from different part of the probe.

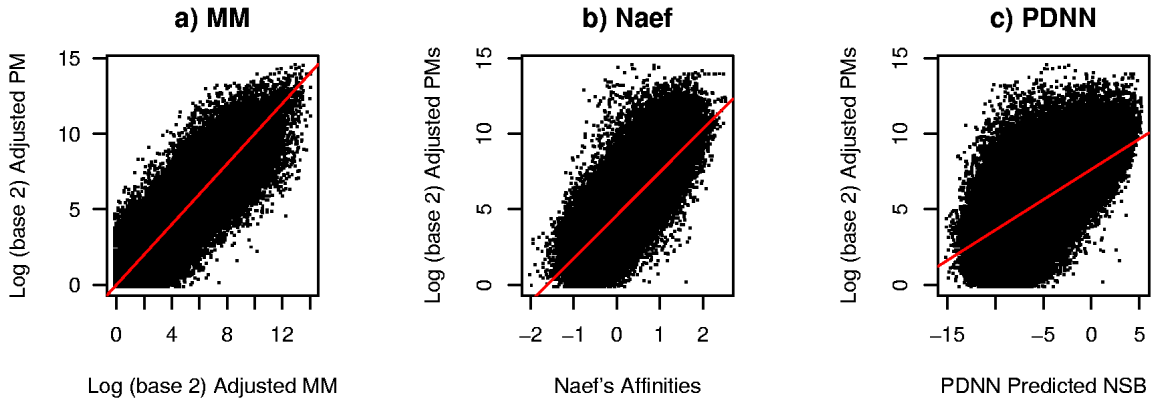

Figure 4: Intensity prediction ability comparison. a). Optical noise adjusted $\log_2(PM)$ intensity plotted against optical noise adjusted $\log_2(MM)$ intensity. b). Optical noise adjusted $\log_2(PM)$ against Naef's predicted affinity. c).Optical noise adjusted $\log_2(PM)$ against PDNN predicted non-specific binding.

The energies for gene-specific binding and NSB of the i-th probe in j-th probe-set is thus calculated as,

$$\begin{aligned} E_{ij} &= \sum \omega_k \varepsilon(b_k, b_{k+1}) \\ E^*_{ij} &= \sum \omega^*_k \varepsilon^*(b_k, b_{k+1}), k = 1,2,...,24, \end{aligned}$$

respectively, where ω_k, ω^*_k are weights, and $\varepsilon(b_k, b_{k+1}), \varepsilon^*(b_k, b_{k+1})$ are nearest-neighbor stacking energies.

They describe the PM intensity of the i-th probe in j-th probe-set as

$$N_j/\{1+\exp(E_{ij})\} + N^*/\{1+\exp(E^*_{ij})\} + O,$$

where N_j is the number of expressed mRNA molecules of gene j, and $N_j/\{1+\exp(E_{ij})\}$ is the contribution from gene specific binding. N^* is population of RNA molecules contributing to NSB for the entire array, and $N^*/\{1+\exp(E^*_{ij})\}$ is the contribution to intensity of i-th probe in j-th probe-set.

Naef and Magnasco (2003) propose a simpler model for the NSB by considering only sequence composition of the probes. Affinity of a probe is described as the sum of position-dependent base affinities:

$$\sum_{l_j} S_{l_j} A_{l_j},$$

where $l = A,C,G,T$ is the letter index and $j = 1,\ldots,25$ indicates the position along the probe. L is a Boolean variable equal to 1 if the j-th base matches the respective position. Thus the A's are per-site, per-base affinities. Their model is fitted to many arrays at once to obtain an affinity value for each sequence.

In Figure 4b and 4c we plot the optical noise adjusted $\log_2(PM)$s from the NSB data set against Naef and Magnasco's affinities and Zhang's PDNN $\log_2(N^*) - \log_2(1+\exp(E^*_{ij}))$. Notice that Naef and Magnasco's affinities predict the NSB almost as well as the MM. The R^2 is 0.64. Zhang's PDNN also does relatively well with an R^2 of 0.38. However, notice that the slope of the PDNN model scatter plot is not 1.

Figure 4 demonstrates that these physical models can not predict NSB perfectly. However, they motivate a simple stochastic model. In Section 4 we propose a model that describes the NSB contribution as a log-normal distributed with log-scale mean directly proportional to Naef and Magnasco's affinities. The two parameters describing the linear relationship is estimated from the data and not predicted using physical models. The model works similarly with using Zhang's $-\log_2(1+\exp(E^*_{ij}))$ as the affinity measure.

3.1 Specific Signal

Let P be the total number of probes on a GeneChip microarray, $p(i)$ the number of probes belonging to an intensity interval $\Delta(i)$ centered on i, and $q(i)$ the quantiles of the ranked measured intensities i. Our goal is the derivation of an analytical form for the intensity density

$$\rho(i) = \frac{1}{\Delta(i)} \frac{p(i)}{P}. \tag{1}$$

The Langmuir adsorption model for a compound in solution in contact with an adsorptive surface stipulates that, at thermodynamic equilibrium,

$$\Theta = \frac{\beta}{\beta+1} = \frac{KC}{KC+1}, \tag{2}$$

where $\beta = KC$ with K the ratio of adsorption and desorption rate constant, C the concentration of the compound in solution, and Θ is the fraction of binding sites which have been adsorbed (P.W., 1994). Obviously, Langmuir adsorption kinetics are characterized by saturation effects as the occupancy fraction Θ increases from zero to one. How one is to attribute a value to Θ depends on the experimental setup. For oligonucleotide microarray, one assumes that Θ is proportional to the fluorescence intensity i, that is,

$$i \propto \Theta = \frac{\beta}{\beta+1}. \tag{3}$$

Suppose for a moment that P is so large that the limit

$$\rho(i) = \lim_{\Delta(i)\to 0} \frac{1}{\Delta(i)} \times \lim_{P\to\infty} \frac{p(i)}{P} \equiv \frac{dq(i)}{di} \tag{4}$$

is meaningful, that is, we consider the quantile $q(i)$ as a continuous function of the intensity i. Since $\beta = KC$ is the true experimental controllable variable, it is more meaningful to consider

$$\rho(i) = \frac{dq}{d\beta}\frac{d\beta}{di} = \rho(\beta)\frac{d\beta}{di} \propto (\beta+1)^2\rho(\beta) \tag{5}$$

where the last proportionality stems from the derivative of equation (3) with respect to β. Since it is observed that $\rho(i) \sim 1/P$ for GeneChip at very high intensities which we assume correspond to points of probe saturation (Hekstra et al., 2003), we expect $\rho(\beta)$ to decrease at least as fast as $\rho(\beta) \sim 1/\beta^2$ as $\beta \gg 1$. Consequently, we postulate

$$\rho(i) = (\beta+1)^2\rho(\beta) \sim \left[\frac{\beta+1}{\beta}\right]^2 \sim \frac{1}{i^2} \tag{6}$$

for mid- to high-intensity data points. As we shall demonstrate below, the intensity density $\rho(i)$ can in fact be described to a high degree of accuracy over the entire range of observed intensities by the equality

$$\rho(i) = \frac{1}{(i+1)^2}. \tag{7}$$

This is a power-law which has been observed empirically in SAGE data (Blades et al., 2003).

4 Unified Physical/Stochastic Model

The described physical models perform relatively well at predicting NSB and, as will become apparent, the distribution of the specific signal. However, the predictions are not perfect and are complimented well with stochastic versions. The system producing intensities is very complicated and we argue that one can use physical models to approximate the process relatively well, but the lack-of-fit is best described with a stochastic model.

Our model for the *PM* intensity contains NSB and specific signal components that on the probe sequence composition as described by the physical models. The model can be written as

$$PM_j = B_j + f(A_j)N_j + g(A_j)S_j$$

where A_j is the probe affinity for the j-th probe, $B \sim \text{Normal}(b_o, \sigma_o^2), \log(N_j) \sim \text{Normal}(\mu_N, \tau^2), \log(S_j) \sim \text{Exponential}(1)$, and we approximate $\log(f(A_j))$ and $\log(g(A_j))$ with linear functions of A_j. Notice that we use an exponential with rate 1 because when exponentiated this distribution approximates very well the distribution described in Section 3.1 and it has a simple parametric form.

Notice that this model only has seven parameters and we have over 200,000 probe intensities to fit them. One can use maximum likelihood estimation to do this. However, writing down the likelihood for this model is complicated as it involved a convolution of 3 densities. We have developed some ad-hoc procedures to estimate the parameters that yield very good fits.

To estimate the parameters we assume that the *MM*s do not measure specific signal and this can be modeled with $MM_j = B_j + f(A_j)N_j$. We then estimate b_o by the 0.02 quantile of the MM intensities, following Affymetrix convention. σ_o is estimate assuming *MM* intensities less than $\widehat{b_o}$ as the left half of a normal distribution. To obtain *PM* and *MM* intensities adjusted for optical noise that are not negative (we know *N* + *S* is positive) we use the posterior mean of $\text{Uniform}(0, 2^{16})\text{given}(\text{Uniform}(0, 2^{16}) + \text{Normal}(\widehat{b_o}, \widehat{\sigma_o}))$ (2^{16} is the scanner upper limit).

To obtain an estimate of the linear function $f(A)$, the log *MM* intensities (after optical background adjustment) are regressed on the *MM* probe affinities described by Naef and Magnasco. For a series of given affinities A_k, we find the $\widehat{g(A)}$ that makes the .8 quantile of $f(A_k)\exp(\text{Normal}(\mu_N, \tau)) + \widehat{g(A)}\text{Exponential}(1)$ the same as the .8 quantile of optical background adjusted PM. We then regress the $\widehat{g(A_k)}$ values on the A_k's to obtain linear approximation of $g(A)$.

5 Results

We fitted the model as described in the previous Section. The model fits extremely well. Figure 5a shows kernel density estimates of the *PM* intensities for one of the spike-in arrays along with the predicted distribution from the model. Notice that there are only 7 parameters and 200,000 data points so over-fitting is not a concern. Furthermore, the model is based on molecular hybridization theory. Figure 5b shows a quantile-quantile plot that confirms the good fit.

Figure 6 we present the results shown in Figure 3 but instead of real data we use data simulated from our 7 parameter model. Notice the similarity between the real and simulated results. This suggests that our proposed model can be used for simulations related to statistical procedures based on Affymetrix data. For example, one could use it decide among different test statistics (Wilcoxon, t-test, SAM, etc...)

Finally, we point out that under this model, as fitted to this array, the probability of a $MM > PM$ is 0.40 which is exactly what we see in practice. Thus having many $MM > PM$ is not necessarily a bad thing. It is just a consequence of the noisy character of the system.

6 Discussion

We have presented a stochastic model motivated by molecular hybridization theory that fits Affymetrix GeneChip probe level data very well. Apart from giving a theoretical explanation for various facts observed in practice, this model can also be used to improve expression measures. For example, once we have fitted the model, we could correct for optical noise and NSB by computing the expectation of S given that we have observed a PM and the all the parameters have been estimated. An approach such as this has been used by Wu et al. (2003) with very encouraging results. Wu et al. describe an expression measure algorithm similar to RMA but using a model such as the one described here to adjust for background. Their expression measure is about as precise as RMA but much more accurate. In fact, it is more accurate than MAS 5.0.

Notice that this model can also be used for normalization and summarization. The fact that we have a prior distribution for the specific signal component suggest that one could use the log-exponential as the reference distribution used in quantile normalization. Furthermore, by incorporating information about probe-sets in the mode (i.e. which probes represent which genes) one could directly obtain MLE estimates of expression measures from the model. One aspect that is not described by our model is the existence of outliers probes. This is subject of future work.

Finally, we would like to point out that the model described can be fitted relatively well using only PM probes. The correction for non-specific binding can be done with Naef and Magnasco's affinities. To see this we include in Figure 3a the results of optical noise adjusted PMs adjusted for NSB using the prediction of our model (based on Naef and Magnasco's affinities). The slope is 0.76 which is comparable to that obtained with $PM - MM$. Although there is complimentary information in the MM and in the affinities, PM-only measures are attractive for various reasons, for example: 1) We can have twice as many probes on the chips and 2) the MM seems to detect signal as demonstrated by Figure 1b.

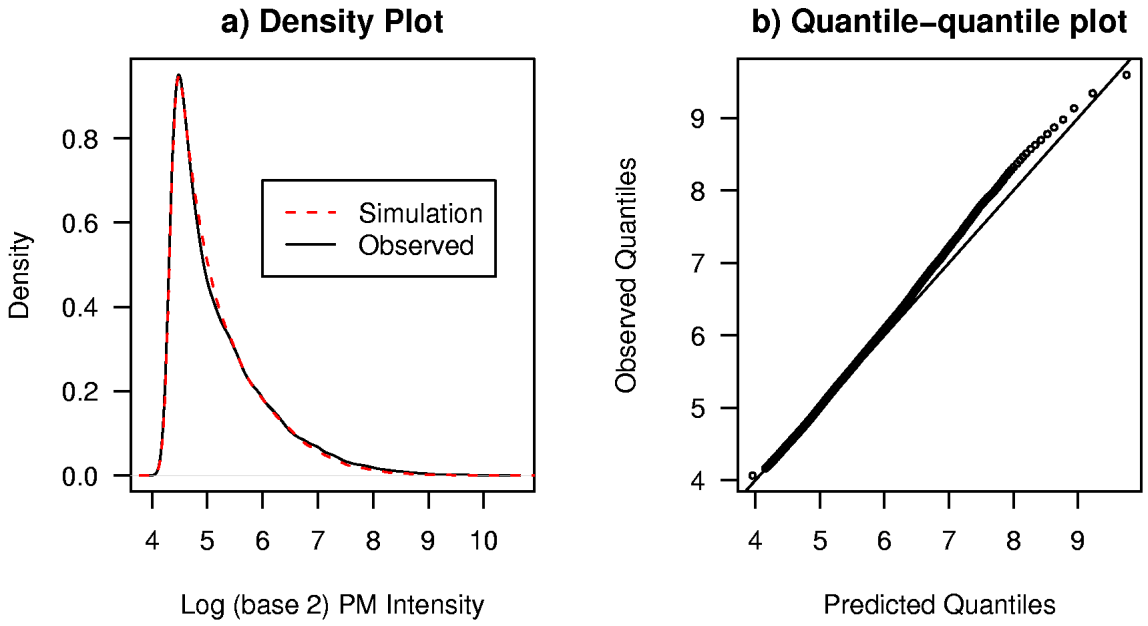

Figure 5: Simulation result. a). Kernel density estimates of simulated PM intensity and PM intensity from Latin-square spike-in experiment. b). Quantile quantile plot of PM intensity from Latin-square spike-in experiment against simulated PM intensity.

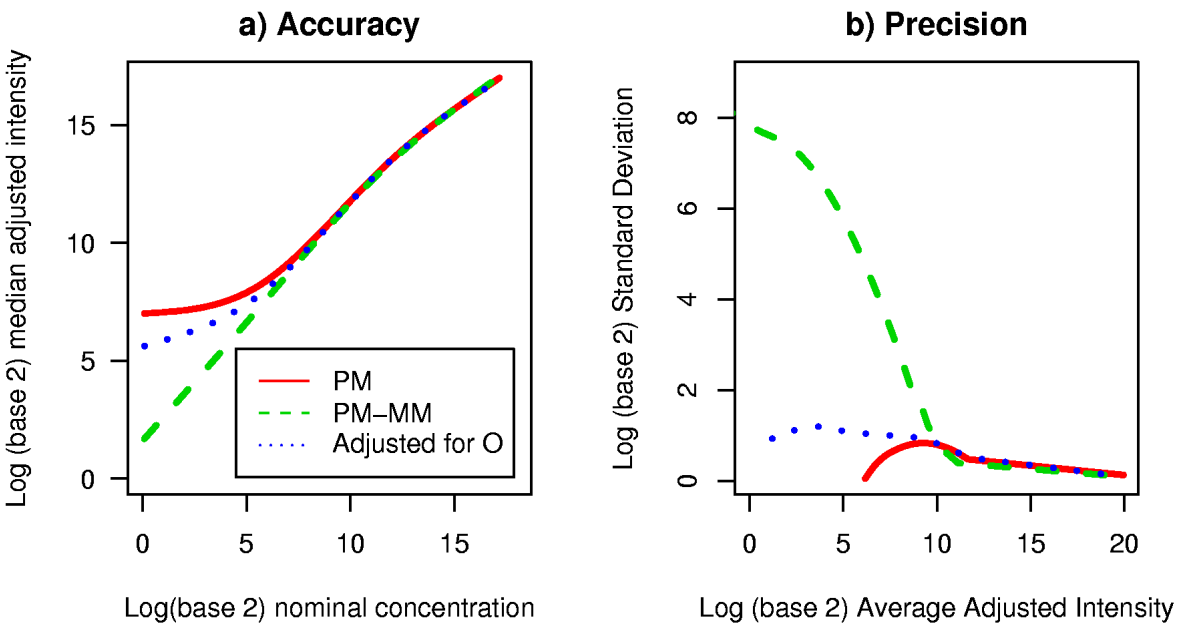

Figure 6: As Figure 3 but with simulated data.

References

Blades, N., Jones, J., Kern, S., and Parmigiani, G. (2003). Denoising of data from serial analysis of gene expression. *Bioinformatics* To appear.

Bolstad, B., A., I. R., Astrand, M., , and Speed, T. (2003). A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. *Bioinformatics* **19**(2).

Chu, T., B., W., and Wolfinger, R. (2002). A systematic statistical linear modeling approach to oligonucleotide array experiments. *Math Biosci* **176**, 35–51.

Cope, L., Irizarry, R., Jaffee, H., Wu, Z., and Speed, T. . (2003). A benchmark for affymetrix genechip expression measures. *Bioinformatics* http://www.biostat.jhsph.edu/ ririzarr/papers/index.html. In press.

Cui, X., Kerr, M. K., and Churchill, G. A. (2003). Data transformations for cDNA microarray data. http://www.jax.org/staff/churchill/labsite/research/expression/Cui-Transform.pdf.

Dudoit, S., Yang, Y. H., Callow, M. J., and Speed, T. P. (2002). Statistical methods for identifying differentially expressed genes in replicated cdna microarray experiments. *Statistica Sinica* **12**(1), 111–139.

Durbin, B. P., Hardin, J. S., Hawkins, D. M., and Rocke, D. M. (2002). A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics* **18**(Suppl. 1), S105–S110.

Hekstra, D., Taussig, A. R., Magnasco, M., and Naef, F. (2003). Absolute mrna concentrations from sequence-specific calibration of oligonucleotide array. *Nucleic Acids Research* **31**(7), 1962–1968.

Holder, D., Raubertas, R. F., Pikounis, B. V., Svetnik, V., and Soper, K. (2001). Statistical analysis of high density oligonucleotide arrays: a SAFER approach. In: *Proceedings of the ASA Annual Meeting, Atlanta, GA 2001*.

Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **1**, 1:9.

Irizarry, R., B. Hobbs, F. C., Beaxer-Barclay, Y., Antonellis, K., Scherf, U., and Speed, T. (2003a). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264.

Irizarry, R., Gautier, L., and Cope, L. (2003b). An r package for analyses of affymetrix oligonucleotide arrays. In: R. I. G. Parmigiani, E.S. Garrett and S. Zeger (eds.), *The Analysis of Gene Expression Data: Methods and Software*. Springer, Berlin.

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003c). Summaries of affymetrix genechip probe level data. *Nucleic Acids Research* **31**.

Kerr, M. K., Martin, M., and Churchill, G. A. (2000). Analysis of variance for gene expression microarray data. *Journal of Computational Biology* **7**, 819–837.

Li, C. and Wong, W. (2001). Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proceedings of the National Academy of Science U S A* **98**, 31–36.

Naef, F., Lim, D. A., Patil, N., and Magnasco, M. O. (2001). From features to expression: High density oligonucleotide array analysis revisited. *Tech Report* **1**, 1–9.

Naef, F. and Magnasco, M. O. (2003). Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays. *Physical Review E* **68**, 011906.

Newton, M., Kendziorski, C., Richmond, C., and Blattner, F. (2001). On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *Journal of Computational Biology* **8**, 37:52.

P.W., A. (1994). *Physical Chemistry*. Oxford, UK: Oxford University Press, 5th edition.

Wolfinger, R., Gibson, G., Wolfinger, E., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C., and Paules, R. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology* **8**(6), 625–637.

Workman, C., Jensen, L. J., Jarmer, H., Berka, R., Gautier, L., Nielsen, H. B., Saxild, H., Nielsen, C., Brunak, S., and Knudsen, S. (2002). A new non-linear normalization method for reducing variability in dna microarray experiments. *Genome Biology* **3**.

Wu, Z., Irizarry, R. A., Gentleman, R., Murillo, F. M., , and Spencer, F. (2003). *A Model Based Background Adjustment for Oligonucleotide Expression Arrays*. Technical report, Johns Hopkins University, Dept. of Biostatistics Working Papers. http://www.bepress.com/jhubiostat/paper1.

Zhang, L., Miles, M. F., and Aldape, K. D. (2003). A model of molecular interactions on short oligonucleotide microarrays: implications for probe design and data analysis. *Nature Biotechnology* To Appear.

Zhang, L., Wang, L., Ravindranathan, A., and Miles, M. (2002). A new algorithm for analysis of oligonucleotide arrays: application to expression profiling in mouse brain regions. *J Mol Bio* **317**, 227–235.