Measurement Errors in Gene Expression Data from Spotted DNA Microarrays

Peter D. Wentzell, Tobias Karakach, Robert Flight, and Joe Boutilier
Trace Analysis Research Centre, Department of Chemistry
Dalhousie University, Halifax, Nova Scotia, Canada
Peter.Wentzell@Dal.Ca
Uncertain knowledge
+ Knowledge of the amount of uncertainty in it
= Usable knowledge

Beautiful Noise?
The Research Group
A Workflow for Timecourse Analysis

- Experimental Design
  - Problem formulation
  - Number of samples
  - Time intervals
  - Randomization

- Experiment

- Microarray Slides

- Image Measurement

- Scanning
  - Pixel Data (TIFF files)
The Workflow (cont.)

Pixel Data (TIFF Files)

Image Processing

Gridding, Flagging
Ratio calculation
Uncertainty analysis

Ratio Data

Data Preprocessing

Censoring
Correlation
Normalization
Transformation

Clean Data, Uncertainties

time
time

R/G
σR/G
The Workflow (cont.)

Clean Data, Uncertainties

Information Extraction

Profiles

Knowledge Synthesis

Curve resolution

Ontological associations
Data mining
Expert assessment

Gene Function
Gene Networks
Transcription Factors
The Meaning of Life
Outline

- Introduction to DNA Microarrays
  - Basic Principles of Microarray Analysis
  - Microarray Spot Morphology
  - Ratio Calculation Methods

- Errors Microarray Measurements
  - Normalization in Comparator Experiments
  - A Model for Errors in Microarray Measurements
  - Bootstrap Method for Error Estimation
  - Validation of Bootstrap Method

- Errors from Gridding and Flagging

- Summary and Conclusions
DNA Microarray Analysis - Signal Transduction

Expressed mRNA

Reverse Transcriptase

Fluorescent Label

Hybridization with Microarray

(Immobilized Single Stranded DNA)
Some Common Microarray Formats

- **Lithographic, Single Channel, Short Oligomers**
  - Affymetrix strategy ("Affy arrays")
  - High density (ca. 40,000 genes/chip)
  - Typically 25-mer oligonucleotide sequences
  - Genome required

- **Spotted, Two Channel, Long Oligomers**
  - Custom fabrication common
  - Lower density (ca. 10,000 genes/chip)
  - Typically 70-mer oligonucleotide sequences
  - Genome required

- **Spotted, Two Channel, cDNA**
  - Custom fabrication from DNA libraries
  - Lower density (ca. 10,000 genes/chip)
  - Genome not required
A Typical DNA Microarray Experiment

Test
mRNA → Extract → Label → Hybridize → Scan

Reference
mRNA → Extract → Label → Hybridize → Analyze

cDNA
Gene Expression by DNA Microarray Analysis

DNA Microarray

In gene expression analysis the objective is to determine which genes change their expression levels under different conditions, in which direction, and by how much.

Comparator Experiments – two or more discrete conditions

Time Course Experiments – a continuum of conditions
Spotted Microarray Fabrication
Ratio Measurements in Microarray Analysis
What Microarray Spots Really Look Like

Red Channel

Green Channel
Spot Morphology (cont.)
Spot Morphology (cont.)

Green

Red

Green
Spot Morphology (cont.)
Anatomy of a Microarray Spot – Ideal Case

• The cartoon above represents the cross-sectional profiles of four “ideal” microarray spots.

• Green = Reference, Red = Test

• Note that one of the genes in the test sample is up-regulated.
Spot dependent intensity variation arises because:

- Not all genes are expressed equally
- Not all ssDNA’s are labelled with the same sensitivity
- Not all ssDNA’s hybridize with the same efficiency
Channel Dependent Sensitivity

- Detector sensitivity to Cy3 and Cy5 dyes is not the same
- Different amounts of mRNA may have been extracted
- May be differences in cDNA conversion and labelling
Variations in Spot Morphology

- The intensity can vary across the spot due to changes in the accessibility of the oligomers at a particular location.
- The profiles on each channel are equally affected by this and remain correlated.
Variations in Background Intensity

- Background fluorescence signals can arise from reagents, substrate, unwashed or non-spectically bound target DNA, impurities, etc.

- Background varies across the slide and is usually estimated from local measurements around the spot and subtracted.
There is strong evidence to suggest that the background intensity under the spot can be substantially different from that around it.
• Instrument noise is expected to be shot-noise limited.

• Indications are that this is not a major source of variability.
Other effects that can be observed include:

- Outlier pixels (e.g. dust spikes)
- Channel shifts
- Saturated pixels
- Cross-hybridization
Common Ratio Calculation Methods

**Ratio of Means:**  
- Ratio of mean signal intensity on each channel  
- Widely used; good signal averaging, but sensitive to outliers

**Ratio of Medians:**  
- Ratio of median signal intensity on each channel  
- Perhaps most widely used method; more robust except for poorly represented spots (e.g. “donuts”)

**Mean of Ratios:**  
- Mean of individual pixel ratios  
- Rarely used; gives statistics, but unreliable

**Median of Ratios:**  
- Median of individual pixel ratios  
- Rarely used; more robust but still unreliable

**Regression Method:**  
- Regress pixel intensities on one channel against the other - slope is ratio measurement  
- Not widely used, but has many advantages
Examples of Regression Ratio Method

Note that the differences between ordinary least squares and orthogonal least squares become more apparent at higher slopes.
Errors in Microarray Measurements

“Systematic” Errors:
• Normalization of channel intensity variations
• Background correction

“Random” Errors:
• Biological variation (biological replicates)
• Slide-to-slide variation (technical replicates)
• Spot-to-spot variation (spot replicates)
• Ratio measurement uncertainty
Case Study 1 – A Comparator Experiment

• In this study, a mutant strain of yeast was compared to wild-type yeast to detect differences in gene expression.

• Ratios were measured for approximately 6000 genes with side-by-side replication.

• Typically for this type of comparator experiment, most genes are expected not to undergo changes in expression.
Distribution of Ratio Measurements

- Note measurements follow a (somewhat) log-normal distribution.
- Note also – distribution is not centered on zero.
Normalization Strategies for Comparator Experiments

**Global Normalization:**
- In global normalization, the mean of the log-normal distribution is forced to zero.

**LOWESS (or LOESS) Normalization:**
- LOWESS normalization is based on the observation that the normalization required appears to depend on intensity. (Why ???)
- In LOWESS an “MA plot” is first constructed where:
  \[ M = \log_2 \left( \frac{\text{Red}}{\text{Green}} \right) \] (y-axis)
  \[ A = 0.5 \times \left[ \log_2(\text{Red}) + \log_2(\text{Green}) \right] \] (x-axis)
  (a measure of “average” intensity)
- A locally weighted scatterplot smooth (LOWESS) is then used to estimate a local normalization adjustment (offset).
Example of Global and LOWESS Normalization
A Model for Random Errors in Microarrays

The total variance in the ratio can be represented as:

\[
\sigma_{\text{total}}^2 = \sigma_{\text{bzl}}^2 + \sigma_{\text{slide}}^2 + \sigma_{\text{spot}}^2 + \sigma_{\text{meas}}^2
\]

- **biological replicates**
- **between slide (technical replicates)**
- **between spots**
- **ratio measurement**

Want to estimate this term

\[
\sigma_{\text{total}}^2 = \sigma_{\text{expt}}^2 + \sigma_{\text{meas}}^2
\]

Can be estimated from replicates

Depends on spot morphology, signal intensity, background
Some Hypotheses

\( \sigma_{\text{expt}} \) : - There is much evidence to suggest that this contribution approximates a log-normal distribution.
- There is also evidence to support a proportional (multiplicative) error model.

\( \sigma_{\text{meas}} \) : - This contribution is not commonly considered.
- Unique for each spot; can be small or large.
- Easiest estimate comes from side-by-side replicates, but has some problems:
  o Expensive for multiple replicates
  o No guarantee replicates have same morphology
- Postulate that this can be estimated by bootstrap approach.
Normally, the ratio measurements do not carry implicit or explicit information about measurement uncertainty. As exemplified below, measurements with the same ratio can have radically different characteristics.

1-7-1106, Ratio=1.00

1-7-6153, Ratio=1.01
The bootstrap method is applied to the estimation of ratios by the regression method.

The Bootstrap Method

Pick N points at random with replacement

Sampled Population

\[ \hat{\theta} \]

Bootstrap Samples

\[ \hat{\theta}_1^* \quad \hat{\theta}_2^* \quad \hat{\theta}_3^* \quad \hat{\theta}_B^* \]
Bootstrap Results – Example 1

Orthogonal Regression Plot, Spot 980

Slope = 11.36 ± 0.40
Bootstrap Results – Example 2

Orthogonal Regression Plot, Spot 4420

Slope = 0.349 ± 0.027
Are the Bootstrap Estimates Reliable?

Validation of the bootstrap approach to estimating uncertainty in ratio measurement was carried out in several ways.

(1) **Simulation studies**
   - Allowed precise control over spot morphology and intensity noise to validate basic principles of the method.

(2) **Comparison to multiple spot replications on same slide**
   - The most direct experimental comparison, although not ideal.

(3) **Qualitative observations**
   - Does the agreement of replicates improve when spots with high uncertainty are removed?

(4) **Modeling of residual distributions in replicate experiments**
   - Integration of bootstrap error estimates with overall error model for replicate experiments.
Simulated Profiles

0% Noise

Profile 1
A

Profile 2
B

5% Noise

C
D

40% Noise

E
F
Simulation Results

Profile 1

A

Profile 2

B

5% Noise

C

40% Noise

D

Simulation Results

Profile 1

A

Profile 2

B

5% Noise

C

40% Noise

D

“true” s.d.

bootstrap estimate

bias estimate

Standard Deviation

Run Number

Standard Deviation

Run Number
Case Study 2 – Halibut Microarray Data

- Study involved 15 microarray slides following the development of halibut.
- Each slide had a total of 9,625 unique features, each with four side-by-side replicates.
- The availability of replicates allowed the experimental estimation of $\sigma_{\text{meas}}$, assuming similar morphology.
- Experimentally measured uncertainties could be compared to bootstrap estimated values for each feature on each slide.
- Following filtering for extreme criteria (flags, negative ratios, etc.), 42,318 spots remained (x 4 replicates each).
- Experimentally measured standard deviations are expected to have an RSD of about 41% based on a $\chi^2$ distribution.
Comparison of Bootstrap with Replicate Spots

(Only 2000 points shown)
Comparison of Bootstrap with Replicate Spots

Density Plot

\[ \log_{10}(\text{Measured RSD}) \]

\[ \log_{10}(\text{Bootstrap Estimated RSD}) \]
Case Study 3 – Yeast Time Course Data

- Study involved 19 microarray slides following changes in *Saccharomyces cerevisiae* for one hour after exit from stationary phase.
- Microarray slides included four duplicate and one triplicate set of time points.
- There were ca. 6300 spots on each array, with no replicated spots.
- Since this was a time course rather than a comparator experiment, ratios were highly variable.
- Duplicate experiments can be compared by log-log plots of ratio measurements. This should give a straight line with a slope of unity. A proportional error structure should give a normal distribution of residuals.
Effect of Removing “Bad” Measurements

- \( \leq 0 \) removed
  - \( N = 6229 \)
  - Slope = 0.900

- \( \leq 0 \), Flagged
  - \( N = 4754 \)
  - Slope = 0.898

- \( \leq 0 \), rsd>30%
  - \( N = 5463 \)
  - Slope = 0.904

- \( \leq 0 \), rsd>30%, Flagged
  - \( N = 4612 \)
  - Slope = 0.921
Note: If \( y = \log x \) and \( \sigma_x = \alpha x \) (proportional error),

then \( \sigma_y = \sigma_x / 2.303x = \alpha / 2.303 \)
Illustration of Proportional Error Contribution

![Graph showing the relationship between ratios on Slide 1 and Slide 2. The x-axis represents the ratio on Slide 1, and the y-axis represents the ratio on Slide 2. The data points are dispersed along the diagonal line, indicating a proportional error contribution.]
A Proposed Error Model

\[ \sigma_R^2(i) = \sigma_{\text{exp}}^2(i) + \sigma_{\text{meas}}^2(i) \]

\[ = \alpha^2 R^2(i) + \sigma_{\text{boot}}^2(i) \]

• Seems reasonable, but Monte Carlo simulations did not produce good fits to experimental distributions of residuals.
A Revised Error Model

\[ \sigma_R^2(i) = \sigma_{\text{expt}}^2(i) + \sigma_{\text{meas}}^2(i) \]

\[ = \alpha^2 R^2(i) + \beta^2 \sigma_{\text{boot}}^2(i) \]

- Although it is not certain why this adjustment should be necessary, Monte Carlo simulations produced excellent fits to experimental distributions of residuals.
Model Fit to Residuals - Yeast

Number of Spots

Orthogonal Residual

- Yellow: Observed Residuals
- Red: Gaussian Fit
- Blue: Monte-Carlo Fit
## Model Parameters – Yeast Duplicates

<table>
<thead>
<tr>
<th>Slide Pair</th>
<th>% Proportional Error (100(\alpha))</th>
<th>Scale Factor ((\beta))</th>
<th>(\chi^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>18.9 ±0.5</td>
<td>4.24 ±0.11</td>
<td>23.7 ±0.6</td>
</tr>
<tr>
<td>1,3</td>
<td>21.7 ±0.8</td>
<td>4.33 ±0.14</td>
<td>30.5 ±1.7</td>
</tr>
<tr>
<td>2,3</td>
<td>9.5 ±0.5</td>
<td>1.93 ±0.07</td>
<td>19.4 ±0.8</td>
</tr>
<tr>
<td>4,5</td>
<td>37.2 ±1.6</td>
<td>3.61 ±0.40</td>
<td>44.0 ±2.5</td>
</tr>
<tr>
<td>7,8</td>
<td>0.14 ±0.11</td>
<td>4.20 ±0.04</td>
<td>57.2 ±1.6</td>
</tr>
<tr>
<td>10,11</td>
<td>5.1 ±0.6</td>
<td>3.40 ±0.06</td>
<td>44.6 ±2.8</td>
</tr>
<tr>
<td>14,15</td>
<td>7.6 ±0.6</td>
<td>2.24 ±0.12</td>
<td>21.8 ±1.1</td>
</tr>
</tbody>
</table>

\[\chi^2_{\text{crit}} = 42.6\]

\[(\alpha = 0.05)\]
Case Study 4 – Malaria Developmental Cycle

- Study involved 55 microarray slides following changes in *Plasmodium falciparum* during the intraerythrocytic developmental cycle. Data were obtained at one hour intervals for 48 hours.
- Microarray slides included seven duplicate and one triplicate set of time points.
- There were ca. 7300 spots on each array, with no replicated spots.
- These data were analyzed in a manner similar to the yeast data, with similar conclusions.
Model Fit to Residuals - Malaria

Graph showing the distribution of observed residuals compared to Gaussian and Monte-Carlo fits.
# Model Parameters – Malaria Duplicates

<table>
<thead>
<tr>
<th>Slide Pair</th>
<th>% Proportional Error (100(\alpha))</th>
<th>Scale Factor ((\beta))</th>
<th>(\chi^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>5.8 ±0.3</td>
<td>3.74 ±0.06</td>
<td>15.7 ±0.7</td>
</tr>
<tr>
<td>1,3</td>
<td>7.4 ±0.7</td>
<td>3.16 ±0.09</td>
<td>42.0 ±2.1</td>
</tr>
<tr>
<td>2,3</td>
<td>8.7 ±0.3</td>
<td>2.47 ±0.06</td>
<td>22.9 ±0.9</td>
</tr>
<tr>
<td>9,10</td>
<td>19.6 ±0.2</td>
<td>2.54 ±0.03</td>
<td>42.3 ±2.7</td>
</tr>
<tr>
<td>14,15</td>
<td>4.6 ±0.3</td>
<td>2.57 ±0.04</td>
<td>53.2 ±1.4</td>
</tr>
<tr>
<td>18,19</td>
<td>5.3 ±0.8</td>
<td>1.95 ±0.06</td>
<td>23.2 ±0.7</td>
</tr>
<tr>
<td>23,24</td>
<td>8.6 ±0.3</td>
<td>2.23 ±0.05</td>
<td>25.7 ±0.8</td>
</tr>
<tr>
<td>26,27</td>
<td>14.6 ±0.4</td>
<td>2.14 ±0.04</td>
<td>34.3 ±1.6</td>
</tr>
<tr>
<td>33,34</td>
<td>10.7 ±0.2</td>
<td>2.07 ±0.02</td>
<td>20.4 ±1.1</td>
</tr>
<tr>
<td>37,38</td>
<td>7.1 ±0.4</td>
<td>3.46 ±0.04</td>
<td>18.7 ±0.9</td>
</tr>
</tbody>
</table>
What are the differences in data sets?
Errors in Gridding and Flagging

**Gridding** is the process where the boundaries of spots are identified. This is usually done in an automated fashion, followed by manual adjustment of locations and diameters.

**Flagging** is the process where spots with unacceptable characteristics (low intensity, poor background, scratches, etc.) are marked as “bad”.
Case Study 5 – Gridding and Flagging

• One slide from a widely studied microarray experiment on the yeast cell cycle (Spellman data set) was used to evaluate gridding and flagging reproducibility.

• The slide, consisting of ca. 6000 spots, was gridded and flagged by three different operators and a commercial software package (SpotReader), in addition to the original grids given for the data.

• The reproducibility of the grids (location, size) and the flagged spots were examined.
Summary of Results

Surprisingly few spots were gridded in common among the three operators, and only one spot was gridded in the same way by all five approaches.

Likewise, there was significant variability in the spots flagged by different operators.
Despite variations in gridding and flagging, differences in final results were relatively small, contributing to a proportional error of about 3-4%.
In general, errors from microarray experiments appear to approximate a log-normal distribution.

> Are the errors log-normal or are their standard deviations log-normal?

Normalization needs to be used to remove systematic errors from comparator experiments.

> LOWESS?
> What about time course experiments?

A proposed error model: \( \sigma_R^2 = \sigma_{\text{expt}}^2 + \sigma_{\text{meas}}^2 \)

For the experiments examined here, \( \sigma_{\text{expt}} \) appears to follow a proportional error model: \( \sigma_{\text{expt}} = \alpha R \).

> The value of \( \alpha \) was found to vary considerably among replicate data sets, typically 5-30%.
A bootstrap procedure can be used to estimate $\sigma_{meas}$, but in some cases, this appears to underestimate $\sigma_{meas}$ by a factor of 2-4.

> Provides the ability for spot-by-spot assessment of measurement quality.

Although there seems to be significant differences in gridding and flagging by different operators, the contribution of this to the overall uncertainty seems to be relatively small.

It should be noted that the objective of this work was not to obtain precise estimates of measurement uncertainty, but rather to provide an adequate description of error structure suitable for higher level modeling.
Additional Thoughts

- Although log transformations of ratio data can provide measurements with an approximately normal distribution, the effect that this has on the structure of the data itself should not be ignored.

- The availability of measurement error information raises the question of how we use it in higher level analysis. A number of methods currently exist, but more are needed.

  Weighted Principal Components Analysis (WPCA)
  Maximum Likelihood Principal Components Analysis (MLPCA)
  Total Least Squares (TLS)
  Multivariate Curve Resolution by Weighted Alternating Least Squares (MCR-WALS)
Acknowledgements

Dalhousie University
Robert Flight, Tobias Karakach
Joe Boutilier, Kaitlyn MacIntosh,
Yazeed Sobaih, Dr. Rick Singer

University of New Mexico
Prof. Maggie Werner-Washburne
Dr. Juanita Martinez, Dr. Chris Allen,
Dr. David Haaland, Sushmita Roy

IMB·IBM
Institute for Marine Biosciences
NRC Halifax

Dr. Sue Douglas & Co.

Financial Support:

Natural Sciences and Engineering Research Council of Canada

The Dow Chemical Company

MITACS