Statistical Monitoring of Heteroscedastic Dose-Response Profiles from High Throughput Screening

J. D. Williams
GE Global Research Center

Jeffrey B. Birch
Virginia Tech

William H. Woodall
Virginia Tech

Nancy M. Ferry
DuPont Crop Protection

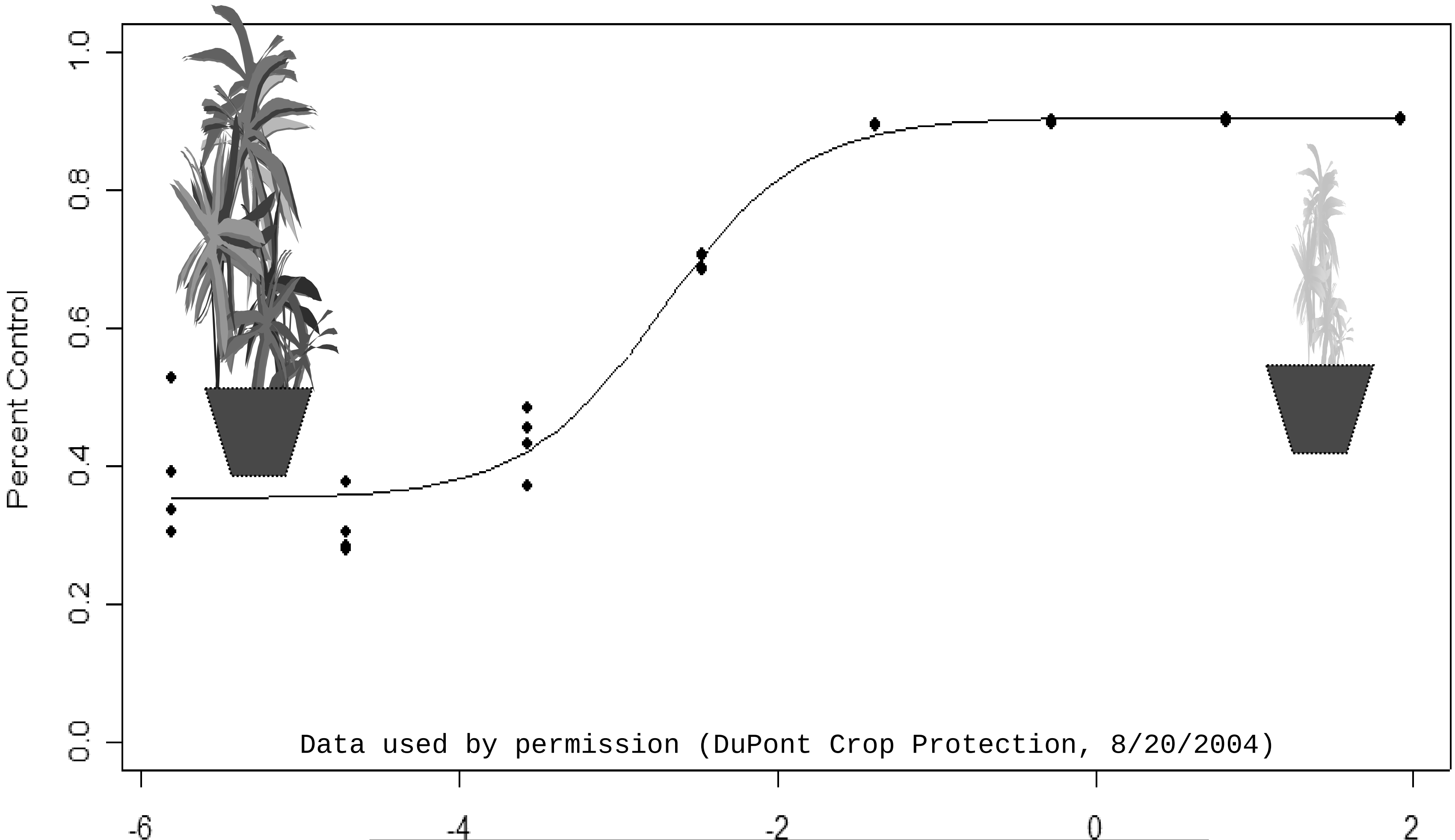
Introduction

Background and Motivation

- **High Throughput Screening (HTS): Use bioassays to test thousands of experimental compounds**
- **Bioassay quality is key to successful HTS**
- **Test commercial chemicals alongside experimental compounds to check the quality of bioassay**
- **If a bioassay experiment of the commercial chemical is undesirable for a given week, then the data generated by the HTS warrants further consideration for that run.**

Introduction

High Throughput Screening (HTS)



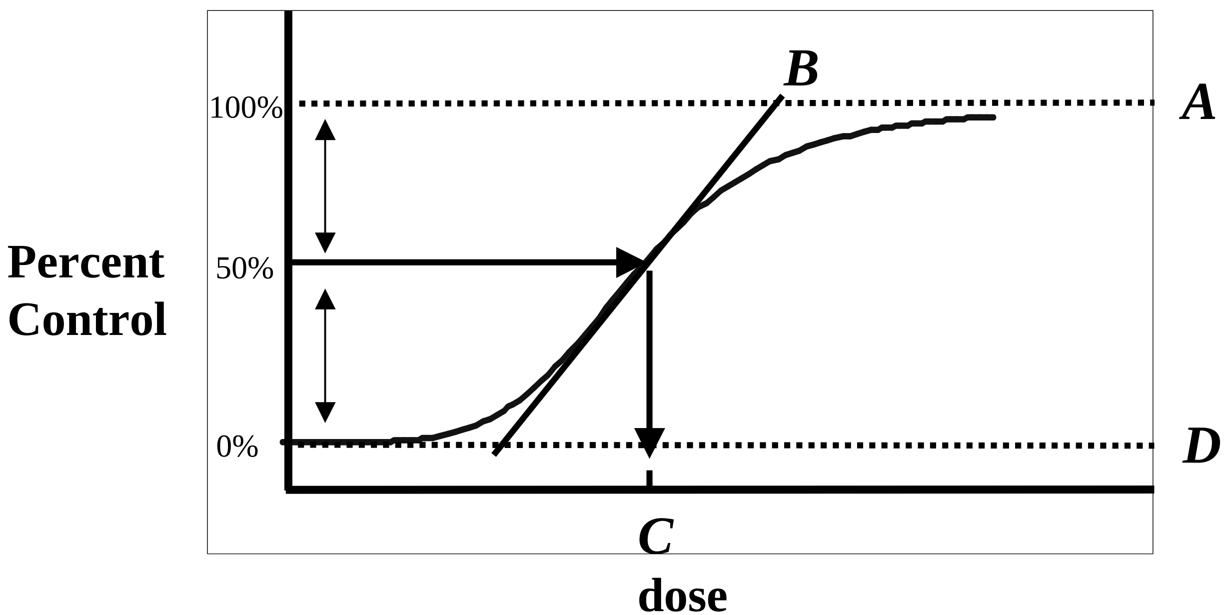
log(dose) of commercial chemical

Profile Estimation

Dose-response Model

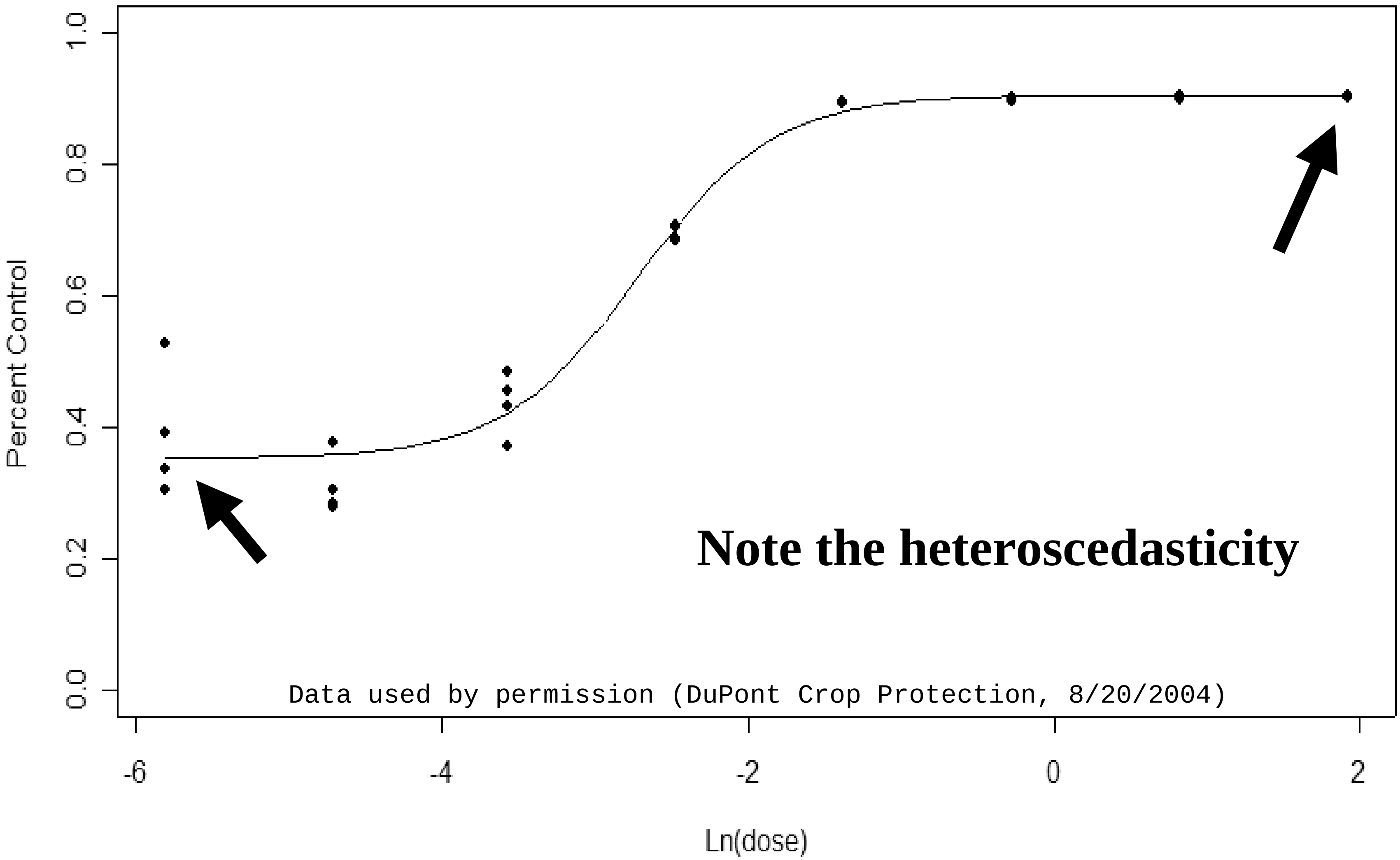
For profile i, dose j, and replication k:

$$PercentControl_{ijk} = A_i + \frac{D_i - A_i}{1 + \left(\frac{dose_{ij}}{C_i}\right)^{B_i}} + \varepsilon_{ijk}$$



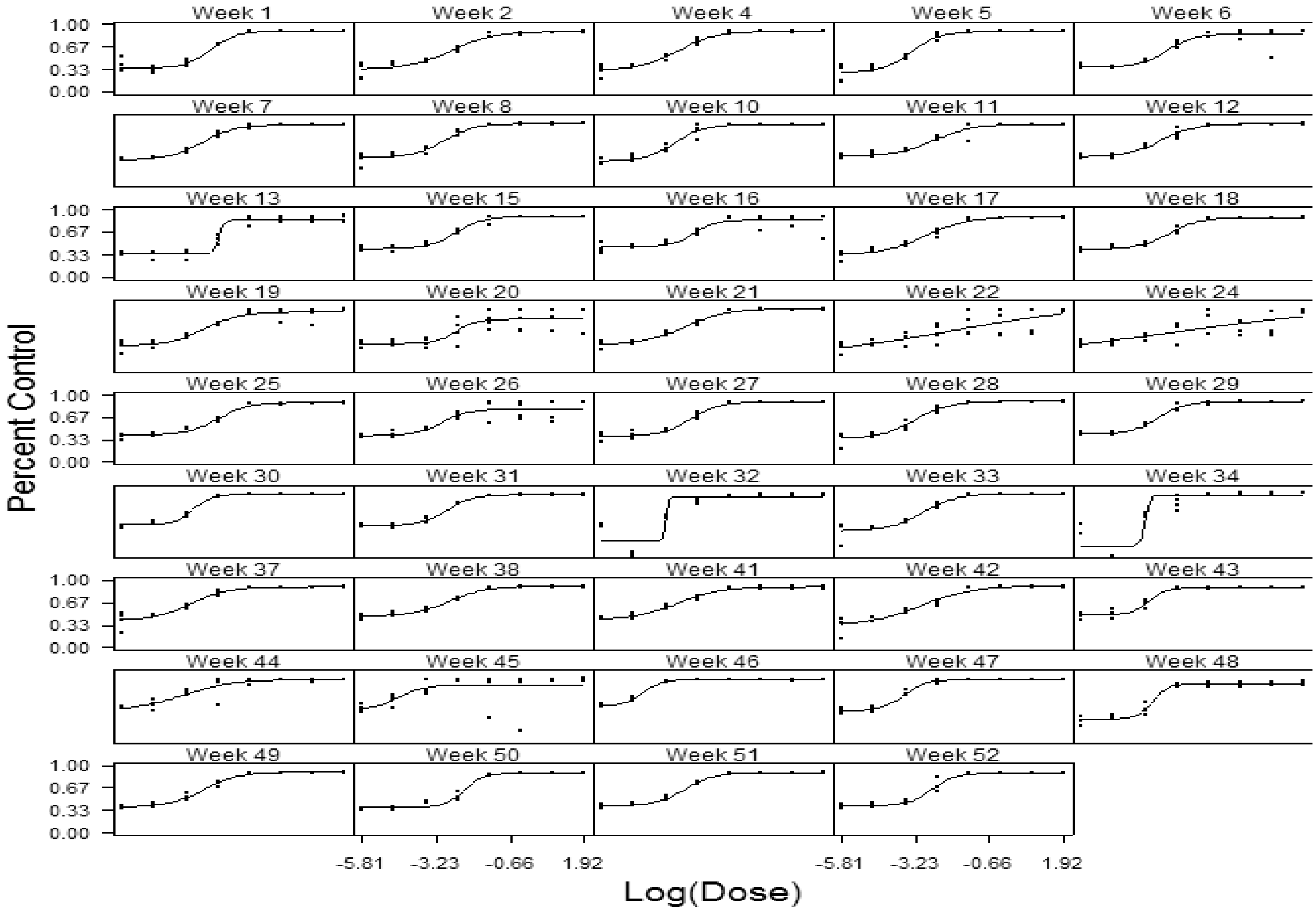
Profile Estimation

Estimated Mean Profile for Week 1



Profile Estimation

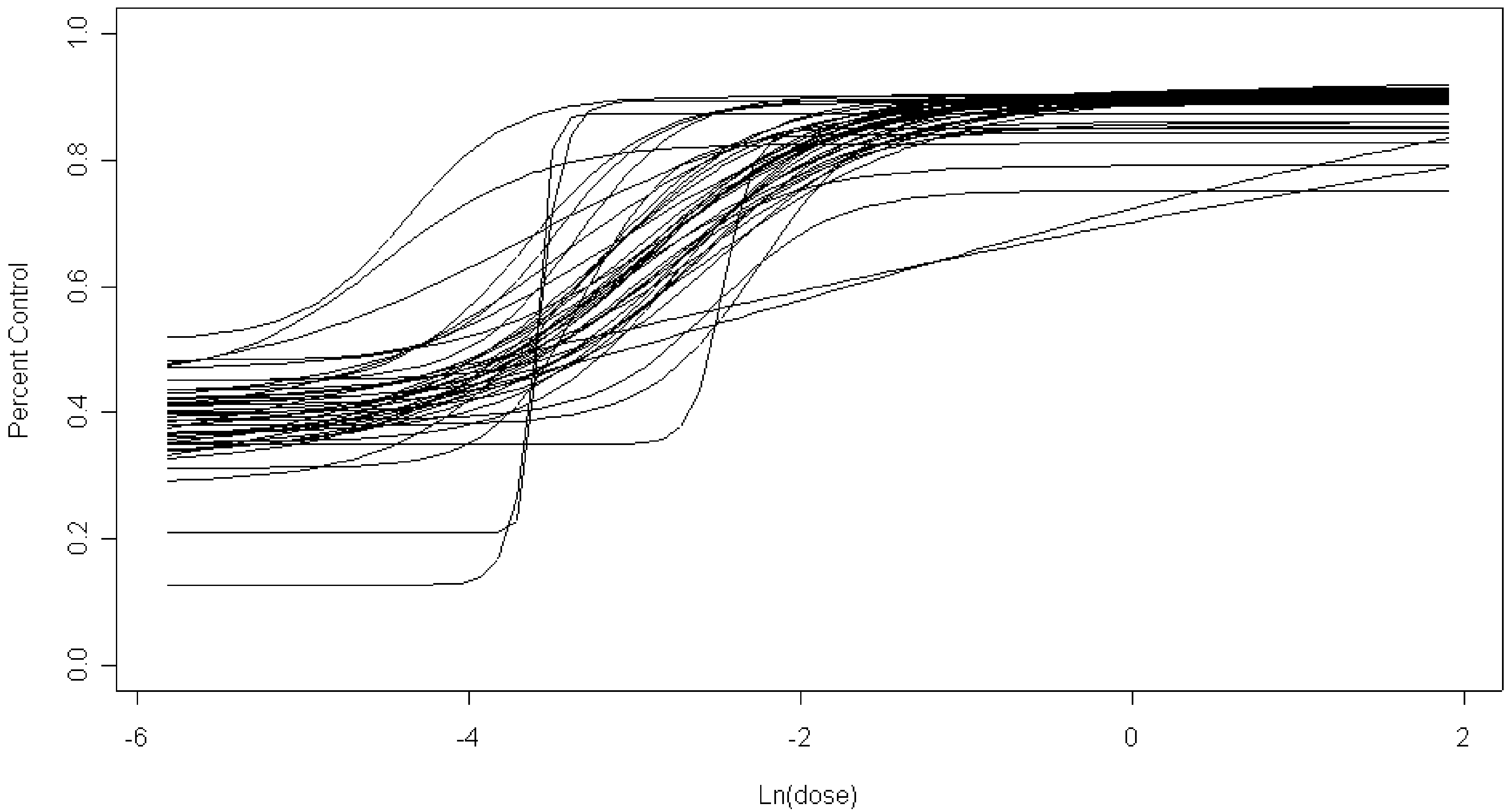
44 Estimated Mean Profiles



Data used by permission (DuPont Crop Protection, 8/20/2004)

Profile Estimation

44 Estimated Mean Profiles – overlaid



Profile Estimation

Variance Function Model

General Variance Function Model:

$$Var(y_{ijk}) = \sigma_i^2 g(z_{ij}, \beta_i, \theta_i)$$

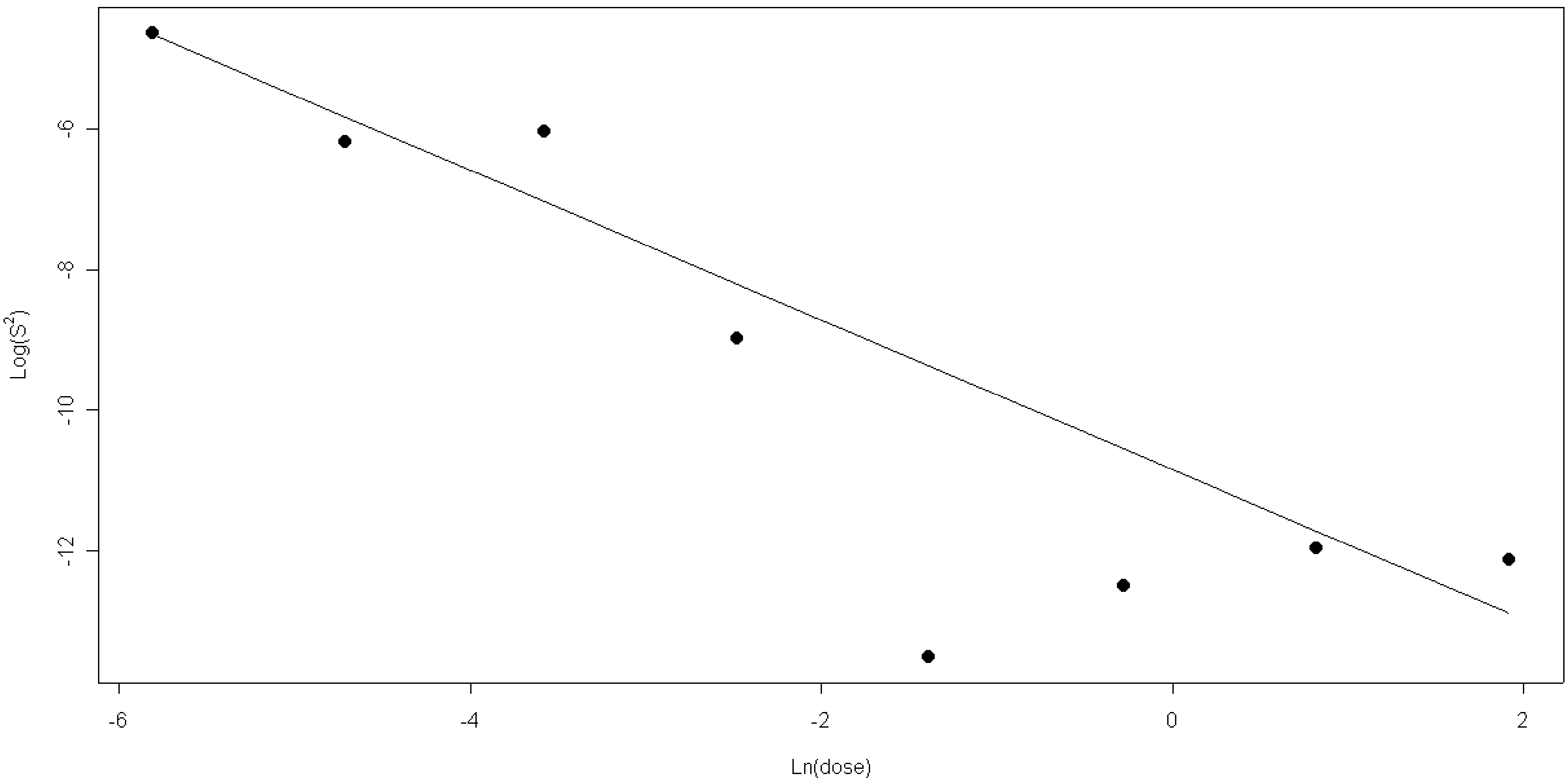
(Davidian and Carroll, 1987)

Can be written as a Generalized Linear Model (GLIM):

$$Var(y_{ijk}) = e^{\theta_{0,i} + \theta_{1,i} \log(dose_{ij})}$$

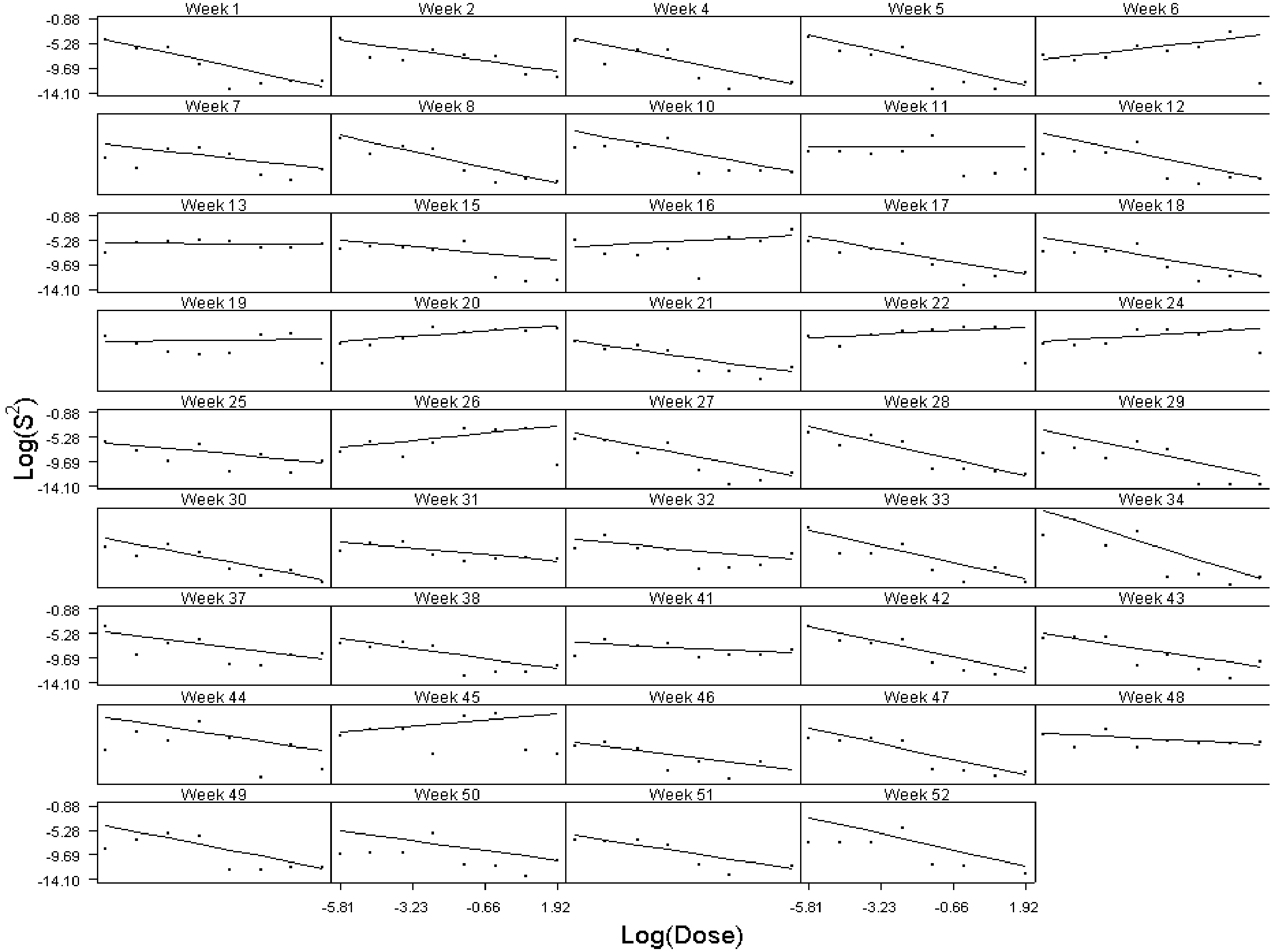
Variance Profile Estimation

Estimated Variance Profile for Week 1



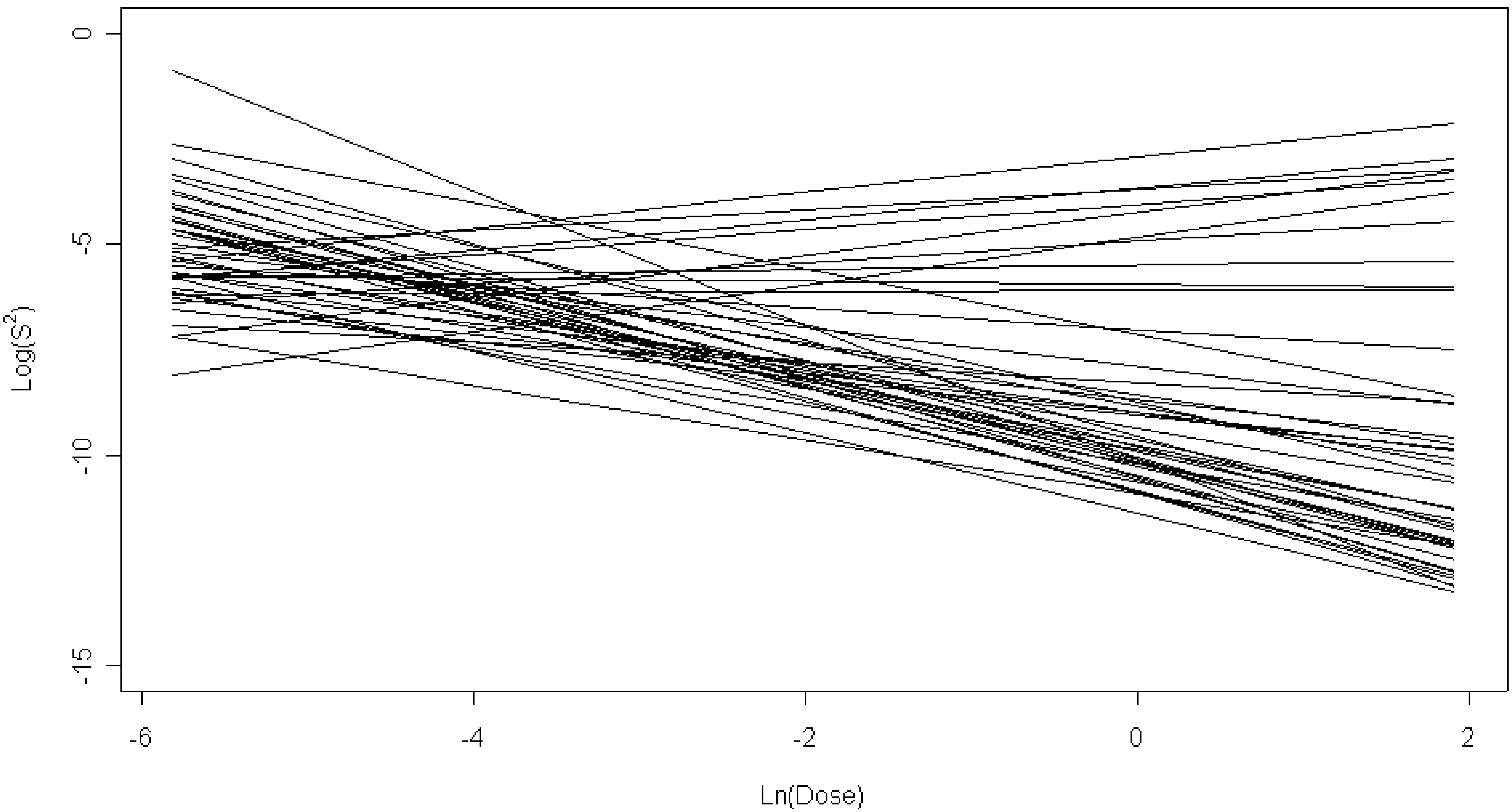
Variance Profile Estimation

44 Estimated Variance Profiles



Variance Profile Estimation

44 Estimated Variance Profiles – overlaid



Control Charts

<u>Goal</u>: Establish in-control model parameters for subsequent continuous bioassay monitoring

1. **<u>Variance Profile Chart</u>**
   - **Check for abnormal variances profiles**
2. **<u>Lack-of-Fit Chart</u>**
   - **Check for adequacy of model**
3. **<u>Hotelling's T^2 Chart</u>**
   - **Check for abnormal estimates of *A*, *B*, *C*, and *D***

Control Charts

Variance Profile Chart: Highlights

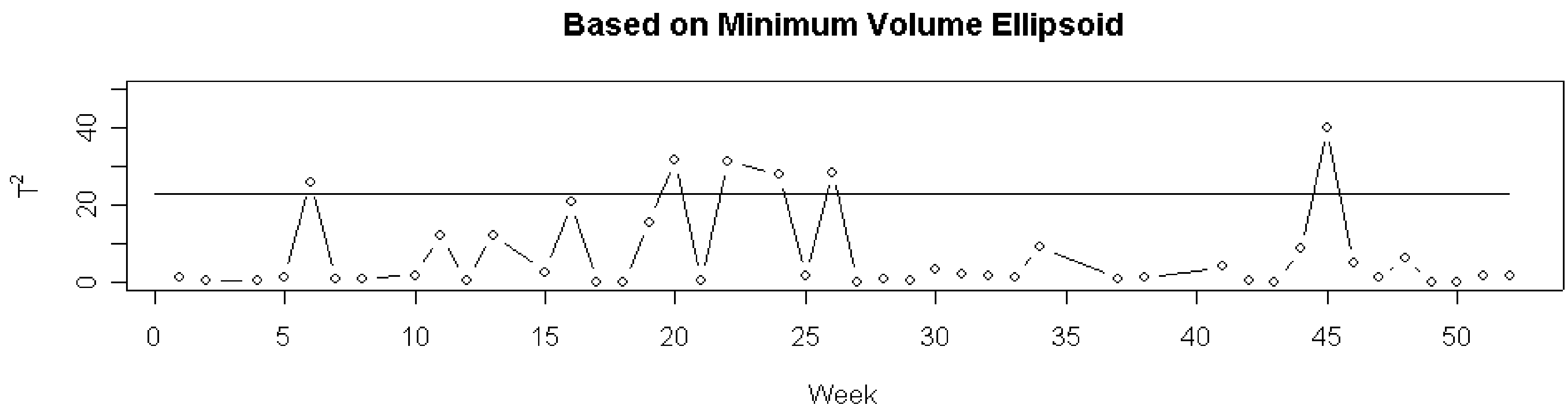
- **Need to identify abnormal values of θ_0 and θ_1**
- **However, θ_0 and θ_1 are correlated**
- **Solution: Use Hotelling's T^2 statistic**

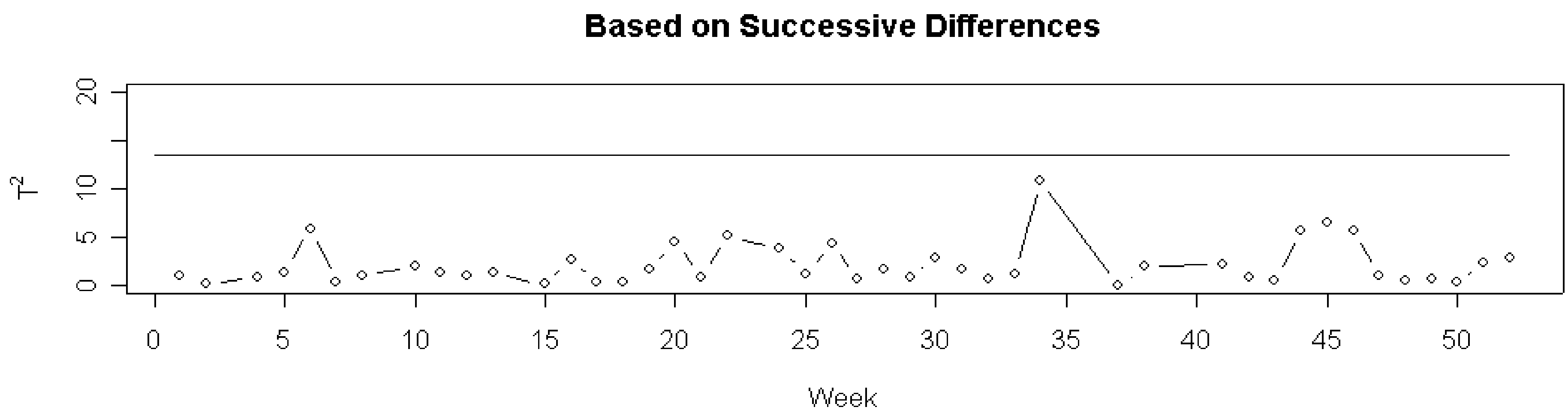
$$T_i^{\,2} = \left(\hat{\theta}_i - \hat{\mu}_\theta\right)' \mathbf{S}^{-1} \left(\hat{\theta}_i - \hat{\mu}_\theta\right) \quad \textbf{where} \quad \hat{\boldsymbol{\theta}}_i = \begin{pmatrix} \hat{\theta}_{0,i} \\ \hat{\theta}_{1,i} \end{pmatrix}$$

- **Various versions of T^2 statistic available**
- **Here we consider T^2 based on the minimum volume ellipsoid and successive differences estimator (see Williams, Woodall, Birch, and Sullivan (JQT, July 2006))**

Control Charts

Variance Profile Charts





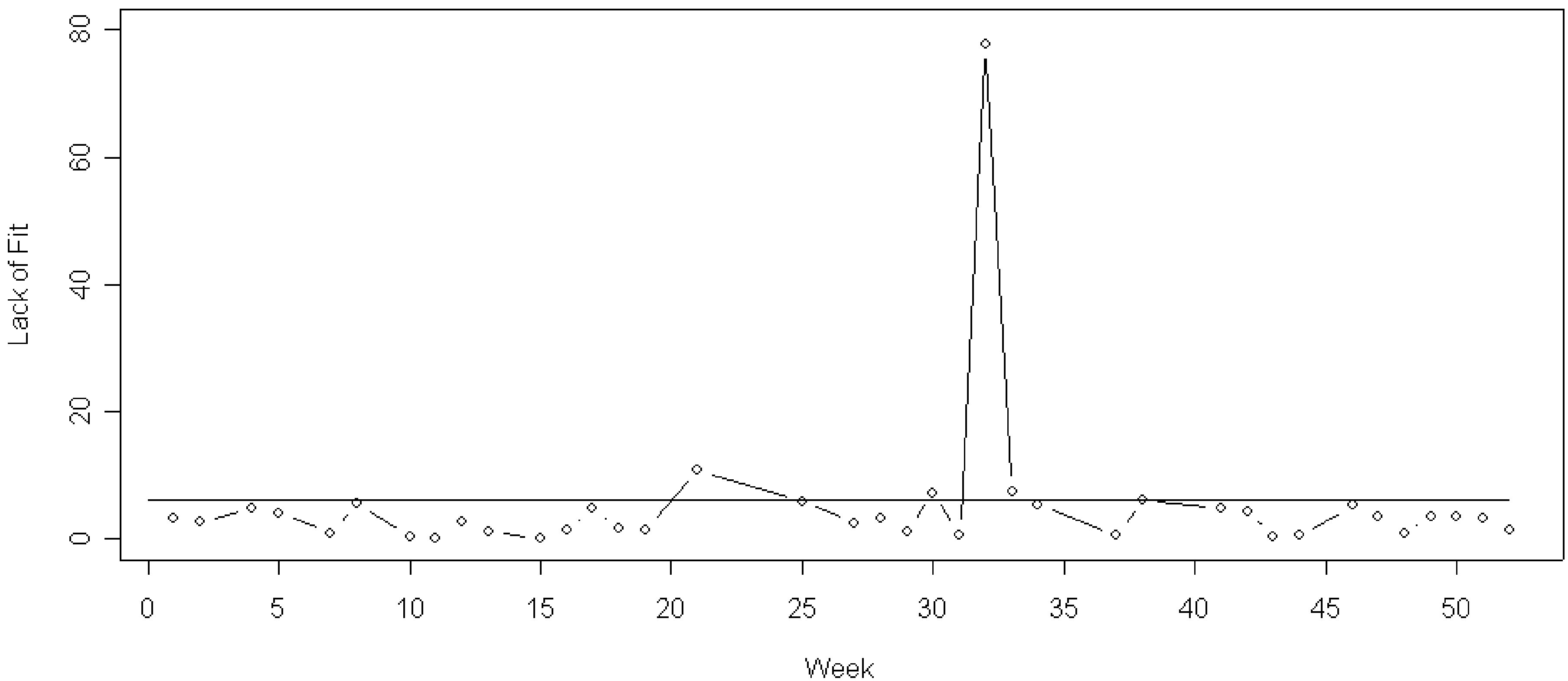
Control Charts

Lack-of-fit Chart: Highlights

- **Need to identify changes in model form**
- **For replicated data, we can calculate lack-of-fit (LOF)**
- **Problem: Heteroscedastic data**
- **Solution: Replace sums of squares with weighted sums of squares in the lack-of-fit F statistic**
- **Resulting statistic has an approximate F distribution**

Control Charts

Lack-of-fit Chart



Control Charts

T^2 Chart: Highlights

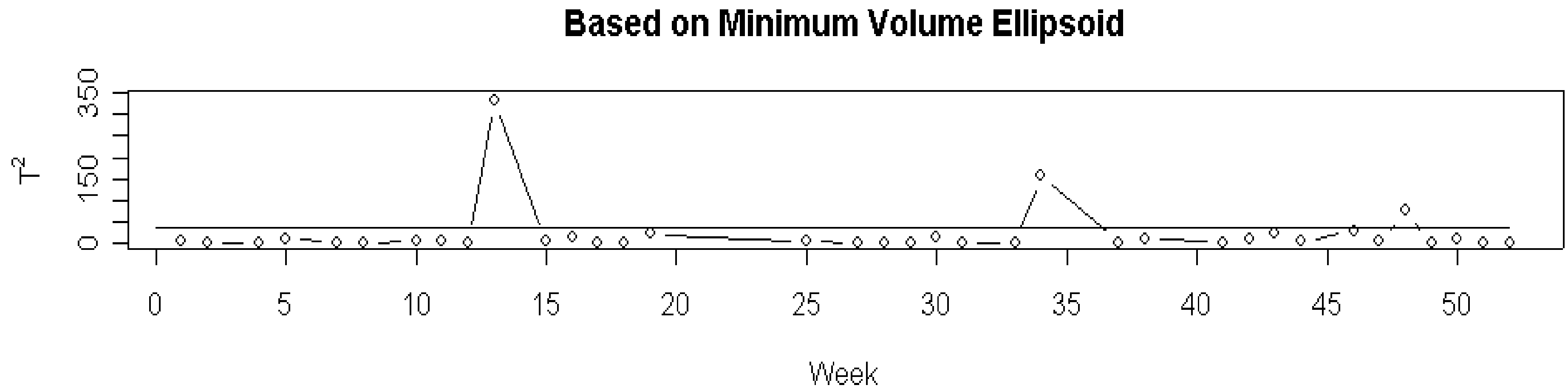
- **Need to identify abnormal estimated values of *A*, *B*, *C*, and *D***
- **However, their estimators are correlated**
- **Solution: Use Hotelling's T^2 statistic**

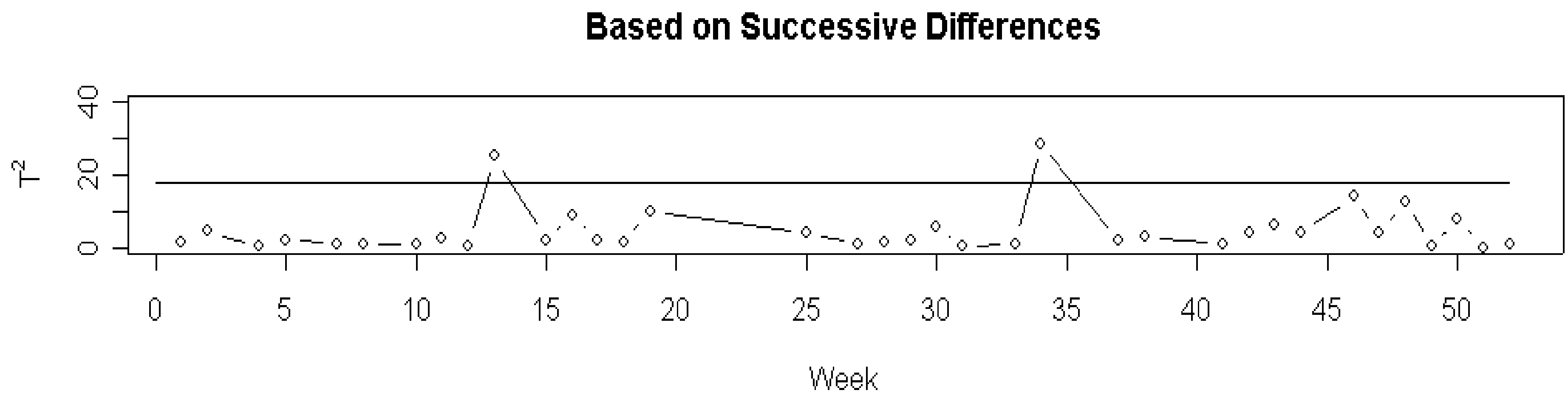
$$T_i^2 = \left(\hat{\boldsymbol{\beta}}_i - \overline{\hat{\boldsymbol{\beta}}}\right)' \mathbf{S}^{-1} \left(\hat{\boldsymbol{\beta}}_i - \overline{\hat{\boldsymbol{\beta}}}\right) \quad \textbf{where} \quad \hat{\boldsymbol{\beta}}_i = \begin{pmatrix} \hat{A}_i \\ \hat{B}_i \\ \hat{C}_i \\ \hat{D}_i \end{pmatrix}$$

- **Various versions of T^2 statistic available**
- **Here we consider T^2 based on the minimum volume ellipsoid and successive differences estimator**

Control Charts

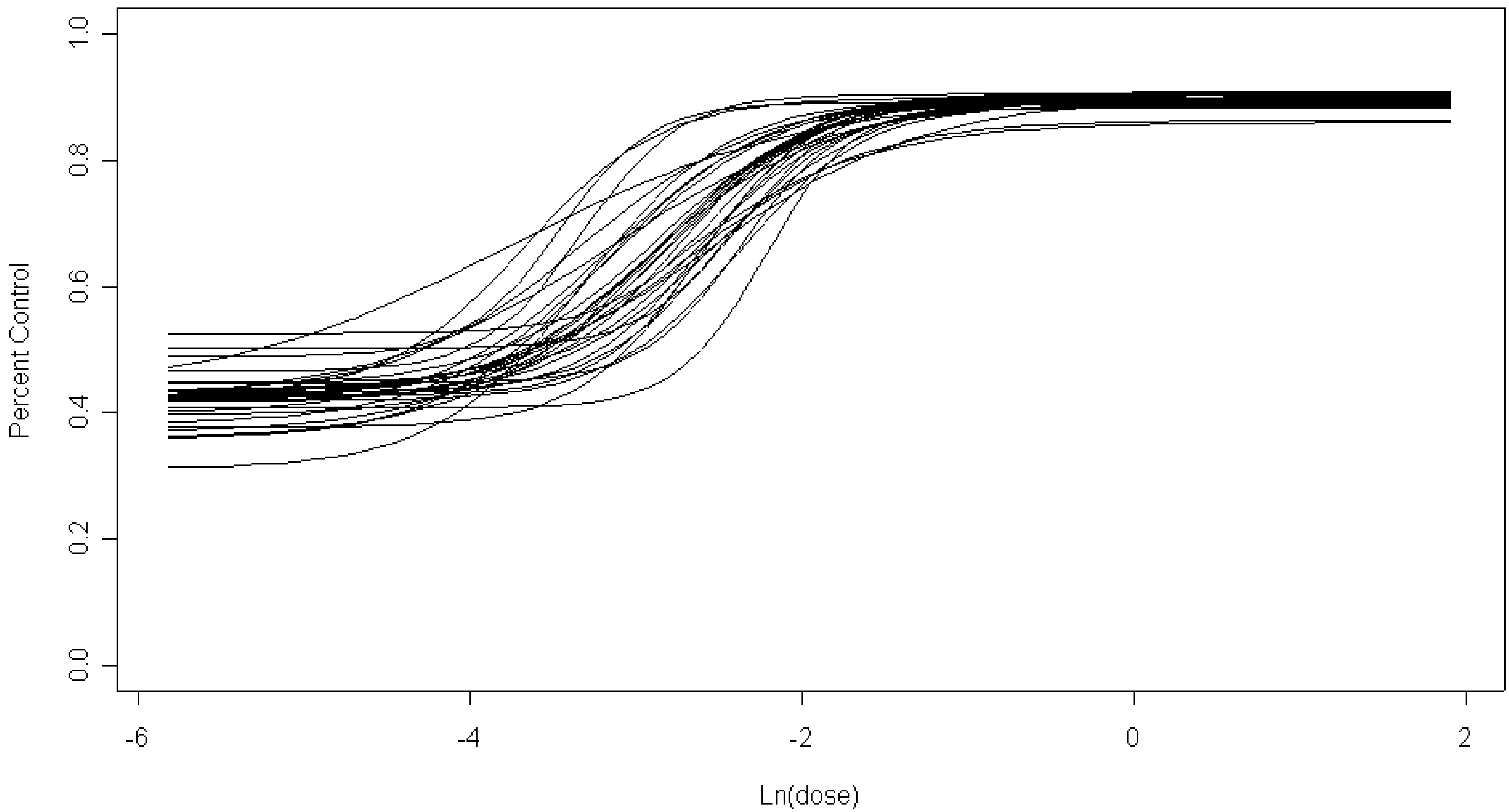
T^2 Charts





Control Charts

In-control Estimated Profiles



Conclusions

Nonlinear Profile Monitoring

- **Nonlinear profile monitoring requires 3 charts:**
  - **Variance Profile Chart**
  - **Lack-of-Fit Chart**
  - **T^2 Chart**
- **Nonlinear profile monitoring is part of the relatively new field of profile monitoring**
- **Many engineering and biological applications of profile monitoring**
- **Profile monitoring papers can be obtained upon request**

Some References

Davidian, M., and Carroll, R. J. (1987), "Variance Function Estimation" *Journal of the American Statistical Association* 82, pp. 1079-1091.

Janzen, W. P. (2002), *High Throughput Screening: Methods and Protocols*, Totowa, New Jersey: Humana Press.

Williams, J. D., Woodall, W. H., Birch, J. B. (2003), "Phase I Analysis of Nonlinear Product and Process Quality Profiles," Technical Report No. 03-5, Department of Statistics, Virginia Polytechnic Institute and State University.

Williams, J. D., Woodall, W. H., Birch, J. B., and Sullivan, J. H. (2006), "On the Distribution of T^2 Statistics Based on Successive Differences" *Journal of Quality Technology*, In print.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004), "Using Control Charts to Monitor Process and Product Quality Profiles" *Journal of Quality Technology* 36, pp. 309-320.

Contact Information

J.D. Williams, Ph. D.

Statistical Leader

GE Global Research Center

1 Research Circle, K1-4C-27A

Schenectady, NY 12309

James.Williams@research.ge.com