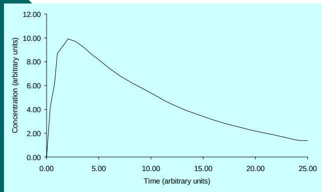
Assessing Nonlinear Mixed Effects Model Parameter Estimates via Profiling of the True Likelihood

Paolo Vicini, Ph.D.
Resource Facility for Population Kinetics (RFPK)
Department of Bioengineering
University of Washington, Seattle, WA
http://www.rfpk.washington.edu



Nonlinear Mixed Effects Model Building

Hierarchical Variability and Averaging

$$\dot{A}_1(t) = -K_a A_1(t) + Dose(t)$$

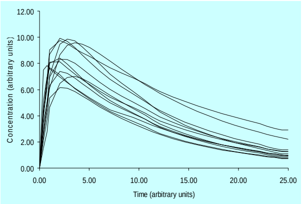
$$\dot{A}_2(t) = +K_a A_1(t) - (CL/V) A_2(t)$$

$$s(t, K_a, CL, V) = A_2(t)/V$$

$Ka, CL, V \sim N(\theta, \omega)$

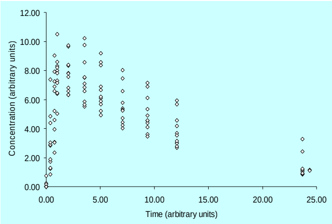
Between-individual (BSV)
"Nonlinear "

Residual unknown (RUV)
"Linear"



$y = s(t, \theta, \omega)$

$\varepsilon \sim N(0, \sigma)$



$y = s(t, \theta, \omega) + \varepsilon$

A Major Obstacle

- The individual model prediction

  $$y_i = s_i(\theta, \eta_i) + \varepsilon_i$$

  implies the population likelihood:

  $$L(\theta) = \prod_{i=1}^{N} L_i(\theta, \eta_i, \varepsilon_i)$$

- But: L depends on the nonlinear function s!
- An integral is needed to express the likelihood as a function of fixed effects:

  $$L(\theta, \omega, \sigma) = \int_{-\infty}^{+\infty} \prod_{i=1}^{N} L_i(\theta, \eta_i, \varepsilon_i) d\eta_i$$

Nonlinear Mixed Effects Models

- Challenge: separating variability (BSV) from noise (RUV)
- Reason: the underlying model is nonlinear
- Solution: *Approximate* methods
  - First-Order
  - First-Order Conditional (Expected Hessian)
  - Laplacian
- *Other* methods are also available
  - Two-Stage (STS, GTS, ITS)
  - Nonparametric

Challenges of Linearization

- A simple example about linearization ~0
- The exponential function: $\exp(\theta) \sim 1 + \theta$

θ	$\exp(\theta)$	$1+\theta$
0	1	1
0.01	1.01005	1.01
0.1	1.105171	1.1
1.0	2.718282	2
5.0	148.4132	6

Maximum Likelihood in Practice

- Linearization of the model function allows for explicit solution of the integral
- However, every flavor of linearization has its own objective function (figure of merit)
- If the model is nonlinear, all available objective functions yield different results
- General rules on linearization’s impact are difficult to define

Monte Carlo Integration

- Monte Carlo integration is a common approach to the evaluation of complex integrals
- This approach would provide the true marginal likelihood evaluated at and near the estimate, *however obtained*
- This can be used to evaluate methods that optimize *different approximations* to the marginal likelihood

Monte Carlo Samplers

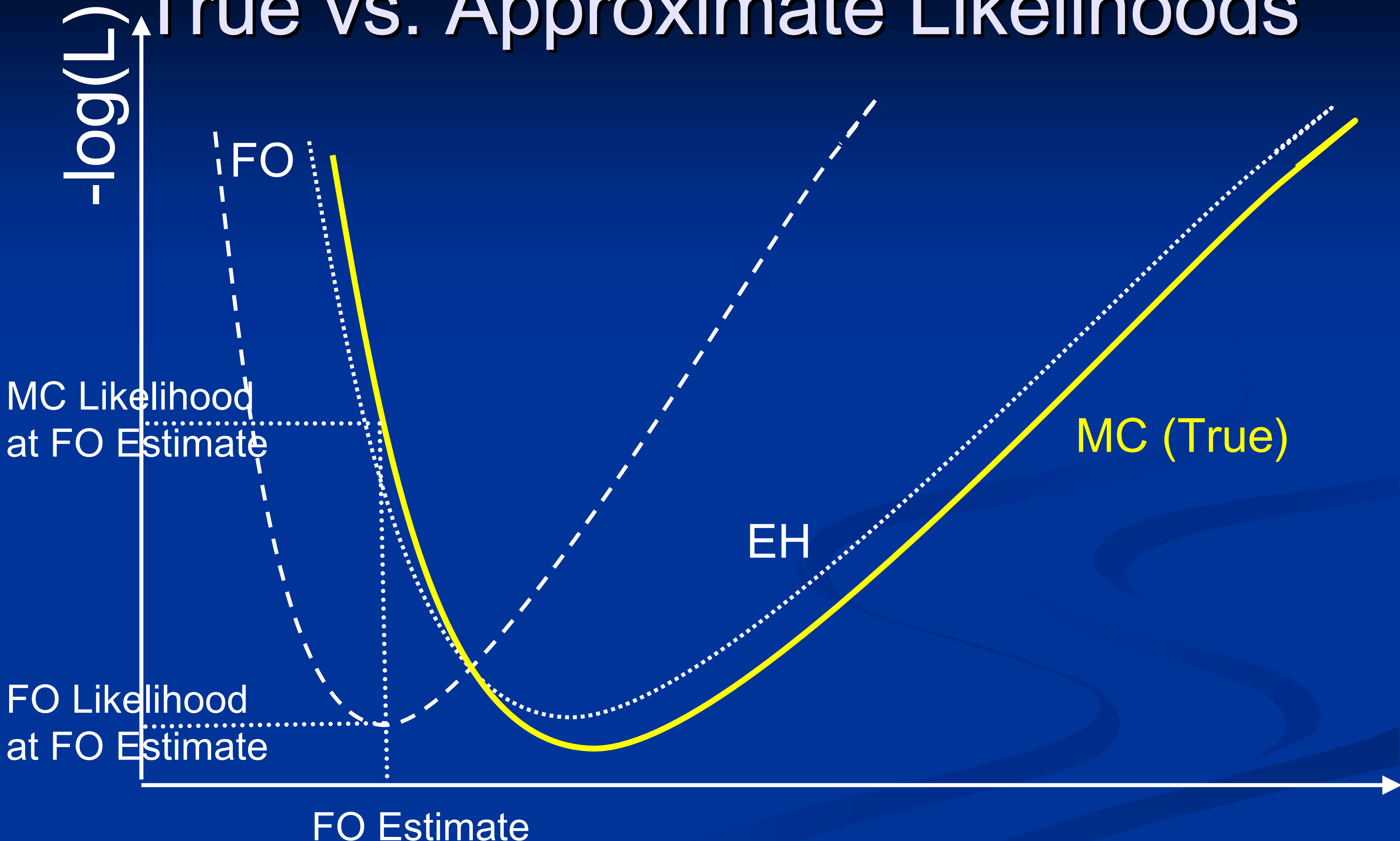
- There is a number of potential samplers
- Some of those we used/explored are:
  - Plain Monte Carlo
    - Uniform sampling over the integration region
  - Grid
    - Samples on a grid over the random effects
  - MISER
    - Samples in regions of highest variance
  - VEGAS
    - Samples in regions that most contribute to the integral
  - ADAPT
    - Globally adaptive algorithm (Genz)

Algorithms: The GNU Scientific Library, http://www.gnu.org/software/gsl/; Genz A.: Computing in the 90s, Lecture Notes in Computer Science Vol. 507, Springer-Verlag, New York, 1991, pp. 279-292.

Likelihood Profiling

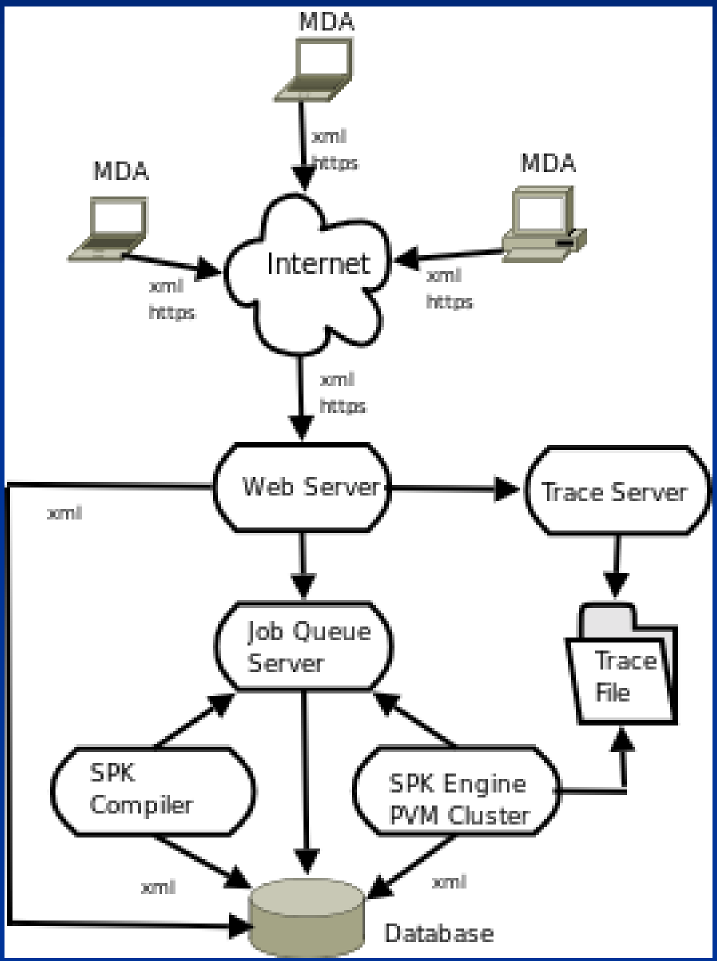
- The idea is to evaluate the ML integral at a putative solution for the nonlinear mixed effects problem, and at two bias positions
- The profiling would allow one to conclude whether the approximate solution is indeed “close” to the solution of the mixed effects problem, without approximation (except as dictated by the accuracy of the simulation)

True vs. Approximate Likelihoods



True, FO and EH negative log-likelihoods superimposed (qualitative plot); for profiling, evaluation is done at the estimate and two bias positions [e.g. 2x(upperL – lowerL)/100]

The System for Population Kinetics



- The SPK web service was used for the preliminary calculations described in this presentation
- The free system can be accessed via a web browser and uses a Java interface for model building
- For more information:
  - http://spk.rfpk.washington.edu

Simple Example

- Nondestructive stress test in railway rails (ultrasound travel times)
- Data excerpt:

Rail	travel	
1	1	55
2	1	53
3	1	54
4	2	26
…		

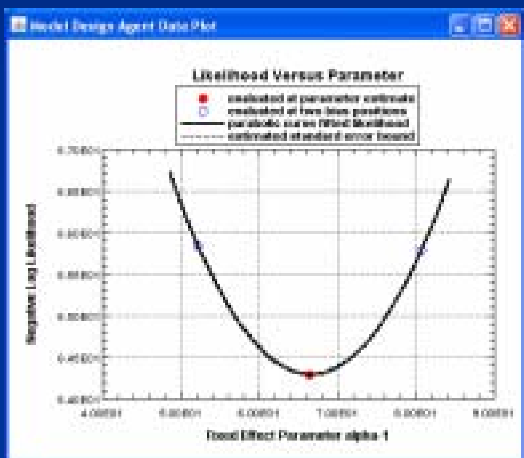
$$\boxed{y_{ij} = \theta + \eta_i + \varepsilon_{ij}}$$

- Linearized ML objective functions: OF = 64.28
  - The model is linear

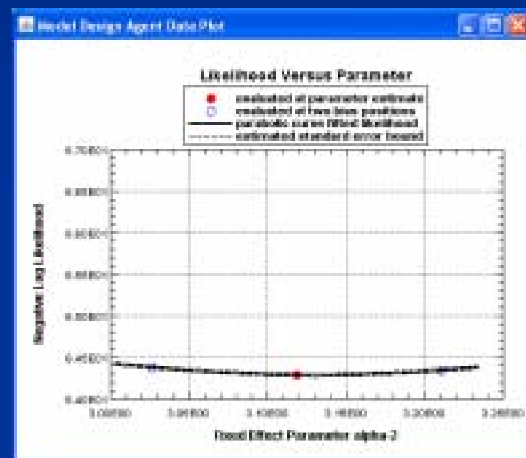
8718 (Vegas)

Linear Model Profiling

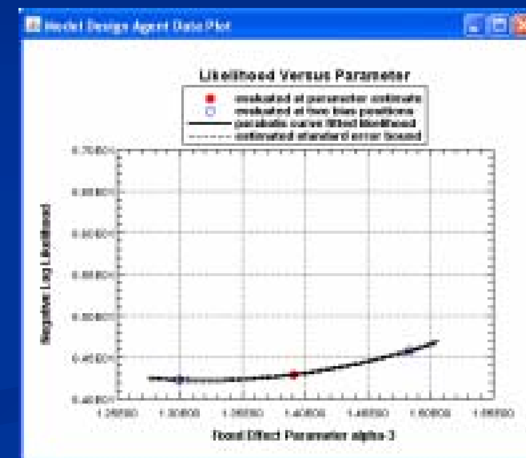
THETA(1)



OMEGA(1,1)

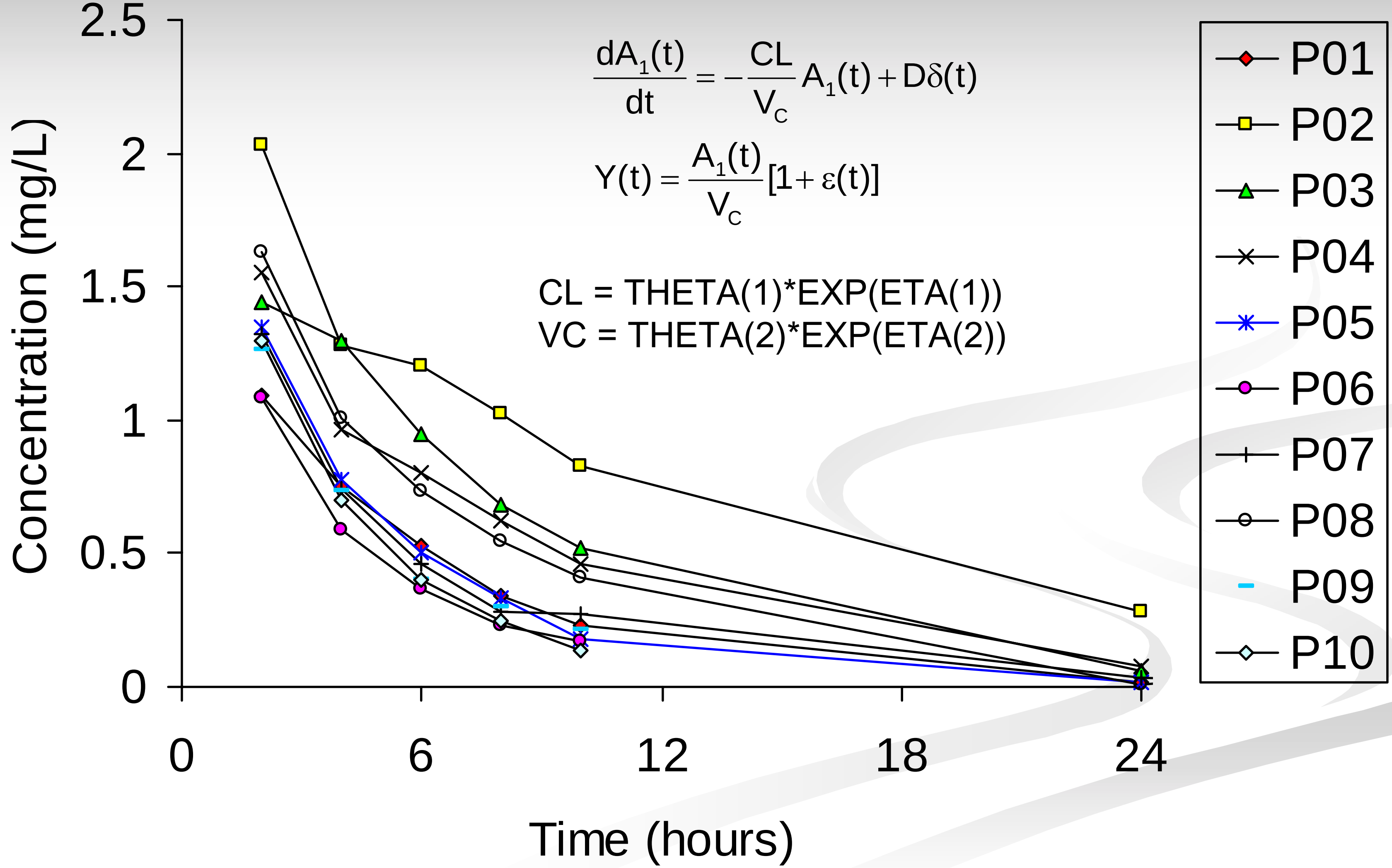


SIGMA(1)



- The fixed effects precision varies
- Precision is related to profile curvature

Cadralazine Data (N=10)



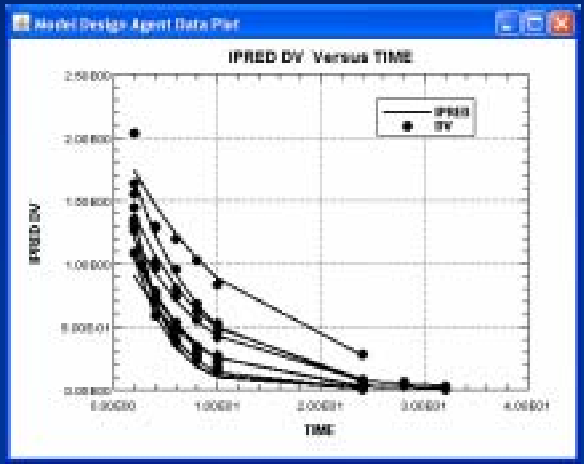
Applied Statistics 43, No 1, 201-221,1994

Parameter Estimates (Base)

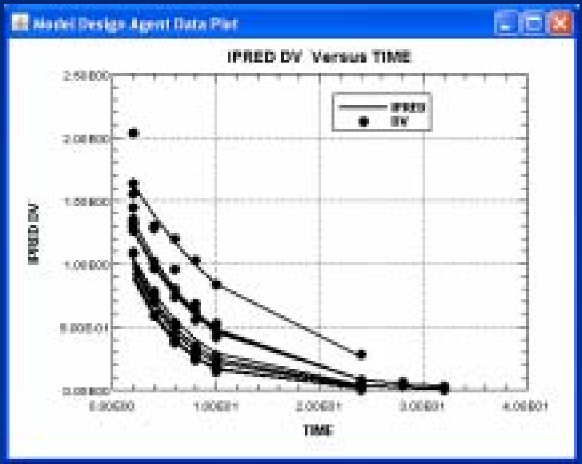
- Standard Two Stage
  - THETA(1) = 3.14
  - THETA(2) = 16.9
  - OMEGA(1,1) = 0.105
  - OMEGA(1,2) = 0.0141
  - OMEGA(2,2) = 0.0448
  - SIGMA(1,1) = 0.109
- First Order
  - THETA(1) = 2.22
  - THETA(2) = 18.5
  - OMEGA(1,1) = 0.168
  - OMEGA(1,2) = 0.0490
  - OMEGA(2,2) = 0.0171
  - SIGMA(1,1) = 0.0316
- Iterative Two Stage
  - THETA(1) = 3.21
  - THETA(2) = 17.9
  - OMEGA(1,1) = 0.146
  - OMEGA(1,2) = 0.0475
  - OMEGA(2,2) = 0.0203
  - SIGMA(1,1) = 0.0864
- Expected Hessian
  - THETA(1) = 3.11
  - THETA(2) = 18.8
  - OMEGA(1,1) = 0.133
  - OMEGA(1,2) = 0.0226
  - OMEGA(2,2) = 0.00386
  - SIGMA(1,1) = 0.121

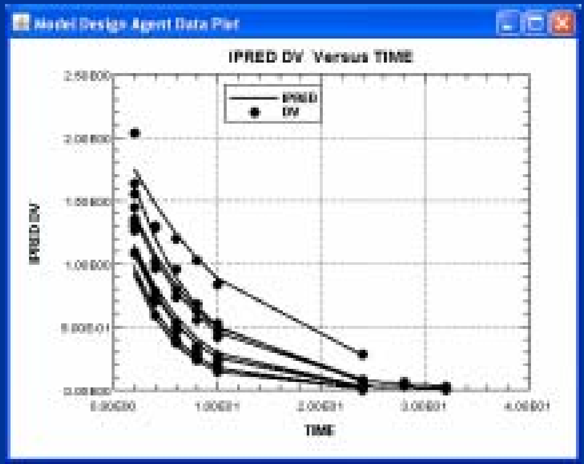
Cadralazine Example - IPRED

Iterative Two Stage (no OF)

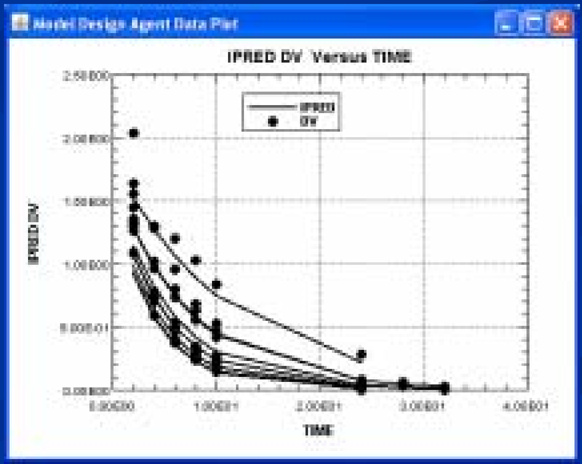


First Order (OF -46.29)





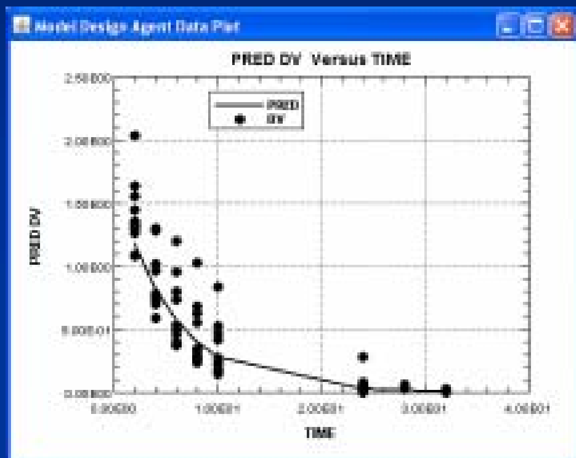
Standard Two Stage (no OF)



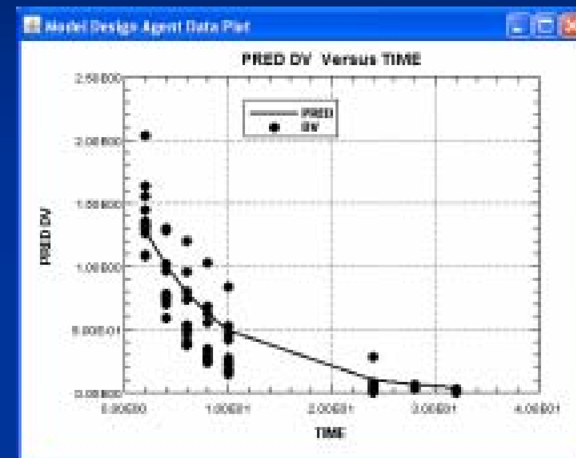
Expected Hessian (OF -39.91)

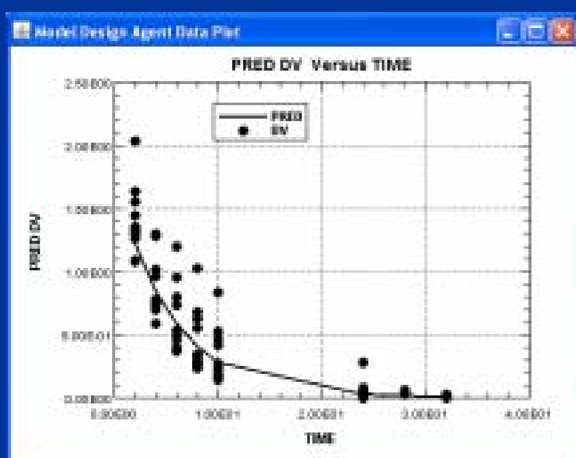
Cadralazine Example - PRED

Iterative Two Stage (no OF)

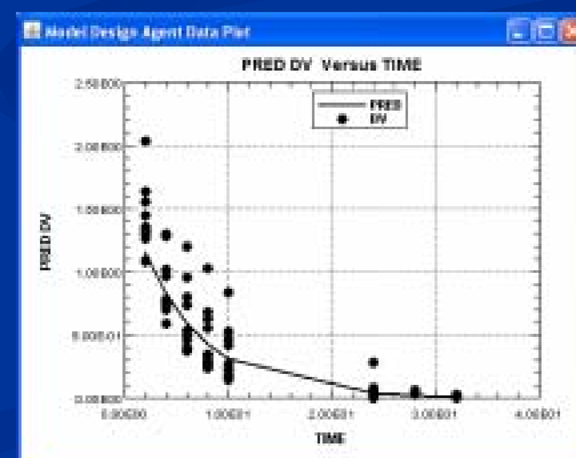


First Order (OF -46.29)





Standard Two Stage (no OF)

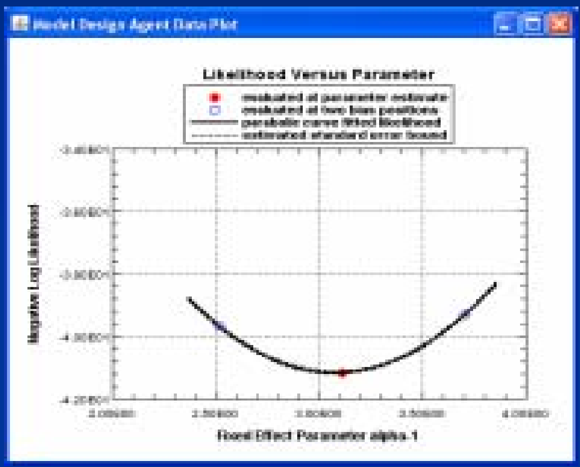


Expected Hessian (OF -39.91)

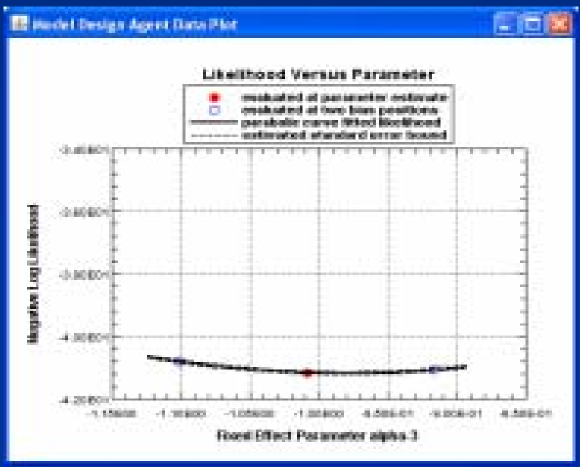
Cadralazine Example – EH

OF = -39.91, -LL = -41.17±0.06

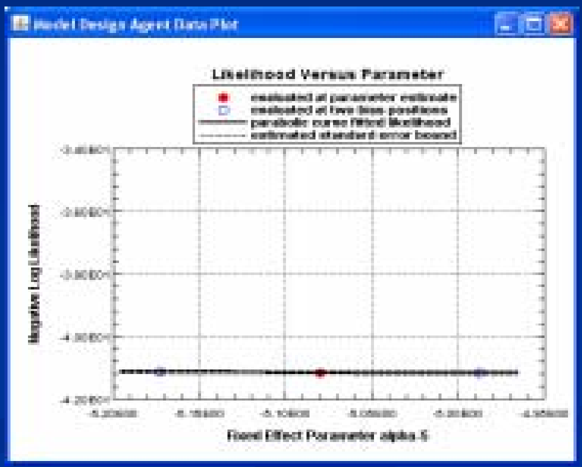
THETA(1)

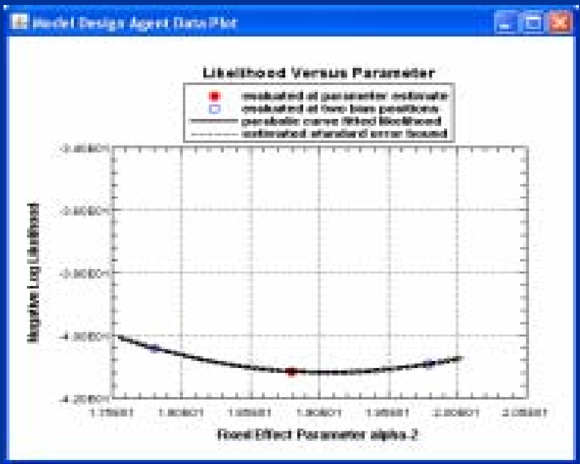


OMEGA(1,1)

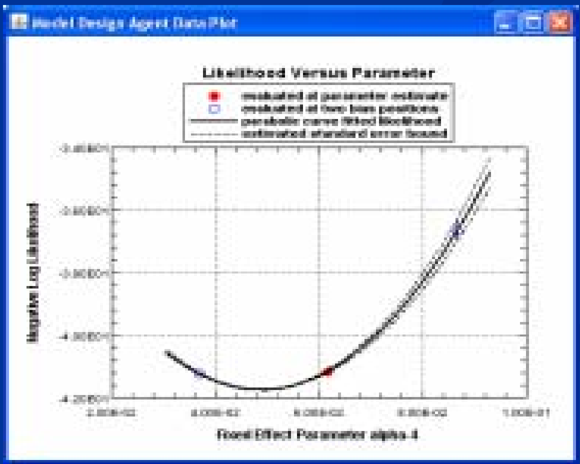


OMEGA(2,2)

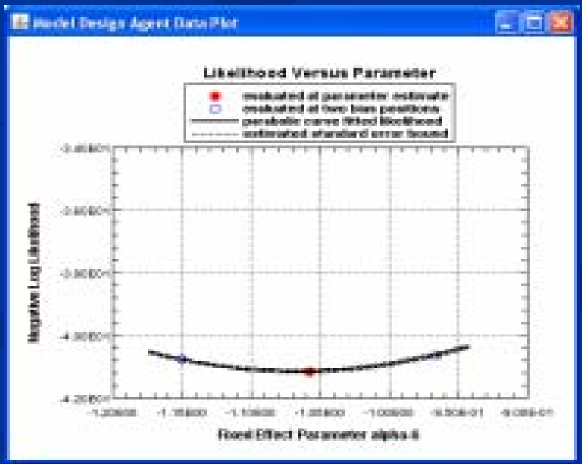




THETA(2)



OMEGA(1,2)



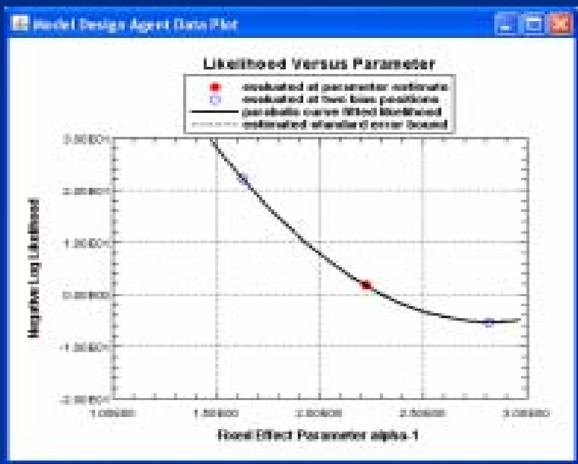
SIGMA(1)

8723 (Vegas)

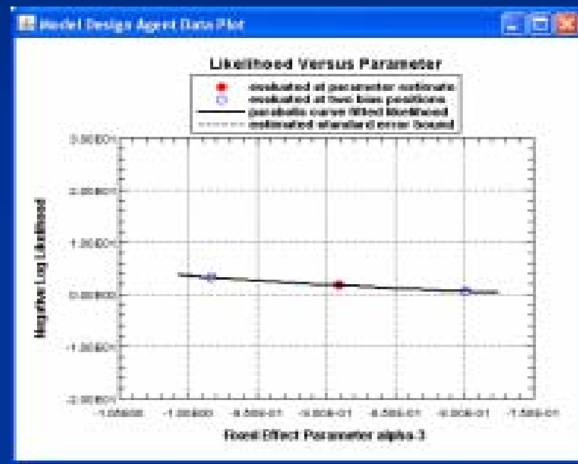
Cadralazine Example - FO

OF = -46.29, -LL = 1.60±0.03

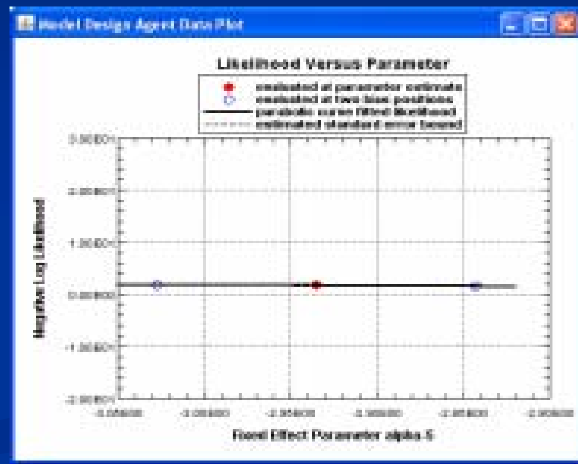
THETA(1)

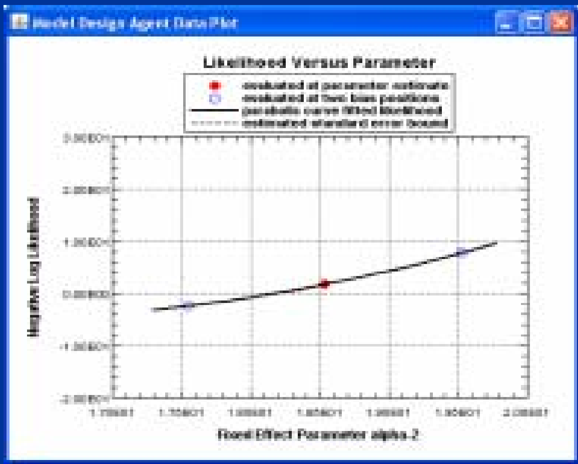


OMEGA(1,1)

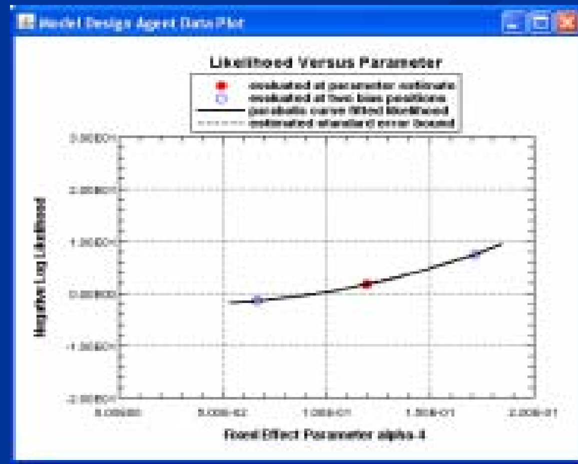


OMEGA(2,2)

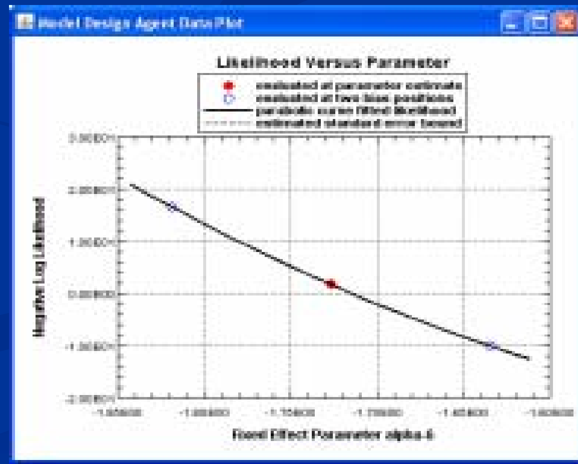




THETA(2)



OMEGA(1,2)

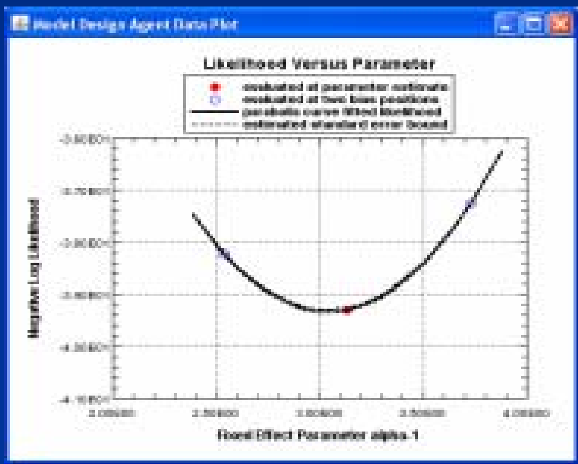


SIGMA(1)

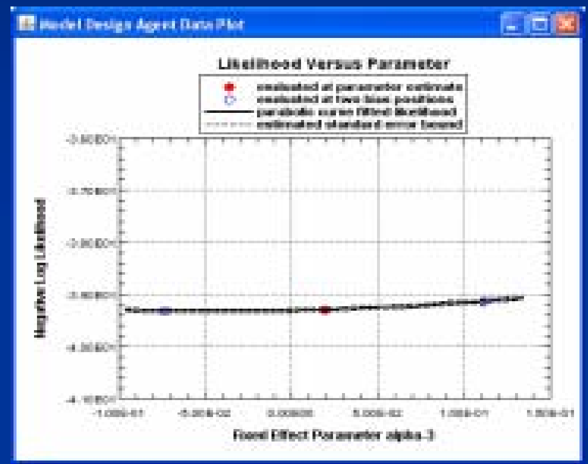
Cadralazine Example - STS

-LL = -39.30±0.04

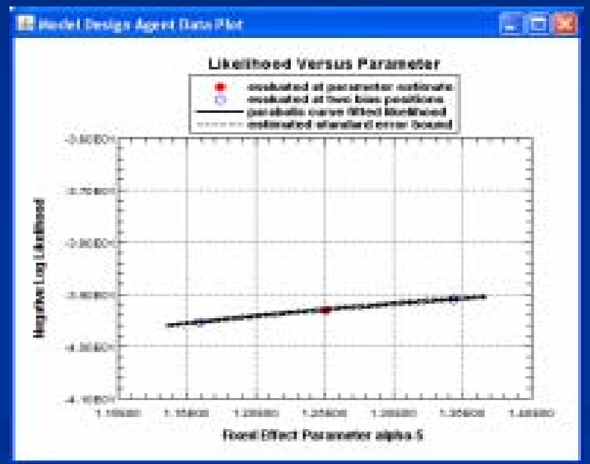
THETA(1)

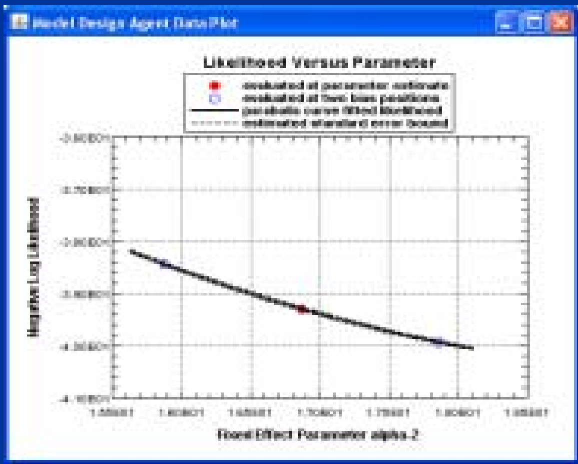


OMEGA(1,1)

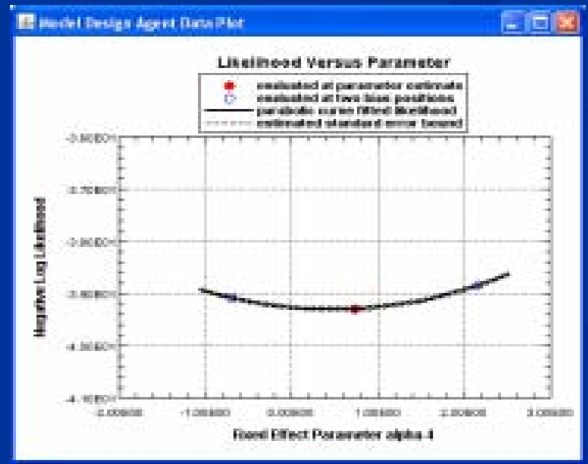


OMEGA(2,2)

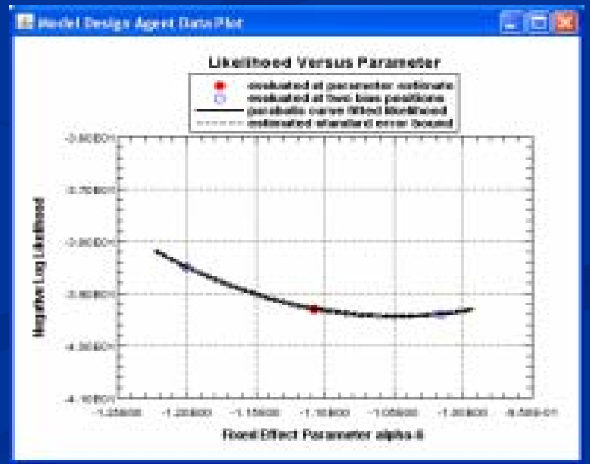




THETA(2)



OMEGA(1,2)

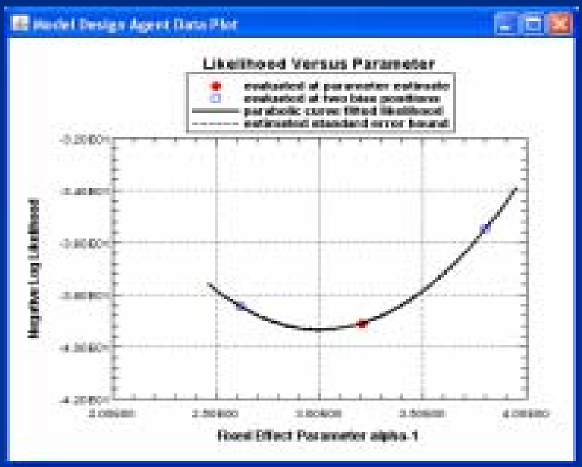


SIGMA(1)

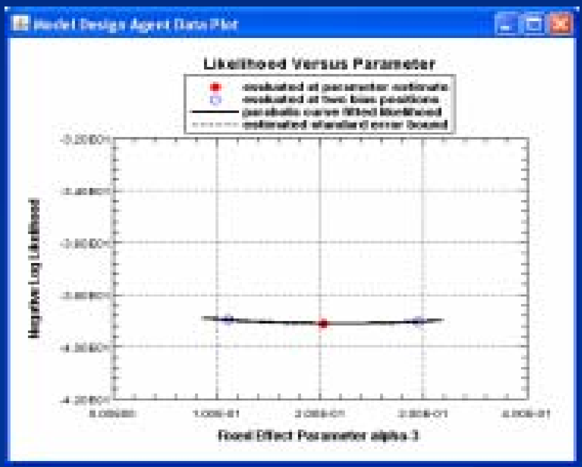
Cadralazine Example - ITS

-LL = -39.12±0.03

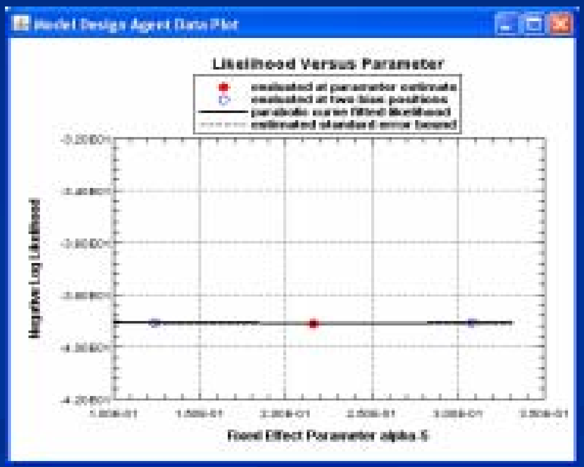
THETA(1)

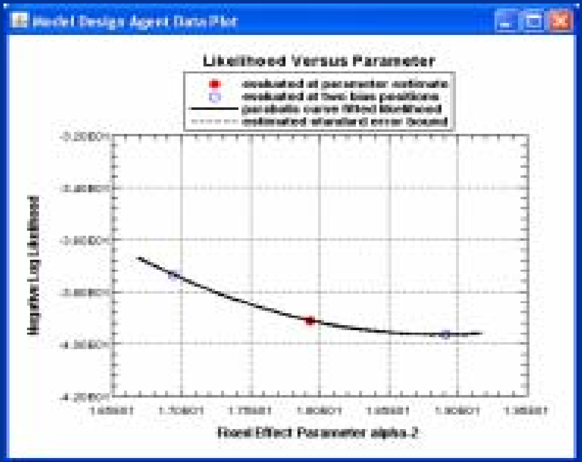


OMEGA(1,1)

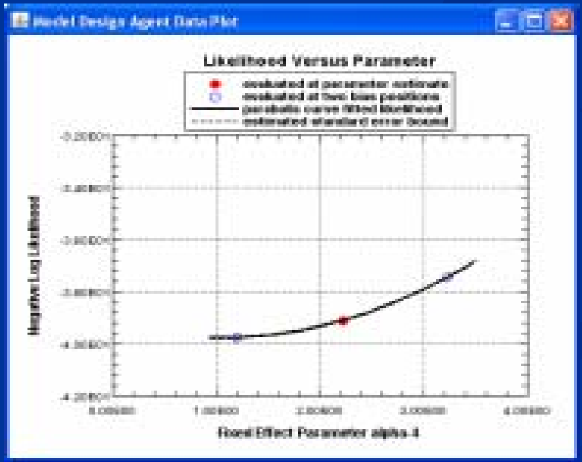


OMEGA(2,2)

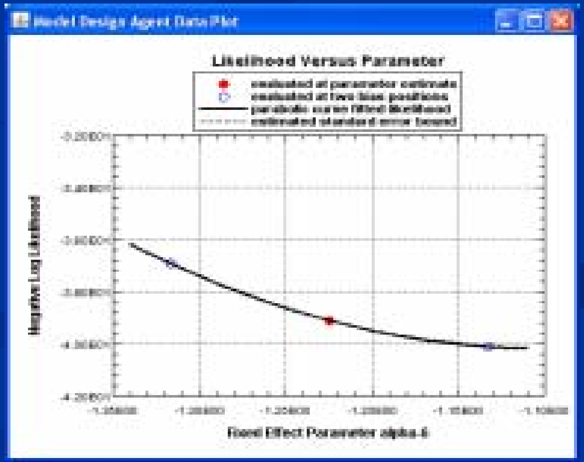




THETA(2)



OMEGA(1,2)

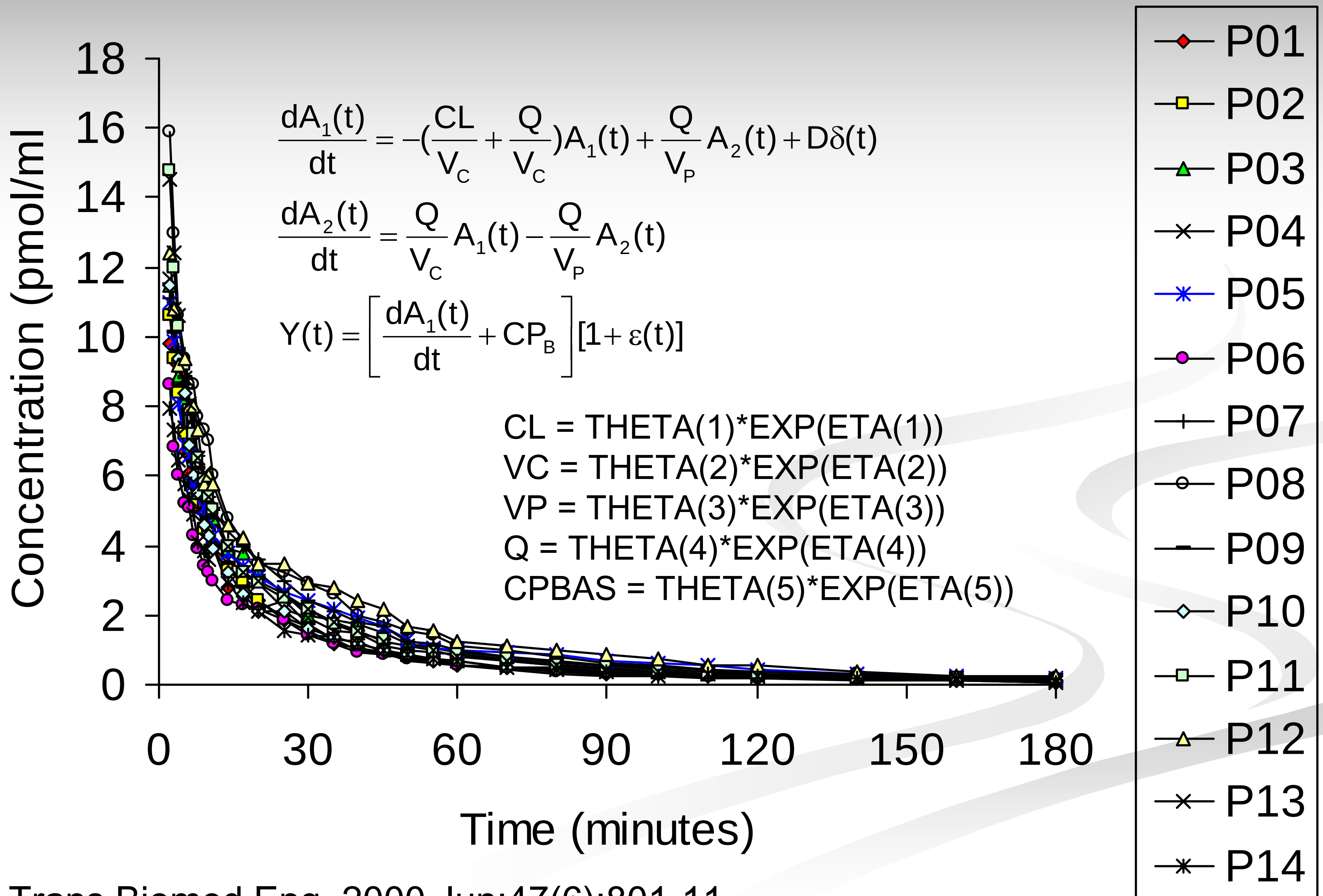


SIGMA(1)

Preliminary Observations

- The profiling informs the user about the location of the estimate on the overall true likelihood landscape
- Likelihood values are affected by randomness due to the Monte Carlo technique
- Two-Stage methods, when applicable, may not be too far off the mark
- As it is well known, the FO method does not provide good estimates when variability is substantial
- Next example: small BSV, intensive sampling

C-Peptide Data (N=14)



IEEE Trans Biomed Eng. 2000 Jun;47(6):801-11

Parameter Estimates (Base)

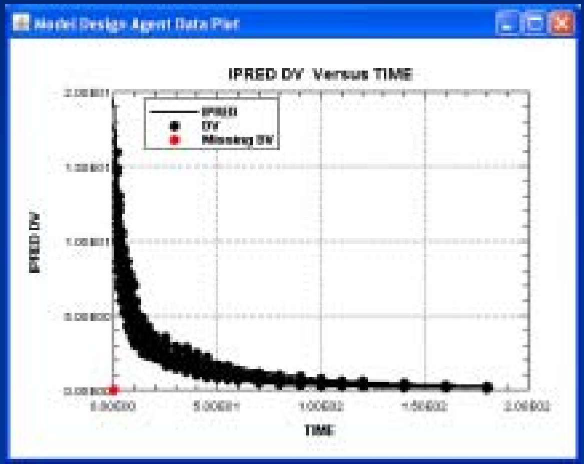
- First Order
  - THETA(1) = 234
  - THETA(2) = 3900
  - THETA(3) = 247
  - THETA(4) = 3550
  - THETA(5) = 0.114
  - OMEGA(1,1) = 0.0431
  - OMEGA(2,2) = 0.0519
  - OMEGA(3,3) = 0.0290
  - OMEGA(4,4) = 0.0285
  - OMEGA(5,5) = 0.0235
  - SIGMA(1,1) = 0.00503
- Expected Hessian
  - THETA(1) = 250
  - THETA(2) = 3610
  - THETA(3) = 228
  - THETA(4) = 3460
  - THETA(5) = 0.102
  - OMEGA(1,1) = 0.0455
  - OMEGA(2,2) = 0.0400
  - OMEGA(3,3) = 0.0376
  - OMEGA(4,4) = 0.0340
  - OMEGA(5,5) = 0.0431
  - SIGMA(1,1) = 0.00531

8703

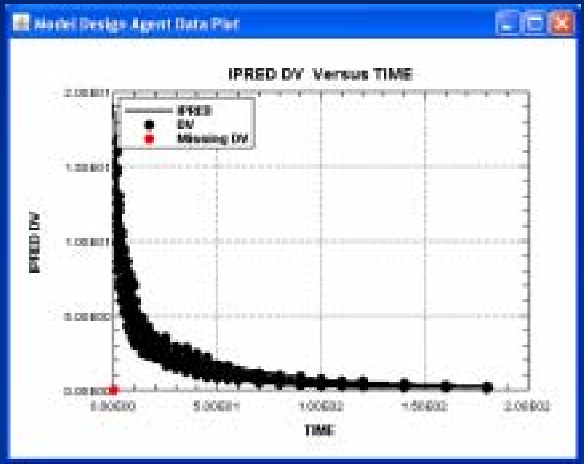
8704

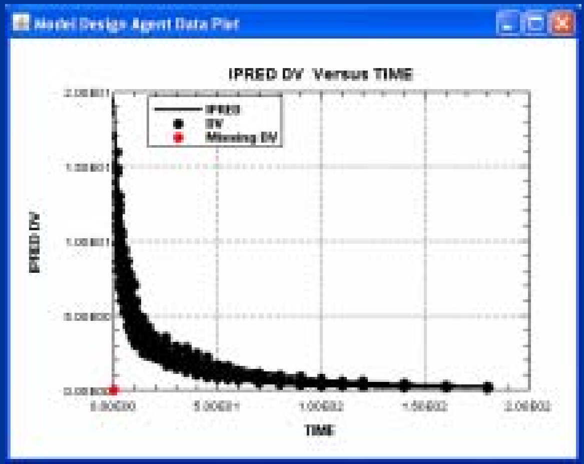
C-Peptide Example - IPRED

Iterative Two Stage

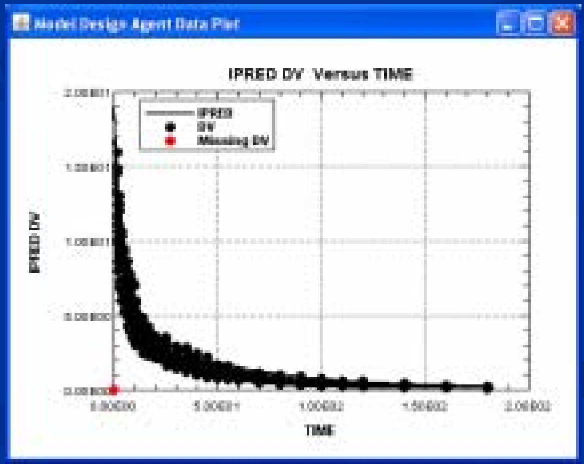


First Order





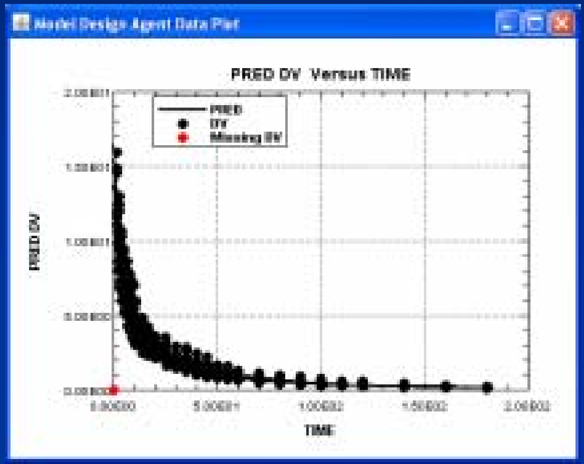
Standard Two Stage



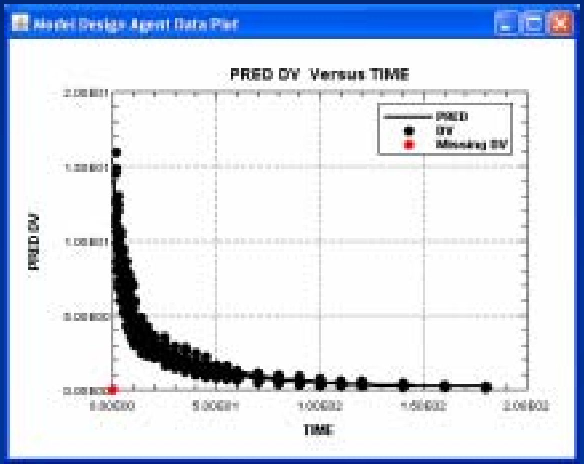
Expected Hessian

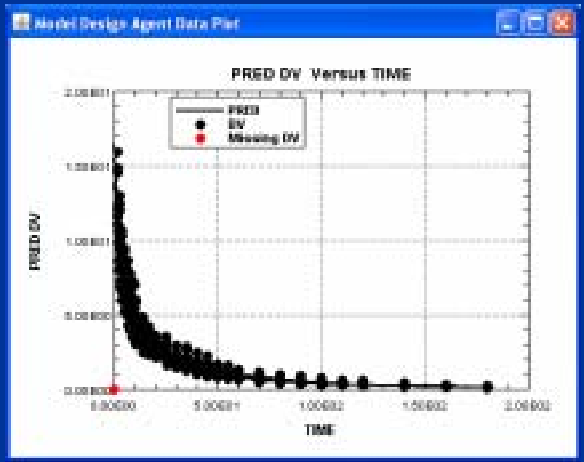
C-Peptide Example - PRED

Iterative Two Stage

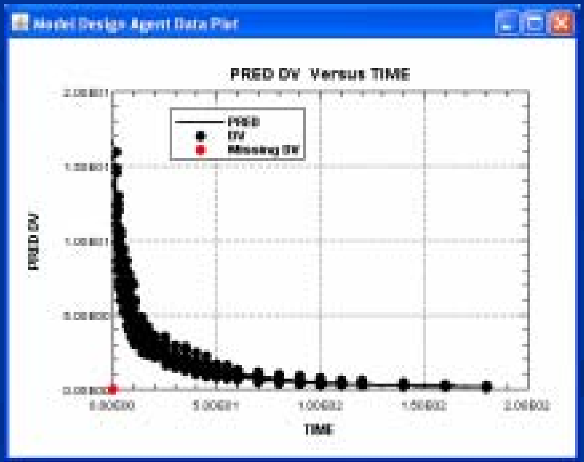


First Order





Standard Two Stage

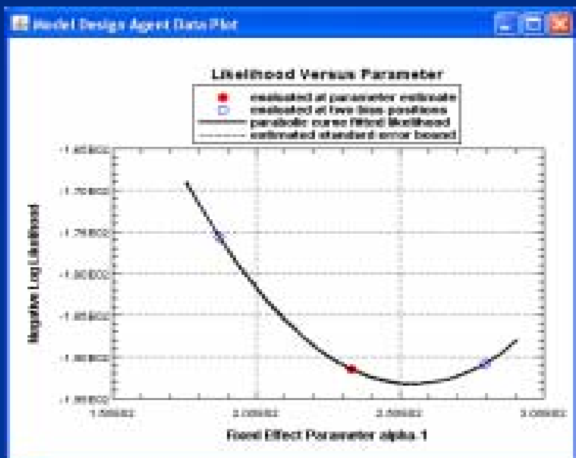


Expected Hessian

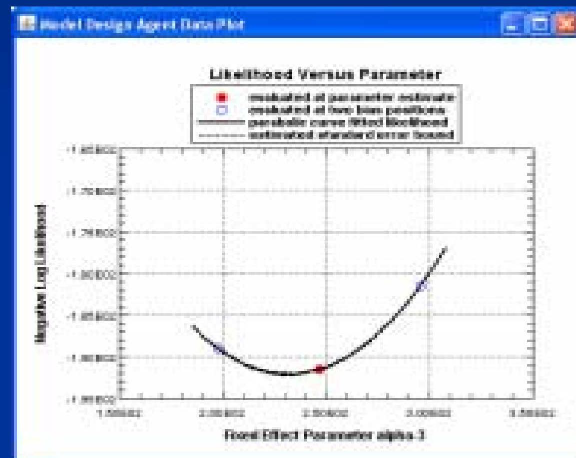
C-Peptide Example – FO (θ)

OF = -173.45, -LL = -191.60±0.13

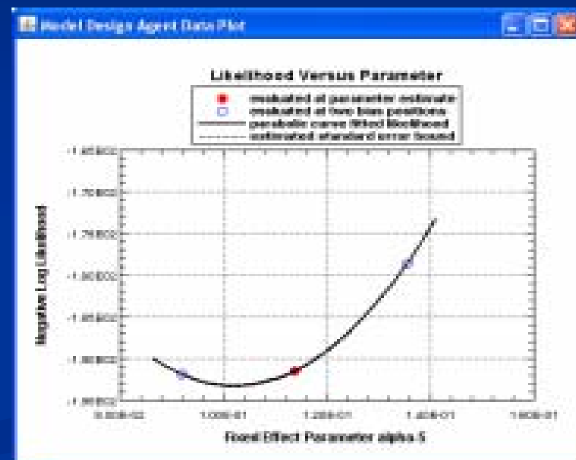
THETA(1)

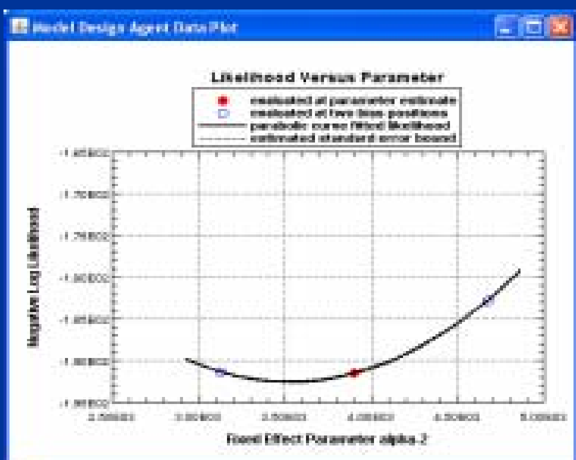


THETA(3)

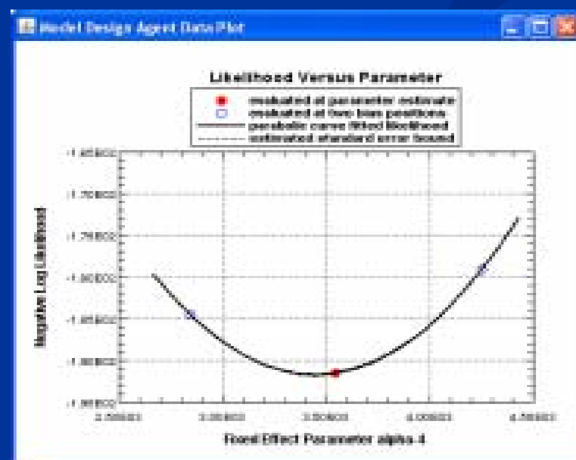


THETA(5)

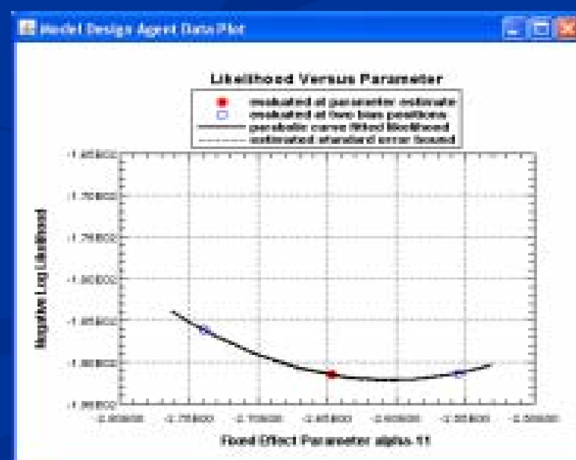




THETA(2)



THETA(4)

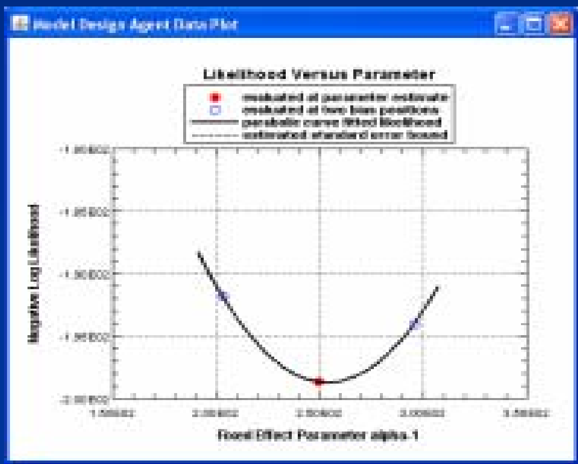


SIGMA(1)

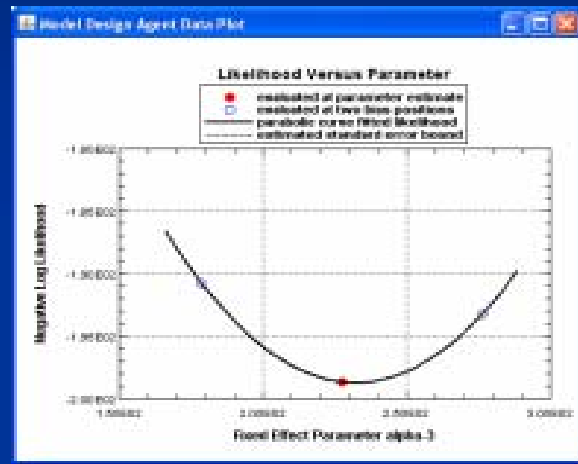
C-Peptide Example – EH (θ)

OF = -196.46, -LL = -198.72±0.02

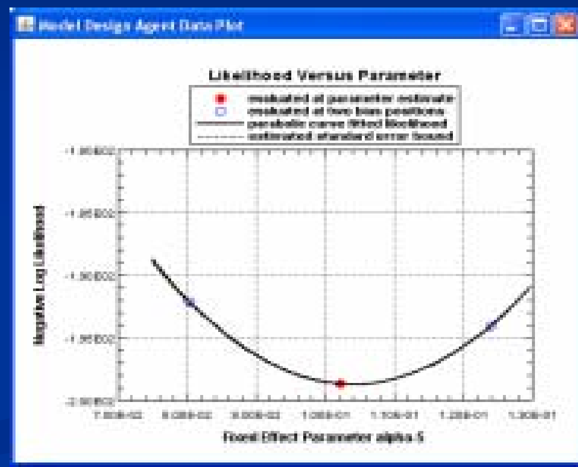
THETA(1)

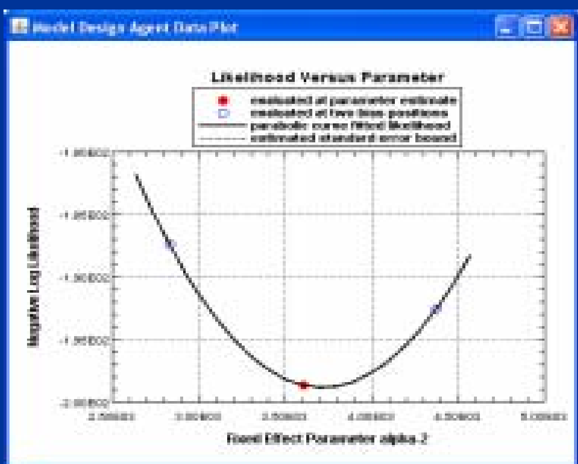


THETA(3)

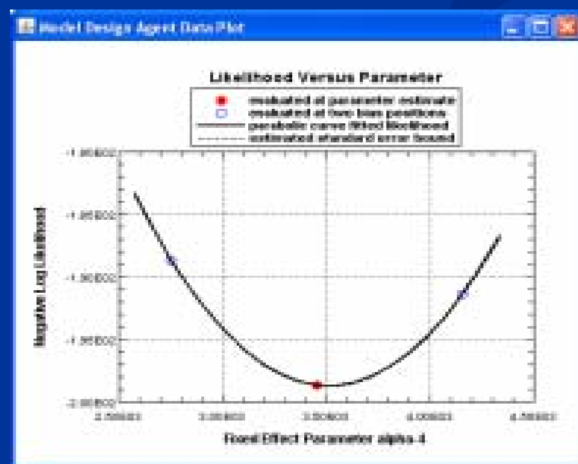


THETA(5)

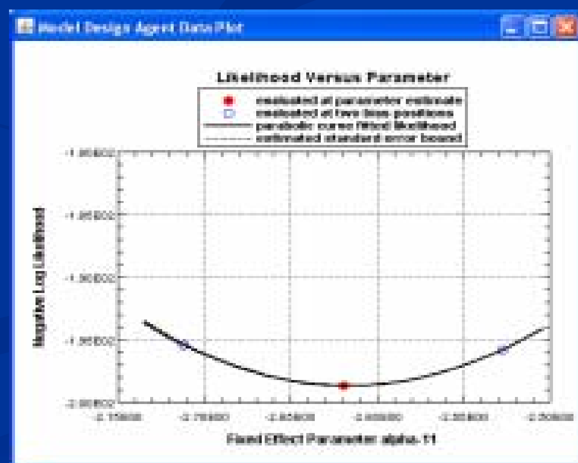




THETA(2)



THETA(4)



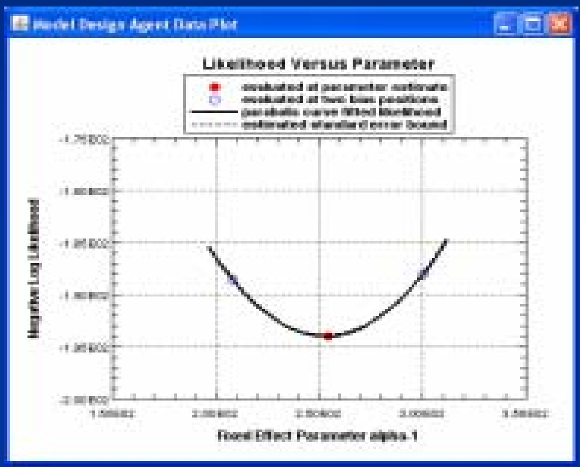
SIGMA(1)

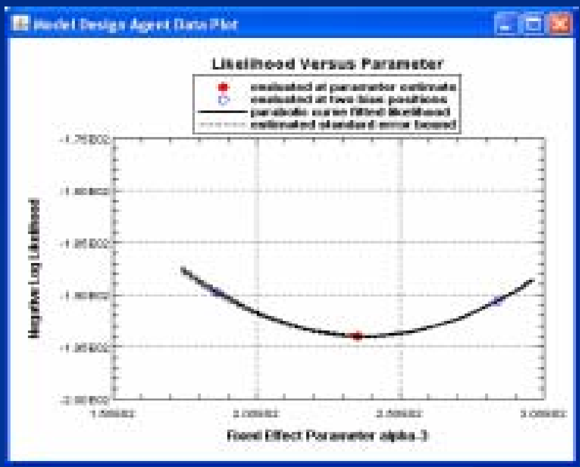
8862 (Vegas)

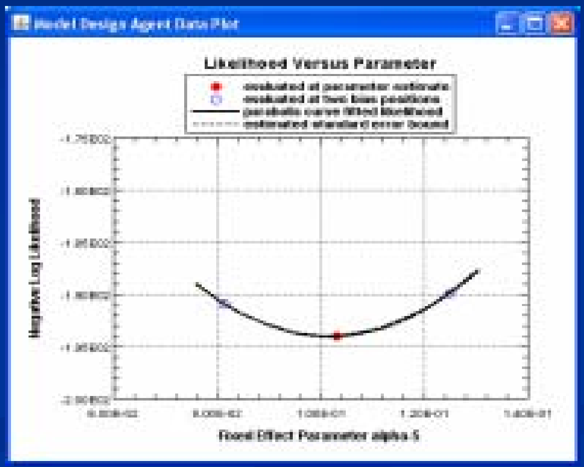
C-Peptide Example – STS (θ)

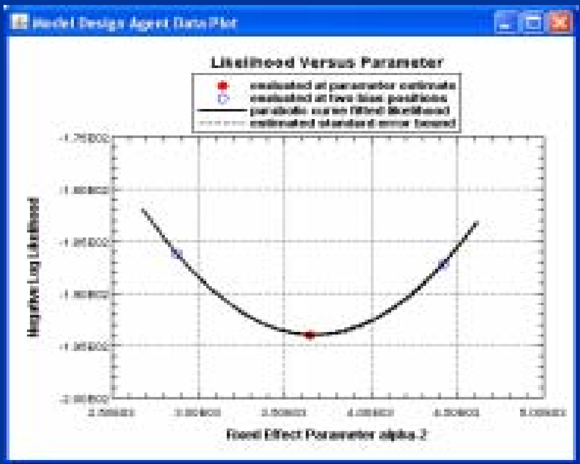
$-LL = -194.02 \pm 0.11$

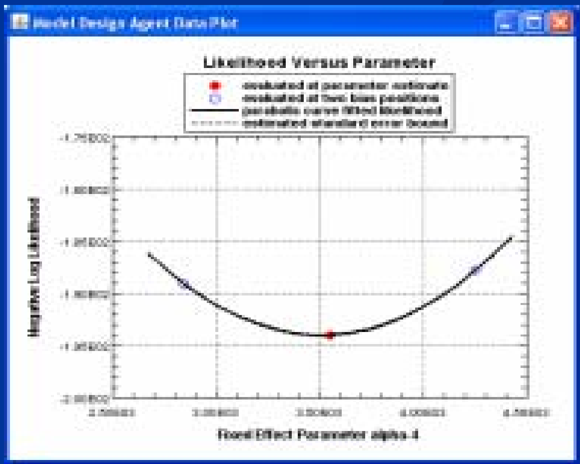
THETA(1) THETA(3) THETA(5)

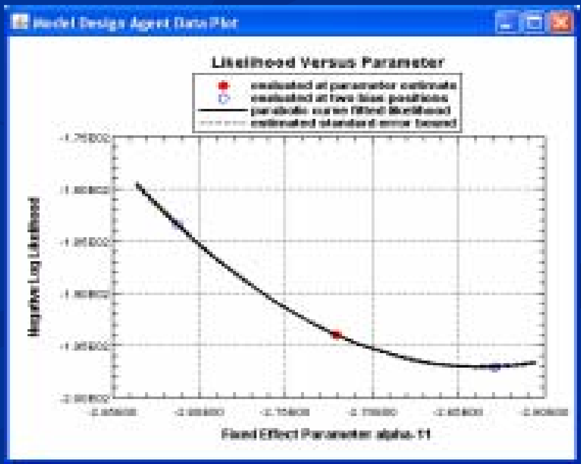










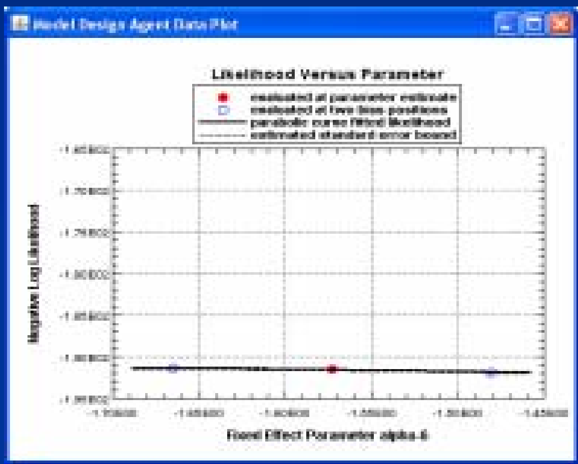


THETA(2) THETA(4) SIGMA(1)

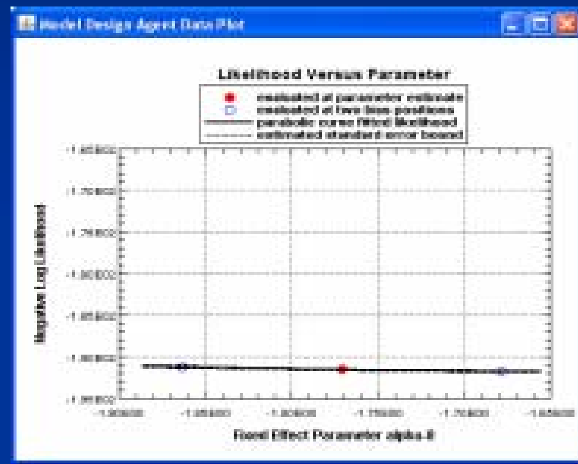
C-Peptide Example - FO (ω)

OF = -173.45, -LL = -191.60±0.13

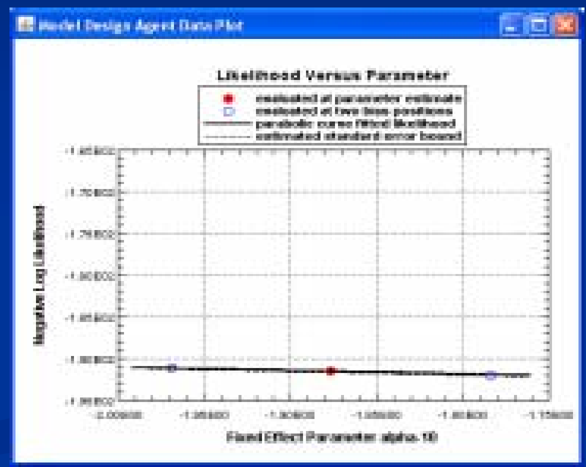
OMEGA(1,1)

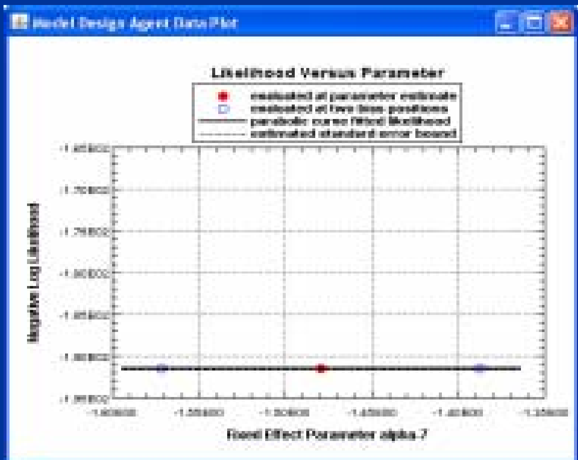


OMEGA(3,3)

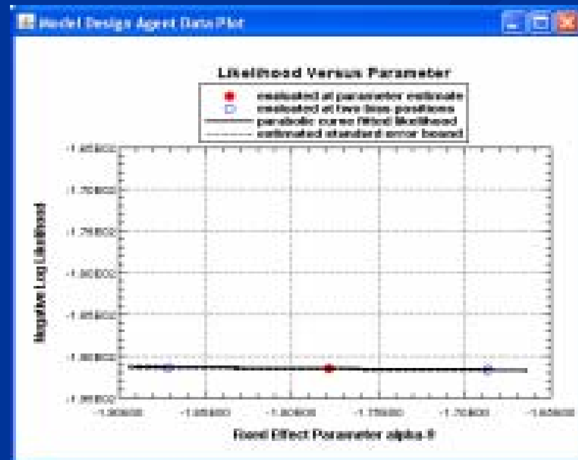


OMEGA(5,5)

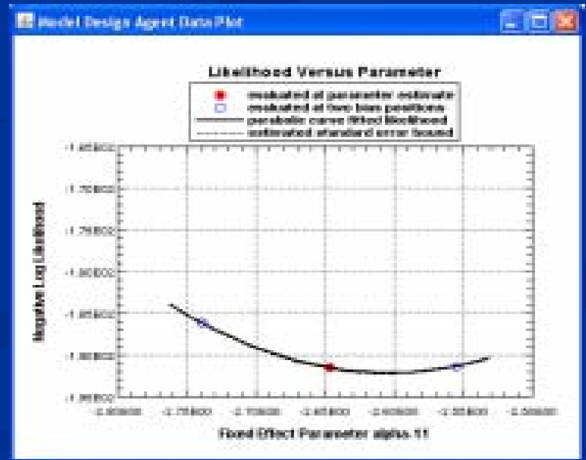




OMEGA(2,2)



OMEGA(4,4)



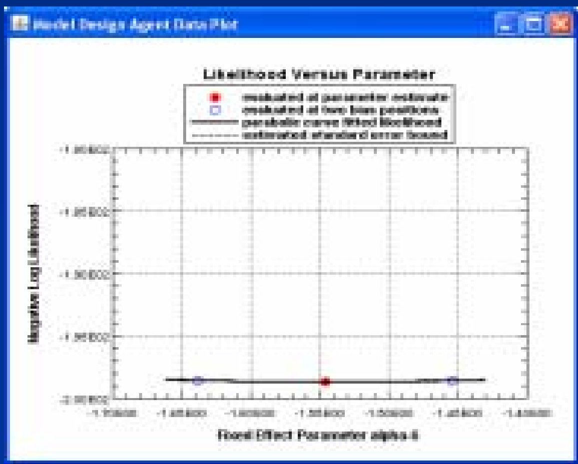
SIGMA(1)

8721 (Vegas)

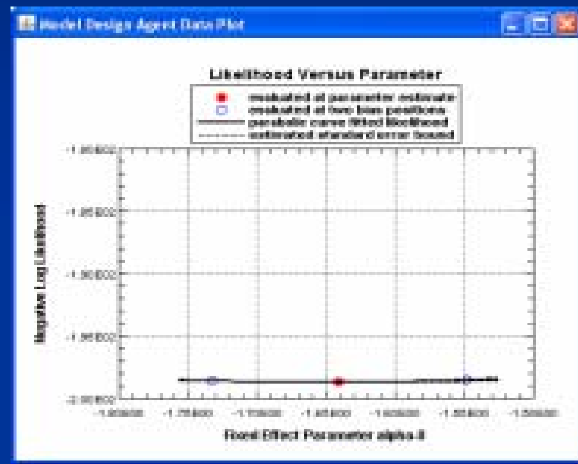
C-Peptide Example - EH (ω)

OF = -196.46, -LL = -198.72±0.02

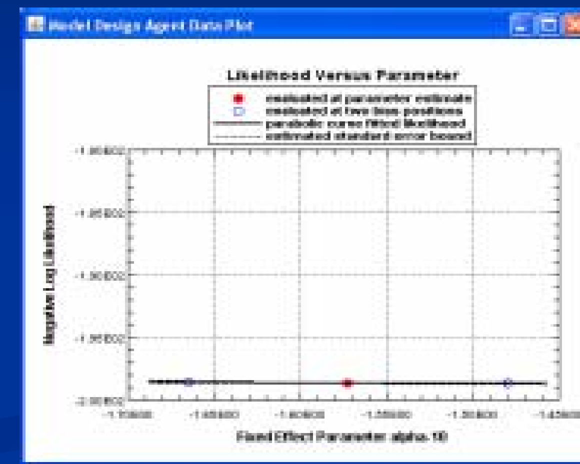
OMEGA(1,1)

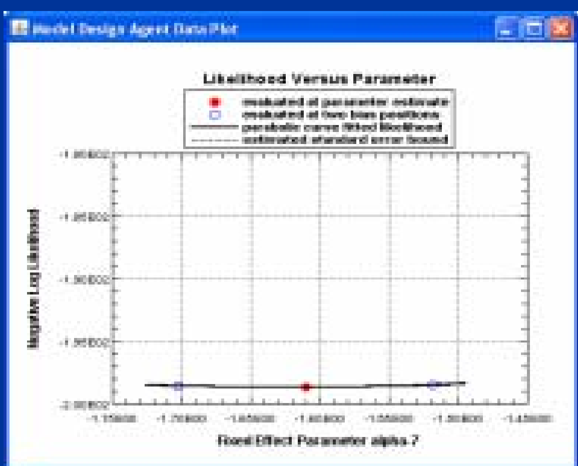


OMEGA(3,3)

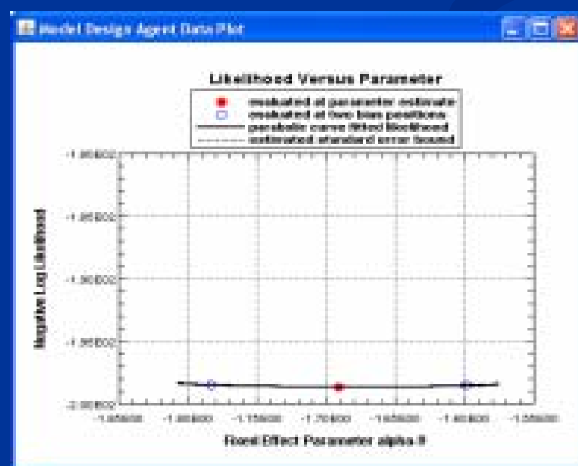


OMEGA(5,5)

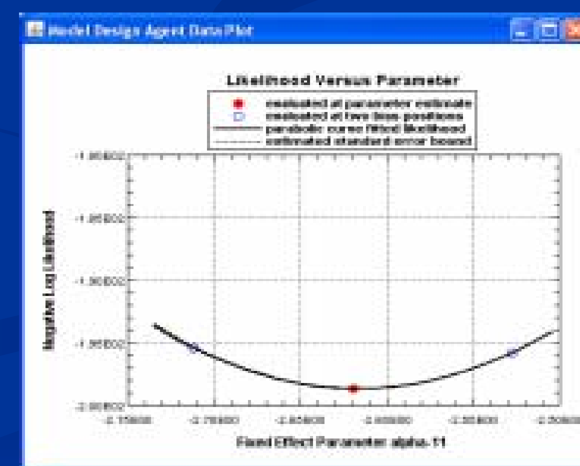




OMEGA(2,2)



OMEGA(4,4)

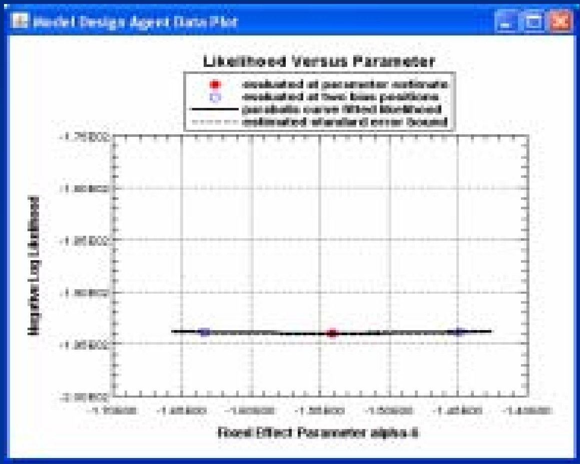


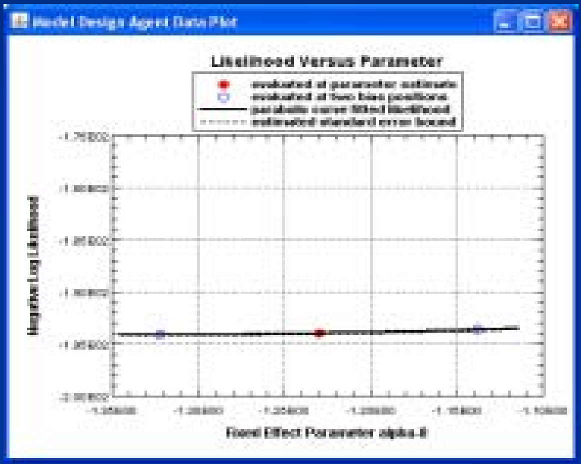
SIGMA(1)

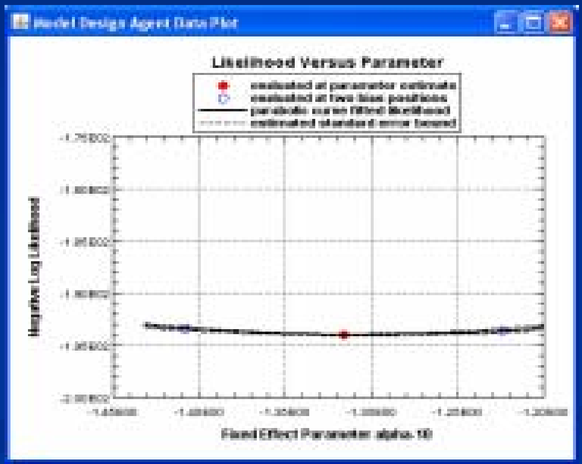
C-Peptide Example - STS (ω)

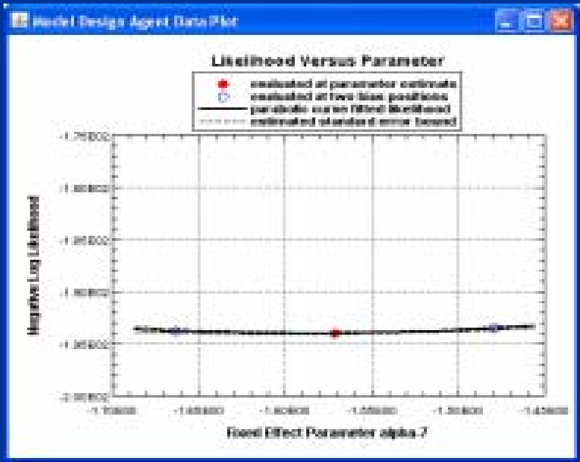
-LL = -194.02±0.11

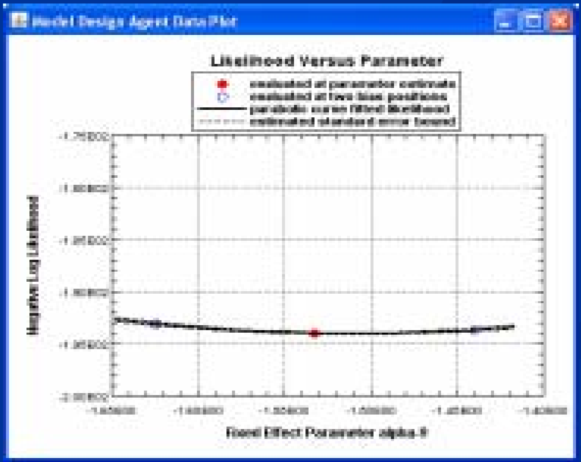
OMEGA(1,1) OMEGA(3,3) OMEGA(5,5)

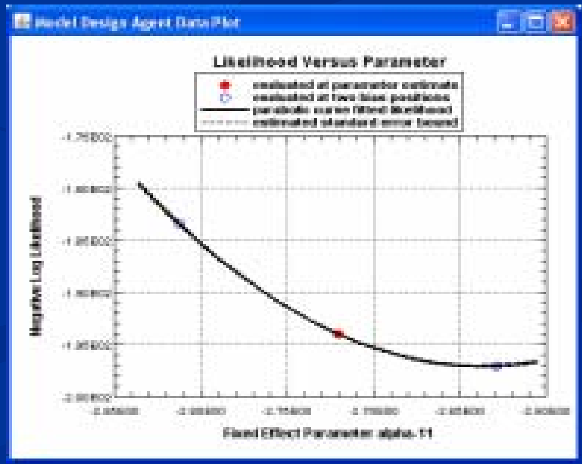






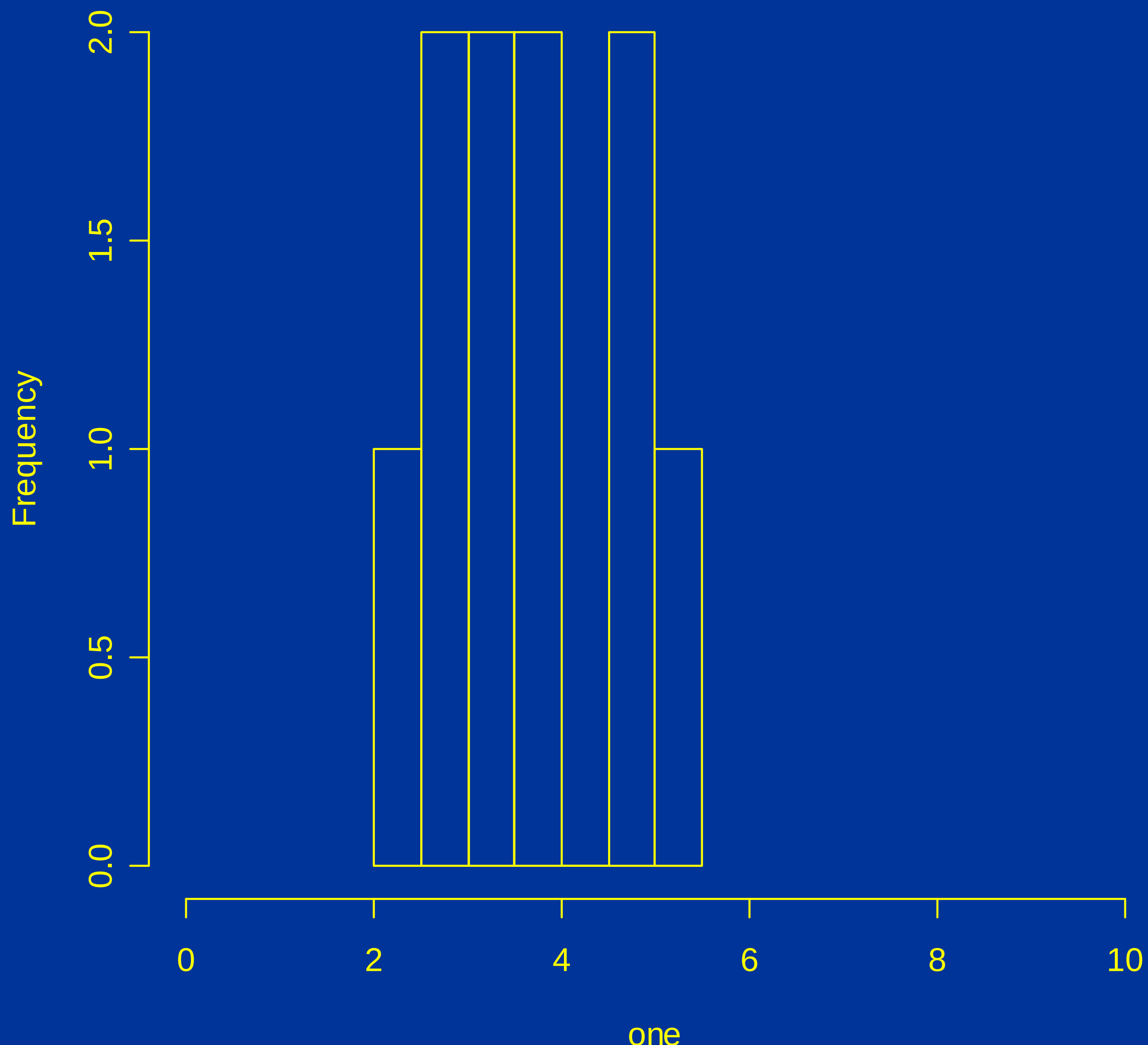




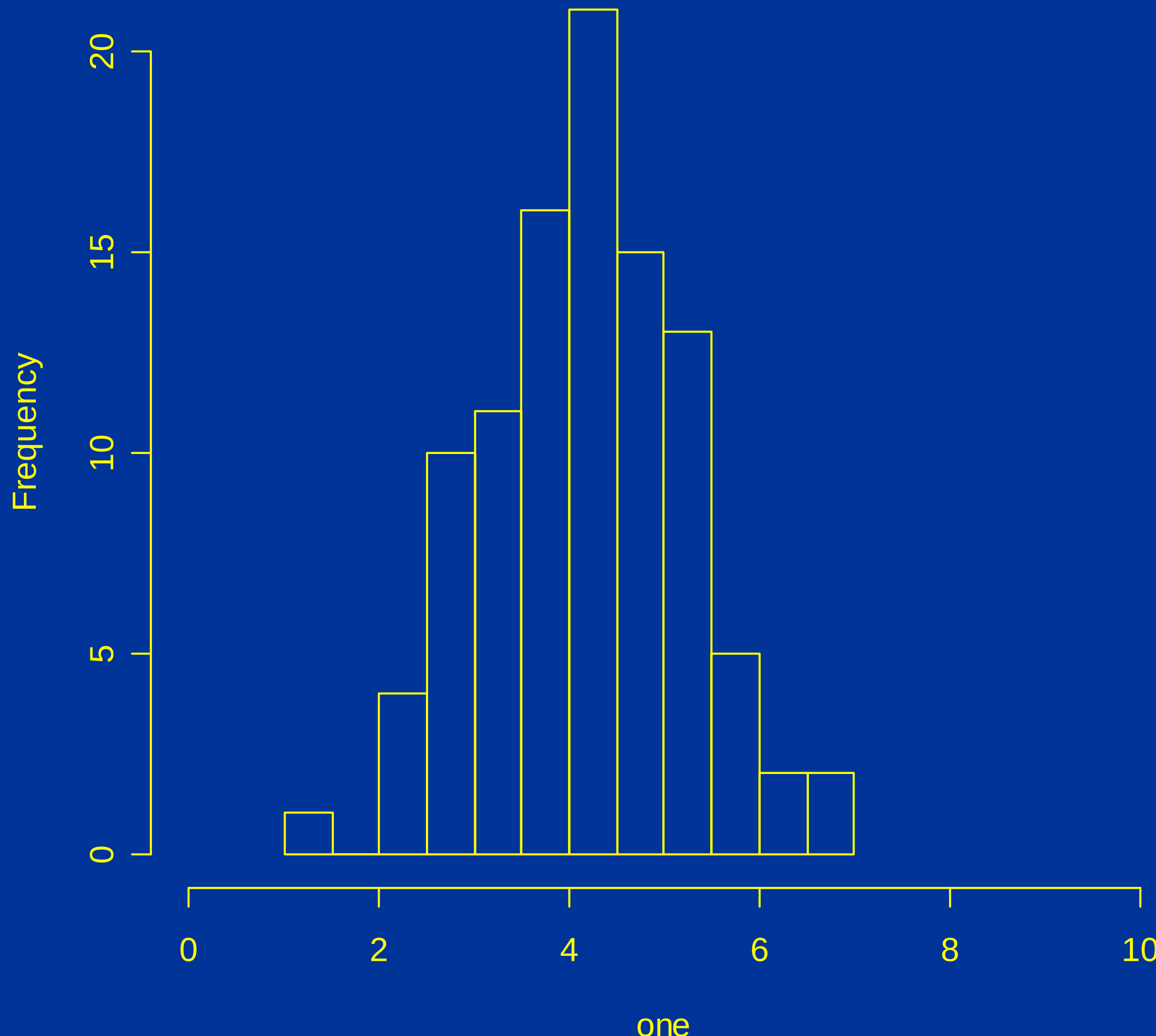


OMEGA(2,2) OMEGA(4,4) SIGMA(1)

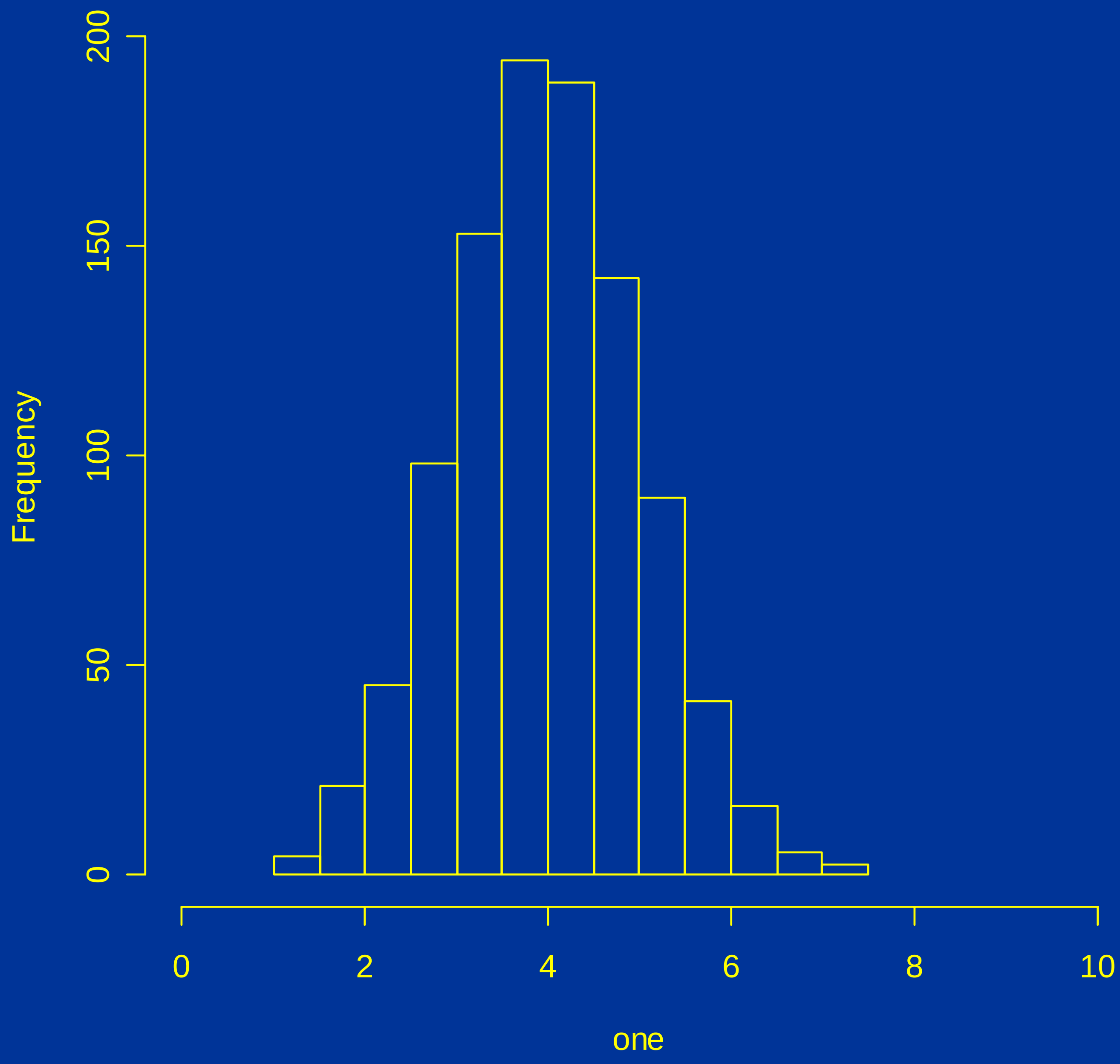
Preliminary Observations



Preliminary Observations



Preliminary Observations



Preliminary Observations

- In presence of small BSV and intensive sampling, various linearization approaches provide similar results
- However: the true likelihood profile for the BSV covariance matrix elements is rather flat (this is hardly surprising given the small number of subjects)

Conclusions and Future Work

- Conclusions
  - We have presented an approach for post-optimality check of approximate nonlinear mixed effects models
  - The approach is applicable to any model and any parameter set
  - The computational burden varies according to the model complexity and the required precision for the likelihood

Conclusions and Future Work

- Future Work
  - Evaluation by simulation
  - The availability of the exact likelihood allows to calculate, in principle both
    - Bias (related to gradient), and
    - Precision (related to Hessian)

    of the maximum likelihood (or any other) solution
  - Hybrid methods (deterministic and stochastic optimization) may be used to rapidly achieve robust solutions to the original ML problem

Acknowledgments

- Former and current members of the RFPK Software Team
  - Bradley Bell
  - James Burke
  - Jiaji Du
  - Andrew Ernst
  - Robert Goddard
  - Sachiko Honda
  - David Salinger
  - Mitch Watrous
  - Alan Westhagen
- RFPK Collaborators and Advisory Committee
- RFPK Graduate Students
- UW Department of Bioengineering
- NIH/NIBIB Grant P41 EB001975