Abstract

In this review we discuss methodology to ascertain the amount of information in given data sets with respect to determination of model parameters with desired levels of uncertainty. We do this in the context of least squares (ordinary, weighted, iterative reweighted weighted or "generalized", etc.) based inverse problem formulations. The ideas are illustrated with several examples of interest in the biological and environmental sciences.

Keywords: Information content, weighted least squares, sensitivity, model comparison techniques, Akaike Information Criterion

AMS classification: 34A55, 45Q05, 90C31, 92B05
1 Introduction

One of the most prevalent and persistent occurrences in inverse problem practice is the “overfitting” of data. That is, often, in eagerness to validate a given model or suites of models, one produces fits with models with increasingly larger numbers of estimated parameters without further analysis. This was accepted practice by numerous investigators in earlier years of inverse problem efforts—see [19, 20, 21, 23] to cite just a few. This is often done with little regard to any efforts on the level of confidence one might place in the parameter estimates. A slight variation to this procedure is the penchant to use increasingly sophisticated models with more parameters/mechanisms in attempts to obtain better fits to a given data set. A slightly different scenario arises when one adds additional state variables to increase goodness of fit. More recently, efforts on uncertainty quantification [4, 17, 32, 33, 46, 56, 58, 63] have led to increased expectations among inverse problems investigators and users. We review here methodologies for use in determining one or more such modeling shortcomings made in these earlier (and unfortunately continuing) contexts and illustrate the resulting analysis with specific examples.

In this review we turn to a fundamental question: how much information with respect to model validation can be expected in a given data set or collection of data sets? Our interest in this topic was stimulated by a concrete example involving previous HIV models [1, 11] with 15 or more parameters to estimate. Using recently developed parameter selectivity tools [8] based on parameter sensitivity-based scores, we found in [7] that a number of these parameters could not be estimated with any degree of confidence. Moreover, we discovered that quantifiable uncertainty levels vary among patients dependent upon the number of treatment interruptions (perturbations of therapy) that a patient had experienced. (While this is not very useful to our physician colleagues with respect to therapy design, it does provide scientific understanding that could be rather succinctly stated: the more dynamic changes represented in the data set, the more “information” in that particular data set).

Our interest was also motivated by our efforts in [24] involving large models (38 state variables and more than 100 parameters). As such mathematical models of interest in applications become more complex with numerous states, increasing numbers of parameters need to be estimated using experimental data. These problems necessitate critical analysis in model validation related to the reliability of parameter estimates obtained in model fitting.

In our discussions here we illustrate the use of several tools for interrogation of data sets with respect to their usefulness in estimation of parameters in complex models. Among these are parameter sensitivity theory, asymptotic theories (as the data sample size increases without bound) of standard error quantification using appropriate statistical models and the Fisher Information Matrix (FIM), bootstrapping (repeated sampling of synthetic data similar to the original data), and statistical (analysis-of-variance type) model comparison techniques as well as theoretical information criteria (AIC, etc.)—some of these are discussed in some detail in the recent monograph [17] as well as in numerous
statistical texts such as [32]. Such techniques can be employed in order to
determine the relative information content in data sets. We pursue this in the
context of recent models [13, 52] for nucleated polymerization in proteins as well
as models to describe decay in size histograms for aggregates in amyloid fibril
formulation [6]. Other examples concern efforts of interest to entomologists,
and include the growth dynamics of the algae as this is a major food source for
_Daphnia magna_, as well as efforts of scientists, pesticide control advisors (PCA),
and farmers to track pests in cotton fields.

Before addressing our main task of determining the information content in a
given data set, we summarize some background material on useful mathematical
and statistical techniques.

2 Standard Errors, Asymptotic Analysis and Bootstrapping

2.1 Estimation Using Weighted Least Squares (IRWLS)

We summarize the asymptotic theory (as the number of observations \( n \to \infty \))
related to parameter uncertainty based on the Fisher Information Matrix for
calculation of standard errors as detailed in [17, 28, 35, 39] and the references
therein. In the case of Generalized Least Squares (GLS), (or more precisely,
as used here Iterative Reweighted Weighted Least Squares (IRWLS)), the
associated standard errors for the estimated parameters \( \hat{\theta} \) (vector length \( \kappa_{\theta} \)) are
given by the following construction (for details see Chap. 3.2.5 and 3.2.6 of
[17]). We consider inverse or parameter estimation problems in the context of a
parameterized (with vector parameter \( q \in \Omega^{\kappa_q} \subset \mathbb{R}^{\kappa_q} \)) \( m \)-dimensional vector
dynamical system or mathematical model given by

\[
\begin{align*}
\frac{dx}{dt}(t) &= g(t, x(t), q), \\
    x(t_0) &= x_0,
\end{align*}
\]

with scalar observation process

\[
f(t; \theta) = C x(t; \theta),
\]

where \( \theta = (q^T, x_0^T)^T \in \Omega^{\kappa_\theta} \subset \mathbb{R}^{\kappa_q + \tilde{m}} = \mathbb{R}^{\kappa_\theta}, \tilde{m} \leq m, \) and the observation
operator \( C \) maps \( \mathbb{R}^m \) to \( \mathbb{R}^1 \). In some of the discussions below we assume without
loss of generality that some subset \( \tilde{x}_0 \) of \( \tilde{m} \leq m \) of the initial values \( x_0 \) are
also unknown and must be estimated. The sets \( \Omega^{\kappa_\theta} \) and \( \Omega^{\kappa_\theta} \) are assumed
known restraint sets for the parameters. Moreover, our data corresponds to
observations at points \( \{ t_i \}_{i=1}^n \) in the compact interval \([0, T]\). The observations
themselves from (3) are corrupted by nontrivial random process observation
error processes \( E_i \).

We make some standard statistical assumptions (see [17, 28, 39, 55]) under-
lying our inverse problem formulations.
A1) Assume \( \mathcal{E}_i \) are independent identically distributed i.i.d. with \( \mathbb{E}(\mathcal{E}_i) = 0 \) and \( \text{Var}(\mathcal{E}_i) = \sigma_0^2 \), where \( i = 1, \ldots, n \) and \( n \) is the number of observations or data points in the given data set.

A2) We assume that there exists a true or nominal set of parameters \( \theta_0 \in \Omega \equiv \Omega^{\kappa \theta} \).

A3) \( \Omega \) is a compact subset of Euclidian space of \( \mathbb{R}^{\kappa \theta} \).

A4) The observation function \( f(t, \theta) \) is continuous in \( t \) and \( C^2 \) in \( \theta \) on \( [0, T] \times \Omega \).

Denote by \( \hat{\theta} \) the estimated parameter for \( \theta_0 \in \Omega \). The inverse problem is based on statistical assumptions on the observation error in the data. If one assumes some type of absolute or generalized relative error data model, then the error is proportional in some sense to the measured observation. This can be represented by a statistical model \( Y_i \) with proportional errors of the form \( f(t_i; \theta_0)^\gamma \mathcal{E}_i \) (note \( \gamma = 0 \) corresponds to absolute error and corresponding Ordinary Least Squares (OLS) which will also be considered below); that is,

\[
Y_i = f(t_i; \theta_0) + f(t_i; \theta_0)^\gamma \mathcal{E}_i, \quad \gamma \in [0, 1], \tag{4}
\]

with corresponding realizations

\[
y_i = f(t_i; \theta_0) + f(t_i; \theta_0)^\gamma \epsilon_i, \quad \gamma \in [0, 1], \tag{5}
\]

where the \( \epsilon_i \) are realizations of the \( \mathcal{E}_i, \ i = 1, \ldots, n \). We chose \( \gamma \in [0, 1] \) to illustrate the ideas, but larger values of \( \gamma \) may also be appropriate for some data sets. For an example where \( \gamma > 1 \) may be appropriate, see [7].

For relative error models one could use inverse problem formulations with Generalized Least Squares (GLS) cost functional

\[
J^n(Y; \theta) = \sum_{i=1}^{n} \left( \frac{Y_i - f(t_i; \theta)}{f(t_i; \theta)^\gamma} \right)^2. \tag{6}
\]

The corresponding estimators would be defined by

\[
\hat{\theta} = \arg\min_{\theta \in \Omega} \sum_{i=1}^{n} \left( \frac{y_i - f(t_i; \theta)}{f(t_i; \theta)^\gamma} \right)^2, \quad \gamma \in (0, 1], \tag{7}
\]

with realizations

\[
\hat{\theta} = \arg\min_{\theta \in \Omega} \sum_{i=1}^{n} \left( \frac{y_i - f(t_i; \theta)}{f(t_i; \theta)^\gamma} \right)^2, \quad \gamma \in (0, 1]. \tag{8}
\]

However, we actually use an iterative form of a weighted least squares procedure which solves for the weights \( w_i(\theta) = f(t_i; \theta) \) in an iterative manner (described in the next section) so that at each stage an iterative weighted least squares

\[
J^n(Y; \theta) = \sum_{i=1}^{n} w_i(\theta)^{-2 \gamma} (Y_i - f(t_i; \theta))^2 \tag{9}
\]
is solved. We shall denote this iterative solution throughout our subsequent discussions as

$$\Theta = \widehat{\arg \min_{\theta \in \Omega}} \sum_{i=1}^{n} \left( \frac{Y_i - f(t_i; \theta)}{f(t_i; \theta)^\gamma} \right)^2, \quad \gamma \in (0, 1], \quad (10)$$

to remind readers that this is NOT the same as the minimization process in (7) or (8).

### 2.2 Implementation of the IRWLS Procedure

We note that an estimate $\theta$ can be solved either directly according to (8) (which is not an easy minimization problem and is seldom used in practice!) or iteratively using an iterative algorithm. This iterative procedure as described in [35, 39] (often referred to as the “GLS algorithm” although in the version presented and used here, is more properly called the “IRWLS algorithm”) is summarized below:

1. Solve for the initial estimate $\hat{\theta}^{(0)}$ obtained using the OLS minimization (7) with $\gamma = 0$. Set $l = 0$.
2. Form the weights $\hat{w}_j = f^{-2\gamma}(t_j; \hat{\theta}^{(l)})$.
3. Re-estimate $\hat{\theta}$ by minimizing

$$\sum_{j=1}^{n} \hat{w}_j (y_j - f(t_j; \theta))^2,$$

over $\theta \in \Omega$ to obtain $\hat{\theta}^{(l+1)}$.
4. Set $l = l+1$ and return to step 2. Terminate the process and set $\hat{\theta} = \hat{\theta}^{(l+1)}$ when two of the successive estimates are sufficiently close.

We note that the above iterative procedure was formulated by the equivalent of minimizing for a given $\tilde{\theta}$ and then updating the weights $w_j = f^{-2\gamma}(t_j; \tilde{\theta})$ after each iteration. One would hope that after a sufficient number of iterations $\hat{w}_j$ would converge to $f^{-2\gamma}(t_j; \tilde{\theta})$. Further discussions of these issues can be found in a number of statistical texts including [35, 39].

### 2.3 Asymptotic Theory for Weighted Least Squares

For the general weighted least squares formulations, we may define the standard errors by the formula

$$SE_k = \sqrt{\Sigma_{kk}(\tilde{\theta})}, \quad k = 1, ..., \kappa_{\theta}.$$
where the covariance matrix $\Sigma$ is given by
\[
\Sigma(\hat{\theta}) = \hat{\sigma}^2(\chi^T(\hat{\theta})W(\hat{\theta})\chi(\hat{\theta}))^{-1} = \hat{\sigma}^2 F^{-1}.
\]
Here $F = (\chi^T(\hat{\theta})W(\hat{\theta})\chi(\hat{\theta}))$ is the Fisher Information Matrix defined in terms of the sensitivity matrix
\[
\chi = \frac{\partial f}{\partial \theta} = \left(\frac{\partial f(t_1; \hat{\theta})}{\partial \theta}, \ldots, \frac{\partial f(t_n; \hat{\theta})}{\partial \theta}\right)
\]
of size $n \times \kappa_\theta$ (recall $n$ is the number of data points and $\kappa_\theta$ is the number of estimated parameters) and $W$ is defined by
\[
W^{-1}(\hat{\theta}) = \text{diag}(f(t_1; \hat{\theta})^2, \ldots, f(t_n; \hat{\theta})^2).
\]
We use the approximation of the variance given by
\[
\sigma_0^2 \approx \hat{\sigma}^2(\hat{\theta})^2 = \frac{1}{n - \kappa_\theta} \sum_{i=1}^{n} \frac{1}{f(t_i; \hat{\theta})^2}(f(t_i; \hat{\theta}) - y_i)^2.
\]

2.4 Bootstrapping for IRWLS

In each of our inverse problems we may attempt to ascertain uncertainty bounds for the estimated parameters using both the asymptotic theory described above and a generalized least squares version of bootstrapping [35, 36, 38, 41, 43]. An outline of the appropriate bootstrapping algorithm is given next.

We suppose that we are given experimental data $(t_1, y_1), \ldots, (t_n, y_n)$ from the underlying observation process
\[
Y_i = f(t_i; \theta_0) + f(t_i; \theta_0)^\gamma \epsilon_i, \quad i = 1, \ldots, n, \quad (11)
\]
where the $\epsilon_i$ are again i.i.d. with mean zero and constant variance $\sigma_0^2$. Then we see that $\mathbb{E}(Y_i) = f(t_i; \theta_0)$ and $\text{Var}(Y_i) = \sigma_0^2 f(t_i; \theta_0)^{2\gamma}$, with associated corresponding realizations of $Y_i$ given by
\[
y_i = f(t_i; \theta_0) + f(t_i; \theta_0)^\gamma \epsilon_i.
\]

A standard algorithm can be used to compute the corresponding bootstrapping estimate $\hat{\theta}_{\text{boot}}$ of $\theta_0$ and its empirical distribution. We treat the general case for nonlinear dependence of the model output on the parameters $\theta$. The algorithm is given as follows.

1. First obtain the estimate $\hat{\theta}^0$ from the entire sample $\{y_i\}$ using the IRWLS given in (10) with $\gamma = 1$. An estimate $\hat{\theta}_{\text{boot}}$ can be solved for iteratively as follows.

2. Define the nonconstant-variance standardized residuals
\[
\hat{s}_i = \frac{y_i - f(t_i; \hat{\theta}^0)}{f(t_i; \hat{\theta}^0)^\gamma}, \quad i = 1, 2, \ldots, n.
\]
Set the counter $m = 0$. 

"
3. Create a bootstrapping sample of size \( n \) using random sampling with replacement from the data (realizations) \( \{s_1, \ldots, s_n\} \) to form a bootstrapping sample \( \{s_1^m, \ldots, s_n^m\} \).

4. Create bootstrapping sample points

\[
y_{i}^m = f(t_i; \hat{\theta}^0) + f(t_i; \hat{\theta}^0)^\gamma s_i^m,
\]

where \( i = 1, \ldots, n \).

5. Obtain a new estimate \( \hat{\theta}_{m+1} \) from the bootstrapping sample \( \{y_i^m\} \) using IRWLS.

6. Set \( m = m + 1 \) and repeat steps 3–5 until \( m \geq M \) where \( M \) is large (e.g., \( M=1000 \)).

We then calculate the mean, variance, standard error (SE), and confidence intervals using the formulae

\[
\hat{\theta}_{\text{boot}} = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}^m,
\]

\[
\text{Var}(\hat{\theta}_{\text{boot}}) = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}^m - \hat{\theta}_{\text{boot}})^T (\hat{\theta}^m - \hat{\theta}_{\text{boot}}),
\]

\[
\text{SE}_k(\hat{\theta}_{\text{boot}}) = \sqrt{\text{Var}(\hat{\theta}_{\text{boot}})_{kk}},
\]

where \( \hat{\theta}_{\text{boot}} \) denotes the bootstrapping estimate.

3 Example from a Nucleated-polymerization Model

3.1 Prions and Protein Polymerization

Prions are misfolded proteins associated with a variety of fatal, untreatable neurodegenerative disorders in mammals [3, 37, 48]. It is well-known that several neuro-degenerative disorders, including Alzheimers, Hunttingsons and Prion (e.g., mad cow), diseases, are related to aggregations of proteins presenting an abnormal folding. These protein aggregates, called amyloids have been the focus of numerous modeling efforts in recent years [34, 52, 64, 65, 66]. A major challenge in this field is to understand (at both qualitatively and quantitatively levels) the key aggregation mechanisms. We choose to illustrate our methodology on polyglutamine (PolyQ) containing proteins. This was also chosen to illustrate the fairly general ODE-PDE model proposed in [52]; the reason for our choice here and in [13] is that, as shown in [52], the polymerization mechanisms prove to be simpler for PolyQ aggregation than for other types of proteins [44].

The data sets (DS1-DS4) of interest to us here (from experiments carried out by Human Rezaei and his team at INRA (Virologie et Immunologie Moleculaires) and used in [52]), are depicted in Figure 1 above and record the evolution of normalized total polymerized mass in time. The total polymerized mass is measured by Thioflavine T (ThT) (which is one of the most common experimental
tools for \textit{in vitro} protein polymerization—see \cite{52,64}), and is presented in these graphs as the normalized or non-dimensionalized total mass.

In \cite{52} and subsequent efforts in \cite{12,13}, the authors sought to investigate several questions including (i) understanding the key polymerization mechanisms, (ii) numerical approximation of the models, and (iii) selection of parameters and calibration of the model. Here we focus on illustrating the methodology used in (iii).

### 3.2 An Infinite Dimensional ODE Model

We first outline a model that is given in \cite{13,52}. The main variables of interest are given, respectively, by concentrations of the normal proteins (denoted by $V$) known as \textit{monomers} (basic subunits that are repeated in a chainlike fashion), monomeric proteins exhibiting an abnormal configuration, denoted by $V^*$ and known as \textit{conformers}, and \textit{i-polymers} made up of $i$ aggregated abnormal proteins denoted here by $c_i$. The dynamics as modeled in \cite{52} are given by (see \cite{52} for further details)

1. A monomer-conformer exchange: $V \xrightleftharpoons[k^*_i]{k^*_i} V^*$. This models the spontaneous formation-dissociation of an active form of the monomer $V$, and conformer $V^*$, from the initially present inert form or monomer $V$. The inert form $V$ cannot react and form fibrils, whereas the active conformer $V^*$ may.
2. A nucleation reaction: \( V^* + V^* + \ldots + V^* \xrightarrow{k_{i_0}^{N}} k_{i_0}^N \) \( c_{i_0} \). This models the spontaneous formation of the smallest stable polymer, formed by the addition of a certain number \( i_0 \) of active conformers. This resulting smallest stable polymer is called a nucleus.

3. A polymerization by conformer addition: \( c_i + V^* \xrightarrow{k_{i_0}^{i}} c_{i+1} \). Once a nucleus is formed, its size may grow progressively by addition of active conformers.

As explained and justified in [52], other reactions like fragmentation and coalescence are negligible for the case of polyglutamine containing proteins. The law of mass action in the deterministic framework (see [5, 28, 54] and the numerous references therein), yields the ordinary differential equation for concentrations \([A], [B]\), etc. as \( \frac{d[A]}{dt} = -k_f^+ [A][B] + k_f^- [A'][B'] \). Using these basic ideas one can derive the infinite system of ordinary differential equations studied in [52] and given by

\[
\begin{align*}
\frac{dV}{dt} &= -k_f^+ V + k_f^- V^*, \\
\frac{dV^*}{dt} &= k_f^+ V - k_f^- V^* + i_0 k_{i_0}^N c_{i_0} - V^* \sum_{i \geq i_0} k_{i_0}^i c_i, \\
\frac{dc_{i_0}}{dt} &= k_{i_0}^N (V^*)^{i_0} - k_{i_0}^N c_{i_0} - k_{i_0}^{i_0} c_{i_0} V^*, \\
\frac{dc_i}{dt} &= V^* (k_{i_0}^{i-1} c_{i-1} - k_{i_0}^i c_i), \quad i = i_0 + 1, \ldots
\end{align*}
\]

with initial conditions

\[
V(0) = c_0, \quad V^*(0) = 0, \quad c_{i_0}(0) = c_i(0) = 0.
\]

A mass balance equation yields

\[
\frac{d}{dt} \left( V + V^* + \sum_{i=i_0}^{\infty} ic_i \right) = 0.
\]

As noted in [13], the experiments we considered measure the a-dimensional or non-dimensional total polymerized mass (abbreviated as Madim) which are our observables in the inverse problem formulation below and are given by

\[
f(t) \equiv \sum_{i \geq i_0} ic_i(t).
\]

Amyloid formations are characterized by very long polymers (a fibril may contain up to \(10^6\) monomer units). A PDE version of the standard model, where a continuous variable \(x\) approximates the discrete sizes \(i\) for sufficiently...
large values of \(i\), is thus a reasonable approximation for large amyloid polymers. However, for small polymer sizes this resulting continuum approximation does not work very well. Thus we take a “hybrid approach” of employing our ODE for smaller polymers and using a PDE for larger fibrils.

### 3.3 A Hybrid ODE/PDE Model

Following [12, 53], we define a small parameter \(\varepsilon = \frac{1}{i_M}\), and let \(x_i = i\varepsilon\) with \(i_M \gg 1\) being the average polymer size defined by

\[
i_M = \frac{\sum_{i > i_0} ic_i}{\sum c_i}.
\]

Let \(\chi_A\) be the characteristic function of the set \(A\), and define the dimensionless quantities

\[
c^\varepsilon(t, x) = \sum c_i \chi_{[x_i, x_{i+1}]}.
\]

We then may derive a hybrid ordinary-partial differential equation system to replace the infinite ODE system. A formal derivation for a full model, also including nucleation, is carried out in [52].

For a fixed integer \(N_0\) we obtain after some arguments [12, 13, 53] the system

\[
\begin{align*}
dV &= -k_1^+ V + k_1^- V^*, \\
V^* &= c_0 - V - \sum_{i=i_0}^{N_0} ic_i - \int_{N_0}^{\infty} xe^\varepsilon dx, \\
\frac{dc_{i_0}}{dt} &= k_{on}^N (V^*)^{i_0} - k_{off}^N c_{i_0} - k_{on}^{i_0} c_{i_0} V^*, \\
\frac{dc_i}{dt} &= V^*(k_{on}^{i-1} c_{i-1} - k_{on}^i c_i), \quad i \leq N_0, \\
\partial_x c^\varepsilon(x, t) &= -V^* \partial_x (k_{on}(x)c^\varepsilon(x, t)), \quad x \geq N_0,
\end{align*}
\]

with initial conditions

\[
V(0) = c_0, \quad V^*(0) = 0, \quad c_{i_0}(0) = c_i(0) = 0, \quad c^\varepsilon(x, 0) = 0,
\]

and the boundary condition

\[
c^\varepsilon(N_0, t) = c_{N_0}(t).
\]

In this resulting model one has passed to the continuous representations for chain lengths larger than \(i = N_0\).

We developed methodology for efficient forward solutions in [12]. We note that the desired spatial computational domain is very large as determined by
the maximum size of observed polymers, with range up to $10^6$. The peak in the distribution is at the left side of the domain of interest; for larger polymer sizes, the distribution is almost linearly decreasing. Based on these and other considerations discussed in [12], the PDE was approximated by the Finite Volume Method (see [51] for discussions of Upwind, Lax-Wendroff and flux limiter methods) with an adaptive mesh, refined toward the smaller polymer sizes. Further details on these schemes including examples demonstrating convergence properties may be found in [12].

3.4 Parameterizations and The Resulting Inverse Problem

An important question in formulating the model for use in inverse problems is how to best parametrically represent the polymerization parameters $k^i_{on}$ of (20) and the polymerization function $k_{on}$ of (21) for our application. We do this in a piecewise continuous formulation of the function $k_{on}(x)$ given by the piecewise linear representation (see Figure 2)

$$
k_{on}(x) = \begin{cases} 
  k_{on}^{\min} + x \frac{k_{on}^{\max} - k_{on}^{\min}}{x_1^{i_{max} - i_0}}, & x \leq x_1^{i_{max}} \\
  k_{on}^{\max}, & x_1^{i_{max}} \leq x \leq x_2^{i_{max}} \\
  k_{on}^{\max} - x \frac{k_{on}^{\max}}{x_2^{i_{max}} - x}, & x_2^{i_{max}} \leq x \leq i_{max} \\
  0, & x \geq i_{max}.
\end{cases}
$$

In our numerical approximations, we chose $i_0 = 2, N_0 = 500$. The discrete polymerization parameters $k^i_{on}, i = i_0, \ldots, N_0$, are then obtained as

$$
k^i_{on} := k_{on}(x = i).
$$

We followed [52] in choosing to approximate $k_{on}$ by a function as depicted in Figure 2. Other choices like a Gaussian bell curve are also possible (based on our discussions with S. Prigent, H. Rezaei and J. Torrent), but as we will subsequently conclude, the presently available data will not support estimation of parameters in these representations. Thus with this parametrization we have 5 more parameters, $k_{on}^{\min}, k_{on}^{\max}$, the fractions $x_1, x_2$, and $i_{max}$, in addition to the 4 basic parameters $k^+_I, k^-_I, k_{on}^N, k_{off}^N$ to be estimated using our data sets. That is, $\theta = (k^+_I, k^-_I, k_{on}^N, k_{off}^N, k_{on}^{\min}, k_{on}^{\max}, x_1, x_2, i_{max})$ with scalar observations $f(t, \theta) = \sum_{i \geq i_0} ic_i(t)$.

The goal then in [53] was to estimate the 9 parameters $k^+_I, k^-_I, k_{off}^N, k_{on}^N$, and $k_{on}$ (represented in parametrical form depicted in Figure 2 with the 5 additional unknowns $k_{on}^{\min}, k_{on}^{\max}, x_1, x_2, i_{max}$) that best fit the data. Equally important from a scientific viewpoint was to carry out this estimation with some acceptable quantification of uncertainties in the estimated parameters. To do this we utilized an efficient discretization method as discussed above for the forward
problem as well as a correct assumption on the measurement errors in the inverse problem. For this we use the ideas from residual plot analysis [17, 28] in attempts to obtain an acceptable statistical model as in equation (4).

3.5 Use of Residual Plots for Statistical Model Evaluation

To pursue a correct statistical model for the polymerized mass data, we carried out (as detailed in [13]) a series of inverse problems and residual plots with data set DS 4 of the experimental data collection. We first used DS 4 on the interval $t \in [0, 8]$. Based on some earlier calculations, we also chose the nucleation index $i_0 = 2$ for all of our subsequent calculations. The residual plots given in [13] strongly support the conclusions that neither of the initial assumptions for statistical models and corresponding cost functionals (absolute error with $\gamma = 0$ and OLS or relative error with $\gamma = 1$ and IRWLS) are correct.

Based on these initial results and the speculation that early periods of the polymerization process may be somewhat stochastic in nature, we chose to subsequently use all the data sets on the intervals $[t_0, 8]$ where $t_0$ is the first time when $f(t_0) > 0.12$ (thus larger than 12% of the non-dimensional total polymerized mass where it is supposed that the polymerization process becomes more deterministic). Moreover, we used other values of $\gamma$ between 0 and 1 to test data set DS 4. Setting $i_0 = 2$, we focused on the question of the most appropriate values of $\gamma$ to use in a generalized least squares approach (again see [17] for further motivation and details). Analysis of the resulting residuals for randomness suggested that either $\gamma = 0.6$ or $\gamma = 0.7$ might be satisfactory for use in a generalized least squares formulation.

Motivated by these results, we further investigated the corresponding inverse
problems for each of the 4 experimental data sets with initial concentration \( c_0 = 200 \ \mu\text{mol} \) and \( i_0 = 2 \). We carried out the optimization over all data points with \( f(t_k) \geq 0.12 \) and used the generalized least squares method with \( \gamma = 0.6 \). The resulting graphics depicted in [13] again suggested that \( \gamma = 0.6 \) is a reasonable value to use in any subsequent analysis of the polyglutamine data for inverse problem estimation and associated parameter uncertainty quantification.

### 3.6 Summary of Findings

In the problem outlined above, the authors of [13] (as did the authors of [52]) did indeed obtain a good fit of the curve and reasonable residuals. However, they also found that the condition number of the \( \kappa_\theta \times \kappa_\theta = 9 \times 9 \) approximate covariance matrix \( F = \chi^T(\hat{\theta})W(\hat{\theta})\chi(\hat{\theta}) \) is \( \kappa \approx 10^{24} \). Looking more closely at the matrix \( F \) revealed a near linear dependence between certain rows, hence the large condition number. One thus can readily draw the following summary conclusions:

1. One obtains a set of parameters for which the model fits well, but one cannot have any reasonable confidence in them using the asymptotic theories from statistics, e.g., see the references [13, 52].

2. We suspect that it may not be possible to obtain sufficient information from our data set curves to estimate all 9 parameters with a high degree of confidence! This is based on our calculations with the corresponding covariance matrices as well our prior knowledge that the graphs depicted in Figure 1 are very similar to Logistic, Gompertz or other bounded growth curves. These curves can usually be sufficiently well-fit with parameterized models with at most 2 or 3 carefully chosen parameters!

To assist in the understanding of these issues, we turn to consider components of the associated sensitivity matrices

\[
\chi = \left[ \frac{\partial f}{\partial \theta} \right]
\]

### 3.7 Sensitivity Analysis

For the sensitivity analyses, we follow [17, 28] and carry out all computations using the differential system of sensitivity equations as detailed in those references. Our subsequent sensitivity analyses were carried out using data set DS 4 and the best estimate \( \hat{\theta} \) obtained for this randomly chosen data set. As we see in the figures below, the model is most sensitive to 4 parameters: \( k_{+}, k_{-}, k_{\text{on}}, k_{\text{off}} \). The sensitivities for the remaining parameters are on the order of magnitude of \( 10^{-6} \) or less (e.g., see the plots in [13]). The observation \( f(t; \hat{\theta}) \) also exhibits some sensitivity with respect to \( x_1 \). However, the parameter \( x_1 \) appears in the model only as the factor \( x_1 i_{\text{max}} \). The sensitivities depicted below and in [13] use
for the 9 best fit GLS parameters, i.e., \( \hat{\theta} \) for \( \kappa_\theta = 9 \). We note that since we use the non-dimensional quantity \( f(t; \hat{\theta}) \) (or Madim) in the cost functionals, it is the sensitivity of this quantity with respect to the parameters \( \hat{\theta} \) (rather than any relative sensitivities), that will determine changes in the cost functionals to be minimized with respect to changes in the parameters.

Figure 3: DS 4: Sensitivity of \( f = Madim \) with respect to (a) \( k_I^- \); (b) \( k_I^+ \)
Figure 4: DS 4: Sensitivity of $f = Madim$ with respect to (a) $k^N_{on}$; (b) $k^N_{off}$
3.8 Inverse Problems Motivated by Sensitivity Findings

Based on the sensitivity findings depicted above, we investigated a series of inverse problems in which we attempted to estimate an increasing number of parameters beginning first with the fundamental parameters $k^+_I$ and $k^-_I$. In each of these inverse problems we attempted to establish uncertainty bounds for the estimated parameters using both the asymptotic theory and a generalized least squares version of bootstrapping described above in Section 2.4.

3.8.1 Estimation of $k^+_I$ and $k^-_I$

We first carried out estimation for the 2 parameters $k^+_I$ and $k^-_I$. We use the IRWLS formulation with $\gamma = 0.6$. Based on previous estimations with DS 4, we fixed globally the parameter values $k^+_N = 4616.962$, $k^-_N = 93.332$, $k^\min = 1684.381$, $k^\max = 1.5152 \times 10^9$, $x_1 = 0.0626$, $x_2 = 0.859$, $i_{\max} = 3.542 \times 10^5$. In carrying out the inverse problem we used the initial guesses (again based on previous estimations with DS 4) $q^0$ for the parameters given by $(k^+_I)^0 = 2.16$, $(k^-_I)^0 = 10.927$.

We also used the bootstrapping algorithm given above with $M = 1000$ to compute means and standard errors. These are given in the table below and compare quite well with the asymptotic theory estimates. The corresponding bootstrapping distributions are depicted in Figures 5 and 6.

<table>
<thead>
<tr>
<th></th>
<th>$k^+_I(boot)(GLS)$</th>
<th>$k^-_I(boot)(GLS)$</th>
<th>$k^+_I(asymp)(GLS)$</th>
<th>$k^-_I(asymp)(GLS)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>2.158</td>
<td>10.911</td>
<td>2.157</td>
<td>10.911</td>
</tr>
<tr>
<td>SE</td>
<td>0.0044</td>
<td>0.0247</td>
<td>0.00396</td>
<td>0.0225</td>
</tr>
</tbody>
</table>
Figure 5: DS 4: Two parameter estimation \( (k_I^+, k_I^-) \). Bootstrapping distribution for \( k_I^+ \) with \( M = 1000 \) runs.

Figure 6: DS 4: Two parameter estimation \( (k_I^+, k_I^-) \). Bootstrapping distribution for \( k_I^- \) with \( M = 1000 \) runs.
3.8.2 Estimation of 3 parameters

We tried next to estimate 3 parameters using the IRWLS formulation, again with $\gamma = 0.6$. Once more we fixed all the parameters describing the domain and the polymerization function $k_{on}$ and we also fixed either $k_{off}$ or $k_{on}$ in the corresponding inverse problems.

3.8.3 Estimation for $k_I^+$, $k_I^-$ and $k_{on}^N$

We fixed values of $k_{off}^N$, $k_{on}^{min}$, $k_{on}^{max}$, $x_1$, $x_2$, $i_{max}$ as before and used initial values 2.1600, 10.9270, 4616.962 for $k_I^+$, $k_I^-$, $k_{on}^N$, respectively.

We obtained the estimated parameters together with the corresponding standard errors, variances and the condition numbers $\kappa$ of the corresponding covariance matrices for the 4 data sets as reported below. The resulting 95% confidence results based on the asymptotic theory were quite acceptable.

<table>
<thead>
<tr>
<th></th>
<th>$k_I^+$</th>
<th>$k_I^-$</th>
<th>$k_{on}^N$</th>
<th>$SE$</th>
<th>$\sigma^2$</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>2.26</td>
<td>13.49</td>
<td>4616.96</td>
<td>(0.12, 0.99, 53.925)</td>
<td>$8.52 \cdot 10^{-b}$</td>
<td>$8.89 \cdot 10^{10}$</td>
</tr>
<tr>
<td>DS2</td>
<td>2.99</td>
<td>16.20</td>
<td>4616.96</td>
<td>(0.21, 1.51, 56.691)</td>
<td>$9.67 \cdot 10^{-b}$</td>
<td>$4.37 \cdot 10^{10}$</td>
</tr>
<tr>
<td>DS3</td>
<td>2.18</td>
<td>15.76</td>
<td>9840.31</td>
<td>(0.11, 1.03, 90.466)</td>
<td>$6.45 \cdot 10^{-b}$</td>
<td>$3.94 \cdot 10^{11}$</td>
</tr>
<tr>
<td>DS4</td>
<td>2.16</td>
<td>10.91</td>
<td>4616.96</td>
<td>(0.0089, 0.0649, 45.262)</td>
<td>$6.36 \cdot 10^{-b}$</td>
<td>$7.14 \cdot 10^{10}$</td>
</tr>
</tbody>
</table>

To compare these asymptotic results with bootstrapping, we carried out bootstrapping with data set DS 4 for the estimation of $k_I^+$, $k_I^-$ and $k_{on}^N$ with the same initial values as above. We then obtained the following means and standard errors (SE) for a run with $M = 1000$, which is compared to the asymptotic theory in the table below.

<table>
<thead>
<tr>
<th></th>
<th>$k_I^+$ (boot)</th>
<th>$k_I^-$ (boot)</th>
<th>$k_{on}^N$ (boot)</th>
<th>$k_I^+$ (asymp)</th>
<th>$k_I^-$ (asymp)</th>
<th>$k_{on}^N$ (asymp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>2.153</td>
<td>10.887</td>
<td>4616.962</td>
<td>2.157</td>
<td>10.910</td>
<td>4616.962</td>
</tr>
<tr>
<td>SE</td>
<td>0.0039</td>
<td>0.0219</td>
<td>0.00003</td>
<td>0.0089</td>
<td>0.0649</td>
<td>45.262</td>
</tr>
</tbody>
</table>

Of noticeable interest are the values obtained for $k_{on}^N$ and the bootstrapping standard errors for $k_{on}^N$, which are extremely small. It should be noted that the sensitivity of the model output with respect to $k_{on}^N$ is also very small. Thus one might suspect that the iterations in the bootstrapping algorithm do not change the values of $k_{on}^N$ very much and hence one observes the extremely small SE that are produced for the bootstrapping estimates. In particular we note the extremely fine scale on the abscissa axes in Figures 7, 8 and 9, where the graphed figures are much more spike-like than one might realize at first glance.
Figure 7: DS 4: Estimation for $k^+_I, k^-_I$ and $k^N_{on}$: Bootstrapping distribution for $k_I$ for $M = 1000$ runs.

Figure 8: DS 4: Estimation for $k^+_I, k^-_I$ and $k^N_{on}$: Bootstrapping distribution for $k_I$ for $M = 1000$ runs.
Figure 9: DS 4: Estimation for $k_I^+$, $k_I^-$ and $k_{on}^N$: Bootstrapping distribution for $k_{on}^N$ for $M = 1000$ runs.
3.8.4 Estimation for $k_I^+, k_I^-$ and $k_N^{off}$

In another estimation, we fixed $k_N$ at 4616.962 along with the other fixed parameter values used above and instead estimated $k_N^{off}$ along with $k_I^+$ and $k_I^-$. Initial values for the 3 parameters were 2.16, 10.9270, 108.256 for $k_I^+$, $k_I^-$, $k_N^{off}$, respectively.

We obtained the estimated parameters and corresponding SE for the 4 data sets reported in tabular form below.

<table>
<thead>
<tr>
<th></th>
<th>$k_I^+$</th>
<th>$k_I^-$</th>
<th>$k_N^{off}$</th>
<th>$SE$</th>
<th>$\sigma^2$</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>2.203</td>
<td>12.997</td>
<td>99.861</td>
<td>(.011,.091,1.208)</td>
<td>8.165$\cdot$10$^{-6}$</td>
<td>4.912$\cdot$10$^4$</td>
</tr>
<tr>
<td>DS2</td>
<td>2.893</td>
<td>15.474</td>
<td>100.019</td>
<td>(.019,.137,1.279)</td>
<td>9.323$\cdot$10$^{-6}$</td>
<td>2.486$\cdot$10$^4$</td>
</tr>
<tr>
<td>DS3</td>
<td>2.168</td>
<td>15.631</td>
<td>41.935</td>
<td>(.011,.102,0.424)</td>
<td>6.435$\cdot$10$^{-6}$</td>
<td>9.125$\cdot$10$^4$</td>
</tr>
<tr>
<td>DS4</td>
<td>2.181</td>
<td>11.090</td>
<td>90.536</td>
<td>(.009,.066,0.936)</td>
<td>6.289$\cdot$10$^{-6}$</td>
<td>3.043$\cdot$10$^4$</td>
</tr>
</tbody>
</table>

In addition, we carried out bootstrapping for DS 4. The bootstrapping distributions for $k_I^+$, $k_I^-$ and $k_N^{off}$ are given in Figures 10-12. We then obtained the following means and standard errors for a bootstrapping run with $M = 1000$ as compared to the asymptotic theory.

<table>
<thead>
<tr>
<th></th>
<th>$k_I^+$ (boot)</th>
<th>$k_I^-$ (boot)</th>
<th>$k_N^{off}$ (boot)</th>
<th>$k_I^+$ (asymp)</th>
<th>$k_I^-$ (asymp)</th>
<th>$k_N^{off}$ (asymp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>2.169</td>
<td>11.013</td>
<td>91.254</td>
<td>2.181</td>
<td>11.090</td>
<td>90.536</td>
</tr>
<tr>
<td>SE</td>
<td>0.0094</td>
<td>0.0699</td>
<td>1.0392</td>
<td>0.009</td>
<td>0.066</td>
<td>0.936</td>
</tr>
</tbody>
</table>

Figure 10: Three parameter estimation ($k_I^+$, $k_I^-$ and $k_N^{off}$): Bootstrapping distribution for $k_I^+$. We again used $M = 1000$ runs.
Figure 11: Three parameter estimation ($k^+_I$, $k^-_I$ and $k^N_{off}$): Bootstrapping distribution for $k^-_I$, $M = 1000$ runs.

Figure 12: Three parameter estimation ($k^+_I$, $k^-_I$ and $k^N_{off}$): Bootstrapping distribution for $k^N_{off}$, $M = 1000$ runs.
3.8.5 Estimation of 4 main parameters

In light of the sensitivity analysis discussed above, we tried to estimate a combination of the parameters $k_{I+}^+, k_{I-}^-, k_{N_{on}}^N, k_{N_{off}}^N$ (the parameter sets with $\kappa_{\theta} = 4$). From the original $\kappa_{\theta} = 9$ parameter fits, parameters were fixed for $k_{min}^{min}, k_{max}^{max}, x_1, x_2, t_{max}$ as $1684, 1.5 \times 10^9, 0.062, 0.859, 3.5 \times 10^5$, respectively. We obtained the following results for the estimation of the 4 parameters using the data sets DS 1 to DS 4. In all of these results, the condition number $\kappa$ of the Fisher’s Information Matrix is too large to carry out the required inversion in order to compute standard errors. This, along with the sensitivity results above, strongly supports the conclusion that the data sets do not contain sufficient information to estimate 4 or more parameters with any degree of certainty attached to the estimates. Again, this supports our expectation that the curves depicted in Figure 1 can each be readily and accurately modeled using simple growth models with at most 2 or 3 parameters. We will revisit this example and these conclusions after the next section where we introduce model comparison techniques as a tool in information content analysis.

<table>
<thead>
<tr>
<th></th>
<th>$k_{I+}^+$</th>
<th>$k_{I-}^-$</th>
<th>$k_{N_{on}}^N$</th>
<th>$k_{N_{off}}^N$</th>
<th>$\sigma^2$</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>2.1431</td>
<td>12.4751</td>
<td>4616.962</td>
<td>108.259</td>
<td>$8.7219 \cdot 10^{-6}$</td>
<td>$6.1226 \cdot 10^{19}$</td>
</tr>
<tr>
<td>DS2</td>
<td>2.7995</td>
<td>14.7639</td>
<td>4616.957</td>
<td>108.4308</td>
<td>$9.8694 \cdot 10^{-6}$</td>
<td>$1.4442 \cdot 10^{19}$</td>
</tr>
<tr>
<td>DS3</td>
<td>2.180</td>
<td>15.757</td>
<td>4618.599</td>
<td>41.369</td>
<td>$6.4622 \cdot 10^{-6}$</td>
<td>$1.881 \cdot 10^{17}$</td>
</tr>
<tr>
<td>DS4</td>
<td>2.161</td>
<td>10.9278</td>
<td>4617.3316</td>
<td>93.3265</td>
<td>$6.374 \cdot 10^{-6}$</td>
<td>$2.144 \cdot 10^{18}$</td>
</tr>
</tbody>
</table>
4 Model Comparison: Nested Restraint Sets

Below we will demonstrate the use of statistically based model comparison tests in several examples of practical interest. In these examples we are interested in questions related to information content of a particular given data set and whether the data will support a more detailed or sophisticated model to describe it. In the next subsection we recall the fundamental statistical tests to be employed here.

4.1 Statistical Comparison Tests

In general, assume we have an inverse problem for the model observations \(f(t, \theta)\) and are given \(n\) observations. As in (9), we define

\[
J^n(Y; \theta) = \sum_{i=1}^{n} w_i(\hat{\theta})^{-2\gamma} (Y_i - f(t_i; \theta))^2
\]

where our statistical model has the form (4). Here, as before, \(\theta_0\) is the nominal value of \(\theta\) which we assume to exist. We use \(\Omega\) to represent the set of all the admissible parameters \(\theta\). We make some further assumptions:

A5) Observations are taken at \(\{t_j\}_{j=1}^{n}\) in \([0, T]\). There exists some finite measure \(\mu\) on \([0, T]\) such that

\[
\frac{1}{n} \sum_{j=1}^{n} h(t_j) \longrightarrow \int_{0}^{T} h(t)d\mu(t)
\]

as \(n \to \infty\), for all continuous functions \(h\).

A6) \(J_0(\theta) = \int_{0}^{T} (f(t; \theta_0) - f(t; \theta))^2 d\mu(t) = \sigma^2\) has a unique minimizer in \(\Omega\) at \(\theta_0\).

Let \(\Theta^n = \Theta^n_{IRWLS}(Y)\) be the IRWLS estimator for \(J^n\) as defined in (10) so that

\[
\Theta^n(Y) = \underset{\theta \in \Omega}{\text{argmin}} J^n(Y; \theta)
\]

and

\[
\hat{\theta}^n = \underset{\theta \in \Omega}{\text{argmin}} J^n(y; \theta),
\]

where \(y\) is a realization for \(Y\).

One can then establish a series of useful results (see [14, 17, 22] for detailed arguments and proofs).

**Result 1:** Under A1) to A6), \(\frac{1}{n} \Theta^n = \frac{1}{n} \Theta^n_{IRWLS}(Y) \longrightarrow \theta_0\) as \(n \to \infty\) with probability 1.

We will need further assumptions to proceed (these will be denoted by A7–A11) to facilitate reference to [14, 17]). These include:
A7) The nominal parameter $\theta_0 \in \mathbb{R}^p$ satisfies $\theta_0 \in \text{int}(\Omega)$.

A8) $f : \Omega \to C[0,T]$ is a $C^2$ function.

A10) $J = \frac{\partial^2 J_0}{\partial \theta^2} (\theta_0)$ is positive definite.

A11) $\Omega_H = \{ \theta \in \Omega | H\theta = c \}$ where $H$ is an $r \times p$ matrix of full rank, and $c$ is a known constant. Here $r$ is the number of constraints placed on the reduced model parameters.

In many instances, including the motivating examples discussed here, one is interested in using data to question whether the “nominal” parameter $\theta_0$ can be found in a subset $\Omega_H \subset \Omega$ which we assume for discussions here is defined by the constraints of assumption A11). Thus, we want to test the null hypothesis $H_0$: $\theta_0 \in \Omega_H$, i.e., that the constrained model provides an adequate fit to the data.

Define then

$$\Theta^n_H(Y) = \arg\min_{\theta \in \Omega_H} J^n(Y; \theta)$$

and

$$\hat{\theta}^n_H = \arg\min_{\theta \in \Omega_H} J^n(y; \theta).$$

Observe that $J^n(y; \hat{\theta}^n_H) \geq J^n(y; \hat{\theta}^n)$. We define the related non-negative test statistics and their realizations, respectively, by

$$T_n(Y) = J^n(Y; \theta^n_H) - J^n(Y; \theta^n)$$

and

$$\hat{T}_n = T_n(y) = J^n(y; \hat{\theta}^n_H) - J^n(y; \hat{\theta}^n).$$

One can establish asymptotic convergence results for the test statistics $T_n(Y)$—see [14]. These results can, in turn, be used to establish a fundamental result about much more useful statistics for model comparison. We define these statistics by

$$U_n(Y) = \frac{nT_n(Y)}{J^n(Y; \theta^n)},$$

with corresponding realizations

$$\hat{u}_n = U_n(y).$$

We then have the asymptotic result that is the basis of our analysis-of-variance–type tests:

**Result 2:** Under the assumptions A1)–A11) above and assuming the null hypothesis $H_0$ is true, then $U_n(Y)$ converges in distribution (as $n \to \infty$) to a random variable $U(r)$, i.e., $U_n \overset{D}{\to} U(r)$, with $U(r)$ having a chi-square distribution $\chi^2(r)$ with $r$ degrees of freedom.

We note that if one is dealing with vector observations where $n = n_1 + n_2$ is the total component observations as we do in the examples in [6, 7] and in the
last example in Section 6.4, then asymptotic theory requires that both \( n_1 \to \infty \)
and \( n_2 \to \infty \). In any graph of a \( \chi^2 \) density there are two parameters \((\tau, \alpha)\)
of interest. For a given value \( \tau \), the value \( \alpha \) is simply the probability that the
random variable \( U \) will take on a value greater than \( \tau \). That is, \( \text{Prob}\{U > \tau\} = \alpha \) where in
hypothesis testing, \( \alpha \) is the \textit{significance level} and \( \tau \) is the \textit{threshold}.

We then wish to use this distribution \( U_n \sim \chi^2(r) \) to test the null hypothesis, \( H_0 \),
that the restricted model provides an adequate fit to represent the data. If the test statistic,
\( \hat{u}_n > \tau \), then we \textit{reject} \( H_0 \) as false with confidence level \((1 - \alpha)100\%\). Otherwise, we \textit{do not reject} \( H_0 \). For our examples below,
we use a \( \chi^2(1) \) table, which can be found in any elementary statistics text or
online. Typical confidence levels of interest are \( 75\% \), \( 90\% \), \( 95\% \), \( 99\% \) and \( 99.9\% \) with
corresponding \((\alpha, \tau)\) values of \((.25, 1.32)\), \((.1, 2.71)\), \((.05, 3.84)\), \((.01, 6.63)\),
and \((.001, 10.83)\), respectively.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \tau )</th>
<th>confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>1.32</td>
<td>75%</td>
</tr>
<tr>
<td>.1</td>
<td>2.71</td>
<td>90%</td>
</tr>
<tr>
<td>.05</td>
<td>3.84</td>
<td>95%</td>
</tr>
<tr>
<td>.01</td>
<td>6.63</td>
<td>99%</td>
</tr>
<tr>
<td>.001</td>
<td>10.83</td>
<td>99.9%</td>
</tr>
</tbody>
</table>

Table 1: \( \chi^2(1) \)

To test the null hypothesis \( H_0 \), we choose a significance level \( \alpha \) and use \( \chi^2 \) tables to obtain the corresponding threshold \( \tau = \tau(\alpha) \) so that \( \text{Prob}\{\chi^2(r) > \tau\} = \alpha \). We next compute \( \hat{u}_n = \tau \) and compare it to \( \tau \). If \( \hat{u}_n > \tau \), then we
\textit{reject} \( H_0 \) as false; otherwise, we do not reject the null hypothesis \( H_0 \).

5 Model Comparison: Non-Nested Restraint Sets

There are a number of model comparison or model selection criteria in the statisti-cal/mathematical literature that can be used to select a “best” approximating
model from a prior collection of competing candidate models. These criteria are
based on either hypothesis testing, e.g., log-likelihood ratio test, and residual
sum of squares based model selection criterion such as the model comparison
techniques for nested models as outlined in the previous section as well as information
theory based techniques (e.g., the Akaike Information Criterion as well as its variations [29, 30, 31, 47]). In these latter techniques the general goal in
the model selection is to minimize both modeling error (bias) and estimation
error (variance).

Most of these model selection criteria are based on the likelihood function
and thus these types of model selection criteria are referred to as likelihood
based model selection criteria, and include one of the most widely used model
selection criteria, the Akaike Information Criterion (AIC) and its variants. In
1973, Akaike found a relationship between Kullback-Leibler information (a well-
known measure of “distance” between two probability distribution models) and
the maximum value of log-likelihood function of a given approximating model
(this relationship is referred to as the Akaike Information Criterion). These
criteria can be used to measure information lost when an approximating proba-
bility distribution model is used to approximate the true probability distribution
model \( p_0 \), which is tacitly assumed to exist for the model comparison/selection
techniques we describe here. That is, let \( p_0 \) denote the probability distribution
model that actually generates the data (that is, is the true probability density
function of some observations \( Y \)), and \( p \) be a probability distribution model that
is presumed to approximate the data. In addition, \( p \) is assumed to be depen-
dent on a parameter vector \( \theta \in \mathbb{R}^{\kappa} \) (that is, \( p(\cdot|\theta) \) is the specified probability
density function of observations \( Y \), and is used to approximate the true model
\( p_0 \)). Then the K-L information between these two models is given by

\[
\mathcal{I}(p_0, p(\cdot|\theta)) = \int_{\Omega_y} p_0(y) \ln \left( \frac{p_0(y)}{p(y|\theta)} \right) dy
\]

\[
= \int_{\Omega_y} p_0(y) \ln(p_0(y)) dy - \int_{\Omega_y} p_0(y) \ln(p(y|\theta)) dy,
\]

where \( \Omega_y \) denotes the set of all possible values for \( y \), and the second term in the
right side, \( \int_{\Omega_y} p_0(y) \ln(p(\cdot|\theta)) dy \), is referred to as the relative K-L information.

5.1 A Large Sample AIC

As we have noted, the AIC is based on K-L information theory, which measures
the “distance” between two probability distribution models. In establishing the
AIC, the maximum likelihood estimation method is used for parameter estima-
tion. Note that the K-L information \( \mathcal{I}(p_0, p(\cdot|\theta_{MLE}(Y))) \) is a random variable
(inherited from the fact that \( \theta_{MLE}(Y) \) is a random vector). Hence, we need
to use the expected K-L information \( \mathbb{E}_Y (\mathcal{I}(p_0, p(\cdot|\theta_{MLE}(Y)))) \) to measure the
“distance” between a candidate distribution \( p \) and the assumed true distribu-
tion \( p_0 \), where \( \mathbb{E}_Y \) denotes the expectation with respect to the true probability
density function \( p_0 \) of observations \( Y \).

It was shown (e.g., see [32] for details) that for a large sample and “good”
model (a model that is close to \( p_0 \) in the sense of having a small K-L value) we have

\[
\mathbb{E}_Y \mathbb{E}_X (\ln(p(X|\theta_{MLE}(Y)))) \approx \ln(L(\hat{\theta}_{MLE}|Y)) - \kappa_\theta.
\]

Here \( \hat{\theta}_{MLE} \) is the maximum likelihood estimate of \( \theta \) given sample outcomes
\( y \) (that is, \( \hat{\theta}_{MLE} = \theta_{MLE}(y) \)), \( L(\hat{\theta}_{MLE}|y) \) represents the likelihood of \( \hat{\theta}_{MLE}
\) given sample outcomes \( y \) (that is, \( L(\hat{\theta}_{MLE}|y) = p(y|\hat{\theta}_{MLE}) \)), and \( \kappa_\theta \) is the total
number of estimated parameters (including mathematical model parameters \( q \)
and statistical model parameters). It is worth pointing out here that having
a large sample and “good” model are used to ensure that the estimate \( \hat{\theta}_{MLE}
\) provides a good approximation to some true value \( \theta_0 \) (involving the consistency
property of the maximum likelihood estimator).
For historical reasons, Akaike multiplied (25) by $-2$ to obtain his criterion, which is given by

$$AIC = -2 \ln \mathcal{L} (\hat{\theta}_{MLE} | y) + 2\kappa_\theta.$$  

(26)

We note that the first term in the $AIC$ is a measure of the goodness-of-fit of the approximating model, and the second term gives a measure of the complexity of the approximating model (i.e., the reliability of the parameter estimation of the model). Thus, we see that for the $AIC$ the complexity of a model is viewed simply as the number of parameters in the model.

Based on the above discussion, we see that to use the $AIC$ to select a best approximating model from a given prior set of candidate models, we need to calculate the $AIC$ value for each model in the set, and the “best” model is the one with the minimum $AIC$ value. Note that the value of the $AIC$ depends on data, which implies that we may select a different best approximating model if a different data set arising from the same experiment is used. Hence, the $AIC$ values must be calculated for all the models being compared by using the same data set. That is, the $AIC$ cannot be used to compare models for different data sets. For example, if a model is fit to a data set with $n = 140$ observations, one cannot validly compare it with another model when 7 outliers have been deleted, leaving only $n = 133$.

Under reasonable assumptions (essentially normality of the measurement errors) [17, 18], one can use ordinary least squares (OLS), weighted least squares (WLS), or iterative reweighted weighted least squares (IRWLS) estimators in place of the usual maximum likelihood estimators in formulating $AIC$ comparison factors.

### 5.2 A Small Sample $AIC$

The discussion in Section 5.1 reveals that one of the assumptions made in the derivation of the $AIC$ is that the sample size must be large. Hence, the $AIC$ may perform poorly if the sample size $n$ is small relative to the total number of estimated parameters. It is suggested in [32] that the $AIC$ can be used only if the sample size $n$ is at least 40 times of total number of estimated parameters (that is, $n/\kappa_\theta \geq 40$). In this section, we introduce a small sample $AIC$ (denoted by $AIC_c$) that can be used in the case where $N$ is small relative to $\kappa_\theta$.

The $AIC_c$ was originally proposed in [57] for a scalar linear regression model, and then was extended in [47] for a scalar nonlinear regression model based on asymptotic theory. In deriving the $AIC_c$, it was assumed in [47] that the measurement errors $e_j, j = 1, 2, \ldots, n$, are independent and normally distributed with mean zero and variance $\sigma^2$. In addition, the true model $p_0$ was assumed to be known with measurement errors being independent and normally distributed with zero mean and variance $\sigma_0^2$. With these assumptions, the small sample $AIC$ is given by

$$AIC_c = AIC + \frac{2\kappa_\theta (\kappa_\theta + 1)}{n - \kappa_\theta - 1},$$  

(27)

where the last term in the right-hand side of the above equation is often referred
to as the bias-correction term. We observe that as the sample size $n \to \infty$ this bias-correction term approaches zero, and the resultant criterion is just the usual $AIC$. From the remarks in the previous section and the results of [18], we again note that one can equivalently use OLS, WLS, etc. estimators in the formulation of the $AIC_c$ model comparison terms.

It should be noted that the bias-correction term in (27) changes if a different probability distribution (e.g., exponential, Poisson) is assumed for the measurement errors. However, it was suggested in [32] that in practice $AIC_c$ given by (27) is generally suitable unless the underlying probability distribution is extremely non-normal, especially in terms of being strongly skewed.

The $AIC_c$ in the multivariate observation case was derived in [29] and discussed more fully in [32, 17].

5.3 Akaike Weights and the Selected “Best” Model

As we have noted above, the selected “best” model is the one with the minimum $AIC$ value. It should be noted that the selected model is specific to the set of candidate models. It is also specific to the given data set. In other words, if one has a different set of experimental data arising even from the same experiment, one may select a different model. Hence, in practice, the absolute size of the $AIC$ value may have limited use in supporting the chosen best approximating model. In addition, the $AIC$ value is an estimate of the expected relative K-L information (hence, the actual value of the $AIC$ is meaningless). Thus, one may often employ other related values such as the Akaike difference and the Akaike weights.

The Akaike difference is defined by

$$\Delta_i = AIC_i - AIC_{min}, \quad i = 1, 2, \ldots, l,$$

(28)

where $AIC_i$ is the $AIC$ value of the $i$th model in the set, and $AIC_{min}$ denotes the $AIC$ value for the best model in the set, and $l$ is the total number of models in the candidate set for comparison. We see that the selected model is the one with zero Akaike difference. The larger $\Delta_i$, the less plausible it is that the $i$th model is the best approximating model given the data set.

Akaike weights are defined by

$$w_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{r=1}^{l} \exp(-\frac{1}{2}\Delta_r)}, \quad i = 1, 2, \ldots, l.$$  (29)

We note that the weights of all candidate models sum to 1, so the weight gives a probability that each model is the best model. Furthermore, the evidence ratio

$$\frac{w_i(AIC)}{w_j(AIC)}$$

(30)

indicates how much more likely model $i$ is compared to model $j$. In addition, if there are two models, say models $i$ and $j$, which have the largest and second
largest weights respectively, then the normalized probability

\[
\frac{w_i(AIC)}{w_i(AIC) + w_j(AIC)}
\]

(31)

indicates the probability of model \(i\) over model \(j\) [61].

The Akaike weight \(w_i\) is similar to the relative frequency for the \(i\)th model selected as the best approximating model by using the bootstrapping method. It can also be interpreted (in a Bayesian framework) as the actual posterior probability that \(i\)th model is the best approximating model given the data. We refer the interested reader to [32, Section 2.13] for details. Akaike weights are also used as a heuristic way to construct the 95% confidence set for the selected model by summing the Akaike weights from largest to smallest until the sum is \(\geq 0.95\). The corresponding subset of models is the confidence set for the selected model. Interested readers can refer to [32] for other heuristic approaches for construction of the confidence set.

As a followup to the weighting factors explanations construction of confidence sets, we emphasize that information criterion analysis is not a “test”, so one should avoid use of “significant” and “not significant”, or “rejected” and “not rejected” in reporting results. That is, null hypothesis testing should not be mixed with information criterion in reporting the results. In particular, one should not use the AIC to rank models in the set and then test whether the best model is “significantly better” than the second-best model.

6 Model Comparison and Information Content Examples

In the first example we return to the polymerization example of Section 3 to illustrate use these model comparison techniques as interrogating tools for data set content. We discuss these in the context of nested models as well as report on AIC comparison factors. In a second example we compare fits for several different models to describe simple decay in a size histogram for aggregates in amyloid fibril formation; this example is also related to aggregate formulation discussed in Section 3. In a related example we consider population experiments with green algae, formally known as \textit{Raphidocelis subcapitata}. In [10, 18] efforts by the authors were concerned with the growth dynamics of the algae as this is the major food source for \textit{Daphnia magna} [2] in experimental settings. In a fourth example we investigate whether the information content in data sets for the pest \textit{Lygus hesperus} in cotton fields as it is currently collected is sufficient to support a model in which one distinguishes between nymphs and adults.

6.1 Polymerization (again!)

Returning to the polymerization example of Section 3, we use the model comparison tests for nested models to determine if an added parameter yields a
statistically significantly improved model fit (we again use DS 4 for \( y \)). Our null hypothesis in each case is: \( H_0 \): The restricted or constrained model is adequate (i.e., the fit-to-data is not significantly improved with the model containing the additional parameter as a parameter to be estimated). We summarize our findings using the model comparison tests.

A) The model with estimation of \( \{k^+_I, k^-_I\} \) holding the other parameters fixed was compared with the model with estimation of \( \{k^+_I, k^-_I, k^N_{off}\} \). We found (in each case here \( n=699 \)),

\[
J_n(y; \hat{\theta}_H^n) = 0.0044192109
\]

and \( \hat{u}_n = 7.7178 \). Thus, we reject \( H_0 \) at a 99% confidence level. This means that the data set does support at a statistically significant level the model with estimation of the additional parameter \( k^N_{off} \). Note that if we compute the corresponding AIC comparison factors (again \( n=699 \)), we obtain \( AIC = -8362.04 \) for the two parameter model versus \( AIC = -8367.72 \) for the three parameter model. This difference suggests the estimation of 3 parameters yields a better model fit.

B) The model with estimation of \( \{k^+_I, k^-_I\} \) vs. the model with estimation of \( \{k^+_I, k^-_I, k^N_{on}\} \) was compared. We find

\[
J_n(y; \hat{\theta}_H^n) = 0.0044192108
\]

with \( \hat{u}_n = 7.49 \times 10^{-06} \). Thus we cannot reject \( H_0 \) at any reasonable confidence level so that estimation of the additional parameter \( k^N_{on} \) cannot be supported at any meaningful positive statistical level. The corresponding AIC factors are both -8360.04 which offers no support for estimation of the additional parameter.

C) The model with estimation of \( \{k^+_I, k^-_I, k^N_{off}\} \) was compared with the model with estimation of \( \{k^+_I, k^-_I, k^N_{off}, k^N_{on}\} \). To the order of computation accuracy we found no difference in the cost functions (hence no difference in the AIC factors) in this case and therefore we do not reject \( H_0 \) at any reasonable confidence level.

D) The model with estimation of \( \{k^+_I, k^-_I, k^N_{on}\} \) vs. the model with estimation of \( \{k^+_I, k^-_I, k^N_{on}, k^N_{off}\} \) was compared. We found

\[
J_n(y; \hat{\theta}_H^n) = 0.0044192108
\]

and

\[
J_n(y; \hat{\theta}_H^n) = 0.0043709780
\]

with \( \hat{u}_n = 7.7133 \) and hence we reject \( H_0 \) with a confidence level of 99%. The corresponding AIC factors are -8360.04 and -8365.71, respectively, which does offer some support for the increased parameter model.
From these and the preceding results from Section 3, we conclude that the information content of the typical data set for the dynamics considered in the nucleated polymerization models above will support at most 3 parameters $\{k_+^I, k_-^I, k_{off}\}$ estimated with reasonable confidence levels.

### 6.2 Size distributions of aggregates in amyloid fibril formation

In a recent paper [53], a question was addressed about size distribution of aggregates in amyloid fibril formation. While an exponential distribution was shown to provide a reasonable fit to the data depicted in Figure 13, the question arose as to whether another distribution such as the Weibull, Gamma, or some other decay distribution with more parameters might provide a better fit.

#### 6.2.1 The Exponential, Weibull, Gamma and other Decay Distributions

On initial observation, the data appears to be well suited to an exponential distribution. The exponential distribution probability density function is defined as $E(x; \lambda) = \lambda e^{-\lambda x}$. Note that when fitting the data, an additional parameter $A$ was added to the exponential function resulting in a total of two parameters and the function to be defined for these purposes as

$$E(x; A, \lambda) = Ae^{-\lambda x}.$$  

The Weibull distribution probability density function [6, 62] is defined as (for the purposes of modeling the data we again add the additional parameter $A$)

$$W(x; A, \lambda, k) = Ak \lambda (\lambda x)^{k-1} e^{-(\lambda x)^k}, \quad x \geq 0.$$  

Note that if we take $k = 1$ we have that $W(x; A, \lambda, 1) = E(x; A, \lambda)$. This function is shown plotted in [6, 62] with several values of $k$. One can see that when $k = 2$ or $k = 1$ the function also bears a resemblance to the shape of our data.

The probability density function of the Gamma distribution is defined as [6, 45] (we again include the additional parameter $A$ for modeling purposes)

$$G(x; A, k, \lambda) = A \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)} \quad \text{for } x > 0 \text{ and } k, \lambda > 0,$$

where $\Gamma(k)$ is the Gamma function evaluated at $k$. One can see [6, 45] that when $k = 1$ and $\lambda = 0.5$, the Gamma probability density function again has a similar shape to the data. Since we know that $\Gamma(1) = 1$, we can see that when we take $k = 1$ we have that $G(x; A, 1, \lambda) = E(x; A, \lambda)$.

The two final models we consider are the logistic decay model,

$$L(x; a, b, c) = \frac{c}{1 + ae^{-bx}}, \quad (32)$$
and the Gompertz decay model [49],

\[ D(x; A, \lambda, k) = A \exp \left( -\lambda e^{-kx} \right). \]  

(33)

The exponential model has only two parameters which must be estimated while each of the other models involve three parameters. In the fits-to-data depicted in Figure 13, we see that all 5 functions provide reasonable fits to the data. Thus an interesting first question is whether we can obtain a statistically significantly better fit to the data by allowing an additional third free parameter \( k \) in either the Weibull or Gamma distribution in comparison to the 2 parameter \((A, \lambda)\) exponential model. We further consider two additional decay models, the logistic and the Gompertz for which the nested model comparison techniques are not applicable. We thus turn to the \( AIC_c \) factors for comparisons ranking for all five models.

![Fibril Data with Fitted Functions](image)

Figure 13: This figure depicts experimental data \((n = 27)\) along with the fitted models as indicated. For each model the \( AIC_c \) and the model weights are also given.

### 6.2.2 Results using the comparison tests

We carried out fits-to-data using an OLS version of the inverse problems (we established with preliminary computations and residual analysis that \( \gamma = 0 \) provided a correct statistical model) for the exponential, Weibull and Gamma distributions. We first considered nested models and tested the following null
and alternative hypothesis for 2 of the alternative models (a Weibull and a Gamma distribution) as compared to the exponential:

- **H₀**: The fit provided by an alternative model is not statistically significantly different from the fit with an exponential distribution.
- **Hₐ**: The alternative model with an unrestricted additional parameter \( k \) provides a statistically significantly better fit than the exponential model (corresponding to the restriction \( k = 1 \)).

When comparing the best fits of the exponential (e) vs. the Weibull (W) distributions we obtained the following results:

\[
J^W_n = 1.4359 \times 10^{-4},
\]

and

\[
J^e_n = 1.6081 \times 10^{-4},
\]

with

\[
\hat{T}^eW_n = 4.6495 \times 10^{-4},
\]

and

\[
\hat{u}^eW_n = \bar{\tau} = 3.2381
\]

for the exponential-Weibull (eW) comparison test statistic. In this case we cannot reject the null hypothesis at the 95% or higher level. We can reject \( H₀ \) at the 90% confidence level. The corresponding \( AIC_c \) factors are given in Table 2 below and fully support these findings.

When comparing the best fits of the exponential (e) vs. the Gamma (G) distribution, we obtained the following results:

\[
J^G_n = 1.4277 \times 10^{-4},
\]

and

\[
J^e_n = 1.6081 \times 10^{-4},
\]

with

\[
\hat{T}^eG_n = 4.8693 \times 10^{-4},
\]

and

\[
\hat{u}^eG_n = \bar{\tau} = 3.4105
\]

for the exponential-Gamma (eG) comparison test statistic. Again in this case we cannot reject the null hypothesis at the 95% or higher level but we can reject \( H₀ \) at the 90% confidence level. The corresponding \( AIC_c \) factors given in Table 2 below again fully support these findings.

Using a threshold of \( \alpha = 0.05 \), we fail to reject the null hypothesis \( H₀ \) in the cases of comparisons between the Weibull or the Gamma distributions when compared to the exponential function. Thus we conclude that the exponential fit is adequate to describe the size distribution of aggregates in amyloid fibril formation.
Table 2: Comparison of $AIC_c$ values for each candidate model for the fibril size data

<table>
<thead>
<tr>
<th>Model</th>
<th>$AIC_{OLSC}$</th>
<th>$w_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>-227.70</td>
<td>0.138</td>
</tr>
<tr>
<td>Weibull</td>
<td>-228.97</td>
<td>0.259</td>
</tr>
<tr>
<td>Gamma</td>
<td>-229.09</td>
<td>0.276</td>
</tr>
<tr>
<td>Logistic</td>
<td>-228.39</td>
<td>0.194</td>
</tr>
<tr>
<td>Gompertz</td>
<td>-227.64</td>
<td>0.134</td>
</tr>
</tbody>
</table>

6.2.3 Results using $AIC_c$ factors

We next considered additional model fits involving Gompertz and logistic distributions and computed the corresponding $AIC_c$ factors. As indicated in Table 2, the Gamma distribution model is considered the 'best' of the candidate models for this data, closely followed by the Weibull model. However, using the evidence ratio in equation (30), the Gamma model is only 1.1 times more likely to be the best model in terms of Kullback-Leibler discrepancy than the Weibull model with a normalized probability of 0.52 (using equation (31) ) when compared to the Weibull model. However, when comparing the normalized probability of the Gamma distribution model to the worst model out of the candidate models (the exponential distribution model), there is a 0.68 probability of the Gamma distribution model as the preferred model over the exponential distribution model.

6.3 Growth dynamics for green algae *Raphidocelis subcapitata*

In ecology, daphnia (e.g., *D. magna*) can be used as early warning organisms because of changes in the daphnia population dynamics (thus they can be thought of as the modern day “canary in the mine shaft”) in response to changing dynamics in the environment. The authors of [10] were concerned with the growth dynamics of the algae as this is the major feeding source for *Daphnia magna* [2] in experimental settings. In a recent paper [10] longitudinal data were collected from four replicate population experiments with green algae, formally known as *Raphidocelis subcapitata*, and analyzed for growth. Here we compare three different dynamical population models for algae growth: the classical logistic model, the Bernoulli model, and the Gompertz model. In doing this we use here both the model comparison and the $AIC_c$ methodology outlined above.

The logistic model is a special case of the Bernoulli whereas the Gompertz is not directly related to either of the other two.

The logistic model is given by

\[
\frac{dx}{dt} = rx(t) \left(1 - \frac{x(t)}{K}\right), \quad x(0) = x_0
\]

where $r$ is the growth rate and $K$ is the carrying capacity for the population.
The Bernoulli model contains one additional parameter $\beta$ and is given by
\[
\frac{dx}{dt} = rx(t) \left( 1 - \left( \frac{x(t)}{K} \right)^\beta \right), \quad x(0) = x_0.
\]

Note that the logistic growth model is obtained from the Bernoulli growth model by setting $\beta$ equal to 1. In standard form, the parameters $K$ and $\beta$ are found jointly in the denominator, possibly causing a problem with identifiability. To address this issue, we let $\hat{K} = K^\beta$ and instead consider the model
\[
\frac{dx(t)}{dt} = rx(t) \left( 1 - \left( \frac{x(t)}{\hat{K}} \right)^\beta \right), \quad x(0) = x_0,
\]
where $K$ can be obtained from $\hat{K}$ using $K = (1/\beta)\hat{K}$. The third model we consider is the Gompertz model,
\[
\frac{dx(t)}{dt} = \kappa x(t) \log \left( \frac{K}{x(t)} \right), \quad x(0) = x_0,
\]
where $K$ is the carrying capacity as in the other two models and $\kappa$ scales the time. We note that both the logistic and Gompertz models contain only two ($\kappa_q = 2$) model parameters while the Bernoulli model contains three ($\kappa_q = 3$) model parameters, where in the notation of Section 5.2, we have $\theta = [q, \sigma]^T$ and $\kappa_\theta = \kappa_q + 1$.

In terms of modeling the algae data, it is demonstrated in the paper by Banks et al. [10] that the appropriate statistical model for this data is a parameter dependent weighted error statistical model with $\gamma = 1$ in Equation (6). Thus we use $\gamma = 1$ and the GLS (i.e., IRWLS) in computing our minimization result. In Table 3 we present comparison results for the logistic model for each replicate vs. the Bernoulli model with $J_{Ber} = .5857416$ for each replicate. From this table we can conclude from the comparison tests that the Bernoulli model is preferable at all reasonable confidence levels over the logistic model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
<th>Replicate 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>logistic</td>
<td>$J_{log}^n$</td>
<td>$J_{log}^n$</td>
<td>$J_{log}^n$</td>
<td>$J_{log}^n$</td>
</tr>
<tr>
<td>$\hat{u}^n$</td>
<td>1.024769</td>
<td>1.36611</td>
<td>1.368</td>
<td>1.951</td>
</tr>
<tr>
<td></td>
<td>26.92</td>
<td>47.88</td>
<td>48.22</td>
<td>83.79</td>
</tr>
</tbody>
</table>

Table 3: Comparison of model fit cost values $J_{log}^n$ and corresponding $\hat{u}^n$ for each logistic vs. Bernoulli with $J_{Ber} = .5857416$ for the algae data

As this was a small data set with only $n = 36$ data points for each of the four replicates, we used $AIC_{IRWLS_c}$ in Equation (38) to compare models. That
is,  
\[
AIC_{IRWLS} \approx n \ln \left( \frac{\sum_{j=1}^{n} w_j^{-2} (y_j - f(t_j, \hat{q}_{IRWLS}^M))^2}{n} \right) + 2(\kappa_\theta + 1),  
\]

where \( w_j = \hat{w}_j^M \approx f^*(t_j; \hat{q}_{IRWLS}) \) where \( M \) is the number of times the process is enumerated in the IRWLS algorithm of Section 2.2 and \( \kappa_\theta = \kappa_\theta + 1 \). Thus we use here

\[
AIC_{IRWLS_c} = AIC_{IRWLS} + \frac{2(\kappa_\theta + 1)(\kappa_\theta + 2)}{n - \kappa_\theta}.  
\]

The results for each replicate are given in Table 4 with the fitted models and data fits for replicate 1 plotted in Figure 14. As shown in Table 4, there is minimal difference across the four replicates and in each case, the smallest AIC value is given by the Gompertz model followed closely by the Bernoulli model. Recall that the Gompertz model has 2 parameters; whereas the Bernoulli model has three; therefore, although the two curves are lying on top of one another in Figure 14, the Bernoulli model is penalized more by the extra parameter. If we heuristically compare the Gompertz and Bernoulli models using the evidence ratio in equation (30) and the normalized probability in equation (31), we see that the Gompertz model is only 1.03 times more likely with a normalized probability of only 0.51 (only slightly more than equal probability). Therefore, either the Gompertz or the Bernoulli model appears to be a good model of the candidate models examined.

<table>
<thead>
<tr>
<th>Model</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC_{IRWLS_c}</td>
<td>( w_i )</td>
</tr>
<tr>
<td>logistic</td>
<td>-127.42</td>
<td>2.7e-05</td>
</tr>
<tr>
<td>Bernoulli</td>
<td>-147.05</td>
<td>0.492</td>
</tr>
<tr>
<td>Gompertz</td>
<td>-147.11</td>
<td>0.508</td>
</tr>
</tbody>
</table>

Replicate 3

<table>
<thead>
<tr>
<th>Model</th>
<th>Replicate 3</th>
<th>Replicate 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC_{IRWLS_c}</td>
<td>( w_i )</td>
</tr>
<tr>
<td>logistic</td>
<td>-117.02</td>
<td>1.5e-07</td>
</tr>
<tr>
<td>Bernoulli</td>
<td>-147.05</td>
<td>0.492</td>
</tr>
<tr>
<td>Gompertz</td>
<td>-147.11</td>
<td>0.508</td>
</tr>
</tbody>
</table>

Table 4: Comparison of AIC_{IRWLS_c} values and weights for each candidate model for the green algae data
Figure 14: This figure graphs experimental algae data along with the fitted models given by equations (34) - (36). For each model, the AICc (equation (38)) and the model weight (equation (29)) are given.

6.4 Lygus hesperus Population Dynamics: Importance of Nymph Deaths

Our next example concerns the methodology for insect counting and revolves around the question of how detailed field counts need to be in environmental studies. In particular, in labor intensive efforts to track pests in field environments, is it necessary to keep track of nymph mortality? We investigate this question in the context of monitoring of Lygus hesperus, a prevalent insect in California which feeds on cotton and other plants [40]. Given a robust data set of L. hesperus counts from over 500 Californian fields over several years, we hope to provide more information about the L. hesperus and direct future research relating to its effects on crops. But first we hope to inform the data collection process carried out by farmers and their associates. We propose 2 ordinary differential equation models, one of which features nymph mortality, estimate parameters for each model, and perform model comparison techniques to determine which model is more appropriate, given the population dynamics and the nature of the data.

Our main database consists of over 1500 data sets (comprising over 500 distinct fields) of L. hesperus counts. One data set is characterized by the following: a designated pesticide control advisor (PCA) counts the number of L. hesperus found in a sample of field sweeps (50 large net sweeps = 1 sample) at intermittent times from early June to early August. Some PCAs distinguish between nymph and adult specimens whereas others simply count total insects caught; therefore only some data sets consist of nymph and adult counts for each time point. In addition, the fields can vary by the absence or application (and variety) of pesticide treatments. We first assumed that field counts are
independent between years (i.e. if one field is sampled in 2004 and 2005, we consider these data sets to be independent).

To narrow down this vast collection of data, and to start with the simplest case, we chose a sub-collection of the data consisting only of data sets corresponding to fields that were untreated by pesticides for a minimum of 2 uninterrupted months, in which PCAs counted both nymphs and adults [6]. There were at least 40 data sets of this nature. By starting with this subset of data, we are able to study the insect population dynamics which are not directly affected by pesticides. We note that pesticide usage on nearby crops can have an indirect effect on these crops but choose to ignore this potential effect for now, as it is largely unknown and variable. In addition, this allows us to propose a 2-dimensional population model. These pesticide free counts occurred between the months of June and August. In this model, we choose 6 of these data sets as a preliminary study.

6.4.1 Mathematical Models

We begin by assuming there are 2 distinct population classes: nymphs and adults. We denote their populations as $x_1(t)$ and $x_2(t)$ respectively, where $t$ is time measured in months ($t \geq 0$). Given this particular insect and data collection scheme, we consider $t = 0$ to mean June 1 (as no observations in our data sets are made before this date). As noted above, we will ignore the effect of pesticides on the population, and consider the population dynamics of *L. hesperus* in an untreated environment. We do not assume a closed population (i.e. $\frac{dx}{dt} = \frac{dx_1}{dt} + \frac{dx_2}{dt} \neq 0$). In addition, it is assumed that there are at least 3 generations per year. One source [40] stated that the generation times (of the nymphs) varied depending on the time of year. They reported 3 generations of *L. hesperus* nymphs in summer 1998, which is depicted in Table 5. This information may be useful when analyzing parameter estimates.

<table>
<thead>
<tr>
<th>Generation</th>
<th>Gen I</th>
<th>Gen II</th>
<th>Gen III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dates</td>
<td>May 20- July 8</td>
<td>July 15- Aug 5</td>
<td>Aug 12 - Sept 23</td>
</tr>
<tr>
<td>Approximate timespan</td>
<td>6.5 weeks</td>
<td>3 weeks</td>
<td>6 weeks</td>
</tr>
</tbody>
</table>

Table 5: Example of nymphal development time

We first consider a simple 2-dimensional ordinary differential equation model, Model A, given by

\[
\begin{align*}
\frac{dx_1(t)}{dt} &= \beta x_2(t) - \gamma x_1(t) \\
\frac{dx_2(t)}{dt} &= \gamma x_1(t) - \mu_2 x_2(t),
\end{align*}
\]

(39)

where $\beta$ is the birth rate of nymphs, $\gamma$ is the transition rate of nymphs into adulthood, and $\mu_2$ is the adult death rate, all with unit $[1/t]$. Clearly, Model
A assumes that there is no (or essentially trivial) nymph mortality. However, Model B assumes a non-trivial nymph mortality rate, $\mu_1$, and is given by

$$\frac{dx_1(t)}{dt} = \beta x_2(t) - (\gamma + \mu_1)x_1(t)$$
$$\frac{dx_2(t)}{dt} = \gamma x_1(t) - \mu_2 x_2(t),$$

(40)

where $\mu_i$ is the death rate for $x_i$, $i = 1, 2$. For both model A and B, initial conditions $X_1 = (x_1(t_1), x_2(t_1)) := (x_{1,1}, x_{2,1})$ are unknown in our data sets and must be estimated. We remark that $t_1$, the time of the first observation, varies between data sets but is known. Our goal is to estimate parameters in Model B, $\tilde{q} = \{\beta, \gamma, \mu_1, \mu_2, x_{1,1}, x_{2,1}\}$ using our chosen data sets (note that the parameters in Model A reduce to those in Model B, with the constraint that $\mu_1 = 0$). We used MATLAB’s constrained optimization tool, fmincon and both ordinary least squares (OLS) and weighted least squares (WLS) techniques.

In addition to our main database, we have a supplementary set of data consisting of 9 fields in which nymphs and adults counts were recorded by PCAs and subsequently counted again by our team within 7 days of the original count. Although these 9 fields are not the same as those found in the large database, they are characteristically similar, and thus can be used to make an inference on data collection error when performed by PCAs. As previously mentioned, some PCAs do not bother to count nymphs or distinguish between age classes. This is largely because the nymphs are smaller and thus harder to see amidst net debris and because the nymphs tend to cling more tightly to the plants during sweeps.

Our early analysis of the available data sets lead us to believe that using weighted least squares in our parameter estimation is important. To estimate parameters, one must search within an admissible parameter space $\Omega$ for the model parameters that produce a model output most similar to the data. In other words, one must minimize the cost functional, $J_n$ defined to be

$$J_n = J_n(y, q) = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{k} \omega_j(y_{ij} - m_{ij})^2,$$

(41)

where $y_{ij}$ is the data point from the $j^{th}$ class at the $i^{th}$ time point, and $m_{ij}$ is the model output for the $j^{th}$ class at the $i^{th}$ time point, given a parameter estimate. Between fields, $n$ (the number of vector observations in a sample) is variable. Note that $k = 2$ (the total number of classes within the data), and $j = 1$ corresponds to the nymph class and $j = 2$ corresponds to the adult class so that the total number of data points is $2n$. Let $W = \{\omega_1, \omega_2\}$. There are formal ways of choosing $W$, but we only discuss here some basic choices. If we choose $W = \{0, 1\}$, we are ignoring the nymph counts in the search for the best parameter estimates for the model and do not expect this to be useful for the
questions we address here. If we choose $W = \{0.5, 1\}$, we are giving less weight to the nymph class than to the adult class. Note that if we choose $W = \{1, 1\}$, our efforts reduce to an OLS method.

6.4.2 Parameter Estimates and Model Comparison Test

There are differing opinions among PCAs and researchers about whether both nymphs and adults need to be counted and this is the issue we wish to investigate here. The reasons for these differences are varying beliefs regarding the effect of pesticides and other factors on the *L. hesperus* populations. We seek a quantitative measure to determine whether counting both nymphs and adults (in the manner in which it is presently done) is necessary, or if it is sufficient to simply count the total number of insects. We see that the sole difference between Models A and B ((39) and (40), respectively) is the assumption of no nymph mortality in Model A. Note that model A can be more simply written as

$$\frac{dX}{dt} = \alpha x_2(t),$$

where $X(t)$ ($X = x_1 + x_2$), represents the total number of *L. hesperus* at time $t$, and $\alpha = \beta - \mu_2$. This simpler model is exponential in nature. One may wonder how this model could possibly be exponential in nature, when there are 2 state variables, $X$ and $x_2$ in one differential equation. We found consistently among PCA-collected data that the nymph counts were almost always zero. Therefore, given the current collection strategies, $X \approx x_2$, and (42) truly becomes an exponential growth model. A natural question is the following: by allowing nymph mortality to be non-zero, does the model provide a better fit the data? (and hopefully a better description of what actually is occurring in the fields). To address this question, using the model comparison test detailed above, we can test the null hypothesis: is the true set of parameter values, $q_0$, in a constrained subset $\Omega_H$ of $\Omega$, which requires that $\mu_1 = 0$, or do we obtain a statistically significant better fit allowing $\mu_1 \neq 0$? Our constraint set $\Omega$ is defined by the constraints on the physical meaning of the parameters, $\beta, \gamma, \mu_1$, and $\mu_2$. Although the only true constraint is that each of these values must be non-negative, we chose impose the further constraint that each be less than 100. (We chose 100 because we found it unlikely that any true parameter value would fall above this upper bound, and this choice greatly speeds up the parameter estimation process by refining the search space.) Therefore, $\Omega = [0, 100] \times [0, 100] \times [0, 100] \times [0, 100]$. To obtain the equivalent of the constraint that $\mu_1 = 0$ (which simply means that there is no nymph mortality), we took $\Omega_H$ as simply $\Omega_H = [0, 100] \times [0, 100] \times \{0\} \times [0, 100]$. Therefore, by testing the null hypothesis $H_0 : q_0 \in \Omega_H$, we can determine with a definitive amount of confidence whether we can ignore nymph mortality and thus use a simple model such as Model A to describe the data.
6.4.3 Results

We chose to perform this analysis on 6 data sets, with 2 choices of $W$: $W_1 = \{1,1\}, W_2 = \{0.5,1\}$

Even if the nymph data has a large degree of error, it is unreasonable to expect an optimization routine to find nymph population parameters such as $\gamma$ and $\mu_1$ with a complete absence of nymph data. In addition, we found that the weights that returned the best estimates of initial conditions among all data sets were $W_1 = \{1,1\}$ and $W_2 = \{0.5,1\}$. Confidence to reject the null hypotheses $H_0$ ranged from a high of 15.77 to 0 with an average of 4.778. Results are summarized in the Table 6 below. We note that the cost functional values minimized over $\Omega$ and over $\Omega_H$ are in numerous cases the same to three or four decimal places so that computing the respective $AIC_c$ values is not helpful in these cases.

| Data set 1 with estimated i. c. $\{0,0.06\}$ |  |
|---|---|---|
| $W$ | $\hat{u}_n$ | $J_n(y, \hat{q}^n)$ | $J_n(y, \hat{q}^n_H)$ |
| $\{1,1\}$ | 0.002 | 0.2410 | 0.2410 |
| $\{0.5,1\}$ | 0.000 | 0.2309 | 0.2309 |

| Data set 2 with estimated i. c. $\{0.01,0.22\}$ |  |
|---|---|---|
| $W$ | $\hat{u}_n$ | $J_n(y, \hat{q}^n)$ | $J_n(y, \hat{q}^n_H)$ |
| $\{1,1\}$ | 0.0024 | 0.0302 | 0.0302 |
| $\{0.5,1\}$ | 0.0016 | 0.0256 | 0.0256 |

| Data set 3 with estimated i. c. $\{0.03,0.05\}$ |  |
|---|---|---|
| $W$ | $\hat{u}_n$ | $J_n(y, \hat{q}^n)$ | $J_n(y, \hat{q}^n_H)$ |
| $\{1,1\}$ | 0.0396 | 0.1556 | 0.1557 |
| $\{0.5,1\}$ | 0.014 | 0.0895 | 3.0895 |

| Data set 4 with estimated i. c. $\{0.004,0.05\}$ |  |
|---|---|---|
| $W$ | $\hat{u}_n$ | $J_n(y, \hat{q}^n)$ | $J_n(y, \hat{q}^n_H)$ |
| $\{1,1\}$ | 0.0002 | 0.5660 | 0.5660 |
| $\{0.5,1\}$ | 0.0022 | 0.3931 | 0.3931 |

| Data set 5 with estimated i. c. $\{0.06,0.15\}$ |  |
|---|---|---|
| $W$ | $\hat{u}_n$ | $J_n(y, \hat{q}^n)$ | $J_n(y, \hat{q}^n_H)$ |
| $\{1,1\}$ | 0.0044 | 0.1879 | 0.1880 |
| $\{0.5,1\}$ | 0.003 | 0.1600 | 0.1600 |

| Data set 6 with estimated i. c. $\{0.05,0.22\}$ |  |
|---|---|---|
| $W$ | $\hat{u}_n$ | $J_n(y, \hat{q}^n)$ | $J_n(y, \hat{q}^n_H)$ |
| $\{1,1\}$ | 0.0076 | 0.0856 | 0.0856 |
| $\{0.5,1\}$ | 0.000 | 0.0632 | 0.0632 |

Table 6: Model comparison test results for data sets 1-6.

Overall, we found compelling evidence for the untreated fields, by the model comparison test, that we should NOT reject the null hypothesis. In other words, it may be reasonable to ignore nymph mortality (i.e., just count total number of L. hesperus and not distinguish between nymphs and adults), which would
greatly simplify the model, as given in (42), as well as the data collection process. It is important to note that this conclusion may not be reasonable for data sets in which pesticide treatment was used. Our findings to date suggest it may be sufficient to only count the total number of *L. hesperus*, rather than distinguish between adults and nymphs.

7 Concluding Remarks

In this review paper we have discussed the use of sensitivity theory, and model comparison techniques including the use of Akaike type criteria to investigate modeling fits to experimental data sets. A number of examples were given to illustrate use of these methods. Among other examples where some of the ideas and methods presented here have played an important role include [9, 15, 16, 25, 26, 27, 59, 60].

Acknowledgements

This research was supported in part by the National Institute on Alcohol Abuse and Alcoholism under grant number 1R01AA022714-01A1, and in part by the Air Force Office of Scientific Research under grant number AFOSR FA9550-15-1-0298.

References


