Material model inference from experimental data

Ken Hanson

CCS-2, Methods for Advanced Scientific Simulations
Los Alamos National Laboratory

This presentation available at
http://www.lanl.gov/home/kmh/

Overview

- Bayesian analysis
  - ▸ appropriateness for analyzing physics experiments
- Likelihood analysis
  - ▸ relation to chi-squared
  - ▸ estimation of parameters and their uncertainties
- Material characterization experiments
- Data analysis using Zerilli-Armstrong model
  - ▸ difficulties in matching data
  - ▸ importance of expertise to obtain satisfactory result
  - ▸ systematic effects, uncertainties

Acknowledgments

Collaborators

- Shuh-Rong Chen, MST-8
- François Hemez, ESA-WR

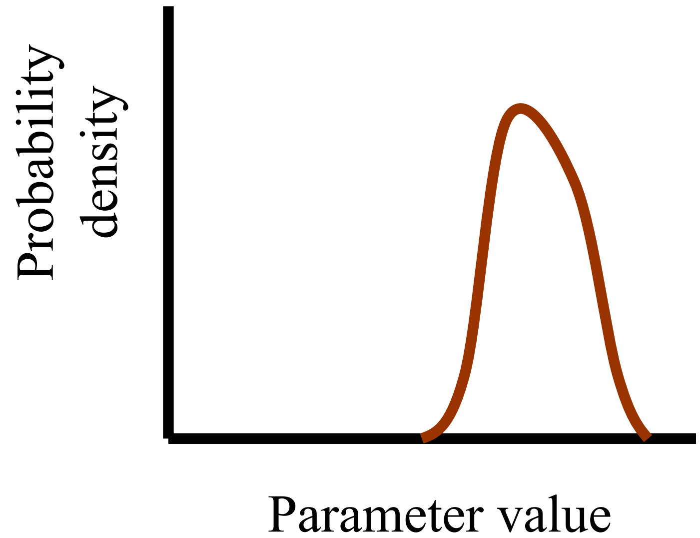
Discussions

- Larry Hull, Eric Ferm, DX-3
- Chris Romero, Tom Duffy, DX-7
- Paul Maudlin, T-3
- Mark Anderson, ESA-WR
- Kathy Campbell, Mike McKay, Dave Higdon, Alyson Wilson, Mike Hamada, D-1

Uncertainties and probabilities

- Uncertainties in parameters are characterized by probability density functions (pdf)
- Probability interpreted as quantitative measure of "**degree of belief**"
- This interpretation sometimes referred to as "subjective probability"
- Rules of classical probability theory apply

Probability density function



Bayesian analysis of experimental data

- Bayesian approach
  - focus is as much on uncertainties in parameters as on their best (estimated) value
  - appropriate for Uncertainty Quantification (UQ)
  - use of prior knowledge, e.g., previous experiments, modeling expertise, subjective
  - model checking –
    does model agree with experimental evidence?
  - compatible with scientific method
- Goal is estimation of **model parameters and their uncertainties**

Bayesian analysis of experimental data

- Bayes theorem

$$p(\boldsymbol{a} \mid \boldsymbol{d}, I) \propto p(\boldsymbol{d} \mid \boldsymbol{a}, I)\ p(\boldsymbol{a} \mid I)$$

  - where
    d is the vector of measured data values
    a is the vector of parameters for model that predicts the data
  - $p(\boldsymbol{d} \mid \boldsymbol{a}, I)$ is called the **likelihood** (of the data given the true model and its parameters)
  - $p(\boldsymbol{a} \mid I)$ is called the **prior** (on the parameters ***a***)
  - $p(\boldsymbol{a} \mid \boldsymbol{d}, I)$ is called the **posterior** – fully describes uncertainty in the parameters
  - *I* stands for whatever **background information** we have about the situation, results from previous experience, our expertise, and the model used

Bayesian analysis – role of the prior

- The prior in Bayes theorem distinguishes Bayesian analysis from “traditional” frequentist statistics
- The prior can be chosen to be non-informative
  - examples: uniform, uniform in log, maximum entropy
  - to reflect complete lack of knowledge about situation, to avoid biasing result
  - to be objective(?); often appropriate for physics analyses
- The prior can be chosen to be informative
  - to enforce physical constraints, e.g., nonnegativity (density)
  - to incorporate information from previous experiments
  - to reflect expert knowledge (elicitation process)
- Choice of prior is subject to discussion and review

The model and parameter inference

- We write the model as

$$y = y(\boldsymbol{x}, \boldsymbol{a})$$

  - where y is a physical quantity, which is modeled as a function of the independent variables $\boldsymbol{x}$ and $\boldsymbol{a}$ represents the model parameters
- In inference, the aim is to determine:
  - the parameters $\boldsymbol{a}$ from a set of n measurements d_i of y under specified conditions x_i
  - and the uncertainties in the parameter values
- This process is called parameter inference, model fitting (or regression) but often uncertainty analysis not done, as in parameter estimation

The likelihood and chi-squared

- The form of the likelihood $p(\boldsymbol{d}\,|\,\boldsymbol{a}, I)$ depends on how we model the uncertainties in the measurements $\boldsymbol{d}$.
- Assuming the error in each measurement d_i is normally (Gaussian) distributed with zero mean and variance of σ_i^2, and the errors are statistically independent,

$$p(\boldsymbol{d}\,|\,\boldsymbol{a}) \propto \prod_i \exp\left[-\frac{[d_i - y_i(a)]^2}{2\sigma_i^2}\right]$$

  where y_i is the value predicted for parameter set $\boldsymbol{a}$
- The above exponent is related to chi squared

$$\chi^2 = -2\log\left[p(\boldsymbol{d}\,|\,\boldsymbol{a})\right] = \sum_i \left[\frac{[d_i - y_i(a)]^2}{\sigma_i^2}\right]$$

- For this error model, likelihood is $p(\boldsymbol{d}\,|\,\boldsymbol{a}) \propto \exp(-\tfrac{1}{2}\chi^2)$

Likelihood analysis

- For a non-informative **flat prior**,
  the posterior is proportional to the likelihood
- Given the relationship between chi-squared and the likelihood, posterior is

$$p(\boldsymbol{a} \,|\, \boldsymbol{d}) \propto p(\boldsymbol{d} \,|\, \boldsymbol{a}) \propto \exp(-\tfrac{1}{2}\chi^2)$$

- Thus, parameter estimation based on **maximum likelihood** is equivalent to that based on **minimum chi squared**

Characterization of chi-squared

- Expand vector $\boldsymbol{y}$ around $\boldsymbol{y}^0$:

$$y_i = y_i(x_i, \boldsymbol{a}) = y_i^0 + \sum_j \left.\frac{\partial y_i}{\partial a_j}\right|_{a^0} (a_j - a_j^0) + \cdots$$

- The derivative matrix is called the *Jacobian*, $\boldsymbol{J}$
- Estimated parameters $\hat{a}$ minimize χ^2 (MAP estimate)
- As a function of $\boldsymbol{a}$, χ^2 is quadratic in $\boldsymbol{a} - \hat{\boldsymbol{a}}$

$$\chi^2(\boldsymbol{a}) = \tfrac{1}{2}\left(\boldsymbol{a} - \hat{\boldsymbol{a}}\right)^{\mathrm{T}} \boldsymbol{K} \left(\boldsymbol{a} - \hat{\boldsymbol{a}}\right) + \chi^2(\hat{\boldsymbol{a}})$$

where K is the curvature matrix (aka the *Hessian*);

$$\left[\boldsymbol{K}\right]_{jk} = \left.\frac{\partial^2 \chi^2}{\partial a_j \partial a_k}\right|_{\hat{a}} = \boldsymbol{J}\boldsymbol{J}^{\mathrm{T}}$$

Parameter inference

- Posterior $p(\boldsymbol{a} \mid \boldsymbol{d}, I)$ can be written as

$$p(\boldsymbol{a} \mid \boldsymbol{d}) = \frac{1}{\det[\boldsymbol{C}]\,(2\pi)^{n/2}} \exp\left[-\tfrac{1}{2}\left(\boldsymbol{a}-\hat{\boldsymbol{a}}\right)^{\mathrm{T}} \boldsymbol{C}^{-1}\left(\boldsymbol{a}-\hat{\boldsymbol{a}}\right)\right]$$

- From known properties of Gaussian distribution, covariance matrix for parameter uncertainties is

$$\mathrm{cov}(\boldsymbol{a}) = \left\langle (\boldsymbol{a}-\hat{\boldsymbol{a}})(\boldsymbol{a}-\hat{\boldsymbol{a}})^{\mathrm{T}} \right\rangle \equiv \boldsymbol{C} = 2\boldsymbol{K}^{-1}$$

- Thus, the chi-squared functionality provides the basis for inference about parameters $\boldsymbol{a}$
- Recall assumptions:
  - linearized model holds for measured quantities ($y = f(\boldsymbol{x},\boldsymbol{a})$)
  - meas. errors indep. & Gaussian distrib. with known variance
  - uniform prior on parameters $\boldsymbol{a}$

Model checking – goodness of fit

- Chi-squared analysis is based on assumption that measurement errors Gaussian distributed, independent
- After minimum χ^2 is found, one can check whether the value of χ^2 is consistent with that assumption
- Chi-squared distribution table gives probability p for obtaining the observed χ^2 value or higher
- Reduced chi-squared is χ^2/ν, where ν is
  # degrees of freedom = # data – # parameters
- Property of χ^2 distribution: $p = 50\%$ is near $\chi^2/\nu = 1$
- Checks self-consistency of models used to explain data (weakly)

Model checking – goodness of fit

- Check of chi-squared value only weakly confirms validity of models used
- Chi-squared value depends on numerous factors:
  - assumption that errors follow Gaussian distribution and are statistically independent
  - proper assignment of standard deviation of errors
  - correctness of model used to calculate measured quantity
  - measurements correspond to calculated quantity (proper measurement model)
- Thus, a reasonable chi-squared *p* value does not necessarily mean everything is OK, because there may be compensating effects

Analysis of multiple data sets

- To combine the data from multiple data sets into a single analysis, the combined likelihood is

$$p_{all}(\boldsymbol{d} \mid \boldsymbol{a}) \propto \prod_k p(\boldsymbol{d}_k \mid \boldsymbol{a})$$

  where $p(\boldsymbol{d}_k \mid \boldsymbol{a}, I)$ is likelihood from kth data set
  - assumes the uncertainties in different data sets are statistically independent
- Thus, because $\chi^2 = -2\log\left[p(\boldsymbol{d} \mid \boldsymbol{a})\right]$, just add χ^2s from each data set

$$\chi^2_{all} = \sum_k \chi^2_k$$

Inclusion of Gaussian priors

- To include priors, use Bayes theorem

$$p(\mathbf{a} \mid \mathbf{d}, I) \propto p(\mathbf{d} \mid \mathbf{a}, I)\ p(\mathbf{a} \mid I)$$

- For a Gaussian prior on a parameter a

$$p(a \mid I) = \frac{1}{\sigma_a\ (2\pi)^{1/2}} \exp\left[-\frac{(a - \tilde{a})^2}{2\sigma_a^2}\right]$$

  where $\tilde{a}$ is the default value for a and σ_a^2 is the assumed variance

- The minus-log-posterior for the parameter a is

$$-\log p(a \mid d, I) = \varphi(a) = \tfrac{1}{2}\chi^2 + \frac{(a - \tilde{a})^2}{2\sigma_a^2}$$

Motivating example

- Problem statement
  - design containment vessel using high-strength steel, HSLA 100
  - one design criterion relates to wall penetration by schrapnel
  - predict degree of wall penetration by specified projectile
  - estimate uncertainty in this prediction to estimate safety factor
- Our present goal is to determine for HSLA 100 the parameters and their uncertainties for the Zerilli-Armstrong plastic strength model for
  - strains up to fracture for use at
  - room temperature
  - high strain rates
- These conditions match the intended application

HSLA 100

- Material under study is the high-strength, low-alloy steel designated as HSLA 100
  - used in critical structural applications
- Manufacture of this steel is done under tight specifications
  - composition is certified and uniform
  - properties should be quite reproducible
- For most metals, processing can affect properties of the material
  - processing often involves rolling of billets into sheets and subsequent heat treatment

Stress-strain relation for plastic deformation

- Zerilli-Armstrong model describes strain rate- and temperature-dependent plasticity in terms of stress σ (or s) as function of plastic strain ε_p

$$\sigma = \alpha_1 + \alpha_5 \varepsilon_p^{\alpha_6} + \alpha_2 \exp\left[\left(-\alpha_3 + \alpha_4 \log \frac{\partial \varepsilon_p}{\partial t}\right) T\right]$$

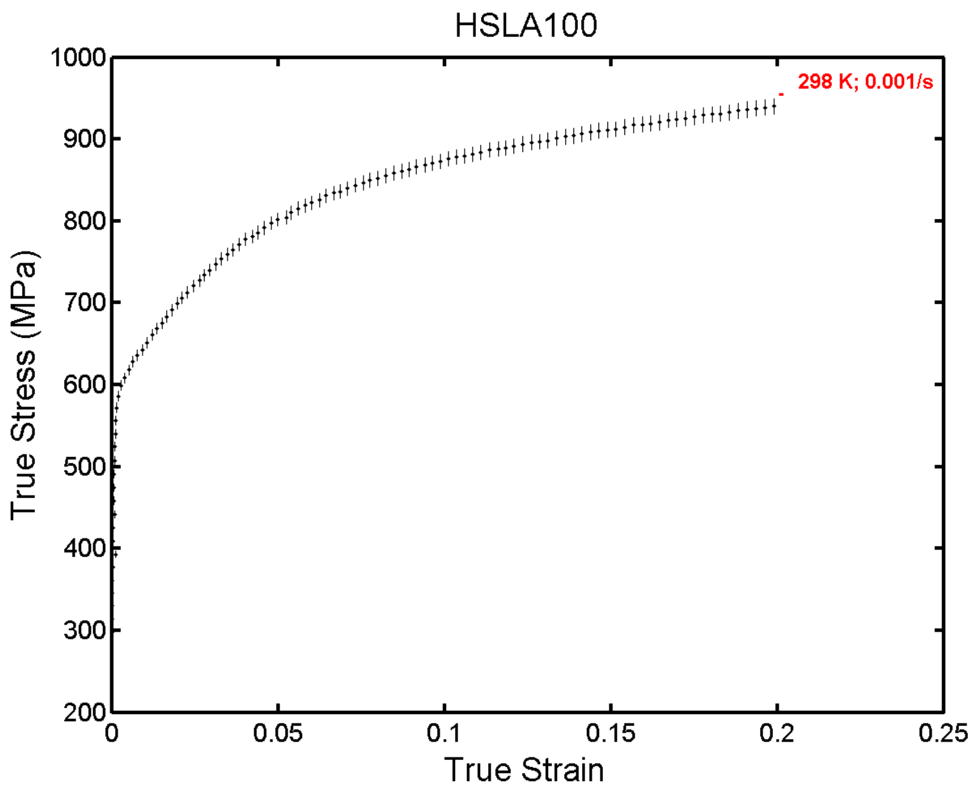
- Six parameters -
  - 2 parameters (α_5 & α_6) specify dependence of stress on strain
  - 4 remaining parameters specify additive offset as function of temperature and strain rate
- Z-A formula based on dislocation mechanics model
  - may not hold for all experimental conditions

Material characterization experiments

- Quasi-static experiments
  - subject material specimen to tension or compression
  - measure force and corresponding sample length
  - convert to true stress and true strain
- Hopkinson-bar experiments
  - send shock wave into thin disc of material
  - measure length of specimen as a function of time
  - interpret in terms of true stress and true strain at a calculated strain rate using simulation code
  - correct measured temperature of specimen for work done on sample, assuming adiabatic process

Quasi-static experiments

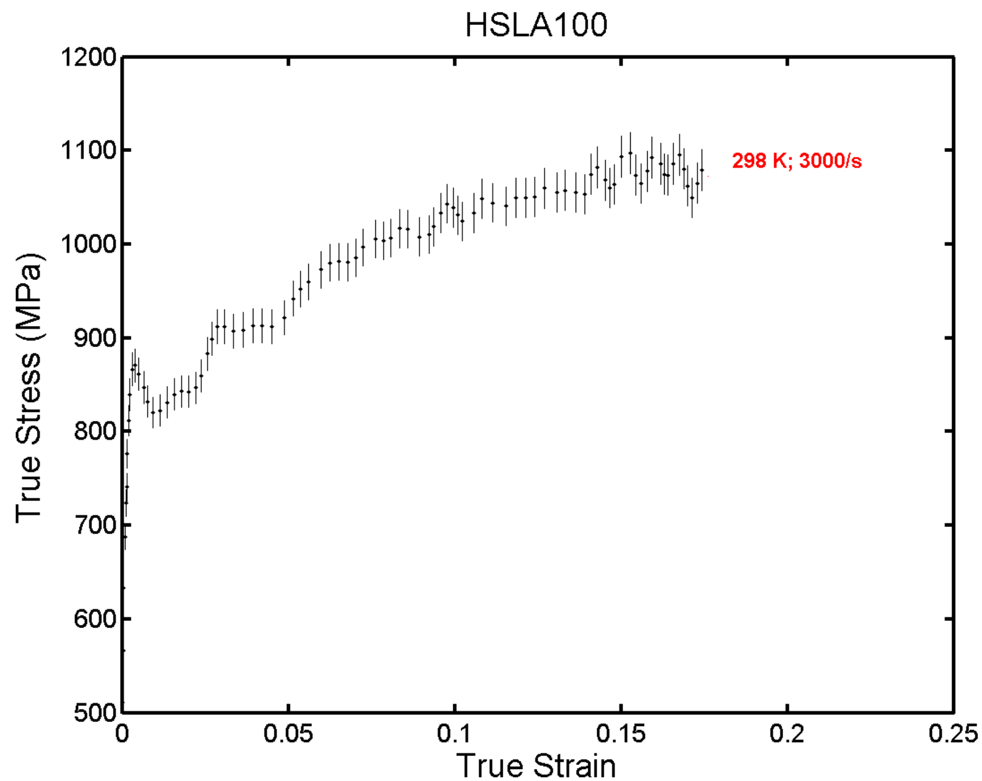
- Data from quasi-static compression experiments tend to be of high quality
- Systematic uncertainties in the basic measurements should be very small
- Example shows data at room temperature
  - elastic region
  - yield stress
  - plastic region
- Error bars shown are 1% or ~ 10 Mpa
  - error bars seem too large!



†data supplied by S-R Chen, MST-8

Hopkinson-bar experiments

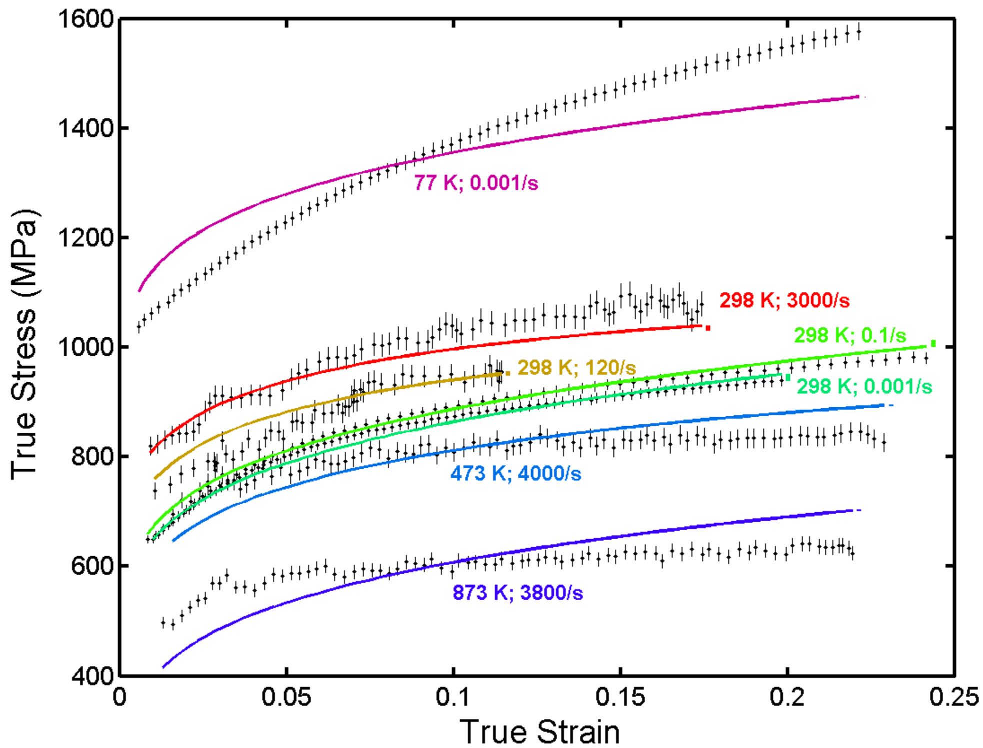
- Data from Hopkinson-bar experiments tend to be of medium quality
- Systematic uncertainties in the basic measurements should be small
- Observe artifacts in the data
  - arise from reflected shocks
  - should exclude these
- Must reply on simulation code to calculate strain rate
- Error bars shown are 2% or ~20 MPa
  - plausible uncertainty level



†data supplied by S-R Chen, MST-8

Fit ZA model to all data

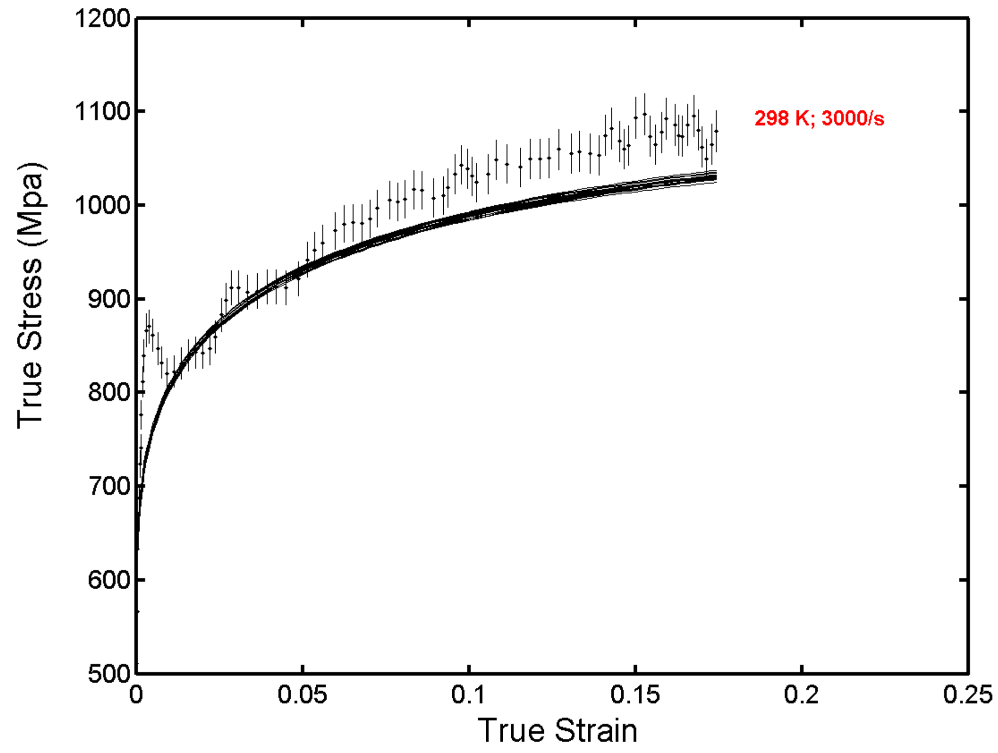
- 7 data sets at various strain rates and temperatures
- Fit to all data above elastic region or after first bump in Hopkinson-bar data
- Model does not reproduce stress-strain curves at high and low temperatures
- Fit is far from expt. measurements for target conditions of room temp., high strain rate
- Uncertainties are highly correlated



†data supplied by S-R Chen, MST-8

Monte Carlo from posterior

- Use Monte Carlo technique to draw random ZA parameter vectors from their uncertainty distribution
- Plot corresponding curves for room temperature, high strain rate and compare to measurements
- Conclude that the parameters inferred from last slide do not plausibly represent the data for target conditions

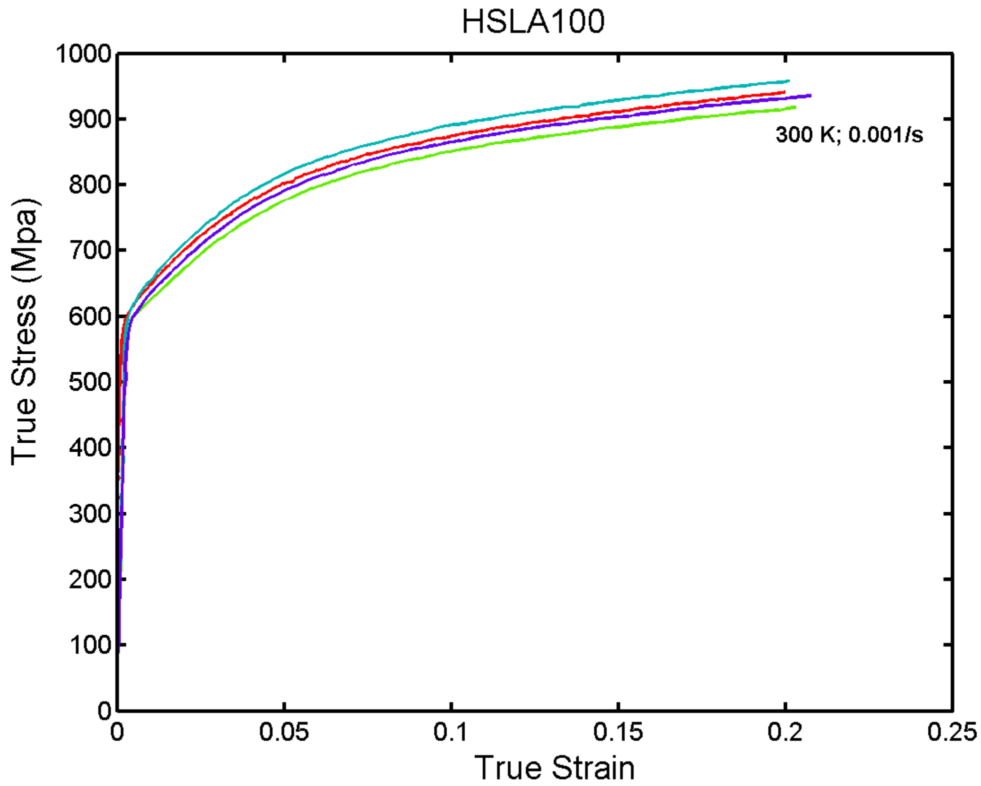


Refine analysis to accommodate data

- Need to improve analysis for intended operating conditions (moderate strain, high strain rate, and room temperature)
- Approach is
  - limit the data for high and low temps to low strain region (<0.06); reasoning is that dislocation mechanics behavior at high strain values is clearly different than at room temperature, but would like to capture behavior near yield points.
  - can not just ignore these data – they are needed to determine temp and strain rate dependence in ZA model
  - strain rate dependence seen in experimental data do not conform with ZA model, or any other smoothly varying model
  - inclusion of sample-to-sample uncertainties into analysis accomodates these differences
  - treat sample-to-sample variability as systematic uncertainty

Repeated experiments

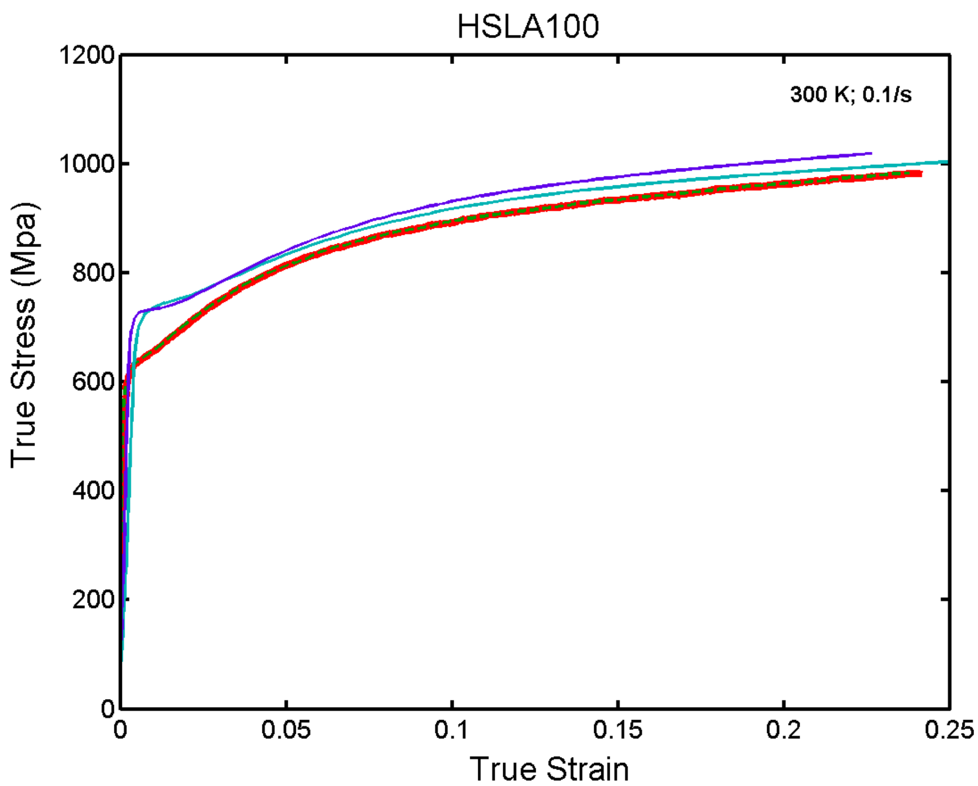
- Repeated experiments
  - stability of apparatus
  - indication of random component of error
  - may or may not indicate systematic error
- Figure shows curves obtained from four samples taken from random positions in thick plate
- Sample-to-sample rms deviation is around 20 MPa at strain of 0.1
- Treat this variability as **systematic uncertainty**



†data supplied by S-R Chen, MST-8

Repeated experiments

- Figure shows curves from four samples
  - nearly identical response for two taken from nearby position and tested together (red and green dashed lines)
  - but disagree with previous tests on samples from different stock, perhaps caused by different processing
- Observe sample-to-sample differences of around 20 MPa for strains > 0.03
- Treat this variability as **systematic uncertainty**



†data supplied by S-R Chen, MST-8

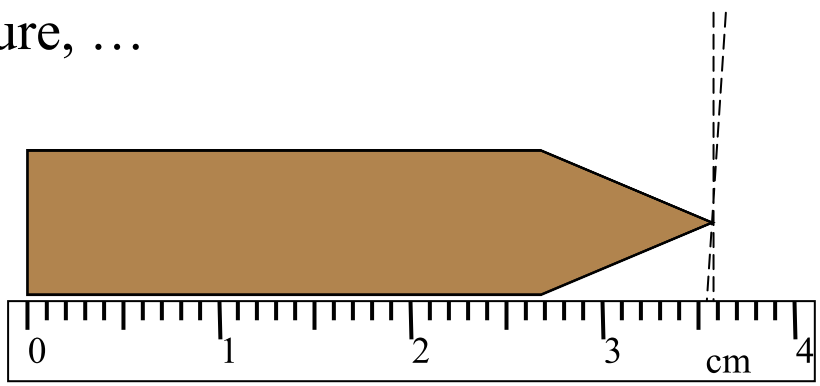
Types of uncertainties in measurements

- Two major types of errors
  - ► random error – different for each measurement
    - in repeated measurements, get different answer each time
    - often assumed to be statistically independent, but often aren't
  - ► systematic error – same for each measurement within a group
    - component of measurements that remains unchanged
    - for example, caused by error in calibration or zeroing
- Nomenclature varies
  - ► physics – random error and systematic error
  - ► statistics – random and bias
  - ► metrology standards (NIST, ASME, ISO) –
    random and systematic uncertainties (now)

Types of uncertainties in measurements

- Simple example – measurement of length of a pencil
  - random error
    - interpolation between ruler tick marks
  - systematic error
    - accuracy of ruler's length; manufacturing defect, temperature, …
- Parallax
  - reading depends on how person lines up pencil tip
  - random or systematic error?

    depends on whether measurements always made by same person in the same way or made by different people



Include offsets for each data set

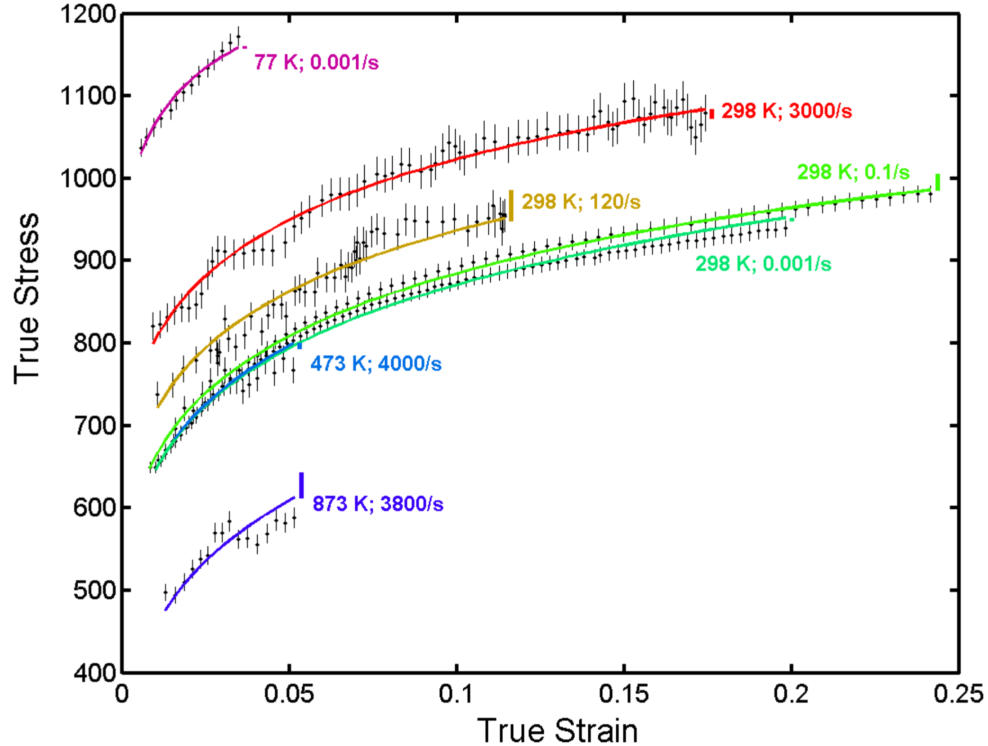
- Represent offset of *k*th data set with a parameter Δ_k
- Treat offset as **systematic effect** for each curve, but as random effect when combining curves
- Information about Δ_k is a prior – Gaussian distributed
- Assume that most probable value of Δ_k is zero and that uncertainty distribution has an rms deviation of σ_k
- Then, the posterior is

$$-\log p(\boldsymbol{a} \mid \boldsymbol{d}, I) = \varphi(\boldsymbol{a}) = \tfrac{1}{2}\sum_k \chi_k^2 + \tfrac{1}{2}\sum_k \frac{\Delta_k^2}{\sigma_k^2}$$

- For HSLA 100 analysis, we have 7 data sets and ZA model has 6 parameters; thus 13 variables in fit

Fit ZA model to selected data

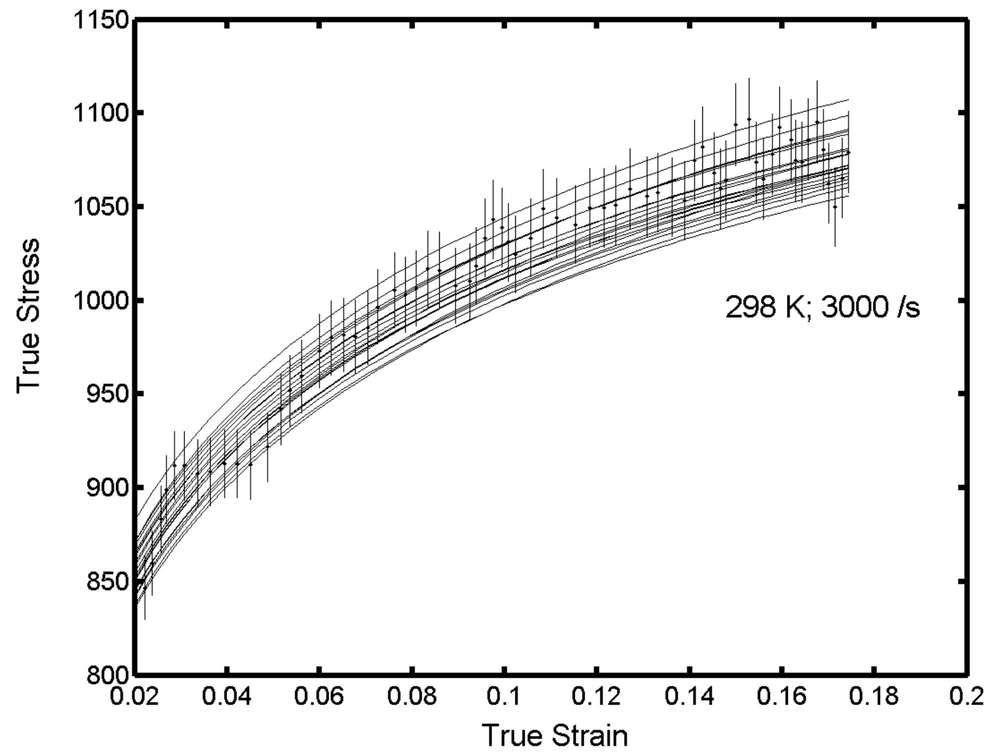
- Use data above elastic region or after first bump in Hopkinson-bar data
- Additionally, restrict data at high and low temps. to low strain (near yield point)
- Add offset parameter for each curve to represent **sample-to-sample variation**
- Fit reasonably represents data for target conditions of room temp., high strain rate



†data supplied by S-R Chen, MST-8

Monte Carlo sampling from posterior

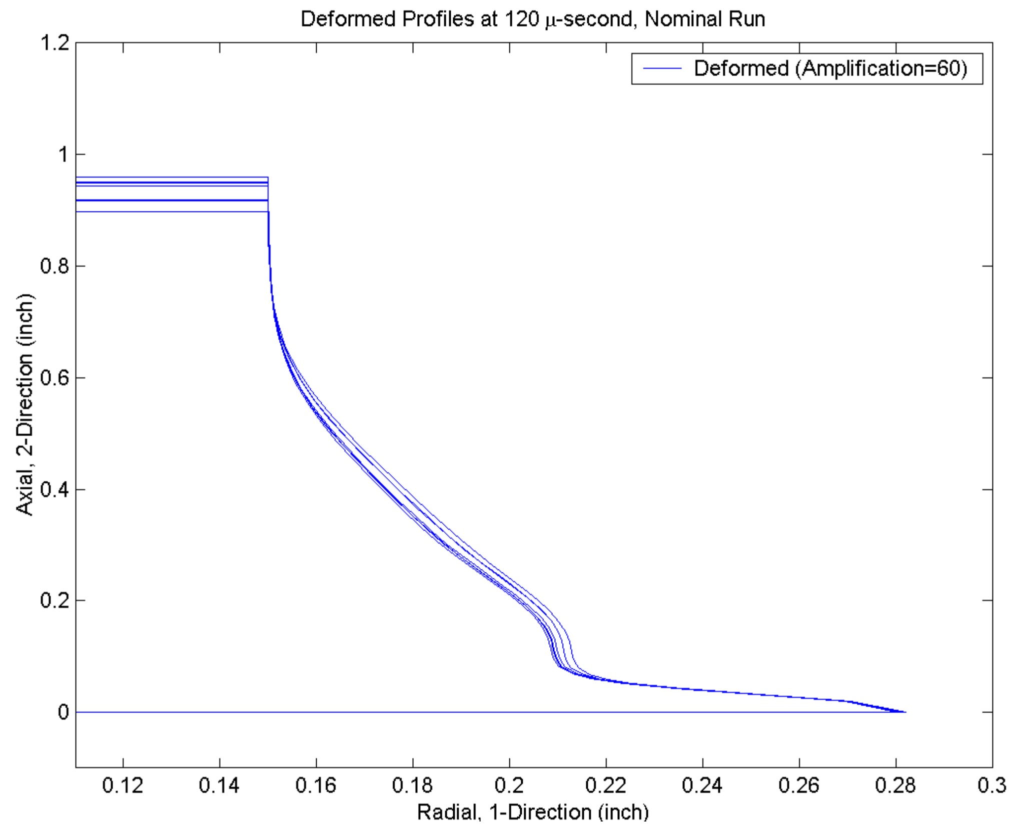
- Use Monte Carlo technique to draw random ZA parameter vectors from their uncertainty distribution
- Plot corresponding curves for room temperature, high strain rate and compare to measurements
- Conclude that parameters and their uncertainties inferred from last slide plausibly represent the data for target conditions



Monte Carlo example - Taylor test

- Use MC technique to propagate uncertainties through deterministic simulation code
  - Draw value for each of four parameters from its assumed Gaussian pdf
  - Run Abaqus code for each set of parameters
- Figure shows range of variation in predicted cylinder shape

Initial NESSUS/Abaqus results

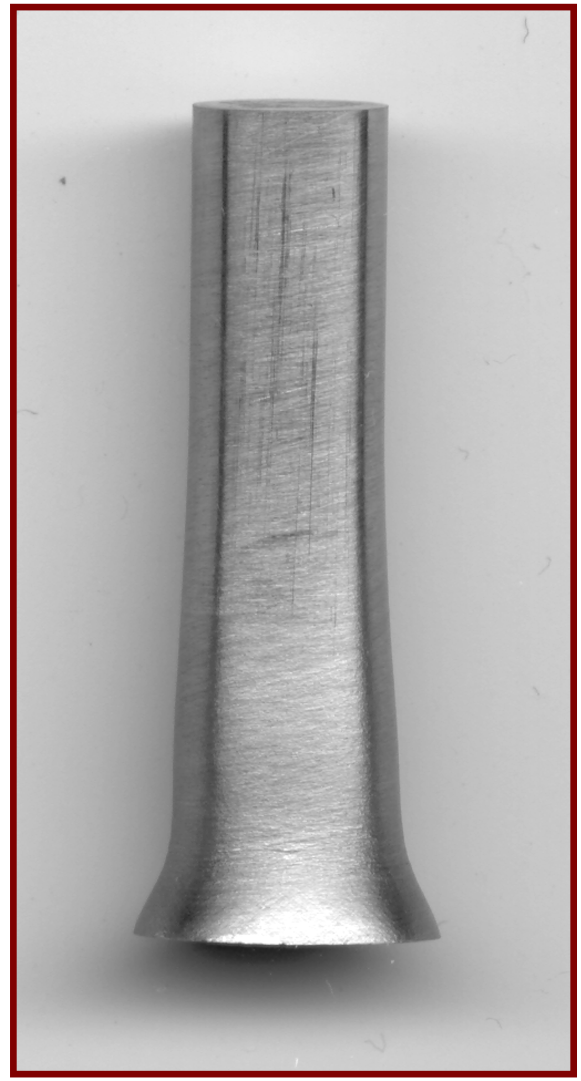


High-strength steel HSLA 100
260 m/s impact velocity

Taylor test experiment

- Taylor impact test specimen
  - high-strength steel HSLA 100
  - impact velocity = 245.7 m/s
  - dimensions, final/initial

    length   31.84 mm / 38 mm
    diameter 12.00 mm / 7.59 mm



Future work

- Demonstrate how model inference can be done through analysis of Taylor experiments using a simulation code
- Hierarchical Bayesian modeling
  - use distributions for unknown parameters, e.g., priors on variance in systematic errors (as opposed to specific, fixed values)
  - infer all parameters and their uncertainties from data & priors
  - provides more flexibility in modeling uncertainties
- Develop statistical approach to minimize uncertainty for targeted range of variables
- Application to other materials and strength models

Bibliography

- *Data Analysis: A Bayesian Tutorial*, D. S. Sivia (Clarendon, 1996); excellent introduction to Bayesian analysis for physicists & engineers
- *Data Reduction and Error Analysis for the Physical Sciences,* P. R. Bevington and D. K. Robinson (Boston, WCB/McGraw-Hill, 1992); good summary of conventional data analysis for physical scientists and engineers
- "A framework for assessing confidence in simulation codes," K. M. Hanson and F. M. Hemez, *Experimental Techniques* **25**, pp. 50-55 (2001); application of uncertainty quantification to simulation codes with Taylor test as example
- "A framework for assessing uncertainties in simulation predictions," K. M. Hanson, *Physica D* **133**, pp. 179-188 (2000); integrated approach to determining uncertainties in physics modules and their effect on predictions

Last two papers available at http://www.lanl.gov/home/kmh/