Calibration Statistics
Selecting a Statistic and Setting a Standard

1996 System Dynamics Conference

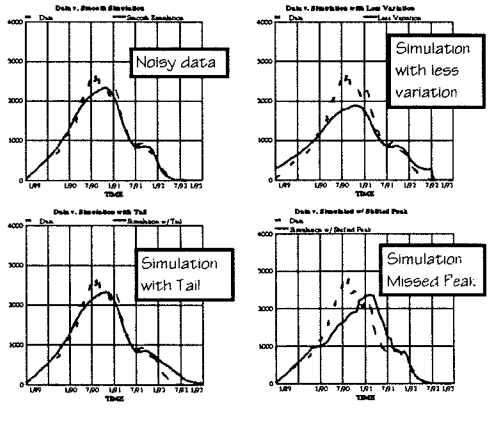
Kimberly Sklar Reichelt,
Dr. James M. Lyneis & Carl G. Bespolka
PUGH-ROBERTS ASSOCIATES

Calibration Statistics

- Why We Care About Statistics
- Alternative Statistics to Consider
- Selecting a Best Statistic
- Setting Standards

Why We Care about Statistics

- To enhance the confidence of others
- For comparison against other calibrations
- To assure that minimum standards are met
- For use in "automatic" calibration and sensitivity analysis software



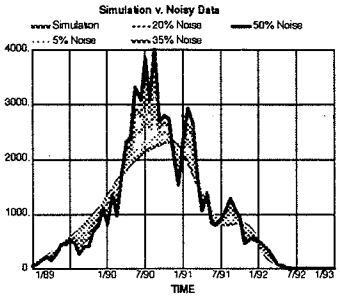
The Ideal Statistic

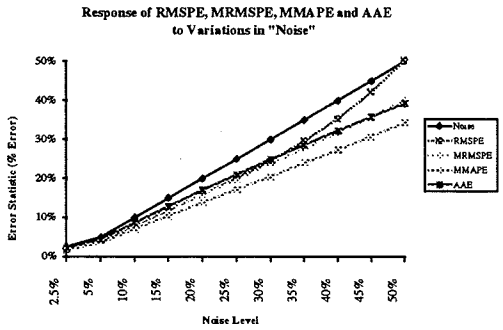
- Provides meaningful results in a variety of circumstances
- Easily interpreted
- Easy to explain ("intuitive")
- Provides a consistent measure
- Is sensitive to important differences
- Is relatively insensitive to small differences when either the simulation or the data is close to zero

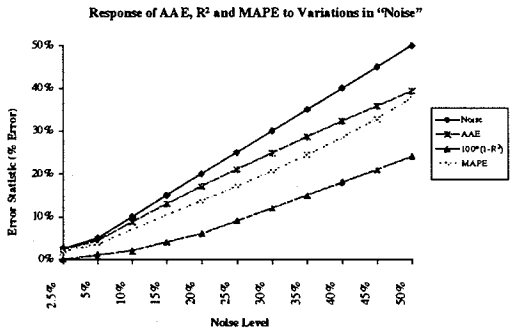
Alternative Statistics

- R^2
- Mean Absolute Percent Error (MAPE)
- Root Mean Square Percent Error (RMSPE)
- Modified Mean Absolute Percent Error (MMAPE)
- Modified Root Mean Square Percent Error (MRMSPE)
- Theil Statistics (U)
- Average Absolute Error (AAE)

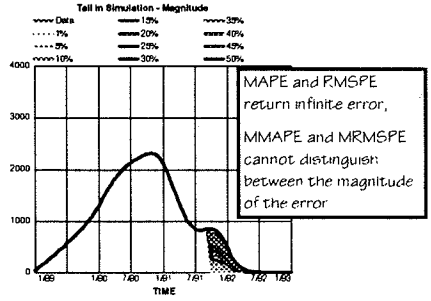
Handling Noisy Data



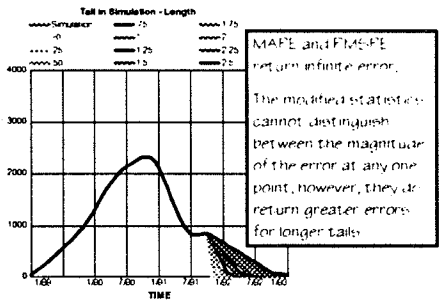




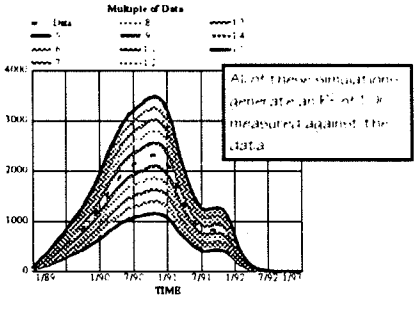
Measuring Sensitivity to Tail Height



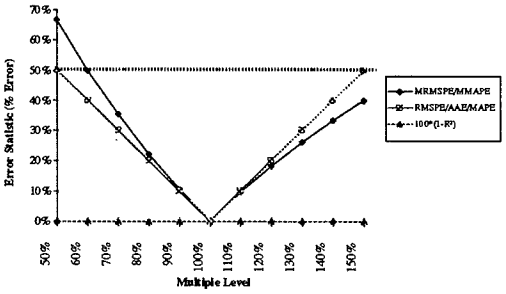
Measuring Sensitivity to Tail Length



R² Ignores *Magnitude* of the Data



Most Statistics Do Not Produce Symmetrical Results for Multiples



AAE Provides the Best Measure for Life Cycle Models

- Returns symmetric errors for bias
- Gives reasonable errors for increasing magnitude and duration of any "tails"
- Provides predictable response to increasing levels of noise
- Yields similar values to other statistics "the rest of the time"
- Is intuitive and easy to explain

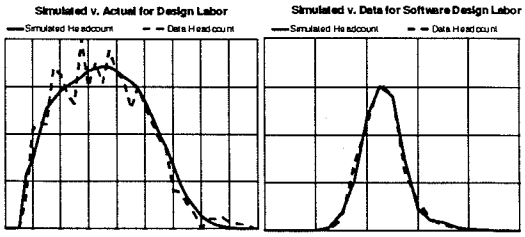
All other measures suffer some fatal flaw

How Good a Fit is Necessary?

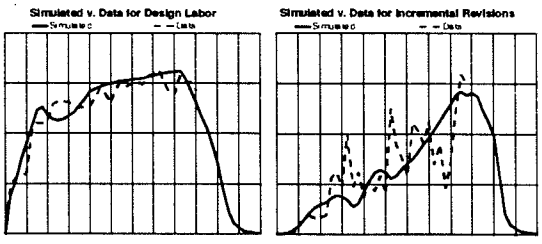
Standard will depend on ...

- The purpose for which the model is constructed
- The amount of noise in the data
- Time and budget available

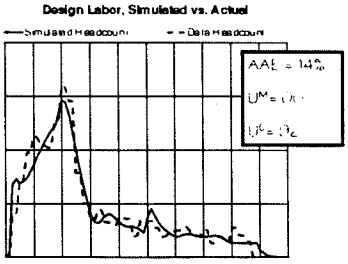
Noise Levels Vary Between Projects



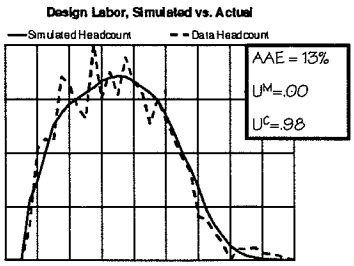
Noise Levels Vary Within a Project



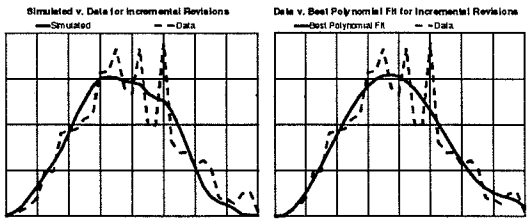
Examples of Good Fit



Examples of Good Fit



Smoothed Data May Provide a More Fair Basis for Comparison



Summary

- Statistics cannot substitute for a careful visual inspection and analysis of the fit to historical data
- We should continue to compute and report all statistics to allow comparison to other results
- The "Average Absolute Error" statistic should be used as the standard
- What fit is "good enough" is inherently subjective.
- The Theil components should be computed and monitored to ensure that "bias" is small (<=10%) and most error is due to covariance (>=70%).

References

- Barlas, Yaman. "Multiple Tests for Validation of System Dynamics Type of Simulation Models," European Journal of Operational Research 42 (1989).
- Forrester, Jay W. Industrial Dynamics. The MIT Press, Cambridge, Massachusetts, 1961.
- Mathews, Brian P. and Adamantios Diamantopoulos. Towards a Taxonomy of Forecast Error Measures: A Factor-comparative Investigation of Forecast Error Dimensions Journal of Forecasting, Vol. 13, 409-416, 1994.
- Sterman, John, Nelson Repenning and Fred Kofman. Unanticipated Side Effects of Successful Quality Programs: Exploring a Paradox of Organizational Improvement. Management Science (forthcoming).
- Lyneis, James M. and Alexander L. Pugh III. "Hand vs. Automated Tuning: An Experimental Analysis," Proceedings of the 1996 International System Dynamics Conference, Cambridge, Massachusetts, July 1996.
- Sterman, John D. Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models. Dynamica. 10(2) pp. 51-66, 1984.
- Sterman, John D., George P. Richardson and Pal Davidsen. Modeling the Estimation of Petroleum Resources in the United States Technological Forecasting and Social Change 33, 219-249, 1988

For a copy of the complete paper, contact the authors at:

Pugh-Roberts Associates
41 William Linskey Way
Cambridge, MA 02142

telephone: (617) 864-8880
facsimile: (617) 864-8884