EPA 22nd Annual National Conference on MANAGING ENVIRONMENTAL QUALITY SYSTEMS

Session: Ambient Air I

Ozone Data Quality Objective Development

Presented at:

Sheraton New Orleans Hotel
New Orleans, Louisiana

Presented on:

April 15-17, 2002

Presented by:

Basil W. Coutant

Battelle

Acknowledgments

Chris Holloman, &
Kristen Swinton
of Battelle

Shelly Eberly &
Mike Papp
of US EPA/OAQPS

Funded by EPA, Contract Number 68-D-02-061

Battelle

Introduction

- The DQO goal
- Simulation Models
- Simulation Examples
- Parameter Estimates from AQS Data
- Progress report

Battelle

INTRODUCTION

- Data Quality Objective Process
  - The goal is to ensure that the data collected are relevant to and meet decision-maker needs.
  - The hardest part - finding out decision maker needs.
  - Once the needs are specified (and quantified) statistical models (simulation models in this case) can be used to quantify data quality that ensure the decision-maker needs or demonstrate how various data quality issues affect the quality of the end product.

Ozone Measurement Goals

- The primary National Ambient Air Quality Standard (NAAQS) for ozone is based on an average from three consecutive years of the fourth highest daily value of rolling 8-hour average hourly concentrations.
- These are compared to a standard of 0.08 ppm. (Rounding effectively makes the standard 0.085 ppm.)

DQO Goals

- Since the measurements are on the hourly level, the objective will be to get the DQO statements to refer to the hourly level measurements.
- We are also working on making sure that CFR QA measures have a direct relevance to the DQO criteria. At the moment, there is an indirect relationship. (There is nothing wrong with the QA measurements, but rather what is done with them.)

The Modeling Process

- A two step modeling process is used.
  - First, the daily maximums are simulated directly. They are NOT calculated from simulated hourly measurements.
    - This is used to establish DQO like requirements for the daily maximums.
  - Second, hourly measurements are simulated and aggregated to the daily level.
    - This is used to convert the DQO requirements from above to more relevant statements at the hourly measurement level.

The Modeling Process (Cont.)

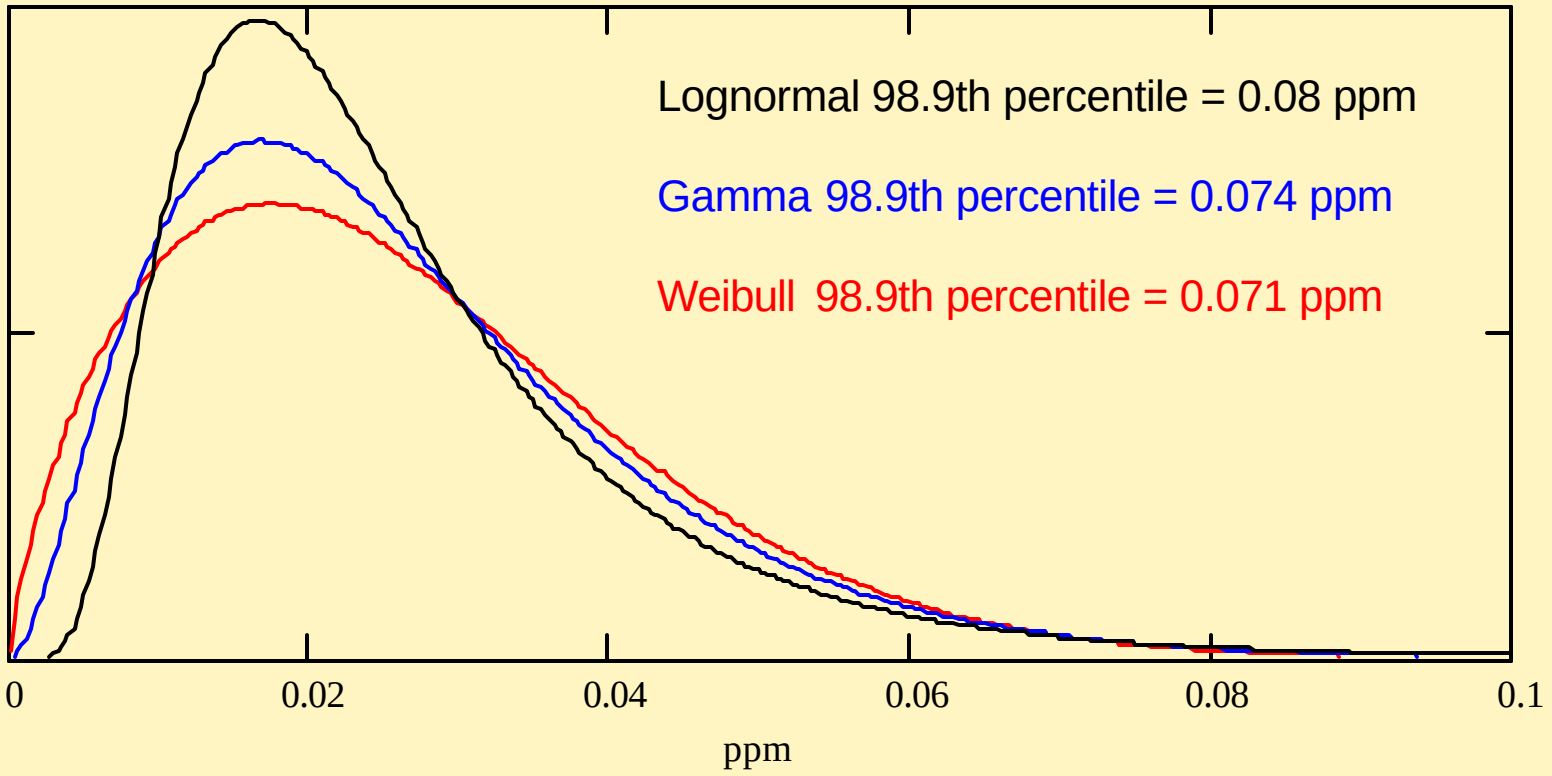
- The two step modeling approach:
  - Gives more control over the simulations to make sure that the simulated values behave like the observed values.
  - Allows us incorporate more of CFR the completeness requirements.
  - Greatly reduces the computations needed. (Ideally, the final DQOs would be based on simulating at least 5,000 3-year periods.)
  - Allows us to look at and understand intermediate steps.

The Modeling Process Details

- First specify a model for the quantities of interest, in this case the maximum 8-hour averages:
  - Long term seasonal patterns are represented by sinusoidal patterns.
  - Small Log-normal random shifts from year to year in the mean.
  - Random Weibull deviations from the seasonal patterns are assumed. These deviations are allowed to be correlated in time.

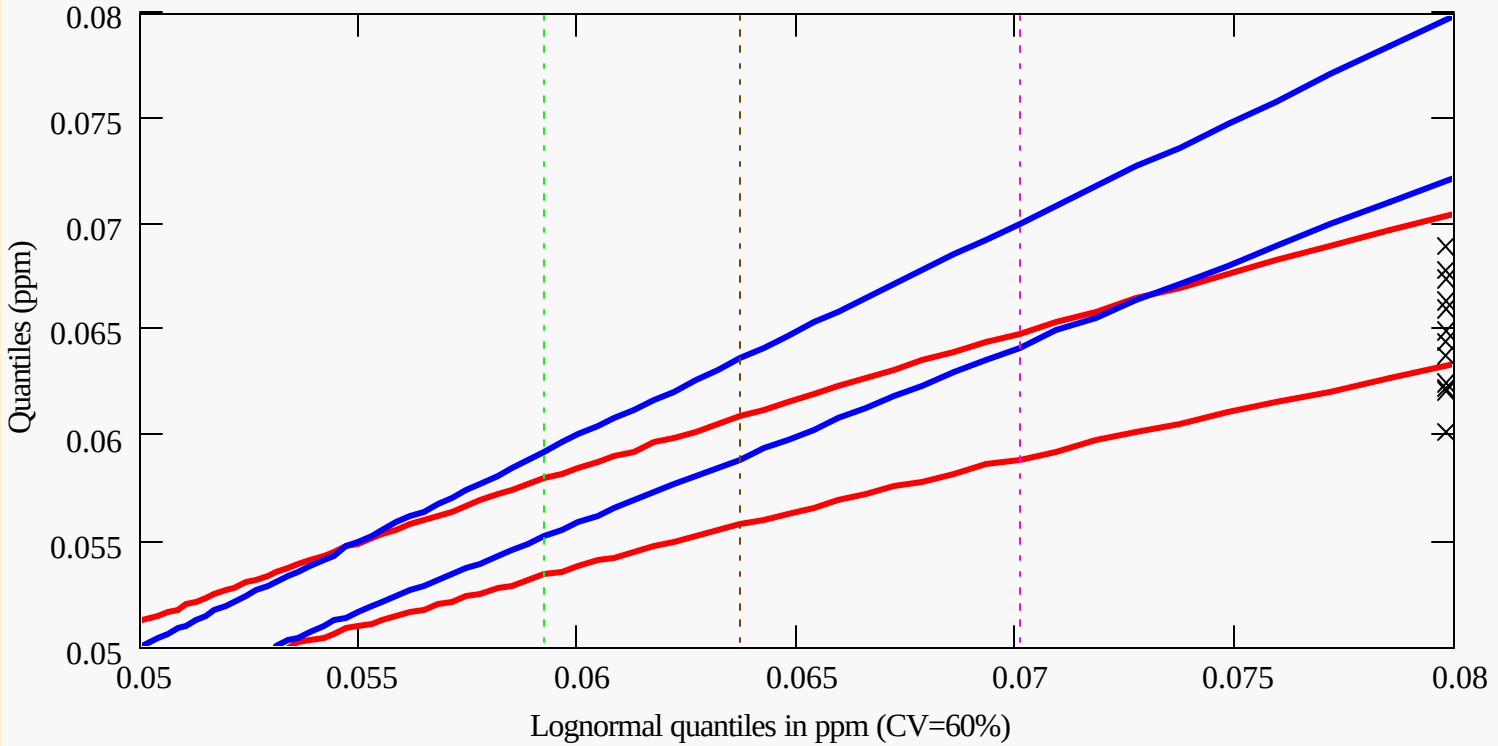
Why Weibull?

Three non-negative distributions with the same mean (0.026 ppm) and a CV of 60%



Battelle

Q-Q Plots
with each distribution scaled to a mean of 0.026 ppm



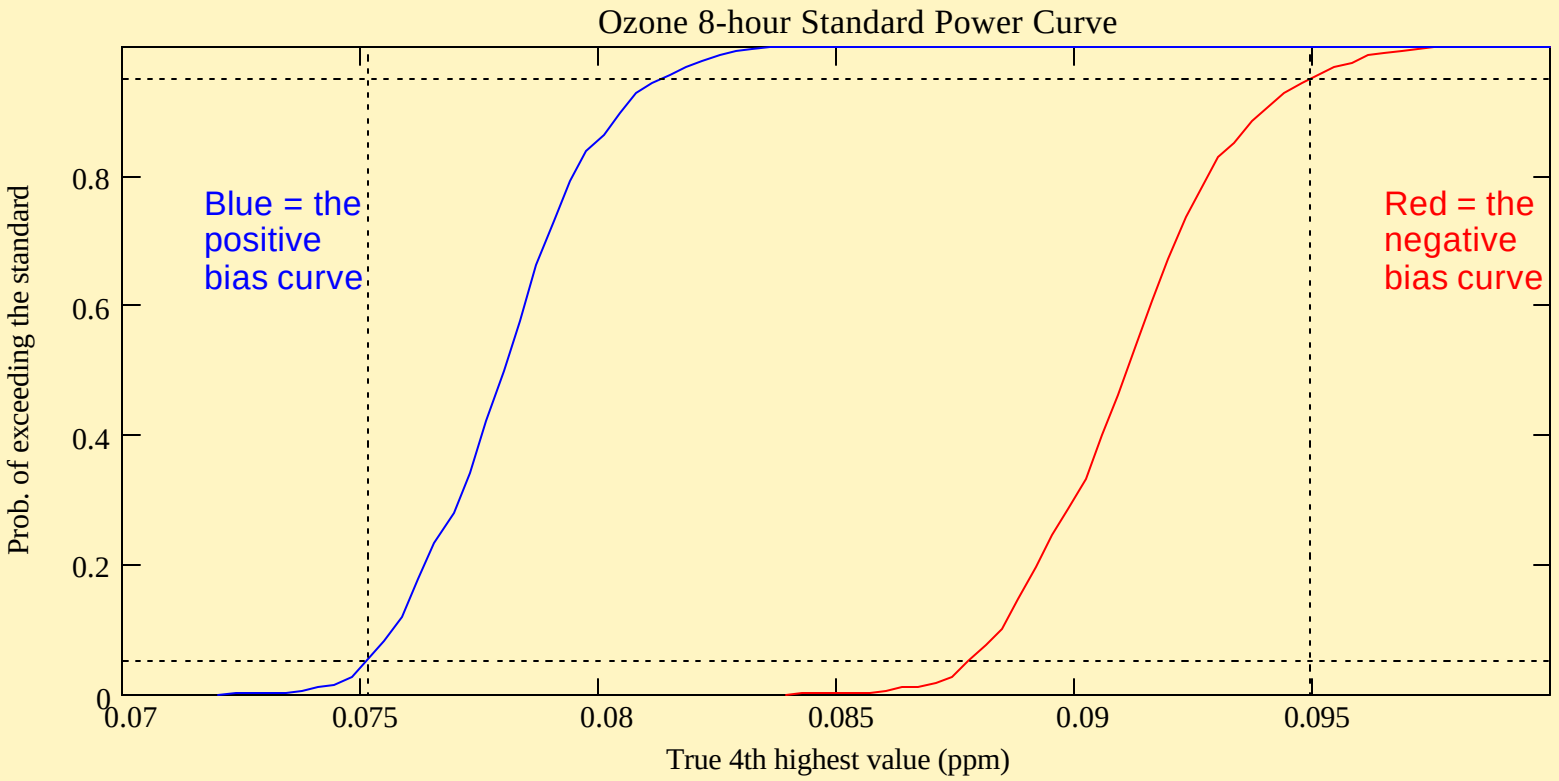
98.9th percentile

Battelle

The Modeling Process Details (Continued)

- The second modeling process is used to “convert” the DQO statements for the daily maximums to statements relevant to the hourly measurements.
  - A sinusoidal diurnal pattern is used. (Other forms looked at, but very little difference was noted.)
  - Random, autocorrelated log-normal deviations from the diurnal pattern.

Simulation Example



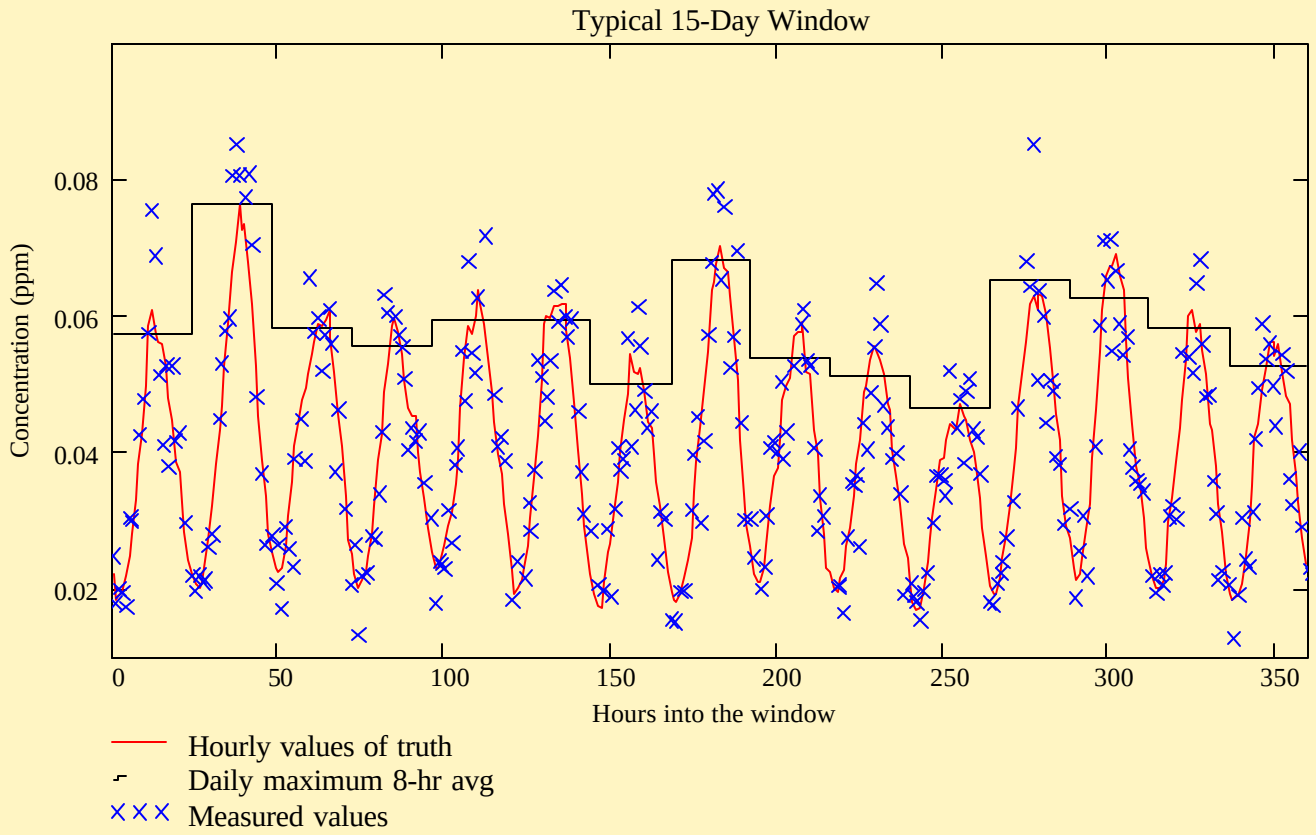
Battelle

Simulation Example Parameters

- Type 1 and Type 2 error = 5%
- Daily Completeness = 95%
- Mid season = June 3rd.
- Seasonal ratio = 3.5
- Weibull gamma = 2.5
- Bias = 7.75%
- Measurement error (of the daily max) = 7.75%

Battelle

Example Part 2



Battelle

Parameter Values for part 2

- Hourly Measurement CV = 15%
- Only 20 hours per day.

- Corresponds to a 5.76% CV for the daily maximums.
- But introduces a 1.65% bias in the daily maximums.

- *Both within day incompleteness and the hourly measurement CV can introduce bias in the estimate of the daily maximums (and hence the 3-year aggregate).*

Battelle

AQS Data for Parameter Estimates

- The raw data database for this effort is an archived collection of hourly ozone concentrations reported in AQS for the years 1999, 2000, and 2001 extracted on September 23, 2002, and available at

www.epa.gov/ttn/airs/airsaqs/archived%20data/archivedaqsdata.htm.

- The P and A data for 1999-2001 have also been extracted and summarized.

Ambient Behavior Parameter Estimates

Decile	Annual Mean (ppm)	Gamma	Ratio	Seasonal Peak	Pop CV (Based on Gamma)
10	0.032	**2.99**	1.33	**7/27/2000**	36.4%
20	0.035	3.26	1.49	7/12/2000	33.8%
30	0.037	3.48	1.62	7/6/2000	31.8%
40	0.039	3.67	1.76	7/1/2000	30.3%
50	0.041	3.87	1.94	6/27/2000	28.9%
60	0.043	4.10	2.14	6/24/2000	27.4%
70	0.045	4.34	2.39	6/20/2000	26.1%
80	0.048	4.61	2.79	6/14/2000	24.7%
90	0.051	5.32	**3.46**	**6/3/2000**	21.6%

Battelle

Parameter Estimation

- The ambient behavior estimates are not based on algorithms that could be easily duplicated in Excel (a goal for the PM estimators). Instead, SAS routines for fitting non-linear models (with non-normal errors) are used. (Proc NLMIXED)
- Consequently, no DQO tool is planned for ozone.

Progress

- A national database of ambient behavior parameter estimates has been built with existing data so that reasonable ranges can be inferred.
- A national database of QA parameters has been built to make sure that the proposed DQO requirements are reasonably attainable. (Not shown.)
- A simulation tool has been developed to test scenarios.
- Several scenarios will be developed for decision-maker consideration.

Battelle