

Beltsville PBL Air Quality Modeling – Atmospheric Chemistry
Evaluation of Meteorological and Air Quality Simulations with Observations
(William Stockwell/Rosa Fitzgerald)

Objective: The first objective is to review basic statistics. Next statistics to evaluate air quality simulations with observations will be discussed. Finally, students will apply these statistics to a dataset that is provided.

Statistics Overview

Simulations and measurements have errors associated with them. There are random errors and systematic errors. First consider random errors and suppose that a certain measurement with random error was repeated a great number of times. Random error leads to the results plotted in the following figure. The x-axis shows the values measured as the experiment was repeated. The y-axis shows the number of times the experiment yielded a particular measured value. The plot is known as a Gaussian or normal distribution. The mean average is the most frequently measured value. The mean, \bar{x} , is given by: $\bar{x} = \frac{\sum x_i}{N}$ where the x_i are the measured values and N is the total number of measurements.

The standard deviation of the mean, σ , is given by $\sigma = \sqrt{\frac{(x_i - \bar{x})^2}{N-1}}$.

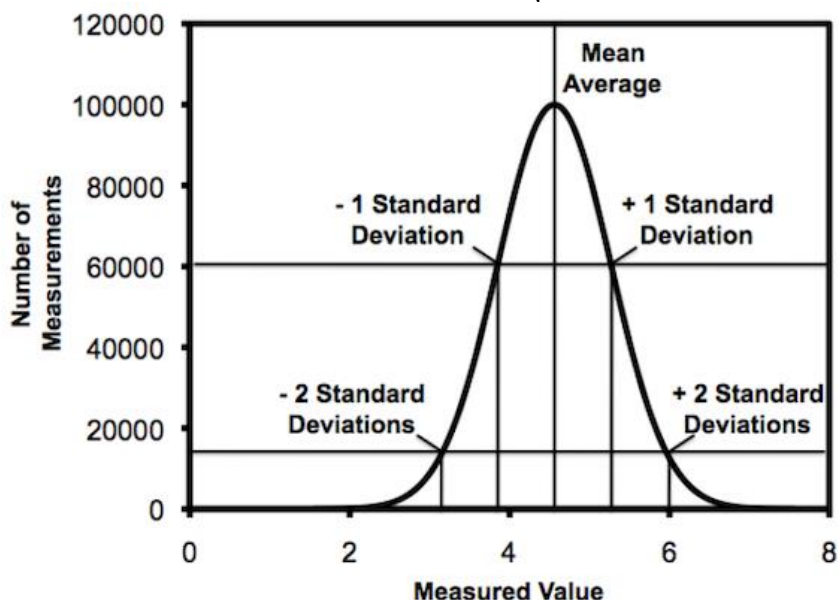


Figure 1. A hypothetical experiment was repeated many times. The measured values are plotted on the x-axis and the number of times a given value was measured is plotted on the y-axis. In this experiment the average measured value is 4.56. The standard deviation, σ , in this particular case is 0.707.

Many measurements are reported as the mean plus or minus one standard deviation, $\bar{x} \pm \sigma$, and in this example 4.56 ± 0.707 . For a Gaussian distribution 68.26% of the measurements are between one standard deviation less than mean and one standard deviation greater than the mean. In general, it is better to report measurements reported as the mean plus or minus two standard deviations, $\bar{x} \pm 2\sigma$, and in this example 4.56 ± 1.414 . For a Gaussian distribution 95.44% of the measurements are between two standard deviations less than mean and two standard deviations greater than the mean. In either case it should be clearly stated which choice has been made to report on the basis of one or two σ .

Accuracy and Precision: This can be turned around to address accuracy. Accuracy is a measure of the degree to which the result of a series of measurements conforms to the true value. If many measurements are made, and there is only random error, there is a 68.26% probability that the true result will be in the range $\bar{x} \pm \sigma$. If many measurements are made, and there is only random error, there is a 95.44% probability that the true result will be in the range $\bar{x} \pm 2\sigma$. In any case you need to state the number of standard deviations you are reporting as the uncertainty range.

Further, the standard deviation is a measure of the probability that an additional measurement would fall within a range. It is not a measure of the precision of the mean. That is the standard error of the mean $SD_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$.

The standard deviation is a measure of precision of individual measurements, the standard error of the mean is a measure of the precision of the entire set of measurements. Because physical and chemical measurements are usually very precise the tendency is to use the standard deviation. In fields where the spread of individual measurements is large one tends to state the precision in terms of standard error.

Figure 2 gives the relationship between accuracy and precision. Remember accuracy relates to the agreement between measurements and the true value while precision is a measure of the agreement between different measurements. High precision does not mean that an experiment is accurate but an experiment with low precision and high accuracy in measuring the mean requires a large number of measurements.

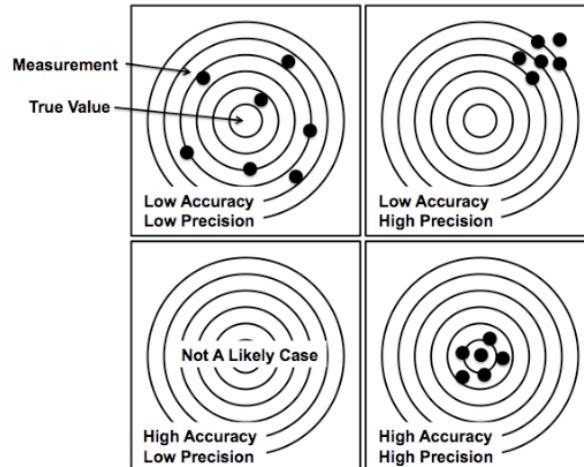


Figure 2. There are three relationships between accuracy and precision in measurements. The measurements may have low accuracy and low precision; high accuracy and high precision; and low accuracy and high precision. An experiment with low accuracy and high precision is likely to suffer from systematic errors, also called bias.

A measurement apparatus may have systematic errors. Systematic errors are not random. For example, the apparatus may have an offset that gives high or low values. Bias is a measure systematic error. Bias, B , is defined by the following equation where N is the number of measurements, x_i represents the measurement values and x_T is the true value $B = \frac{1}{N} \sum (x_i - x_T)$. A major difficulty with this formula is knowledge of the true value. However, in the case of calibration with known standards it is possible to account for bias in an apparatus.

Linear Regression: Determination of a calibration equation is easiest if the device that responds linearly to the quantity that it measures. The equation of a line is: $Y = mX + b$ where X is the independent variable, Y is the dependent variable. How fast the line is rising or falling is given by the slope, $Slope = m = \frac{\Delta y}{\Delta x}$ and the intercept, b , is where the line crosses the y -axis when $x=0$, Figure 3. As with any other quantities calculated from data, the slope and intercept have a standard deviation associated with each one of them. In the sections below the standard deviation of slope and the standard deviation of intercept are denoted by σ_m , and σ_b respectively.

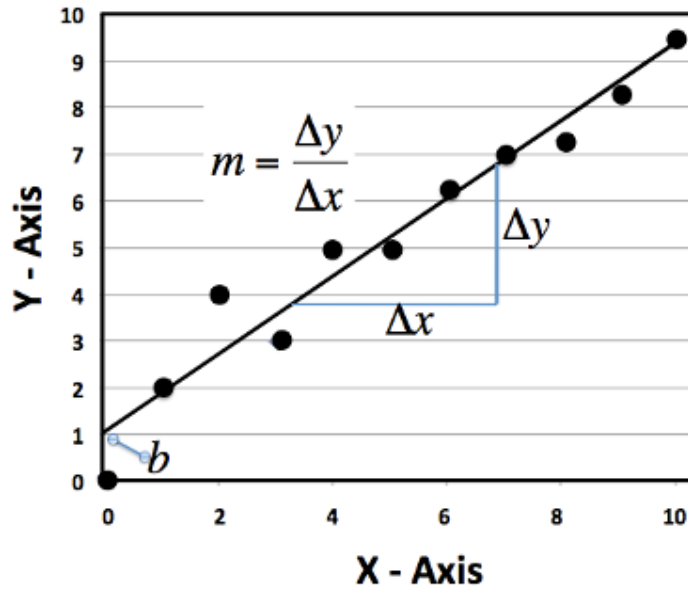


Figure 3. This figure shows the definition of the intercept, b , and the slope, m , of a line.

A correlation coefficient is associated with the fit of a line to data. The correlation coefficient is a measure of the quality (or tightness) of the fit of the line to the data, Figure 4. A positive correlation coefficient means that Y increases as X increases while a negative correlation coefficient means that Y decreases as X increases. The square of the correlation coefficient is the shared variance between the X and Y variables. Although the definition of shared variance is highly mathematical, you can think of it as the variability that is common between two variables.

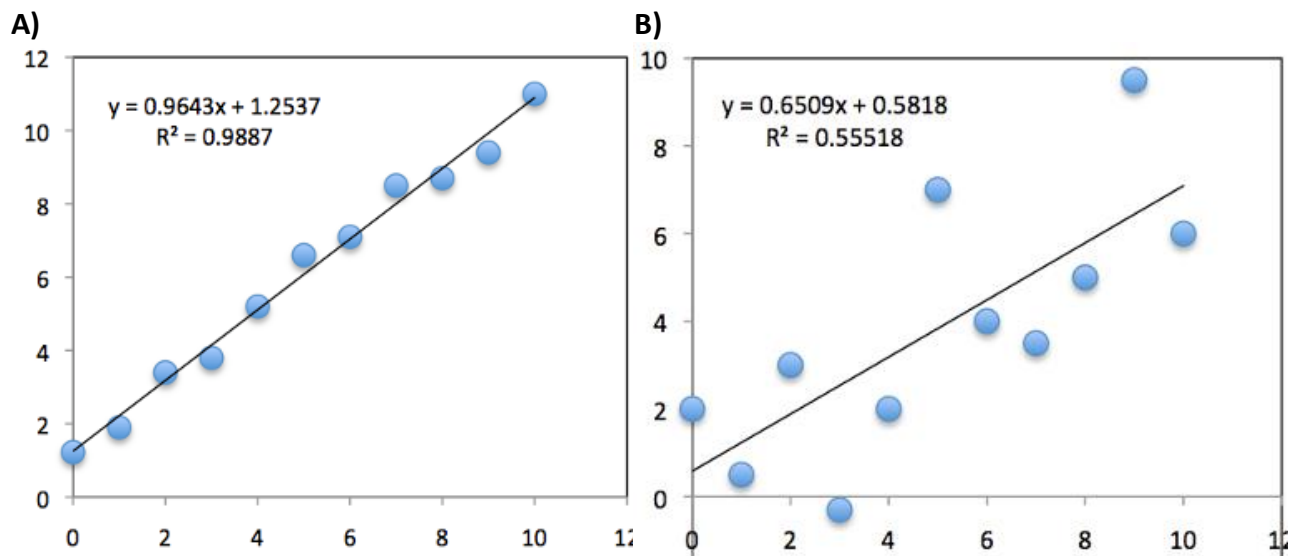


Figure 4. This figure shows two sets of data. Plot A shows data with a high degree of correlation and plot B shows data with a much lower degree of correlation.

Another way of looking at this is that if the correlation coefficient is close to + 1, given a value of x you can confidently predict y. On the other hand, if the correlation coefficient is near zero, e.g. the data has a wide-spread OR the fitted line is horizontal, given x there predicting y is chancy. People often forget the last issue

Now there are statistical computer programs that calculate slope, intercept, standard deviation of slope and intercept, correlation coefficient and shared variance. These computer programs to calculate the statistical values use the following formulas.

Slope

$$m = \frac{\{N \times \sum(X \times Y) - \sum X \times \sum Y\}}{\{N \times \sum X^2 - [\sum X]^2\}}$$

Intercept

$$b = \frac{\{\sum Y - m \sum X\}}{N}$$

Standard Deviation of Slope

$$\sigma_m = \frac{\sqrt{\{N \sum Y^2 - Y^2\} - m^2\{N \sum X^2 - [\sum X]^2\}}}{\sqrt{(N - 2)} \times \sqrt{N \sum X^2 - [\sum X]^2}}$$

Standard Deviation of Intercept

$$\sigma_b = \sigma_m \sqrt{\left\{ \frac{\sum X^2}{N} \right\}}$$

Correlation Coefficient

$$r = \frac{\{N \sum X \times Y - \sum X \sum Y\}}{\sqrt{\{N \times \sum X^2 - [\sum X]^2\} \times \{N \sum Y^2 - [\sum Y]^2\}}}$$

Shared Variance

$$v = r^2$$

Statistics for Evaluating Air Quality Simulations with Observations

Eder and Yu (2006) present statistics that are commonly used to evaluate CMAQ and WRF-Chen air quality simulations. These statistics include mean bias (MB), normalized mean bias (NMB), root mean square error (RMSE), normalized mean error (NME), unpaired peak ratio (UPRR), paired mean normalized gross error (PMNGE) and paired mean normalized bias (PMNB). Seven statistical equations are given below. The observed concentrations are labeled as $[X]_o$, model simulated concentrations are labeled as $[X]_m$ and there are N pairs of observations. These statistics may be calculated for the stations individually and for all stations together.

The first statistic is mean bias (MB).

$$MB = \frac{1}{N} \sum_1^N ([X]_m - [X]_o)$$

Normalized mean bias (NMB) is defined by:

$$NMB = \frac{\sum_1^N ([X]_m - [X]_o)}{\sum_1^N ([X]_o)} 100\%$$

Root mean square error (RMSE) is defined by:

$$RMSE = \sqrt{\frac{1}{N} \sum_1^N ([X]_m - [X]_o)^2}$$

Normalized mean error (NME) is defined by:

$$NME = \frac{\sum_1^N |[X]_m - [X]_o|}{\sum_1^N ([X]_o)} 100\%$$

Unpaired peak ratio (UPRR) is the ratio of model simulated maximum value to the observed maximum value.

$$UPRR = \frac{\text{Maximum} ([X]_m)}{\text{Maximum} ([X]_o)}$$

The statistic paired mean normalized gross error (PMNGE) is defined by:

$$PMNGE = \frac{1}{N} \sum_1^N \left(\frac{|[X]_m - [X]_o|}{[X]_o} \right) * 100\%$$

Finally paired mean normalized bias (PMNB) is given by:

$$PMNB = \frac{1}{N} \sum_1^N \left(\frac{[X]_m - [X]_o}{[X]_o} \right) * 100\%$$

In general air quality model performance for simulating PM is not as high as it is for simulating ozone. Air quality models often under-predict PM_{2.5} mass concentrations and the shared variance between the observed and the simulated concentrations might not be high as high as those for ozone.

References

Eder, B. and S. Yu (2006), A performance evaluation of the 2004 release of Models-3 CMAQ, *Atmos. Environ.*, 40, 4811-4824.

Assignments

You have a dataset titled "Model Evaluation Exercise". There is a text version and an Excel version. They contain the same numbers, please use the format that is easiest for you. The files contain initial concentrations, simulated concentrations and hypothetical observed values. You may consider the results coming from a box model or one grid square from a larger model. Note that we assume that the simulated aerosol is equal to the total of the produced nitric acid and sulfuric acid (SULF).

1. Plot the simulated concentrations as time series. Based on your knowledge of atmospheric chemistry discuss the concentration changes you find.
2. Make a correlation plot of the simulated O_3 (y-axis) as a function of the observed O_3 (x-axis). Make correlation plot for the simulated (y-axis) and observed (x-axis) aerosol. Use your preferred software to calculate correlation coefficients and shared variance for your O_3 and aerosol plots and discuss. Note that for the correlation coefficient calculations you do not need to use directly the equations given here.
3. Calculate the model evaluation statistics: mean bias (MB), normalized mean bias (NMB), root mean square error (RMSE), normalized mean error (NME), unpaired peak ratio (UPRR) and paired mean normalized gross error (PMNGE) and paired mean normalized bias (PMNB) and discuss.