CAE 553 Measurements and Instrumentation in Architectural Engineering

Fall 2018

August 28, 2018

Introduction to data acquisition, uncertainty analysis, error propagation, QA/QC, and data analysis

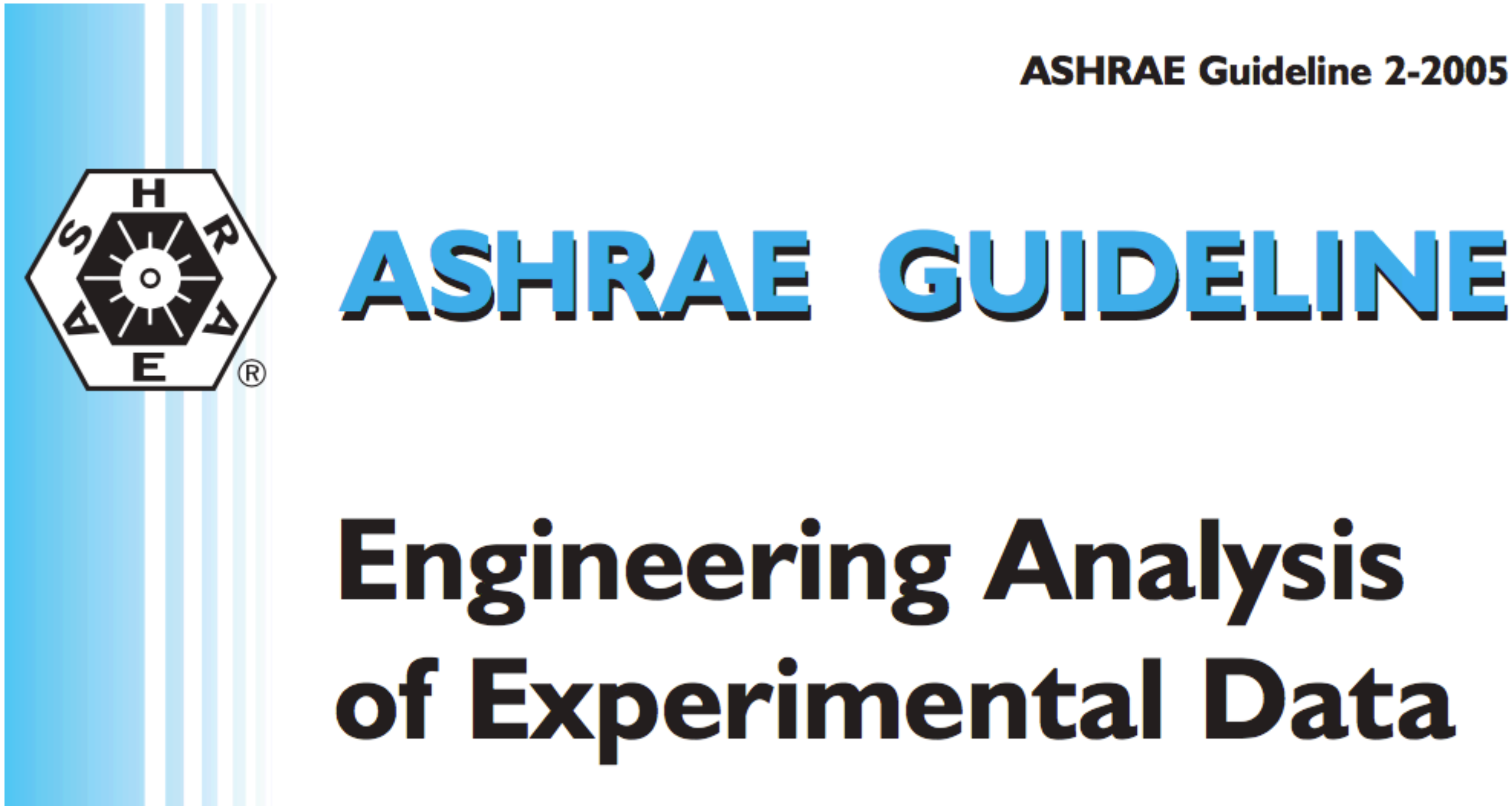
Built
Environment
Research
@ IIT

Advancing energy, environmental, and sustainability research within the built environment

www.built-envi.com

Twitter: @built_envi

Dr. Brent Stephens, Ph.D.
Civil, Architectural and Environmental Engineering
Illinois Institute of Technology
brent@iit.edu



INTRODUCTION TO BUILDING MEASUREMENTS

So you want to make a measurement in a building…

There are **two types of measurements**:

- **Primary measurements**
  - One that is obtained directly from the measurement sensor
  - It is a single item from a specific measurement device
  - Examples: temperature, pressure, velocity, etc.

- **Derived measurements**
  - One that is calculated using one or more measurements
  - The calculation can occur at the sensor level, or by a data logger, or during data processing
  - Derived measurements can use both primary and other derived measurements
  - Examples: an energy meter uses flow and temperature difference to report an energy use rate

So you want to make a measurement in a building…

There are **two categories of data:**

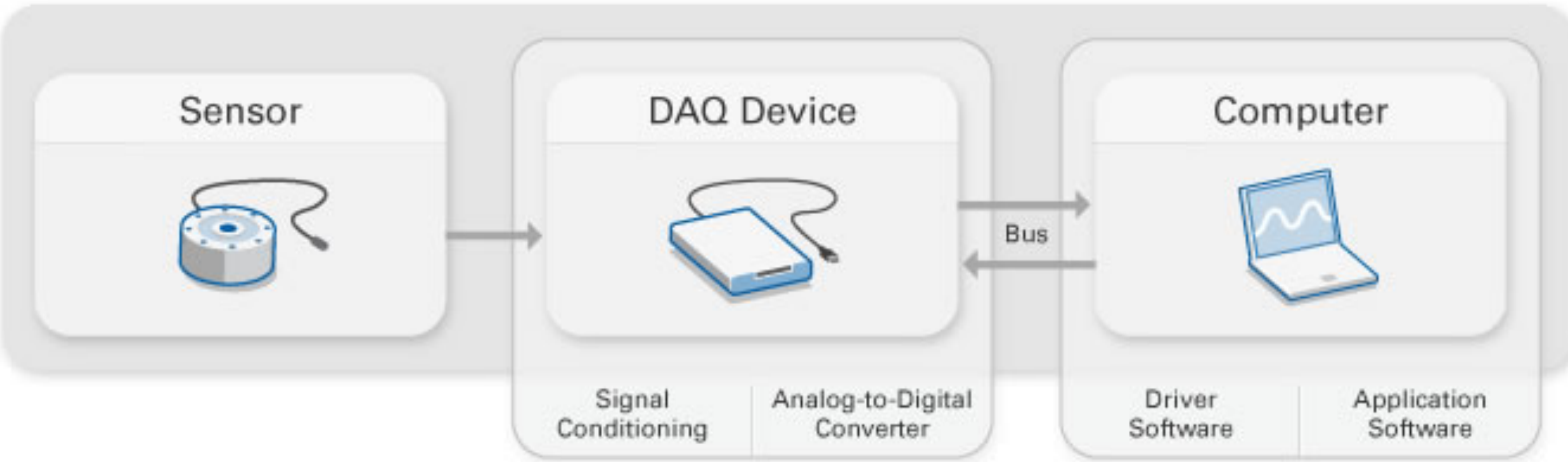
- **Type (i.e., the measured value)**
  - Stationary
    - Data do not change in time
    - Examples: area of a room, length of ductwork
  - Time-dependent
    - Data do change in time
    - Examples: temperature, chilled water flow, electrical power
- **Sample (i.e., how you measure it)**
  - Single-sample
    - One or more readings taken under identical conditions at the same or different times
    - Example: repeated measurements using the same instrument
  - Multi-sample
    - Repeated measurement of a fixed quantity using different instruments
    - Example: repeated measurements using different instruments

The experimental process

- A detailed experimental plan and measurement methodology must be developed to obtain the results you need
- The experimental process:
  1. Identify experimental goals and acceptable accuracy
  2. Identify the important variables and appropriate relationships
  3. Establish the quantities that must be measured and their expected range of variation
  4. Tentatively select sensors/instrumentation appropriate for the task
  5. Document uncertainty of each measured variable
  6. Perform a preliminary uncertainty analysis
  7. Study uncertainty results and reassess the ability of the measurement methods and instrumentation to meet acceptable accuracy
  8. Install selected instrumentation in accord with standards/best practices
  9. Collect experimental data and subject data to ongoing quality control criteria
  10. Reduce and analyze data, perform final uncertainty analysis, and report results

(Some) real world concerns

- First cost and operating cost
- Ease of use
- Safety of use
- Durability
- Flexibility
- Reliability
- Power requirements
- Environmental conditions requirements



INTRODUCTION TO DATA ACQUISITION

http://www.ni.com/data-acquisition/what-is/

Data Acquisition

- The interface system makes every transducer and sensor computer-compatible
- Integration of the transducer to the system leads to loss of its identity
- A data acquisition system usually consists of:
  - Sensors
  - Data acquisition measurement hardware
  - A computer with programmable software

Data Acquisition

- The transducer follows the linearization, offset correction, self-calibration, and so forth, of the system

- Computers are integrated into every aspect of data recording, which allows:
  - Sophisticated graphics
  - Data acquisition hardware
  - Control
  - Data analysis
  - Additional processing power
  - Additional flexibility and connectivity capabilities

Data Acquisition

- *Chart recorders* are a type of data acquisition designed for multipurpose or specifically for a specific sensor
- Chart recorders usually:
  - Provide a visual indication and/or a hard copy record of the data
  - Its output is rarely used to process data
  - Provide simple visual indicators and readouts to monitor the output
- Industrial environments commonly use signal transmitters for control to convert the signal output (e.g., the standard 4-20 mA current loop)

Dicksondata.com

Data Acquisition

- A digital data acquisition system contains an interface:
  - Involves one or several analog-to-digital converters
    - In the case of multichannel inputs (Multiplexers)
  - The interface may also provide excitation for transducers, calibration, and conversion of units
  - Once the input signals have been digitized, the digital data are essentially immune to noise and can be transmitted over great distances

- The digital data are arranged into one or several standard digital bus formats

Data Acquisition

- Pulses are transmitted as 4-, 8-, 12-, 16-, or 32-bit words
- 8-bit word is a byte (is a common communications method)
- Digital data are transferred in either serial or parallel mode
- Serial transmission means:
  - The data are sent as a series of pulses, one bit at a time
  - Makes it slower than the parallel systems
  - Serial interfaces require only two wires (less cabling cost)
  - The speed of serial transmission is rated by the # of symbols per second
- Parallel transmission means:
  - The entire data word is transmitted at one time
  - Each bit of a data word requires its own transmission line
  - Other lines are needed for clocking and control
  - Better for short distances or when high data transmission rates are required

Data Acquisition

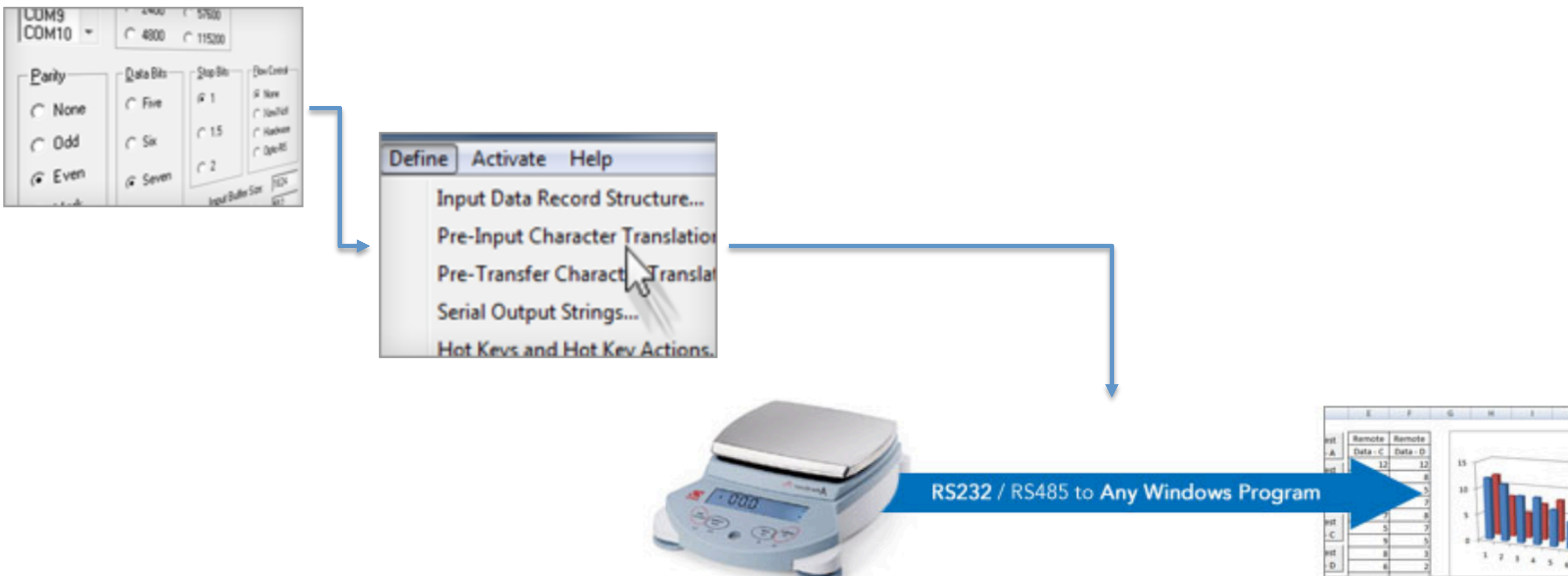
- The two most popular interface bus standards are:
  - The IEEE 488, or general-purpose interface bus (GPIB)
  - The RS232 serial interface
- The IEEE 488 bus system:
  - Feeds data down eight parallel wires (one data byte at a time)
  - Allows to transfer data rapidly at up to 1 million characters per second
  - Limit to a cable length of 65 ft
  - Require an interface connection on every meter for proper termination
- The RS232 system:
  - Feeds data serially down two wires, one bit at a time
  - May be over 1000 ft long
  - May feed a modem to send data over standard telephone lines
  - May use a local area network (LAN) for transmitting information
  - Makes the data available to any computer connected to the network

Data Acquisition

- Bus measurements can simplify three basic applications:
  - Data gathering
  - Automated limit testing
  - Computer-controlled processes
- Data gathering collects readings over time
- A controller can monitor any output indefinitely and then display the data directly on screen or record it
- Automated limit testing requires comparison each measurement with programmed limits and provide readings to a good/bad readout

Data Acquisition

- Examples of existing data acquisition systems: “WinWedge”
  - *Data collection software for serial devices*
  - *Designed for meters, balances, scales or any RS232 instrument*
  - *Sends data to directly to Excel, Access or any Windows application or web page*



http://www.taltech.com/winwedge

Data Acquisition

- Examples of existing data acquisition systems:
  - National Instruments

- Limited channel count
- Applicable for single data acquisition devices

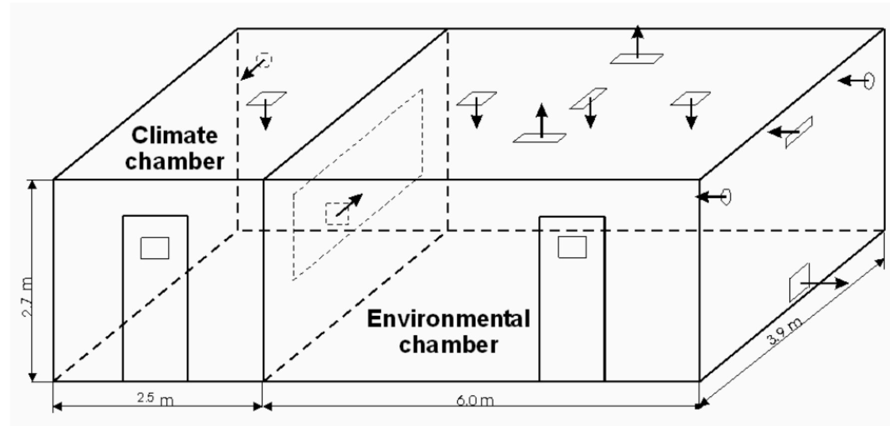
- Medium channel count
- Modular data acquisition system
- USB, Ethernet, and Wi-Fi connectivity

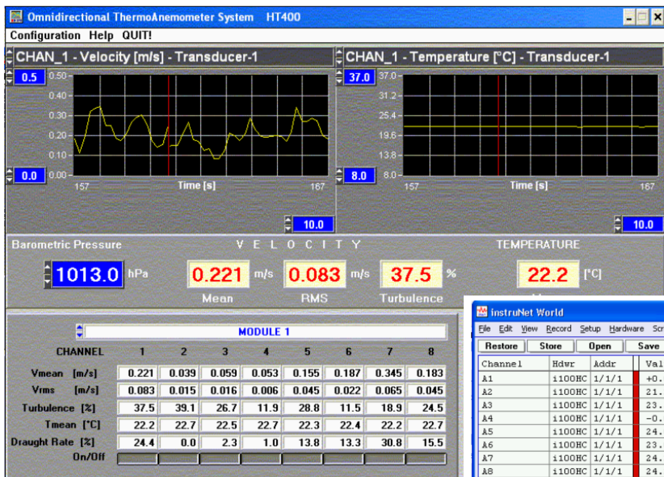
- High channel count
- Applicable for single data acquisition devices
- Synchronization and stand-alone operation or remote control

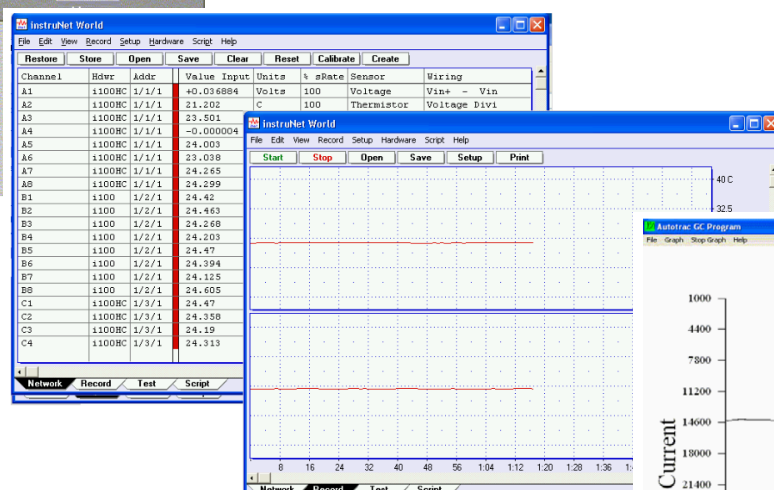
http://www.ni.com/data-acquisition/

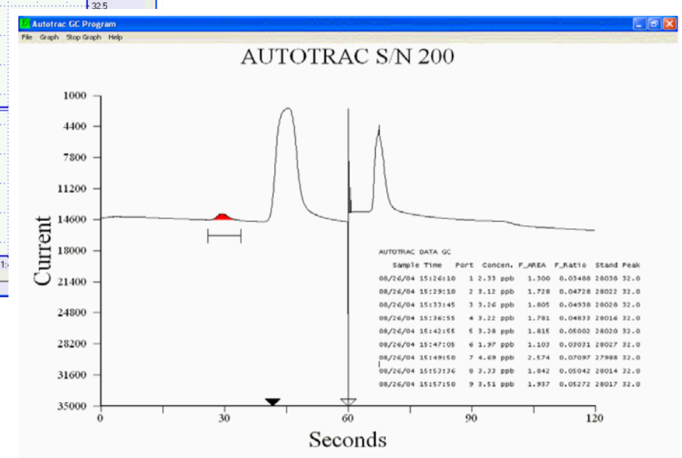
Data Acquisition

- Examples of existing data acquisition systems:









http://www.ce.utexas.edu/prof/Novoselac/Atila_Novoselac_thesis.pdf

Data Loggers

- Data loggers digitally store signals (analog or digital):
  - Store data at different intervals
  - May store data based on an event (e.g., button push)
  - May perform linearization or other signal conditioning
  - Assign the time/date with transducer signal information
  - Have different channel configurations (from single channel input to 256 or more)
  - Can work for general-purpose devices that accept a multitude of analog and/or digital inputs
  - May be more specialized to a specific measurement (e.g., a portable anemometer with built-in data-logging capability) or application (e.g., a temperature, relative humidity, CO_2, and CO monitor with data logging for IAQ applications)
  - Download the stored data using a serial interface with a temporary direct connection to a personal computer

Data Loggers

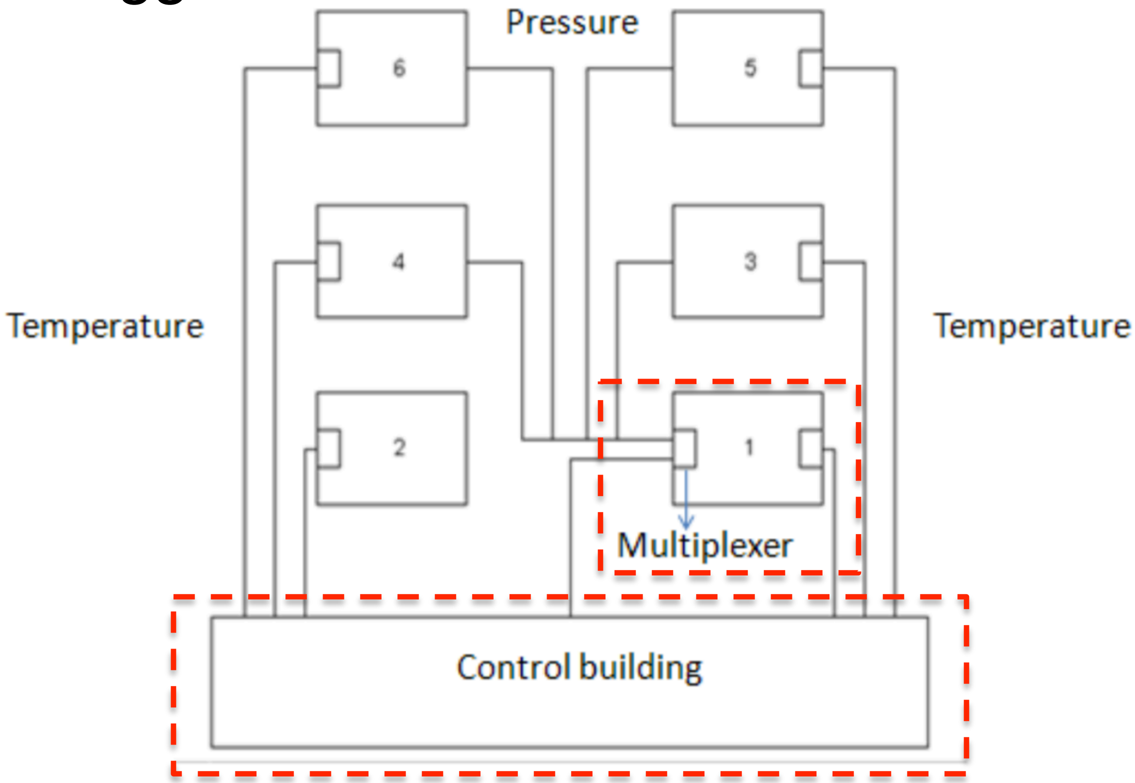
- Examples of using data loggers:

Telaire 7001

HOBO U12

Data Loggers

- Examples of using data loggers:

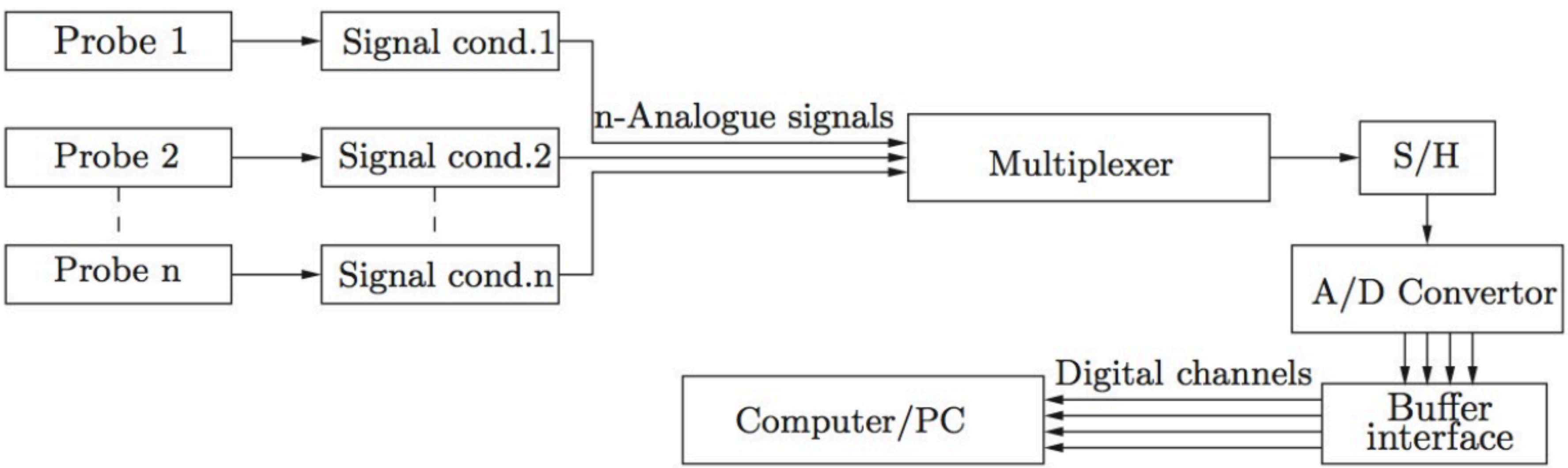


- Multiplexer: AM 16/32 Relay Multiplexer, Campbell Scientific, Inc.
- Data Logger: Campbell Scientific, Inc. 23x

Zhao, PhD Dissertation, Penn State

Data Loggers

- Examples of using multiplexers:
  - Appropriate for low cost system design
  - S/H: Sample and Hold
  - A/D Convertor: Analog to Digital Convertor



Rathakrishnan, Instrumentation, Measurements, and Experiments in Fluids

Commercial Data Loggers

Select, configure, deploy indoor data loggers - and get the data you need!

View Online

ONSET

Register Now

Webinar: Tips for Selecting and Deploying Data Loggers for Building Performance Monitoring

Two Sessions Available

Wednesday, Sep. 19, 2018
Wednesday, Nov. 14, 2018

2:00 PM to 3:00 PM EDT (Sep 19)
2:00 PM to 3:00 PM EST (Nov 14)

Register

Need to monitor a building's energy consumption, heating, air conditioning, lighting, or other system? Figuring out which data logger is best for your indoor application can be confusing, so join us for this **free one-hour webinar**, get tips on how to select the best logger for your monitoring needs, and learn how to:

- Configure and deploy data loggers
- Export and review data from loggers

http://on.onsetcomp.com/webmail/22402/479929823/3db192b081ed917f69cc424d507572a6

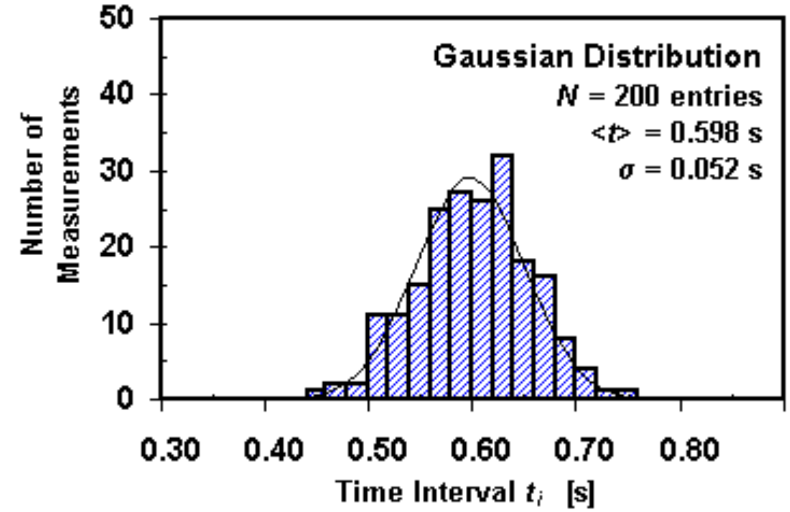
UNCERTAINTY ANALYSIS

Uncertainty Analysis

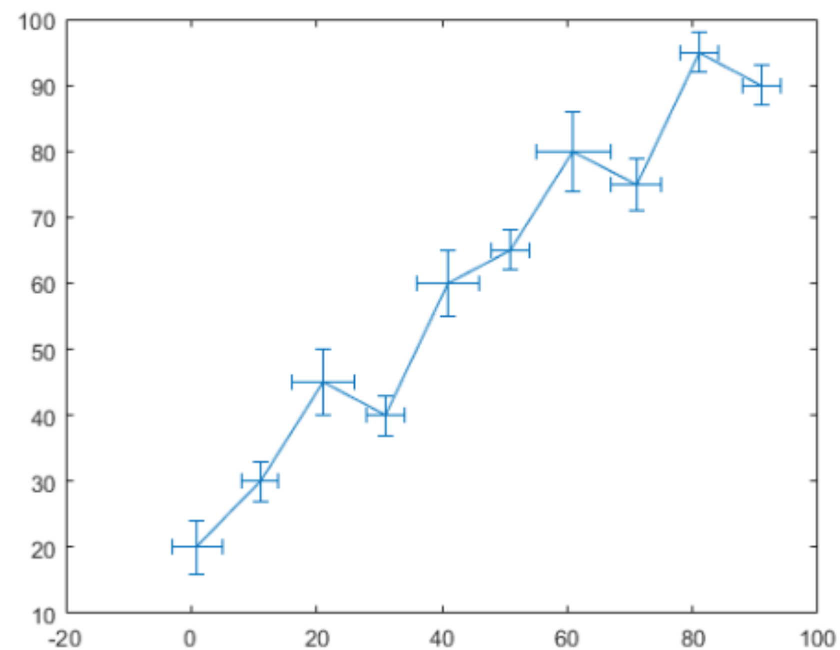
- No measurement is accurate
- The inaccuracy can be expressed in uncertainty
- Uncertainty is NOT the same as an error
  - An error in a measurement is the difference between the correct value and the measured value
  - An error is fixed and not a statistical variable
- Uncertainty is a possible value that the error might take on in a given measurement

Uncertainty Analysis

- Uncertainty:
  - Is a statistical variable
  - Can be represented as a histogram of values



- Similarity of uncertainty with statistical analysis:
  - Measured value describes a central tendency (similar to Mean)
  - Uncertainty represents dispersion usually in terms of probability (similar to Standard Deviation)



How is accuracy reported?

- Onset HOBO U12 internal temperature sensor
  - ± 0.35°C from 0 to 50°C (± 0.63°F from 32° to 122°F)

- Onset HOBO U12 internal relative humidity sensor
  - ± 2.5% RH from 10% to 90% over the range 10° to 50°C (50° to 122°F) typical; ± 3% RH from 5% to 95%; conditions above 80% RH and 60°C (140°F) may cause additional error

- Telaire 7001 CO_2 sensor
  - ±50 parts per million (ppm) or ±5% of reading, whichever is greater

- Setra differential air pressure sensor
  - ±1% FS (Root Sum Squares of Non-Linearity, Non-Repeatability, Hysteresis)

Uncertainty Analysis

- Source of uncertainties:
  - Inaccuracy in the mathematical model
  - Inherent stochastic variability of the measurement process
  - Uncertainties in measurement standards and calibrated instrumentation
  - Time-dependent instabilities caused by gradual changes in standards and instrumentation
  - Effects of environmental factors such as temperature, humidity, and pressure
  - Values of constants and other parameters obtained from outside sources
  - Uncertainties arising from interferences, impurities, inhomogeneity, inadequate resolution, and incomplete discrimination
  - Computational uncertainties and data analysis
  - Incorrect specifications and procedural errors

Uncertainty Analysis

- Some popular terms used in uncertainty analysis

Current	Earlier Version
Precision	Repeatability Random error Random component of uncertainty Probable error
Bias	Fixed error Fixed component of uncertainty Systematic error

Type of Errors

- Errors in the measurements are in two forms of:
  - Fixed or systematic error (*Bias*)
    - It is the same amount for each trial
    - It is the same for each reading
    - Can be removed by calibration or correction
  - Random of non-repeatable error (*Uncertainty*)
    - It is random by nature
    - It is different for every reading
    - Cannot be removed by calibration or correction

CHAPTER 14

MEASUREMENT AND INSTRUMENTS

HVAC engineers and technicians require instruments for both laboratory work and fieldwork. Precision is more essential in the laboratory, where research and development are undertaken, than in the field, where acceptance and adjustment tests are conducted. This chapter describes the characteristics and uses of some of these instruments.

TERMINOLOGY

The following definitions are generally accepted.

Deviation, standard. Square root of the average of the squares of the deviations from the mean (root mean square deviation). A measure of dispersion of a population.

Distortion. Unwanted change in wave form. Principal forms of distortion are inherent nonlinearity of the device, nonuniform response at different frequencies, and lack of constant proportionality between phase-shift and frequency. (A wanted or intentional change might be identical, but it is called **modulation**.)

Drift. Gradual, undesired change in output over a period of time that is unrelated to input, environment, or load. Drift is gradual; if variation is rapid and recurrent, with elements of both increasing

TERMINOLOGY

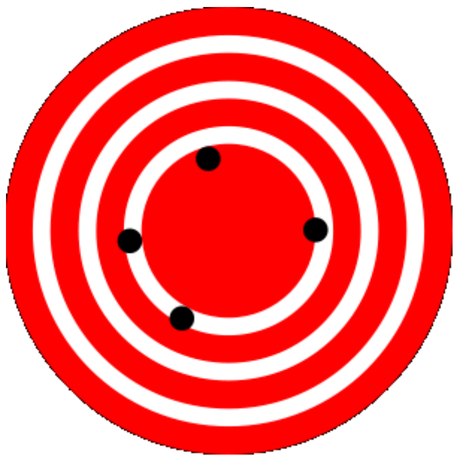
Terminology

- Precision vs. Accuracy:
  - These terms are often confused and conflated with other terms
- Accuracy is "*Capability of an instrument to indicate the true value of a measured quantity*"
- Precision is "*Repeatability of measurements of the same quantity under the same conditions; not a measure of absolute accuracy*"
  - Precision not often reported

Reference *ASHRAE Guideline 2*

Terminology

- Example of accuracy and precision:



High **accuracy**,
low **precision**

Low **accuracy**,
High **precision**

- A good measurement result is both accurate and precise

Terminology

- Do not confuse inaccuracy with accuracy
- Inaccuracy defines as “departure from the true value due all causes of error such as:
  - Hysteresis
  - Nonlinearity
  - Drift
  - Temperature effects
  - Other sources

Reference *ASHRAE Guideline 2*

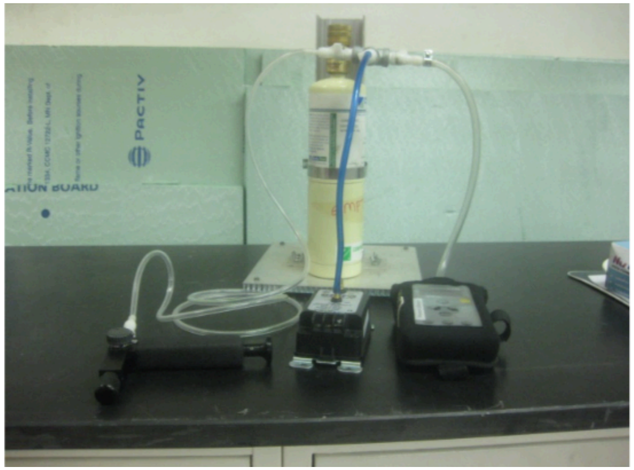
Terminology

- Calibration is defined as:
  - Process of comparing a set of discrete magnitudes or the characteristic curve of a continuously varying magnitude with another set or curve previously established as a standard
  - Process of adjusting an instrument to fix, reduce, or eliminate the deviation defined due to the comparisons with the references curve

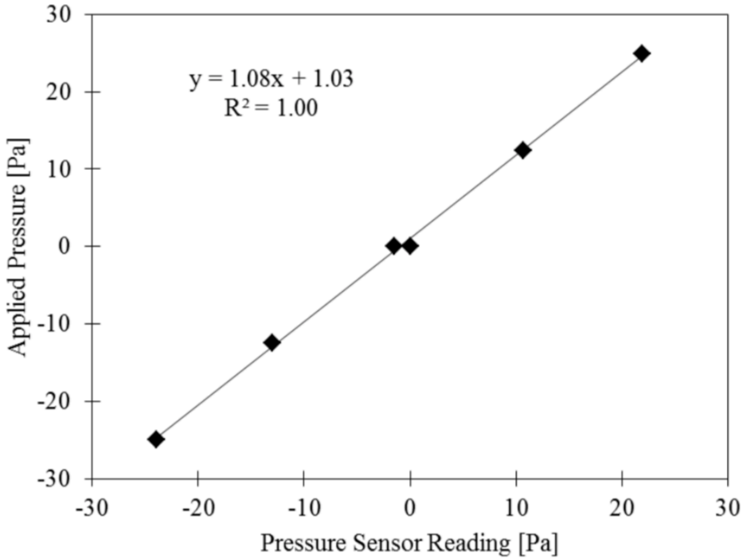
- The process of calibration usually aims to provide calibration curves:
  - Path or locus of a point that moves so that its graphed coordinates correspond to values of input signals and output deflections
  - Plot of error versus input (or output)

Terminology

- Example of calibration:
  - Setup:

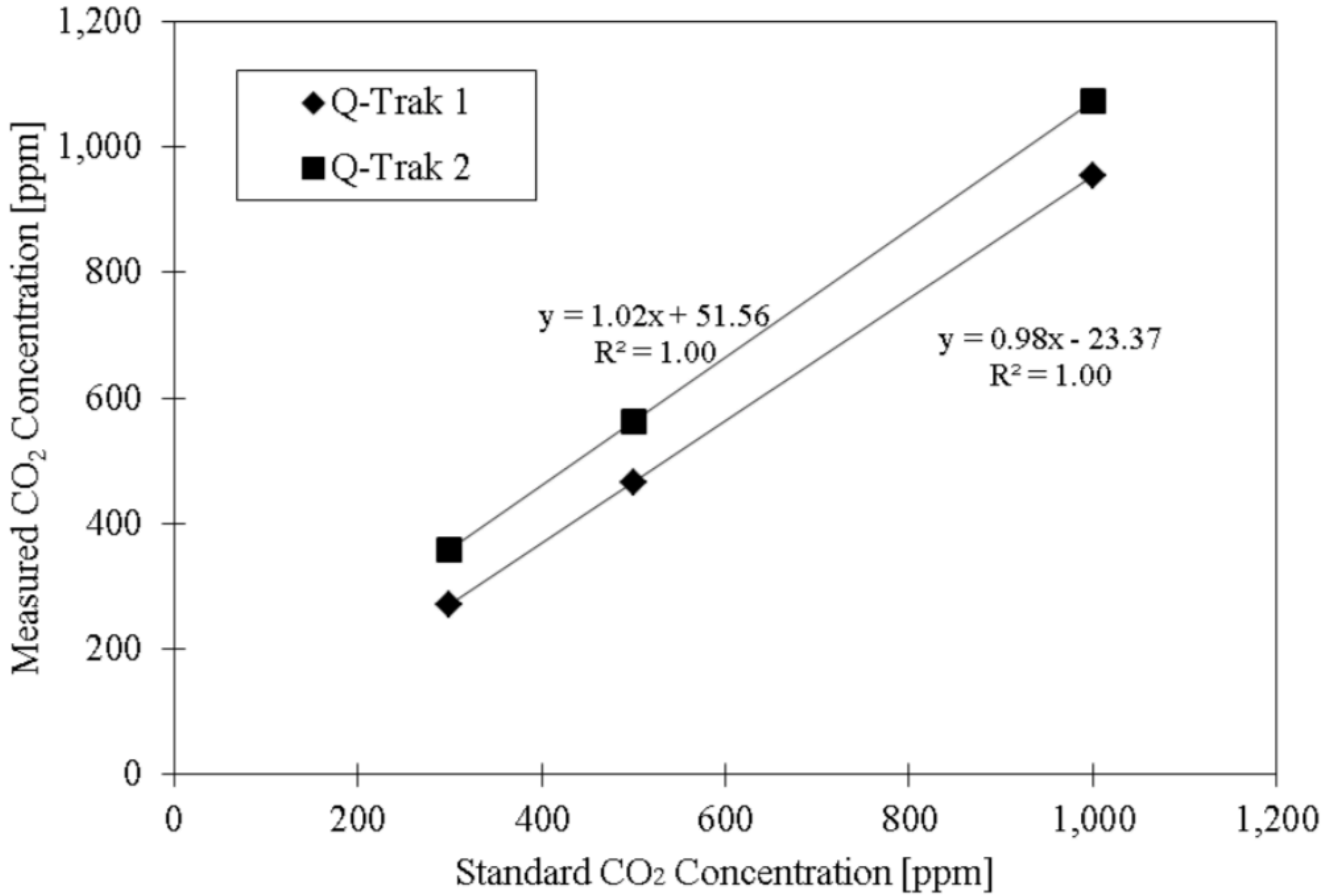


  - Calibration curve:



Terminology

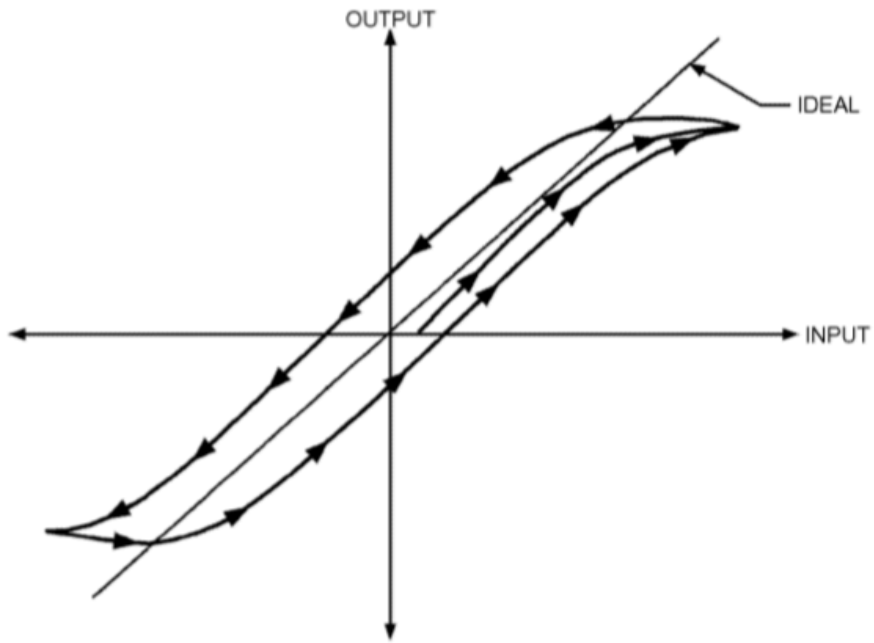
- Example of calibration:
  - Calibration curve:



RP1596, Technical Report

Terminology

- Hysteresis is defined as “summation of all effects, under constant environmental conditions, that cause an instrument’s output to assume different values at a given stimulus point when that point is approached with increasing or decreasing stimulus”

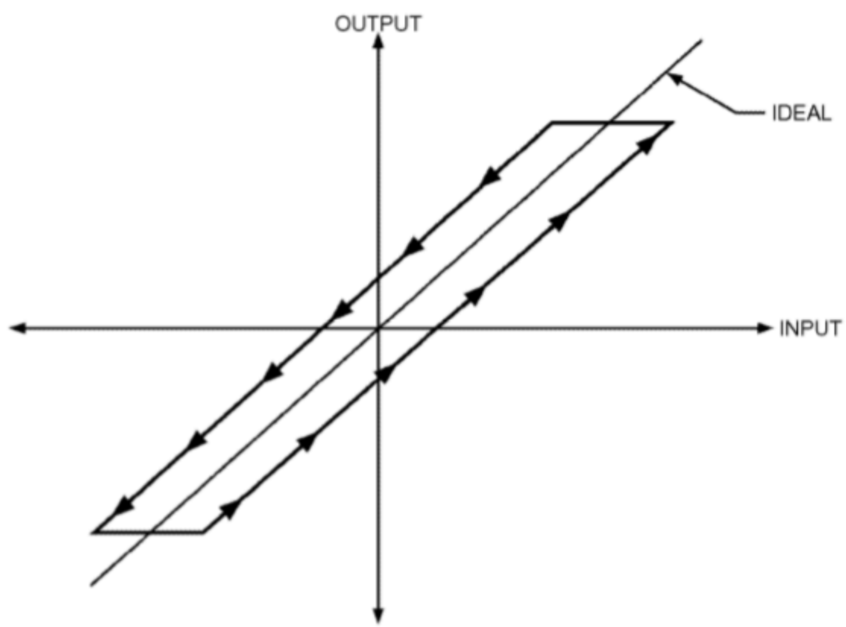


B. HYSTERESIS

ASHRAE Fundamentals, Chapter 14

Terminology

- Deadband is range of values of the measured variable to which an instrument will not effectively respond
- The effect of deadband is similar to hysteresis



ASHRAE Fundamentals, Chapter 14

A. DEADBAND

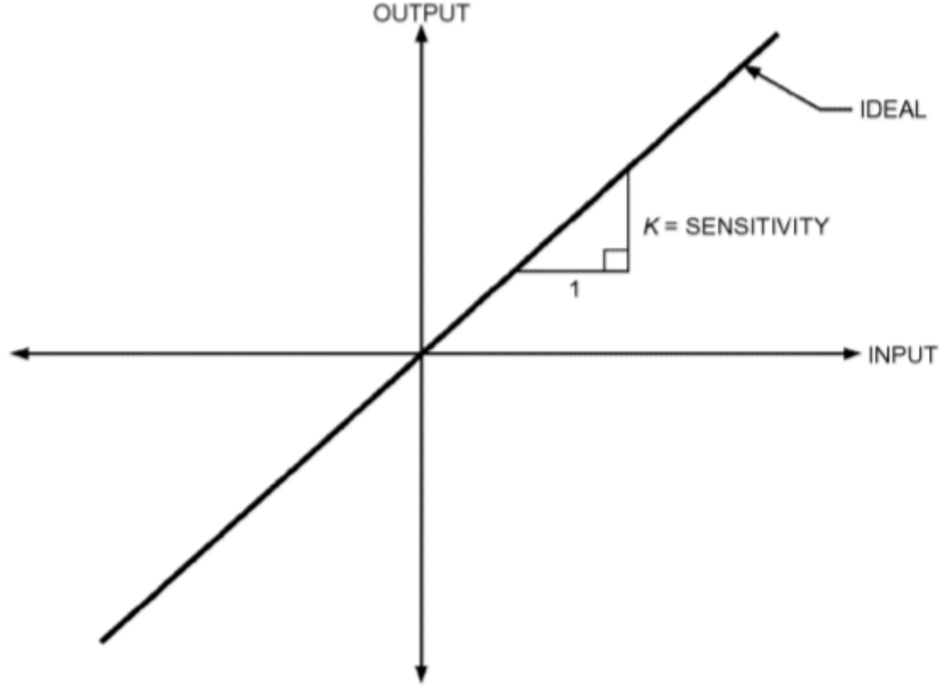
Terminology

- Drift is a gradual, undesired change in output over a period of time that is unrelated to input, environment, or load

- Drift is gradual; if variation is rapid and recurrent, with elements of both increasing and decreasing output, the fluctuation is referred to as cycling

Terminology

- Sensitivity is slope of a calibration curve relating input signal to output
- For linear instruments, sensitivity represents the change in output for a unit change in the input



C. SENSITIVITY

ASHRAE Fundamentals, Chapter 14

Terminology

- Resolution is “the smallest detectable incremental change of input parameter that can be detected in the output signal”

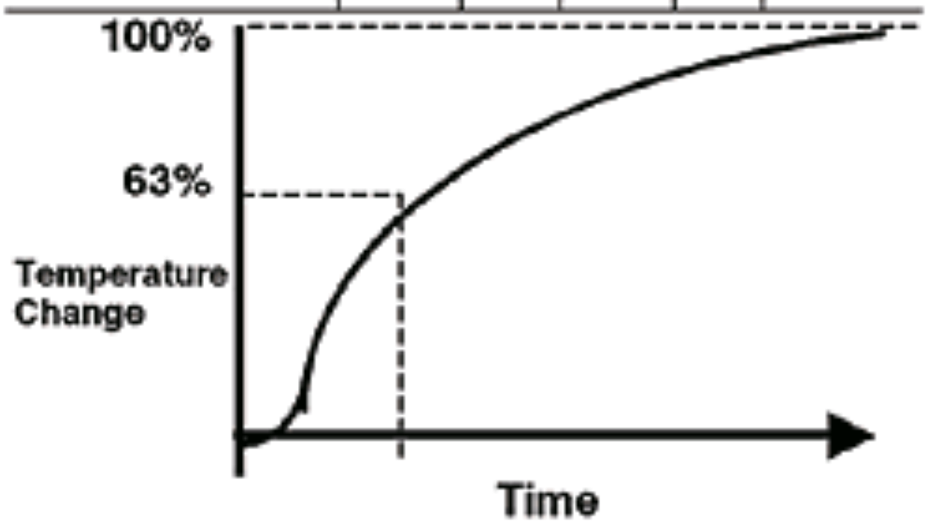
- Resolution unlike precision is a psychological term For example:
  – A CO_2 sensor has a resolution of one part per million (ppm) of full scale
  – The Accuracy is no better than 25 ppm (0.0025%)

- An instrument cannot be any better than the resolution of the indicator or detector

Terminology

- “Time constant” is the time required for an exponential quantity to change by an amount equal to 0.632 times the total change required to reach steady state for first-order systems

% of Final Temperature	50	63	80	95	99.99
Time	0.7 T	T	1.6 T	3 T	4.6 T



Terminology

- Range is a statement of upper and lower limits between which an instrument's input can be received and for which the instrument is calibrated

- Accuracy is rarely constant over a range
- Consider frequent calibration
  - Using standards
  - Calibrate over range of interest
  - Don't use complicated calibration curves
    - Anything other than linear requires justification

- Consider arrangement with multiple sensors

Terminology

HOBO® MX CO_2 Logger (MX1102) Manual

HOBO Data Loggers

The HOBO MX CO_2 data logger records carbon dioxide, temperature, and relative humidity (RH) data in indoor environments using non-dispersive infrared (NDIR) self-calibrating CO_2 sensor technology and integrated temperature and RH sensors. This Bluetooth® Low Energy-enabled logger is designed for wireless communication with a mobile device and also supports a USB connection. Using the HOBOmobile® app on your phone or tablet or HOBOware software on your computer, you can easily configure the logger, read it out, and view plotted data. The logger can calculate minimum, maximum, average, and standard deviation statistics and can be configured to trip audible or visual alarms at thresholds you specify. In addition, it supports burst logging in which data is logged at a different interval when sensor readings are above or below certain limits. This logger also has a built-in LCD screen to display the current CO_2 level, temperature, RH, logging status, battery use, memory consumption, and more.

HOBO MX CO_2 Logger

MX1102

Included Items:

- Four AA 1.5 V alkaline batteries

Required Items:

- HOBOmobile app and Device with iOS or Android™ and Bluetooth

OR

- HOBOware 3.7.3 or later and USB cable

Accessories:

Specifications

Temperature Sensor

Range	0° to 50°C (32° to 122°F)
Accuracy	±0.21°C from 0° to 50°C (±0.38°F from 32° to 122°F), see Plot A
Resolution	0.024°C at 25°C (0.04°F at 77°F), see Plot A
Drift	<0.1°C (0.18°F) per year

RH Sensor*

Range	1% to 70% RH when CO_2 sensor is enabled (non-condensing) 1% to 90% RH when CO_2 sensor is disabled (non-condensing)
Accuracy	±2% from 20% to 80% typical to a maximum of ±4.5% including hysteresis at 25°C (77°F); below 20% and above 80% ±6% typical
Resolution	0.01%
Drift	<1% per year typical

CO_2 Sensor

Range	0 to 5,000 ppm
Accuracy	±50 ppm ±5% of reading at 25°C (77°F), less than 70% RH and 1,013 mbar

Terminology

Warm-up Time	15 seconds
Calibration	Auto or manual to 400 ppm
Non-linearity	<1% of FS
Pressure Dependence	0.13% of reading per mm Hg (corrected via user input for elevation/altitude)
Operating Pressure Range	950 to 1,050 mbar (use Altitude Compensation for outside of this range)
Compensated Pressure Range	-305 to 5,486 m (-1,000 to 18,000 ft)
Sensing Method	Non-dispersive infrared (NDIR) absorption

Response Time

Temperature	12 minutes to 90% in airflow of 1 m/s (2.2 mph)
RH	1 minute to 90% in airflow of 1 m/s (2.2 mph)
CO_2	1 minute to 90% in airflow of 1 m/s (2.2 mph)

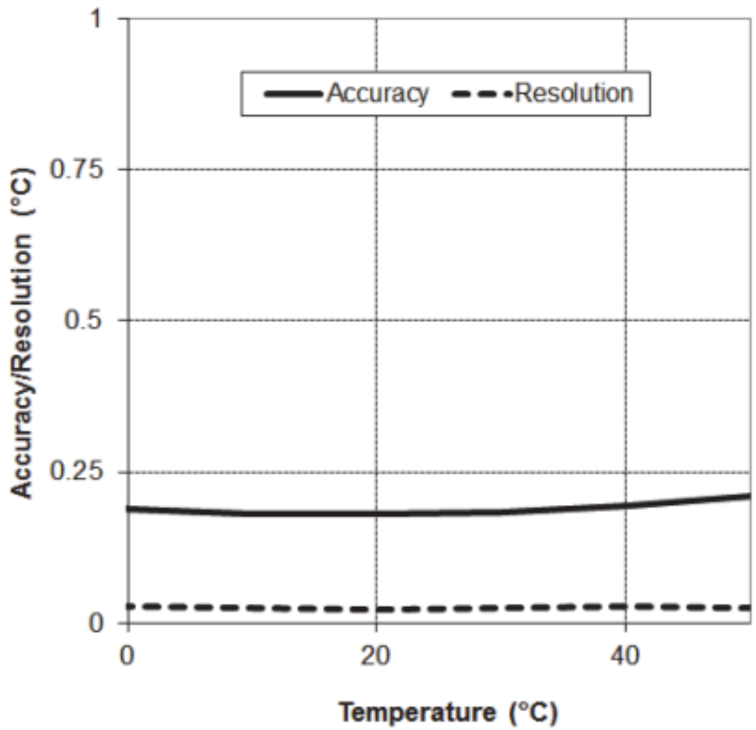
Logger

Radio Power	1 mW (0 dBm)
Transmission Range	Approximately 30.5 m (100 ft) line-of-sight
Wireless Data Standard	Bluetooth Low Energy (Bluetooth Smart)
Logger Operating Range	0° to 50°C (32° to 122°F); 0 to 95% RH (non-condensing)

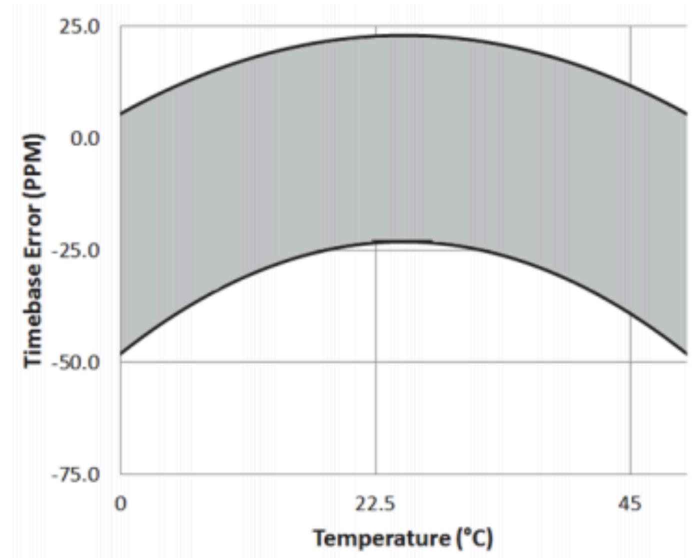
*Per RH sensor manufacturer data sheet

Note: The HOBO U-Shuttle (U-DT-1) is not compatible with this logger.

Terminology



Plot A: Temperature Accuracy and Resolution



Plot B: Time Accuracy

ERROR PROPAGATION

Error Propagation

- The quantification of uncertainty for a measurement requires considering influence of the all variables
- Example: Imagine a pressure measurement use pressure transductors of diaphragm type

$$\frac{\Delta \mathrm{R}}{\Delta \delta} = \mathrm{a}(1 \pm \epsilon_a) \qquad \frac{\Delta \mathrm{V_{in}}}{\Delta R} = \mathrm{b}(1 \pm \epsilon_b) \qquad \frac{\Delta \mathrm{V_{out}}}{\Delta V_{in}} = \mathrm{c}(1 \pm \epsilon_c) \qquad \frac{\Delta \mathrm{y}}{\Delta V_{out}} = \mathrm{d}(1 \pm \epsilon_d)$$

  - ϵ_a to ϵ_d :Proportional error for each step
  - R: Resistance
  - V_{in} and $\mathrm{V_{out}}$: Input and output voltage
  - δ: Diaphragm displacement
  - “a” to “d”: measurement of each step

Rathakrishnan, Instrumentation, Measurements, and Experiments in Fluids

Error Propagation

- Final measurement:

$$X = \frac{\Delta y}{\Delta \delta} = abcd(1 \pm \epsilon_a)(1 \pm \epsilon_b)(1 \pm \epsilon_c)(1 \pm \epsilon_d)$$

- Assume all errors are very small compared to unity:

$$\epsilon = \epsilon_a + \epsilon_b + \epsilon_c + \epsilon_d$$

- In general:

$$X = F(a, b, c, d)$$ (X is our variable)

$$dX = \sum_{n=a}^{d} \left(\frac{\partial F}{\partial n}\right) dn$$ (error in X)

$$\epsilon = \sqrt{\epsilon_a^2 + \epsilon_b^2 + \epsilon_c^2 + \epsilon_d^2}$$

Rathakrishnan, Instrumentation, Measurements, and Experiments in Fluids

Resources for Error Propagation

- Wikipedia: https://en.wikipedia.org/wiki/Propagation_of_uncertainty
- NIST: https://www.itl.nist.gov/div898/handbook/mpc/section5/mpc55.htm
- Notes on Taylor text: https://courses.washington.edu/phys431/propagation_errors_UCh.pdf

General Formula for Error Propagation

1. If x and y have independent random errors δx and δy, then the error in $z = x + y$ is

$$\delta z = \sqrt{\delta x^2 + \delta y^2}.$$

2. If x and y have independent random errors δx and δy, then the error in $z = x \times y$ is

$$\frac{\delta z}{z} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2}.$$

https://courses.washington.edu/phys431/propagation_errors_UCh.pdf

General Formula for Error Propagation

3. If $z = f(x)$ for some function $f()$, then

$$\delta z = |f'(x)|\delta x.$$

https://courses.washington.edu/phys431/propagation_errors_UCh.pdf

General Formula for Error Propagation

We measure $x_1, x_2 \ldots x_n$ with uncertainties $\delta x_1, \delta x_2 \ldots \delta x_n$. The purpose of these measurements is to determine q, which is a function of $x_1, \ldots, x_n$:

$$q = f(x_1, \ldots, x_n).$$

The uncertainty in q is then

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x_1}\delta x_1\right)^2 + \ldots + \left(\frac{\partial q}{\partial x_n}\delta x_n\right)^2}$$

https://courses.washington.edu/phys431/propagation_errors_UCh.pdf

General Formula for Error Propagation

If

$$q = x_1 + x_2,$$

we recover rule 1:

$$\begin{aligned} \frac{\partial q}{\partial x_1} &= 1, \\ \frac{\partial q}{\partial x_2} &= 1, \\ \delta q &= \sqrt{\delta x_1^2 + \delta x_2^2} \end{aligned}$$

https://courses.washington.edu/phys431/propagation_errors_UCh.pdf

General Formula for Error Propagation

If

$$q = x_1 \times x_2,$$

we recover rule 2:

$$\begin{aligned} \frac{\partial q}{\partial x_1} &= x_2, \\ \frac{\partial q}{\partial x_2} &= x_1, \\ \delta q &= \sqrt{x_2^2\,\delta x_1^2 + x_1^2\,\delta x_2^2} \\ &= \sqrt{q^2\left[\left(\frac{\delta x_1}{x_1}\right)^2 + \left(\frac{\delta x_2}{x_2}\right)^2\right]} \\ \frac{\delta q}{q} &= \sqrt{\left(\frac{\delta x_1}{x_1}\right)^2 + \left(\frac{\delta x_2}{x_2}\right)^2} \end{aligned}$$

https://courses.washington.edu/phys431/propagation_errors_UCh.pdf

Error Propagation Example

Problem: Suppose you measure three numbers as follows:

$$x = 200 \pm 2, \quad y = 50 \pm 2, \quad z = 40 \pm 2,$$

where the three uncertainties are independent and random. Use step-by-step propagation to find the quantity $q = x/(y-z)$ with its uncertainty.

Solution: Let $D = y - z = 10 \pm 2\sqrt{2} = 10 \pm 3$. Then

$$q = \frac{x}{D} = 20 \pm 20\sqrt{0.01^2 + 0.3^2} = 20 \pm 6.$$

https://courses.washington.edu/phys431/propagation_errors_UCh.pdf

Error Propagation Exercises

- ASHRAE Guideline 2: Appendix A and B
  - Examples B.1, B.2, B.3, and B.5
  - Work in groups

DATA ANALYSIS

Data analysis starts with data validation

- Limit checks
  - Physical limits
  - Expected limits
  - Theoretical limits
- Independent checks
  - Mass/energy balances
- Checks on outliers
  - Various criteria for rejecting data

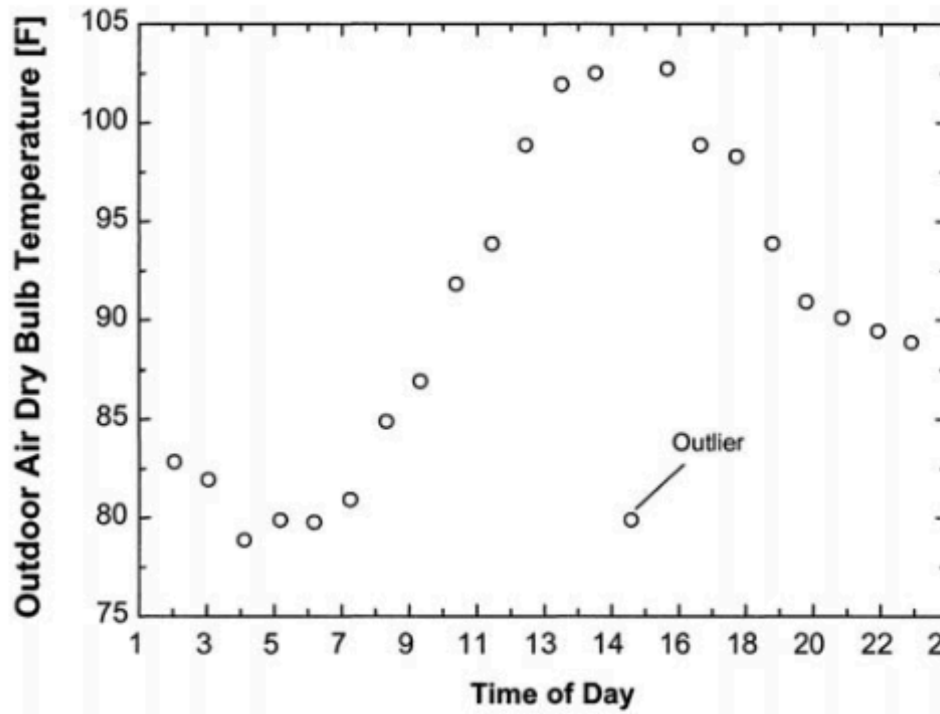


Figure 7-2 Time-dependent data set with single outlier near hour 15 (3:00 p.m.).

Identifying (and rejecting) outliers in the data

7.5 Chauvenet's Criterion for Rejecting Data Points. A statistics-based method that can be used as a basis for rejecting outliers is Chauvenet's criterion. Chauvenet's criterion states that a suspect data point or reading can be rejected if the probability of obtaining its particular deviation from the mean exceeds a specified threshold. To apply the criterion, a trial mean and trial standard deviation are computed using all data points (including suspected points). Then the deviations of individual points from the mean are divided by the trial standard deviation. All values that exceed the criterion values given in Table 7-1 (or, alternatively, from Equation 7-1) can be rejected and removed from the data set. After the data set has been purged of outliers, a new mean and standard deviation should be calculated.

$$\frac{d_{max}}{\hat{\sigma}} = 0.819 + 0.544\ln(n) - 0.02346\ln(n^2) \qquad (7\text{-}1)$$

Also:

- Z-score testing

ENGINEERING STATISTICS HANDBOOK

HOME TOOLS & AIDS SEARCH BACK NEXT

http://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm

TABLE 7-1 Chauvenet's Criterion for Rejecting Outliers

Number of Readings n	Deviation Ratio $d_{max}/\hat{\sigma}$
2	1.15
3	1.38
4	1.54
5	1.65
6	1.73
7	1.79
10	1.96
15	2.13
25	2.33
50	2.57
100	2.81
300	3.14
500	3.29
1000	3.48

Regression model development

- Sometimes you need to analyze your measured data versus some other data to identifying and quantifying meaningful relationships (i.e., regression analysis)
  - Understand relationship between *response variables* and *predictor variables*

$$\hat{Y} = b_0 + b_1 X$$

- Types of regression models:
  - Linear
    - Make residuals as close to zero as possible using least squares method
    - Minimize the sum of the squares of all residuals:

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

  - Polynomial
    - Usually 2^{nd} or 3^{rd} order

$$\hat{Y} = b_0 + b_1 X + b_2 X^2 + \ldots + b_n X^n$$

  - Multivariate models
    - Rely on more than one predictor variable

$$X = C_0 + C_1 \cdot T_{evap} + C_2 \cdot T_{cond} + C_3 \cdot T_{evap}^2 + C_4 \cdot T_{cond}^2 + C_5 \cdot T_{evap} \cdot T_{cond}$$

Regression example: linear

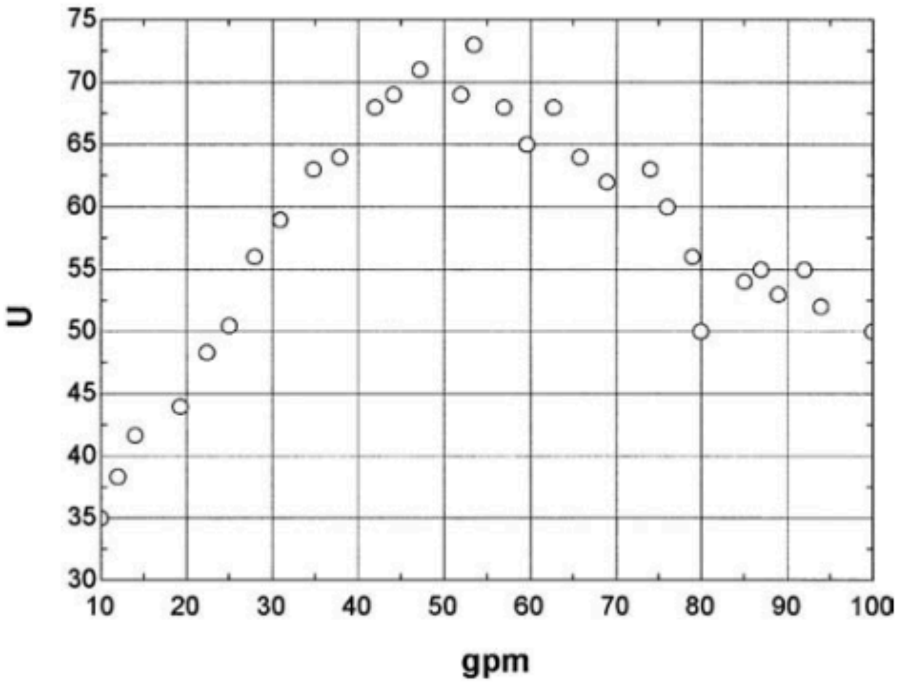


Figure 8-1 Example of heat transfer data.

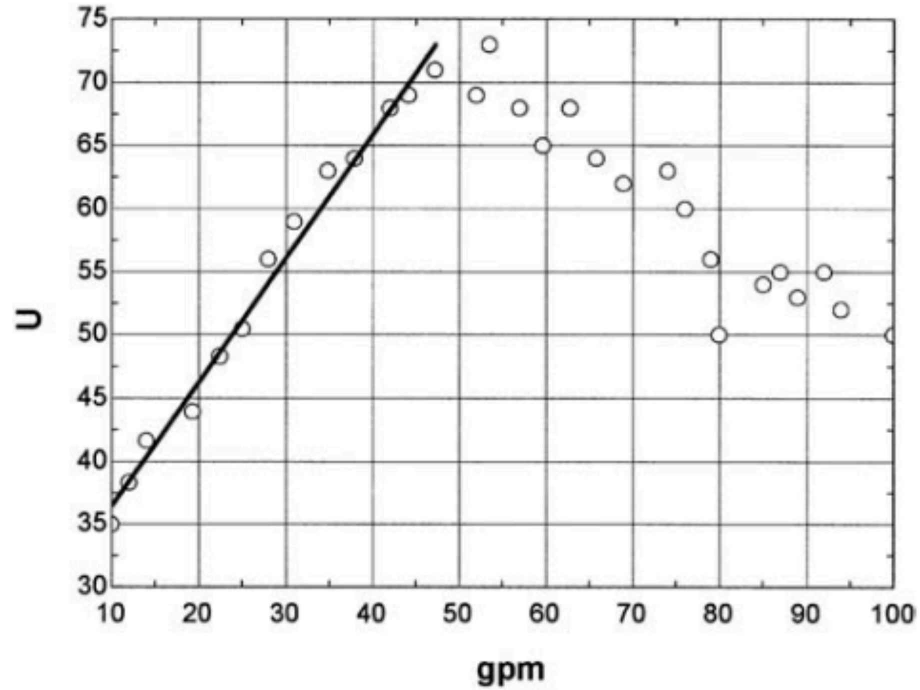


Figure 8-2 Example of heat transfer data including first-order curve fit for all data in the flow rate range from 10 to 50 gpm [0.63 - 3.2 l/s].

$$\hat{U} = 27.44 + 0.9598 \cdot gpm \qquad (10 \le gpm \le 50)$$

Regression example: polynomial

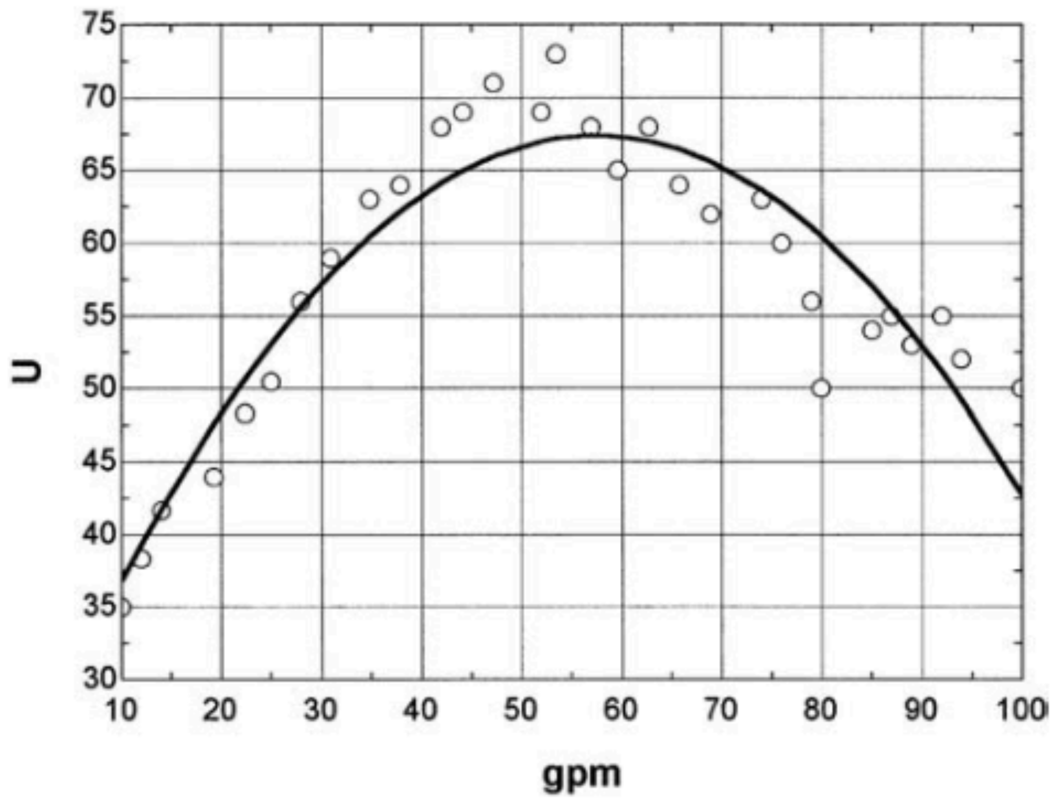


Figure 8-3 Example of heat transfer data including second-order curve fit for all data in the flow rate range from 10 to 100 gpm [0.63 to 6.3 l/s].

$$\hat{U} = 22.6678 + 1.5594 \cdot gpm - 0.013587 \cdot gpm^2$$

Regression example: multivariate

$$Tons = C_0 + C_1 \cdot T_{evap} + C_2 \cdot T_{cond} + C_3 \cdot T_{evap}^2 + C_4 \cdot T_{cond}^2 + C_5 \cdot T_{evap} \cdot T_{cond}$$

Coefficient	Value	t-value
C_0	152. 50	6.3
C_1	3.71	36.1
C_2	-0.335	-0.6
C_3	0.0279	52.4
C_4	-0.000940	-0.3
C_5	-0.00683	-6.1

Model diagnostics: Residuals

8.5.1.1 The residuals, $\varepsilon_i = (Y_i - \hat{Y}_i)$, should be plotted:

1. Overall
2. In sequential order (if known)
3. Against fitted values, $\hat{Y}_i$
4. Against each predictor variable, X_i

Do the residuals exhibit the behavior required by the model?

- Mean of zero?
- Normal distribution?

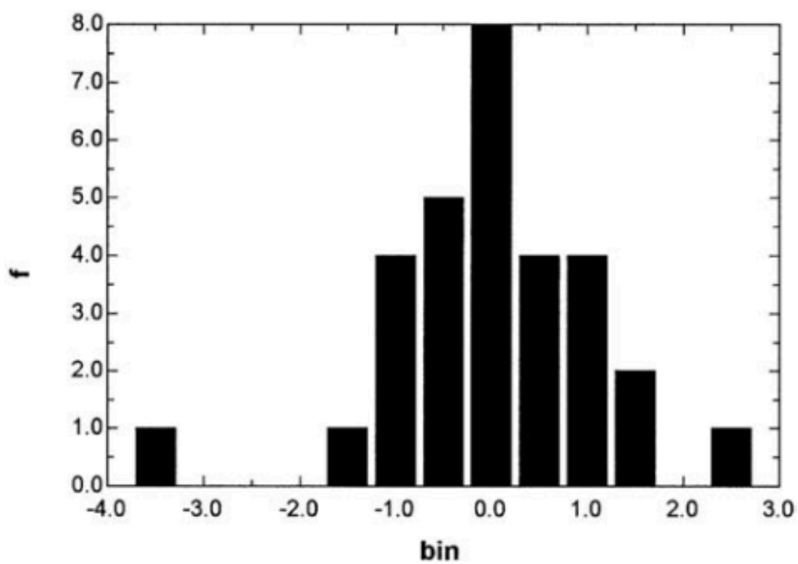


Figure 8-5 Overall or histogram plot of the residuals that resulted from the quadratic curve fit shown in Figure 8-3.

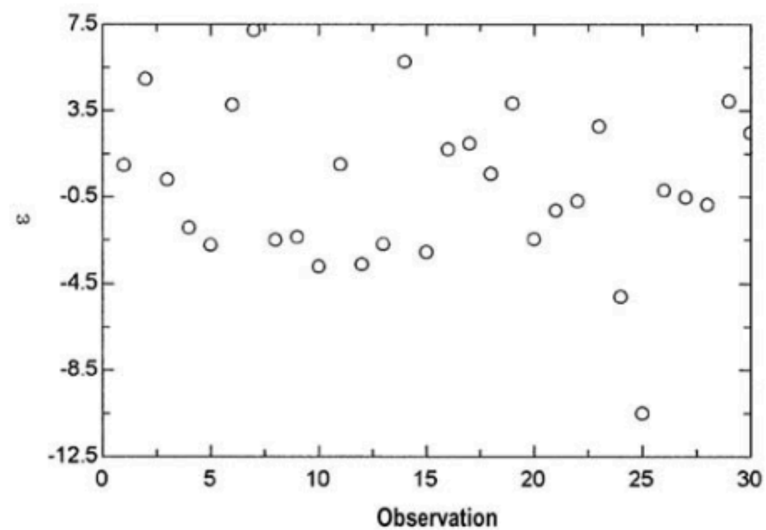


Figure 8-6 Residuals plotted in order of observation.

Model diagnostics: Residuals

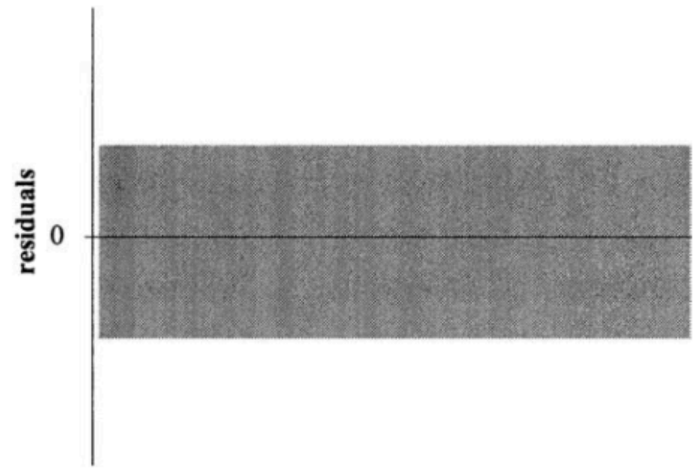


Figure 8-7a Residuals exhibitng acceptable behavior.

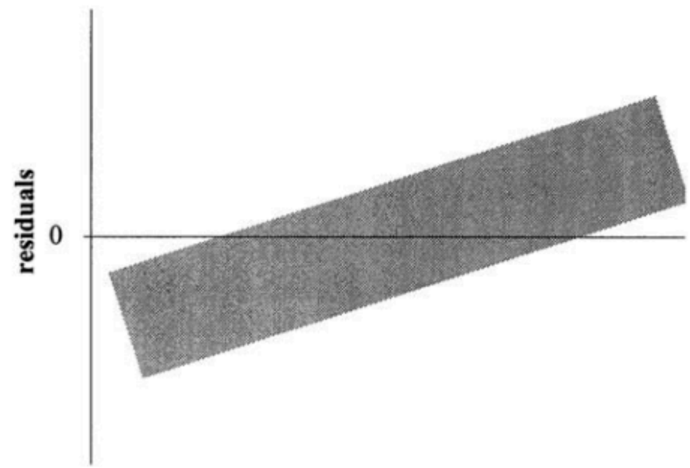


Figure 8-7b Residuals exhibiting behavior that illustrates the model is missing a linear term.

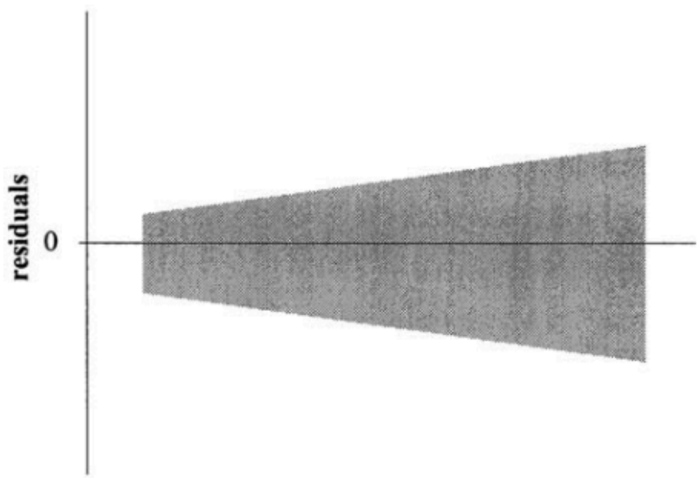


Figure 8-7c Residuals exhibiting behavior that illustrates the model's variance is nonconstant and increasing.

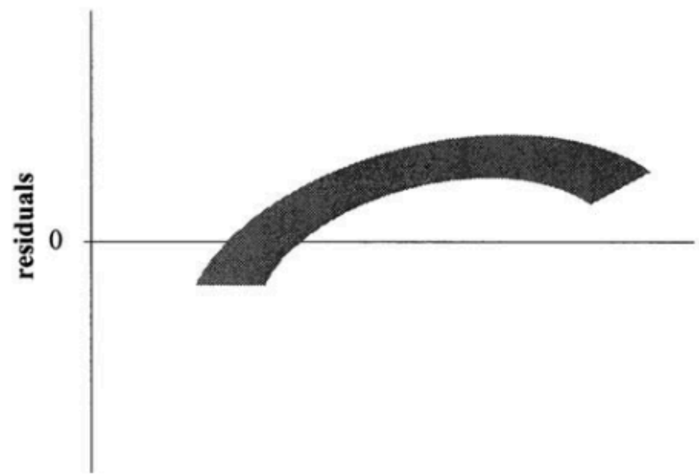


Figure 8-7d Residuals exhibiting behavior that illustrates the model is missing a quadratic term.

R-squared: Be careful!

- Does an $R^2 = 0.82$ represent a good model fit?

Coefficient of determination (R^2): $$R^2 \equiv 1 - \frac{SS_{\rm err}}{SS_{\rm tot}}$$

Total sum of squares:

$$SS_{\rm tot} = \sum (y_i - \bar{y})^2$$

Mean:

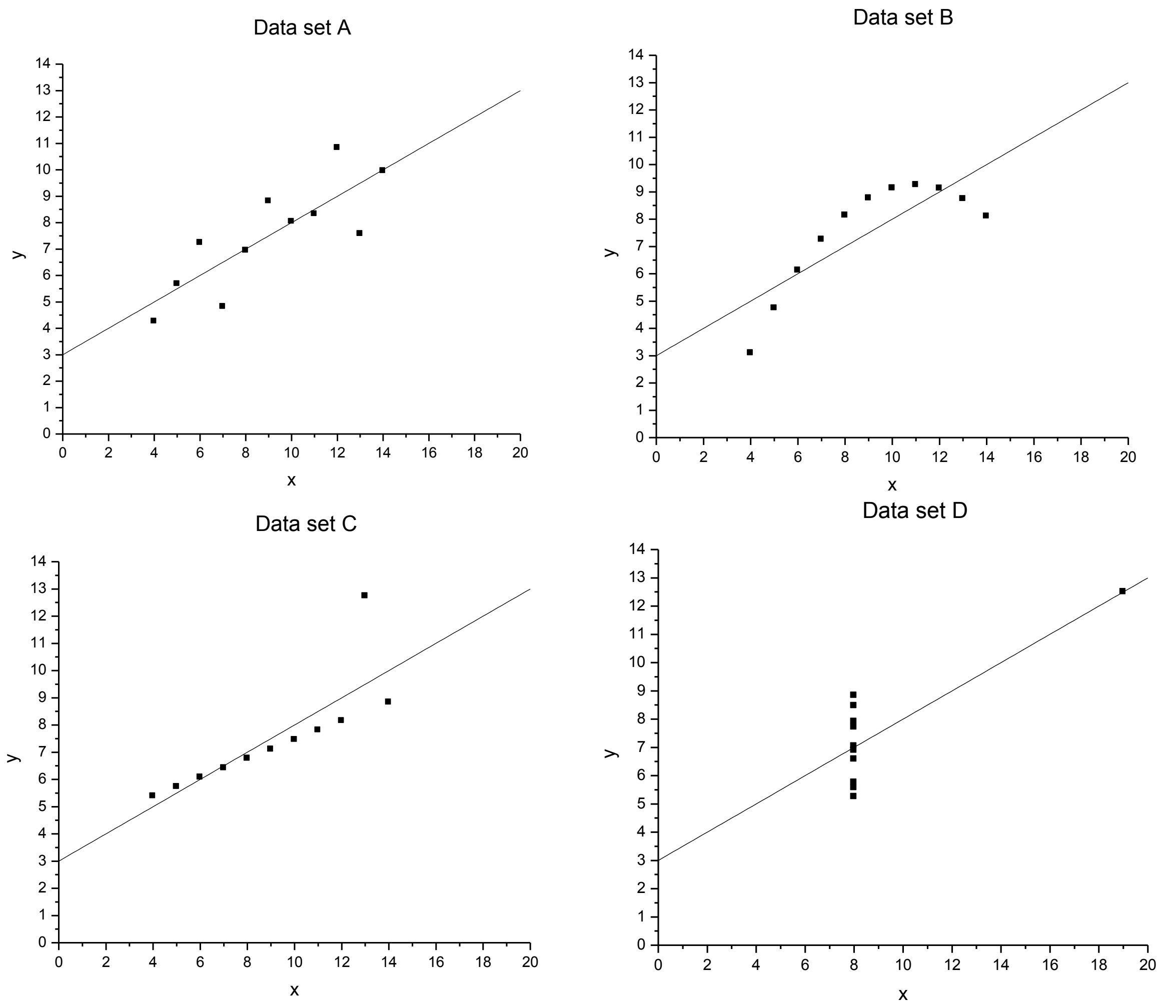
$$\bar{y} = \frac{1}{n} \sum_i^n y_i$$

Sum of squares of residuals :

$$SS_{\rm err} = \sum (y_i - f_i)^2$$

R-squared: Be careful!

- Anscombe's quartet: statistical misinterpretation of data
  - All 4 model fits have the same mean, variance, and $R^2 = 0.82$

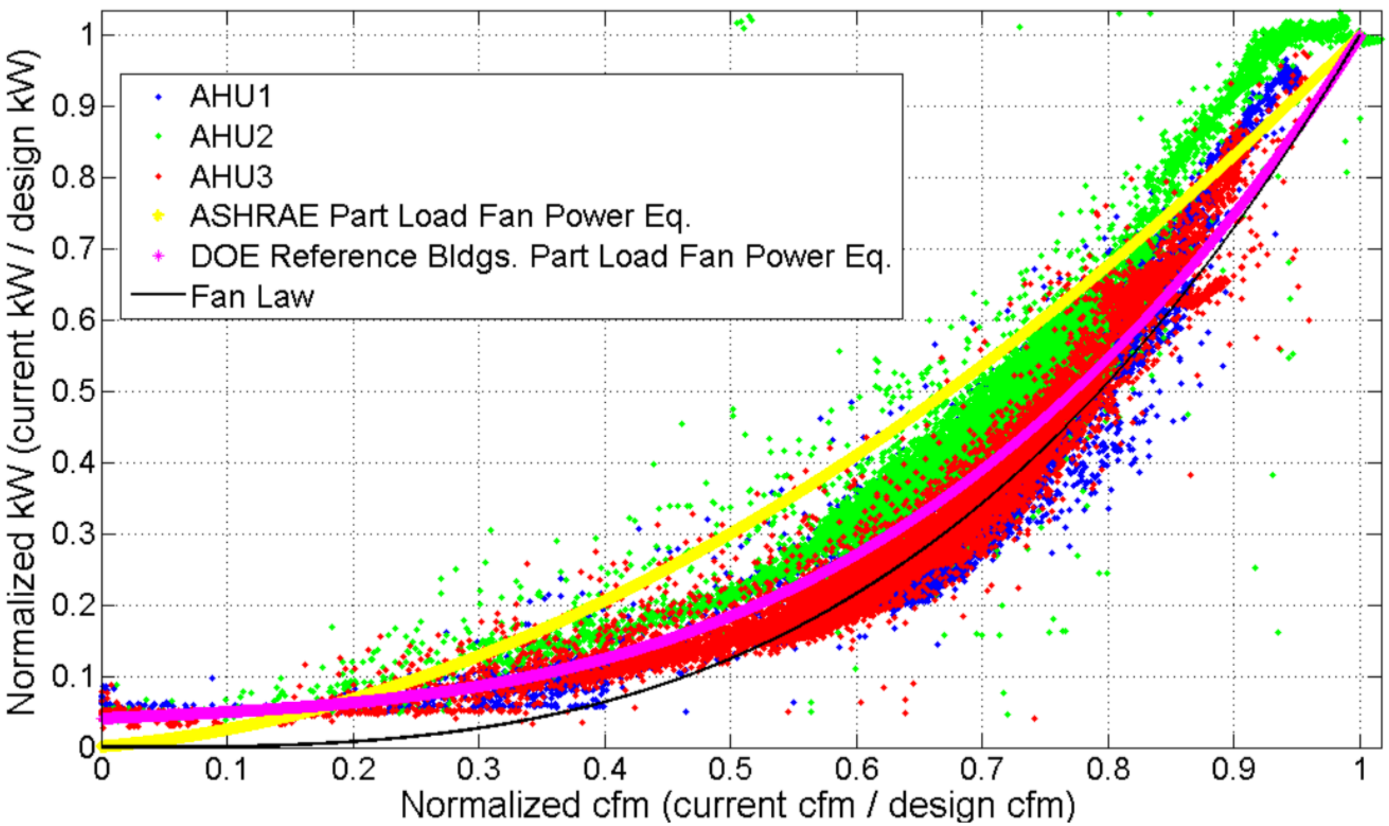


Be sure to graph and visually analyze your data before you do anything else!

Example Model Fit

$$P_{fan} = 0.0408 + 0.088X + 0.073X^2 + 0.943X^3$$

- P_{fan}: Fraction of full load power (current power/design power)
- X: Part load ratio of fan operation (actual cfm/design cfm)



Salimifard, Masters Thesis, Penn State

QUALITY ASSURANCE (QA)
QUALITY CONTROL (QC)

QA/QC

How do we consider QA/QC in our experiments?

Design of Experiment Phase

- Define objective(s)
  - What question(s) are you trying to answer?
  - How will you know you are finished?
  - What is the design of experiment process?

- Choose the following items:
  - Factors of interest
  - Parameters to measure
  - Experiments control method(s)
  - Data analysis techniques

QA/QC

Measuring Phase

- Use appropriate measuring techniques
- Utilize compatible sensors
- Collect sufficient data (including repetition)
- Record all available conditions / parameters
- Use experiment control methods

Data Analysis Phase

- Establish a hypothesis testing
- Develop graphs and tables
- Omit outliers
- Identify inaccurate readings
- Use statistical tools (e.g., regression)
- Interpret the results and draw conclusions
- Prepare to revise the hypothesis

QA/QC

- For more information, consult EPA Requirements for Quality Assurance Project Plans
  - EPA QA/R-5

United States Environmental Protection Agency

Office of Environmental Information Washington, DC 20460

EPA/240/B-01/003 March 2001

EPA Requirements for Quality Assurance Project Plans

EPA QA/R-5

https://www.epa.gov/quality/epa-qar-5-epa-requirements-quality-assurance-project-plans