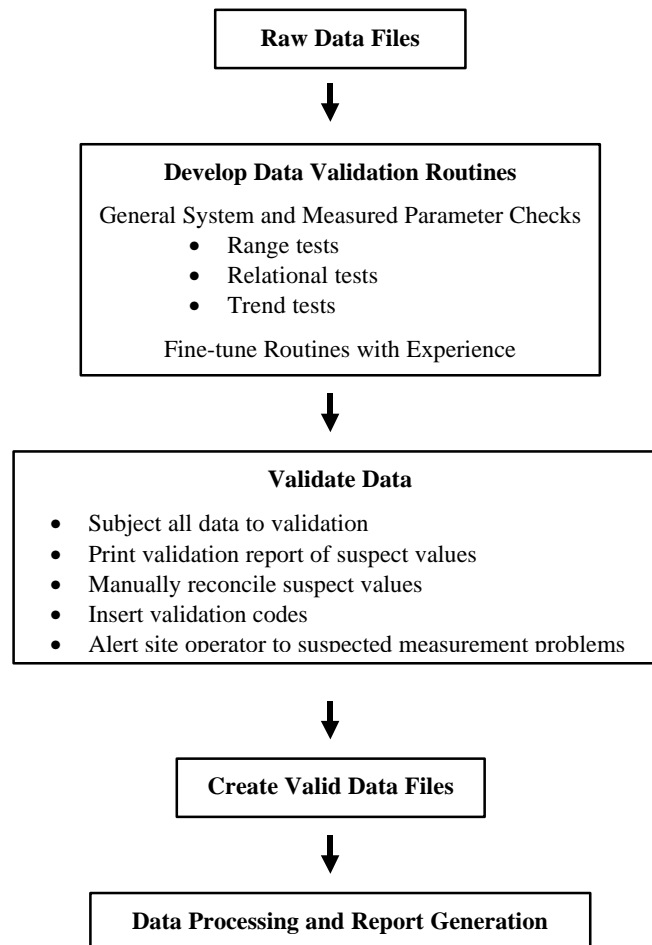


DATA VALIDATION, PROCESSING, AND REPORTING

After the field data are collected and transferred to your office computing environment, the next steps are to validate and process data, and generate reports. The flow chart presented in Figure 9.1 illustrates the sequence and roles of these steps.

Figure 9.1
Data Validation Flowchart



Data validation is defined as the inspection of all the collected data for completeness and reasonableness, and the elimination of erroneous values. This step transforms raw data into validated data. The validated data are then processed to produce the summary reports you require for analysis. This

step is also crucial to maintaining high rates of data completeness during the course of the monitoring program. Therefore data must be validated as soon as possible, within one - two days, after they are transferred. The sooner the site operator is notified of a potential measurement problem, the lower the risk of data loss.

9.1 DATA VALIDATION METHODS

Data can be validated either manually or automatically (computer-based). The latter is preferred to take advantage of the power and speed of computers, although some manual review will always be required. Validation software may be purchased from some data logger vendors, created in-house using popular spreadsheet programs (e.g., Microsoft Excel, Quatro Pro, Lotus 123), or adapted from other utility environmental monitoring projects. An advantage of using spreadsheet programs is that they can also be used to process data and generate reports. These programs require an ASCII file format for imported data; the data logger's data management software will make this conversion if binary data transfer is used.

There are essentially two parts to data validation, data screening and data verification.

- **Data Screening:** The first part uses a series of validation routines or algorithms to screen all the data for suspect (questionable and erroneous) values. A suspect value deserves scrutiny but is not necessarily erroneous. For example, an unusually high hourly wind speed caused by a locally severe thunderstorm may appear on an otherwise average windy day. The result of this part is a data validation report (a printout) that lists the suspect values and which validation routine each value failed.
- **Data Verification:** The second part requires a case-by-case decision on what to do with the suspect values — retain them as valid, reject them as invalid, or replace them with redundant, valid values (if available). This part is where personal judgment by a qualified person familiar with the monitoring equipment and local meteorology is needed.

Before proceeding to the following sections, you should first understand the limitations of data validation. There are many possible causes of erroneous data: faulty or damaged sensors, loose wire connections, broken wires, damaged mounting hardware, data logger malfunctions, static discharges, sensor calibration drift, and icing conditions, among others. The goal of data validation is to detect as many significant errors from as many causes as possible. Catching all the subtle ones is impossible. For example, a disconnected wire can be easily detected by a long string of zero (or random) values, but a loose wire that becomes disconnected intermittently may only partly reduce the recorded value yet keep it within reasonable limits. Therefore, slight deviations in the data can escape detection (although the use of redundant sensors can reduce this possibility). Properly exercising the other quality assurance components of the monitoring program will also reduce the chances of data problems.

To preserve the original raw data, make a copy of the original raw data set and apply the validation steps to the copy.

The next two subsections describe two types of validation routines, recommend specific validation criteria for each measurement parameter, and discuss the treatment of suspect and missing data.

A. Validation Routines

Validation routines are designed to screen each measured parameter for suspect values before they are incorporated into the archived data base and used for site analysis. They can be grouped into two main categories, general system checks and measured parameter checks.

1. General System Checks

Two simple tests evaluate the completeness of the collected data:

- **Data Records:** The number of data fields must equal the expected number of measured parameters for each record.
- **Time Sequence:** Are there any missing sequential data values? This test should focus on the time and date stamp of each data record.

2. Measured Parameter Checks: These tests represent the heart of the data validation process and normally consist of range tests, relational tests, and trend tests.

- **Range Tests:** These are the simplest and most commonly used validation tests. The measured data are compared to allowable upper and lower limiting values. Table 9.1 presents examples of range test criteria. A reasonable range for most expected average wind speeds is 0 to 25 m/s. However, the calibration offset supplied with many calibrated anemometers will prevent zero values. Negative values clearly indicate a problem; speeds above 25 m/s are possible and should be verified with other information. The limits of each range test must be set so they include nearly (but not absolutely) all of the expected values for the site. Technicians can fine-tune these limits as they gain experience. In addition, the limits should be adjusted seasonally where appropriate. For instance, the limits for air temperature and solar radiation should be lower in winter than in summer.

Table 9.1
Sample Range Test Criteria

*Sample Parameter	Validation Criteria
Wind Speed: Horizontal	
• Average	offset < Avg. < 25 m/s
• Standard Deviation	0 < Std. Dev. < 3 m/s
• Maximum Gust	offset < Max. < 30 m/s
Wind Direction	
• Average	0° < Avg. ≤ 360°
• Standard Deviation	3° < Std. Dev. < 75°
• Maximum Gust	0° < Max. ≤ 360°
Temperature	(Summer shown)
• Seasonal Variability	5°C < Avg. < 40°C
Solar Radiation	(Optional: Summer shown)
• Average	offset ≤ Avg. < 1100 W/m ²
Wind Speed: Vertical	(Optional)
• Average ^{**} (F/C)	offset < Avg. <± (2/4) m/s
• Standard Deviation	offset < Std. Dev. <± (1/2) m/s
• Maximum Gust	offset < Max. < ± (3/6) m/s
Barometric Pressure	(Optional: sea level)
• Average	94 kPa < Avg. < 106 kPa
ΔT	(Optional)
• Average Difference	>1.0° C (1000 hrs to 1700 hrs)
• Average Difference	<-1.0° C (1800 hrs to 0500 hrs)

* All monitoring levels except where noted.

** (F/C): Flat/Complex Terrain

If a value meets a criterion, that check considers the value valid. However, most parameter values should have several criteria checks, because a single criterion is unlikely to detect all problems. For example, if a frozen wind vane reports an average direction of exactly 180° for six consecutive ten-minute intervals, the values would pass the 0-360° range test, but the stationary vane would report a standard deviation of zero and be flagged as suspect.

Table 9.2

Sample Relational Test Criteria

*Sample Parameter	Validation Criteria
Wind Speed: Horizontal	
• Max Gust vs. Average	Max Gust \leq 2.5 * Avg.
• 40 m/25 m Average Δ **	\leq 2.0 m/s
• 40 m/25 m Daily Max Δ	\leq 5 m/s
• 40 m/10 m Average Δ	\leq 4 m/s
• 40 m/10 m Daily Max Δ	\leq 7.5 m/s
Wind Speed: Redundant	(Optional)
• Average Δ	\leq 0.5 m/s
• Maximum Δ	\leq 2.0 m/s
Wind Direction	
• 40m/25 m Average Δ	\leq 20°

* All monitoring levels except where noted.

** Δ : Difference

- **Relational Tests:** This comparison is based upon expected physical relationships between various parameters. Table 9.2 gives examples of relational test criteria. Relational checks should ensure that physically improbable situations are not reported in the data without verification; for example, significantly higher wind speeds at the 25 m level versus the 40 m level.
- **Trend Tests:** These checks are based on the rate of change in a value over time. Table 9.3 lists sample trend test criteria. An example of a trend that indicates an unusual circumstance and a potential problem is a change in air temperature greater than 5°C in one hour.

The examples of validation criteria in Tables 9.1, 9.2, and 9.3 are not exhaustive, nor are they necessarily apply to all sites. With use, technicians will learn which criteria are most often triggered and under what conditions. For example, some criteria may almost always be triggered under light wind conditions, yet the data are valid. This occurrence may argue for one set of criteria under light wind conditions (below 4 m/s) and another set for stronger winds. Therefore the technician(s) should modify criteria or create new ones as needed.

Table 9.3

Sample Trend Test Criteria

*Sample Parameter	Validation Criteria
Wind Speed Average	(All sensor types)
• 1 Hour Change	$<$ 5.0 m/s
Temperature Average	
• 1 Hour Change	\leq 5°C
Barometric Pressure Average	(Optional)
• 3 Hour Change	\leq 1 kPa
Δ Temperature	(Optional)
• 3 *Hour Change	Changes sign twice

* All monitoring levels except where noted.

A secondary benefit of the data validation process is that the person(s) directly involved in the validation process will become very familiar with the local wind climatology. The behavior of the wind under various weather conditions will become apparent, as will the relationship between various parameters. This is an invaluable experience that cannot be appreciated solely by poring over monthly summary tables and may be important for evaluating the impact of the local meteorology on wind turbine operation and maintenance.

Special Note: Some data loggers and their data retrieval software record the system battery voltage for each averaging interval. Range and relational tests can be incorporated into your wind data validation routines to check for a reduction in battery voltage that may indicate a system problem.

B. Treatment of Suspect and Missing Data

After the raw data are subjected to all the validation checks, what should be done with suspect data? Some suspect values may be real, unusual occurrences while others may be truly bad. Here are some guidelines for handling suspect data:

1. Generate a validation report (printout or computer-based visual display) that lists all suspect data. For each data value, the report should give the reported value, the date and time of occurrence, and the validation criteria that it failed.

2. A qualified person should examine the suspect data to determine their acceptability. Invalid should be assigned and replaced with a validation code. Table 9.4 gives some examples. A common designation for data rejection is assigning a -900 series validation code, with numbers that represent various rejection explanations. Operation and maintenance logs or site temperature data should be reviewed to determine the code.
3. If redundant sensors are used, replace a rejected value from the primary sensor with a substitute one from the redundant sensor as long as the redundant sensor's data passed all the validation criteria.
4. Maintain a complete record of all data validation actions for each monitoring site in a Site Data Validation Log, an example of which is provided at the end of this chapter. This document should contain the following information for each rejected and substituted value:
 - File name
 - Parameter type and monitoring height
 - Date and time of flagged data
 - Validation code assigned and explanation given for each rejected datum
 - The source of the substituted values.

Table 9.4
Sample Validation Codes

Code	Rejection Criteria
-990	Unknown event
-991	Icing or wet snow event
-992	Static voltage discharge
-993	Wind shading from tower
-995	Wind vane deadband
-996	Operator error
-997	Equipment malfunction
-998	Equipment service
-999	Missing data (no value possible)

Important: Maintain raw and validated data files separately. Differentiate the files by assigning different extensions to the file names. For example, the file extension for the raw data file could be designated as “.raw” and the verified data file as “.ver.” Valid data can then be compiled into a master data file for further data reporting and archiving.

C. Data Recovery

The data recovery rate is defined as the number of valid data records collected versus that possible over the reporting period and should be determined for each primary wind sensor (for all levels at each site). The method of calculation is as follows:

$$\text{Data Recovery Rate} = \frac{\text{Data Records Collected}}{\text{Data Records Possible}} \times 100$$

where

$$\text{Data Records Collected} = \text{Data Records Possible} - \text{Number of Invalid Records}$$

For example, the total possible number of ten-minute records in December is 4,464. If 264 records were deemed invalid, the number of valid data records collected would be 4,200 (4,464 - 264). The data recovery rate for this example would be:

$$\text{Data Recovery Rate} = \left(\frac{4,200}{4,464} \right) \times 100 = 94.1\%$$

9.2 DATA PROCESSING AND REPORTING

When the data validation step is complete, the data set must be subjected to various data processing procedures to evaluate the wind resource. This typically involves performing calculations on the data set, as well as binning (sorting) the data values into useful subsets based on your choice of averaging interval. From this, informative reports can be produced, such as summary tables and performance graphs. Data processing and reporting software are available from several sources, including certain data logger manufacturers and vendors of spreadsheet, data base, and statistical software.

Hourly averages are normally used for reporting purposes. Compiling validated ten-minute data subsets into an hourly average data base can be performed using available data processing and reporting software. Whatever method is used, care must be taken to not include a flagged data point or -900 series flag code when computing the hourly average.

The basic parameter set allows for the determination and presentation of many valuable wind characterization tools. Table 9.5 presents the recommended monthly data reports. Examples of selected reports are provided at the end of this chapter.

Except for fully programmable data loggers, the wind shear exponent, turbulence intensity, and wind power density are typically not internal processing functions of most data loggers. These parameters can be easily calculated using a spreadsheet software application to obtain hourly and monthly averages. A description of each parameter and calculation method is presented in detail below.

1. Vertical wind shear exponent: Wind shear is defined as the change in horizontal wind speed with a change in height. The wind shear exponent (a) should be determined for each site, because its magnitude is influenced by site-specific characteristics. The 1/7th power law (as used in the initial site screening) may not be applied for this purpose, as actual shear values may vary significantly from this value. Solving the power law equation for a gives

$$a = \frac{\text{Log}_{10} \left[\frac{v_2}{v_1} \right]}{\text{Log}_{10} \left[\frac{z_2}{z_1} \right]}$$

where

v_2 = the wind speed at height z_2 ; and

Table 9.5
Sample Monthly Data Reporting

Report Products	Presentation	Reporting Heights
Mean hourly diurnal wind speed	Graph/Table	10 m, 25 m and 40 m
Joint wind speed & direction frequency distribution (16 sectors)	Table	25 m and 40 m
Wind speed frequency distribution: 0.5 m/s (1 mph) bins	Graph/Table	10 m, 25 m and 40 m
Mean hourly diurnal temperature	Graph/Table	3 m
Mean hourly wind shear	Graph/Table	Between 25 m and 40 m
Mean turbulence intensity	Graph/Table	10 m, 25 m and 40 m
Wind rose	Graph	25 m and 40 m
Mean hourly wind power density	Graph/Table	40 m

v_i = the wind speed at height z_i .

2. Turbulence intensity: Wind turbulence is the rapid disturbances or irregularities in the wind speed, direction, and vertical component. It is an important site characteristic, because high turbulence levels may decrease power output and cause extreme loading on wind turbine components. The most common indicator of turbulence for siting purposes is the standard deviation (\mathbf{s}) of wind speed. Normalizing this value with the mean wind speed gives the turbulence intensity (TI). This value allows for an overall assessment of a site's turbulence. TI is a relative indicator of turbulence with low levels indicated by values less than or equal to 0.10, moderate levels to 0.25, and high levels greater than 0.25. TI is defined as

$$TI = \frac{\mathbf{s}}{V}$$

where

\mathbf{s} = the standard deviation of wind speed; and
 V = the mean wind speed.

3. Wind power density: Wind power density (WPD) is a truer indication of a site's wind energy potential than wind speed alone. Its value combines the effect of a site's wind speed distribution and its dependence on air density and wind speed. WPD is defined as the wind power available per unit area swept by the turbine blades and is given by the following equation:

$$WPD = \frac{1}{2n} \sum_{i=1}^n \mathbf{r} \times v_i^3 \quad (\text{W/m}^2)$$

where

n = the number of records in the averaging interval;
 \mathbf{r} = the air density (kg/m^3); and
 v_i^3 = the cube of the i^{th} wind speed (m/s) value.

As is evident by the summation sign (Σ), this equation should only be used for all wind speed values ($n > 1$) during an averaging period and not for a single long term average (e.g., monthly, yearly). The reason is based on the normal variability of the wind and the cubic wind speed relationship. The following example is provided to illustrate this point.

Assume that over a two hour period the average wind speed at a site is 6.7 m/s (15 mph); 4.5 m/s (10 mph) the first hour and 8.9 m/s (20 mph) the next. Calculate the WPD using the combined two hour average ($n=1$) and then with the two distinct hourly average values ($n=2$). Applying standard temperature and pressure (101,325 Pa and 288 °K) to the above equation, the calculated WPD using the overall average value is 184 W/m^2 while that using the two average values is 246 W/m^2 . The latter value represents the average of that calculated for 4.5 m/s (55 W/m^2) and 8.9 m/s (438 W/m^2). There is actually 34% more power available at the site than would have been realized if the equation was used incorrectly.

The air density term in the WPD must be calculated. It depends on temperature and pressure (thus altitude) and can vary 10% to 15% seasonally. If the site pressure is known (e.g., measured as an optional parameter), the air density values with respect to air temperature can be calculated from the following equation:

$$\mathbf{r} = \frac{P}{R \times T} \quad (\text{kg/m}^3)$$

where

P = the air pressure (Pa or N/m²);
 R = the specific gas constant for air (287 J/kg·K); and
 T = the air temperature in degrees Kelvin (°C+273).

If site pressure is not available, air density can be estimated as a function of site elevation (z) and temperature (T) as follows:

$$r = \left(\frac{P_o}{R \cdot T} \right) \exp\left(\frac{-g \cdot z}{R \cdot T}\right) \text{ (kg/m}^3\text{)}$$

where

P_o = the standard sea level atmospheric pressure (101,325 Pa), or the actual sea level-adjusted pressure reading from a local airport;
 g = the gravitational constant (9.8 m/s²); and
 z = the site elevation above sea level (m).

Substituting in the numerical values for P_o , R , and g , we get:

$$r = \left(\frac{353.05}{T} \right) \exp^{-0.034\left(\frac{z}{T}\right)} \text{ (kg/m}^3\text{)}.$$

This air density equation can be substituted into the WPD equation for the determination of each hourly average value.

9.3 QUALITY ASSURANCE REPORTING

A component of your monitoring program documentation should be a periodic report on the program's adherence to the quality assurance plan (described in Section 2.4). The field and/or data analysis staff should prepare the report and submit it to the project manager (or quality assurance coordinator) monthly or quarterly. The report should address the following topics:

- Dates of operation and maintenance site visits: activities and findings
- Description of monitoring problems and corrective actions taken
- Record of equipment calibration activities (if applicable)
- Data validation findings and actions taken
- Data recovery rate.

A sample Quality Assurance Report Form is provided at the end of this chapter.

SAMPLE MONTHLY DATA REPORTS

Wind Direction Frequency: Wind Rose

SAMPLE QUALITY ASSURANCE REPORT FORM

Reporting Period		*Page ____ of ____	
Prepared By		Date Prepared	

**If more than four sites, copy and attach additional forms.*

For each monitoring site, please provide the following:

Site QA Information	Site:	Site:	Site:	Site:
1. Were O&M Site Visits Performed? (Circle)	YES NO	YES NO	YES NO	YES NO
<ul style="list-style-type: none"> <i>• If YES, attach copy(s) of O&M documentation for each site visit describing the activities, findings, monitoring problems and corrective actions taken. Documentation is attached: Check here ___(✓).</i> <i>• If NO, provide a detailed explanation (e.g., harsh weather conditions) in the Comments & Information section provided below.</i> 				
2. Were instruments in compliance? (Circle)	YES NO	YES NO	YES NO	YES NO
<ul style="list-style-type: none"> <i>• If NO, provide a detailed explanation in the Comments & Information section provided below.</i> 				
3. New/replacement equipment installed? (Circle)	YES NO	YES NO	YES NO	YES NO
<ul style="list-style-type: none"> <i>• If YES, attach revised Site Information Log. Documentation is attached: Check here ___(✓).</i> 				
4. Raw data appended to site raw data base? (Circle)	YES NO	YES NO	YES NO	YES NO
<ul style="list-style-type: none"> <i>• If NO, provide a detailed explanation in the Comments & Information section provided below.</i> 				
5. Data validated and appended to validated data base? (Circle)	YES NO	YES NO	YES NO	YES NO
<ul style="list-style-type: none"> <i>If YES, attach Site Data Validation Log. Documentation is attached: Check here ___(✓). If NO, provide a detailed explanation below.</i> 				
6. Data back-up performed and properly stored? (Circle)	YES NO	YES NO	YES NO	YES NO
<ul style="list-style-type: none"> <i>• If NO, provide a detailed explanation in the Comments & Information section provided below.</i> 				
7. Was data recovery [≥] 90%? (Circle)	YES NO	YES NO	YES NO	YES NO
<ul style="list-style-type: none"> <i>• If NO, provide a detailed explanation (e.g., icing conditions) in the Comments & Information section provided below.</i> 				

Comments & Information:
<i>(Continue on back if necessary)</i>