

Data Quality Control

GADS Wind Training Module 14
April 2019 - Final

RELIABILITY | ACCOUNTABILITY



- This module will review:
 - Dividing time
 - Data Quality Process
 - Data Quality Levels Defined
 - Level 1 - Credentials
 - Level 2 – Record Identification
 - Level 3 – Data Checks
 - Level 3 – Land Mine
 - Level 4 – Component ID
 - Level 5 – Component Cross Checks

Calendar Turbine Hours (CaTH) 744 hrs.

Inactive Turbine Hours (ITH) 44 hrs.

10 hrs. Inactive Reserve (IRTH)

30 hrs. Mothball (MBTH)

4 hrs. Retired (RTH)

Available Turbine Hours (ACTH) 700 hrs.
Previously Known as Period Turbine Hours (PDTH)

600 hrs. Available

100 hrs. Unavailable

400 hrs. Contact (CTH)

50 hrs. Reserve Shutdown (RSTH)

60 hrs. Forced Outage (FTH)

30 hrs. Maintenance Outage (MTH)

10 hrs. Planned Outage (PTH)

150 hrs. Resource Unavailable (RUTH)

10 hrs. Non OMC (FO)

10 hrs. OMC (FO)

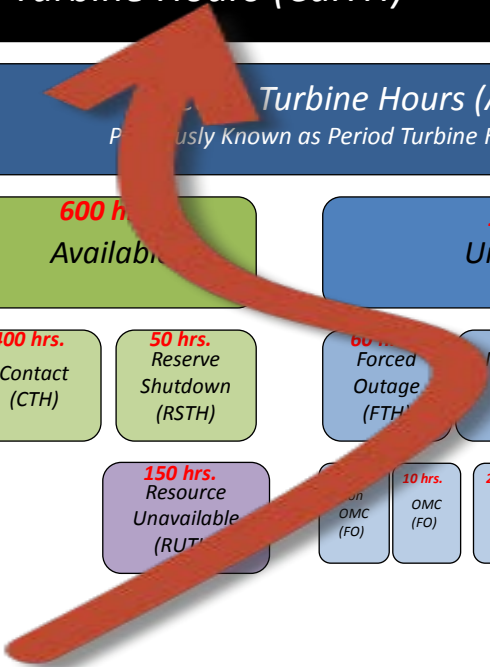
28 hrs. Non OMC (MO)

2 hrs. OMC (MO)

8 hrs. Non OMC (PO)

2 hrs. OMC (PO)

Data Quality is ensuring that you are an authorized user for the plant being submitted and that all the hours roll-up through the various hourly levels correctly.



Data check points are identified in appendix J of the Wind DRI.

Using the Dividing Time chart, most of the check points are simply a roll-up from bottom to top equaling the Calendar Turbine Hours for the period.

- As data is submitted there are actually 5 levels in the data quality process
- Many of the quality checks can be found in Appendix J with their definition and how the indicator is calculated
- If an error is found at a particular level, checking is stopped and a message displayed identifying the location of the error
- Many of the error messages will reference **specific items** in Appendix J
- Sub-group will only reject rows that fail. All the other rows will be accepted
- Performance and Component will reject all rows if one row fails

Appendix J – Data Quality Control

GADS data should be reviewed for the following potential discrepancies before submission. This list is by no means comprehensive, but data not meeting these minimum requirements will be rejected. Reporting is done on a monthly basis, submitted no later than 45 days after the end of the quarter. At the time of the writing of this document, reporting is voluntary.

1. **Calendar Turbine Hours (CalTH)** – The total number of turbine hours in a month is equal to the number of turbines times the number of hours in a month. Example: January has 31 days or 744 hours. If the plant has 100 turbines then there would be 74,400 turbine hours for January.

$$\text{CalTH} = (\text{Hours in Month}) \times (\text{Number of Turbines})$$

- The data quality process is multi-layered beginning with security check to component verification
- Many of the data checks can be found in appendix J
- Error checking works its way through each of the three records as submitted. If an error is found immediate feedback is provided
- Error messages will reference specific errors in appendix J. The error will be defined and the method of calculation shown
- The Sub-group record will reject only rows with errors
- The Performance and Component records will reject the entire record if any rows are found with an error

Data quality checks can be identified into 5 distinct sections

1. Credentials – Are you authorized
2. Record Identification – All three records – Which Sub-group and when
3. Data Check – Do the numbers make sense
4. Component Identification – Correct component codes
5. Does component data match performance data

In order for the data checks to work properly, they must be submitted or presented in the proper order. Failure to follow the proper sequence will cause the data to be rejected.

- 1. First there must be a valid Sub-group record for the data**
- 2. Performance data can be submitted**
- 3. Component data can be submitted after performance**

If revised Performance or Component data is submitted, steps 2 and 3 must be repeated in order

The order that records are submitted is important. If there is no Sub-group record, Performance and Component data as nothing to check against. Component data is checked against Performance totals. If the totals do not match an error is generated.

Data can be revised at any time with a few qualifiers:

- Correct the Performance record first, then correct the component record
- Make sure that the number of turbines in the sub-group has not changed from current to the previous period
- If the number of turbines is different for the correction period than what is current, errors will occur because the calendar hours will not add up. In this case adjust the number of turbines in the Sub-group record for the time period of the correction. Correct the Performance and Component reports. Change the turbine count back to its current value.

Level 1 – Credentials determine if you are an authorized user and the utility you can report for. There are several steps:

1. Is the company registered with the NERC Wind GADS program
2. Does it have a valid NERC compliance number or Voluntary reporting number (aka Utility ID)
3. Is the user authorized to report for this company (Correct E-mail address)
4. Is the user's PIN correct

Each reporting company is required to have a NERC Compliance ID or a Voluntary reporting number. Only one is need and if a NERC Compliance ID is available, it should always be used. Multiple plants can be reported under one NERC Compliance ID. There is no need for ID's for each plant.

The compliance officer identified for the NERC Compliance number is responsible for identifying the individuals responsible for reporting data. These individuals can be employees or contractors that have been identified by the company.

Level 2 – Record Identification – All data is submitted at the Sub-group level for the 3 record types (Sub-group, Performance and Component). The first few columns in each record provides identification information and the following checks are made:

1. Is the Utility ID valid
2. Is the user authorized to report for this Utility ID
3. Is the Sub-group ID valid and associated with the Utility ID
4. Is the Sub-group ID active
5. Is this a new Sub-group ID request
6. Is there already a Sub-group with the same configuration
7. Are the column headers correct and in the right order
8. Are there any special characters or formatting
9. Some fields require specific inputs and are checked against allowable entries located in Appendix F of the Wind DRI. Also, some columns have range limits

For the Performance and Component records there are 2 additional identification columns, Reporting Month and Year. Future years and months will cause an error.

Each record is check for:

- Is the NERC Compliance number or voluntary ID valid
- Can the user report for the identified Utility ID
- Does the Sub-group ID belong to the Utility ID and valid
- Is the Sub-group ID active. Performance and Component data cannot be reported for inactive Sub-groups
- If there is no Sub-group ID is this a request for a new Sub-group
- When requesting a new Sub-group is there a duplicate?
- Are the column headers correct and in the right order?
- No special characters like commas, dashes, underlines, question, stars etc
- Are entries checked against allowable entries in appendix F and are they within range. Like Latitude -180 to +180

Performance and Component records will also require month and year in the correct format

Level 3 – The Performance data checks are detailed in Appendix J of the Wind DRI. When an error occurs during submission a message is generated indicating the data check number that failed and a brief description. As an example:

1. Data Check 1 checks to see if all Calendar Turbine Hours (CalTH) are accounted for
2. From the Sub-group data the number of turbines (WTG) is obtained (100 in the example below)
3. The reporting year (leap year) and month is obtained from the Performance record and the number of days identified (31 in the example)
4. CalTH is calculated (31days x 100 WTG x 24 Hrs./Day = 74,400 hours)
5. If the total hours as demonstrated on the dividing time page (3) do not add up to the CalTH an error is generated and the entire data record is rejected.

1. **Calendar Turbine Hours (CalTH)** – The total number of turbine hours in a month is equal to the number of turbines times the number of hours in a month. Example: January has 31 days or 744 hours. If the plant has 100 turbines than there would be 74,400 turbine hours for January.

$$\text{CalTH} = (\text{Hours in Month}) \times (\text{Number of Turbines})$$

2. **Turbine State** – Turbines are either in an Active state (PDTH) or an Inactive state (ITH). The sum of the two equals the CalTH.

Data Checks – Performance Record

At this point the recorder has been verified and an active Sub-Group is verified. Actual hours and MWH are verified as legitimate (Not necessarily correct) using various test as described in Appendix J. The object is to catch as many errors as possible before data is entered into the data base. Most of the checks are common sense. See the Dividing Time module.

The hourly data checks all revolve around CalTH. When adding or retiring turbines, care must be used when updating the Sub-group record:

1. In the previous example there were 100 WTG in the Sub-Group or 74,400 CalTH
2. One WTG is retired mid month and the Sub-group record immediately adjusted. This reduced the WTG count to 99 or CalTH equal to 73,656.
3. Because the WTG was retired midmonth, the reported hours were somewhere between 74,400 and 73,656, generating an error

The proper steps are:

1. Report all the hours for 100 turbines during the month of retirement. Make no adjustment to the Sub-group record
2. Submit a revised Sub-group record (100 to 99 WTG) with an effective date the first full month after retirement
3. Report on 99 WTG or 73,656 CalTH the first full month after retirement.

When repowering, retiring or adding turbines care must be taken when adjusting the Sub-group record. When submitting Performance data the current version of the Sub-group record will be used. If monthly reporting occurs in the wrong order, calendar hours will not match and the record will fail testing.

Level 4 – Component reporting breaks down the 3 outage types (PO, MO, FO) into individual systems and components. The different systems and components are described in Appendix C. Level 4 checks for valid entry codes. If an invalid entry code is used, an error is generated and the entire record rejected.

Appendix C – System-Component Codes

The following tables list available components for each system and the code to enter for each component:

System	Component	Entry
Balance of Plant	General	674
	Battery Systems	1042
	Breakers	1037
	CT / PT	1041
	Fuses	1044
	Main Substation Transformer	1038
	Metering and Relays	655
	Overhead Lines	652
	Pad Mount-Multiple WTG	651
	Preventative Maintenance	682
	SCADA	656
	Site Communication	1043
	Site Reactive Power Comp	654
	Substation	653
	Switches	1039

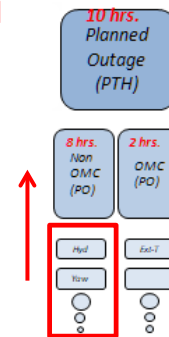
The component codes are very general and cover many types of turbines. They are not meant to be exhaustive. If there is the need for additional codes please let GADS know.

Level 5 – Performance records are entered first. Component records break down the 3 outage states into (PO, MO & FO) into systems and components. When the component record is submitted several data checks occur:

1. Do all the PO hours in the Component record add up to the PO hours in the Performance report. Same for MO and FO
2. Do all the PO Outside Management Control (OMC) hours in the Component record add up to the PO OMC hours in the Performance report. Same for MO OMC and FO OMC

Remember that OMC is a sub-set of each outage type. The FO bucket contains all the FO OMC. If FO < FO OMC occurs, an error will be generated

3. If derates are submitted, do they roll-up properly in the Performance report
4. If delays are submitted, do they roll-up properly in the Performance report



The Performance report is the summary report. Component reporting is the detail behind many of the Performance report numbers. When reporting component data will fail if it does not roll-up to match the Performance report.



gadswind@nerc.net