Analysis of the Approaches Seeking Quality in Anemometric and Solarimetric Data for Energy Production

Angélica Luana Linhares; Luiz Claudio Ribeiro Galvão, Miguel Edgar Morales Udaeta; André Luis Veiga Gimenes
GEPEA - Energy Group of the Department of Energy and Electrical Automation Engineering of the EPUSP
São Paulo, Brazil
angelicallinhares@gmail.com ; udaeta@pea.usp.br

***Abstract*— The objective of this work is to examine the methods of quality analysis, used in anemometric and solarimetric data, through comparing with a compendium of national and international recommendations, and identify from a data set which used literally the methods on for treatment and what can cause the lack of the use these recommendations.**

The data analyzed refer to two measuring stations, both anemometric and solarimetric, belonging to a Strategic Project of Research and Development (R&D) in São Paulo. The criteria and recommendations to attain quality data are based on articles elaborated by the WMO - World Metereological Organization, NREL - National Renewable Energy Laboratory, BSRN - Base-line Surface Radiation Network). Both the anemometric data as the solarimetric data analyzed had followed only part of the rec-ommendations that aim at the attainment of good data. The results of the analysis showed that for a number of quality data, it is necessary a detailed work, which passes through data validation criteria suitable for the location of measurement, analyses of consistency limits, of historical surrounding series, comparison to estimated values and comparison between sensors, also following the recommendations of analyses. As regards to the presentation of three data sets (raw, marked and treated), it was noted it is essential in any series of data and characterizes an efficient way to be applied by the user, who can clearly, according to his need opt to which of the sets desires to use. Besides, afterwards the whole quality data, which had passed through a process of adequate tests to identify errors and imperfections in the series of data.

Keywords* — *Data Quality Analysis; Anemometric Data; Solarimetric Data.

I. INTRODUCTION

An essential factor for the development of projects on the use of energy potentials is to know the availability of the energy resource. A great scientific effort has been being developed in the last decade by various Brazilian groups of research to reduce the barrier due to scientifically reliable information of the available energy potential in domestic territory [1].

Studies guided by data from satellites supply an indication of the energy availability for a region, but they present due uncertainties of the adopted methodology and the space resolution where they had been carried through. Certainly, the most accurate estimate of the energy potential for a specific place is determined by the installation of precise instruments of measurement in the interest area.

Wissmann (2006) quotes that it is not enough to only measure,, it is necessary to process, to correct, to generate and to guarantee time and space consistency to the measured data as effective as possible.

Data used for energy potential need to pass through a rigorous process of quality control, once these data will be, later, used for energy forecast.

The measurement directed toward the calculation of solar and wind power potential energy, is generally added to a station or tower with specific tools, and defined by international standards in order to generate data close to the real available in loco.

Taking into account the need of good data for the analysis of energy potential, this work will analyze the development of the process of analysis of quality, based as standard the recommendations and methods for analysis of data quality from highly regarded institutes, such as WMO (World Meteorological Organization), BSRN (Baseline Surface Radiation Network) and NREL (National Renewable Energy Laboratory), for the solarimetric and anemometric data assessment, in data in loco

II. ANALISES OF DATA QUALITY

To obtain appropriate quality data, suitable measures must be taken in each one of the phases, such as: a good planning in accordance with the needs of the users, the election and installation of the equipment, the operation, the calibration of the equipment, the validation and the treatment of the data, the maintenance and the activities of training. Even facing these measures, it is still necessary to take into account that errors and imperfections can address a series due the factors that are out of the control as, for instance: the delay in implantation due to climatic changes or rains; or possible equipment exchange that in the installation had soon presented imperfections; or the interference of insects and birds with sensitive measures, amongst others.

The objective of controlling the data quality is to identify erroneous or questionable registers in these and to prevent that bad data are used in the analyses.

A complete evaluation of a quality system is of particular importance, therefore obvious errors, sometimes, remain in the data sets and interesting climatic characteristics are occasionally identified as errors [2]

The data quality assessments verify, identify and correct the data that seem to be incorrect or inconsistent. According to Golz (2005) and Golz, Einfalt and Chr. Michaelides (2006), the verification of the data quality is the process to analyze the data to characterize them as being suspect or non-suspect. The correction of the data modifies the comments that had been marked as being suspect in a way that they pass through the process of quality control after the correction. Thus, the effects of a good quality data analysis are: (i) to assure that the data are homogeneous, consistent and continuous and (ii) to signal any abnormal data that need more inquiry and eventual correction [3].

The quality control happens by means of the use of a series of filters applied to the identification of errors in data collected for sensors and comments. Particular care must be taken when controlling the quality. The filters cannot exclude real extreme registers, but at the same time they must exclude illegitimate registers, that if kept in the historical series would increase the frequency of extreme cases [4].

III. Evaluation And Consistency of the measured Data Quality

According to the Electrical Energy Research Center (CEPEL), there are several procedures for raw data evaluation and the qualification of these data can be made by criteria that become more elaborated along with the time of development of the studies, such as:

- If the measured value is contained inside of the physical limits of the variable;
- If there is consistency between different sensors;
- If there is historical consistency;
- Comparison with values estimated by models.

A. *Consistency between different sensors*

After the recognition of measurement range of an instrument or sensor it is feasible to observe the consistency between two sensors to inquire the veracity of the registered value. This technique uses redundancy to prove the consistency of the measured data and can be called Sensor Fusion.

In search of the increase of the reliability and the integrity of the sensory data, the use of more than a sensor is frequently carried out. The process of combination of the data from multiples sensors of the same kind or different kinds is called Sensor Fusion. Its main objective is to supply the systems with data with bigger quality, thus allowing to reduce the imperfections in decision-making processes [5].

B. *Consistency with the physical limits*

Consistency control of physical limits is in the detection of values that exceed a predetermined physical limit, the values identified as abnormal can be excluded by the specialized observer.

WMO (2004) describes that the objective of examining the consistency of the physical limits of a variable as being an act to verify if the instant data values (average of one or the sum) are within the acceptable reaching boundaries. The limits of different meteorological parameters depend on the climatic conditions of the place and the type of station and instruments.

1) Physical Limits of the Variable According to Measuring tool

Each equipment of measurement has a minimum and maximum limit registered. By means of these information, it is possible to detect if the equipment is operating properly or not. This is a first test of consistency of physical limits of the variable.

This range of measurement deals with the limit established by the sensor, in a way that any register out of this value certainly presents an error diagnosis, which can be originated from diverse sources such as the lack of calibration , or incorrect offset and slope . The range of each sensor or equipment is supplied in technical documents from the manufacturer by means of datasheet and/or manuals.

Error! Reference source not found. and Table 2 show the illustration of the measurement range of two sensors: an anemometer and a temperature and humidity sensor, in which it is possible to notice the specifications of limits of each one.

Table 1. Example of measurement Range of a Speed of the wind Sensor (Anemometer). Source: Kintech Engineering, 2015.

ANEMOMETER Technical Data	
FEATURE	DESCRIPTION/VALUE
Measuring range	0,3...75m/s
Precision	0,3...50 m/s 1% of measurement or < 0,2 m/s
Maximum speed	80 n/s (minimum 30 minutes)
Acceptable environmental conditions	-50...+80 C, any situation of relative humidity dew

Table 2. . Example of measurement Range of a temperature and humidity Sensor. Source: Kintech Engineering, 2015.

TEMPERATURE AND HUMIDITY - GALLTECK KPC 1/5 - Technical Data	
HUMIDITY	
FEATURES	DESCRIPTION/VALUE
Measuring Range	0…100% rh
Precision	(5…95%rh at +10…+40 °C) ±2% rh
Temperature Influence	<+10 °C, >+40°C; ,+0,1%/°C adcional
TEMPERATURE	
FEATURES	DESCRIPTION/VALUE
Measuring element	Pt100 class B (class 1/3 DIN under demand)
Measuring Range	-30…+70°C

2) Physical Limit of the variable according to Climatological Standard

Another consistency that must be taken into account are the values of limits according to Climate Normals in Brazil - Period 1961-1990, developed by the National Institute of Meteorology (INMET). The climate normals exhibit average values calculated for a period relatively long and uniform, covering three consecutive decades.

According to the World Meteorological Organization (WMO) itself, the climatic data are frequently more useful when compared with standardized normal values, thus considered those obtained according to its proper technical recommendations. Thus, it becomes of utmost importance the calculation and publication of Climate Normals [6].

These data are available by the INMET on the website, in the section Climate, where it is possible to get them ac-cording to the need of the user, being then generated an .XLS archive.

The consistency of the variable boundaries in focus with the climate normals offer a good analysis of the reality and behavior of the variable, thereby assisting the data to be-come most reliable and close to the real value.

3) Physical limit of the variable according to climatic conditions of the locality

The consistency of the variable physical limit must be based and also evaluated in accordance with the climatic conditions of the locality, since altitude, atmospheric pressure and the place where it is inside the planet Earth, substantially modify the climatic conditions. In the face of this fact, it is essential a deepened analysis of the conditions of the analyzed locality. It is possible the use of data from stations in the surroundings and data of satellites for analysis of this consistency.

The limits of each acceptable interval must be defined so that almost all (but not absolutely) the expected values for the place. The technicians can adjust these limits to the measure as they gain experience. Moreover, the limits must be adjusted seasonally. For example, the limits of air temperature and solar radiation must be lower in the winter than in summer [7].

C. Consistency with historical series

The consistency with historical series of data is one of the observing phases of more significant, because homogeneity tests can be applied and correlation among the data in hands. This historical series must cover a period enough for observing the seasonality of the variables in study.

The basic idea of this type of test is to determine the behavior of a data set compared with a similar set from another period. The test of historical consistency can be of three main types: checking the limits, successive or behavioral differences and relation of parameters. For this purpose, empirical or subjectively limits are set, from technical engineering experiences. [7]

This analysis is done in order to look for atypical behaviors for the locality of study. Normally, depending on the season, the locality or variable, the hourly values of concentration change without great jolts. If, when analyzing the data, sudden very high peaks are perceived, and with low values in the previous and also posterior hour, this value must be detached for the possible study of the reason that generated.

D. Comparison with values estimated by models

The comparison between estimated models can also be used for analysis of consistency of data. The methods of values estimate for the climatic data (solar radiation, wind speed, of winds direction, air humidity, rainfall, temperature) change according to the variable to be estimated.

These models of estimated values can be by methods: (i) physical, (ii) traditional statistics and (iii) hybrid. The physical method uses equations to describe mathematically the physical processes that occur in the atmosphere, physical models generally use global databases of meteorological measures or atmospheric mesoscale models, and require computational systems to obtain more precise results [8]; the statistical method, in turn, works as part of applied mathematics in order to find plausible results; and the hybrid method has the characteristic of mixing the methods in order to remedy the weaknesses found to estimate a value.

This comparison between models and the measured data serves for an analysis of how close or not the data are to the estimated value models, in order to also parameterize the phenomena and abnormalities detected in the data.

IV. DESCRIPTION OF MEASURING, ANEMOMETRIC AND SOLARIMETRIC PILOT UNITS

Measurement units are linked to the ANEEL (National Electric Energy Agency) Strategic Project of Research and Development R&D-0061-0043/2014, titled "Integration of Wind Power Plants and Photovoltaic Solar Plants to Existing Hydroelectric Power Plants: An approach through Energetic Complementation with the practical support of pilot plants with different technologies ", developed in the city of Rosana-SP, and installed in the vicinity of the Engineer Sergio Motta Hydroelectric Power Plant. [8]

A. Description of the pilot plant Anemometric Station

The anemometric station is installed since the December 16, 2014 in the vicinity of hydroelectric power Engineer Sergio Motta, located near the cities of Rosana and Primavera, State of São Paulo in Brazil. Table 3 shows information about the location of tower installation.

The tower of the station had two configurations of height and place of installation, with a tower of 100 meters and one of 150 meters height. Moreover, the 100 meters tower had two compositions of height of the anemometers.

Table 3. Localization of the anemometric station. Source: Own elaboration, 2017.

	100 meter height tower	**150 meter height tower**
Model	EOL Zenith	EOL Zenith
Place	Primavera /SP	Primavera /Sp
Latitude	-22,484853	-22,490425
Longitude	-52,954542	-52,953286
Altitude	277meters	277meters

The 100 meters tower presented three anemometers and the 150 meters tower, 5 anemometers. Figure 1 brings the compositions and heights adopted by the two towers.

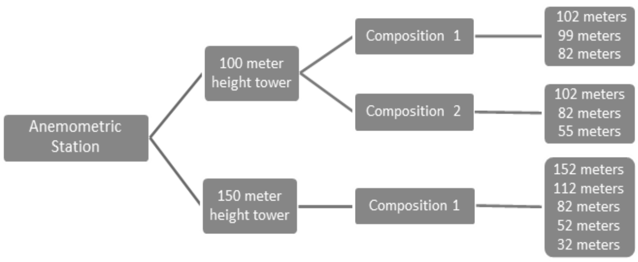

Figure 1. Configurations of the anemometric station and height of the anemometers. Source: Own elaboration, 2017.

The composition 1 of the 100 meters tower had 3 installed anemometers as ANEEL norm (Table 4.), however for a better exploration and possible modeling of the primary energy, as much bigger the vertical covering, better will be the comments of the behavior of the winds. In the face of this observation, the exchange for heights of one of the sensors installed was carried out, in which the 99 meters high sensor now is 52 meters high (Table 4).

From December 17, 2016 the tower moved place and happened to be 150 meters height, with 2 more anemometers, having a total of 5 anemometers.

Slope and Offset are parameters of adjustment of calibration of the raw data (Hertz, Volts) in which the data are parameterised in order to indicate the difference of a value of reference. Errors of Slope and Offset can compromise the validity of the data and also to make impracticable the use of these. It is recommended a verification in manuals of the sensors to check if the addressed values are correct.

The anemometric station has three data sets that had been divided according to the compositions of heights of the anemometric station, illustrated in Table 4.

Table 4. List of sensors of the anemometric station from 10/04/2015 to 16/12/2015 and from 17/12/2015 to 13/12/2016. Source: Own elaboration, 2017.

User Name	**Height**	**Type**	**Model**	**Slope**	**Offset**
Anemo1	102	Anemometer	Thies First Class	0,046040	0,235700
Anemo2	99	Anemometer	Thies First Class	0,046100	0,237200
Anemo3	82	Anemometer	Thies First Class	0,046010	0,245600
Wind Vane	99	Windvane	Thies 2k	1,000000	325,0000
Windvane	80	Windvane	Thies 2k	1,000000	325,0000
Temperature	92	Temperature	Galltec kpc 1/5	100,0000	-30,0000
Rel humidity	92	Rel humidity	Galltec kpc 1/5	100,0000	0,000000
Pressure	92	Pressure	AB60	60,00000	800,0000
Photodiode	15	Radiation	Photodiode	598,8024	0,000000
User Name	**Height**	**Type**	**Model**	**Slope**	**Offset**
Anemo1	102	Anemometer	Thies First Class	0,046060	0,232200
Anemo2	99	Anemometer	Thies First Class	0,046100	0,237200
Anemo3	82	Anemometer	Thies First Class	0,046010	0,245600
Wind vane	99	Wind vane	Thies 2k	1,000000	325,0000
Wind vane	80	Wind vane	Thies 2k	1,000000	325,0000
Temperature	92	Temperature	Galltec kpc 1/5	100,0000	-30,0000
Rel humidity	92	Rel humidity	Galltec kpc 1/5	100,0000	0,000000
Pressure	92	Pressure	AB60	60,00000	800,0000
Photodiode	15	Radiation	Photodiode	598,8024	0,000000

Table 5 illustrates the heights of installation of the equipment and the five anemometers in the 150 meters height tower

Table 5. List of sensors of the anemometric station from 17/12/2016 to 30/06/2017. Source: Own elaboration, 2017.

User Name	Height	Type	Model	Slope	Offset
Anemo1	152	Anemometer	Thies First Class	0,045910	0,239500
Anemo2	112	Anemometer	Thies First Class	0,045920	0,249400
Anemo3	82	Anemometer	Thies First Class	0,046060	0,232200
Anemo4	52	Anemometer	Thies First Class	0,046100	0,237200
Anemo5	32	Anemometer	Thies First Class	0,046010	0,245600
Pressure	92	Pressure	K611P	1,000000	600,0000
Wind vane1	99	Wind vane	Thies 2k	1,000000	0,000000
Wind vane2	80	Wind vane	Thies 2k	1,000000	0,000000
Temperature	92	Temperature	Galltec kpc 1/5	100,0000	-30,0000
Rel humidity	92	Rel humidity	Galltec kpc 1/5	100,0000	0,000000
Photodiode	15	Radiation	Photodiode	598,8024	0,000000

B. Description of the Solarimetric Station pilot Unit

The solarimetric station is installed since the December 18, 2016 in the vicinity of hydroelectric power Engineer Sergio Motta, located near the cities of Rosana and Primavera, State of São Paulo. Table 6 shows the information of the locality of installation of the solarimetric station.

Table 6. Localization of the solarimetric station. Source: Report of monthly data Analysis - Solarimetric Station, 2016.

Item				
Name	Solarimetric Station UHE Porto Primavera			
Municipality	Rosana			
State	São Paulo			
Geographical coordinates SIRGAS 2000	Latit.	-22,483°	Long.	-52,995°
Altitude	260			
Magnetic Declination (*)	-17,70 in 31/12/2014 (reference date)			
Prevailing wind direction (*)	75 east - northeast			
UTC hour				

The solarimetric measurement is composed by three bases of lift: 1) the 2,3 meters height; 2) one with 10 meters height and 3) a connecting rod of 1,15 meters height. Base 1 is support for the specific equipment of the tracker which consists of: pyranometers, Pyrheliometer, thermometer, and barometer. Higher base 2, is destined to the sensors of measurement of wind: anemometer and Anemoscope. And base 3 is a connecting rod for the Pluviometer lift.

Table 7 shows the list of the sensors that belong to the solarimetric station.

Table 7. List of sensors of the solarimetric station. Source: Report of monthly data Analysis - Solarimetric Station, 2016.

Manufacturer/ Model	Equip.	Height	Numb. serial
Thies First Class Advanced 4.3351.00.000	Anemometer	10	03151207
Thies Wind Vane Compact 4.3129.70.701	Windvane	8	01151485
Galltec Mela/KP S52100	Temperature	1,4	144933
Galltec Mela/KP S52100	Humidity	1,4	144933
AB60	Barometer	0,9	SQ07632
Young/52203	pluviometer	1,15	TB12493
Kipp&Zonen/CMP6	Pyranometer	2,3	140693
Kipp&Zonen/CMP6	Pyranometer	2,3	140693
EKO MS-56	Pyreliometer	2,1	P15008

C. Validation and data quality assessment of the stations data: solarimetric and anemometric

The validation and the treatment of the data are essential processes for the quality and reliability of any data group and are part of the quality analysis.

The climate limits used in one of the tests of validation in the installed locality had been established using the limits developed by the NREL, BSRN and WMO.

Is verified that the acceptable range checks of measurement according to WMO, in the document Guidelines on Quality Procedures Control will be Date from Automatic Weather Stations of 2004, is more permissible of what the tests of the NREL, Wind Resource Assessment Handbook - Fundamentals will be Conducting the Successful Monitoring Program, 1997, placing to the test the climatic conditions in the influence area where the sensors of measurement of the climate variable are installed.

In this sense, this topic presents the indication of acceptable range checks for the measured variable of a context of the localization of installation of the equipment of measurements, but this does not exclude the importance of the accomplishment of other routines of tests as recommended by the bibliographical references of the subject.

According to NREL (1997) the tests presented in Table 8 are validation tests most used for speed and direction of the wind, bringing in them the criteria of interval for the speed of the wind (horizontal and vertical) and of direction of the wind.

A reasonable interval for the average speed of the most waited wind more is of 0 the 25 m/s. However, the calibration of offset supplied in many calibrated anemometers will hinder values zero. Negative values clearly indicate a problem. Speeds above are possible of m/s and must be verified with other information. [9]

Table 8. Acceptable range check for speed and direction of the wind.

Parameter		Validation Criterion
Wind Speed Horizontal	Average	offset < Avg. < 25 m/s
	Standard Deviation	0 < Std. Dev. < 3m/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°
Wind Speed: Vertical line (optional)	Average	offset < Avg. <± (2/4) m/s
	Standard Deviation	offset < Std. Dev. <± (1/2) m/s
	Maximum Gust	offset < Max. <± (3/6) m/s

Source: NREL, 1997.

The tests of air humidity and rainfall are based on recommendations from the WMO (2014). However, the validation criteria for climate data are merged between WMO and NREL for a proper fit, shown in the Table 9

Table 9. Acceptable ranges to test other climatological data.

Parameter		Validation Criterion
Temperature	Seasonal Variability	5° C < Avg. 40° 1659 <
Barometric Pressure	Average	94 kPa < Avg. <106 kPa
Humidity *	Seasonal Variability	0 to 100%
Precipitation **	Amount	0 to 11 mm/min

Source: NREL, 1997.

* According to WMO, 2004.

** Maximum Limit of sensor limit Pluviometer Young/52203

For the acceptable ranges of variables of the solar radiation, focusing on the location in study, tests developed by BRSN were designated, which brings a set of tests for the observation of values measured more real.

Stage 1 to be followed for validation and treatment of the data is the verification of the physically possible data. In Table 10 are the limits considered possible for global solar radiation and its direct and diffuse components. The objective of this first step is the elimination of gross errors committed during the measurements or the handling of the data. [10]

Table 10. Method of validation - Physically possible.

Physically possible (Algorithm 1)		
Global horizontal (W/m^2)	Direct Normal (W/m^2)	Diffuse (W/m^2)
1 Shunting line standard ≠ 0 (zero)	1 Shunting line standard ≠ 0 (zero)	1 Shunting line standard ≠ 0 (zero)
2 Minimum: -4 Maximum: S x 1,5 x μ01,2 + 100 W/m^2	2 Verification of the state of operation of the solar tracker (where it is))	2 Verification of the state of operation of the solar tracker (where it is)
	3 Minimum: -4 Maximum: S	3 Minimum: -4 Maximum: Sa x 0.95 x μ01,2 + 50 W/m^2

Stage 2 is the verification of the existence of measured values that fit inside of limits that hardly happen. This step serves so that special attention is given to the considered data improbable. The limits established for this step are shown in Table 11. [10]

Table 11. Method of validation - Extremely rare

Extremely rare (Algorithm 2)		
Global horizontal (W/m^2)	**Direct Normal (W/m^2)**	**Diffuse (W/m^2)**
Minimum: -2 Maximum: Sa x 1,2 x μ0 1,2 + 50 W/m^2	Minimum: -2 Maximum: Sa x 0,95 x μ0 0,2 + 10 W/m^2	Minimum: -2 Maximum: Sa x 0.75 x μ01,2 + 30 W/m^2

Stage 3 deals with the establishment of intervals for objective comparison to detect misalignments of the sensors of measurement of direct and diffuse radiation.

Table 12. Method of validation - Comparison between sensors.

Extremely rare (Algorithm 3)		
Global horizontal (W/m²)	Direct Normal (W/m²)	Diffuse (W/m²)
(Global)/(Sum) it must be between ± 10% of 1,0 for AZS <75°, Sum >50 (Global)/(Sum) it must be between ± 10% of 1,0 for AZS <75°, Sum >50 Obs: this test is not possible for Sum < 50	Direct x µ0 - 50 W/m ² ≤ direct horizontal line (Global - Diffuse) ≤ Direct x µ0 + 50 W/m ²	(Diffuse)/(Global) < 1,05 for AZS < 75°, Global > 50 W/m² (Diffuse)/(Global) < 1,05 for AZS < 75°, Global > 50 W/m ² Obs: this test is not possible for Global < 50 W/m²

The used intervals of comparison are listed. On the assumption that the measurements of global radiation are more reliable, in the first case, values below the inferior limit mean misalignment of pyrheliometer and values above of the superior limit must to the misalignment of the shading ring. In the second case the inverse occurs, that is, values below the inferior limit mean misalignment of the shading ring and above the superior limit they mean misalignment of pyheliometer. [10]

V. ANALYSIS AND DESCRIPTION OF THE PROCEDURES OF VERIFICATION OF THE QUALITY OF THE DATA IN THE STUDY OF CASE

The analysis is based on the verification and the description of the procedures and used methods to evaluate the quality of the data of two units measurement pilot: an anemometric station and a solarimetric station.

A. Solarimetric data

In this topic it will be described how it was developed the process of analysis of quality of the collected data of the solarimetric station, called solarimetric data.

1) Raw data

The raw data are collected remotely via manufacturer's server of the datalogger. These are data measured in seconds and integrated in one minute, that is, it supplies registers minute by minute of all the variable measured per the station.

All the registers measured by the equipment possess: average, minimum, maximum, standard deviation and a count. For the rain variable sum and meter are used.

The raw data are given already measured in the reference unit (m/s, W/m ², mBar, among others), not havin raw data in Hertz and Volts.

The registers of raw data are separate in daily archives, with extension CSV (Comma-separated values). The extension CSV needs to be worked to be legible, coherent and friendly to the user. There are software of data analysis that accept the direct use of the data in CSV. Figure 2 illustrates as the raw data are visualized in format CSV, without any type of intervention.

After the raw is collected, they pass through the process of data validation, that generates the data set that signals possible imperfections and errors found in the measurements.

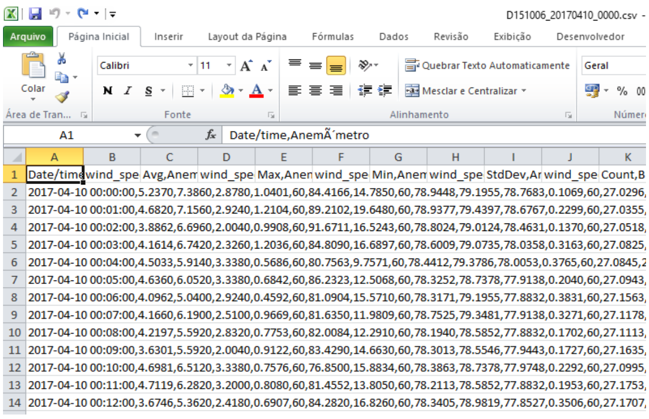

Figure 2. Raw data at the solarimetric station. Source: Own elaboration, 2017.

2) Signaled data

The data validation of the is made by means of the test application that evaluate the veracity of the measured registers. The values that are not accepted as valid in the tests are expunged. This data set possesses a monthly base of analysis.

The following criteria presented are adjusted for time series of 1 in 1 minute and had been elaborated in agreement with climatic variations in the measurement locality:

1) Temperature

a) Values out of the interval from 5 to 40°C;

b) Values that not to vary more than 0,03°C during one hour;

c) Values that vary more than 3°C between one measurement and the next one;

2) Wind speed

a) Values out of the interval of 0,21 the 25 m/s;

b) Values that do not vary more than 0.2m/s during one hour;

c) Values that vary more than 10m;s between a measurement and the next one;

3) Relative humidity

a) Values out of the interval from 0 to 100%;

b) Values that do not vary more than 0,05°C during one hour;

c) Values that vary more than 10% between a measurement and the next one;

4)Wind Direction

a) Values out of the interval from 0 to 360°;

b) Values that do not vary more than 3;5° during one hour;

5) Atmospheric Pressure

a) Values out of the range from 907.5 to 1092.5 mb (sensor limits);

b) Values that do not vary more than 0.001 mb during one hour;

c) Values that vary more than 0.5mb between one measurement and the next one;

6) Rainfall

a) Values out of the range from 0 to 11 mm/minute (sensor limits);

7) Irradiation/irradiance

a) Global Horizontal Irradiance 1 (GHI 1), Global Horizontal Irradiance 2 (GHI 2), Diffuse Horizontal Irradiance (DHI) and Direct Normal Irradiance (DNI)

i) Irradiation nocturnal values bigger and lesser;

b) GHI 1 and GHI 2

i) Values out of the interval of: -2a {1,2 x GTOA [cos (AZS)] ^1,2 +50} W/m ^{2}, for GHI;

Where:

AZS = solar zenithal angle;

GTOA = Irradiation in the horizontal plan in the top of the atmosphere;

ii) Values of irradiation registered with lesser solar rise than the 1st above the registered in the horizon;

iii) Values in which the absolute difference between GHI1 and GHI2 will be greater than the uncertainty standard of the pyranometers (5%).

c) DHI

i) Values out of the interval of:

-2 to {0,75 × GTOA × [cos(AZS)^1,2] + 30} W/m^2, for DHI;

ii) T2C - compatibility test between GHI and DHI;

iii) Kd/Kt Relation;

iv) test to stop the tracker: If the coefficient Kd is greater than 0.9 all day and Kt exceeds 0.5 at least once in the day, the whole day is excluded.

d) DNI

i) Values out of the interval of:

-2 a {0,95 × GTOA × [cos(AZS)^1,2] + 10} W/m^2, for DNI;

ii) T3C-compatibility test between GHI_measured and GHI_calculated;

iii) test to stop the tracker: If the direct irradiance is close to zero all day and the Kt coefficient exceeds the value of 0.24 at least once a day, the whole day is excluded (tracker stood all day).

Records not validated by the tests are expunged from the series of flagged data thus leaving gaps in the series. Figure 3 and Figure 4 illustrate the air temperature variable before and after the application of the validation tests, in which it is possible to note falls to zero which characterize data gaps. This type of lack in data directly affects the final result of analysis and it is recommended the use of a technical job of imputation of data in order to replace the purged data.

The validation includes specific tests for solarimetric data that assess directly the solar components.

After application of the validation tests, an analysis of the results is performed in order to visualize the data quality and data recorded not registered or rejected by quality control.

The tests and results are printed on a document that brings the necessary explanations and observations in quality tests.

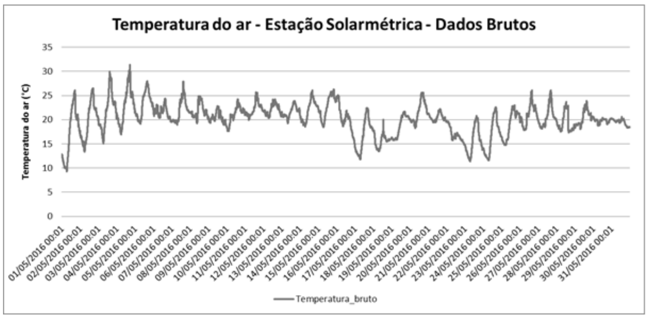

Figure 3. Air temperature records with raw data at the engineer Sergio Motta HPP. Source: Own elaboration, 2017.

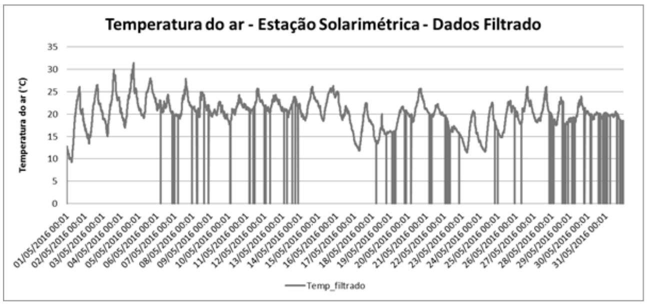

Figure 4. Air temperature records with raw data at the Engineer Sergio Motta HPP. Source: Own elaboration, 2017.

B. Anemometric data

In this topic it will be described how it was developed the process of quality analysis of the data collected from the anemometric station, called anemometric data.

1) Raw data

The anemometric data had been delivered to users via email and link via lodging in the cloud (Dropbox), however without remote access to the data. The raw data do not possess a definite frequency of time (days, month).

As part of the raw data three archives are generated by the equipment:

- An archive in txt, that possesses: Information of decoding, Information of Download, Configuration of the sensors and errors, Diagnosis of Information of Sensor.]
- An archive in log, with encrypted information.
- A file in wnd, an extension that loads the data generated by the sensors. This file can be opened by EOLCharting ® software, or through it generates a copy file with txt extension, which can be manipulated by various software and carry the same records.

The data are measured minute by minute and spaced out every 10 minutes.

The records of each equipment of the 100 meters high tower include:

- Anemometric Sensors-average, minimum, maximum and standard deviation. For two anemometers the TI30 (turbulence index).
- Climate sensors - average, minimum and standard deviation.

The records of the 150 meters high tower include:

- Anemometric Sensors-average, minimum, maximum and standard deviation. For two anemometers TI30 (turbulence index).
- Weather sensors - average

2) Calibrated data

The anemometric data had passed through a process of calibration of Slope and Offset, a process in which the raw data in Volts and Hertz are displayed in a mathematical formula that corrects possible errors, come after passing through the datalogger for conversion in the units of measurement.

The calibrated data had been divided in three groups:

- Group 1. Anemometers at 102, 99 and 82 meters height;
- Group 2. Anemometers at 102, 82 and 55 meters height;
- Group 3. Anemometers at 152, 112, 82, 52 and 32 meters height;

These data possess extension xlsx.

The criteria of quality analysis used in the anemometric data had been marked in adequate criteria according to the company executor of the data handling. The applied criteria of validation had been:

1) Average speed of the wind (10 min)

a) Values of speed between 0 m/s and 50 m/s;

2) Maximum speed of the wind (1 second)

a) Values of speed between 0 m/s and 70 m/s;

3) Atmospheric Pressure

a) Values that vary between 800 hPa and 1060 hPa

4) Temperature

a) Values of temperature that are between -15°C and 50°C

5) Humidity of air

a) Values of humidity that vary between 0% and 110%

The criterion of validation on the variable solar radiation is not defined in the reports.

In Group 1 (Dec/16/2014 to Dec/16/2015) of the calibrated data there were gaps in the registers due to lack of data, as illustrated in Figure 5. This lack reached all the variables measured by the station.

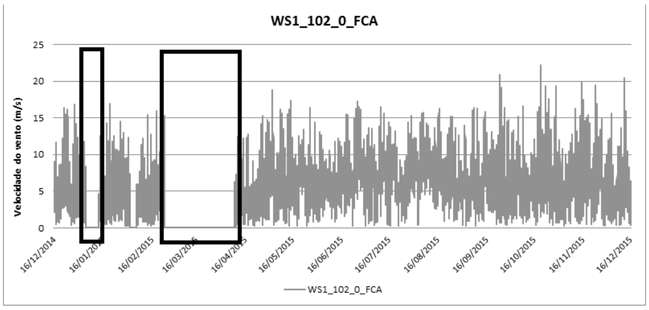

Figure 5. Gaps of data in Group 1 of the anemometric data. Source: Own elaboration, 2017.

As already mentioned in the previous topic, the gaps found in the data series should apply to data imputation technique to complete the series.

The Report of Data Quality Analysis brings the developed one with the previously cited tests, and specific tests for the wind power resource, with Correlation between anemometers, Distribution of Weibull, Directional Analysis, Intensity of turbulence in 10 minutes and 30 seconds and Shearing of the wind.

The tests and results are printed on a document that brings the necessary explanations and observations carried out in data quality tests.

VI. CONCLUSION

The development and the good findings of an investigation are coupled to the conduction of the quality in the di-verse areas and stages inside of a project. This involves since the planning until the conclusion phase.

And a factor of great impact is the data obtained during all process of inquiry. For good results, obtaining data is not sufficient, it is necessary that there is quality in these data and that they are reliable data. What, consequently, will bring greater credibility and it will add a fine knowledge in the developed area.

Figure 6 shows the three types of basic data sets for the development of any analysis that involves a series of data, which goes from rude data, (collected and without no type of treatment), to the data observed (the abnormalities are signaled) until the development of data assessment, that expunges, fills and corrects errors presented by the observed data.

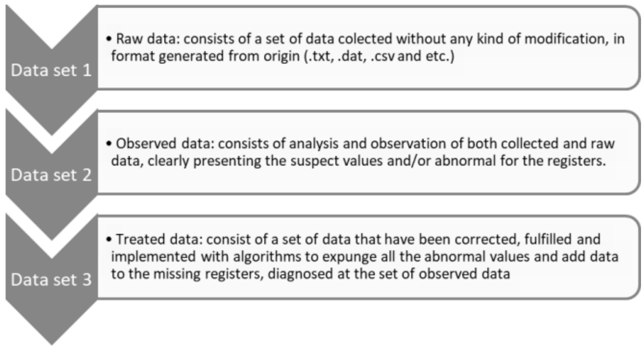

Figure 6. Data set for research. Source: Own elaboration, 2017.

These three data sets assist the user in the diverse necessary stages for the development of an analysis and express the reality of the one measured. This form of data organization assists the user in a clear way to define which set adjusts to his needs.

The results show that the recommendations and techniques of analysis of quality of the data had been followed partially in the two types of data: anemometric and solarimetric.

The tests of solarimetric data validation had been finer, following the adjusted tests less permissible and more adequate to the climatic variations of the measurement locality. Beyond bringing specific tests for the solar energy resource in its three components: a) Global Horizontal Irradiance 1 (GHI 1), Global Horizontal Irradiance 2 (GHI 1), Diffuse Horizontal Irradiance (DHI) and Direct Normal Irradiance (DNI).

The recommendations of test of consistency of limits, historical serial consistency, consistency between different sensors and comparison with values estimated by models had not been elaborated with the solarimetric data.

As regards to the three data sets defined as of extreme importance in any data series (raw data, filtered data, treated data), the solarimetric data obtained two of them: raw and filtered. The data assessment have not been concluded since the gaps left by the tests that remove the suspicious data had been filled with the imputation of data, thus leaving the empty gaps in the series of solarimetric data.

The validation tests of the anemometric data were simpleton and less adequate to the climate variations found in the measurement locality. The analyses with the specific tests on the wind power resource had been satisfactory, in which there was comparison between the anemometers and the turbulence registered in two of the anemometers.

The recommendations of test of consistency of limits, historical serial consistency, consistency between different sensors and comparison with values estimated by models that had not been elaborated with the solarimetric data.

The anemometric data had only gotten the raw data of the three definite sets as extremely useful in any series of data. The calibrated data that correct imperfections from the raw data had been added, with use of Slope and Offset. The gaps left in the data had not been substituted, what generates empty values in the series of data.

Thus, we understand that the methods of analysis for the qualities of the data addressed, both solarimetric and anemometric, in this work, in general, are simple and do not explore all the possible ways of uncertainties that involve a measured one. The tests of validation, consistency and treatment can be broadly analyzed once several research institutes bring the due recommendations for such use.

When we speak of quality data it is necessary to observe that the standards need to be as finest as possible, analyzing the locality of installation of the tools, what generates a series of more coherent data and next to the real.

VII. Acknowledgements

To CESP (Companhia de Energia de São Paulo / São Paulo Energy Company), for funding the R&D ANEEL PE-0061-0037/2012 project which, for instance, enabled the use some data and reports without any compromise for this work.

REFERENCES

[1] MARTINS, Fernando Ramos, PEAR TREE, Enio Bueno. Comparative degree of the trustworthiness of estimates of solar irradiation for the Brazilian Southeast gotten from data of satellite and for interpolation/extrapolation of surface data comparative data Brazilian magazine of Geophysicist (2011) 29 (2): 265-276. São José dos Campos, Brazil, 2011.

[2] DURRE, I.; MENNE, M.J.; VOSE, R.S. Strategies for evaluating quality control procedures. In: SYMPOSIUM ON METEOROLOGICAL OBSERVATION AND INSTRUMENTATION, 14, 2007, San Antonio, 2007.

[3] FERRARI, Gláucia Tatiana. Imputation of rainfall data and its application in the modeling of extreme events of agricultural drought / Gláucia Tatiana Ferrari. Piracicaba, 2011.

[4] MAGINA; Flávio de Carvalho. Automatic Acquisition and Treatment of Meteorological Data Applicable to Design and Operation Airlines Electricity Transmission. (Master's thesis). FEDERAL UNIVERSITY OF ITAJUBÁ. Itajubá; Minas Gerais; 2007.

[5] SALUSTIANO, Rogério Esteves. Application of sensor fusion techniques in environmental monitoring / Rogério Esteves Salustiano. Campinas, 2006.

[6] BRASIL. Ministry of Agriculture and Agrarian Reform. National Secretariat of Irrigation. National Department of Meteorology. Normal Climatological (1961-1990). Brasília: 1992.

[7] FRONDIZI , Carlos Alberto. Air quality monitoring: theory and practice. Rio de Janeiro, E-papers, 2018.

[8] CEPEL, CRESESB. Manual of Engineering for Photovoltaic Systems. organizers: João Tavares Pinho e Marco Antônio Galdino. Rio De Janeiro, 2014.

[9] NATIONAL RENEWABLE ENERGY LABORATORY - NREL. WIND RESOURCE ASSESSMENT HANDBOOK - Fundamentals for Conducting a Successful Monitoring Program. Albany, 1997.

[10] ABREU, S. L., COLLE, S., ALMEIDA,A. P., MANTELLI NETO, S. L. Qualification and recovery of solar radiation data measured in Florianópolis -SC. In: Brazilian Congress of Thermal Engineering and Sciences, 8 th, 2000, Porto Alegre. Anais. Porto Alegre: ENCIT, 2000.

[11] WISSMANN, J. A. Computational Tool for Data Consistency Analysis Pluviometric. Varia Scientia, 2006.

[12] GEPEA. Executive Technical Document. DTE-Phase2: Mass data and its magnitudes generated by the primary and secondary measurement units. ANEEL Research and Development Project PD-0061-0043 / 2014 - Research and Development Program of Energy Company of São Paulo – CESP. São Paulo, 2016 Energy Company of São Paulo.

[13] VIANA, Inês Susana Ribeiro. Synthetic Data Generation Method for the Creation of Publicly Used Microdata. Master's thesis. University of Porto. Portugal, 2014.

[14] PAULESCU; M. et al.; Weather Modeling and Forecasting of PV Systems Operation: green energy and technology. London: Springer-Verlag; 2013.

[15] FENG, Song; QIAN, Weihong. Quality Control Of Daily Meteorological Data In China,1951–2000: A New Dataset. INTERNATIONAL JOURNAL OF CLIMATOLOGY. Int. J. Climatol. 24: 853–870, 2004.

[16] REEK, Thomas; DOTY, Stephen R.; OWEN, Timothy W. A deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network. Bulletin of the American Meteorological Society 73: 753–762. National Oceanic and Atmospheric Administration, National Climatic Data Center. Asheville, North Carolina 2012.

[17] STOOKSBURY DE, Idso CD, Hubbard KG. The effects of data gaps on the calculated monthly mean maximum and minimum temperatures in the continental United States: a spatial and temporal study. Journal of Climate 12: 1524–1533. 1999.

[18] WMO. World Meteorological Organization. Baseline Surface Radiation Network (BSRN): operations manual; version 2.1. Geneva: WCRP-121; WMO/TD-No. 1274; 2005.

[19] WMO. World Meteorological Organization. Guide to Meteorological Instruments and Methods of Observation. Geneva: Chairperson; Publications Board; 2004.

[20] WMO. World Meteorological Organization. Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. Geneva, Switzerland, 2004.

[21] GARCÍA-PEÑA, Marisol; ARCINIEGAS-ALARCÓN, Sergio; BARBIN, Décio. Climate Data Imputation Using Decomposition by Singular Values: An Empirical Comparison. Brazilian Journal of Meteorology, v.29, n.4, 527 - 536, 2014.

[22] GEPEA. "Technical and Scientific Report (RTC): Description of Measuring and Generation Units Envisaged for Installation at the Engenheiro Sergio Motta Hydroelectric Power Plant". ANEEL PD-0061-0043 / 2014: Integration of Wind and Solar Photovoltaic Plants to Existing Hydroelectric Power Plants: An approach through Energy Complementation with the practical support of pilot plants with different technologies, Energy Group of the Department of Energy Engineering and Electrical Automation, University of São Paulo, São Paulo, 2016.

[23] FOLEY, A.M., LEAHY, P.G., MARVUGLIA, A. and MCKEOGH, E.J. Current Methods and Advances in Forecasting of Wind Power Generation. Renewable Energy, 37, 1-8. http://dx.doi.org/10.1016/j.renene.2011.05.033. 2012.

[24] LONG C.N. and DUTTON E.G. “BSRN Global Network recommended QC tests V2.0”, Baseline Surface Radiation Network (BSRN), 2002. Disponivel em: http://bsrn.awi.de/fileadmin/user_upload/redakteur/Publications/BSRN_recommended_QC_tests_V2.pdf

[25] GEPEA. "Technical and Scientific Report (RTC): Determination and Characterization of the Model of Primary Wind Energy". ANEEL PD-0061-0043 / 2014: Integration of Wind and Solar Photovoltaic Plants to Existing Hydroelectric Power Plants: An approach through Energy Complementation with the practical support of pilot plants with different technologies, Energy Group of the Department of Energy Engineering and Electrical Automation, University of São Paulo, São Paulo, 2016

[26] GEPEA. "Technical and Scientific Report (RTC): Calculation and Presentation of the Elements of Modeling of Primary Wind Energy". ANEEL PD-0061-0043 / 2014: Integration of Wind and Solar Photovoltaic Plants to Existing Hydroelectric Power Plants: An approach through Energy Complementation with the practical support of pilot plants with different technologies, Energy Group of the Department of Energy Engineering and Electrical Automation, University of São Paulo, São Paulo, 2016.