Improving Forecasting Through Accurate Data

By

Jim Blatchford
Sr. Policy Issues Representative

Rizwaan Sahib
Telemetry Engineer

CAISO

Abstract

Poor meteorological and production data quality can raise significant barriers to the development of accurate energy forecasts from wind power facilities for use by the California Independent System Operator Corporation (CAISO) in maintaining grid reliability and operating efficient markets. This paper describes (1) the technical challenges faced by the CAISO and the wind power industry as a result of poor meteorological and production data quality, (2) the causes or sources of the poor meteorological and production data quality at wind sites, and (3) how such poor data quality affects energy forecast accuracy and how such inaccuracies can and do affect market operations and grid reliability. Finally, this paper recommends implementation of data and process modifications to ensure quality information is reported and available to enhance the accuracy of wind production forecasts. Those recommendations include improved Scheduling Coordinator (SC) outage reporting, data redundancy, independent power sources for telemetry data, improved communication between SC's and field personnel, and development of a more aggressive compliance program by the CAISO.

Background

The PIRP Program

The CAISO and stakeholders developed the Participating Intermittent Resource Program (PIRP) to facilitate the participation of intermittent resources in CAISO markets. PIRP accomplishes this objective by mitigating, through special settlement provisions, the economic risks associated with deviations between scheduled and actual metered output that are inevitable given the inherent variability or unpredictability of intermittent resources, such as wind generators. The success of PIRP, therefore, is highly dependent on the accuracy of the energy forecasts utilized by Participating Intermittent Resources (PIRs) to schedule energy. The need for accurate forecasts from intermittent resources will become increasingly critical as load serving entities comply with California's aggressive renewable portfolio standards. With greater intermittent resource penetration, inaccuracies in forecasts and the resulting energy schedules can lead to inaccurate unit commitment and excessive dispatches of energy in real time that negatively impact reliability and system efficiency.

Currently, PIRs provide real time generation, meteorological, and other data to the CAISO. This data is forwarded by the CAISO to a forecast service provider (FSP), which develops high quality relationships between concurrent meteorological conditions and power production to develop an energy forecast for each PIR. The generation and meteorological data provided by PIRs consists of real time MW production, air temperature, wind direction, wind speed, barometric pressure and turbine outage information. The data is transmitted to the CAISO via a virtual private network and retrieved by the FSP every 10 minutes. The FSP then provides the forecast to the SC for the PIR every 2 hours and 45 minutes before the operating hour. (Note: this will change to 105 minutes in MRTU) This is defined as the Hour Ahead ("HA") forecast, which is used by the SC as the wind facility's HA generation schedule.

Development of PIRP II

PIRP originally used an informal manual email system to report errant data. Under this system, the FSP would perform the data quality validation and communicate any errant data issues to the CAISO, who would then forward the information to the SC for the PIR. This process helped improve forecasts to a small extent. However, serious delays in reporting errant data to the PIRs were observed because data checking occurred too far downstream in the process. Delays of several days or longer were not uncommon for identifying and correcting communication problems, missing data, and errant data from devices at various wind farms.

In 2006, the CAISO upgraded the PIRP software application system to reduce delays in reporting errant data to wind parks SCs. The primary objective was to provide SCs with timely information regarding data availability without the need for manual intervention. Information provided to the SCs give immediate errant data feedback when the SC retrieves their HA Forecast, thus eliminating the need for a manual feedback system.

Root Cause Analysis

Root Cause Study

After the implementation of PIRP II, the CAISO began studying ways to improve data quality throughout the PIRP program by re-examining current processes and requirements. The goal of the study was to develop a list of recommendations for both the CAISO and SCs to improve data quality.

The importance of studying the root cause of data quality failures was essential in identifying gaps in current procedures and policies. Through a better understanding of the root causes of errant data, the CAISO can provide recommendations and work with SCs to reduce errant data and improve forecast accuracy.

Data Availability and Forecast

A report provided by AWS True Wind found that the Annual Mean Absolute Error (MAE) for the Next Operating Forecast Hour for 2007? range from a low of 6.8% to a high of 15.3% with an overall average of 13.3%. Greater data availability correlated to lower MAE.

Table 1.0 represents three similar wind parks and demonstrates the relationship between data availability and forecast error. Facility A.1 with the highest data availability, 98.37%, has the more accurate forecast of 11.3% In contrast; facility A.2 with data availability of 87.18% had a forecast error of 14.59%.

Facility	Data Availability (Good Data Quality)	Next Operating Hour Forecast Annual MAE	Next operating Hour Forecast Annual Net Deviation
A.1	98.37%	11.30%	-0.18%
A.2	87.18%	14.59%	2.18%
A.3	86.92%	12.43%	1.07%

Table 1.0: Relationship between data availability and forecast deviation at selected sites

Data Analysis

The study initially reviewed the current data validation process from the FSP and the PIRP application. The CAISO also began developing a system that normalized the validation process. A normalized system makes certain that both the FSP and the CAISO, via the PIRP application, are all using a similar data validation process. The CAISO recommended simulating the validation process currently used by the FSP due to their higher level of accuracy in identifying errant data.

The CAISO then used the new validation system to identify errant data for 6 weeks. In each instance of errant data, the CAISO performed a root cause analysis, categorized each failure, and tracked all communication between the CAISO and SCs related to errant data to examine how issues were transmitted and the time needed to correct the problems.

Data Validation

A comprehensive data validation process is important to accurately identify errant data. As stated above, working with the FSP, AWS Truewind, the CAISO found that normalizing the data validation procedure used by the FSP would be the most efficient use of resources.

The normalized data validation rules, jointly developed by the CAISO and the FSP, are listed below in Table 1.1. The table describes the “as is” processes for PIRP application, AWS Truewind data validation, and the normalized validation. The new validation process uses the data stored in the CAISO PI system as its source. The CAISO used these procedures to perform a detailed study on identifying the root cause of all errant data.

The normalized data validation process used in the CAISO’s data quality study to identify the root cause of errant data is very accurate and greatly[1] reduces falsely identified errant data by utilizing forecasting data. There remain a limited number[2] of occurrences of falsely identified errant data. The new system is not currently implemented, and will be discussed in the recommendation section of this paper.

The CAISO also identified that data, such as barometric pressure and air temperature, depending on site location, may not show significant changes in a one hour period. Consequently, the PIRP II application began identifying both barometric pressure and air temperature data as bad quality or flat lining. To resolve this problem, the CAISO removed both barometric pressure and air temperature from the validation process. This greatly improved the accuracy of the data validation.

[1] Previous Estimates were greater than 20% false validation errors.
[2] Only 1.24% of the sample period from Table 1.2

DATA TYPE	PIRP II DATA VALIDATION	TRUEWIND DATA VALIDATION	NORMALIZED VALIDATION
Quality Point	All data points that are not "Normal" are identified as bad data quality (TE or telemetry error means data not received form site)	All data points that are not "Normal" or in "Alarm" are identified as bad data quality (TE or telemetry error means data not received form site)	All data points that are not "Normal" or in "Alarm" are identified as bad data quality (TE or telemetry error means data not received form site)
Meteorological Data (Flatline)	All individual data points that do not change for more then 1 hour are considered to be bad data quality. Barometric Pressure and air temperature are exempt	If all data points simultaneously do not change their current value for more than 1 hour, then the data set is considered bad data quality	If all data points simultaneously do not change their current value for more than 1 hour, then the data set is considered bad data quality
Generation Data (Flatline)	If wind speed in greater then 5 m/s and the generation values does not change for more then 1 hour then consider data to be bad data quality.	If wind speed in greater then 5 m/s and the generation values does not change for more then 1 hour then consider data to be bad data quality Uses Meteorological and outage information to determine flatlining on Low and High generation ranges.	If the wind speed is greater than or equal to 5m/s and generation is greater than or equal to 1 MW, and data does not change for 1 hour then data is bad quality If the wind speed is greater than or equal to 8m/s and generation is less than or equal to 1 MW, and data does not change for 1 hour then data is bad quality

Table 1.1 AWST and CAISO normalized

Errant Data Identification

The CAISO identified errant data by using the normalized data validation procedures as developed in Microsoft Excel to poll data from the PI system [3] every 10 minutes and perform data validation on a weekly basis. All the errant data identified by the validation process in Excel was manually confirmed.

After confirming all errant data, the CAISO began its root cause analysis. By using the **S**cheduling **L**ogging for the **I**SO of **C**alifornia system or SLIC[4], the CAISO was able to determine if the root cause of errant data resulted from a generation or transmission outage at a particular site. If the cause of the errant data remained unclear, the site was

3 The Plant Information System or PI system archives data every 4 seconds from the DPG at each site.

4 The **S**cheduling **L**ogging for the **I**SO of **C**alifornia (SLIC) accepts transmission and generation outage requests electronically submitted via the SLIC Web Client.

directly contacted to determine the cause. After determining the root cause of the errant data, the CAISO categorized the errant data into five categories: (1) site outages, (2) communication, (3) equipment malfunction, (4) unknown, and (5) false indication.

Site outages are due to generator maintenance or transmission line work that causes power failures on meteorological equipment, or communication equipment resulting in errant data.

Communication outages occur when there are issues with the Virtual Private Network circuit at a site or networking or communication issues with the CAISO. Communication outages are difficult to predict and are not currently being reported through SLIC. Equipment failures include malfunctions with the meteorological, communication links, or telemetry equipment like the Data Processing Gateway ("DPG").

Unknown failures are errant data whose root cause can not be determined.

False indications are errant data which was identified during the validation process, but during the manual check processes, were found to be valid.

Findings

The CAISO was able to identify trends of errant data and develop recommendations intended to improve the quality of data gathered from the sites. The results of the root cause study are laid out in Table 1.2. The table identifies the total length of time in hours each type of failure occurred in a 1 week period and the last row aggregates the errant data failure of all 13 studied wind sites.

The results demonstrate that the majority, almost 100% of the failures occurred due to telemetry equipment failures and unreported outages. In particular, based on further investigation with site engineers, the CAISO concluded that the **primarily reason** for errant data during telemetry outages is the absence of an independent or backup power supply for the telemetry equipment including the DPG. A majority of wind parks power their telemetry and meteorological equipment by back feeding from the transmission lines, or directly from a feeder line connected to a group of wind generators. Accordingly, when a site is forced into an outage resulting from utility work on the transmission lines, the DPG and other equipment lose power. Similarly, if the telemetry equipment is powered by a the wind turbine feeder line, when the line is forced off, due to maintenance or other equipment problems, then all the associated telemetry equipment using the feeder line also fails and the site is not able to provide real time telemetry to the FSP.

	OUTAGE (HR)		COMM FAILURE (HR)	EQUIPMENT FAILURE (HR)		UNKNOWN (HR)		FALSE		TOTAL FAILURE (HR)
	FLATLINE	TE	TE	FLATLINE	TE	FLATLINE	TE	FLATLINE	TE	ALL
WEEK 1										
Total Failure	119.5	2.67	0	42.67	0	18.67	0	0	0	183.5
% Failure	65.12%	1.46%	0.00%	23.25%	0.00%	10.17%	0.00%	0.00%	0.00%	100.00%
WEEK 2	65.12%	1.46%	0.00%	23.25%	0.00%	10.17%	0.00%	0.00%	0.00%	
Total Failure	711.83	138.83	0	261.5	186	5.33	4.5	0	7.67	1315.67
% Failure	54.10%	10.55%	0.00%	19.88%	14.14%	0.41%	0.34%	0.00%	0.58%	100.00%
WEEK 3	54.10%	10.55%	0.00%	19.88%	14.14%	0.41%	0.34%	0.00%	0.58%	
Total Failure	512	95	0	221	112.6	6.81	4.5	0	7.33	959.24
% Failure	53.38%	9.90%	0.00%	23.04%	11.74%	0.71%	0.47%	0.00%	0.76%	100.00%
WEEK 4	53.38%	9.90%	0.00%	23.04%	11.74%	0.71%	0.47%	0.00%	0.76%	
Total Failure	538.83	180.33	0	0	50.67	0	0	0	11.5	781.33
% Failure	68.96%	23.08%	0.00%	0.00%	6.48%	0.00%	0.00%	0.00%	1.47%	100.00%
WEEK 5	68.96%	23.08%	0.00%	0.00%	6.49%	0.00%	0.00%	0.00%	1.47%	
Total Failure	39.5	111.83	0	2	1.17	0	0	0	23.5	178
% Failure	22.19%	62.83%	0.00%	1.12%	0.66%	0.00%	0.00%	0.00%	13.20%	100.00%
WEEK 6	22.19%	62.83%	0.00%	1.12%	0.66%	0.00%	0.00%	0.00%	13.20%	100.00%
Total Failure	184.33	79.33	0	4	130	0	0	0	5	402.67
% Failure	45.78%	19.70%	0.00%	0.99%	32.28%	0.00%	0.00%	0.00%	1.24%	100.00%
Total Hours	2106.0	608.0	0.0	531.2	480.4	30.8	9.0	0.0	55.0	3820.4
Total %	55.12%	15.91%	0.00%	13.90%	12.58%	0.81%	0.24%	0.00%	1.44%	100.00%

Table 1.2 Errant Data Study Results

The simple malfunction of equipment, such as weather monitoring equipment, revenue meters, control system equipment, or telemetry equipment, for reasons other than power outages constitutes a secondary source of errant data. Some of the most common failures occur because control systems like programmable logic controllers (PLC), remote terminal units (RTU), or the DPG stops updating data.

Other less common issues include weather equipment failure and DPG failure due to wear and tear of the equipment. These failures can lead to flat lining of data or telemetry failures.

Recommendations

Site Communication

Improving communication between the CAISO, FSP and SCs is an important component of reducing errant data and its impact. Currently the CAISO monitors DPGs for complete communication failure for all generators with a DPG. Only in these circumstances would the CAISO directly contact the site operator to resolve problems. To improve forecast accuracy, the CAISO, along with the FSP, took an active role in identifying errant data for each site. When data issues were identified, the CAISO took steps to contact the SCs and make them aware of the telemetry issues. This notification was a manual process and was possible because of the limited amount of SCs in PIRP. As PIRP grows, monitoring of the telemetry must become an explicit responsibility of the SC and the SC must provide the CAISO with outage information[5]. In part, the improvements implemented by the CAISO through PIRP II have provided SCs with additional information to appropriately monitor the PIRs telemetry.

The CAISO, however, has observed a lack of communication between some of the SCs and the underlying sites during times of equipment failure. SCs must have direct contact with site technicians when the SC identifies or is made aware of errant data. If the SC is not in contact with the site technicians, there will necessarily be delays in correcting errant data issues.

The CAISO recommends that site engineers be given access to the DPG data being provided to the CAISO. This will supplement communication from the SC through its monitoring the PIRP forecast webpage. Further if a site is having data problems, direct notification to the CAISO Engineering along with SLIC reporting will aide in rapidly resolving data issues.

Availability Reporting

As a result of performing the root cause study, the CAISO also determined that SCs frequently fail to accurately report unit availability. As maintenance schedules change and turbines are forced offline, the maximum output for the site can change from hour to hour. Consequently, the accuracy of the forecast is affected. The FSP requires accurate reporting of generation MW capacity in order to develop accurate forecasts. If generation capacity is inaccurately stated, the associated forecast will reflect the inaccurate capacity. For instance there have been cases where the SC MW forecast has exceeded the capacity of the park due to unreported outages.

[5] EIRP 2.2.3 states, "A Participating Intermittent Resource must install and maintain the communication equipment required pursuant to EIRP 3, and the equipment supporting forecast data required pursuant to EIRP 6".

It is, therefore essential for SCs to accurately report site capacity via SLIC. In the case of line outages, generator derates, and maintenance at a park, the park's SC should ensure that all parties, including the CAISO, are accurately notified of the outage. The CAISO recommend all outages down to 1 MW must be reported.

Independent Power Sources

As noted, at many sites, power for telemetry equipment is obtained either from back feeding through transmission lines or directly from the wind generators. Unfortunately, when the wind turbines or a transmission lines go down for maintenance (or fail to produce power), all telemetry equipment fails. A simple solution is to have all telemetry equipment (i.e. meteorological equipment and the DPG) install an independent backup power source. The backup power source can be small gas generator or photovoltaic systems with battery storage for night time. Similar systems have been implemented for revenue meters and other telemetry equipment in remote areas.

Accordingly, the CAISO recommends that a site must install an independent power source. This will greatly improve data availability and reliability without significant infrastructure cost to the participants.

Data redundancy

The second most common cause of errant data is equipment failure as noted in Table 1.2. Another important step in reducing the effect of equipment failure on data used in forecasting is by adding data reporting redundancy. The CAISO, FSPs and the wind industry have observed that sites with more meteorological equipment spread throughout the site have improved forecast accuracy.

Using multi-telemeter sites as a benchmark, CAISO recommends the PIR must provide anemometer data from no less than one (1) turbine for every 5 turbines within the footprint of the park. The turbine providing the nacelle anemometer data will be located to represent the topology of the adjacent group of five (5) turbines. For example, if a wind farm has 5 turbines on a ridgeline a turbine that is least affected buy the wake effect of the other turbines would be designated to provide the nacelle anemometer information. Exception to this standard could be made if there were less than 5 units in the same area. If there were more than five (5) adjacent turbines, but less than ten (10), then two (2) turbines would be designated to provide nacelle anemometer information.

Compliance

The final recommendation advanced by the CAISO is to develop a more aggressive compliance program. In the past, errant data issues have been allowed to persist for weeks at time. Other than the issuance of a letter of non-compliance by the CAISO, it has been difficult to enforce data availability requirements. The CAISO recognizes that it must implement a more well defined and effective compliance program to enforce data availability requirements.

Conclusion

At the request of the stakeholders, the CAISO performed a study to determine the root cause of errant data which affects the quality of the forecasts from the FSP. The CAISO found that there are four basic causes for errant data. Those causes are:

- Outages
- Communications Failure
- Equipment Failure
- False Validation

The CAISO is making the following recommendations to address these root causes of errant data.

Site Communication – Site technicians should be able to direct access to DPG data and the SCs must have means to immediately recognize and communicate to the site technician the errant data condition as reported on their PIRP Forecast page or DPG feedback.

Availability Reporting –The SC is responsible and must report all data anomalies and outages from the site via SLIC. These anomalies include MW availability and all telemetry problems with the CAISO.

Independent Power Supply - electrical interruption of telemetry equipment causes errant data which must be eliminated and therefore an independent power supply should be mandatory.

Data Redundancy – Receiving anemometer data from multiple sites within the wind park will add two important components to the meteorological data streams. Those components are redundancy of data from the site along with a more representative collection of data from the site to develop an energy forecast.

Compliance – The CAISO has also developed a new process to enforce the tariff requirements for telemetry. If PIRs do not provide quality telemetry for more the 5 days, they are sent a non-compliance letter and given 30 days to comply or they will be disqualified from PIRP. The compliance processes has greatly improved the over all turnaround time it takes PIRs to correct errant data issue.

Biographies

Jim Blatchford is a Senior Policy Issue Representative in the External Affairs Division with the California Independent System Operator – CAISO. He is responsible for integrating the Participating Intermittent Resource Program (PIRP) into the CAISO Renewable Road Map Project. Jim has been with CAISO for eight years and has worked in the Renewables area for six of those years. He has held Project Management positions within the aerospace, computer industries and at ERCOT. Jim holds a Bachelors of Science degree in Computer Information Systems and a Masters of Science degree in Electronic Commerce.

Jim Blatchford
Sr. Policy Issues Rep.
Stakeholder & Industry Affairs
W 916-351-2411
jblatchford@caiso.com

Rizwaan Sahib is a Telemetry Engineer, and assists with data analysis with the CAISO. He is responsible for working with stakeholder to manage the implementation of telemetry for new generators, and performing root cause analysis for various types of telemetry failures. He graduated California Polytechnic on 2006 with a Bachelor Degree in Electrical engineering, emphasizing in power systems. In 2008 he graduated the University of Phoenix with a Masters in business technology management. Currently, he is working at the CAISO as a telemetry engineer and data analyst.

Rizwaan Sahib
CAISO RIG Engineer
916-351-4454 - office
916-934-9319 - mobile
rsahib@caiso.com