Maximization of the Value of Intra-day Wind and Solar Forecasts for an Island Grid System via Customized Forecasts and Evaluation Metrics

John W Zack
AWS Truepower, a UL company
Albany, NY 12205
jzack@awstruepower.com

Robert Kaneshiro and Lisa Dangelmaier
Hawaii Electric Light Company
Hilo, Hawaii
robert.kaneshiro@hawaiielectriclight.com

***Abstract*—This paper provides an overview of an in-progress study to identify which aspects of intra-day forecast performance are most critical to providing value for the management of the impact of renewable generation variability on an island system with no interconnections and a high penetration of variable renewable generation. The objective is to identify which forecast information provides the value to operational decision-making and to design a customized forecast evaluation metric that more effectively measures the sensitivity of the operational decision-making environment to forecast error than traditional error metrics.**

The platform for the study is the island grid system operated by the Hawaii Electric Light Company (HELCO) on the "Big Island" in the State of Hawaii. Forecasts from a customized wind and solar forecast system have been used in the operational decision making process for several years but the quantification of the actual value of forecast information has been difficult. A customized forecast evaluation system is being built from (1) the identification of the critical time periods and scenarios as well as the key parameters that impact operational decisions at those times, and (2) the formulation a forecast evaluation metric that emphasizes the performance for the prediction of key parameters during the critical time periods and scenarios.

The initial phase of this project has identified three key daily time periods with characteristic operating issues. A categorical forecast structure has been developed to focus on the key information for each of the three key decision-making periods. A generalized skill score has been defined to evaluate the categorical forecasts in a way that emphasizes performance in infrequent but key scenarios.

Keywords-energy forecast value; grid management with high renewable pentration;wind and solar integration

I. INTRODUCTION

The increasing penetration level of non-dispatchable variable renewable generation resources such as wind and solar generators on grid systems have created the need for tools and approaches to assist grid operators in the management of the variability in order to maintain supply-demand balance and grid reliability in an economical manner. The need increases as higher amounts of variable resources increases offline cycling of conventional units, poses more complex unit commitment decisions, and raises uncertainty in the management of energy resources. In addition to online reserves (which have cost implications), there are a number of flexible energy resources that are helpful to grid operators managing variability including: (1) energy storage, (2) demand response, (3) active power control of variable generation resources, and (4) flexible, quick-start generation resources. Short-term forecasting of renewable generation variability provides useful insights into the best use of these energy resources to meet reliability and cost goals through more optimized use of the available energy resources. Some of these resources are not available on specific grid systems, due to the limitations of the current mix of system resources and the high cost or lengthy time to make changes to system assets. Short-term forecasting typically has a low implementation barrier and a very favorable cost/benefit ratio. However, the simple availability of forecast data to the grid operator does not guarantee that the potential value of that information will be realized in the grid management process. A key to realizing value from forecast information is the extraction of components of forecast information that address specific operational issues and inform associated decision-making. It is critical to evaluate and quantify how well the forecast information addresses key operational issues, to instill confidence in forecast users.

Island grid systems with a high penetration of wind and solar generation often have the most acute need for tools to manage variable generation because of their (1) small system size and associated high sensitivity to variations in load and generation, and (2) lack of interconnections to buffer supply-demand imbalances. Thus, they are an excellent venue to develop and evaluate methods to optimize the value of forecast information in operational decision-making.

This study addresses the issue of which aspects of intra-day forecast performance are most critical to the management of renewable generation variability on an island system with a high penetration of variable renewable generation. The objective is to identify which forecast information provides value to operational decision-making and to design customized forecast evaluation metrics that more effectively measures the sensitivity of the operational decision-making to forecast error than traditional error metrics.

The venue for this investigation is the island of Hawaii, which is also known as the "Big Island". The electric system on the island is operated by the Hawaii Electric Light Company (HELCO). HELCO is a subsidiary of Hawaiian Electric Company (HECO), which operates the grids on five of the eight islands of the state of Hawaii.

II. THE OPERATING ENVIRONMENT

The renewable generation assets on the HELCO system during 2017 are listed in Table I. These provide "must take as available" generation that consists of 31 MW of wind generation and 16.2 MW of hydro generation. The output from these must-take resources can be reduced during low demand, but only after non-essential generators are cycled offline and reserves down are at minimum.

In addition to the system-level resources in Table I, there is also approximately 90 MW of "behind-the-meter" (BTM) distributed (mostly residential and commercial rooftop) PV generation that is mostly not visible nor controllable by HELCO. As a result of the combined impact of the utility-scale must-take resources and distributed PV, the system operator must make significantly more unit commitment decisions. Generators that were previously scheduled have been retired, and generators that were operated continuously are subject to offline cycling once or twice daily. This change in resources has created more variability in the demand to be served while also increasing the need for demand forecasting to make unit commitment and decommitment decisions.

TABLE I. SUMMARY OF THE RENEWABLE GENERATION RESOURCES ON THE HELCO SYSTEM

Type of Resource	Capacity
Geothermal	38 MW
Hydro (3 facilities)	16.2 MW
Wind (2 facilities)	31 MW
Solar (distributed behind the meter)	90 MW

The average net and gross weekday load profiles during each quarter of 2017 are depicted in Figure 1. The "net load" is the measurable demand served by the HELCO generation resources and incorporates the BTM PV generation that offsets some of the actual load during the daylight hours. The "gross load" is the true demand by users on the system. The gross load is inferred by adding the estimated PV generation to the measured net load. Therefore, the difference between the gross and net load is the estimated system-wide BTM PV production. The net load profiles indicate that there are four significant features that define the daily system management cycle: (1) a nighttime minimum between midnight and 0900 HST, (2) a morning peak between 0600 HST and 1300 HST, (3) a daytime minimum between 0900 HST and 1400 HST due to distributed PV generation, and (4) an evening peak between 1300 HST and midnight.

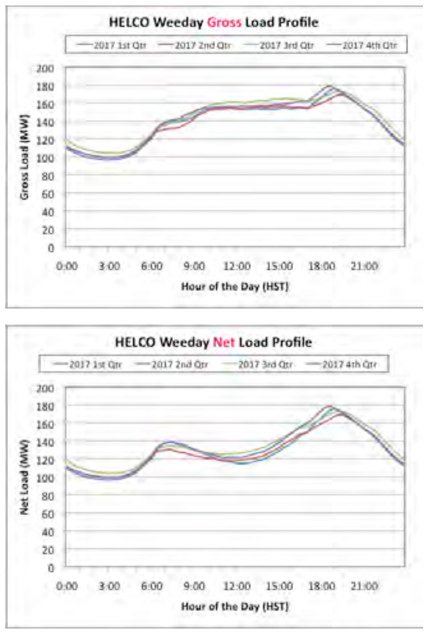

Figure 1. 2017 quarterly average gross (top) and net (middle) load profiles and their difference (bottom) for the HELCO system.

There are no interconnections between the electric grids on any of the islands that comprise the state of Hawaii. Therefore, the HELCO system operates without the ability to export or import power from neighboring systems, which of course increases the difficulty in managing generation or demand variability. All balancing must be done by the resources available on the HELCO system. Any imbalance results in a system frequency excursion.

III. KEY OPERATING ISSUES

The shapes of the gross load and solar generation profiles, along with the potential for significant short-term variability of the solar and wind generation and the attributes of the non-renewable generation resources, combine to create a set of ongoing operating issues that are characteristic of specific times of the day. This section presents three key issues.

A. *Planning for Morning Peak/Midday Minimum*

The first critical time of day is typically before sunrise at about 0500 HST. The challenge is to determine if midday net loads will be low enough to shutdown a unit after the morning peak. If so, a simple cycle CT that has no start/stop restrictions but less efficient can be used to serve the morning peak then shutdown when no longer needed. If net loads are expected to remain high, a more efficient combine cycle CT will be used throughout the day. One of the combine cycle plants has a permit and contract start/stop restriction, and is not allowed to have multiple starts in a calendar day.

A second issue at this time is whether an excess energy situation is expected due to high "as available" (listed in Table I) generation. The forecasting of the duration of the expected excess energy event is needed to determine whether curtailment of the as-available renewable generation or taking a unit offline will best address the situation. Unit will be taken offline if the excess energy event duration is greater than the minimum downtime of the unit.

Wind and solar generation forecasts available at 0500 HST for the middle of the day are important factors in the pre-sunrise decision-making. There are two key forecasting questions for the midday period: (1) will the distributed solar generation rise to its typical midday values or will the weather conditions be much cloudier than usual and this result in much below normal solar generation? (2) will the wind generation increase or decrease from its pre-sunrise level and thus either contribute to or offset a potential excess energy situation?

In order to evaluate the value of forecasts for the 0500 HST decision time frame, two parameters were defined. The first was named the Morning Solar Increase Parameter (MSIP), which was defined as the change in system-wide solar generation from 0800 HST to 1100 HST. MSIP measures the net increase in solar production during the 0800 HST to 1100 HST period. The second parameter was termed the Morning Wind Change Parameter (MWCP) and defined as the change in system-wide wind generation from 0500 HST to the 3-hour average for 0800 to 1100 HST. MWCP measures the change in wind generation between the decision time at 0500 HST and the late morning period.

The distributions of the MSIP and MWCP over the full year of 2017 are shown in Figure 2. This distribution was used to define three categories for each parameter. For MSIP the middle 80% of the distribution was labeled as "Typical". The lowest 10% was labeled as "Significantly Below" and the highest 10% was termed "Significantly Above". Similarly, the middle 80% of the MWCP was labeled "No Change" while the lowest 10% was categorized as "Significant Decrease" and the highest 10% was called "Significant Increase". The reasoning is that the tail events (lowest and highest 10%) had the most impact on operational decisions and therefore, it is useful to measure how well the forecasts can predict the tail events.

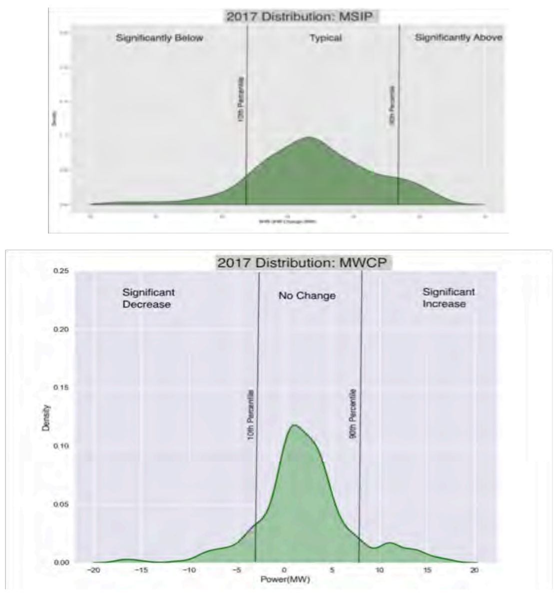

Figure 2. 2017 distribution of the (top) Morning Solar Increase Parameter (MSIP) and (bottom) Morning Wind Change Parameter (MWCP).

B. *Midday Net Load Ramps*

A second operating issue that must be considered is the probability of the occurrence of sudden large amplitude changes (i.e. ramps) in distributed solar production during the midday period (0800 HST to 1400 HST) that induce large ramps in the net load.

In these situations it is important to have adequate ramping capability available with the online units to ensure that the system frequency doesn't go too high or too low. As the PV penetration on the system increases, it is expected that the frequency and amplitude of these type of events will increase. Therefore, the ability to anticipate the periods for which the probability of this type of behavior is high will have considerable value to the grid operators to manage reserve capacity and ramping rate requirements.

A single parameter, named the Mid-day Solar Variability Parameter (MSVP), was used to evaluate the value of the forecasts for the mid-day decision period. The MSVP was defined as the range of the 15-minute average system-wide solar generation during the 1000 HST to 1400 HST period (i.e. Max Solar Generation (1000-1400) – Min Solar Generation (1000-1400)). The range was select as the measure of variability because it is the solar forecast variable most closely aligned the with magnitude of reserves needed during the mid-day period.

The distribution of the MSVP for the full year of 2017 is shown in Figure 3. This distribution was used to define three MSVP categories. The lowest 50% of the distribution was labeled "Low". The middle 20% was specified as "Moderate" variability and the highest 30% of the MSVP values were labeled as "High" variability. These categories were used to evaluate how well the forecasts can distinguish among days with low, moderate and high variability.

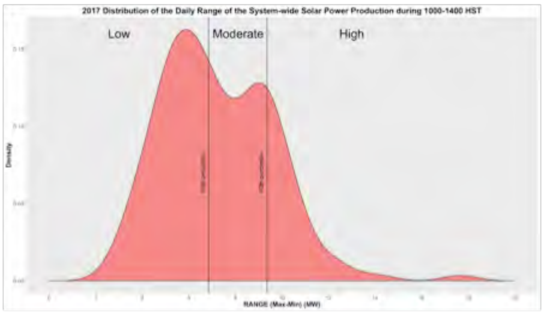

Figure 3. 2017 distribution of the Mid-day Solar Variability Parameter (MSVP).

C. *Preparation for the Evening Demand Peak*

Another important decision-making time is typically in the early afternoon (~1300 HST) when plans have to be made to position the system for the evening peak demand period. The key forecasting issues at this time are (1) will the distributed solar production decrease at a typical rate as evening approaches or will it decrease more quickly than the average rate (i.e. more late afternoon clouds than typical)? and (2) will wind remain constant, increase or decrease as the evening demand peak approaches?

The starting of combine cycle combustion turbines requires the unit be kept at constant load while the heat recovery steam generator is started, limiting the system's regulating capability. Each unit added increases the system minimum dispatch limit. An ability to anticipate the trend and variability of the wind can help to decide the timing of when to start bringing the units online. More accurate forecasts of wind and solar generation during this time period enable a better timing of the unit start-ups.

In order to evaluate the value of forecasts to the preparation for the evening demand peak, two operationally significant parameters were defined. The first was named the Afternoon Solar Decrease Parameter (ASDP) and defined as the change in solar generation from 1300 HST to the average for the 3-hour period from 1500 HST to 1800 HST. ASDP measures the rate of decrease in solar production during the middle to late afternoon as the system is being positioned for the evening peak. The second parameter was named the Afternoon Wind Change Parameter (AWCP and defined as the change in system-wide wind generation from 1300 HST to the average wind generation for the 1600 HST to 1900 HST period. AWCP measures the change in wind production between the decision time at 1300 HST and the start of the evening demand period.

The 2017 distributions of the ASDP and AWCP are shown in Figure 4. This distribution was used to define three categories for each parameter. For ASDP the middle 80% of the distribution was labeled as "Typical". The lowest 10% was labeled as "Significantly Below" and the highest 10% was termed "Significantly Above". Similarly, the middle 80% of the AWCP was labeled "No Change" while the lowest 10% was categorized as "Significant Decrease" and the highest 10% was named "Significant Increase".

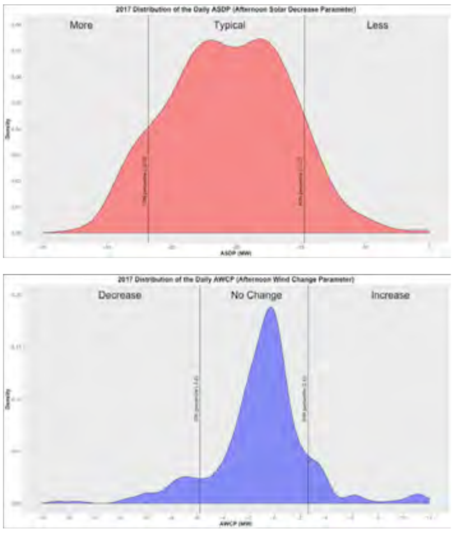

Figure 4. 2017 distribution of the (top) Afternoon Solar Decrease Parameter (ASDP) and (bottom) Afternoon Wind Change Parameter (AWCP).

IV. RENEWABLE GENERATION FORECAST SYSTEM

Renewable generation forecasts are provided to HELCO by a customized prediction system called the Solar and Wind Integration Forecast Tool (SWIFT) [1]. SWIFT is based on the multi-method ensemble approach to forecasting. In this approach, forecasts are generated by multiple forecast algorithms. The ultimate forecast that is delivered to the user is then created by statistically constructing a deterministic or probabilistic composite of the individual forecasts.

SWIFT provides wind and solar forecasts on two different look-ahead time scales. The first is a 6-hour look-ahead period with a forecast increment of 15 minutes that is updated every 15 minutes. This product is targeted for the type of intra-day decision-making described in the previous section. The second look-ahead time frame is 168 hours (7 days), with a 1-hour forecast increment that is updated on an hourly basis. This is targeted for longer term planning activities. The forecast content is the same for both look-ahead time periods.

Solar generation forecasts are provided for each utility-scale facility (there are none at present on the Big Island but some are planned) and for the aggregate of all distributed PV generation resources connected to each substation. Regional and system-wide forecasts are then produced by combining the forecasted production from the substations and utility-scale facilities. Wind generation forecasts are provided for each utility-scale facility. These are combined to produce a system-wide wind generation forecast.

An example of a SWIFT 6-hour ahead system-wide solar forecast is shown in Figure 5. The forecast is expressed in terms of nine probability of exceedance (POE) values. The 50% POE value is often used as a deterministic forecast.

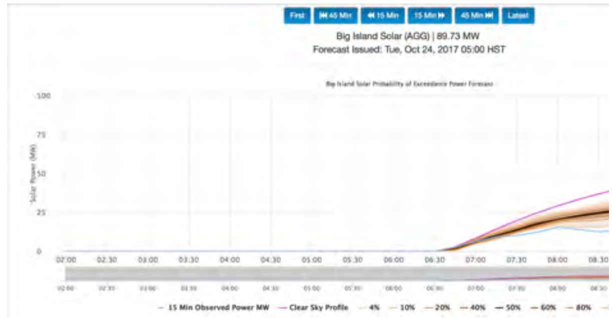

Figure 5. The distributed solar generation (MW) forecast for the HELCO system produced by SWIFT at 0500 HST October 24, 2018. The purple line represents the clear sky generation profile. The blue line is the estimated actual generation (calculated historically at the end of forecast period). The brown lines are the probability of exeedance (POE) forecasts from SWIFT.

V. CUSTOMIZED METRICS

A set of customized metrics was developed to measure the prediction skill with respect to the forecast attributes that are important to each decision-making scenario. An event-oriented category-based approach was employed. Each decision-making scenario and decision component was considered as a separate event. The forecasts for each event were expressed in a 3-category structure. The argument for a category-based structure is that decisions are typically based on an interpretation of information that divides the potential scenarios into bins for which the same decision will be made with little consideration of differences within the bins and the bins will have sharp boundaries that define the transition between alternative decisions.

The simplest way to visualize this concept of forecast evaluation is through the use of a contingency table of forecasted vs. observed (outcome) categories. A depiction of a 3-category contingency table is shown in Figure 6. This illustrates three forecast categories in the horizontal direction and the corresponding 3 outcome categories in the vertical direction. Table cells 3,5 and 7 (shaded in green) represent the correct forecasts. The cells shaded in yellow represent 1-category errors while the cells shaded in red represent 2-category errors.

The distribution of forecasts-outcome pairs among the cells provides an indication of forecast performance. However, it is useful to compress the set of numerical values in the cells into a single performance metric. Two different metrics were used for this purpose: (a) Critical Success Index (CSI) and (b) General Skill Score (GSS).

Bin #	Forecasted Category			
Onserved Category	Category	F_1: Below	F_2: Typical	F_3:Above
	O_3: Above	1	2	3
	O_2: Typical	4	5	6
	O_1: Below	7	8	9

Figure 6. An example of a forecasted vs. observed contingency table with numeric labels (1 to 9) assigned to each table cell.

A. *Critical Success Index (CSI)*

The CSI is a widely used event-oriented forecast performance metric [2] that is defined as

$$CSI = {}^{H}\!\!\diagup_{(H + M + FA)} \tag{1}$$

where H is number of "hits" (event is forecasted and observed), M is the number of misses (event is observed but not forecasted) and FA is the number of false alarms (event is forecasted but not observed).

While the CSI is simple to calculate and understand it is not a comprehensive nor flexible metric. For example, the CSI does not have the ability assigns weights to different types of errors or to account for multiple category errors.

B. *General Skill Score (GSS)*

Many of the issues associated with the CSI can be addressed through the use of a generalized skill score (GSS) in a manner similar to that formulated by [3] and [4]:

$$GSS = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{K} n(F_i, O_j)s_{ij} \tag{2}$$

where N is the total number of forecasts and $n(F_i,O_j)$ is the number of forecasts that fall into the cell identified as foreacst category "i" and observed category "j" (see Figure 6), K is the number of categories and s_{ij} is a scoring matrix that weights the contribution of each cell ($_{ij}$) to the GSS value.

This metric differs from the CSI metric by considering all forecast-outcome combinations (i.e. all cells in the contingency matrix) and also by the use of a scoring parameter (S_{ij}) that weights the contribution of each matrix cell in the calculation of the overall metric. The scoring metric provides the flexibility for metric customization. For example, it can be used to account for the sensitivity of the decision-making process to a particular type of error. For example, misses of particular outcomes can be penalized more than false alarms but misses and false alarms of other outcomes can be treated equally.

The properties of the GSS will depend upon the specification of the scoring matrix. A widely used definition of the scoring matrix treats all errors equally, has a linear penalty for multiple category errors (e.g. 2 category errors are penalized twice as much as 1 category errors), requires a score of 1.0 for all correct forecasts, and formulates the scoring matrix values such that a random forecast of the categories (with knowledge of the relative observed frequencies) will produce a GSS value of zero. In that case, if all forecasts are correct the GSS is 1.0 (100%). A negative score will occur if the forecast is worse than a random forecast. This is the specification that was used for this project.

VI. Forecast Performance results

The event-oriented category-based forecast evaluation paradigm was applied to all of the decision time frames. However, the definitions of the decision-making events and the forecast categories were specific to each time frame.

A. *0500 HST Forecast*

As noted in section IIIA two decision-making parameters (MSIP and MWCP) were defined for the 0500 HST forecast. Scatter plots of the forecasted vs. observed values for both of these parameters are shown in Figure 7. Each of the points in the scatter plot represents the forecasted and observed parameter values for one day of 2017. The horizontal and vertical lines on these charts correspond to the category boundaries (10^{th} and 90^{th} percentiles) defined in section IIIA. These lines essentially transform the scatter plot into a contingency table.

The numbers in each box of Figure 7 correspond to the contingency table labels in Figure 6. Thus, the points in cells #3, #5, #7 represent correct category forecasts. The points in cell #5 represent the correct forecasts of the large number of "Typical" or "No Change" cases that have less operational significance. Successful forecasts ("hits") of the tail events are represented by the points in cells #3 and #7. These are the cases that have considerable value in the decision-making process.

.

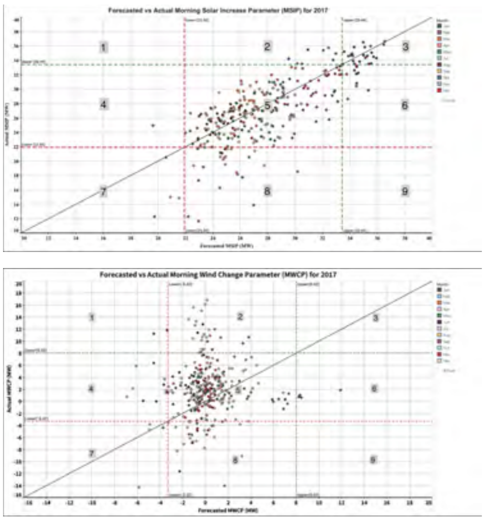

Figure 7. Scatter plots of the 0500 HST forecasted vs actual daily values of the (top) MSIP and (bottom) MWCP for 2017. Each point represents one day during 2017.

The computation of the GSS is shown in Figure 8. As described in Section V, the GSS is computed by multiplying the number of points in each cell by the corresponding scoring matrix value for that cell. Forecasts of the MSIP achieve a GSS of 39.1% while the forecasts of the MWCP only attain a GSS of 6.5%. This indicates that the forecast system, as currently configured, has very little ability to identify the days that will have a significant increase or decrease in wind generation during the morning period. The solar forecasts demonstrate more skill in the prediction of the large departures from typical conditions. However, from the scatter plot it is evident that this is predominantly because of their ability to correctly predict the less cloudy (i.e. more solar generation) than typical days (cell #3). The skill of predicting much cloudier than typical days (cell #7) is much lower.

GSS	Forecasted			
Observed	Category	F_1:Below	F_2: Typical	F_3:Above
	O_3: Above	0	-32	24
	O_2: Typical	0	232	0
	O_1: Below	4	-108	0
Category Scores		4	92	24
Total Score				120
Number of Forecasts				307
GSS (Total Score/# Forecasts)				39.1%

GSS	Forecasted			
Observed	Category	F_1:Decrease	F_2: NC	F_3:Increase
	O_3: Increase	-2	-108	0
	O_2: NC	0	243	0
	O_1: Decrease	4	-116	0
Category Scores		2	19	0
Total Score				21
Number of Forecasts				321
GSS (Total Score/# Forecasts)				6.5%

Figure 8. The GSS scores (bottom lines of each table) and its components for 0500 HST category-based forecasts of the MSIP (top) and MWCP (bottom) for all available forecasts during 2017.

B. *1000 HST Forecast*

Only one decision-making parameter (MSVP) was defined for the 1000 HST time frame. Scatter plots of the forecasted vs. observed values for MSVP are shown in Figure 9. The horizontal and vertical lines on these charts correspond to the category boundaries (50th and 80th percentiles) defined in section IIIB.

The computation of the GSS for the MSVP is shown in Figure 10 and indicates the forecasts of the MSVP achieved a GSS of 14.7%. The scatter plot and the components of the GSS indicate that the minimal skill of the current configuration of the forecast system for this forecast objective is mostly attributable to the ability to identify days with moderate variability (cell #5). Only one of the 10% of days of high variability was correctly forecasted (cell #3). Many of the high variability days had forecasts of low variability (cell #1) and this was a significant contribution to the low GSS. However, it should be emphasized that these variability forecasts were inferred from time series forecasts of the 15-minute average power production and were not optimized for the prediction of the range of variability.

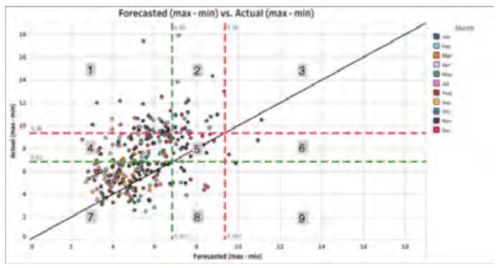

Figure 9. Scatter plots of the 1000 HST forecasted vs actual values of the MSVP for 2017. Each point represents one day during 2017. The numbers in cells correspond to the cell labels in Figure 6

GSS	Forecasted			
Observed	Category	F_1:Low	F_2:Moderate	F_3:High
	O_3:High	-74.3	-9.0	1.0
	O_2:Moderate	-30.9	20.0	0.5
	O_1:Low	143.0	-4.3	-0.5
Category Scores		37.9	6.7	0.9
Total Score				45.5
Number of Forecasts				310
GSS (Total Score/# Forecasts)				14.7%

Figure 10. The GSS scores (bottom lines of each table) and its components for 1000 HST category-based forecasts of the MSVP for all available forecasts during 2017.

C. *1300 HST Forecast*

As described in section IIIC two decision-making parameters (ASDP and AWCP) were defined for the 1300 HST forecast. Scatter plots of the forecasted vs. observed values for each day of 2017 for both of these parameters are shown in Figure 11. The horizontal and vertical lines on these charts correspond to the category boundaries (10th and 90th percentiles) defined in section IIIC.

The computation of the GSS for the 1300 HST forecasts is shown in Figure 12. Forecasts of the ASDP achieved a GSS of 31.7% while the forecasts of the AWCP had a GSS of -1.2%. The modest skill of the solar (ASDP) forecasts was largely attributable to the many cases (19 points in cell #7) in which a larger than typical afternoon decrease in solar generation occurred. The poor score of the wind forecasts was associated with very few correct forecasts (cells #3 and #7) of significant increase or decrease cases. The scatter plot indicates that the forecast system has a large bias towards the prediction of "No Change" cases.

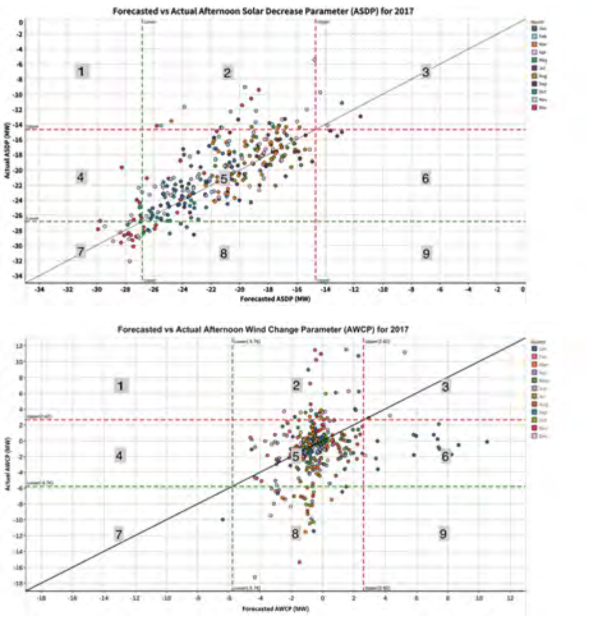

Figure 11. Scatter plots of the 1300 HST forecasted vs actual daily values of the (top) ASDP and (bottom) AWCP for 2017. Each point represents one day during 2017.

GSS		Forecasted		
Observed	Category	F_1:More	F_2: Typical	F_3:Less
	O_3: Less	0	-108	4
	O_2: Typical	0	231	0
	O_1: More	19	-48	0
Category Scores		19	75	4
Total Score				98
Number of Forecasts				309
GSS (Total Score/# Forecasts)				31.7%

GSS		Forecasted		
Observed	Category	F_1:Decrease	F_2: NC	F_3:Increase
	O_3: Increase	0	-124	3
	O_2: NC	0	248	0
	O_1: Decrease	1	-132	0
Category Scores		1	-8	3
Total Score				-4
Number of Forecasts				331
GSS (Total Score/# Forecasts)				-1.2%

Figure 12. The GSS scores (bottom lines of each table) and its components for 1300 HST category-based forecasts of the ASDP (top) and AWCP (bottom) for all available forecasts during 2017.

VII. SUMMARY

A study is in progress to identify which aspects of intra-day forecast performance are most critical for the management of renewable generation variability on an island system with no interconnections and a high penetration of variable renewable generation. The initial objective is to identify which forecast information provides the most value to operational decision-making and to then design customized forecast evaluation metrics that more effectively measure the sensitivity of the operational decision-making environment to forecast error than traditional error metrics. The ultimate objective is to optimize the forecast system to achieve the best possible performance as measured by these metrics. The platform for the study is the island grid system operated by the Hawaii Electric Light Company on the "Big Island" in the State of Hawaii.

A customized forecast evaluation system is being built from (1) the identification of the critical time periods and scenarios as well as the key parameters that impact operational decisions at those times, and (2) the formulation of forecast evaluation metrics that emphasize the performance for the prediction of key parameters during the critical time periods and scenarios.

The initial phase of this project has identified three key daily time periods with characteristic operating issues. A categorical forecast structure has been developed to focus on the key information for each of the three key decision-making periods. A generalized skill score has been defined to evaluate the categorical forecasts in a way that emphasizes performance in infrequent but key scenarios.

ACKNOWLEDGMENT

The authors thank the HELCO operational team for their extensive feedback on their use of forecasts in grid management decision-making.

REFERENCES

[1] J. W. Zack, J. D. Freedman, and D. Nakafuji, "A solar and wind integrated forecast tool (SWIFT) designed for the management of renewable energy variability on island grid systems", Proceedings of the 3rd International Solar Integration Workshop, London, 2013, Paper SIW13-1075.

[2] I.T. Jolliffe and D.B. Stepheson, 2003: Foreacst Verification: A Practitioner's Guide in Atmospheric Science. Wily, Hoboken, NJ, 240 pp.

[3] J.D. Gordon, "Evaluating the skill of categorical forecasts". *Mon Wea Rev*, vol.110, 1982, pp. 657-661

[4] J.P. Gerrity, "A note on Gandin's and Murphy's equitable skill score." *Mon Wea Rev*., vol. 120, 1992, pp. 2709-2712.