Advancing Satellite-based Solar Power Forecasting through Integration of Infrared Channels for Automatic Detection of Coastal Marine Inversion Layer

2nd International Workshop on Integration of Solar Power into Power Systems Lisbon, November 2012

Vladimir Kostylev, Andrey Kostylev, Chris Carter,
Chad Mahoney, Alexandre Pavlovski, Tony Daye
Green Power Labs Inc.,
One Research Dr.
Dartmouth, Nova Scotia, Canada
vkostylev@greenpowerlabs.com

Dallas Eugene Cormier,
Lena Fotland

San Diego Gas and Electric Co,
Mission Control Bldg A. 9060 Friars Road,
San Diego, CA, USA

***Abstract*-- The marine atmospheric boundary layer is a layer or cool, moist maritime air with the thickness of a few thousand feet immediately below a temperature inversion. In coastal areas as moist air rises from the ocean surface, it becomes trapped and is often compressed into fog above which a layer of stratus clouds often forms. This phenomenon is common in many parts of the world and poses a particular challenge for satellite-based solar radiation monitoring and forecasting. Hour ahead satellite-based solar radiation forecasts are commonly using visible spectrum satellite images, from which it is difficult to automatically differentiate low stratus clouds and fog from high altitude clouds. This provides a challenge for cloud motion tracking and cloud cover forecasting. San Diego Gas & Electric® (SDG&E®) Marine Layer Project was undertaken to obtain information for integration with PV forecasts, and to develop a detailed understanding of long-term benefits from forecasting Marine Layer (ML) events and their effects on PV production. In order to establish climatological ML patterns, spatial extent and distribution of marine layer, we analyzed visible and IR spectrum satellite images (GOES WEST) archive for the period of eleven years (2000 - 2010). Historical boundaries of marine layer impact were established based on the cross-classification of visible spectrum (VIS) and infrared (IR) images. This approach is successfully used by us and elsewhere for evaluating cloud albedo in common satellite-based techniques for solar radiation monitoring and forecasting. The approach allows differentiation of cloud cover and helps distinguish low laying fog which is the main consequence of marine layer formation. ML occurrence probability and maximum extent inland was established for each hour and day of the analyzed period and seasonal patterns were described. SDG&E service area is the most affected region by ML events with highest extent and probability of ML occurrence. Influence of ML was the strongest in coastal areas up to 50 km away from coast.**

Extent of ML inland in SDG&E service area was the largest in May and receded towards coast gradually through summer and fall. The ML probability correlated well with accuracy of solar radiation and PV forecasting in the region. Coastal sites appeared the most challenging for accurate forecast while PV forecasts in inland areas were performing consistently well. Improvement of automated cloud classification algorithms through integration of infrared satellite channels as well as better understanding of meteorological drivers of ML events is the key for improvement of satellite-based PV generation forecasting in coastal areas.

***Index Terms* — solar power forecasting, marine layer, coastal fog, distributed generation, remote sensing for power producers.**

I. Introduction

Solar power forecasting is an integral part of utilities' Smart Grid technology envelope. It is focused on improving reliability of power supply and operational efficiency of utilities and system operators with a growing share of solar power generation. Solar power forecasting brings substantial immediate and long-term benefits to utilities, system operators, and power producers by reducing the uncertainty of day ahead solar power generation commitments, minimizing system balancing and operational costs, and providing a operational data for broad ranges of installed solar capacity, including large-scale solar generation facilities and distributed solar installations.

Forecast accuracy strongly depends on the climatic conditions at the forecast site. Because of the high influence of cloud cover on solar radiation reaching ground, cloud regime strongly defines success of forecast performance. Regionally, and seasonally difference in forecast skill between sunny and cloudy climates, when tested using same sets of models, is almost double. Extensive testing of arious models carried in Europe shows RMSE ranges from 20% to 60% depending on location [1] and seasonal variability has similar effect [2, 3]. Locally, difference in forecast quality between coastal and inland sites demonstrates difficulty in accounting for details microclimate specific to a particular location. In coastal southern California, for example, the marine inversion layer common from May till September, is

creating overcast conditions throughout morning until noon. Cloud ceilings of about 1000 m and reduction in average visibilities to 6 miles contribute to forecast bias error of up to 54% on summer mornings [4].

The marine atmospheric boundary layer is generally defined as a layer or cool, moist maritime air with the thickness of a few thousand feet immediately below a temperature inversion. Along the coast of Southern California, the marine layer is created when subsiding air is heated when high pressure associated with the North Pacific Subtropical High passes over the cold waters of the coastal Pacific Current. As moist air rises from the ocean surface, it becomes trapped under a dome of subsiding high pressure and is often compressed into fog above which a layer of stratus clouds often forms. This phenomenon often occurs in late spring in Southern California. The relatively cool moist air in the marine boundary layer can intensify and prolong subsidence inversions which occur due to higher level, broad-scale sinking motions associated with high pressure systems along with lower level turbulence within the surface mixed layer [5]. The increased cloud cover makes a significant effect on the amount of solar radiation reaching the ground which consequently affects performance of solar energy generation systems. Solar power producers and utilities involved in solar energy currently require a significantly better ability to predict occurrence and propagation of marine layer inland.

In order to establish climatological patterns, spatial extent and distribution of marine layer, we analyzed visible and IR spectrum satellite images (GOES WEST) archive for the period of eleven years (2000 – 2010). The objectives of this work were to establish historical extent of marine layer impact and probability of its occurrence at different seasons. We also aimed at improvement of automated cloud classification algorithms through integration of infrared satellite channels for improvement of satellite-based PV generation forecasting in coastal areas. Cross-classification of visible and IR spectrum images is successfully used for evaluating cloud albedo in common satellite-based approaches for solar radiation monitoring and also helps distinguish low laying stratiform clouds and fog which is the main consequence of marine layer formation.

II. Marine Layer Detection Methodology

A. Data sources and analysis

Satellite images were obtained from the two NOAA's Geostationary Operational Environmental Satellites (GOES). GOES 10 images covered time span from May 1, 2000 to June 20, 2006, and GOES 11 - June 21, 2006 to August 31, 2010. The total time scale of analysis comprised 11 years, May 1, 2000 - Aug 31, 2010 with 4 months a year analyzed (May through August) corresponding to high likelihood of ML events. Satellite images are acquired every 30 minutes at 00 and 30 minutes of every hour for local daylight hours. For each acquisition event (sample) we looked at the following Visible Channel 1 (1 km resolution) and Infrared Channel 4 (4 km resolution) data with the total of 386,063 files analyzed.

The Extent of the Study Area for developing Marine Layer classification methodology was from -118.67° to -115.04° by Longitude and from 31.48° to 35.08° by Latitude. Geographical information used in this project included a digital elevation model (DEM) for the area, as well as the vector data for coastline and SDG&E service area boundary.

Marine layer detection and analysis was carried using the System for Automated Geoscientific Analyses (SAGA) Geographic Information System (GIS) software. SAGA has been designed for an easy and effective implementation of spatial algorithms and offers a comprehensive, growing set of geoscientific methods. This Free Open Source Software (FOSS) provides an easily approachable user interface with many visualisation options.

A pair of corresponding VIS and IR images was used to automatically detect ML. The logic behind the methodology was based on the fact that VIS spectrum images can not differentiate low stratus clouds and fog (typical for ML) from high altitude clouds. IR images (i.e. temperature maps) can not detect fog because its temperature is very close to the temperature of bare ground or water, but clearly detect high altitude clouds. Cross-classification of IR and VIS images however allows clear separation of ML. Threshold values were identified for visible spectrum images following frequency distribution analysis of pixel values for the studied years for the area of interest and we chose a threshold value of 200 to separate cloud cover from bare ground (Fig. 1).

The IR spectrum image pixel values correspond to temperature in degrees Kelvin times 10. We set the detection limit for high altitude clouds to temperatures below 272 Kelvin, this value is much colder than the temperature of ML during the months analysed and warmer than high altitude clouds since they are well below 270 K (Fig. 2).

Marine Layer was distinguished from total cloud cover by using Cross-Classification in the following manner: By comparing the two images generated in previous step, one of three classes were assigned to each ground pixel:

- Bare ground
- Marine Layer only
- High altitude Clouds

Since each class had a numeric value assigned to it we isolated Class 2 – ML only through all classified grids. This procedure provided a full set of ML Presence/Absence values for all 30 min satellite images for 2000 - 2010.

The probability of marine layer presence through day time hours was calculated for each day as the average fraction of time ML was present at a given pixel over the time period analysed. False positives generated by the bare

ground reflection in deserts and other high reflectivity areas were further removed through post-processing.

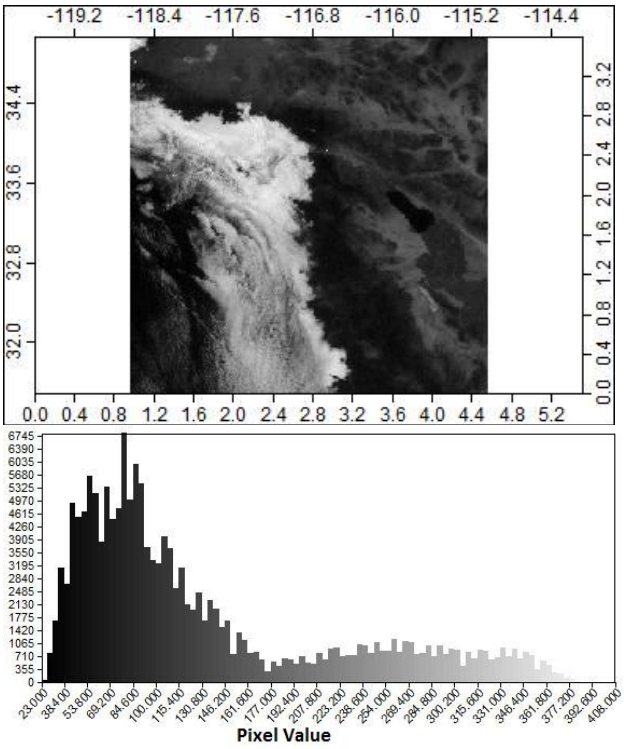

Fig. 1. Example histogram for VIS spectrum images showing two distributions of pixel frequency corresponding to bare ground (left peak) and cloud cover (right peak).

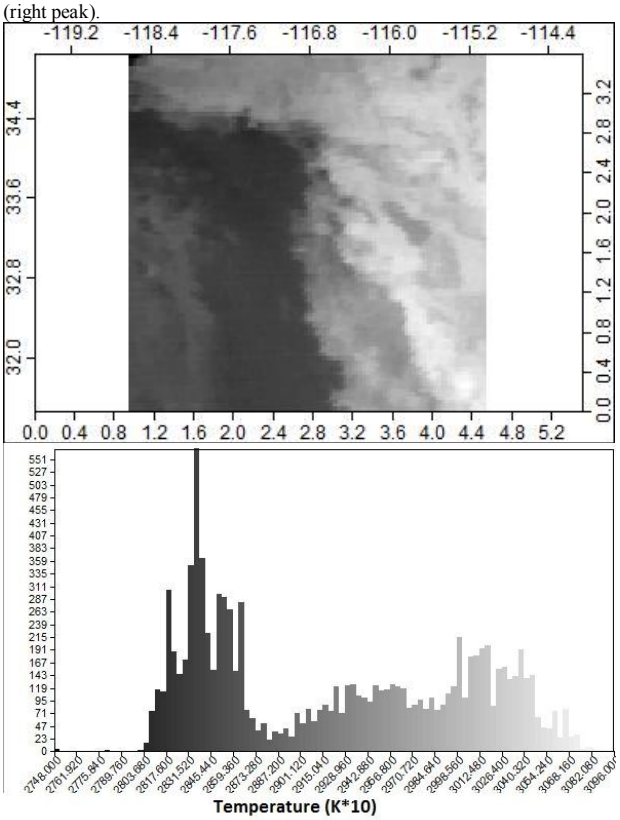

Fig. 2. Example histogram for IR image corresponding to the VIS image on previous figure. Similarly to VIS this histogram shows two peaks in distribution corresponding to bare ground (warm) and marine layer clouds (colder). High altitude clouds are not present in this image.

As a result of we have produced a number of outputs showing overall probability of marine layer occurrence during daylight hours at 1 km grid over service area, averaged over all analyzed images, as well as averages for each analyzed month and year; and overall monthly averages (all years pooled). The same results were produced for the maximum inland extent of marine layer inland defined at 5% probability level.

The figure 3 below shows for Marine Layer probabilities pooled for all months of all analyzed years, plotted as contour lines at 5% intervals. The results indicate that ML occurs with 15 – 20% probability during daylight hours over coastal zone in SDG&E service area during May – August months.

The extent of ML inland leading edge based on 5% probability boundary varies through months of a year (Fig. 4). The ML extends farther inland during the month of May and retreating closer to shore through June to August and is most common in Southern California.

B. Model validation

Accuracy of estimating probability of ML was carried for the extended area California-wide. The objective of this analysis was to estimate potential errors in calculating ML probability in relation to distance inland from coast as well as the variability of model accuracy from North to South along the coast. Validation was carried at 24 sites regularly distributed across the state. METAR station data from these sites were used to compile a set of historical observations for 2005-2010. Using this data, a parameter which indicates the presence of the marine layer was developed and analyzed by hour, month and year to generate observations –based statistics for the probability and variability of the marine layer. The following criteria were used to test for the presence of marine layer: 1) Presence of fog; 2) Presence of mist, and 3) Broken or overcast (5/8ths to complete) cloud coverage under 1500 ft. Any METAR observation that met one or more of these criteria was flagged as having the marine layer present. This historical data set was used as a baseline value with which the satellite derived forecasts were validated.

Validation of satellite-derived predictions of probability of marine layer occurrence by hour of day was carried because the expected probability of the marine layer throughout the day is known to vary. The coastal marine layer begins to develop during the night and moves inland, reaching its peak right around dawn. It then recedes through the day and begins to form once again after sunset. This trend can be seen in the marine layer probability derived from METAR measurements that follows. Since the satellite-derived probabilities use visual spectrum data as part of the forecast, only daytime hours could be compared.

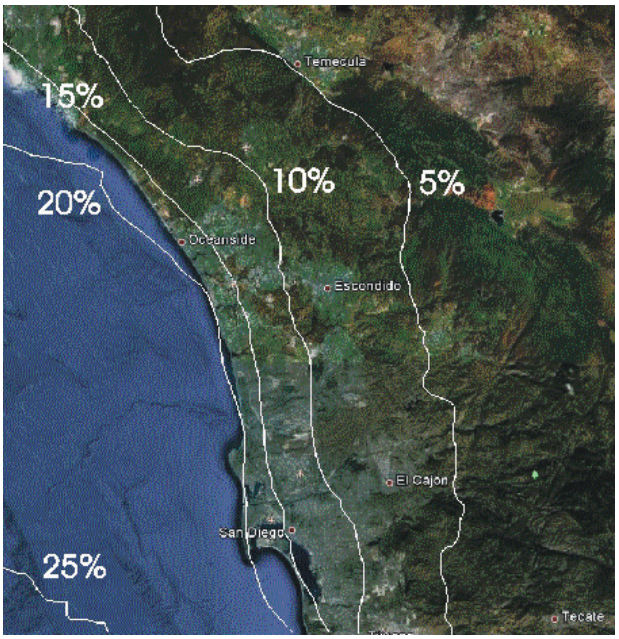

Fig. 3. Probability contours of Marine Layer presence for the full timescale examined, daylight hours. DEM shows land topography for reference.

Fig. 4. Contours showing inland extent of ML based on 5% probability for each month averaged over similar moths of each analysis year. Overall average ML inland extent over the study period is also indicated.

Under sunrise and sunset conditions, it appears that the satellite derived estimates are inadequate at determining the presence of the marine layer and result in a large under-prediction, likely due to deficiencies in the quality of the satellite images. Figure 5 shows marine layer probability by hour for KNKX METAR site for all months pooled. The probabilities derived from observed data (red) closely match the predicted (blue) for most hours except early morning where shadow regions on visible spectrum images confuse classification system.

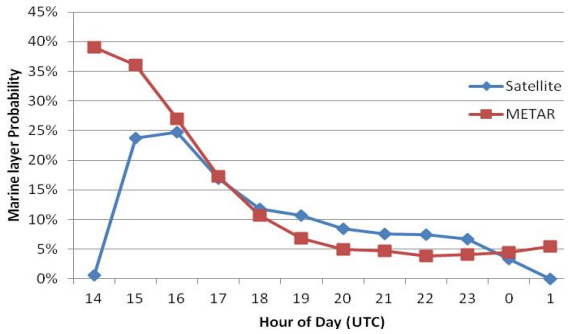

Fig. 5. Marine layer probability by hour for KNKX site. The probabilities derived from observed data (red) closely match the predicted (blue).

While the model performance in general exceeded our expectations, there was variability of model performance by year and month. Both satellite-derived and observed ML probabilities displayed large monthly and intra-annual variability. The following figures show that the use of satellite data can allow the predictions to capture the yearly variability that is shown in the METAR data:

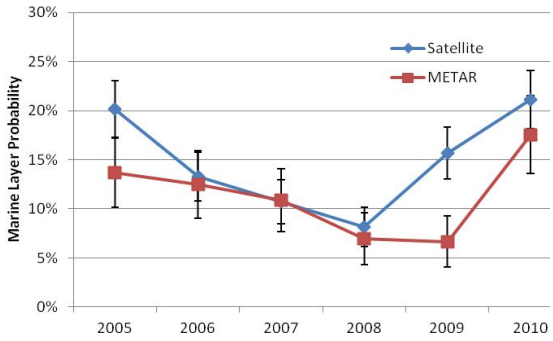

Fig. 6. Comparison of marine layer probabilities for the KSNA site in June. The satellite-based prediction will follow similar trends to the observed but may differ in magnitude.

The analysis shows that the correlation between observed and satellite-derived ML probabilities and associated errors vary greatly from site to site but overall for the sites experiencing ML on a regular basis, the correlations were high. For few sites that are located in the arid inland areas most ML detections were false positives so a low correlation does not indicate a poor performance. Likewise, the MBE and RMSE for those sites are similarly not applicable to evaluation of performance.

Overall the developed methodology is accurately classifying ML events along the coastal extent (within 50 km from shore) of California (Fig. 7), with highly significant correlations (0.65 – 0.99). Beyond this distance from shore this methodology is not applicable. There is also a significant decrease in correlation with increase in site Latitude (Fig. 8) suggesting that the strongest effects of ML are observed in San Diego and Orange county.

TABLE I. VALIDATION OF SATELLITE-DERIVED ML PROBABILITIES

NAME	MBE	RMSE	R2
	Coastal sites		
KSDM	5.30%	5.80%	0.99
KTOA	2.59%	3.71%	0.98
KAPC	-3.21%	5.65%	0.98
KSMX	-5.13%	7.13%	0.97
KCRQ	5.49%	7.02%	0.96
KEMT	7.28%	7.95%	0.95
KNSI	-14.79%	15.73%	0.94
KSBA	-1.73%	5.21%	0.94
KNTD	-2.68%	5.32%	0.94
KAVX	-3.15%	6.56%	0.87
KMRY	-2.36%	8.28%	0.85
KUKI	0.78%	4.70%	0.84
KACV	-9.03%	10.73%	0.79
KPRB	1.38%	8.68%	0.65
KLVK	1.65%	5.15%	0.65
Average	**-1.17%**	**7.17%**	**0.89**
	Inland sites		
KRRD	0.11%	2.56%	0.04
K9L2	15.47%	22.06%	0.01
KTVL	0.12%	1.62%	-0.01
KNID	5.66%	13.63%	-0.04
KPMD	5.44%	12.82%	-0.09
KBFL	4.56%	8.00%	-0.12
KMAE	1.39%	1.79%	-0.13
KMHR	3.47%	8.06%	-0.23
KBLU	-5.38%	5.89%	-0.24
Average	**3.43%**	**8.49%**	**-0.09**

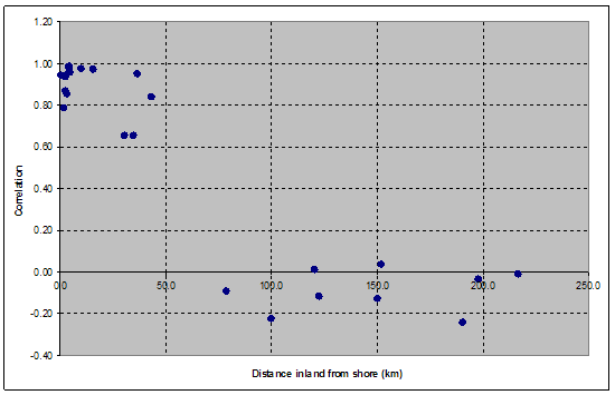

Fig. 7. Relationship between strength of correlation in marine layer probabilities derived from satellite vs. observed and distance to the METAR site from coast inshore.

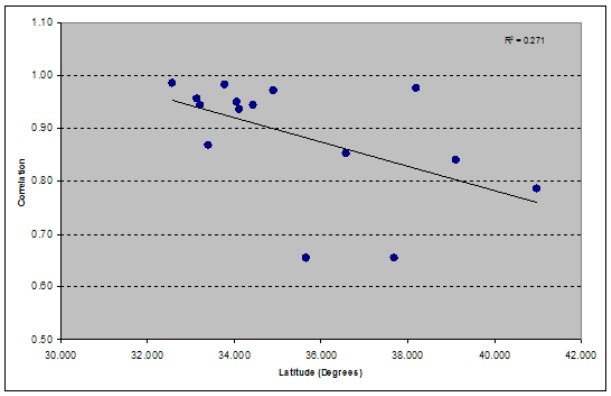

Fig. 8. Relationship between strength of correlation in marine layer probabilities derived from satellite vs. observed and Latitude of the test site.

III. DISCUSSION

Methodology was successfully developed and tested for using visible and IR channel satellite data to estimate probability of ML in California. SDG&E service area was found to be the most affected region by ML events with highest extent and probability of ML occurrence. Influence of ML is the strongest in coastal areas up to 50 km away from coast. Accordingly, the developed methodology can be used successfully in the southern California along the spatial buffer along the coastline.

Extent of ML inland in this area is the largest in May and recedes towards coast gradually through summer and fall. Overall probability of ML at any coastal location is 15-20% during day time hours in May – August inclusively. Satellite-based methodology successfully captures daily dynamics of ML and is highly correlated with ground-based observations.

1) Performance improvement

Satellite derived marine layer probabilities have slight negative bias compared to METAR derived. However, on further inspection, if the correlation between the two sets of data is analyzed, the result is somewhat more favourable. An explanation for the apparent discrepancy between the correlation coefficients and the magnitudes of the two sources is the fact that the METAR station reports clouds based on the entire visible sky whereas the satellite looks only at the points directly above the METAR station. In situations where the cloud cover is less than complete this would result in the METAR station seeing clouds when the satellite would not. The observation-based index developed for this comparison may not have been optimal and the observer-dependent accuracy of METAR stations data could have added additional uncertainty to validation as well.

Spatial registration of satellite images poses a concern for using satellites in operational forecasting without spatial adjustment. Satellite navigation errors are usually the main source of these errors [6]. More importantly georeferencing of the satellite images has a large implication to classifying ground pixels and consequently determining the probability of marine layer occurrence. In order to establish the effect ob bad georeferencing of ML classification we corrected geographic position of poorly references images manually for years 2000 (GOES10) and 2010 (GOES11). The X (east-west) and Y (north-south) offset directions were measured. Offset magnitudes (X and Y vector sum) of 3.25 km (SD=1.61) for GOES10 and 2.82 km (SD=1.79) for GOES 11 were recorded. The offsets are evenly distributed around zero and their magnitudes have no distinct temporal pattern. Comparing the visible spectrum pixel brightness values obtained from the satellite images pre- and post-offsetting at the 12 test sites was carried in order to evaluate the effect on potential ML misclassifications. The effect of the satellite image registration error is seen in -0.75% MBE

and 5.4% RMSE in pixel brightness for both satellites. It is apparent that the error associated with the registration of the satellite images is related to ML validation errors through biased estimates of pixel brightness and has to be accounted for in order to optimize the forecast of the marine layer. Automated techniques for correcting occasional inconsistencies in image navigation during wind production for meteorological needs have been attempted [7] however no automatic adjustment approaches are currently broadly available and development and accessibility of such is an immediate requirement for improving forecast quality.

2) Implications to solar power forecasting

GPLI has tested solar power forecasts at a number of PV generation locations in Southern California. Overall quality of forecasts varied depending on geography (particularly closeness to the coast) and seasonality (effects of coastal marine layer during months of April – October). For intraday forecasting half-hourly visible spectrum satellite images of cloud cover provide a convenient way to evaluate the amount of solar radiation incident of the ground. From 1 to 6 hours ahead satellite-based models produce results superior to NWP-based and persistence based models [2]. Cloud motion vectors are commonly used to interpolate future position of clouds over ground, however poor understanding of cloud types, unrealistic assumption of steady state cloud cover and clouds moving at different altitudes produce an additional challenge in correct interpolation of the next cloud image. All of these challenges are presented to forecaster during marine layer events. Low laying fog and stratus clouds may be advected together with high altitude clouds when a typical motion vector approach is applied, which leads to incorrect extrapolation of cloud cover over forecast location and consequently bad solar power forecast.

The problems of forecasting in the coastal zone affected by ML are illustrated below. Table 2 shows forecast validation results for two sites in Southern California – one on a coast and another inland, beyond the effect of ML. The hour ahead (HA) and day ahead (DA) forecasts are shown.

TABLE II. EFFECT OF ML ON SOLAR POWER FORECAST

	Coastal site		**Inland site**	
	HA	DA	HA	DA
RMSE	16.8%	19.6%	12.2%	14.1%
MAE	13.2%	14.0%	8.5%	9.3%
MBE	9.1%	2.6%	3.2%	-0.2%
R^2	0.76	0.63	0.8	0.75

The coastal site forecasts exhibit lower accuracy and precision than the sites not affected by marine layer. This result is consistent with our identification of variable standards of solar forecasting in cloudy vs. clear sky climates [8]. Incorporation of ML forecast and integration of its results with motion vector forecast of cloud cover is the way to remove a significant amount of forecast uncertainty.

Improvement of the ML classification model and implementation of ML forecast along with solar power forecast requires better understanding of climatological drivers of ML events. Development of air pressure gradients, onshore-offshore temperature gradients, presence of temperature inversion and direction of winds at inversion heights all could provide valuable information for forecasting the probability of ML clouds. Through a combination of satellite-based models, high resolution numerical weather prediction models, and dense measurement and metering infrastructure, understanding of ML phenomenon and strong ML predictive capacity can be gained. This will be an important breakthrough for forecasting location-based and distributed solar power generation and net load in coastal California.

IV. ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of Marlene Moore for help with this manuscript, insight, support and accessing industry needs in solar forecasting. We also thank GPLI staff – Jim Fletcher, Shawn Huang, for their persistent work on developing, implementing and analyzing performance of solar forecasting approaches.

V. REFERENCES

[1] E. Lorenz, J. Remund, S. C. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, J. A. Ruiz-Arias, V. L. Fanego, L. Ramirez, M. G. Romeo, C. Kurz, L. M. Pomares, C. G. Guerrero, "Benchmarking of Different Approaches to Forecast Solar Irradiance," *24th European Photovoltaic Solar Energy Conference*, Hamburg, Germany, pp. 21-25, September 2009

[2] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Renné, T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy* vol. 84, pp. 2161-2172, December 2010.

[3] I. Moradi, R. Mueller, B. Alijani, G. A. Kamali, "Evaluation of the Heliosat-II method using daily irradiation data for four stations in Iran," *Solar Energy,* vol. 83, pp. 150–156, 2009.

[4] A. Nottrott, J. Kleissl, "Validation of the NSRDB–SUNY global horizontal irradiance in California," *Solar Energy,* vol. 84, pp. 1816–1827, 2010

[5] S. F. Iacobellis, J. R. Norris, M. Kanamitsu, M. Tyree, and D. C. Cayan, Climate variability and California low-level temperature inversions. California. Climate Change Center, Publication # CEC-500-2009-020-F, 2009.

[6] B. Gibbs, GOES Image Navigation and Registration. *SatMagazine*, July 2008. Available online at http://www.satmagazine.com/cgi-bin/display_article.cgi?number=1889485593

[7] S.J. Nieman, J. Daniels, D. Gray, S.T. Wanzong, C.S. Velden, and W.P. Menzel, Recent performance and upgrades to the GOES-8/9 operational cloud-motion vectors. Proc. Third Intl. Winds Workshop, Monte-Verita-Ascona, Switzerland, EUMETSAT, 1996.

[8] V. Kostylev and A. Pavlovski, Solar Power Forecasting Performance – Towards Industry Standards. Grid Integration Workshop, Aarhus, Denmark, November 2011.