Mid-term Solar Forecast system based on Geostationary Satellite data

Zhenzhou Peng [#1], Shinjae Yoo [#2], Dantong Yu [#3], Dong Huang [*4]

[#]*Stony Brook University*
100 Nicolls Road, Stony Brook, NY 11794
[1]zhenzhou.peng@stonybrook.edu

[*]*Brookhaven National Laboratory*
50 Bell Avenue, Upton, NY 11973
[2]sjyoo@bnl.gov
[3]dtyu@bnl.gov
[4]dhuang@bnl.gov

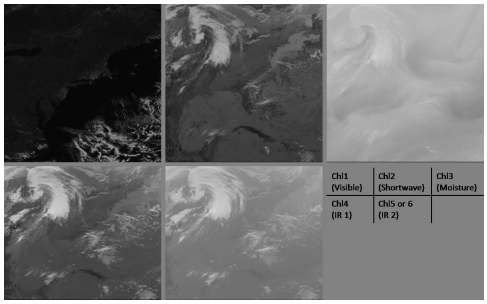
***Abstract*—Prediction of solar energy has become a significant concern in smart grid field. In this paper, a mid-term forecast system is designed with utilizing Optical Flow Motion Estimation on multi-channel geostationary data and Support Vector Regression(SVR) on solar radiation. Different from previous approaches, our system is trying to improve satellite model precision and fill the gap of forecasting in 30 minutes up to 5 hours. The experiment result shows that performance in both estimation and forecasting has been significantly improved.**

I. Introduction

With utilization and economy concern, planning and managing of control in advance acquires great importance in solar grids system. For the purpose of making decisions in near future, a forecast system is needed to provide information of solar energy tendency and distribution. In general, distribution and variation of cloud is the main source of solar radiation fluctuation. Therefore we propose to use Geostationary satellite data as input to extract cloud properties and track the motion.To better understand how cloud influence the radiation, we divide the work into two parts: Estimation of irradiance based on current cloud information from satellite and forecast of cloud motion in near future. In short, modeling and forecasting.

The reason of division is to improve satellite estimation model and at the same time make robust prediction scheme. Although the idea of using satellite data as input of solar energy estimation is not new, there are some challenging issues on cloud related fields: 1) Cloud properties is hard to recognize, 2) More meteorologic variables,such as atmospheric indices are needed, 3) low Spatial and temporal resolution of estimation.

As regards forecasting, the more uncertainties emerge with time of prediction increases. Even in the short period of time, cloud motion is hard to detect due to: 1) motion variation is unpredictable and hard to track,2) Cloud motion may needs cloud classification and detection in advance, 3) formation and deformation of cloud. Therefore the imprecision of motion estimation will be magnified with longer time prediction. In brief, motion estimation algorithm is required to have not only accuracy but also resolution so as to assure the quality of prediction.



With these concerns, our novelties lie in data source, model selecting, and more importantly, cloud motion forecasting:

1) **integration of multi-channel** Besides visible channel, we import other 4 channels multi-channel data of Geostationary Operational Environmental Satellite(GOES) imaging system into model. Since each channel can sense radiant and solar reflected energy in different spectrum range, comprehensive cloud properties can be presented with multispectral view.(Figure **??**).
2) **Data Preprocessing** Abnormalities and noise are common in both satellite images and measured radiation especially when multi-channel data is used. Before feeding to model, filters and other procedures are applied to clean input data. Besides, regarding radiation from ground station, a non-meteorologic method is presented to deal with normalization and integration.
3) **radiation feature with time** Radiation measured in ground station can be treated as a miniature of microscale solar energy change. We propose to add the variation of solar radiation as another concerned feature so as to provide localized knowledge.
4) **New model with regression** In previous works related to irradiance estimation, linear model is widely used with empirical correction for local variation[1]. We address it differently using the idea of SVR which takes both linear and non-linear relation into consideration.
5) **New motion estimation scheme** As motion tracking is the core procedure of modeling, our system implements an Optical Flow(OF) motion estimation algorithm on

the basis gradient of grey-scale change. Compared with traditional algorithms, it is more sensitive to small area change and robust in shape distortion.

In this paper, previous works and potential approaches are discussed in Section II. In Section III, preprocessing scheme and SVR method is presented in details. Later in Section IV, motion estimation techniques are discussed and compared. As for details of satellite model, Section V formulate 7 estimation models convering linear and non-linear, regular and SVR. In Section VI, performance of satellite models is presented on aspects of estimation and forecasting. Based on these results,in Section VII, we conclude that the new forecast system has significant improvement in mid-term radiation forecast field.

II. Background

As estimating solar radiation data from satellite images is more accurate than interpolating data measured by a modern radiometric network,[2], Many works regarding mesoscale range have been develop to do radiation estimation using satellite data. In early years, Satellite models are built up firstly to correlate cloud coverage with Global Horizontal Irradiance(GHI)[3], [4], [5]. Later works follows this idea by studying on linear relationship between Direct Normal Irradiance (DNI) and satellite visible channel[6], [7]. These models use "Cloud Index"(CI) as optical density derived from satellite data to indicate fraction or coverage. By using multiple empirical clear sky models, the local distribution of solar energy is derivable from satellite image[8], [9]. More recent works majorly follow this idea but with different concerns such as terrain factor[10], but with new ideas coming from knowledge of other fields such as statistical approach[11] and Artificial Neural Network(ANN) method[12], [13], moreover even without including meteorological data[14]. In fact, the biggest problem of framework of cloud coverage lies in untrusted and unstable visible channel image, since snow coverage and floating brightness of image with various zenith angle. Though Multispectral analysis is the most known and developed method to detect clouds, this work relies on heuristic thresholds for infrared and visible range are needed for cloud classification[15] and snow detection[16] Another drawback of previous works is that simple linear relation derived from cloud coverage has limitations in describing variation under cloudy condition. Therefore, a customized constant must be used as an estimation compensation to reduce linear biased influence [1].

On the topic of forecast using satellite, studies are mostly around cloud prediction. To describe cloud coverage in advance, cloud motion vector extraction is usually used for estimating cloud movement. As an import input parameter to feed satellite models. Cloud motion tracking algorithm is commonly generated from blockwise cross-correlation matching[17], [18]. It turns out to be highly sensitive to block size and segmentation as it try to represent area of cloud in block unit. In fact, this methodology is based on a basic assumption that cloud motion is stable and identical with no deformation. But in reality, cloud on satellite image is more complicated as with rotation and shape changing. Therefore merging and splitting of cloud will be ignored due to fixed block size. Another extreme case is that when multilayer clouds appear, block matching will fail as its correlation only covers the texture information in a block. Therefore a lot of recent works tried evade the unsolved issue of cloud motion by integration of other source of information,e.g. radar ground measurement. One way to to predict cloud ahead is using ground radar to get continuous radiation fluctuation trend[19]. Another work explores the time series feature of cloud statistically to do prediction[20]. The drawbacks of no-motion methodology is that cloud tracking is of low precision and forecast time can only be either minutes or up to 2 hours as they claimed. Though satellite approach is also in mesoscale, local information can be assimilated through adding ground-based pyranometer and multispectral views.Another drawback of current models is that the precision in both temporal and spacial aspect decrease rapidly with the longer forecast period. Due to the restriction of motion estimation algorithm, cloud

In our approach, we are targeting to use the ground-based pyranometer and satellite image to get estimated radiation in future. The local instrument can provide Direct Normal Irradiance(DNI) per second while satellite image has 30 minutes response time before next scanning. For pre-scheduling need, radiation forecast coverage starts from 30 minutes to 5 hours. Compared with pyranometer, satellite image has low spatial resolution(1 km x 1 km in VIS channel, 4km x 4km in Infrared channel), especially in multispectral and motion vector related applications. To meet to requirement of local forecast, pyranometer data is assimilated into satellite model as another feature.

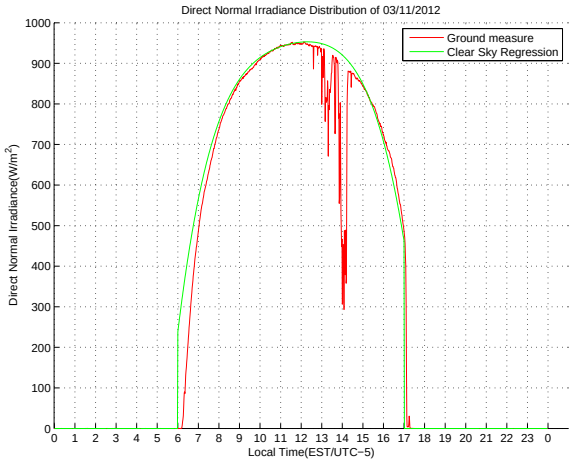
III. Preprocessing and modeling

In this section, we present the idea of preprocessing pipeline at the beginning. This procedure is significant for data selecting and noise filtering. In general, preprocessing includes satellite multi-channel data and radiation data. The need of radiation data processing originates from ground measurement data normalization and clear sky calculation. In later subsection, a new satellite model using SVR is introduced. Compared with previous linear approaches, we propose a new linear relation solution using SVR linear kernel. To utilize the SVR more, we also develop non-linear solution to the forecast problem.

A. Preprocessing

Since remote sensing used in geostationary satellite relies on scanning of radiometer, in some cases, the capture of will fail due to unresolved reasons. From the images postprocessed from raw data, it is more obvious and intuitive to see the errors. Since multi-channel data are proposed to be used as input data, it is necessary to preprocess images before modeling and prediction further. From the dataset collected by GOES project[], we find several patterns of error that appear frequently: 1) black rasters of multispectral images due to failure of sensing or raw data processing, 2) luminance variation, especially on visible channel, 3) Multi-channel timestamp mismatch. One solution of preprocessing to solve these common errors is to use empirical filters such as mean filter and bad-frame filter. Another one is to regulate the brightness with normalization with zenith. The overview of preprocessing is shown in Figure **??**.

Besides preprocessing of satellite data, another key procedure is related to radiation measured by pyranometer. As



ground truth of model output, and maybe input feature, how to remove noise of radiation is crucial in modeling and forecasting. With this concern, raw radiation is processed through two modules in parallel and combined together with using $normalizationandintegration$ scheme to output fraction of radiation over clear sky (Figure **??**). Similar to satellite data preprocessing, One of the modules is responsible for noise removing with filters and thresholds. Differently, the other radiation module is designed for clear sky irradiance calculation.

The importance of clear sky model is that normalization based on clear sky value indicates the fraction of cloud coverage which is more stable and less sensitive than radiation value itself which fluctuates with different atmospheric conditions. But in most clear sky models discussed in II, the calculation of solar direct normal irradiance is based on parameterization model. In other words, model takes a lot of atmospheric input parameters such as O2, CO2, ozone, water vapour and aerosol optical thickness (AOT). The setting up of each model requires peripheral instruments for additional information. As a result, prediction of radiation change becomes more challenging in terms of precision and location. Thus we designed a new method to solve the estimation problem but with less variables involved. With only historical radiation data, a clear sky model can be estimated using statistical method,for instance, regression. In our model, 2-order polynomial regression is used for generation of monthly clear sky curve(figure III-A).

To combine two models together in final step of radiation preprocessing, $normalizationandintegration$ is proposed. $Normalization$ stands for the calculation of fraction over clear sky value while $integration$ requires radiation and clear sky value at certain point to be integral over time. The idea of using integral is because of the mismatch of time range between satellite and ground base station. On one hand, the routine scan of satellite is 30 minutes while regular pulse of pyranometer is in 1s unit. On the other hand, As scanning of takes minutes for the whole plate view, timestamp of satellite is not that accurate. As a result, fluctuation in between and imprecision of time needs to considered in preprocessing. With integral around certain period at satellite timestamp, radiation is less sensitive than picking single value and able to capture the variation during the span. Then fraction of radiation is calculated using equation 1.

$$DNI_{norm} = \frac{\int_{t-N}^{t+N} DNI}{\int_{t-N}^{t+N} CSI} \tag{1}$$

where $2xN$ is the integral period at time t.

B. SVR modeling

With preprocessed input data of both multispectral images and radiation feature, we summarise modeling as a regression problem with a set of training patterns $(x_1, y_1), \ldots, (x_n, y_n)$, where $X_i \in R^N$, i=1,...,n. Specifically in our concern of satellite model, the idea is implemented as: $y_t \leftarrow x_t, x_t = \{Chl1_t, Chl2_t, Chl3_t, Chl4_t, Chl6_t, y_{t-\Delta t}\}$,where $Chln$ stands for satellite channel n, y is DNI value from preprocessing. If the relation is linearly formulated, the estimation output is presented in (2) as a function of (w, b). $\langle . \rangle$ stands for dot product.

$$f_t(x) = \langle w, x_t \rangle + b, w \in R^N, b \in R \tag{2}$$

To solve this problem, we use Support Vector Regression(SVR). SVR utilize the idea of maximum margin with regularization of cost function. The solution of SVR can be treated as a classic quadratic optimization problem. In general, it is equivalent to solve:

$$\begin{aligned} minimize \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ subject\,to \quad & \left\{ \begin{array}{c} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle - y_i + b \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{array} \right\} \end{aligned} \tag{3}$$

SVR is also useful for non-linear problem as it uses non-linear kernels. The kernel of SVR is a mapping method which can be transferred as linear relation. Once the linear relation is defined from kernel mapping, quadratic optimization problem can be solved without additional modification on objective function and constraints. In this paper, we use a kernel named radial basis function(RBF) for our non-linear modeling. The $f(x)$ in RBF concern is rewrite in (4) .With RBF kernel, the estimation formula is different from previous linear SVR approach, since it contains a mapping of $x_i \leftarrow \Phi(x_i)$. In RBF method, $\Phi(x_i)$ is implicit mapping and can be ignored through kernel $k(.,.)$,defined in (5) calculation in optimization.

$$f_t(x) = \langle w, \phi(x_t) \rangle + b \tag{4}$$

$$k(x, x') = \langle \phi(x), \phi(x') \rangle = e^{-\frac{\|x-x'\|^2}{2\delta^2}} \tag{5}$$

In SVR model calculation, an Open source library named LIBSVM[21] is used to solve the optimization problem given satellite data and radiation values. The optimal solution is then generated through tuning regularization parameters.

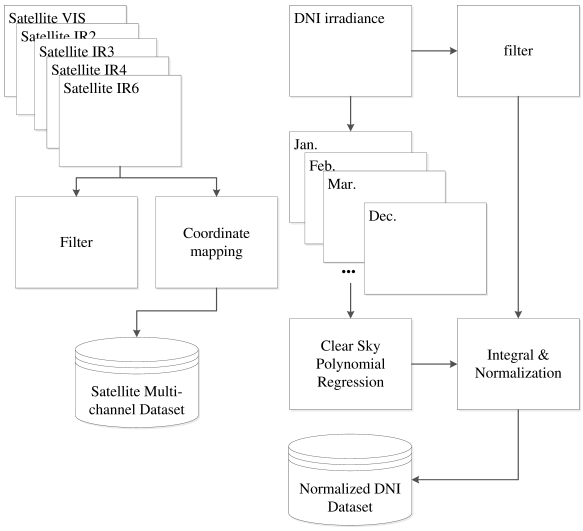


Fig. 1. Data Preprocessing Framework

IV. Motion Estimation and Forecasting

Forecasting part in this paper is equivalent to capturing radiation feature with time and cloud tracking. The radiation related prediction uses previous radiation directly as pre-knowledge of prediction while cloud tracking relies on motion estimation on satellite image. The goal of using satellite image is to get the movement of cloud field and predict cloud location after minutes or even hours. Thus, algorithms of motion estimation are crucial for precision of forecasting output. In this section, we propose to use Optical Flow(OF) motion estimation method as the core of tracking algorithm. To better illustrate the performance of it, a comparison is done with another common method using block-matching.

A. Optical Flow Motion Estimation

As motion estimation is not a new topic since in image processing and multimedia field, a lot of approaches have been carried out. The most common method is blockwise motion estimation which is based on block similarity to estimation displacement of each block. This approach is widely used in video coding and compressing as its properties of fast speed and high compressing ratio. But the performance of block related algorithms is very sensitive to region and feature matching which assume consistence of segmentation and luminance in region. In other words, with cloud variation and uncertainty on satellite image, block matching is very unstable varying with blocksize and similarity criteria. As a result, to get rid of these issues in satellite application, we turn to use Optical Flow algorithm after comparing with traditional block matching.

Optical Flow Motion Estimation is a branch of methodology that utilize the gradient of image. Under the assumption of constant illuminance, displacement of image will be estimated through ingredient change. Optical Flow has a lot of variations as ingredient of image is defined differently. In our approach, we choose Lucas-Kanade Optical Flow(LKOF)[22] as the tracking method. To improve the robustness of motion tracking on satellite image, we implement a pipeline which use recursion to approach the best tracking result. Visible channel image is scaled into several layers size(in experiment

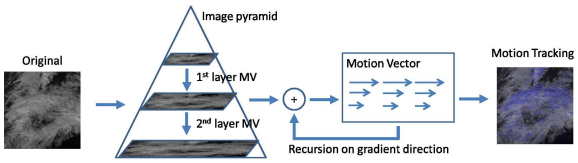


Fig. 2. Optical Flow Motion Estimation Pipeline

1000x1000, 500x500 are used), for smaller image(Upper layers in image pyramid) with higher ratio, OF tracking is applied for detecting motion vectors. This motion vector will be initial motion vector to the next layer. The final layer is normally with no scaling, OF is recursively done with the direction of gradient change. Then final motion vector map is generated for each pixel. In reality, window size and number of layers need be to tuned in case of over-estimation. The overview of pipeline is presented in Figure 2.

B. Optical Flow vs. Block-based matching

To illustrate the difference, we design two experiments to compare Optical Flow with Block Matching method in both image scale and pixel scale. To measure the goodness of motion estimation, the most straightforward method is to use output motion vectors directly to generate predicted image and compare with ground truth image. Based on this idea, Displaced Frame Difference(DFD) is used as benchmark of compare two images. In this paper, DFD_{avg} is generated by averaging DFD over all pixels(**??**).

$$DFD_{avg} = \frac{\sum_{i,j} |I_t(i,j) - I_{t-\Delta t}(i + u_{i,j}\Delta t, j + v_{i,j}\Delta t)|}{N} \tag{6}$$

where u,v is motion vector of pixel (i, j) in direction of x,y, respectively.

Block Matching method used for comparison is Hierarchical Block-Matching(HBM). It uses bigger block(50 x 50 pixels) as the representative of smaller block(10 x 10 pixels) to decrease influence of noise in small area. The similarity is defined as normalized cross correlation. As a result, HBM generates motion vectors in unit of 10 x 10 resolution. In LKOF method, we take 10 X 10 region as gradient calculation, 2 layer pyramid, and up to 100 iterations to generate pixel-scale motion vectors. The comparison of DFD score is shown in Figure IV-B. The result shows that Optical Flow with mean filter is better in motion tracking.

DFD score, however, is a criteria per image which consider the whole image as unit. It means that DFD is not able to present the goodness of slow motion or motion in small scale. In addition, as a sum of absolution error, it puts more weight on no motion detection because motion since movement of pixels leaves gap(black area) in previous position. But this should not be encouraged especially motion is not that obvious. Therefore, we pick another criteria to get evaluation of prediction results in local view. In VI, we use satellite model to output forecast radiation based on predicted images, with ground truth from local station, a evaluation can be presented with deviation in terms of radiation.

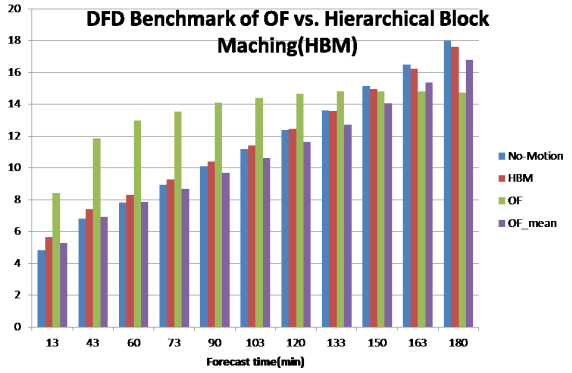


Fig. 3. DFD score of OF vs. HBM

V. Radiation Prediction Models

As mentioned in III-B, satellite model is formulated as $y_t \leftarrow x_t, x_t = \{Chl1_t, Chl2_t, Chl3_t, Chl4_t, Chl6_t, y_{t-\Delta t}\}$. The $y_t = f(x_t)$ mapping can be linear or non-linear. In single-channel model, linear relation is commonly used for cloud coverage extraction. While in multi-channel approaches, non-linear idea is also accepted since input data dimension increases more than 5. In this paper, 7 of satellite models are implemented for estimation and forecasting evaluation. On the whole, 5 of them still stick With linear exploration while 2 of them try to consider the problem as non-linear. In the view of input data, 2 linear models are built directly on visible channel , others utilize multi-dimension input data.

For comparative study, we use Cloud Index(CI) model as our baseline. This is a widely used estimation model on the bais of linear relation between single-channel with radiation. One drawback of this model is the normalization since dynamic range of channel with time is not stable and visible channel may have brightness variation. Driven by this idea, two extensions are implemented with modifications: 1) normalization with zenith, 2) removal of compensation for bias error. The first extension is Linear Regression(LR). It considers visible channel with regression method instead of calculate CI first. The other one is Aggregate Linear Regression(ALR) which is linear regression but with multi-channel data. CI,LR,ALR, are all non-SVR approaches since no SVR optimization is involved.

The rest 4 models are all SVR models. Similar to as ALR, SVR models also take multispectral data into consideration. In first two approaches, SVR uses linear kernel($SVR - Li$) and RBF kernel($SVR - RBF$) to estimation radiation with only satellate data as input. Driven by the idea of radiation with time feature, two other SVR models are developed using satellite data and previous radiation as input. One is extended from linear kernel named as ($SVR - Li_{rad}$) and the other is from RBF kernel($SVR{-}RBF_{rad}$). Moreover, another concern in modeling is that with multiple dimension of input and uncertain quality of motion prediction, output of radiation can be noisy and outbound. In other words, with normalized radiation as training data, model can output abnormal values like below lower bound 0 or beyond upper bound 1. Thus, for the sake of range control, final output of $SVR - Li_{rad}$ and $SVR - RBF_{rad}$ is trimmed with threshold.

1) **Cloud Index(CI)** Implementation of CI model is following [1] but with our statistical clear sky model.

$$CI = \frac{x_1 - bound_{min}}{bound_{max} - bound_{min}}, x_1 \in [0, 255], \quad (7)$$

$$f(x) = w(1 - CI) + b, CI \in [0, 1] \quad (8)$$

2) **Linear Regression(LR) on single channel** LR implements directly on visible channel.

$$f(x) = wx_1 + b, x_1 \in [0, 255] \quad (9)$$

3) **Aggregate Linear Regression(ALR) on multi-channel** 5 GOES channel are used for linear regression for radiation.

$$f(x) = \langle w, x \rangle + b, x \in [0, 255]^5 \quad (10)$$

4) **SVR using Linear kernel($SVR - Li$)** This model use ϵ-SVR method with linear kernel.

$$f(x) = \langle w, x \rangle + b, w \in R^5, x \in [0, 255]^5 \quad (11)$$

5) **SVR using Radial Basis Function kernel($SVR{-}RBF$)** The RBF funtion is used for kernel mapping in SVR before linear optimization.

$$f(x) = \langle w, \phi(x) \rangle + b, w \in R^5, x \in [0, 255]^5 \quad (12)$$

6) **SVR using Linear kernel with radiation ($SVR{-}Li_{rad}$)** Different from SVR-Li, this model uses previous radiation as another input feature. A 0 1 threshold of is applied to forecast output to generate normalized prediction.

$$f_t(x) = \langle w, x_t \rangle + b, w \in R^5, \\ x_t = \{x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t6}, y_{t-\Delta t}\} \quad (13)$$

where x_{tn} is input of channel n at time t,y is ground truth radiation, Δt is the time span.

7) **SVR-Radial Basis Function kernel with radiation ($SVR - RBF_{rad}$)** Same as $SVR - Li_{rad}$ but with RBF kernel as non-linear mapping to linear space.

$$f_t(x) = \langle w, \phi(x_t) \rangle + b, w \in R^5, \\ x_t = \{x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t6}, y_{t-\Delta t}\} \quad (14)$$

VI. Experiment Results

A. Dataset

The time range of our satellite dataset is from April 1st 2012 to November 1st 2012, covering partial Spring ,Summer, and most autumn. All the raw satellite multi-channel data is collected from public FTP server from GOES project[]. The radiation data is from pyranometer on-site measurement. As satellite data from GOES has routine scan(30 min) and rapid scan(15 min), Span between two images is not fixed and motion vector extraction also takes this into account for later forecasting. For both satellite images and pyranometer measurement, raw input has been preprocessed by removing data points which have 1) bad-frame of multi-channels 2) failure of ground radiation sensor 3) multi-channel timestamp mismatch 4) low solar angles. In total, training dataset is of

8477 frames for each satellite channel. The timestamps of data records are continous in each individual day but may not cover all days for each month in range. All radiation data are normalized and integrated over 15 minutes around frame time of satellite data. The detailed information of dataset is listed in Table I.

TABLE I. SATELLITE MODEL DATASET

Dimension	size	temporal resolution	spatial resolution
Channel 1	8477	15min,30min	1km x 1km
Channel 2,3,4,6	8477	15min,30min	4km x 4km
radiation	8477	15min	NA

B. Estimation Evaluation

For whole satellite model dataset, we used 5-fold cross validation for 7 models listed in section V. With the purpose of performance evaluation of models, motion tracking and cloud prediction are skipped in case of bringing forecast errors. In order to measure the deviation of model output from ground truth radiation value, Mean Square Error(MSE) and Mean Absolute Error(MAE) are used to evaluate accurary. They are defined as:

$$MAE = \frac{i}{N}\sum_{i}^{N}|f(x_i) - y_i| \tag{15}$$

$$MSE = \frac{i}{N}\sum_{i}^{N}(f(x_i) - y_i)^2 \tag{16}$$

Where $f(x_i)$ stands for model estimation radiation using i_{th} data record, yi is the actual value in accordance with i_{th}timestamp, N is the size of dataset, specifically, $N = 8477$ in experiment. Moreover, as radiation is normalized as a fraction of clear sky value, the fluctuation range is fixed in 0 1. MAE and MSE value only indicate the percentage deviation of estimation from reality. The final results are shown in figure VI-B and figure VI-B.

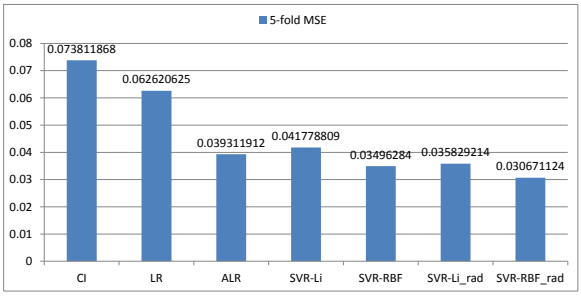


Fig. 4. Averge MSE Plot of 5-fold cross validation

C. Forecast Result

As the key procedure that influence accuray of forecasting, motion tracking and image prediction are taken into evalutaion. Motion vectors extracted from motion estimation methodology is used for image prediction. With targeting the hours's forecast to meet mid-term requirement, the experiment is designed to compare prediction output at 30 mintues,1 hour, 2 hour ,3 hour, 4 hour and 5 hour.

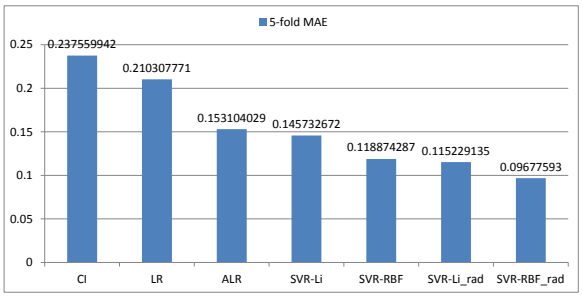
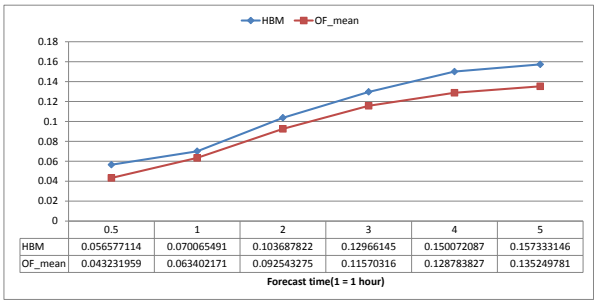


Fig. 5. Averge MAE Plot of 5-fold cross validation

Besides the purpose of comparison among models, forecasting output is also used for evaluation of motion estimation algorithms. As DFD value is a measurement of average error per image, MSE score of local scale(individual pixel) using SVR model is generated for performance comparison of motion estimation with time. The result is presented in Figure VI-C. By using Optical Flow as motion estimation algorithm, the MSE and MAE score of 7 models is shown in Figure VI-C and Figure VI-C.



	0.5	1	2	3	4	5
HBM	0.056577114	0.070065491	0.103687822	0.12966145	0.150072087	0.157333146
OF_mean	0.043231959	0.063402171	0.092543275	0.11570316	0.128783827	0.135249781

Fig. 6. MSE score of Motion Estimation with $SVR{-}RBF_{rad}Forecasting$

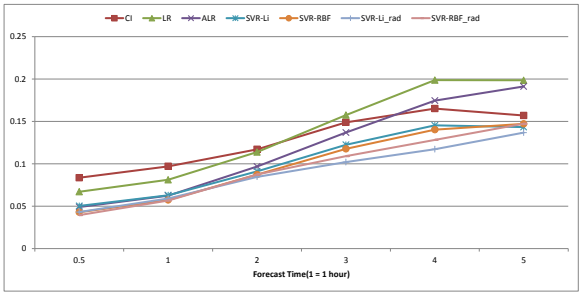


Fig. 7. MSE score of 7 forecasting models

VII. CONCLUSION

In this paper, a new mid-term forecast system is developed with innovations on both modeling and predicting aspect. To choose and compare satellite models, we implement 7 methods with both linear and non-linear concerns. From experiment, we conclude that non-linear SVR model combined with radiation feature is the best model which has more than 50% than the baseline CI method. In cloud motion estimation, Optical Flow with mean filter method is used as it is better than traditional

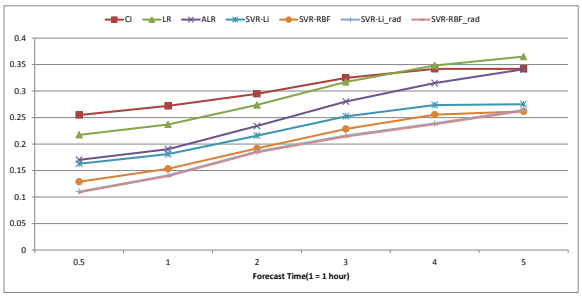


Fig. 8. MAE score of 7 forecasting models

block matching method using both image processing criteria and forecasting MSE score. In 0.5 5 hours forecasting, we find that SVR related models significantly improve the accuracy than others. Although noise from cloud prediction increase exponentially with time, the SVR non-linear model is more robust than the baseline,especially combined with radiation at previous timestamp. The accuracy improvement of is more than 50% in 30 minutes prediction and 10% in 5 hours prediction.

REFERENCES

[1] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F. Vignola, "A new operational model for satellite-derived irradiances: description and validation," *Solar Energy*, vol. 73, no. 5, pp. 307–317, 2002.

[2] A. Zelenka, R. Perez, R. Seals, and D. Renné, "Effective accuracy of satellite-derived hourly irradiances," *Theoretical and Applied Climatology*, vol. 62, no. 3-4, pp. 199–207, 1999.

[3] D. Cano, J.-M. Monget, M. Albuisson, H. Guillard, N. Regas, and L. Wald, "A method for the determination of the global solar radiation from meteorological satellite data," *Solar Energy*, vol. 37, no. 1, pp. 31–39, 1986.

[4] R. Stuhlmann, M. Rieland, E. Raschke, *et al.*, "An improvement of the igmk model to derive total and diffuse solar radiation at the surface from satellite data." *Journal of Applied Meteorology*, vol. 29, no. 7, pp. 586–603, 1990.

[5] J. Schmetz, "Towards a surface radiation climatology: Retrieval of downward irradiances from satellites," *Atmospheric Research*, vol. 23, no. 3, pp. 287–321, 1989.

[6] P. Ineichen and R. Perez, "Derivation of cloud index from geostationary satellites and application to the production of solar irradiance and daylight illuminance data," *Theoretical and Applied Climatology*, vol. 64, no. 1-2, pp. 119–130, 1999.

[7] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller, and H. G. Beyer, "Solar energy assessment using remote sensing technologies," *Remote Sensing of Environment*, vol. 86, no. 3, pp. 423–432, 2003.

[8] S. Janjai, J. Laksanaboonsong, M. Nunez, and A. Thongsathitya, "Development of a method for generating operational solar radiation maps from satellite data for a tropical environment," *Solar Energy*, vol. 78, no. 6, pp. 739–751, 2005.

[9] F. Martins, E. Pereira, and S. Abreu, "Satellite-derived solar resource maps for brazil under swera project," *Solar Energy*, vol. 81, no. 4, pp. 517–528, 2007.

[10] R. Perez, P. Ineichen, M. Kmiecik, K. Moore, D. Renne, and R. George, "Producing satellite-derived irradiances in complex arid terrain," *Solar Energy*, vol. 77, no. 4, pp. 367–371, 2004.

[11] L. F. Zarzalejo, J. Polo, L. Martín, L. Ramírez, and B. Espinar, "A new statistical approach for deriving global solar radiation from satellite images," *Solar Energy*, vol. 83, no. 4, pp. 480–484, 2009.

[12] L. F. Zarzalejo, L. Ramirez, and J. Polo, "Artificial intelligence techniques applied to hourly global irradiance estimation from satellite-derived cloud index," *Energy*, vol. 30, no. 9, pp. 1685–1697, 2005.

[13] O. Şenkal and T. Kuleli, "Estimation of solar radiation over turkey using artificial neural network and satellite data," *Applied Energy*, vol. 86, no. 7, pp. 1222–1228, 2009.

[14] O. Şenkal, "Modeling of solar radiation using remote sensing and artificial neural network in turkey," *Energy*, vol. 35, no. 12, pp. 4795–4801, 2010.

[15] E. Ricciardelli, F. Romano, and V. Cuomo, "Physical and statistical approaches for cloud identification using meteosat second generation-spinning enhanced visible and infrared imager data," *Remote Sensing of Environment*, vol. 112, no. 6, pp. 2741–2760, 2008.

[16] R. Perez, S. Kivalov, A. Zelenka, J. Schlemmer, and K. Hemker Jr, "Improving the performance of satellite-to-irradiance models using the satellites infrared sensors," in *Proc. of American Solar Energy Societys Annual Conference, Phoenix, AZ*, 2010.

[17] J. A. Leese, C. S. Novak, and V. Ray Taylor, "The determination of cloud pattern motions from geosynchronous satellite image data," *Pattern Recognition*, vol. 2, no. 4, pp. 279–292, 1970.

[18] D. Heinemann, E. Lorenz, and M. Girodo, "Forecasting of solar radiation," *) Solar energy resource management for electricity generation from local level to global scale. Nova Science Publishers, New York*, 2006.

[19] U. Görsdorf, A. Seifert, V. Lehmann, and M. Köhler, "Cloud statistics and nwp-model validation based on long term measurements of a 35 ghz radar," in *35th Conference on Radar Meteorology*, 2011, pp. 26–30.

[20] D. Yang, P. Jirutitijaroen, and W. M. Walsh, "Hourly solar irradiance time series forecasting using cloud cover index," *Solar Energy*, 2012.

[21] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[22] B. D. Lucas, T. Kanade, *et al.*, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th international joint conference on Artificial intelligence*, 1981.