Peer Reviewed

Title:
Cloud Characterization for Solar Forecasting Using Sky Imagery

Author:
Chow, Chi Wai

Acceptance Date:
2015

Series:
UC San Diego Electronic Theses and Dissertations

Degree:
Ph. D., UC San Diego

Permalink:
http://escholarship.org/uc/item/78j5h1bs

Local Identifier(s):

Abstract:
Methods for cloud characterization for solar forecasting using sky imagery are presented. Sky images taken every 30 seconds were processed to determine cloud cover using a clear sky library (CSL) method. The CSL method utilizes spectral color information from single-frame sky image and detects clouds reliably outside the circumsolar region. A sky image sequence segmentation (SISS) method combing multi-frame sky images, contour detection, and random walk image segmentation is developed and shown to outperform the CSL method. Cloud motion vectors were generated by cross-correlating two consecutive sky images and variational optical flow (VOF). Cloud locations were forecasted using a frozen cloud advection method at 30 second intervals up to a forecast horizon of 15 min. A month of image data was analyzed to compare the accuracy of VOF forecast with cross-correlation method (CCM) and image persistence method. The VOF forecast with a fixed smoothness parameter was found to be superior to image persistence forecast for all forecast horizons for almost all days and outperform CCM forecast with an average error reduction of 39%, 21%, 19%, and 19% for 0, 5, 10, and 15 min forecasts respectively. Optimum forecasts may be achieved with forecast-horizon-dependent smoothness parameters. Cloud stability and forecast confidence was evaluated by correlating point trajectories with forecast error. Point trajectories were obtained by tracking sub- sampled pixels using optical flow field. Point trajectory length in minutes was shown to increase with decreasing forecast error and provide valuable information for cloud forecast confidence at forecast issue time. A technique for characterizing and predicting hours-ahead solar forecast error using cloud patterns as predictors is also presented. The dependence of forecast root-mean-square error (RMSE) on different cloud patterns is analyzed using multivariate linear regression and analog model. Image entropy or cloud randomness is shown to contribute the most to the forecast RMSE

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Cloud Characterization for Solar Forecasting Using Sky Imagery

A dissertation submitted in partial satisfaction of the requirements for the

degree of Doctor of Philosophy

in

Engineering Sciences (Mechanical Engineering)

by

Chi Wai Chow

Committee in charge:

       Professor Jan Kleissl, Chair
       Professor Ery Arias-Castro
       Professor Serge Belongie
       Professor Raymond De Callafon
       Professor Carlos Coimbra

2015

The Dissertation of Chi Wai Chow is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California, San Diego

2015

DEDICATION

This work is dedicated to my family, my Lord, and Kam Kwong World Mission

Ministries

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Publication: Chow, C. W., Belongie, S., & Kleissl, J. (2015a) Sky image sequence segmentation for solar forecasting. In preparation.

Co-authors: Serge Belongie and Jan Kleissl

Publication: Chow, C. W., Belongie, S., & Kleissl, J. (2015b). Cloud motion and stability estimation for intra-hour solar forecasting. Solar Energy, 115, 645-655.

Co-authors: Serge Belongie and Jan Kleissl

Publication: Urquhart, B., Ghonima, M., Nguyen, D., Kurtz, B., Chow, C. W., Kleissl, J. (2013). Sky-Imaging Systems for Short-Term Forecasting. In J. Kleissl, Solar Energy Forecasting and Resource Assessment (pp. 195-232). Academic Press.

Co-authors: Bryan Urquhart, Mohamed Ghonima, Andu Nguyen, Ben Kurtz, and Jan Kleissl

Publication: Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C. W., Ghonima, M., & Kleissl, J. (2014). Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. Solar Energy, 103, 502-524.

Co-authors: Handa Yang, Ben Kurtz, Andu Nguyen, Bryan Urquhart, Mohamed Ghonima, and Jan Kleissl

VITA

2009   Bachelor of Science, University of California, San Diego

2012   Master of Science, University of California, San Diego

2015   Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Chow, C. W., Belongie, S., & Kleissl, J. (2015) Sky image sequence segmentation for solar forecasting. In preparation.

Chow, C. W., Belongie, S., & Kleissl, J. (2015). Cloud motion and stability estimation for intra-hour solar forecasting. *Solar Energy*, *115*, 645-655.

Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., **Chow, C. W.**, Ghonima, M., & Kleissl, J. (2014). Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. Solar Energy, 103, 502-524.

Gohari, M. I., Urquhart, B., Yang, H., Kurtz, B., Nguyen, D., **Chow, C. W.**, ... & Kleissl, J. (2014). Comparison of solar power output forecasting performance of the total sky imager and the University of California, San Diego sky imager.Energy Procedia, 49, 2340-2350.

Urquhart, B., Ghonima, M., Nguyen, D., Kurtz, B., **Chow, C. W.**, Kleissl, J. (2013). Sky-Imaging Systems for Short-Term Forecasting. In J. Kleissl, Solar Energy Forecasting and Resource Assessment (pp. 195-232). Academic Press.

Ghonima, M. S., Urquhart, B., **Chow, C. W.**, Shields, J. E., Cazorla, A., Kleissl, J., 2012. A method for cloud detection and opacity classification based on ground based sky imagery. Atmospheric Measurement Techniques 5, 2881–2892.

Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., & Washom, B. (2011). Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy, 85(11), 2881-2893.

ABSTRACT OF THE DISSERTATION

Cloud Characterization for Solar Forecasting Using Sky Imagery

by

Chi Wai Chow

Doctor of Philosophy in Engineering Sciences (Mechanical Engineering)

University of California, San Diego, 2015

Professor Jan Kleissl, Chair

Methods for cloud characterization for solar forecasting using sky imagery are presented. Sky images taken every 30 seconds were processed to determine cloud cover using a clear sky library (CSL) method. The CSL method utilizes spectral color information from single-frame sky image and detects clouds reliably outside the circumsolar region. A sky image sequence segmentation (SISS) method combing multi-

frame sky images, contour detection, and random walk image segmentation is developed and shown to outperform the CSL method.

Cloud motion vectors were generated by cross-correlating two consecutive sky images and variational optical flow (VOF). Cloud locations were forecasted using a frozen cloud advection method at 30 second intervals up to a forecast horizon of 15 min. A month of image data was analyzed to compare the accuracy of VOF forecast with cross-correlation method (CCM) and image persistence method. The VOF forecast with a fixed smoothness parameter was found to be superior to image persistence forecast for all forecast horizons for almost all days and outperform CCM forecast with an average error reduction of 39%, 21%, 19%, and 19% for 0, 5, 10, and 15 min forecasts respectively. Optimum forecasts may be achieved with forecast-horizon-dependent smoothness parameters

Cloud stability and forecast confidence was evaluated by correlating point trajectories with forecast error. Point trajectories were obtained by tracking sub-sampled pixels using optical flow field. Point trajectory length in minutes was shown to increase with decreasing forecast error and provide valuable information for cloud forecast confidence at forecast issue time.

A technique for characterizing and predicting hours-ahead solar forecast error using cloud patterns as predictors is also presented. The dependence of forecast root-mean-square error (RMSE) on different cloud patterns is analyzed using multivariate linear regression and analog model. Image entropy or cloud randomness is shown to contribute the most to the forecast RMSE.

1 Introduction

Clouds and aerosol continue to contribute to the largest uncertainty to estimates of the Earth's radiation budget (ERB, Boucher et al., 2013). The quantification of clouds and convective effects in climate models; and modeling of aerosol-cloud interactions, continues to be a challenge. Global climate is very sensitive to changes in clouds and radiation. The horizontal extent of clouds, cloud water path, and cloud particle size, phase, and shape affect the ERB, especially shortwave cloud radiative forcing (SWCRF). For example, a 15-20% relative increase in low-level cloud amount could offset the radiative forcing caused by doubling the CO_2 amount (Slingo 1990). Satellite-based remote sensing remains to be the primary source of global cloud data but issues persist regarding the accuracy of retrieving cloud and aerosol properties (Boucher et al., 2013). Norris (2009) suggested changes in trade cumulus may be more important than changes in marine stratocumulus in global climate models (GCMs) and yet measuring small changes in cloudiness has insufficient precision. Therefore, it is crucial to establish high precision cloud monitoring systems around the globe including ground-based sky imager and development of automatic cloud detection algorithm to monitor clouds.

The ability to detect and characterizes clouds is also vital for solar resource assessment and forecasting. The advent of solar power is met with challenges unique to the solar resource. Namely, variability in ground irradiance makes regulating and maintaining power both challenging and costly, as the uncertainty of solar generation compared to more conventional fossil power sources requires larger regulation and reserve capacities to meet ancillary service requirements. Of particular interest to the energy industry are sudden and widespread changes in irradiance, termed "ramp events,"

typically caused by large clouds or widespread changes in cloud cover (Pfister et al., 2003). Transient clouds cause strong spatio-temporal variability and fluctuating solar power feed-into the grid. Large ramp events are of primary concern for relatively small microgrids and island grids, as their ability to absorb the fluctuations is limited. For example, a 10s solar ramp due to a passing cloud caused up to 20% of power output change and a severe ramp event caused a 80% change in 60s at a 13.2 MW PV plant in Nevada (Mills, 2010). While distributed PV causes less variability (Hoff and Perez, 2012) to the grid in aggregate, it is less controllable by grid operators as it often lacks the ability for power curtailment (Eber and Corbus 2013). The resulting imbalance motivates the need for regulation reserve that scale with both variability and forecast uncertainty (Helman et al., 2010). Ramp events require ancillary services to ramp up or down to meet the change in electrical supply and maintain power quality. Different strategies have been studied to mitigate the operational problems (Eber and Corbus 2013, Ela et al., 2013) and a simulation study by Ela et al. (2013) demonstrated that an increased power dispatch frequency and accurate short-term solar forecasts can reduce regulating reserve requirements and production costs. Therefore, reliable forecast information on the expected power production is essential for efficient integration.

Currently, solar radiation forecasting is tackled by two main approaches: physics-based numerical weather prediction (NWP), and forecasting that is more directly based on real-time measurements from satellites or ground-based instruments, although the boundaries between these approaches are becoming more fluid, and machine learning techniques often synthesize the various approaches Paoli et al. (2010); Chen et al. (2011); Marquez et al. (2013); Huang et al. (2013). Perez et al. (2010) and Lorenz

et al. (2007) found that NWP forecasts generally perform better than satellite forecasts at forecast horizons longer than a few hours. The spatial resoluton of NWP is coarse at about 100 km^2, but there is active research on high-resolution rapid refresh models with grid cell areas of less than 10km^2 (Benjamin et al. 2010, Lara-Fanego et al. 2011). Either way, most clouds will remain unresolved in NWP. The limited resolution and uncertainty in initial conditions makes NWP models presently unable to predict the exact position of clouds and their effect on solar radiation over a specific location. Consequently, there is a need for the development of better short-term measurement-based forecasts at high spatial and temporal resolutions.

Deterministic measurement-based forecasting is typically based on satellite (Marquez et al., 2013)(Perez et al., 2002) or ground-based sky images. Sky imager approaches have only been applied recently for 15 min DNI (Marquez and Coimbra, 2013), 15 min GHI (Yang et al., 2014), and AC power forecasting for 48 MW of photovoltaics using the Total Sky Imager (TSI, Yankee Environmental Systems)(Urquhart et al., 2013).

Sky imagers provide a visual measurement of a whole sky dome with high temporal and spatial resolution. Spatial resolution is determined by the camera's ability to resolve a single picture element (pixel), the solid angle subtended by this pixel, and the distance of the cloud from the imager which is a function of its cloud base height (CBH) and zenith angle (ZA). Temporal resolution is dictated primarily by the operational ability of the image processing algorithm. Atmospheric properties such as aerosol optical depth, cloud optical depth, cloud cover, cloud type, and CBH have been retrieved using sky imagers developed by research groups (Cazorla et al. 2008, 2009, Heinle et al. 2010,

Kassianov et al. 2005, Long et al. 2006, Huo and Lu 2009, 2010, Kalisch and Macke 2008, Pfister et al. 2003, Seiz et al. 2007, Shields et al. 2007, 2009).

The goal in this dissertation is to demonstrate how clouds in sky imagery can be characterized and how that information can improve solar forecasting. In Chapter 2, the experimental setup and specifications of two sky imaging systems are presented. An overview of solar forecasting algorithm developed at University of California, San Diego (UCSD) is described in Chapter 3. Among the solar forecast procedure, only the image-based forecast is focused in this dissertation. Chapter 4 provides a method to detect clouds in sky imagery and special attention is paid to the cloud detection accuracy in proximity to the sun. Since most solar variability and forecast models require cloud velocity as main input, accurate cloud motion estimation has become of increased interest. Chapter 5 presents a method to obtain cloud motion vectors using cross-correlation and variational optical flow and the accuracy of the methods are assessed. Using cloud detection and cloud motion a forecast algorithm is designed and validated. Forecast confidence is related to cloud stability through point trajectories that are constructed by tracking pixel point. Finally, a method to characterize and predict hours-ahead solar forecast error using cloud patterns from satellite images is presented in Chapter 6 and conclusions follow in Chapter 7.

2 Experimental Setup

2.1 Total Sky Imager (TSI)

A rooftop mounted TSI 440A total sky imager (TSI, Yankee Environmental Systems, setup at 32.8852°N, 117.2400°W, 124 m MSL) has taken images of the sky over UCSD since August 2009 (Figure 2.2). The instrument consists of a spherical mirror and a downward pointing camera. Images are taken every 30 seconds when the sun is above an elevation angle of three degrees. The camera provides images that are 640 by 480 pixels and the mirror occupies 420 by 420 pixels.
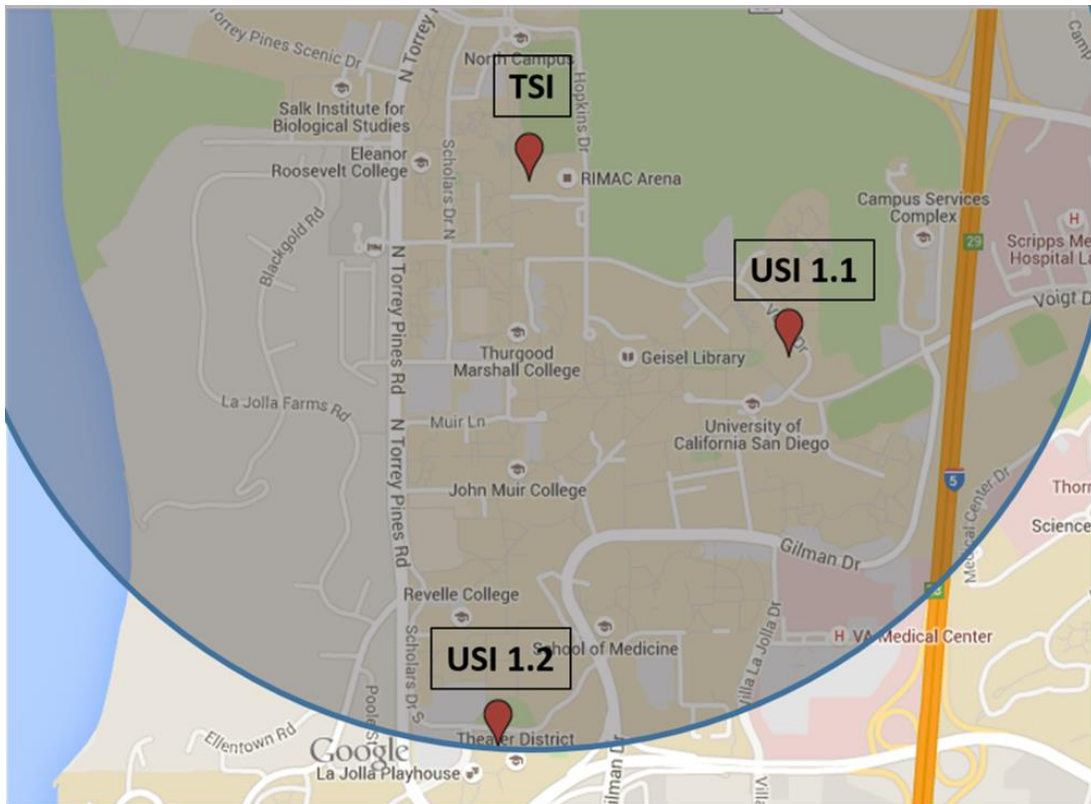
Figure 2.1: Locations of TSI and USI installation. The coverage area of the sky imager is a function of cloud base height. The region displayed here has a radius of approximately 2 km which would apply to a cloud base height of 1150 m at a maximum zenith angle of 65°. Map data 2015 Google.

Figure 2.2: TSI 440A Total Sky Imager.

The image output settings by the TSI data acquisition software are default and cannot be changed. Images are 24-bit compressed jpgs. The compression process induces a small loss of information in the image. Additionally the system has an automatic gain adjustment to provide a larger dynamic range for the 8 bits on each of the red, green and blue channels. This adjustment process causes the intensity histogram of each channel to redistribute and thus the relative spectral composition of the images to fluctuate slightly

with incoming signal strength. To determine sky cover obstructed by the camera arm and the shadowband, image masks are generated, edges of the masks are identified, and pixel values of the edge region are used to interpolate the region within. While all the methods described in this dissertation can be applied to the TSI images, the TSI is replaced by a sky imager described in the next section in 2012 and the results from the TSI are presented only in Section 4.2.3 and 5.2.2.1.

2.2 UCSD Sky Imager (USI)

In cooperation with Sanyo Electric Co. (now Panasonic), LTD., Smart Energy Systems Division, the University of California, San Diego designed and developed a sky imager system specifically for short-term solar forecasting applications. The USI captures images using an upward-facing charge-coupled device (CCD) image sensor sensing RGB channels at 12 bit intensity resolution per channel. A 4.5 mm focal length circular fisheye lens allows imaging of the entire sky hemisphere. Although the CCD image sensor is capable of capturing images at a resolution of 2048 x 2048 pixels, the fisheye lens limits the image circle diameter to 1748 pixels. The USI was also equipped with a neutral density filter with an optical depth of 6.9. Utilizing high dynamic range (HDR) imaging, the USI outputs images at 16 bits per channel with a dynamic range of over 80 dB. Lossless PNG compression is used to store and transmit images for forecast analysis. Complete specifications of the USI system can be found in Urquhart et al. (2013, 2014).

Since cloud cover near the sun provides vital information for short-term solar forecasting, the USI does not employ a solar occulting device. Although improved sky visibility is gained, the high intensity of the direct solar beam causes CCD blooming,

wherein excess electrons from saturated pixels near the solar disk overflow into neighboring pixels and cause a vertical stripe (or "blooming stripe") to appear over the sun's location in the image (Urquhart et al., 2013). If present, the blooming stripe is removed during the USI forecast procedure. We have not noticed any evidence of significant sensor degradation caused by direct exposure of our CCD to the sun for over 2 years of continuous operation. The increased resolution and dynamic range, combined with the ability to image the entire sky hemisphere, has allowed the USI to overcome the primary shortcomings of the TSI.

The images used in this dissertation were taken by a rooftop-mounted USI 1.2 (Figure 2.1). The instrument is set to capture images every 30 seconds during times when the sun is above an elevation of -3°, but only images corresponding to solar zenith angles (SZAs) < 80° were processed for forecasting purposes.

Acknowledgements

Publication: Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C. W., Ghonima, M., & Kleissl, J. (2014). Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. Solar Energy, 103, 502-524.

Co-authors: Handa Yang, Ben Kurtz, Andu Nguyen, Bryan Urquhart, Mohamed Ghonima, and Jan Kleissl

3 Solar Forecast Procedure

Prior techniques for estimating global horizontal irradiance (GHI) over an extended area consisted of satellite-derived cloud cover coupled with an empirical clear sky model (Cano et al. 1986, Schmetz 1989, Zelenka et al. 1999, Perez et al. 2002, Rigollier et al. 2004, Pereira et al. 2000, Martins et al. 2007). On the other hand, ground based instruments are typically point sensors and do not provide spatial irradiance information.

The primary factor modulating GHI from its clear sky value is the presence of a cloud between the ground location and the sun as this directly attenuates the solar beam irradiance. Variations in diffuse irradiance caused by changing cloud distribution and optical depth are smaller and generally negatively correlated with beam irradiance (a possible exception is cloud enhancement effects, which are beyond the scope of this dissertation, Luoma et al. 2012).

Solar forecasts using ground-based sky imagery can be broken into two steps: one that operates purely upon sky images, specifically cloud detection and cloud motion, and one which is specific to the location and equipment at the site of interest. A power or irradiance forecast for a specific pixel on the ground requires executing both sections, but the image-based forecast performance yields insights into the validity of some fundamental assumptions in the forecast process. The forecast procedure is similar for the TSI and USI and improvements are constantly building on this approach. In this chapter, the procedure used to generate USI forecasts for the UCSD DEMROES network is focused and outlined in the flow chart in Figure 3.1.
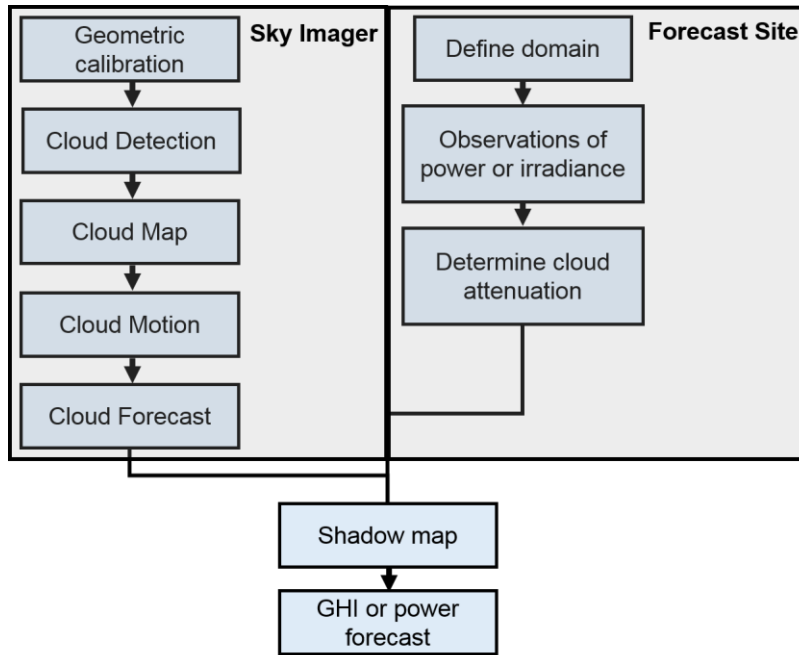
Figure 3.1: Flowchart of solar forecast procedure.

Before image processing, each USI image pixel was calibrated to a geographic azimuth and zenith angle by leveraging the known position of the sun and the known (equisolid angle) projection function of the lens. Geographic azimuth and zenith coordinates were measured from images taken on a clear day by recording the calculated solar position for all saturated pixels (assumed to be within the solar disk). All recorded solar positions for a given pixel throughout the day were averaged to produce a single estimate of solar azimuth and zenith angles for each saturated pixel. Then, the image center (the zenith), the precise focal length, and the orientation of the lens w.r.t. geographic coordinates were obtained from a nonlinear fit which minimizes the Euclidean distance between the measured and modeled directions projected onto a unit sphere. Once the geographic azimuth and zenith angles are known, the angle between the vector to the sun and the direction vector for a given image pixel, or "sun-pixel angle,"

describes the distance of an image pixel from the sun. A comprehensive geometric and camera calibration procedure can be found in Urquhart et al. (2014).

Following the geometric calibration, clouds are detected (Section 4) and cloud base height (CBH) is obtained from METAR weather reports, which are typically generated once per hour (sometimes more frequently) from automated weather observation stations at airports. Stereographic method to determine CBH is also available when there are two or more sky imagers (Nguyen and Kleissl, 2014). The binary cloud/no-cloud information is still in the original image coordinates, but what is needed is georeferenced cloud mapping. A geometric transform similar to the pseudo-Cartesian transform of Allmen and Kegelmeyer (1996) was then performed to map cloud information to a latitude-longitude grid at the selected CBH. The resulting "cloud map" is a two-dimensional planar mapping of cloud position at the obtained CBH above the forecast site, centered at the location of the USI.

Cloud velocity (Section 5) is then used to advect the planar cloudmap to generate a cloud-position forecast out to 15 min forcast lead time for each forecast interval (30 s). The forecast domain was defined as a 2.6 x 3.6 km grid at a resolution of 2.5 m per grid point. Each grid point was mapped to a latitude and longitude, as well as an elevation obtained from the SRTM1 digital elevation model (Farr et al., 2007). The forecast domain contains the georeferenced placement and coverage of ground sites (pyranometers and photovoltaic plants) within the domain, hereby denoted "footprint". The UCSD site consisted of 6 point sensors which were assumed to represent the footprint pixel in which they were located. For each forecast grid point, a ray is traced along the vector to the Sun and the intersection with the cloudmap is determined (Figure

3.2). If the intersected point is clear, that ground location is deemed clear, whereas if the intersection is cloud, the ground point is deemed shade by cloud. Repeating the shadow-mapping process for each forecast grid point constructs a map of cloud shadows (shadowmap), which provides the percentage of the footprint that is shaded. The method to generate power or irradiance output is site specific, and can be found in Urquhart et al. (2013) and Yang et al. (2014). For the rest of the dissertation, forecast procedure operates upon the sky images will be focused and only image-based performance is evaluated. The results of power and irradiance forecast by the author and his colleagues are presented in Urquhart et al. (2013), Yang et al. (2014), and Gohari et al. (2014).
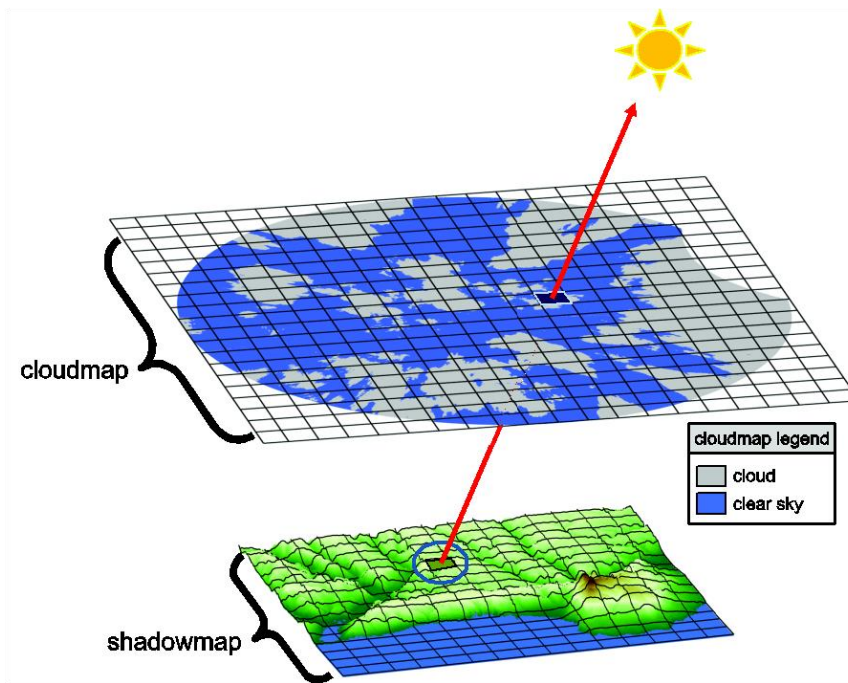
Figure 3.2: Ray tracing to construct a georeferenced mapping of shadows. The shadow value for a given point in the forecast domain grid is determined by tracing a ray along the solar vector and determining the cloud value at the intersection with the cloudmap.

Acknowledgements

In the course of preparing this chapter, the following publications were used:

4 Cloud Detection

4.1 Introduction

The TSI data acquisition software provides a cloud decision output image, however, we developed our own algorithm for solar forecast purpose, and all image processing is done on raw image output. Our cloud detection technique is based on the concept of the ratio of the red channel to the blue channel developed at the Scripps Institution of Oceanography (Johnson et al. 1989, 1991, Shields et al. 2007, 2009). Increased molecular (Rayleigh) scattering of shorter wavelengths causes the clear sky to be blue and the signal measured in the blue channel to be larger. Since clouds scatter the visible wavelengths more evenly, the red signal is similar to the blue signal. The red-blue-ratio (RBR) indicates whether the dominant source of the scattered light incident on a pixel is the clear sky or a cloud.

The RBR in the whole sky hemisphere in clear conditions is not uniform due to aerosol and airmass effects. Due to the size distribution of aerosols relative to the wavelength of incident visible light, scattering by aerosols shows a weaker wavelength dependence than scattering by molecules (McCartney 1976), which results in the scattered light appearing whiter (i.e. the blue is less saturated). This is apparent near the horizon where relative optical airmass increases and more total aerosols occupy the optical path through the atmosphere. In the circumsolar region, the forward scattering effects of aerosols are also apparent as a whiter region around the sun. In general, different instruments will vary in measured RBR distribution throughout an image due to specific measurement hardware characteristics and the level and type of instrument

calibration. Specifically, this may include impacts by such factors as scattering off the front optical surface. With some instruments, a single RBR threshold can be used effectively for opaque clouds, but not with thin clouds.

4.2 Clear Sky Library (CSL)

4.2.1 Two-dimensional CSL and sunshine parameter

With the TSI, we found that a single RBR threshold for the entire image is not appropriate. Rather, a clear sky library (CSL based on Shields et al. 2009) as a function of zenith and sun-pixel-angle was calculated from images on a cloud-free day. A clear sky background image is then generated for each sky image based on the current solar ZA (Figure 4.1a). The CSL threshold is defined to be the RBR in the clear sky background image plus an additional threshold value. A pixel is classified as cloudy if its RBR (Figure 4.1c) is larger than the CSL threshold. In general, the method using the CSL is able to detect white opaque clouds accurately. However, within the circumsolar region, thick dark clouds cannot be identified since they have a lower RBR than the CSL threshold. Therefore, a sunshine parameter (SP, defined by Pfister et al. 2003) is used in addition to the CSL to improve cloud decision in the circumsolar region. SP is computed as the average RBR of the pixel area on the shadowband around the sun position (indicated in Figure 4.1a). SP is a dynamic single-valued RBR threshold that changes from image to image, and is typically small when the sun is obscured. Pixels with RBR > SP are classified as cloudy even if the CSL indicates otherwise. By adding the clouds detected using the CSL (Figure 4.1e) to the clouds detected using SP (Figure 4.1d), the overall cloud decision image is obtained (Figure 4.1f).

17
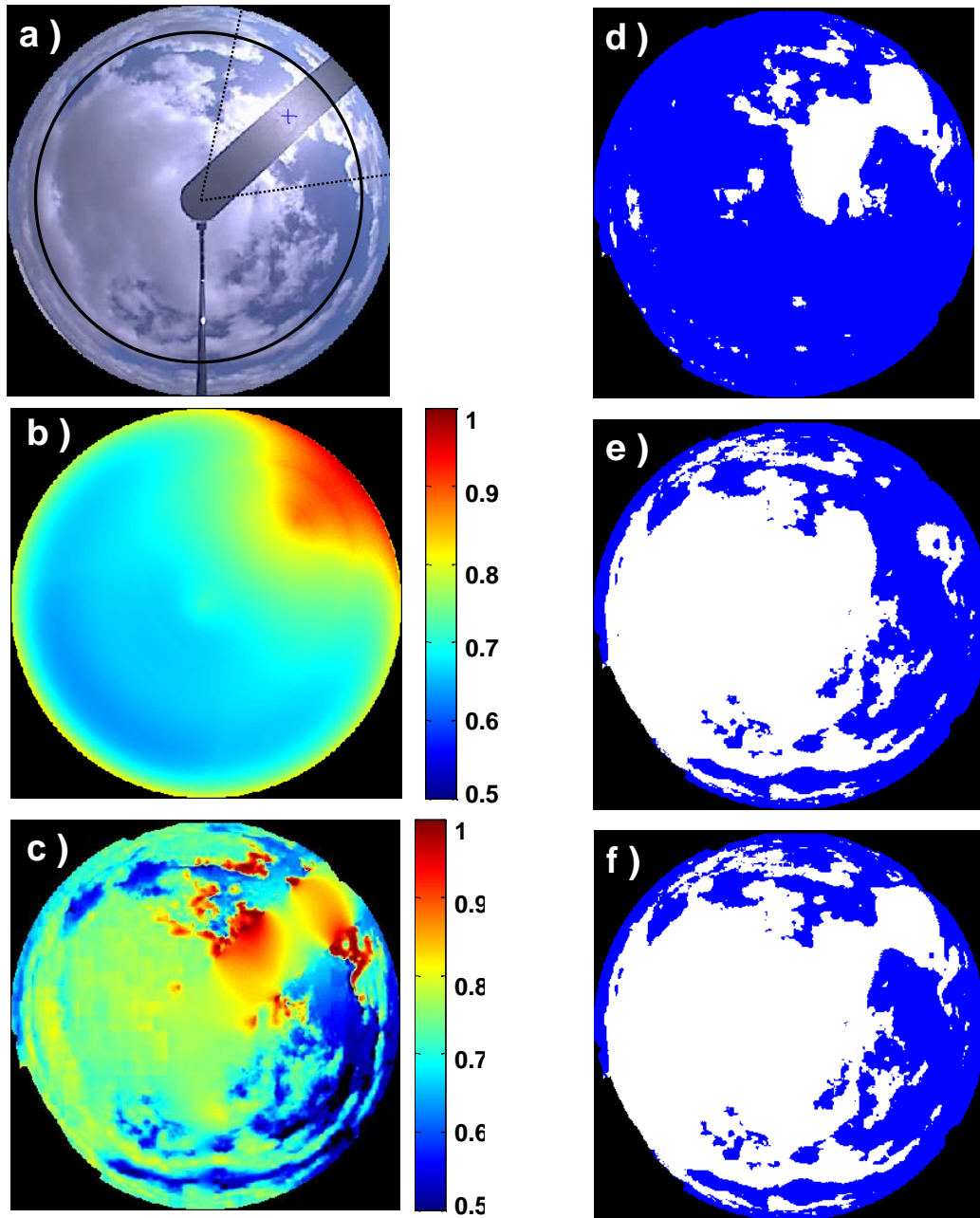
Figure 4.1: Processing chain of a sky image on October 4, 2009 15:45:30 PST (a) to obtain the cloud decision image. The sunshine parameter is 0.85 and is evaluated around the sun position indicated by the blue cross. The dotted black lines show the borders of the circumsolar region defined as solar azimuth ±35° and the solid black line shows ZA at 65°. (b) Clear sky RBR (colorbar) background image plus a threshold (0.19). (c) RBR (colorbar) image. (d) Pixels in (c) with RBR > SP or (e) RBR > CSL threshold are assumed to be cloudy. (f) Shows the final cloud decision image. White areas are clouds and blue areas are clear skies.

The CSL was compiled as a sun-pixel-angle by ZA lookup table from a full day of images on September 24, 2009. In Figure 4.2a, sky imager ZAs larger than 75° correspond to the horizon, and small sun-pixel-angles (approximately < 35°) correspond to the circumsolar region. It can be seen in Figure 4.2a that clear pixels away from the sun and horizon have rather uniform RBR. Decreasing the distance from sun increases RBR because the forward scattering of sunlight turns the circumsolar region whiter and brighter. Near the horizon, the higher aerosol concentration and optical depth (due to the airmass effect) causes increased Mie scattering and thus also acts to increase RBR.
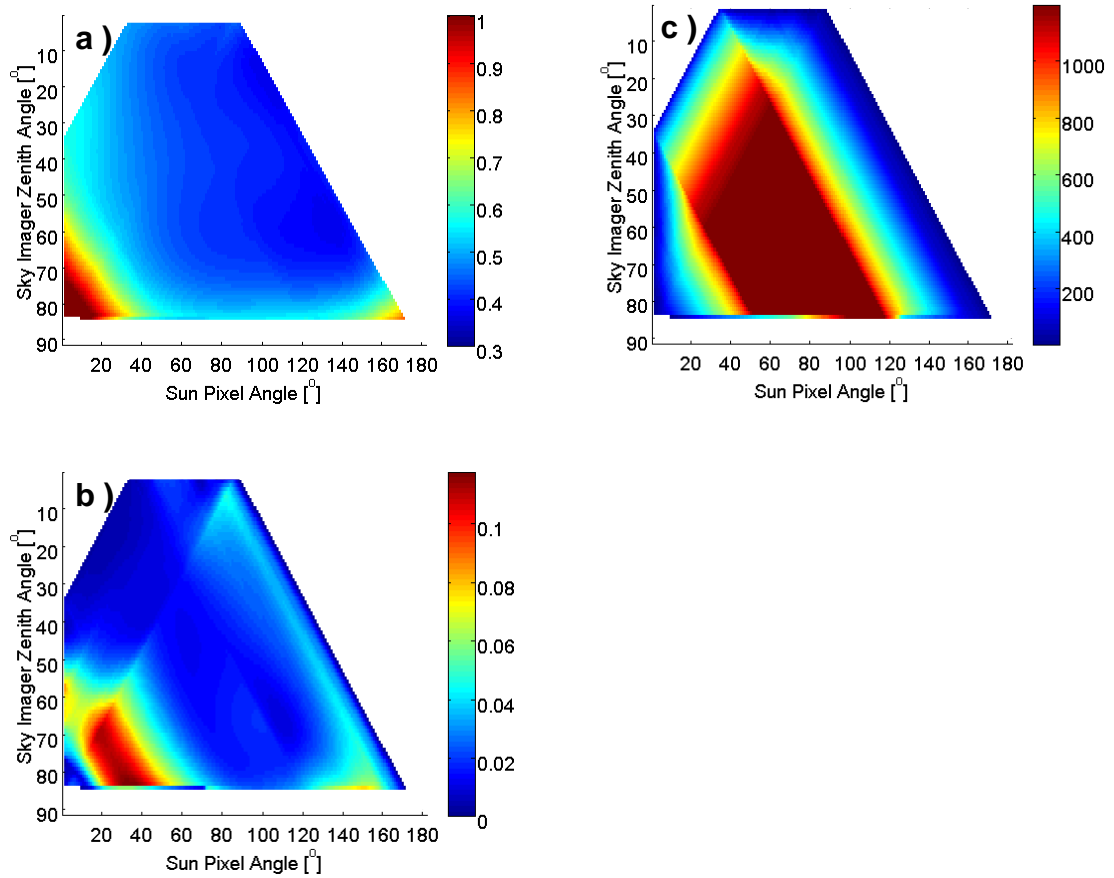
Figure 4.2: (a) Average RBR (colorbar) as a function of zenith angle and sun-pixel-angle based on TSI imagery on a clear day (September 24, 2009). This lookup table is used to generate the clear sky library background image in Figure 4.1b. (b) Standard deviation (colorbar) of clear sky RBR and (c) Number of occurrence (colorbar) for each zenith angle and sun-pixel-angle pair.

Figure 4.2b presents the standard deviation of the clear sky RBR to quantify the RBR variation during the (clear) day in Figure 4.2a. Larger variation of RBR leads to less accurate cloud decisions since it makes it more likely that a clear sky RBR would be larger than the threshold and classified as cloudy. For sun-pixel-angles greater than 50° and sky imager ZA less than 60° (i.e. in the region with the highest occurrence, Figure 4.2c) the variation in the clear sky RBR is small and our method performs well. However, the standard deviation of the RBR in the circumsolar region is found to be largest when

the sun is near the horizon (low sun-pixel-angle and high zenith angle). Anecdotal observations confirm that clear conditions in the circumsolar region at low sun elevation are often misclassified as cloudy while the horizon and circumsolar region appear white. In addition, at larger solar ZAs less sunlight reaches the TSI due to the increased optical path length through the atmosphere. This effect, along with shading by other clouds, can cause clouds to be less bright. This results in a smaller RBR than the CSL threshold leading to incorrect clear classification.

4.2.2 Three-dimensional CSL and thin cloud threshold

The CSL method continues to be used in the USI and the algorithm is further developed to differentiate between thick and thin clouds in various atmospheric conditions (Ghonima et al., 2012). Figure 4.3 shows histograms of the RBR of clear, thick-cloud, and thin-cloud states from a set of 60 manually annotated images captured using the TSI system. Significant overlap of the thin-cloud histogram can be seen with both clear- and thick-cloud histograms, indicating that a single threshold is problematic. Particularly challenging are cases with high concentrations of aerosols or haze. The cause of the difficulty is that large aerosols particles, such as dust or sea salt have nearly uniform scattered intensities within the visible spectrum, which acts to increase the relative red content of a pixel over that observed in clean, clear skies. When the aerosol concentration is high, more of this spectrally uniform scattering occurs, which results in cloud-free pixels having larger than average RBR. To address the problem, an improved three-dimensional CSL, which the RBR of clear-sky pixels was stored as a function of pixel-zenith angle, sun-pixel angle, and solar zenith angle, and a haze correction factor is

applied to the CSL RBR image. Utilizing the haze-corrected difference and a fixed clear-sky threshold value, any pixels not already classified as thick are classified as thin-cloud and clear (Ghonima et al., 2012).
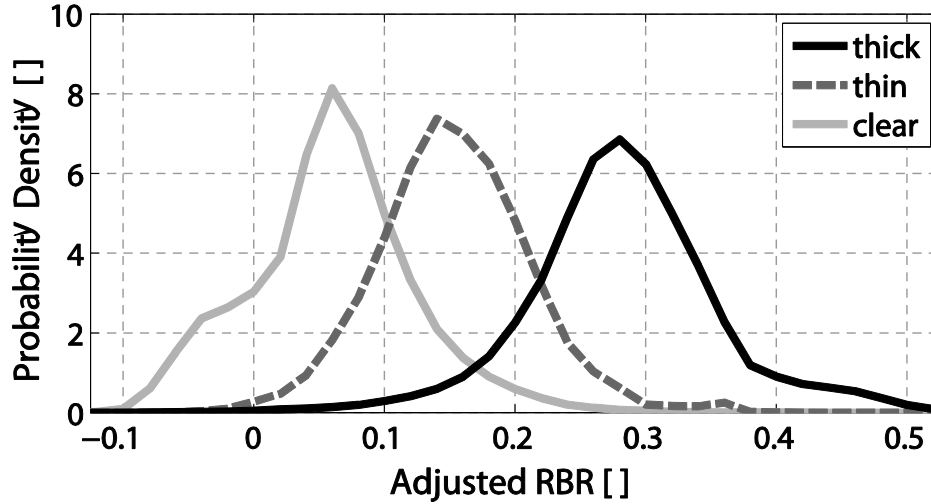
Figure 4.3: Histograms of clear sky, thin cloud, and thick cloud generated from a set of 60 manually annotated images captured by the TSI. These histograms are assessed to select red-blue ratio detection thresholds. Adjusted RBR is RBR - CSL RBR.

Instead of the rigorous procedure performed in Ghonima et al. (2012) to determine thin and thick ΔRBR (let ΔRBR $\equiv$ RBR - RBR$_{clear}$) thresholds, thresholds were visually calibrated by comparing resulting cloud decision images with raw images and their ΔRBR images.

Due to variations in composition of the atmosphere (primarily related to aerosol concentration), clear sky RBR values stored in a given CSL will be less representative of days farther from the day from which the CSL was generated. For this reason, only the nearest CSL will be used by the cloud decision algorithm. In an operational setting, the CSL should be updated on every clear day. In addition, in climates with larger variations in aerosol optical depth, daily corrections to these thresholds are recommended.

Based on our experience, the area near the sun is still prone to cloud decision errors for the USI even with the three-dimensional CSL, as high RBR_{clear} values near the solar disk reduce ΔRBR of clouds and thus their detected thickness. A "CSL bypass" procedure based on the sunshine parameter was used: when the sun was determined to be obstructed (less than 50% saturated pixels in pixels of sun-pixel angle < 1∘), the CSL was not used within regions of sun-pixel angle < 35∘.

4.2.3 Cloud detection with CSL

Since cloud decision images can only be validated visually, we describe here qualitative experiences of applying our method. The combined CSL and SP method works well most of the time for opaque clouds in the outer region of TSI images. However, near the horizon or within the circumsolar region, thick clouds are sometimes misclassified depending on sun obscuration. If the sun is not obscured (not shown), SP is large and gray clouds near the sun are misclassified as clear. When the sun is obscured, gray clouds are correctly classified due to the SP algorithm (Figure 4.4).
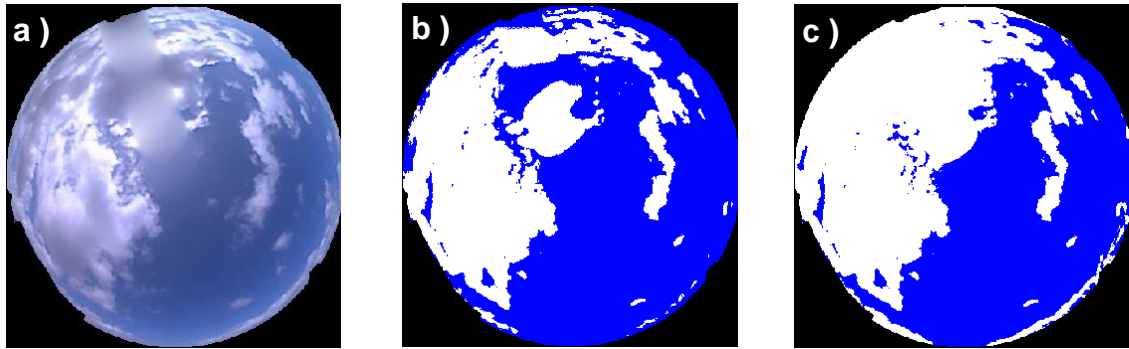
Figure 4.4: Cloud decision comparison between clear sky library (CSL) method (b) and integrated CSL and sunshine parameter method (c) on October 4, 2009 12:43:30 PST.

Optically thin clouds appear bluish because the background clear sky is visible through the cloud and are notoriously difficult to detect automatically (Shields et al.

1998) as their RBR is only slightly larger than the clear sky in our data. If the RBR threshold is lowered to detect thin clouds it will increase false cloud (cloud detected but no cloud exists) detections, especially in the circumsolar region.

Cloud decision errors for the USI are more prevalent within the circumsolar region for the three-dimensional CSL method. Since forward scattering of sunlight by aerosols causes clear sky RBR to increase within the circumsolar region, the clear sky RBR stored in CSLs is elevated within the circumsolar region. Additionally, the intensity of the solar beam causes saturation in pixels immediately within the solar disk, resulting in a clear sky RBR of unity (which is much larger than the clear sky RBR in other regions). These factors tend to reduce the ΔRBR of clouds and therefore the detected optical thickness of clouds within the circumsolar region. Errors due to this effect are difficult to quantify, as thick clouds falsely classified as thin clouds will not impact matching and cap errors (as defined in section 5.2.2), though overprediction of forecast irradiance results if circumsolar thin clouds cast shadows on the site footprint.

Gray clouds near the sun are misclassified as clear and the sunshine parameter and CSL-bypass method is found to be unreliable. Figure 4.7d shows the cloud decision for an approaching cumulus cloud with a gray cloud base using the classical CSL method. In the circumsolar region, the forward scattering effect of aerosols causes high RBR (0.8-1.2) near the sun (star in Figure 4.7) even in clear skies. White clouds tend to have a RBR close to unity since they scatter the visible wavelength evenly which is larger than the clear sky RBR everywhere except over and very close to the solar disk. However, some clouds with large cloud optical depth can appear to be gray. Gray clouds have a smaller RBR than the clear sky background, and lead to false clear sky decision

(black box in Figure 4.7d). Generally, grays clouds are tall water clouds, but even thinner clouds can appear gray under certain conditions. First, the cloud could be in the shadow of other clouds. Second, for large solar zenith angles, the peak of the phase function for forward scattering is directed more horizontally and less radiation is scatter through the cloud bottom. Also, larger water droplets near the bottom of a cloud absorb more light than smaller drops. In general, such thick cumulus clouds create significant ramps in PV power output. For the above reasons, it motivated us to develop a more accurate cloud detection method in the next section.

4.3 Sky Image Sequence Segmentation (SISS)

Before proceed to our proposed method, it is useful to identify a list of important visual features of cumulus cloud and find a way to utilize such information. Figure 4.5 illustrates the importance of visual features from different scales. Clouds are easily identified when looking at the whole image since context and neighboring structures in the image and prior knowledge about the content of the scene is considered. On contrary, pixel points itself shows insufficient to distinguish gray cloud from the clear sky. Since the CSL method developed from the previous sections identifies cloud using only color information at pixel scale, it is easy to understand why it fails. In additional, clouds are observed to be brighter and have high light variations on their lit side and darker, flatter in light variations on their unlit parts. When the sun is behind the cloud, the edges of the cloud are much brighter than the core. It gives a very important visual feature of gray clouds and outlines the contour and edges of the clouds.

To overcome the cloud detection challenges in the previous section, an integrated approach combining multi-frame sky images, contour detection, and random walk image segmentation is presented in this section. Instead of considering cloud decision in terms of pixel points on an image, regions or superpixels are considered to integrate local image structure. A state-of-the-art contour detection method (Arbelaez et al., 2014), which uses multi-scale image features to cluster pixel points, is implemented. Besides static image features, motion information is another feature adequate for cloud detection. Motion cues can be extracted from sky images using two-frame optical flow. Pixels belong to the same cloud move similarly and make them easily detected to human eyes. In a recent paper by Liu et al. (2013), cloud image sequence is modeled by dynamic texture and high accuracy of cloud classification by type is achieved. It illustrates the value of motion cues in learning objects. However, since the sky scene from an image lacks image texture and prior information on the content of the scene is not modeled, the clear sky region appears to move similarly with the cloud and makes motion information from the optical flow alone insufficient for cloud segmentation. Lastly, to combine the CSL method, motion and contour information, a random walk image segmentation method is used to compute the potential of cloud labels.
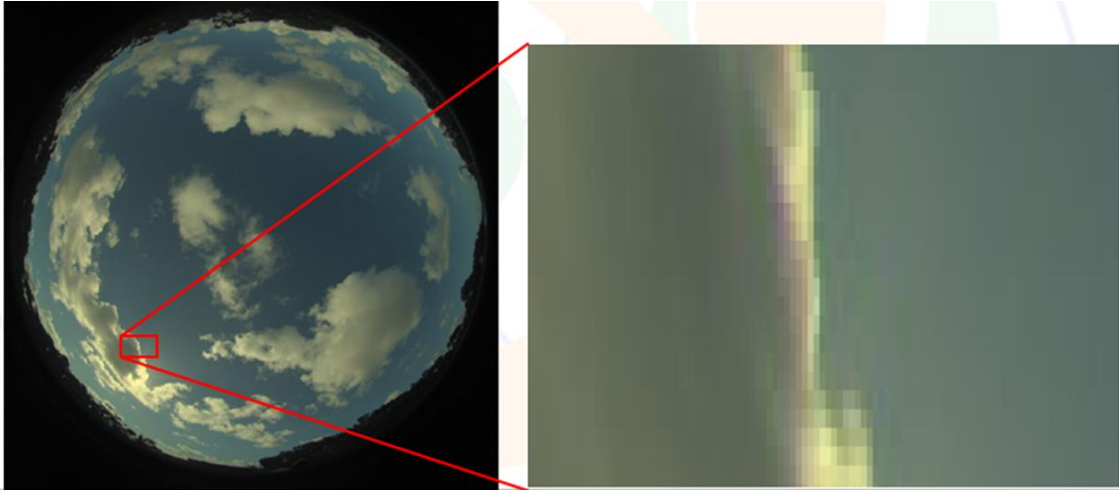
Figure 4.5: Clouds are easily identified by looking at the whole image since local and global image structure is considered. Cloud decision becomes difficult when looking at a smaller scale or pixel scale since local neighborhood information is discarded.

4.3.1 Hierarchical Segmentation

Given a coordinated transformed ground-based sky image, a hierarchical segmentation from Arbelaez et al. (2014) is used to generate an initial segmentation which is then combined with a spatio-temporal affinity graph and random walk for final segmentation (hereinafter sky image sequence segmentation (SISS)). The original multiscale hierarchical image segmentation (Arbelaez et al., 2011, 2014) can be obtained in three steps and is briefly outlined here. First, the original image is subsampled N times to generated an image pyramid. A set of "superpixels" is then obtained on each scale independently through a contour detector and Ultrametric Contour Map (UCM). Second, each hierarchy is rescaled and contours are realigned to produce a set of boundary locations with N strength. Lastly, these boundaries with N strengths are trained with a classifier and combined into a single probability of boundary estimation.Since we found that the multiscale approach produced no significant improvements and is associated a

computational cost by a factor of magnitude, only single scale hierarchical image segmentation is evaluated in this dissertation.

Single scale hierarchical segmentation groups pixels into regions or superpixels to reduce complexity, provide local, coherent visual information, and structure necessary for segmentation at the scale of interest. The aim is to train a machine to understand image content by providing more features than just color and intensity. A combination of local image brightness, color, and texture differences in oriented half-disks of three sizes (Martin et al., 2004), sparse coding on patches (Ren and Bo, 2012), and structured forest contours (Dollar and Zitnick, 2013) are used as inputs and globalized with a fast normalized cut spectral clustering method (Arbelaez et al.,2014) to detect contours. The weight of each feature is pre-trained on a BSDS training set (Martin et al., 2004) and optimum weights for our application are achieved by training on sky images. The resulting contours are often not closed and hence regions needed to be recovered by constructing UCM based on the mean contour strength (Arbelaez, 2006). Closed contours are achieved by iteratively merging adjacent regions based on the contour strength on their common boundary. The UCM are weighted contours, which represent their likelihood of being a true boundary. A given threshold of contour strength then yields a set of closed contours or "superpixels". For a hierarchical segmentation of an image S into a hierarchy of regions $\{S^1, S^2, \cdots, S^L\}$, S^1 is the base level of this hierarchy, respects weak contours, and contains the finest set of superpixels. S^L is the complete domain and regions from coarse levels respect only strong contours and are unions of regions from

fine levels. To illustrate the sensitivity of the hierarchical segmentation, different image filtering and normalizations are used to produce UCM.

4.3.2 Spatio-temporal affinity graph

After the hierarchical image segmentation, the UCM or a set of superpixels is obtained and applied to construct a spatio-temporal affinity graph to be used in the final image segmentation. Affinity graph is a common image processing technique and often used along with spectral clustering in video object segmentation (Galasso et al., 2013). A spatio-temporal region affinity graph is an affinity matrix W formed by evaluating pair-wise similarity scores $w_{i,j}$ between data elements i and j for single or multiple frames. Image pixels and point trajectories are often used as the data elements for image and video segmentation respectively (Shi and Malik, 2000, Brox and Malik, 2010). The UCM is employed here as data elements to derive single frame affinity while optical flow from Section 5.3 is employed to derive cross-frame affinity. Given superpixel p_t^i and p_t^j at frame t, a single-frame affinity (SFA) is defined as

$$w_{p_t^i,p_t^j}^{SFA} = e^{\frac{-\left\|\alpha_{p_t^i}-\alpha_{p_t^j}\right\|^3}{\sigma_\alpha^2}-\frac{\left\|c_{p_t^i}-c_{p_t^j}\right\|^2}{\sigma_c^2}}, \qquad 4.1$$

where α is the contour strength, c is the superpixel centroid, and σ are free parameters. The SFA measures the similarity of superpixels' contour strength and the distance between superpixels' centroids. For cross-frame affinity (CFA), the similarity of superpixels is computed by propagating superpixels with optical flow to the next frame and measuring their fraction by the Dice measure, and is given by

$$w^{CFA}_{p^i_t, p^j_{t+1}} = \frac{2\left|A^{t+1}_{p^i_t} \cap A_{p^j_{t+1}}\right|}{\left|A^{t+1}_{p^i_t}\right| + \left|A_{p^j_{t+1}}\right|},$$

4.2

where $A^{t+1}_{p^i_t}$ indicates the area of propagated superpixel p^i_t to frame $t+1$. The use of UCM drastically reduces the computational complexity, since the number of data elements to be considered is reduced (pixels to superpixels).

4.3.3 Random Walk

A random walk image segmentation is a method for performing multilabel, interactive image segmentation (Grady, 2006). Since the red-blue-ratio (RBR) method is erroneous only within the circumsolar region, one can use the cloud decision outside the circumsolar region as seeds in the random walk image segmentation. Given a number of pixels with user-defined labels, the potentials of labels for each unlabeled pixel are analytically determined by a random walker algorithm. Each pixel is assigned to the label for which the greatest probability is calculated. To begin, a graph G is defined as in Grady (2006) that consists of vertices (nodes) v and edges e. The weight of an edge e_{ij} is denoted by w_{ij} as in section 2.3. The degree of a vertex is $d_i = \sum_j w_{ij}$. The weight or affinity score w_{ij} is a bias that affects a random walker's choice and is undirected ($w_{ij} = w_{ji}$). Grady showed that solving the random walk solution for image segmentation is equivalent to minimizing the energy function

$$E(x) = \frac{1}{2} x^T L x.$$

4.3

where x is the probability of node labels that minimize $E(x)$ and L is a Laplacian matrix given by

$$L_{ij} = \begin{cases} d_i & if\ i = j \\ -w_{ij} & if\ i \neq j' \end{cases}$$ 4.4

where L_{ij} is indexed by vertices v_i and v_j. If we partition the vertices into two sets, V_L (labeled/seeded nodes) and V_U (Unlabeled/unseeded nodes), and reordered L and x such that labeled nodes are in front of unlabeled nodes, equation (3) can be decomposed into

$$E(x_U) = \frac{1}{2}[x_L^T\ x_U^T]\begin{bmatrix} L_{L,L} & L_{L,U} \\ L_{U,L} & L_{U,U} \end{bmatrix}\begin{bmatrix} x_L \\ x_U \end{bmatrix}$$ 4.5

$$= \frac{1}{2}\left(x_L^T L_{L,L} x_L + 2x_U^T L_{U,L} x_L + x_U^T L_{U,U} x_U\right).$$

Since L_{ij} is positive semi-definite and the critical points of $E(x_U)$ will be minima, differentiating Eq. 4.5 with respect to x_U to find the minima yields

$$L_{U,U} x_U = -L_{U,L} x_L,$$ 4.6

where x_U is the only unknown since x_L is user-defined potentials. By solving Eq. 4.6, the potentials of labels for each node, or superpixel in this dissertation, are determined.

For the implementation, the original sky image is subsampled to exclude sun-pixel-angle greater than 50°. For superpixels at a sun-pixel-angle greater than 12°, the cloud decision from the CSL method is used to predefine labels. Since a superpixel may contain a group of pixels with different labels, a superpixel is assigned the label with the highest fraction of pixels. Then for sun pixel angles less than 12°, the superpixels are unlabeled and will be assigned a label with the random walk method.

4.3.4 Evaluation metric

To evaluate the skill of SISS segmentation, the CSL method is used as the baseline method and both are compared against a ground truth dataset produced by

human assisted annotation. Since manually annotating sky images to obtain ground truth segmentation is very costly, a human assisted segmentation method in Arbelaez et al. (2008) is used to produce a ground truth dataset. The idea of human assisted segmentation is very similar to our proposed method except that labels are manually corrected. Human assisted annotation is more efficient than human annotation, since the human only correct errors rather than drawing each cloud contour. Binary cloud decision from the CSL method is used as an initial segmentation. The ground truth is then achieved by either manually adjusting the incorrect labels or partially seeding labels and letting the algorithm in Arbelaez et al. (2008) determine labels of the unlabeled pixels (Figure 4.6). To determine the accuracy of the segmentation, binary cloud decision is evaluated using the success ratio ($SR = \frac{hits}{hits + false\ alarms}$) and hit rate ($HR = \frac{hits}{hits + misses}$) for each pixel. Since the ground truth segmentation is based on the CSL method, the results will bias toward the CSL method. To avoid the bias, cloud decision is quantitatively evaluated only for sun-pixel-angle less than 12°, where the CSL cloud detection error mostly occurs.
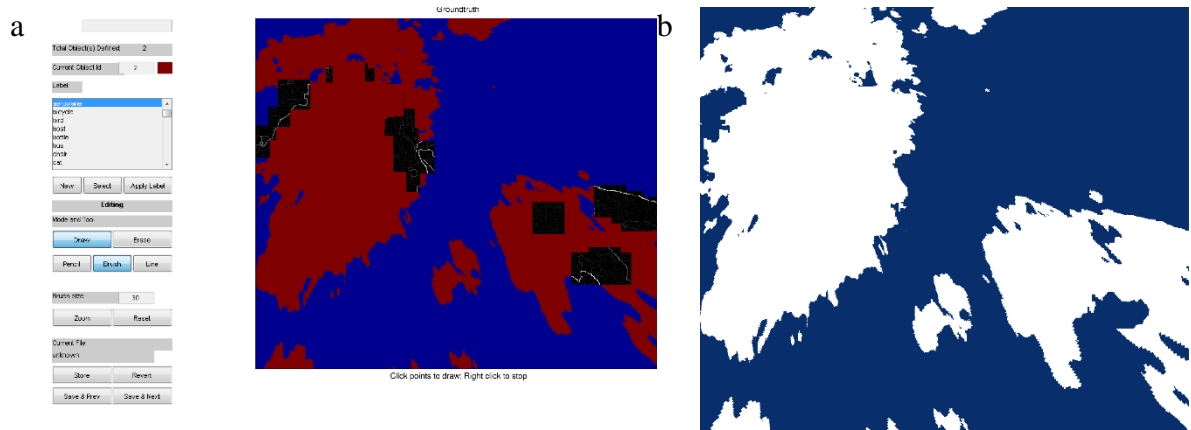
Figure 4.6: Ground truth segmentation (b) can be obtained by removing incorrect labels (as indicated by black tiles in (a)) and propagating the seeded labels (red and blue) to the black regions from hierarchical boundaries. Ground truth segmentation is biased toward the CSL method.

4.3.5 Results
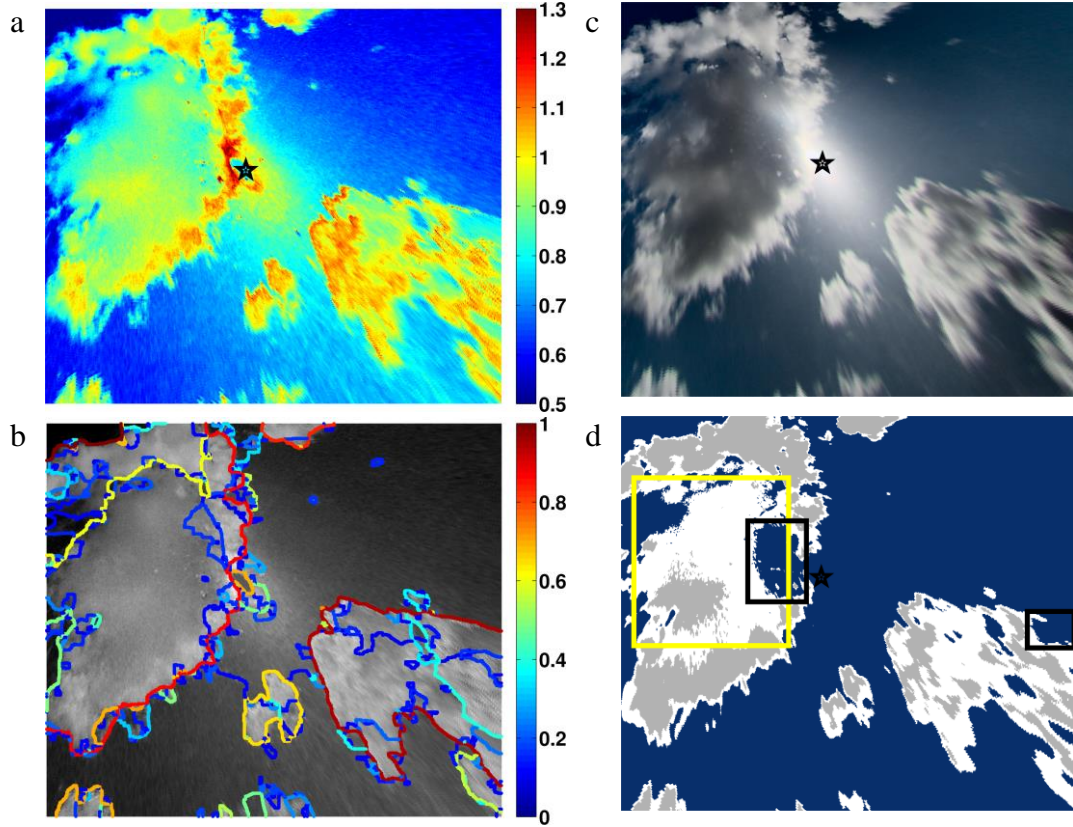
Figure 4.7: Cloud decision of thick cumulus clouds (b, white: cloudy, blue: clear) using the CSL method on November 10th, 2012, 09:11:30 PST. A RBR (colorbar)(a) smaller than the clear sky background is observed in cloudy region with gray color (c) and leads to false clear sky decision and optical depth reduction (as indicated by the black and yellow box in d). The hierarchical segmentation generates weighted contours (colorbar in b) and superpixels on top of the sky image (b). The false clear sky region is grouped into a superpixel with other parts of the cloud, and making it likely that the region will be correctly identified as cloudy at the final segmentation.

Figure 4.7c illustrates the advantage of using the UCM. Even though the RBR of the gray region is indistinguishable when compared to the clear sky at the individual pixel level, the gray region is grouped into a single superpixel with other parts of cloud because of the local image texture. It demonstrates the importance of multi-scale texture features where the spatial dependency between pixels is considered. All the features described in section 4.3.1 can be powerful representations that can be transformed into contours and,

more importantly, clustered together to form superpixels. The abstraction of superpixels from the image allows recognizing coherent objects and properly interpreting clouds.

Since the spatio-temporal affinity graph and random walk segmentation highly depends on the output from the hierarchical segmentation, it is useful to study the sensitivity of the UCM to different inputs. In particular, we examine the RBR versus the ΔRBR image and the impact of smoothing the image. The results of UCM from four different inputs are presented in Figure 4.8. Visually inspection indicates that, UCMs from the ΔRBR image shows a better performance than the RBR image with more accurate boundary locations and stronger contour strength contrast. For the same clouds in the sky image, stronger contour strength is observed in the ΔRBR image. This is because subtracting the clear sky background RBR removes the sun-pixel-angle dependence in the ΔRBR image, making cloud edges appear sharper and more homogeneous across a range of sun-pixel angles.

Improved boundary locations are observed (Figure 4.8) after filtering the image with a 5 by 5 pixel median filter. Filtering noise appears to yield more representative local image textures. A median filtered ΔRBR image is used in this dissertation as the input to the hierarchical segmentation as the filter results in more precise boundary locations and stronger contour strength contrast, which is advantageous when applying the affinity graph and random walk segmentation. To improve the hierarchical segmentation results, weights of the local features as described in section 4.3.1 can be trained on a sky image training set.
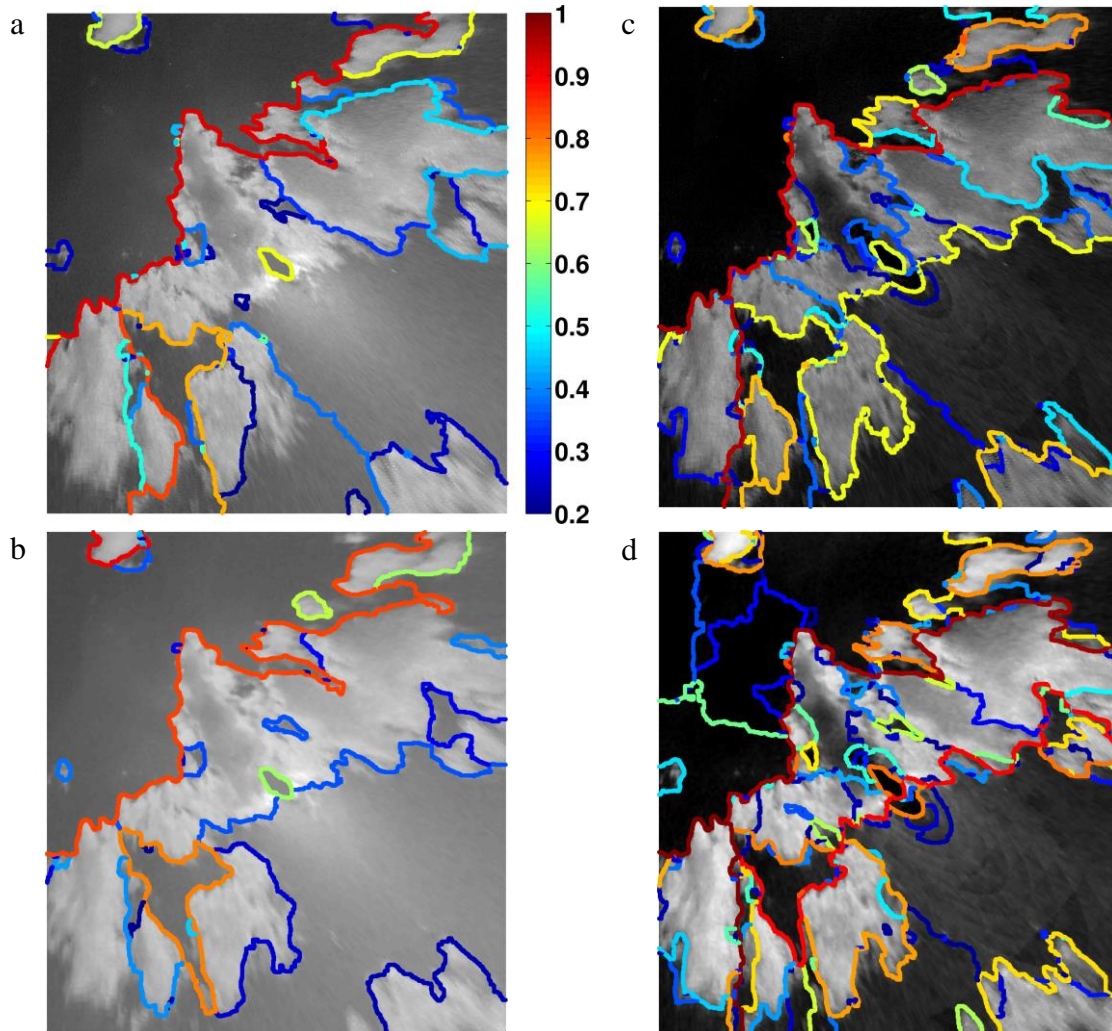
Figure 4.8: The Ultrametric Contour Map (UCM) is overlaid with the RBR image (a & b) and ΔRBR image (b & c) on November 10th, 2012, 08:32:00 PST. Both inputs with a median filter (b & d) have a stronger contour strength (colorbar). All contour strengths less than 0.2 is removed for image clarity. The ΔRBR image with median filter (d) is used in this dissertation as the input.

Even though the hierarchical image segmentation reduced the number of pixels from around 50,000 to 8000 by the use of superpixel, the computational cost for operational solar forecasting would still be prohibitive. Since the UCM contains many superpixels with weak contours that do not improve the final segmentation, all contour weights less than 0.1 (0-1 scale) are removed. This further reduces the number of superpixels to around 800 and gives reasonable computation time.

The hierarchical segmentation alone does not give satisfactory cloud segmentation result. Cloud decision can be obtained from the hierarchical segmentation by labeling each superpixel with the dominant RBR cloud decision within its boundaries. Before obtaining labels for the superpixels, the appropriate scale or granularity of the superpixels has to be determined. Figure 4.9 illustrates that by applying different thresholds on the UCM yields multiple levels of granularity for superpixels. The threshold in the UCM, serves to identify the closed contours with weights above the threshold to define a segmentation. As expected, the lowest threshold (0.15 in Figure 4.9a) results in weak contours oversegmentating the image. Increasing the threshold from 0.15 to 0.3 (Figure 4.9b) removes weak contours and smaller regions are aggregated. Finally, applying a threshold of 0.5 (Figure 4.9c) yields superpixels at the scale of a cloud. The bounding contours of the 0.5 threshold appear to match the cloud boundary locations nicely (as compared to Figure 4.8). The results reinforce that most clouds have strong contours and this characteristic should be taken advantage for cloud segmentation or estimating cloud properties. By determining the label of superpixels at this scale, the initial cloud decision image is obtained. However, the RBR cloud decision is erroneous itself and hence produces unreliable results even with the additional information from the superpixels. Yet, the hierarchical segmentation provides a natural starting point for further refinements that will be described in the following.
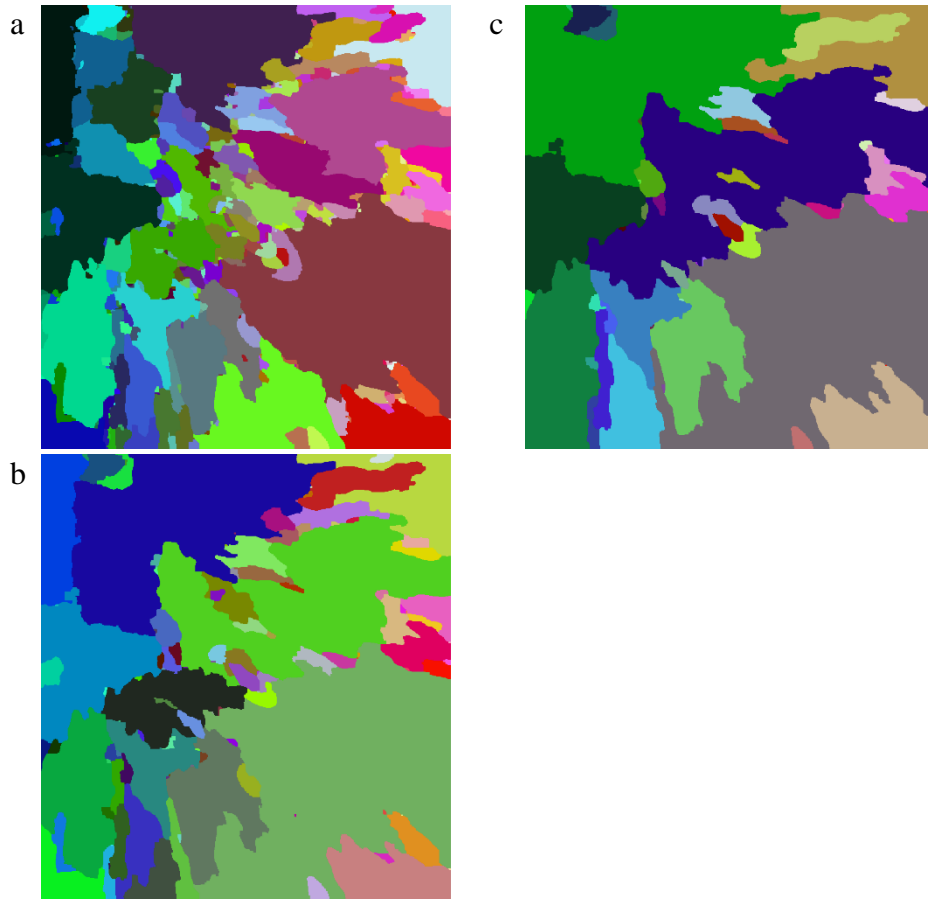
Figure 4.9: Closed contours or superpixels of different scales are formed from the Ultrametric Contour Map (Fig. 3) by applying a threshold of 0.15 (a), 0.3 (b), and 0.5 (c). A low threshold yields an oversegmentation of an image and a high threshold yields superpixels corresponding to the cloud scales.
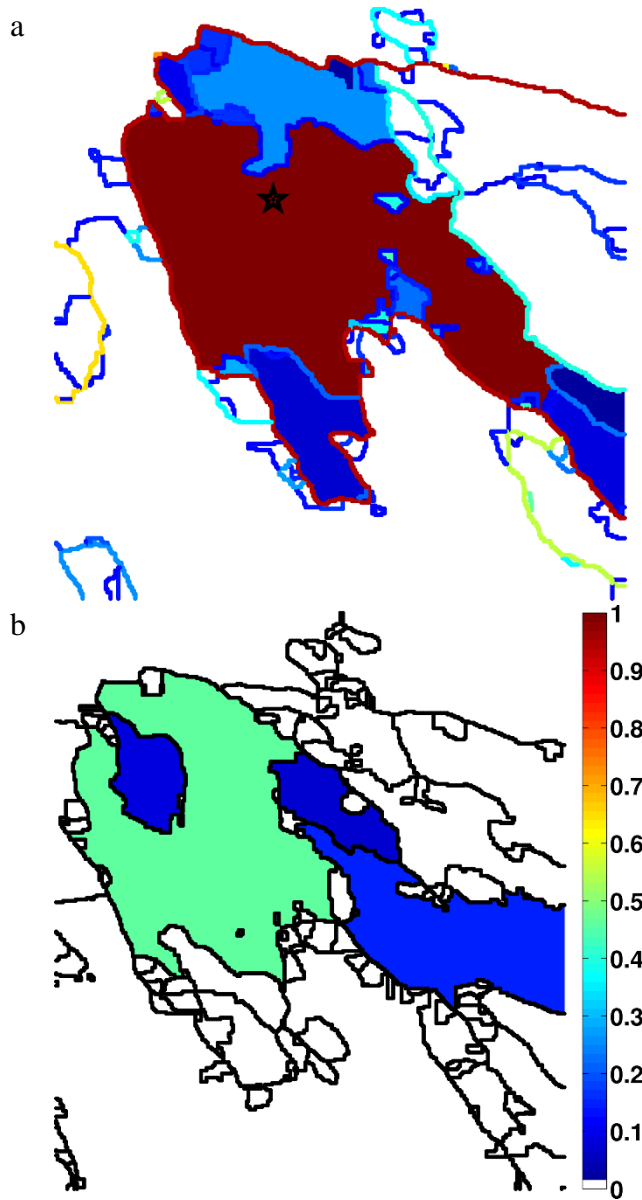
Figure 4.10: The score of single-frame affinity (a) and cross-frame affinity (b) for the red superpixel (containing the star) for November 10th, 2012, 09:11:30 PST. The colorbar indicates both the affinity score and contour strength. Since the cross-frame affinity score does not depend on the contour strength, the contours in (b) are shown in black for clarity.

Figure 4.10 shows the single-frame (Figure 4.10a) affinity score for a superpixel with an erroneous RBR cloud decision. The affinity score is set up in a way that the superpixel itself has a score of unity. The single-frame affinity score between superpixels

depends on the strength of contour that separates them and the distance between their centroids with a stronger contour and larger distance resulting in smaller affinity. The red superpixel shown in Figure 4.10 is bounded by approximately three different contour strengths: red (~0.9) on the left, cyan (~0.4 on the bottom), and blue (~0.15 on the top). The affinity scores between the red superpixel and other superpixels separated by weak contour (blue) are between 0.1 and 0.2. An affinity score of zero (white) is assigned to superpixels that are outside the stronger contours (red and cyan).

These affinity scores serve as most important inputs to the random walk method, which resembles a steady state diffusion process described by a Laplace equation (Grady, 2006). The contour strength and affinity score determines how the label is propagated into the neighboring superpixels. In this case, the label from the red superpixel diffuses into the blue superpixels through the weak contours and the process is stopped by the stronger contours surrounding the blue superpixels. The affinity score can be easily adjusted by changing the value of σ in Eq. 4.1.

On the other hand, the cross-frame affinity score does not depend on the contour strength but on the dice score in 4.2. Figure 4.10b shows the affinity score between the red superpixel (Figure 4.10a) and superpixels from the previous frame. The turquoise superpixel overlap the most with the red superpixel after being propagated to the next frame with the optical flow and has an affinity score of 0.47. The affinity score across frames improve the temporal consistency of labels. For example, the red superpixel is more likely to be labeled as cloudy if the turquoise superpixel in the previous frame has a cloudy label.

The random walk segmentation result of a cloud image sequence is presented in Figure 4.11. The cloud decision image from the CSL method is used as the seeds for the random walk segmentation. A set of points (yellow in Figure 4.11a) within the circumsolar region is not seeded as the cloud decision is unreliable as previously mentioned. The potentials of two labels (cloudy and clear sky) are then calculated for each superpixel by solving Eq. 4.6. One should note that even though the cross-frame affinity is computed between two frames, the random walk solution is solved on a Laplacian matrix, which contains all frames from a user-defined length. The random walk segmentation results can be different if a different length of cloud image sequence is used. In this dissertation, a length of 15 frames or 7.5 min is selected as it is observed to be a common duration for cumulus cloud to be within the field-of-view of the sky imager. A more comprehensive sensitivity study of the segmentation result to the length of image sequence is recommended for the future and it is likely to be a function of trajectory lengths as computed in Section 5.4. Figure 4.11b illustrates the potential of the clear sky label. The label with the highest potential for each superpixel is selected. The final cloud decision (Figure 4.11c) is shown to overcome the problem of false clear sky region when compared to the result in Figure 4.7b.
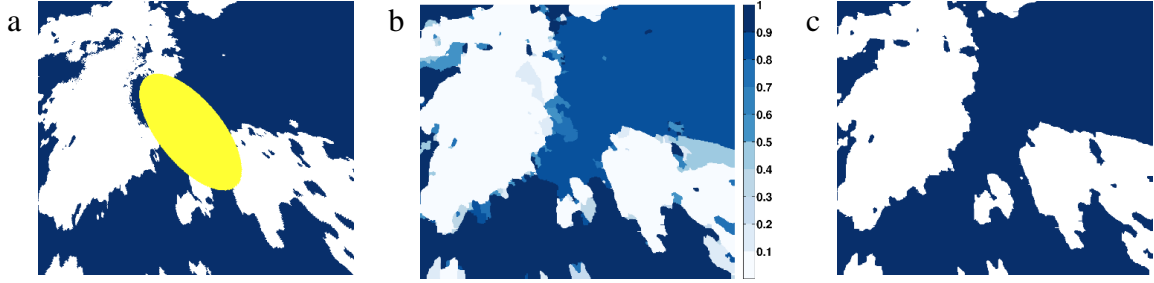
Figure 4.11: The random walk image segmentation is initialized with seeds from the CSL method. A set of points (yellow), which prone to high decision error, is not seeded (a). The potential of clear sky label is computed (b)(the potential of cloudy label is not shown here) and the dominant label for each superpixel is selected as the final cloud decision (c). The SISS segmentation is different from the CSL method even for seeded regions (sun-pixel-angle greater than 12°) since they are prelabeled by the dominant seed for each superpixel as described at the end of Section 4.3.3.

While the random walk segmentation is obtained for the entire image sequence, i.e. 15 frames, we are most interested in the latest frame for operational solar forecast application. Therefore, only the latest frame from the SISS method is used in the validation. Table 4.1 summaries daily success ratio and hit rate from both methods. The result indicates that our proposed method improves upon the CSL method.

Table 4.1: Daily success ratio and hit rate for both methods. All images are evaluated for sun-pixel-angle less than 12° (yellow region in Figure 4.11).

	Success Ratio		Hit Rate	
	CSL	SISS	CSL	SISS
Nov 10th	0.8944	0.9325	0.8619	0.8907
Nov 18th	0.9094	0.9158	0.8810	0.9021

Acknowledgements

Co-authors: Bryan Urquhart, Matthew Lave, Anthony Dominquez, Jan Kleissl, Janet Shields, and Byron Washom

Publication: Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C. W., Ghonima, M., & Kleissl, J. (2014). Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. Solar Energy, 103, 502-524.

Co-authors: Handa Yang, Ben Kurtz, Andu Nguyen, Bryan Urquhart, Mohamed Ghonima, and Jan Kleissl

Publication: Chow, C. W., Belongie, S., & Kleissl, J. (2015a) Sky image sequence segmentation for solar forecasting. In preparation.

Co-authors: Serge Belongie and Jan Kleissl

5 Cloud Motion

5.1 Introduction

Cloud motion estimation and tracking has a long history in visible satellite imagery (e.g. Menzel, 2001) and solar irradiance forecasts up to a few hours ahead are strongly dependent on the apparent motion of cloud structures. Sparse cloud motion vectors (CMVs) are generally obtained by first locating salient image features such as brightness gradients, corners, cloud edges, or brightness temperature gradients (Bedka and Mecikalski 2005, Menzel 2001). Assuming the features do not change significantly over a short interval, CMVs are computed using pattern-matching techniques applied to successive images. The future cloud situation is then estimated by the extrapolation of motion assuming persistence of cloud speed, size, and shape. CMVs as a basis for advecting frozen cloud field derived from ground-based sky imagers is first studied by the author and presented in this dissertation.

5.2 Cross-correlation method (CCM)

5.2.1 Motion estimation

Cloud velocity and direction of motion is determined through the cross-correlation method (CCM) applied to two consecutive sky images (Hamill & Nehrkorn 1993). Before applying the CCM, images are projected into sky coordinates to remove geometric distortion. The projected sky image is partitioned into subsets of pixels of equal size such that each subset is about 1% of the sky image area. The CCM finds the position that best matches each given subset of pixels in the previous sky image within

the current image (Figure 5.1). The CCM yields a wind vector (direction and speed) with the largest cross-correlation coefficient (CCC) that specifies the quality of the match. Since the CCM is computationally expensive, the search area is limited to a search distance corresponding to a velocity of $0.012 \times$ CBH. In general, the CCM obtains an accurate motion on inhomogeneous areas which have high contrast of pixel values, such as a cloud with a sharp boundary in a clear sky. The CCM is performed on the red channel image which has a higher contrast between clear sky and cloud than the blue and green channels. (The CCM is not performed on the cloud decision image to avoid error introduced by incorrect cloud detection).
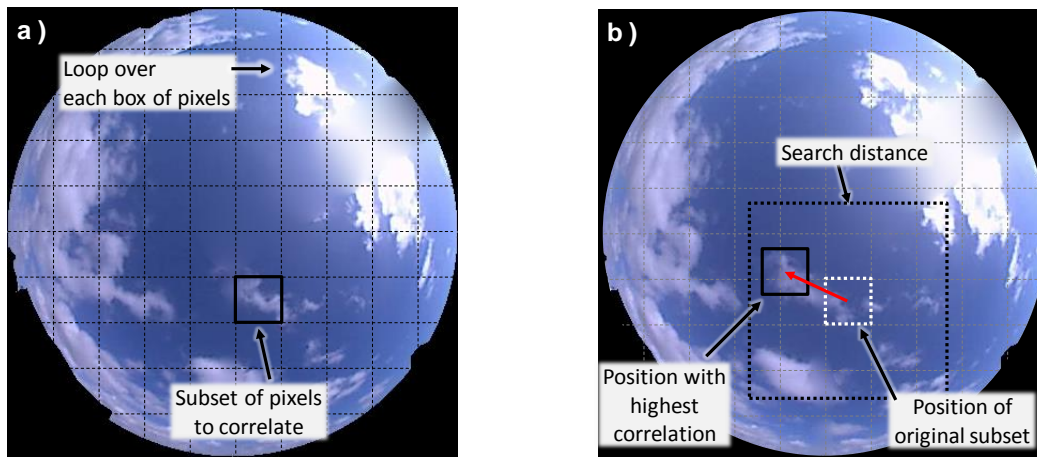
Figure 5.1: Illustration of the cross-correlation method using two images on October 4, 2009 at 16:18:30 (a) and 16:19:00 PST (b). Each subset of pixels from (a) is correlated to (b) within a search distance. The location of the highest correlation is found and a motion vector is defined for each box. The cloud velocity determined from this pair of images was 5.2 m s-1. While the method is illustrated here on a full color sky image, the method is actually applied to the coordinate transformed red channel image.

Assuming spatial homogeneity of cloud velocity, the vector field obtained through the CCM is further processed using several quality controls to yield an average cloud velocity across the image. The circumsolar region is often offscale bright (when a pixel registers the highest possible 8-bit value) and shows little variation over time in

clear conditions. This results in erroneously small motion vectors and hence the CCM is not performed in this region. Using the information in the cloud decision image, the CCM is performed only on areas that contain clouds. Only motion vectors with CCC higher than 0.8 are considered. Finally, vectors that are more than one standard deviation from the mean in either the x or y component of the velocity are eliminated. The remaining vectors are vector-averaged to yield the advection velocity of the cloud field. We note that this method needs further development for the case of several cloud layers with different velocity vectors.

Figure 5.2 illustrates the computed motion vector field at two in the quality control process. A reduction in the CCC usually implies smaller accuracy of motion vectors. Reduced CCC can be the result of cloud formation, deformation, and evaporation. Also the appearance of a cloud may change over time due to different camera white balancing, lighting on the cloud, and different viewing geometries.

After performing the CCM only on cloudy region, and vectors with CCCs less than 0.8 are removed, most remaining vectors are uniform (Figure 5.2a). By computing the average and standard deviation of the vector field, the remaining outlier vectors are eliminated (Figure 5.2b).
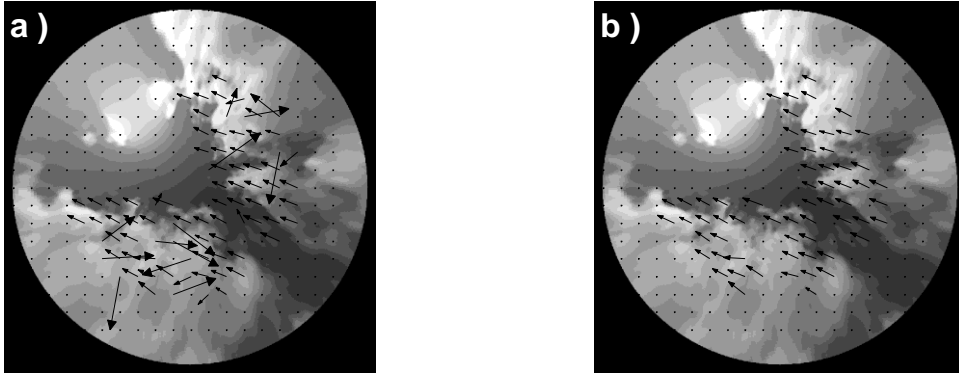
Figure 5.2: Illustration of procedures for obtaining a motion vector field on a sky image in sky coordinates on March 4, 10:21:30 PST. (a) Vector field resulting from the cross-correlation method after removing vectors with a correlation of less than 0.8, and (b) after removing outliers. The average cloud velocity was 4.0 m s-1.

Incorrect motion vectors are found primarily in overcast and nearly clear sky conditions (Figure 5.2). In overcast skies, the sky contrast is reduced resulting in low CCC. In nearly clear sky conditions, clouds are often thin and evaporate. If the sky contains clouds with sharp cloud boundaries, CCM generally performs well. Due to the general persistence of cloud velocities (usually within 2 m s^{-1} over several minutes), erroneous decision vectors can be identified since they are associated with large cloud velocity fluctuations. If the cloud velocity changes by more than two standard deviations of velocity over the past 5 minutes, the current cloud velocity is replaced by the average velocity over the past 5 minutes.

5.2.2 Cloud forecasting

To forecast cloud cover, the cloud map at time t_o is advected at the speed and direction of the global vector determined from cross-correlating the images at time t_o and $t_o - dt$ ($dt = 30$ seconds). To determine accuracy, the actual cloud map at time $t_o + dt$ (Figure 5.3b) is overlaid onto the advected cloud map (Figure 5.3c) to determine the

pixel-by-pixel forecast error (Figure 5.3d). The matching error between the two cloud maps is

$$e_m = \frac{P_{false}}{P_{total}} \times 100\% \, ,$$

<div align="right">5.1</div>

which is the ratio between the number of falsely forecasted pixels and the number of total pixels in a region. The cloud-advection-versus-persistence (cap) error:

$$e_{cap} = \frac{e_m}{e_{m,p}} \times 100\% \, ,$$

<div align="right">5.2</div>

describes the forecast error obtained by cloud advection (Figure 5.3d) divided by the error obtained if the image at t_o was assumed to persist until $t_o + dt$ (no advection). An $e_{cap} < 1$ implies that the cloud advection improves the forecast compared to persistence.
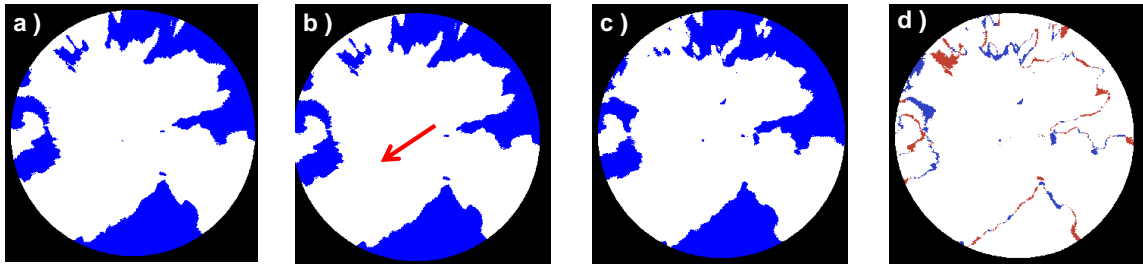
Figure 5.3: The forecast cloud map (b) is produced by advecting the cloud map at time t_o (a) in the direction of the motion vector (indicated by the red arrow). To determine the forecast error (d), the future cloud map at time $t_o + 30$ seconds (c) is compared to the forecast (b). Blue and red colors in (d) show forecast errors (blue: pixel forecast cloudy and but actually clear; red: pixel forecast clear but actually cloudy) and white indicates accurate forecasts.

Matching errors for mostly uniform sky conditions (i.e. completely clear or completely overcast) are by default close to zero and are not an interesting test of forecast skill, so the mean and standard deviation were only computed using matching errors for times corresponding to 5% < cloud fraction < 95%. Similarly, daily cap error, scalar average cloud speed, and average cloud height were computed only for the same time

periods. In Section 5.2.2.1, the following days were analyzed: September 14, 2009; October 4, 2009; March 4, 2010; and March 10, 2010. In Section 5.2.2.2 and 5.3, the entire month of November 2012 was selected for analysis. November 12, 2012 was excluded due to maintenance and replaced by December 2, 2012. These 31 days provided a variety of sky conditions with which to validate USI forecasts, including clear skies, marine layer overcast stratus and stratocumulus, passage of frontal systems, and partly cloudy conditions with low-level cumulus, mid-level altocumulus, and high-level cirrus clouds. Six days were clear and are not analyzed further, but most days had challenging partly cloudy conditions with cloud fractions between 20% and 65% (see Table 5.1 for a detailed summary).

Table 5.1: Summary of sky conditions listed chronologically. Sky conditions joined by an ampersand indicate simultaneous occurrence. All times in PST (UTC - 8 hours).

Date	Morning (7:00 - 9:59)	Midday (10:00 - 12:59)	Afternoon (13:00 - 16:00)
Nov 1, 2012	OVC	Cu-EF	CLR, Ci
Nov 2, 2012	OVC, Cu-EF	CLR	CLR
Nov 3, 2012	CLR	CLR, Cu	Cu
Nov 4, 2012	CLR	CLR	CLR
Nov 5, 2012*	CLR	CLR	CLR
Nov 6, 2012	CLR, Fog (brief)	CLR	CLR
Nov 7, 2012	OVC	Cu-EF	Cu-EF
Nov 8, 2012	Ci, Ac	Cu-EF & Ac	Cu & Ac
Nov 9, 2012	Cu & Ac, Ac	Ac, Cu & Ac	Cu-EF
Nov 10, 2012	Cu	Cu-EF	Cu-EF, CLR
Nov 11, 2012	CLR	CLR	CLR
Nov 13, 2012	CLR	CLR	CLR
Nov 14, 2012	Ci	Ci, Cc	Cc
Nov 15, 2012	OVC	OVC	OVC
Nov 16, 2012	Ac, CLR	CLR, Ac, CLR	CLR, Ac
Nov 17, 2012	Cu-EF	Cu-EF	Ac
Nov 18, 2012	OVC, Cu-EF & Ac	Cu-EF, Ci, Cu & Ci	Cu
Nov 19, 2012	CLR	CLR	CLR, Cu
Nov 20, 2012	CLR	CLR	Cu
Nov 21, 2012	OVC	Cu-EF, CLR	CLR
Nov 22, 2012	Fog, Cu-EF	≈ CLR	Cu-EF
Nov 23, 2012	OVC	CLR, Cu-EF, CLR	CLR, Cu-EF, OVC
Nov 24, 2012	CLR	CLR	CLR, Cu, Haze
Nov 25, 2012	OVC	OVC, Cu-EF, CLR	CLR, Cu, CLR
Nov 26, 2012	Cu-EF	CLR, Cu-EF, Cu-EF & Ci	Cu-EF & Ci
Nov 27, 2012	OVC	Cu-EF & Ci, Ci	Cc, Cc & Cu, OVC
Nov 28, 2012	OVC	Cu-EF, Haze	CLR
Nov 29, 2012	OVC	OVC	OVC, Ac
Nov 30, 2012	OVC	Cu-EF & Ac, OVC	OVC
Dec 1, 2012	OVC, Cu-EF & Ci	Cu-EF & Ci	Ci, Cu & Ci
Dec 2, 2012	OVC	OVC, Cu	CLR

CLR: Clear sky. Cloud fraction < 5%.
OVC: Overcast. Cloud fraction > 95%.
-EF: Denotes periods of prominent cloud evaporation and formation.
Cu: Cumulus clouds. Low-level (< 2 km) clumpy clouds with sharp edges.
Ac: Altocumulus clouds. Mid-level (2 km to 6 km) clumpy clouds with sharp edges.
Cc: Cirrocumulus clouds. High-level (> 6 km) clumpy clouds with sharp edges.
Ci: Cirrus clouds. High-level (> 6 km) thin, wispy clouds.

5.2.2.1 TSI

The performance of minutes ahead forecast is discussed in this section for four days following the metrics established in Figure 5.3. Since the circumsolar region has a large potential for erroneous cloud decision, the sky image is divided into the circumsolar

region (within 35° of the solar azimuth angle) and the outer region (Figure 5.4) for the

TSI. Figure 12 show the time series of cloud speed, cloud fraction, and errors for a 30

second forecast on September 14. Since both clear and overcast skies always have perfect

forecasts due to the use of binary cloud maps, cloud speeds and forecast errors are not

assigned for these times.
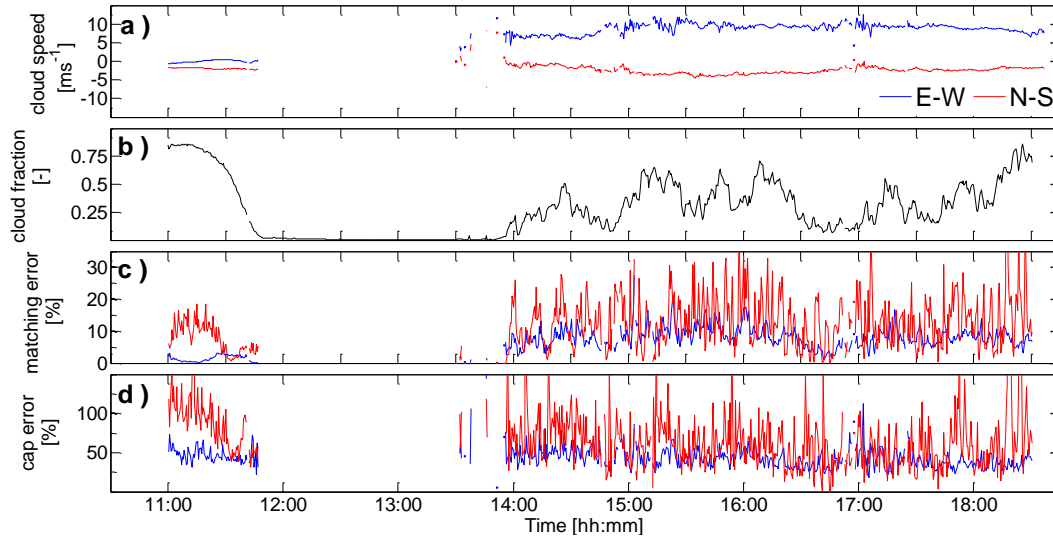
Figure 5.4: Time series of (a) eastward and southward cloud speed, (b) cloud fraction, (c) matching error, and (d) cloud-advection-versus-persistence (cap) error obtained every 30 seconds on September 14, 2009. The red line in c) and d) indicates the error in the circumsolar region, while the blue line indicates the error in the outer region. In clear conditions (e.g. 1145 – 1350 PST) no cloud speed is determined and the errors are zero.

On September 14, 2009 a transition from overcast to clear sky occurred between

1100 PST and 1200 PST with clouds at an altitude of 914 m moving from north to south

at less than 2 m s^{-1}. In the afternoon, cumulus clouds moved at high speed eastward and

were evaporating and deforming rapidly, as evident by the variable cloud fraction. As a

result, the matching error in the afternoon is much larger than between 1100-1200 PST.

Despite the poor matching error, September 14 has similar cap error as the other days

(Table 5.2). The matching error is larger than on other days, but the large cloud speed

causes large persistence errors which reduces the cap error. The invariance of the cap error over time suggests that the normalization by the persistence image provides a universal performance metric. While matching errors become small when the cloud cover fraction is small or large (the edge-to-area ratio is small in the latter case), the cap error is not sensitive to cloud cover fraction. The cap error is however sensitive to cloud speed.

In the circumsolar region, the large errors are a result of the poor cloud decision. While the SP increases the accuracy of the cloud decision in the circumsolar region, its high variability occasionally causes cloud decision errors. Additionally, a false cloud detection results when two different clouds bordering the shadowband on both sides are interpolated across. Both issues cause a step change in the cloud decision in a large area between consecutive images, which leads to erroneous forecasts.

Mean and standard deviation of outer region error metrics for a 30 second forecast are summarized in Table 5.2. The mean matching errors are 3 to 7% and the cap errors are 40 to 45%. The cap error of less than unity indicates that translating the cloud map improves the forecast compared with persistence. The mean matching error scales with cloud speed. Otherwise, the comparison of different days does not yield strong correlations between different variables, presumably due to intra-day variability.

Table 5.2: Mean and standard deviation of matching and cap error (5.1 and 5.2) for the outer region on four different days. The computed errors are independent of ground irradiance measurement and are averaged only during conditions that were not clear or overcast shown in the sky images. Since errors during overcast and clear conditions are zero, the errors in the table are biased high.

	mean e_m [%]	mean e_{cap} [%]	std e_m [%]	std e_{cap} [%]	avg cloud fraction [-]	avg cloud speed [m s^{-1}]	avg cloud height [m]
Sept 14, 2009	6.26	42.8	3.24	14.2	0.135	7.79	840
Oct 4, 2009	4.09	40.0	1.48	7.45	0.473	3.75	1090
Mar 4, 2010	3.16	44.7	1.33	10.5	0.276	2.95	610
Mar 10, 2010	3.85	41.1	2.40	20.2	0.427	4.49	1220

Table 5.3 shows the mean total matching error and total cap error for a 30 seconds to 5 minutes ahead forecast. Mean forecast matching errors increase less than linearly from around 6% for 30 second forecasts to 23-30% for 5 minute forecasts. Mean cap errors are less than 80% out to 5 minutes, except on September 14, 2009 when large cloud speeds were observed. Note that while the outer region matching error was considered for 30 second forecasts, here the total (circumsolar and outer region) errors are considered as clouds are frequently advected between the circumsolar region and the outer region (depending upon cloud speed and direction).

Table 5.3: Mean total matching error and total cap error for a 30 seconds to 5 minutes ahead forecast. The computed errors are independent of ground irradiance measurement. The errors are averaged only during conditions that were not clear or overcast shown by the sky images and when the entire scene had not been advected out of the field-of-view. Since errors during overcast and clear conditions are zero, the errors in the table are biased high.

	30 sec	1 min	2 min	3 min	4 min	5 min	30 sec	1 min	2 min	3 min	4 min	5 min	Time until advection out of scene [min]
			e_m [%]						e_{cap} [%]				
Sept 14, 2009	8.67	13.5	20.1	23.8	27.2	28.0	45.0	51.5	63.9	70.0	76.5	123	4 - 27
Oct 4, 2009	6.01	9.84	15.9	20.5	24.2	27.1	47.2	49.8	55.6	61.5	66.6	70.3	8 - 18
Mar 4, 2010	5.36	8.68	13.6	17.3	20.4	23.4	54.6	55.3	59.3	63.4	67.7	71.8	9 - 24
Mar 10, 2010	6.41	10.7	17.8	23.1	27.2	30.0	48.8	53.9	62.3	68.8	75.1	78.0	9 – 15

For the low cloud heights and cloud speeds observed, the maximum forecast horizon was about 5 to 25 minutes. The following sources of forecast error are observed:

(i) In general, cloud deformation contributes most to forecast error. In our low latitude coastal areas with mostly mesoscale weather patterns, changing cloud patterns are more common than in more synoptically driven areas. Large solar irradiance can lead to short convection and cloud evaporation time scales. The rapid condensation and evaporation and associated deformation of clouds introduce complexity to deterministic cloud forecasting.

(ii) Perspective and occlusion effects on clouds are another source of error resulting from the two-dimensionality of the sky image. Clouds are assumed to have a negligible vertical extent at a uniform height; the sides of approaching clouds are interpreted as cloud bottoms. Consequently, as the clouds approach the sky imager, the

perspective changes until only the cloud bottom is visible leading to a perceived decrease in cloud sizes and change in cloud shape. The perspective error may also create the illusion that there is one cloud while there are two nearby but separate clouds. Generally, the perspective error is inversely proportional to cloud vertical extent; it will be largest for tall thunderstorm clouds and small for thin stratus clouds. A network of sky imagers with different perspectives could reduce this forecast error.

(iii) Heterogeneity of cloud velocity. Usually, the cloud velocity computed from CCM is representative of the velocity of all clouds in the image. In rare cases, however, cumulus clouds to the east of the sky imager were moving at a different velocity than stratiform clouds to the west. Since our forecast assumes a homogeneous cloud velocity (roughly the size weighted average velocity of the two cloud layers), large forecast errors result.

The method's accuracy is also limited by the shortcomings of the sky imager used. Part of the data within each image is lost due to obscuration by the camera arm and shadowband. Low dynamic range causes images to saturate easily near the sun, even with automatic gain control (AGC) of the CCD camera. AGC causes its own problems because the output signal strength fluctuates between images, leading to shifting intensity histograms and making spectral composition based cloud decisions less reliable. Inaccuracies in the forecast sky conditions are also due to cloud deformation, evaporation, and condensation, as well as uncertainty in cloud base height. The effect of these hardware limitations is evident by the increased errors in the circumsolar region (Figure 5.4).

5.2.2.2 USI

Summaries of image-based forecast performance for USI are presented in Table 5.4 for forecast horizons of 30 seconds and 5 minutes. The image-weighted average 5-minute cap error is 91.9%, which implies 5-minute advection forecasts performed better than image persistence over the entire month. Inspection of daily 5 min cap errors reveals that 11 of 22 days exhibited cap errors higher than 100%. Ergo, pure advection of clouds is not appropriate for all sky conditions

The image-weighted average 30-second cap error of 67.0% is larger than cap errors in the forecast by TSI (Section 5.2.2.1), which ranged from 45.0% to 54.6%. The larger validation set analyzed in this section (31 consecutive days versus 4 ideal days) presented a wider variety of cloud conditions, causing a greater range in cap errors and a larger (but more representative) average cap error. Additionally, new features of the USI such as thin cloud detection and an unobstructed circumsolar region (area immediately surrounding sun) result in more accurate kt assignment and greater visibility of the sky dome, but increase cap errors. Since thin cloud detection is very sensitive to the ΔRBR threshold, and that threshold is not as distinct from typical clear sky RBRs as the thick cloud threshold, thin cloud detection fluctuates more from image to image – for example, due to sun obstruction or stray light in the optics. Cloud decision errors in the circumsolar region, along with the sun obstruction CSL bypass also cause fluctuations in the number of cloudy pixels between images. These effects increase the total number of false pixels in both advection and image persistence forecasts, causing their ratio to be closer to unity.

Table 5.4: Mean and standard deviation of matching errors, along with total daily cap errors (Eqn. 3 and Eqn. 4) for fh = 30 s and fh = 5 min. Reported average cloud speed is a scalar average.

Date	e_m(%) 30 s	5 min	Std. e_m(%) 30 s	5 min	e_{cap}(%) 30 s	5 min	Avg. cloud fraction (%)	Avg. cloud speed (m/s)	Avg. CBH (m)
Nov 01	4.6	15.5	1.6	7.5	90.8	102.9	45.0	3.5	1798
Nov 02	3.9	12.3	1.9	6.8	90.6	96.0	12.0	14.7	5796
Nov 03	6.1	28.1	2.8	17.2	62.3	379.6	8.7	23.3	1340
Nov 04	-	-	-	-	-	-	0.0	-	-
Nov 05	-	-	-	-	-	-	0.0	-	-
Nov 06	-	-	-	-	-	-	0.6	-	-
Nov 07	5.0	37.2	1.8	18.4	65.3	240.0	54.9	3.7	349
Nov 08	4.3	15.2	2.5	8.4	64.5	79.4	64.4	2.5	769
Nov 09	4.6	19.0	2.2	8.4	49.7	78.7	62.8	8.1	1522
Nov 10	4.1	19.1	1.7	6.1	37.3	67.3	20.2	9.6	1372
Nov 11	-	-	-	-	-	-	0.0	-	-
Nov 13	-	-	-	-	-	-	0.0	-	-
Nov 14	3.9	12.7	2.7	6.3	54.4	57.4	35.5	20.2	5486
Nov 15	-	-	-	-	-	-	99.1	-	-
Nov 16	2.4	10.9	1.7	5.2	43.0	63.0	22.0	12.1	2606
Nov 17	3.4	16.6	1.4	6.3	53.8	83.7	27.9	4.8	1269
Nov 18	6.7	23.1	2.9	11.3	69.4	99.2	54.8	6.9	1058
Nov 19	-	-	-	-	-	-	2.5	-	-
Nov 20	7.5	29.4	3.1	13.8	52.4	119.2	7.2	62.4	2980
Nov 21	7.8	17.8	6.1	10.1	82.9	90.5	37.0	37.5	6045
Nov 22	6.1	31.3	2.6	18.6	66.4	264.6	14.6	3.4	402
Nov 23	10.8	20.2	4.8	13.0	103.4	117.9	49.1	7.8	1695
Nov 24	7.7	21.0	6.7	21.5	101.0	101.0	6.8	0.1	274
Nov 25	6.3	20.4	4.2	17.6	78.6	106.1	44.7	2.4	429
Nov 26	4.7	15.6	3.2	8.2	83.5	100.4	42.9	14.7	4269
Nov 27	3.6	8.7	3.7	5.8	68.8	48.0	86.6	0.9	488
Nov 28	5.6	21.3	5.4	11.8	98.1	144.0	40.5	11.7	2563
Nov 29	-	-	-	-	-	-	97.7	-	-
Nov 30	-	-	-	-	-	-	99.0	-	-
Dec 01	6.3	17.0	4.4	14.5	95.3	103.2	52.2	26.6	2364
Dec 02	3.6	15.9	1.8	8.4	50.7	81.6	58.6	3.5	546
Mean*	5.0	18.9	2.8	12.4	67.0	91.9	36.6	11.2	1991

* Mean values are weighted by number of images.

Days with cap error exceeding 100% demonstrated unfavorable phenomena, such as stationary conditions (advection performs worse than persistence) and rapidly

deforming clouds (cap error near 100%, as advection performs just as poorly as persistence). On the other hand, the cumulus clouds present on the day with the smallest cap error (Nov 10) presented heterogeneous cloud cover moving at uniform velocities, which is the ideal scenario for the cloud motion algorithm, as the sharp edges characteristic of cumulus clouds allow the cross-correlation method to obtain robust results. Statically advecting cumulus clouds conforms with the forecast assumptions, and cumulus clouds are optically thick, so few thin pixels were present in each cloud decision image, thereby minimizing the fluctuations in cloudy pixels caused by thin cloud classification. Hence, Nov 10 was an ideal day for USI forecast performance, and its average 30 sec (5 min) cap error of 37.3% (67.4%) is superior to the smallest cap errors for similarly ideal days in Section 5.2.2.1 of 45.0% (70.3%).

5.3 Variational Optical Flow (VOF)

To overcome the above challenges for estimating multiple independent and non-rigid motions, a variational optical flow (VOF) technique is evaluated. Similar nonrigid registration techniques have also been implemented successfully to estimate cloud motion on sky images last year (Bernecker et al, 2014, West et al, 2014). Optical flow techniques estimate the two-dimensional dense motion field (i.e. every image pixel) with sub-pixel accuracy between two consecutive images (Szeliski, 2010). The variational technique is used in optical flow to minimize an objective function composed of a data model and a regularization term (Horn and Schunck, 1981). The objective function can be solved by well-founded and optimized numerical methods due to the theory of the calculus of variations. As a result of the flexibility of the data modelling process, VOF became a

popular technique for motion estimation for fluid imagery such as satellite meteorological images (Corpetti et al., 2002, Héas et al. ,2007, 2008) and experimental fluid mechanics (Corpetti et al, 2006, Heitz et al, 2010).

5.3.1 Motion estimation

The fundamental assumption behind optical flow is that an image pixel values does not change over consecutive frames, but only shift position. Mathematically,

$$I_t(x_t, y_t) = I_{t+1}(x_t + u_t(x_t, y_t), y_t + v_t(x_t, y_t)), \qquad 5.3$$

where I represents pixel values, such as color (R, G and B) or gray scale intensity, x_t and y_t are the Cartesian pixel indices, and $u_t(x_t, y_t)$ and $v_t(x_t, y_t)$ are the motion vector components for pixel (x_t, y_t) in frame t. The goal is to compute the optical flow field $\{u_t(x_t, y_t), v_t(x_t, y_t)\}$ between two successive frames of an image sequence. The brightness constancy equation is a nonlinear equation in u and v. To simplify the nonlinear equation and solve for the optical flow field, the equation is linearized by a first order Taylor expansion leading to the well-known optical flow constraint (OFC) equation

$$0 = \frac{\partial I}{\partial t} + \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v. \qquad 5.4$$

The OFC equation is often violated in a realistic cloud scene due to changing illumination, occlusion, non-Lambertian reflectance, etc. For example, clouds in proximity to the sun are whiter than in other locations due to forward scattering and this leads to significant deviations from the brightness constancy assumption. Therefore, to remove the sun-pixel-angle dependence of pixel intensity, a ΔRBR is used in this dissertation to represent I to correct for background heterogeneity. Clouds are known to

leave stronger signatures in the red channel (Shields et al., 2013) and normalization by the blue channel contributes to normalizing out general brightness deviations.

The OFC is an ill-posed problem, i.e. an under-determined system that has one equation with two unknowns, u and v, for which a unique motion cannot be recovered locally without additional constraints. It suggests the system has insufficient information to recover a reliable estimate of the optical flow field. Intuitively, if we rewrite the OFC equation into vector form,

$$0 = I_t + \nabla I \cdot (u, v), \tag{5.5}$$

we can see that merely the flow normal to the gradient can be determined because if (u, v) satisfies the equation, so does $(u+j,\ v+k)$, if $\nabla I \cdot (j, k) = 0$. Therefore, a second assumption and information from spatial neighbors is needed to provide a unique solution.

An early approach to handle the OFC problem, known as the local method, was proposed by Lucas and Kanade (1981). They evaluated the OFC equation within a neighborhood where the flow field is assumed homogenous. Nevertheless, choosing an appropriate neighborhood size is challenging and this approach does not solve the ambiguity in homogenous regions. Horn and Schunck (1981) proposed another method based on the assumption that the optical flow field is similar locally and spatially smooth for the whole image. They minimized an objective function by using brightness constancy and global smoothness as model assumptions in a variational method.

In a variational method, model assumptions $S_1 + \cdots + S_m$ are formulated in terms of an energy functional

$$E\big(f_1(\omega), \dots, f_n(\omega)\big) = \int (S_1 + \dots + S_m)d\omega \qquad 5.6$$

and the functions $f_1(\omega), \dots, f_n(\omega)$ should minimize the energy E, where $\omega = (x, y)$ denotes a point in the image domain. In this research, the algorithm by Liu (2009) is used due to its simple implementation, relatively low computational intensity, and flexibility in parameters. The global deviations from the brightness constancy and smoothness assumption are measured by

$$E(u, v) = \int \varphi\big(I_t(x_t, y_t) - I_{t+1}(x_t + u, y_t + v)\big)d\omega \qquad 5.7$$

$$+ \alpha \int \varphi(|\nabla_2 u|^2 + |\nabla_2 v|^2)\, d\omega,$$

where α is a parameter that weighs the second term (regularization term) relative to the OFC term, ∇ is the gradient operator, and φ is a robust function (Black et al., 1996). The regularization term models the spatial smoothness of the optical flow and penalizes high variation across an image. The goal is to find the optical flow field u and v that minimizes E.

In addition to the variational approach, a multi-scale approach is used to avoid local minima of energy. If displacements between two images are large, the first order Taylor expansion of the brightness constancy equation becomes invalid and the solution of the energy function 5.7 can be trapped in a local minimum. To avoid such situations, the multi-scale approach initializes the energy minimization in 5.7 on a coarse scale to find the global minimum and propagates the solution gradually to the finer scale (). Finally, the theory of the calculus of variations leads to a system of Euler-Lagrange equations and they are solved by successive over-relaxation (SOR) numerical

approximation. Details on minimization of the energy and the numerical approximation can be found in Brox (2005) and Liu (2009).

To shift the cloud map with the motion vector field the optical flow method obtains pixel positions with a heterogeneous flow field $u_t(x_t, y_t)$ and $v_t(x_t, y_t)$. In general, there are two ways to warp an image: forward and inverse mapping. Let $U(x, y)$ and $V(x, y)$ be a mapping by the optical flow field between coordinate (x, y) and (x', y'). For forward mapping, each coordinate pair (x, y) in the source image is copied to the output image location (x', y') or in vector notation as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} U_f(x, y) \\ V_f(x, y) \end{pmatrix}. \tag{5.8}$$

The primary limitation of forward mapping is that it produces holes and overlapped pixels in the output image that need to be handled with interpolation and averaging. Therefore, inverse mapping is used to warp the image and is expressed as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} U_i(x', y') \\ V_i(x', y') \end{pmatrix}. \tag{5.9}$$

Inverse mapping scans through each output pixel and samples the correct input pixel from the source image. In this way, every pixel in the output image is guaranteed to be mapped to a pixel from the source image.

5.3.2 Cloud forecasting

To evaluate the VOF forecast, the VOF forecasts of the binary cloud decision of a sky image transformed to Cartesian coordinates (hereinafter cloud map) is compared against the CCM forecasts described in section 5.2.1. The nowcast (i.e. 0 minute forecast) is obtained by shifting the cloud map at time $t_0 - dt$ ($dt = 30$ s throughout this analysis)

with the motion vector field determined from the VOF and CCM applied to the images at time t_0 and $t_0 - dt$ (Fig. 1). While the nowcast performance is not of practical relevance since "future" information is used in generating the nowcast, it serves as a useful benchmark since the assumption of cloud speed persistence is not required. The forecast skill (FS)

$$FS = 1 - \frac{e_{m,VOF}}{e_{m,CCM}}$$ 5.10

is defined to measure the improvement in the matching error of the VOF forecast compared to the CCM forecast. Positive values of FS indicate that the VOF forecast is superior to the CCM forecast, with a maximum possible value of one. Since the two advection methods produce spatially different forecast maps, only the common points are compared in both metrics.
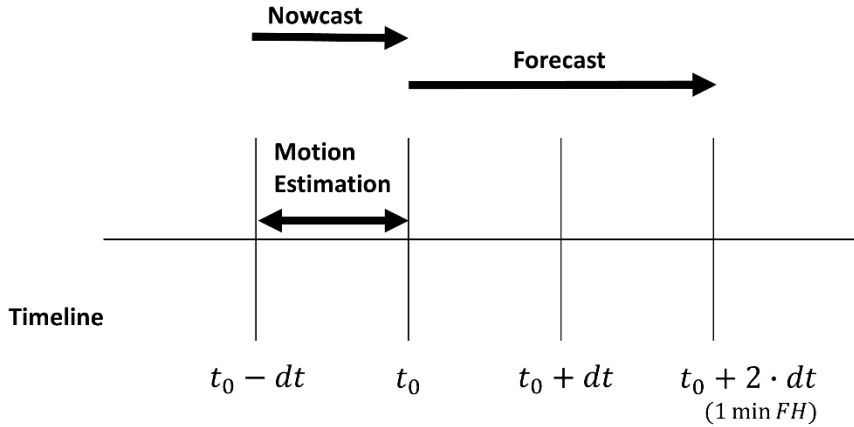
Figure 5.5: Timeline for nowcasting and forecasting the cloud map. $dt = 30$ s throughout this dissertation.

5.3.2.3 Variational optical flow forecast (VOF forecast)

Figure 5.6b shows an example image demonstrating the ability of the optical flow technique to capture the non-uniform cloud motion. The cloud in the center moves about

10% faster than the clouds in the east and south areas of the forecast domain. Table 5.5 shows a summary of the CCM and VOF forecast performance for each partly cloudy day in November 2012 using the error metrics introduced in section 5.2.2 and 5.3.2. For the 0 min forecast (nowcast), FS is positive for all days, which confirms that the VOF method is better able to estimate the motion than CCM.

Since the average FS for 5, 10, and 15 min horizons is 0.21, 0.19, and 0.19 respectively, the VOF forecast performed better than CCM forecast on average. The VOF forecast is superior to the CCM forecast for 5 and 10 min FHs for all days. The VOF forecast also outperformed image persistence forecast ($e_{cap,VOF} < 100\%$) for the 5-15 min FHs for more than 18 out of 20 days, but less than half of the days for the CCM forecast. For the 15 min forecast, the cloud map is always advected out of 70% of the sky imager field-of-view on 8 days (resulting in missing entries in Table 5.5) revealing the general limitation of sky imager forecast for FH beyond 15 minutes. For most days, the FS decreases with increasing FH. We believe that errors not related to cloud motion equally increase the matching error of both methods with longer FH, resulting in a larger matching error ratio and smaller FS. Examples of such errors are cloud decision errors and violations of the frozen cloud map assumption, such as perspective errors (Yang et al, 2014).

For the remainder of the section only the VOF technique is considered and the ability to track points with the VOF technique is exploited to quantify correlations between cloud forecast accuracy and cloud stability.

Table 5.5: Summary of cap errors and forecast skill for 0, 5, 10, and 15 min forecast horizon for 20 days in November 2012. The errors are only computed under partly cloudy conditions (5% < cloud fraction < 95%) only. Asterisks mark cases when the cloud map was advected more than 70% out of the USI field-of-view on average.

	$e_{cap,CCM}$[%]				$e_{cap,VOF}$[%]				FS[-]			
	Forecast Horizon [min]											
Date	0	5	10	15	0	5	10	15	0	5	10	15
11/01	59.5	91.2	109	118	37.4	59.4	75.9	85.7	0.37	0.35	0.30	0.27
11/02	96.6	110	121	123	54.4	68.5	76.9	84.3	0.44	0.38	0.36	0.31
11/03	112	102	100	100	77.0	85.9	95.0	111	0.31	0.16	0.05	-0.11
11/07	93.5	253	*	*	70.9	89.0	*	*	0.24	0.65	*	*
11/08	63.3	87.6	112	125	42.0	61.1	76.1	87.3	0.34	0.30	0.32	0.30
11/09	50.3	76.1	80.1	*	28.7	58.4	65.7	*	0.43	0.23	0.18	*
11/10	35.4	63.4	*	*	23.8	53.6	*	*	0.33	0.16	*	*
11/14	44.8	60.8	73.9	87.7	24.5	40.5	52.5	65.5	0.47	0.30	0.23	0.16
11/16	44.1	60.3	69.8	56.9	24.3	45.9	58.8	50.4	0.45	0.24	0.16	0.11
11/17	55.4	86.6	10	12	33.4	63.3	82.3	94.2	0.40	0.27	0.23	0.23
11/18	102	101	*	*	58.6	82.5	*	*	0.43	0.18	*	*
11/20	74.9	97.3	99.6	*	70.5	83.4	89.6	*	0.06	0.14	0.10	*
11/21	76.4	98.4	96.7	*	49.2	65.2	72.4	*	0.36	0.34	0.25	*
11/22	96.0	429	*	*	62.3	198	*	*	0.35	0.54	*	*
11/23	100	100	100	100	76.4	80.0	81.9	82.1	0.24	0.20	0.18	0.18
11/24	100	101	99.4	97.5	76.9	78.6	85.0	84.4	0.23	0.22	0.14	0.13
11/25	93.0	311	388	*	58.6	126	197	*	0.37	0.59	0.49	*
11/26	87.4	130	126	133	47.6	71.3	82.3	94.1	0.46	0.45	0.35	0.29
11/27	47.0	46.5	43.2	45.7	26.9	29.4	28.3	31.0	0.43	0.37	0.35	0.32
11/28	226	469	345	382	58.9	61.8	64.6	68.1	0.74	0.87	0.81	0.82
Mean									0.39	0.21	0.19	0.19

5.3.2.4 Sensitivity to VOF smoothness

The optical flow estimation result with two different smoothness parameters α (Eq. 5.7) is shown in Figure 5.7. As expected, more spatial variations of cloud motion and sharper motion boundaries are observed for a smaller α. The larger α yields a smoother flow field with a tradeoff of possible less accurate local flow estimation.

VOF forecast from a less smoothed flow field ($\alpha = 0.01$) is shown to perform better for FH of 1 and 5 min for Nov 10 and Nov 14, respectively (Figure 5.7). The less smoothed flow field better captured the small-scale motion near the cloud edge and

forecasts more accurately for a very short forecast horizon. Yet, the physical lifetime of these small scales is usually only a few minutes after which they become unpredictable leading to deviations from the frozen cloud map assumption and larger forecast errors as the forecast horizon increases. Especially for the cumulus case (Nov 10) cloud turbulence is stronger and small scales are expected to have shorter life times leading to the earlier cross-over between the different smoothness parameters in Figure 5.7. This implies that the optimal cloud motion field should be smooth enough to avoid extrapolating localized motion beyond its lifetime, yet able to capture spatially heterogeneous cloud motion. Therefore, a FH-dependent smoothness parameters may be advantageous to yield the optimum forecast over all time horizons, at least for days with boundary layer clouds that are associated with turbulent motion on a large range of time scales. However, the implementation of such a procedure would require significant fine tuning and was beyond the scope of this paper and $\alpha = 0.1$ was used for the remainder of manuscript.
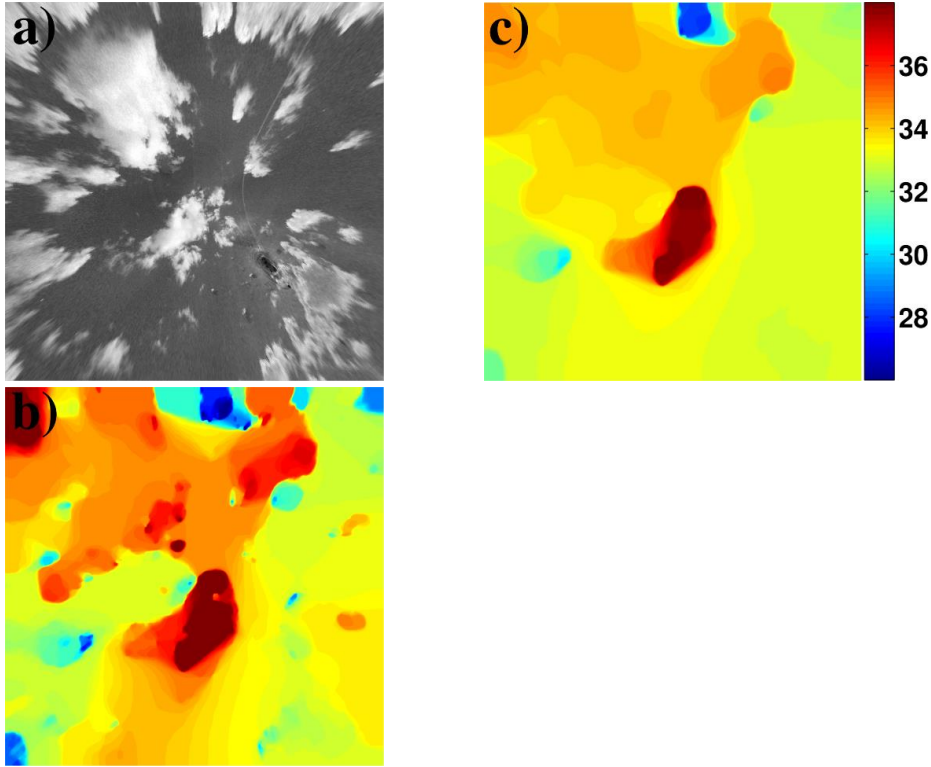
Figure 5.6: Optical flow estimation of the ΔRBR RBR image on November 10th, 2012, 08:46:00 PST out to zenith angles of 75° (a) with spatial smoothness $\alpha = 0.01$ (b) and $\alpha = 0.1$ (c). The colorbar indicates the motion magnitude in pixel per frame.
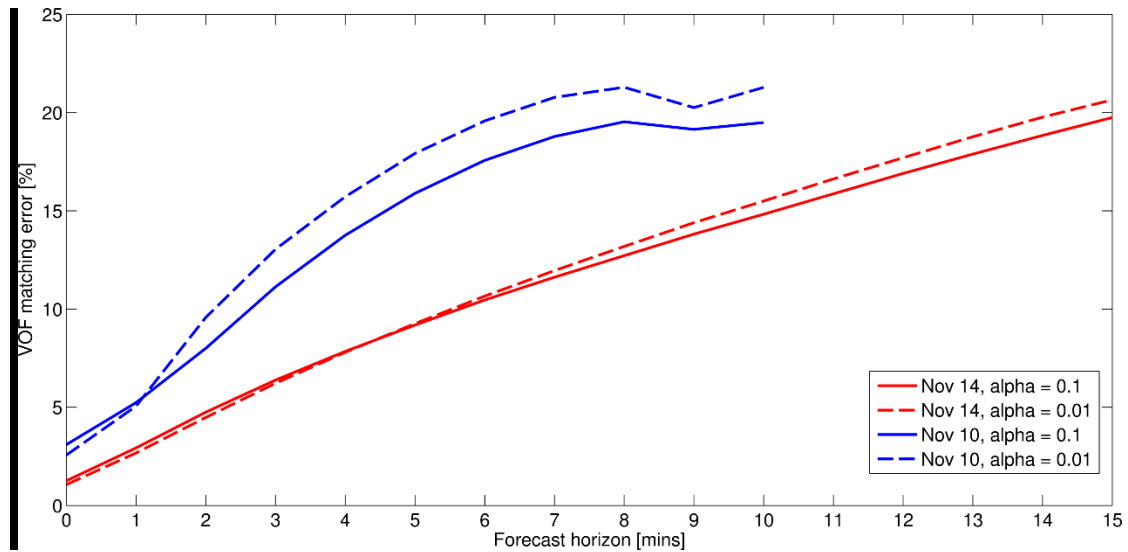
Figure 5.7: Cloud forecast performance for Nov 10 (cumulus) and 14, 2012 (cirrus clouds) with smoothness $\alpha = 0.01$ and 0.1. Cloud matching error for FH beyond 10 min is not shown for Nov 10 since more than 70% of the cloud map is advected out of the USI field-of-view.

5.4 Cloud stability

Even though many advanced techniques to estimate cloud motion exist, little attention has been paid on cloud stability, i.e., how rapidly a cloud is changing, which is a key challenge to "frozen" cloud map advection. While time series of cloud fraction and brokenness in the sky imager field-of-view provide information about the cloud cover variability, changes in these metrics are often dominated by the Eulerian framework (i.e. the advection of clouds in and out of the sky imager field-of-view) and present little information on cloud stability in a Lagrangian sense.

In most cloud advection forecast models, cloud features are assumed constant over the forecast horizon. The validity of this assumption is scale-dependent. Over the sub-30 minute forecast horizon of ground-based sky imagery, this assumption often holds for synoptic and even mesoscale cloud systems but is usually violated for individual clouds or small scale features. From highly granular imagery, clouds – especially those located in the atmospheric boundary layer – have often been observed to significantly deform, evaporate, and develop over time scales of a few minutes in the San Diego coastal area (Chow et al, 2011, Yang et al, 2014). Cloud dynamics are driven by cloud and boundary layer turbulence as well as topographic effects and present challenges to deterministic cloud forecasting. Therefore, a method to identify such circumstances and quantify cloud stability is highly desired.

Temporal invariance of cloud features is a characteristic of cloud stability. For that reason, we propose to establish a forecast confidence metric based on dynamic image features and the optical flow field extracted from the VOF method to infer cloud stability and the validity of the frozen-cloud advection technique. In fact, dynamic features have

been shown to be of importance in many applications such as object segmentation (Brox and Malik, 2010), cloud classification and synthesis (Liu et al, 2013), and camera calibration (Jacobs et al, 2013).

5.4.1 Point trajectories and forecast confidence

Point trajectories are obtained by developing an optical flow tracker based on Sundaram et al. (2010). Tracking points are initialized (sub-sampled) every twenty pixels from the first frame of an image sequence. Points located in homogenous (e.g. clear or overcast) regions are difficult to track and are therefore removed. Homogeneous regions are identified by the second eigenvalue of the structure tensor, $J_o = \sum_{i=1}^{2} \nabla I_i \nabla I_i^T$ (hereinafter image structure)(Sundaram et al., 2010). Each of the points is then tracked using the optical flow field

$$(x_{t+1}, y_{t+1}) = (x_t, y_t) + (u_t(x_t, y_t), v_t(x_t, y_t)). \qquad 5.11$$

Tracking of a point is terminated if one of the following three circumstances is encountered:

1. Point is advected out of the forecast domain.

2. Forward and backward optical flow yield inconsistent results (Sundaram et al., 2010). Tracking is stopped if the inconsistency is larger than a threshold, which varies as a linear function of motion magnitude.

3. The image structure around the trajectory point decreases. The local image structure can capture the dynamics of the cloud evolution as, for example, cloud evaporation decreases the local RBR gradient. However, image structure can also

decrease due to measurement errors or optical effects; for example, clouds moving into the solar region appear to have less structure due to pixel saturation.

Lastly, to fill the empty areas due to terminated trajectories, new tracks are initialized in unoccupied areas in each new frame. The trajectory length represents the duration of a pixel point that stays in an image sequence, and the average time length of terminated trajectories for a frame, $\bar{T}_{len}$, is used to quantify cloud stability. Since clouds entering the field-of-view of the sky imager are by default associated with a shorter trajectory length, a marginal space / transition zone is used between the image domain and region of interest (see Figure 5.8 later) and the trajectories in the transition zone are not considered in $\bar{T}_{len}$. Furthermore, the minimum value of $\bar{T}_{len}$ is 1 min as newly initialized tracks are not considered in the average.

The tracking result of an image sequence of cumulus clouds on Nov 10 is illustrated in Figure 5.8. A set of points on a regular grid was initialized in areas with high image structure (see beginning of this section). High image structure exists primarily in partly cloudy conditions, while image structure is small in clear sky region and homogenous cloud areas (Figure 5.8a). Points were tracked from frame to frame according to the optical flow field to construct point trajectories. In Figure 5.8b, nearly half of the points had been tracked for 5 min (light green color), while the remaining points have a shorter trajectory length (blue) due to cloud break-up (yellow box of Figure 5.8b) and formation and entering the sky imager field of view. Only a small fraction of points were tracked for 7.5 min (Figure 5.8c) and two clouds were merging to form a larger cloud (yellow box of Figure 5.8c). In this image sequence, the longest point

trajectories (orange color) are associated with the largest cloud from the first frame. Generally, large clouds were observed to have longer point trajectories presumably since they are thicker on average and therefore contain more liquid water. Therefor cloud evaporation and deformation requires more energy and time, which results in less change of appearance.

The tracking result of an image sequence of cirrus clouds on Nov 14 is shown in Figure 5.9. This image sequence presented temporally stable clouds without deformation and with uniform motion across the sky. Only a small fraction of points were terminated and not tracked to for 7.5 min (Figure 5.9c).
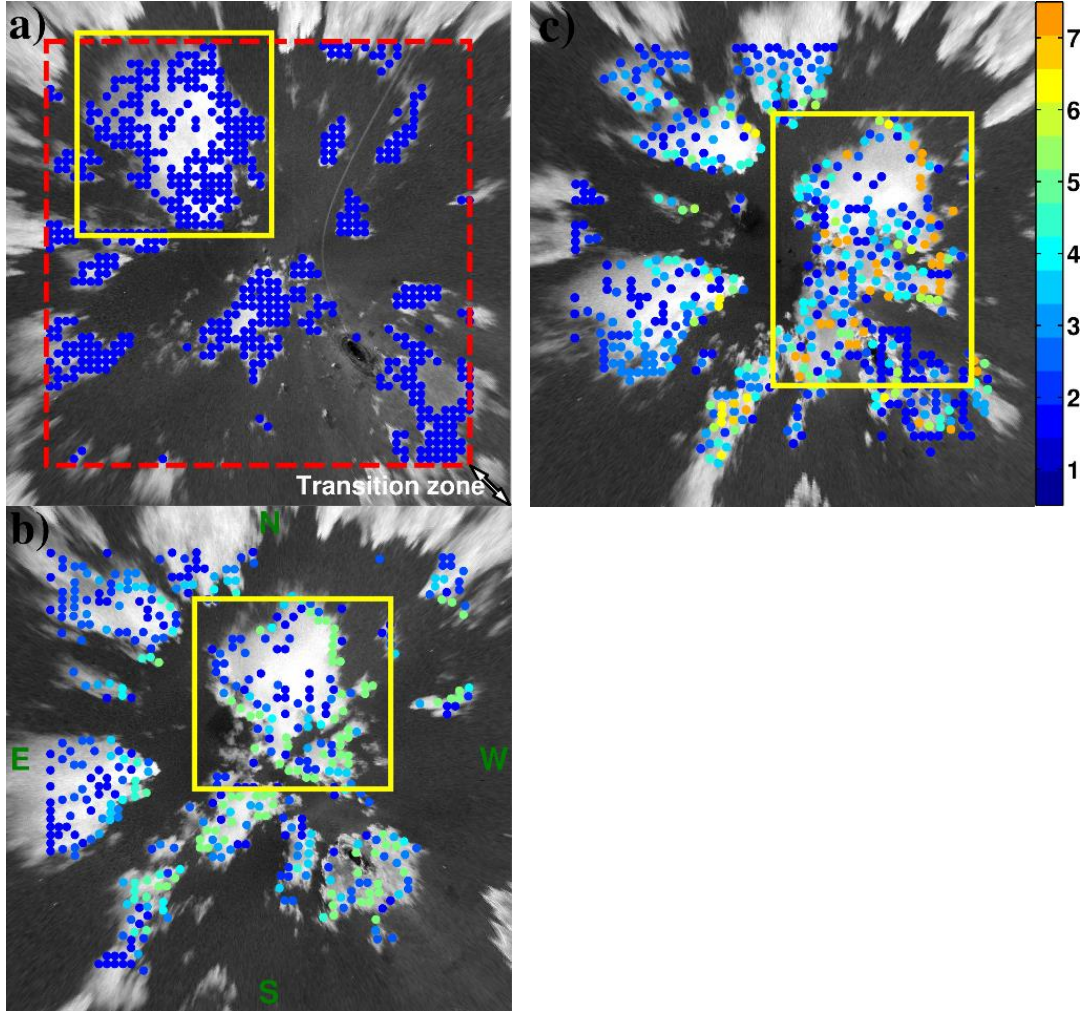
Figure 5.8: A set of points is initialized with a sub-sampled grid on a ΔRBR RBR image (a) and tracked for 5 min or 10 frames with $\bar{T}_{len} = 3.18\ min$ (b) and 7.5 min or 15 frames with $\bar{T}_{len} = 3.79\ min$ (c) on November 10th, 2012. For ease of following the cloud motion the largest cloud is demarcated by the yellow box. While the sequence illustrated here begins at 16:46:00 UTC, the point trajectory method is implemented at the beginning of each day when solar zenith angle (SZA) $<75°$. Trajectories are initialized everywhere but $\bar{T}_{len}$ (colorbar) is not counted in the transition zone (as indicated by the red dashed box) to avoid $\bar{T}_{len}$ being biased low from new clouds entering the forecast domain. The colorbar indicates the point trajectory length in minutes.
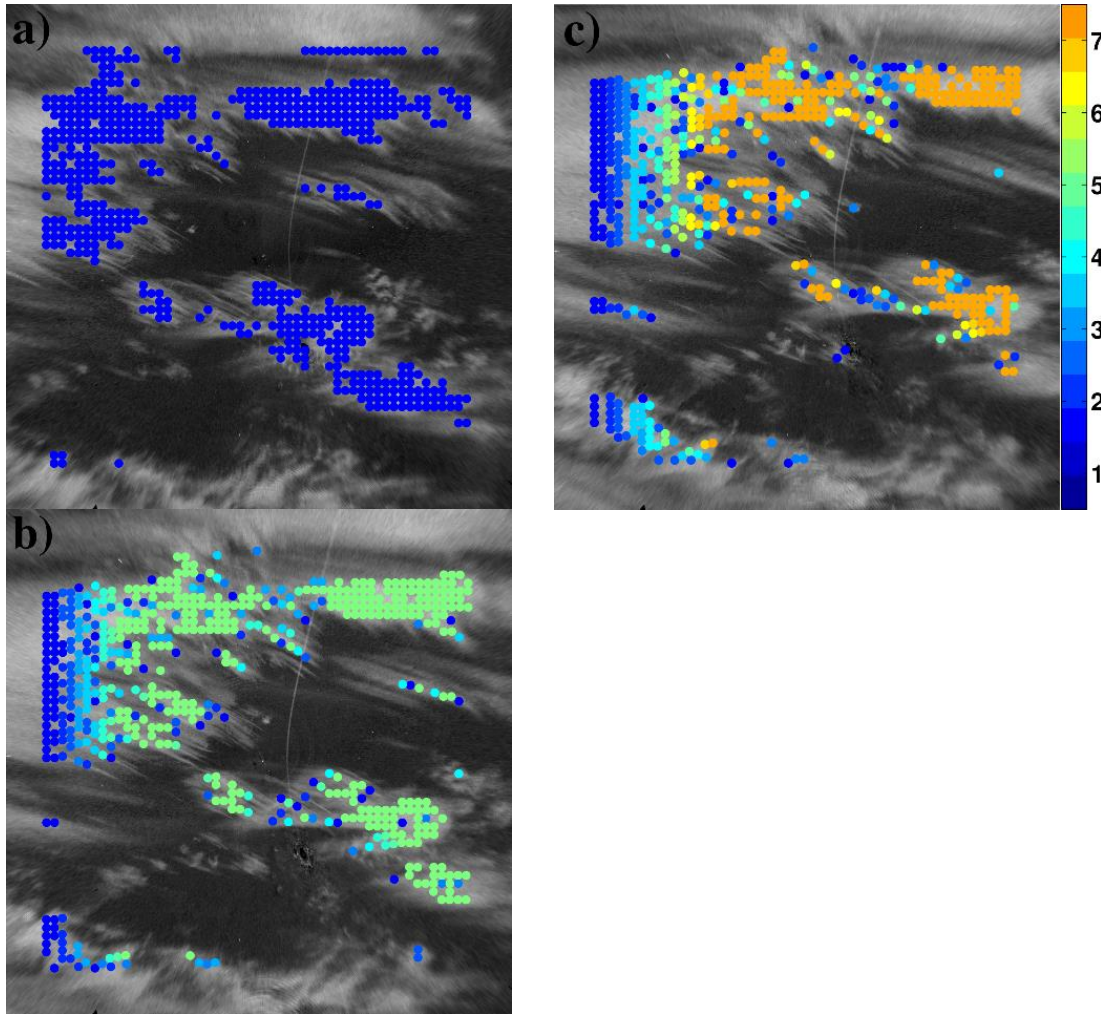
Figure 5.9: A set of points is initialized (a) and tracked for 5 min or 10 frames with $T_{len} = 4.38$ min (b) and 7.5 min or 15 frames with $T_{len} = 6.71$ min (c) on November 14th, 2012 at 18:00:00 UTC. Cirrus clouds were observed to be stable with almost constant image structures that yield long trajectory points.

Using the point trajectory technique, point trajectory lengths were sampled at the forecast issue time on the 20 days listed in Table 5.5. To investigate the relationship between point trajectories lengths and cloud stability, time series plots of VOF matching error, cloud fraction, and average trajectory length $\bar{T}_{len}$ for 5 min forecast of two different days are presented in Figure 5.10. While the cloud matching error is presented at forecast valid time, the trajectory length sampled at forecast issue time is shifted to the

forecast valid time to facilitate comparison. November 7 (Figure 5.10a) illustrates a day with poor forecast accuracy (cap error of 89%). Starting at 1900 UTC highly variable cumulus clouds are observed in the sky images (not shown here) which is reflected in the short trajectory length ($\bar{T}_{len}$ = 1.18 min). Therefore, image persistence and VOF forecasts performed similarly on this day.

On the other hand, on November 8 multiple cloud layers are observed. Between 1600 - 1800 UTC stable altocumulus and cirrus clouds brought forth a $\bar{T}_{len}$ between 2 and 10 min (Figure 5.10b). Cloud-advection outperformed persistence forecast during this period as indicated by the smaller matching error. Shallow cumulus clouds along with altocumulus clouds were predominant from 2000 UTC until 2300 UTC. $\bar{T}_{len}$ is between 1 and 5 min and persistence and cloud advection forecasts performed similarly. From 2300 UTC until sunset, single layer altocumulus clouds with a trajectory length around 5 min allowed cloud advection to outperform persistence again. These observations indicate a correlation of the point trajectory length with cloud stability and cap error.
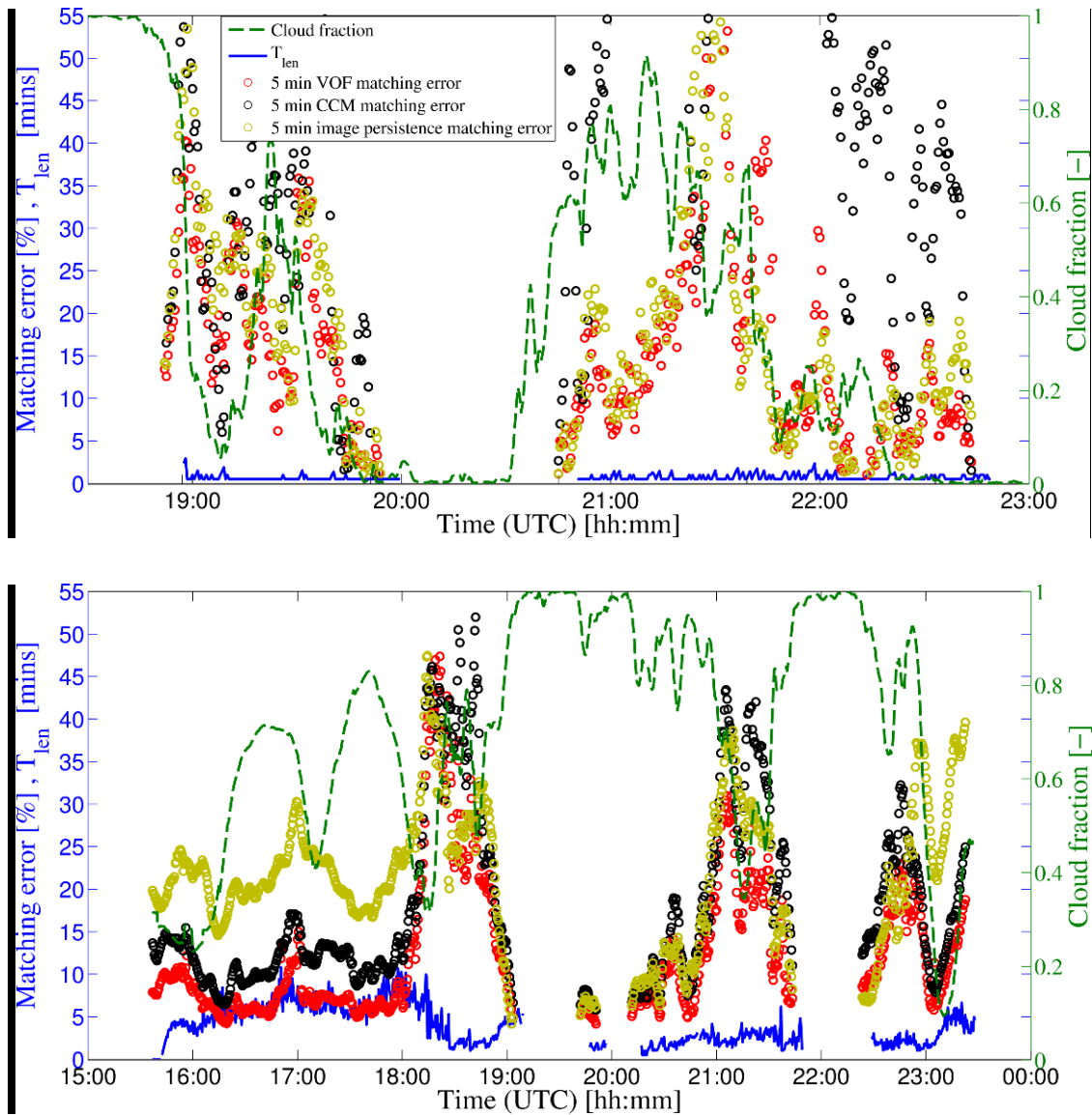
Figure 5.10: Time series of cloud matching error e_m for VOF and image persistence, average trajectory length $\bar{T}_{len}$, and cloud fraction on November 7, 2011 (a), and November 8, 2011 (b). Large and small cloud fractions are typically associated with small matching errors.

To determine the general applicability of trajectory length as a forecast confidence metric for an operational forecast setting, boxplots of 5 and 10 min cap error against average trajectory length for all days are presented in Figure 5.11. The results indicate that both 5 and 10 min cap errors decrease as $\bar{T}_{len}$ increases. For 5 min forecasts, VOF forecast began to perform 50% better (as measured by the median) than image

persistence for trajectory lengths of 5 min. For 10-min VOF forecast to outperform 50%
persistence, a minimum length of 6 min is required. It is expected that the critical
trajectory length for VOF to outperform persistence increases with forecast horizon, as
the frozen cloud assumption must hold over the forecast horizon. While the critical
trajectory length may be expected to be larger than the forecast horizon, it is important to
note that many trajectory points may terminate due to small scale deformation of the
cloud, while the large scale feature of the cloud will remain intact and is most relevant for
cap error.

Table 5.6 summarizes daily 5 min cap errors along with $\bar{T}_{len}$. For the 8 out of 20
days that have cap error greater than 75% , the average $\bar{T}_{len}$ is 1.53 min, while it
averages 3.93 min for the remaining days. The results indicate that trajectories are longer
($\bar{T}_{len} > 2$ min) for days when cloud advection improves 25% over image persistence
forecast confirming the hypothesis that short trajectories imply unstable clouds. However,
the daily averaged values may not be representative since cloud conditions often vary
during the day (e.g. Nov 8 in Figure 5.10b).
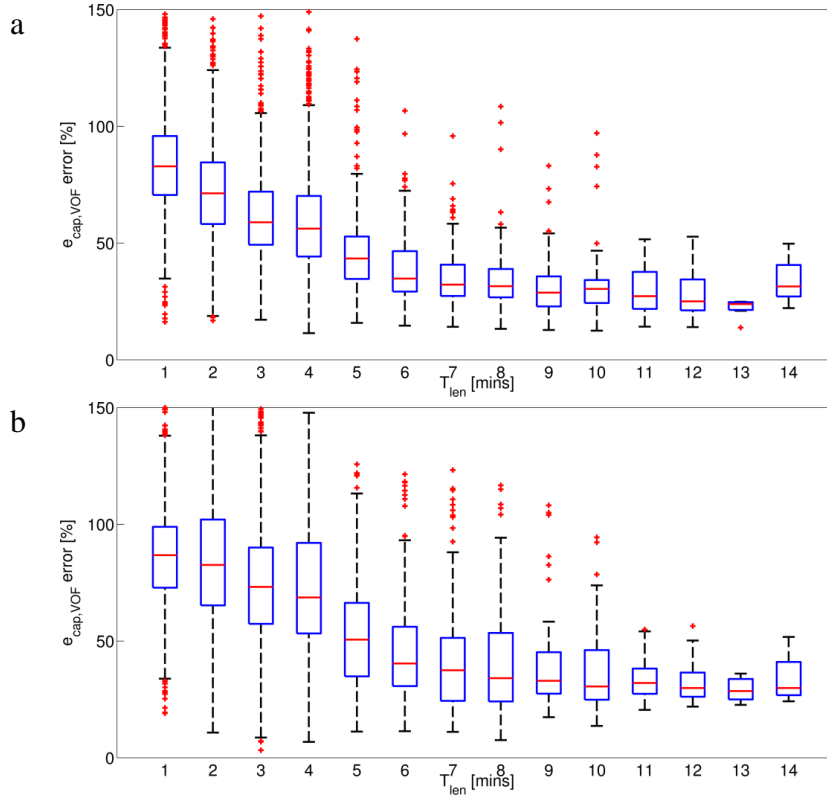
Figure 5.11: Boxplot of VOF cap error as a function of average trajectory length T_{len} for 5 min (a) and 10 min (b) forecast horizons. Red lines and dots indicate the median and outliers respectively. The top and bottom of the box shows the 75th and 25th percentile. The end of the whisker marks the value of the data set that are within 1.5 times the interquartile range.

Table 5.6: Daily 5-min cap errors of VOF forecasts and average trajectory length. No trajectory length greater than 1 min is observed on Nov 28 due to false cloudy sky caused by inaccurate cloud decision.

	$e_{cap,VOF}$ [%]	$\bar{T}_{len}$ [min]		$e_{cap,VOF}$ [%]	$\bar{T}_{len}$ [min]
Nov 01	59.4	3.13	Nov 18	82.5	1.56
Nov 02	68.5	2.41	Nov 20	83.4	1.17
Nov 03	85.9	1.16	Nov 21	65.2	2.76
Nov 07	89.0	1.18	Nov 22	197.5	1.35
Nov 08	61.1	4.05	Nov 23	80.0	1.04
Nov 09	58.4	3.02	Nov 24	78.6	1.05
Nov 10	53.6	2.95	Nov 25	126.1	2.62
Nov 14	55.8	4.57	Nov 26	71.3	3.33
Nov 16	45.9	4.74	Nov 27	29.4	8.57
Nov 17	63.3	3.72	Nov 28	61.8	NaN

5.5 Conclusion

CMVs as a basis for advecting such a frozen cloud field derived from ground-based sky imagers were developed. Similar to Hamill and Nehrkorn (1993), the cross-correlation method (CCM) was applied to obtain an average CMV for the entire cloud field. Forecast cloud field position is obtained by shifting the cloud decision images along the corresponding motion vector.

Despite these hardware limitations of the TSI, cloud-advection-versus-persistence (cap) errors were found to be a universal metric to compare forecast performance; for 30 second forecast, a 50-60% reduction in forecast error compared to persistence was found. The error would be further reduced if it were spatially averaged over a large solar array rather than computed on a pixel-by-pixel basis. Matching forecast errors increase with cloud speed and forecast horizon.

Techniques for dense motion estimation and point trajectories were presented. A month of image data captured by a sky imager at UC San Diego was analyzed to determine the accuracy of variational optical flow (VOF) forecasts and infer cloud stability. The VOF method not only resulted in better motion estimation (0 min forecasts), it was also able to produce accurate cloud forecasts (>0 min) by capturing multiple independent cloud motions while maintaining a spatially smooth motion field across an image. The VOF forecast was found to be superior to CCM forecast for with an average error reduction of 39%, 21%, 19%, and 19% for 0, 5, 10, and 15 min forecasts respectively. While image persistence outperformed VOF forecast for forecast horizons of 5 and 10 min on 2 out of 20 days these days were associated with highly variable clouds that make any cloud advection approach challenging.

The VOF analysis demonstrated that unstable clouds make accurate cloud motion forecasts impossible and such conditions need to be identified to quantify forecast confidence. Cloud stability was successfully quantified using point trajectory lengths by tracking points in an image sequence to form trajectories. Trajectory lengths were correlated with forecast cap errors for both daily averaged metrics and individual forecast realizations. Consequently, short trajectory lengths are associated with large cap error of VOF forecast (or low confidence), while long trajectory lengths and small matching errors were related to high cloud stability. The VOF forecast was found to be 50% superior to image persistence forecast for 5 and 10 min forecast horizons with a minimum trajectory length of 5 and 6 min respectively. Point trajectory length was proven to be a valuable forecast confidence and cloud stability metric that can provide information on the applicability of the frozen cloud map assumption at forecast issue time.

Acknowledgements

Publication: Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C. W., Ghonima, M.,
& Kleissl, J. (2014). Solar irradiance forecasting using a ground-based sky
imager developed at UC San Diego. Solar Energy, 103, 502-524.

Co-authors: Handa Yang, Ben Kurtz, Andu Nguyen, Bryan Urquhart, Mohamed
Ghonima, and Jan Kleissl

Publication: Chow, C. W., Belongie, S., & Kleissl, J. (2015b). Cloud motion and stability
estimation for intra-hour solar forecasting. Solar Energy, 115, 645-655.

Co-authors: Serge Belongie and Jan Kleissl

6 Cloud Patterns

6.1 Introduction

In this chapter, a technique for characterize and predict hours-ahead solar forecast error using image features or cloud patterns is presented. Hours-ahead solar forecasting is important for grid operators to manage intra-day solar variability through procurement of reserves and operation of energy markets. As Geostationary Operational Environmental Satellite (GOES-15) was deployed few years ago and the more advanced GOES-R will be launched in 2016, the ability of solar forecasting using satellite imagery is expected to improve and to be widely used by forecast providers.

Many forecast techniques have been developed for solar irradiance using satellite imagery and Numerical Weather Prediction (NWP) (e.g. Hammer et al, 2000, Perez et al., 2010, 2011, Marquez, Pedro, and Coimbra, 2012). Perez et al (2010) showed that the satellite cloud motion-based forecasts improve upon NWP forecasts up to forecast horizons of 4-5 hours on average over the United States. For operational forecasting, however, the relative skill of different forecasts depends on meteorological conditions. The motivation of this chapter is to determine this "cross-over time" under different conditions and determine parameters (cloud cover fraction, cloud speed, etc.) that are good predictors for this cross-over time.

For NWP, good predictors of solar irradiance forecast error are clear sky index and solar zenith angle (Lorenz et al. 2009, Mathiesen and Kleissl 2011). For satellite forecasts no published results on predictors of error exist. Knowledge of the cross-over time and determining parameters will allow the composition of a forecast across all time

horizons, which is based on optimal combinations of inputs from different models. For example, a satellite model would be chosen as the forecast if it had been determined to be most accurate for frontal passages with homogeneous cloud cover and persistent cloud motion. More accurate and certain solar forecasts would result that facilitate more economical high solar penetration on the electric grid.

6.2 Data

6.2.1 Goes-15 Imager data

The Geostationary Operational Environmental Satellite-15 (GOES-15) Multispectral Imager satellite dataset from April 24th to May24[th] is used. GOES-15 generates a full disk image every 15 minutes with a spatial resolution of 1km and 4km for the visible and infrared bands, respectively. All of our results are presented at 4 km resolution. The domain of interest covers an open ocean region with an approximate size of 234 km x 470 km from $32°22'$ to $34°31'$ N and from $122°62'$ to $127°62'$ W (Figure 6.1). Open ocean area is selected to avoid complex terrain effects that introduce stationary (mountain) clouds and high spatial variability in satellite measured reflectivity. GOES-15 images starting at 1645 and 1700 UTC (09 PST) are used to determine cloud speed and the 1700 UTC clear sky index image is advected with hourly time steps out to 6 hours forecast horizon.
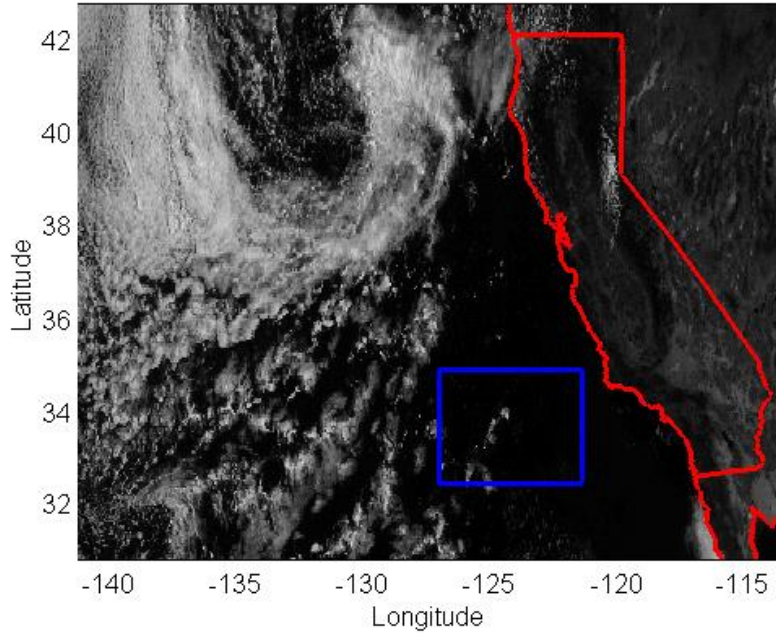
Figure 6.1: The GOES-15 visible channel image that shows the region of interest (blue) with the coast and California state lines (red).

6.2.2 AVHRR Pathfinder Atmospheres-Extended (PATMOS-x) algorithms

The GOES-15 data is processed with the Clouds from AVHRR Pathfinder Atmospheres-Extended (PATMOS-x) algorithm developed by the GOES-R Advanced Baseline Imager (ABI) Cloud Algorithm Working Group (AWG) to provide cloud-related information. PATMOS-x generates a cloud mask and classifies clouds into different types such as water, supercooled water, mixed phase, cirrus, opaque ice, etc (Pavolonis et al 2004, 2005). Among over thirty variables, cloud transmission and cloud mask with 4 km spatial resolution from the PATMOS-x are used. Assuming the atmosphere without clouds is transparent to the global horizontal irradiance (GHI), cloud transmission can be approximated as the clear sky index defined as:

$$kt = \frac{GHI_{measured}}{GHI_{clear\ sky}},$$ 6.1

where $GHI_{clear\ sky}$ is computed using a clear sky model that depends on location, time, elevation, and Linke turbidity (Ineichen and Perez, 2002). Since the spatial resolution and coordinate system is different between GOES-15 and NAM data. GOES-15 data is up-scaled and collocated with the NAM by simple averaging and linear interpolation. While the author is writing this dissertation, the Global Solar Insolation Project (GSIP) from the National Oceanic Atmospheric Administration (NOAA) is publicly available as an operational model and generates GHI from the GOES visible and infrared channel measurement as one of the outputs. The analysis in this chapter is done before GSIP become available.

6.2.3 North American Mesoscale (NAM) Model

The total downward short wave radiation (GHI) NAM forecast is published by the National Oceanic and Atmospheric Administration's (NOAA) NCEP with a 12 km x 12 km grid spanning the continental United States (CONUS). Forecasts published at 12 UTC (shortly before sunrise at the region of interest) with hourly time steps out to 36 hours are used. Forecast kt is computed as in Eq. 6.1. Since satellite forecast issue time is 1700 UTC (Section 6.2.1), the sixth hour of the NAM forecast is compared to the first hour of satellite forecast and so on. Since the NAM forecast error varies only weakly in time for the first 24 hours (Perez et al., 2009), the difference in forecast horizon is not expected to impact the results.

6.3 Methods

6.3.1 Cloud Motion Vectors (CMV) Forecast

Atmospheric motion vectors (AMV) have been derived from geostationary satellites since the 1960s (Fujita 1968) and different tracers and meteorological scales have been selected to derive AMV. In this paper, the cloud transmission or kt is used as a tracer. The cross-correlation method (CCM) described in Section 5.2 and two consecutive (15 min apart) kt images are utilized to derive cloud motion vectors (CMVs). A 15 x 15 pixels (60 x 60 km) target box is used to cross-correlate within a 100 x 100 km search box (Figure 6.2). This limits the maximum detectable velocity to 200 km h^{-1}. To increase the computational efficiency, CCM only computed at 12 km resolution (every 3 pixels).

Assuming kt and CMVs remain constant over the forecast horizon, computed pixel-by-pixel CMVs are used to advect the kt pixels to each forecast horizon (Figure 6.2). Nearest neighborhood interpolation is applied on sections of the forecast kt image where gaps and overlapping pixels are introduced by converging and diverging flow.

Two quality control procedures are applied. First, CMVs in cloud-free regions (as per the PATMOS-x cloud mask) are removed. Second, AMVs that significantly differ from the neighboring vectors are replaced using a median filter technique. These vectors might represent errors in the cross-correlation algorithm or (less likely) depict actual local (microscale) motion associated with convective clouds (Mecikalski 2002; Rabin et al. 2004).
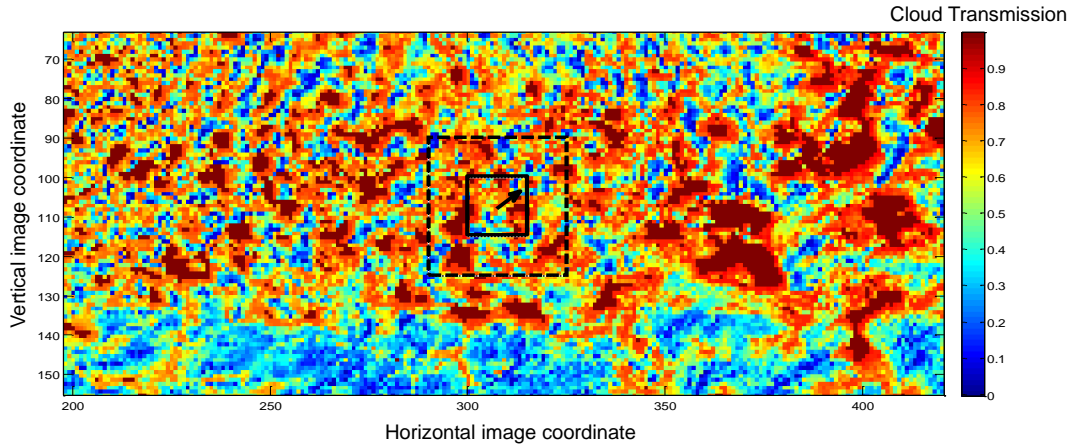
Figure 6.2: Illustration of CCM on an image of cloud transmission (4 km resolution) at 10 PST on 7th June, 2012. A 60 x 60 km pixels target box (solid black line) is cross-correlated within a 100 x 100 km pixels search box (dash black line) from a cloud transmission image 15 minutes later (not shown here). A cloud motion vector is obtained for the center pixel of the target box.

6.3.2 Cloud Patterns

Seven different regional features (Table 6.1) extracted from the forecasted kt maps are used to characterize the CMV and NAM forecast error. For CMV forecast, spatially averaged cloud speed, kt (AVGKT), correlation coefficient (CC) obtained as described in section 3.1, standard deviation of kt (std), cloud fraction (Cf from PATMOS-x), entropy, and uniformity are computed over the domain (Figure 6.1) to represent the state of atmosphere (Figure 6.3). All features are computed on the forecast kt image to account for likely local conditions at the forecast valid time. For NAM, the same forecast features are computed except cloud speed, std, and CC.

Table 6.1: 7 cloud related patterns are extracted from the kt map to characterize forecast error.

Cloud Features	Equation	Description and Relevance
Cloud fraction (Cf)	$\dfrac{\#\ of\ cloudy\ pixels}{\#\ of\ pixels}$	An image with a very small or very large cloud fraction (clear or overcast) is easier to forecast
Cloud Speed	$\sum_{i=1}^{n} p(cs_i)cs_i$	cs_i is the cloud speed of each pixel, p is the probability of cs_i, and n is the number of pixels in an image. An image with higher cloud speed would be more difficult to forecast since error in cloud direction will cause clouds to diverge faster from their actual position.
Correlation Coefficient (CC)	$\sum_{i=1}^{n} p(cc_i)cc_i$	Spatially averaged cross-correlation coefficient measuring the degree of correspondence between the 1645 and 1700 images that are used to determine CMVs (Haralick et al, 1992). An image with smaller CC indicates larger shape or cover changes in the cloud field making CMVs less accurate and violating the assumption of a 'frozen' cloud field. Reduced forecast accuracy would result.
AVGKT	$\sum_{i=1}^{n} p(kt_i)kt_i$	Spatially averaged cloud transmission.
Std	$\sqrt{\sum_{i=1}^{n} p(kt_i)(kt_i - \langle kt_i \rangle}$	Average contrast of an image (see Figure 6.3). For the same Cf an image with larger std would be more difficult to forecast.
Entropy	$\sum_{i=1}^{n} p(kt_i)log_2 p(kt_i)$	Randomness of an image (see Figure 6.3). For the same std an image with larger Entropy would be more difficult to forecast.
Uniformity	$\sum_{i=1}^{n} p(kt_i)$	Uniformity in image pattern (see Figure 6.3). For the same Cf an image with smaller Uniformity would be more difficult to forecast.
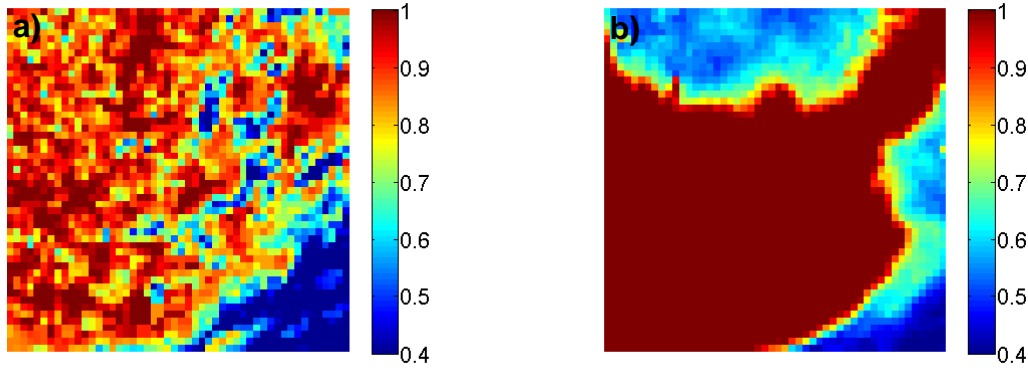
Std:	0.19	0.19
Entropy:	0.134	0.07
Uniformity:	0.0122	0.069

Figure 6.3: Illustration of two different kt (colorbar; blue: thick clouds, red: clear) patterns. Both (234 x 470 km) images have the same average contrast (std) but different entropy and uniformity. The high entropy kt image (a) shows a random cloud field, while the low entropy kt image (b) shows a clear sky with a two well defined clouds.

6.3.3 Multivariate linear regression

To find out the important determinants of forecast error, a multivariate linear regression analysis of forecast error versus different combinations of features is performed as in

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_i x_i + \epsilon. \qquad 6.2$$

y and x_i represent the RMSE of the forecast and features at a specific forecast hour respectively, β_i is the fit coefficient solved for in the regression analysis, and ϵ is the error term representing the mismatch between the linear model and the observations. The coefficient of determination (R^2) quantifies the goodness of fit and is defined as the ratio of the variance of the regression model and the variance of the observations:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2} \times 100\%, \qquad 6.3$$

where $\hat{y}_i$ is the mean modeled y (6.2) and $\bar{y}$ is the mean measured y.

6.3.4 Analog scheme

To address the limitations of the multivariable linear regression model, an artificial intelligence (AI) method called analog method is also used to estimate the forecast error. The analog of a forecast for a given location and time is defined as a past prediction that matches selected features of the current forecast. Analyzing historical forecasts errors for the analogs, forecast error can be inferred (DelleMonache et al. 2011).

The search for analog forecast is controlled by a K-nearest neighborhood (KNN) algorithm consisting of a weighted normalized Euclidean distance metric that ranks past predictions by how similar they are to the current forecast

$$\|E\| = \sqrt{\sum_{i=1}^{N} \frac{(F_i - A_i)^2}{\sigma_i}},$$

6.4

where F_i is the current forecast feature value; A_i is the feature value in an analog, N is the number of features, σ_i is the standard deviation of each feature respectively.

K analogs (K = 10 here) with the shortest Euclidean distance are then used to predict the error of the current forecast. A weighted average of the analog errors is computed as

$$AN = \sum_{i=1}^{K} \gamma_i E_i$$

6.5

$$\gamma_i = \frac{\frac{1}{\|E_i\|}}{\sum_i^K \|E_i\|}$$

6.6

where AN is the weighted average error, E_i is the analog distance in (3), and γ_i is the weight that is inversely proportional to the distance of the analog from the forecast. Therefore, a higher weight will be assigned to the analog with the shortest distance from the current forecast.

6.4 Results

6.4.1 Forecast performance

The performance of CMV forecast is compared to the basic persistence forecast that equates future with current weather condition, i.e. the current kt satellite image is assumed to persist. Forecasts are validated against the truth satellite kt image. Choosing satellite data as a validation source may artificially increase the NWP and decrease the satellite error, since identical (satellite) and different (NWP) methods are used to determine kt. Comparing forecast results to (independent) ground irradiance measurements would be preferable, but measurement sites are not available at our oceanic site and would also be very sparse over land. The root mean square error between the forecast and truth image is given as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(kt_f - kt_t)^2}{N}},$$

6.7

where kt_f is the forecast kt, kt_t is the truth kt, and n is the number of pixels in an image.

Figure 6.4 shows the RMSE of kt at different forecast horizons for each forecast model. The NAM forecast error is shown to vary weakly with forecast horizon, which confirmed our earlier assumption. The CMV forecast error increases with increasing forecast horizon as expected. The persistence forecast errors are slightly larger than CMV

forecast errors up to a 6 hours forecast horizon. The cross-over time between satellite CMV and NAM forecast is found to be in between forecast horizons of 6 hours and 1 day, which is longer than in Perez et al. (2010) probably because Perez used ground measurements for validation.
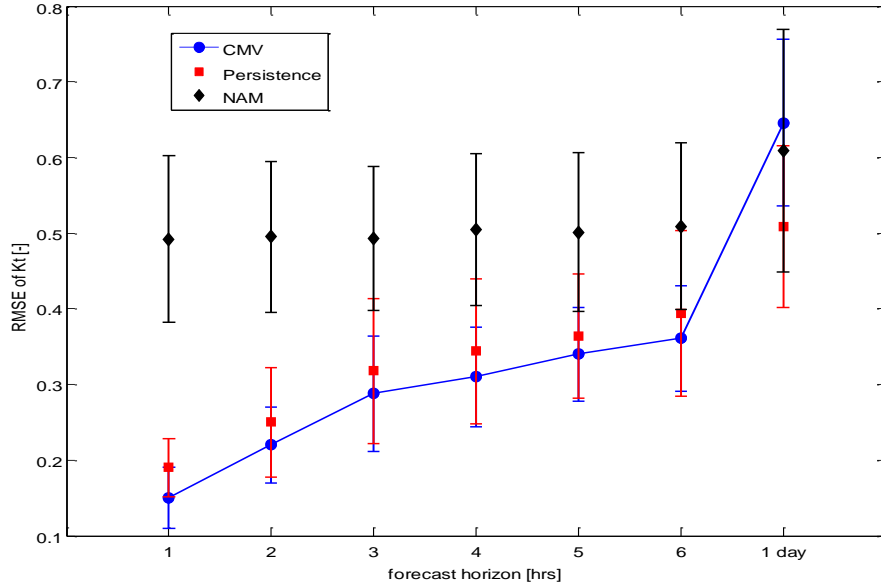
Figure 6.4: RMSE of kt as a function of forecast horizon for each forecast model. Error bars show the standard deviation of the RMSE. Note the jump from 6 hours to 1 day on the x axis.

6.4.2 Satellite forecast RMSE versus kt standard deviation, cloud speed, and cross-correlation coefficient

To motivate the analysis in Section 6.3.3, dependencies of RMSE on a few different features are shown in Figure 6.5 and Figure 6.6. The features are computed over 20x20 km (5x5 pixels) rather than the whole region described in Section 6.2. Figure 6.5 shows that RMSE increases with increasing std for all forecast horizons. While RMSE also increases with increasing cloud speed, the increase is smaller (~0.04) than for the std (~0.1). Dependencies between the RMSE and a combination of two features are

examined in Figure 6.6; RMSE increases with increasing std and decreasing CC. The sensitivity of RMSE to CC increases with std.
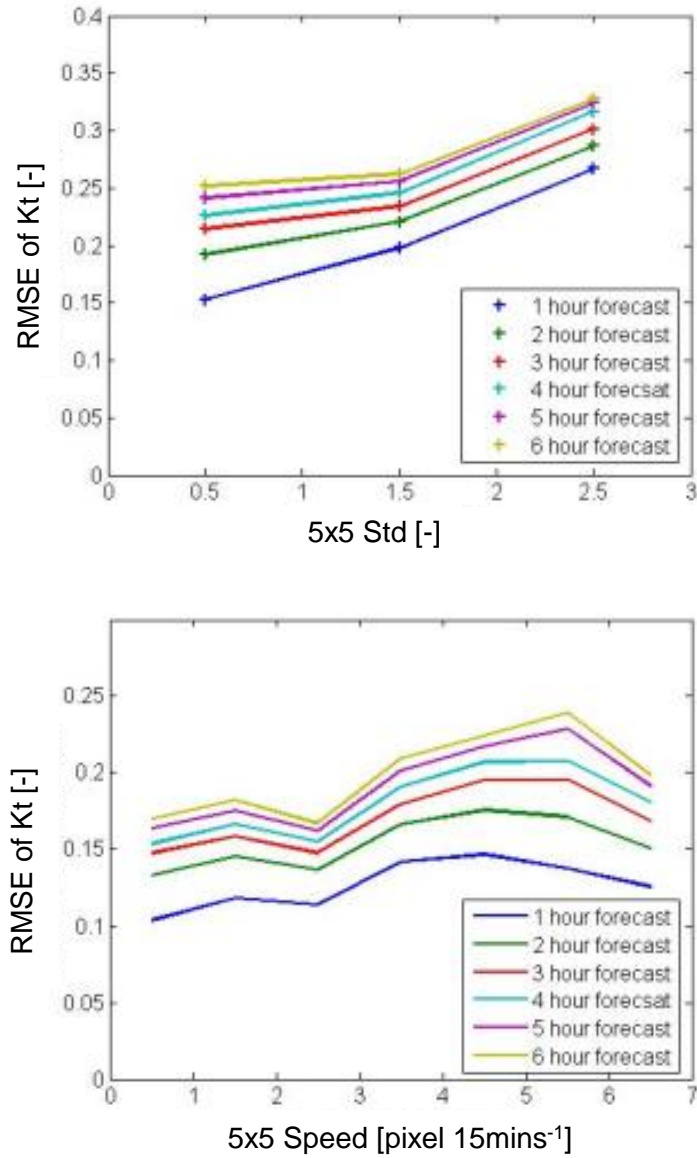
Figure 6.5: RMSE of kt as a function of std (left) and cloud speed (right) with different forecast horizons (colors).
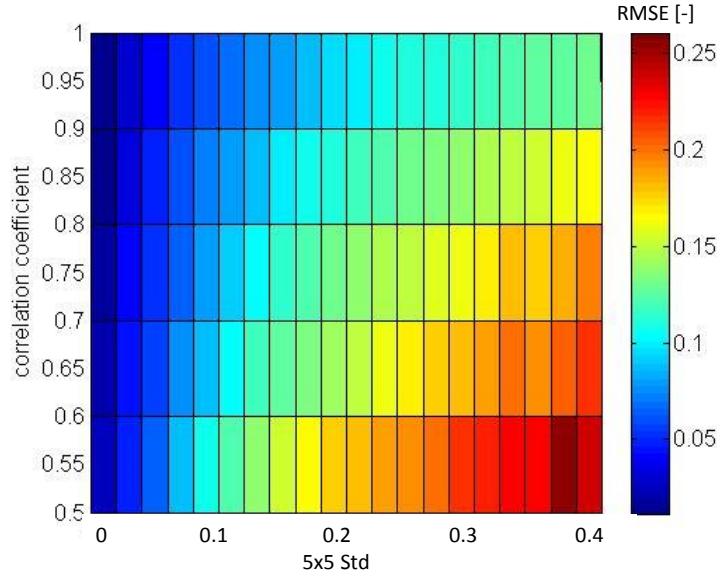
Figure 6.6: RMSE of kt (colorbar) as a function of std and CC for the 1st hour of CMV forecast.

6.4.3 Multivariate linear regression for satellite forecast RMSE

Figure 6.7 shows the coefficient of determination (R^2) for several multivariate linear models (Eq. 6.2) at different forecast horizons, each representing a different combination of features. R^2 describes the goodness of multivariate linear fit of RMSE. For the first forecast hour RMSE modeled with a single feature, the satellite forecast RMSE is most correlated to Entropy (R^2=73.4%), followed by Uniformity (R^2=65.7%). For two features, the highest correlated combination of features is Entropy and cloud fraction with a 3% increase (R^2=76.4%) compared to the single feature using Entropy. R^2 further increases to 83% with a combination of Entropy, Uniformity and Cf or AVGKT. Adding extra variables only slightly increases R^2 to 85.4%. The R^2 for single and double feature model has a similar trend for forecast horizons of 1-3 hours. While cloud speed replaces Cf or AVGKT in triple variable models to become the highest correlated

combination, the difference in R^2 is not significant. Entropy and Uniformity are the most important features for 1-3 hour forecasts.

At the fourth to sixth hour of forecast, the combination of important features changes (Figure 6.7 bottom). Entropy is still the feature with the largest R^2. The combination of CC, AVGKT, and Entropy or Uniformity is found to have the highest correlation. Since entropy and uniformity are correlated and cloud fraction is correlated to AVGKT, the main difference between the 1-3 hours and 4-6 hours models is due to the CC. CC has a very low R^2 in a single variable model but has a larger impact on the forecast RMSE for longer (> 3 hours) forecast horizon.
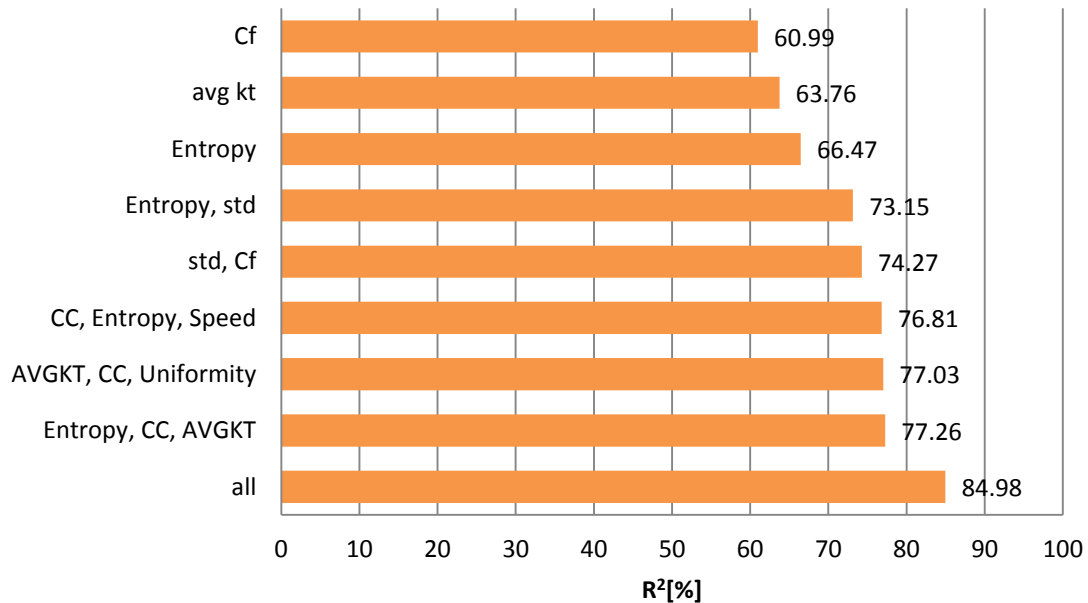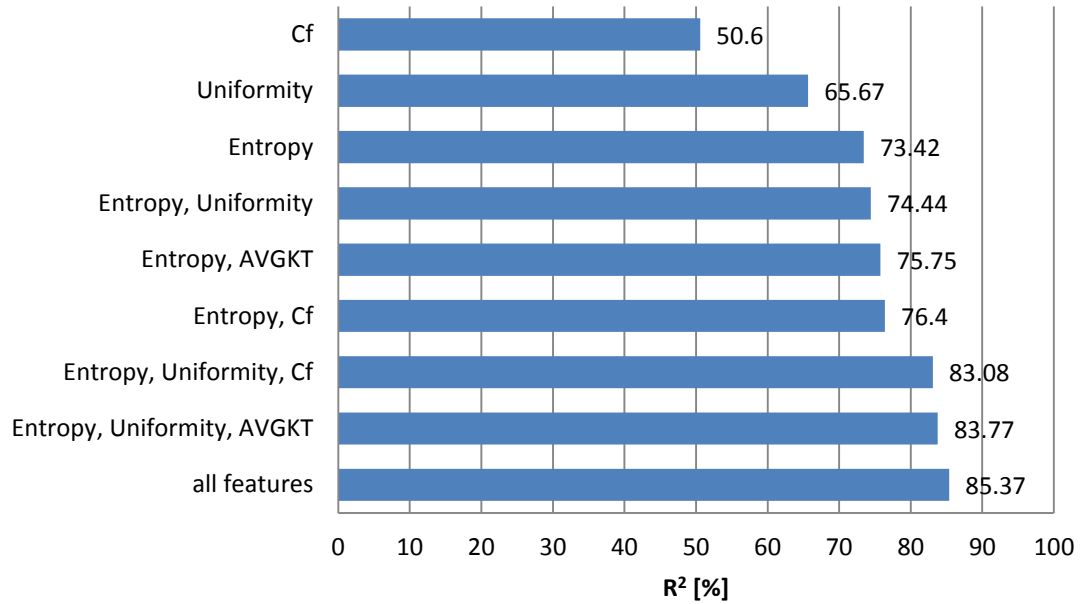
Figure 6.7: Coefficient of determination (R2) for different combination of features for the 1st hour (top) and the 5th hour of CMV forecast (bottom).

The multivariate regression analysis suggests that Entropy (i.e. the randomness of the kt field, Figure 6.3) is the primary predictor CMV forecast RMSE. A high randomness typically indicates that the cloud pattern is made of many small scale clouds

or clouds with high spatial variability of optical thickness. Since smaller scale clouds tend to have a smaller 'half-life' than large clouds, the assumption of a 'frozen' cloud field in the CMV method is violated, resulting in larger forecast error.

6.4.4 Predicting satellite versus NAM performance

To demonstrate the performance of the analog algorithm, forecasts for May 25th - June 25th, 2012 are analyzed where analogs are extracted from the April 24 – May 24, 2012 period. The RMSE predicted by the weighted analog in Eq. 6.5 is validated against the truth RMSE. The relative RMSE (rRMSE) of the analog model is computed as

$$RMSE_{model} = \frac{1}{RMSE_{mean}} \sqrt{\frac{\sum_{i=1}^{n}(\text{RMSE}_{predicted} - RMSE_{truth})^2}{N}} \times 100\% \qquad 6.8$$

where $RMSE_{mean}$ is the mean of the truth RMSE.

The analog scheme is not restricted by the linearity assumption and able to model the relative performance between CMV and persistence or NAM forecast under different conditions described by the features. The relative performance (RP) is defined as:

$$RP_{NAM,i} = \frac{RMSE_{NAM,i}}{RMSE_{CMV,i}}, \qquad 6.9$$

where i indicates forecast horizon. The idea of RP is similar to the cross-over time mentioned earlier. RP describes the relative performance of two models in terms of RMSE at a particular forecast horizon: RP < 1 indicates that CMV performs better than the NAM forecast, and vice versa. Under certain conditions, a cross-over time (RP = 1) earlier than the one in Figure 6.4 can be found (Figure 6.8).
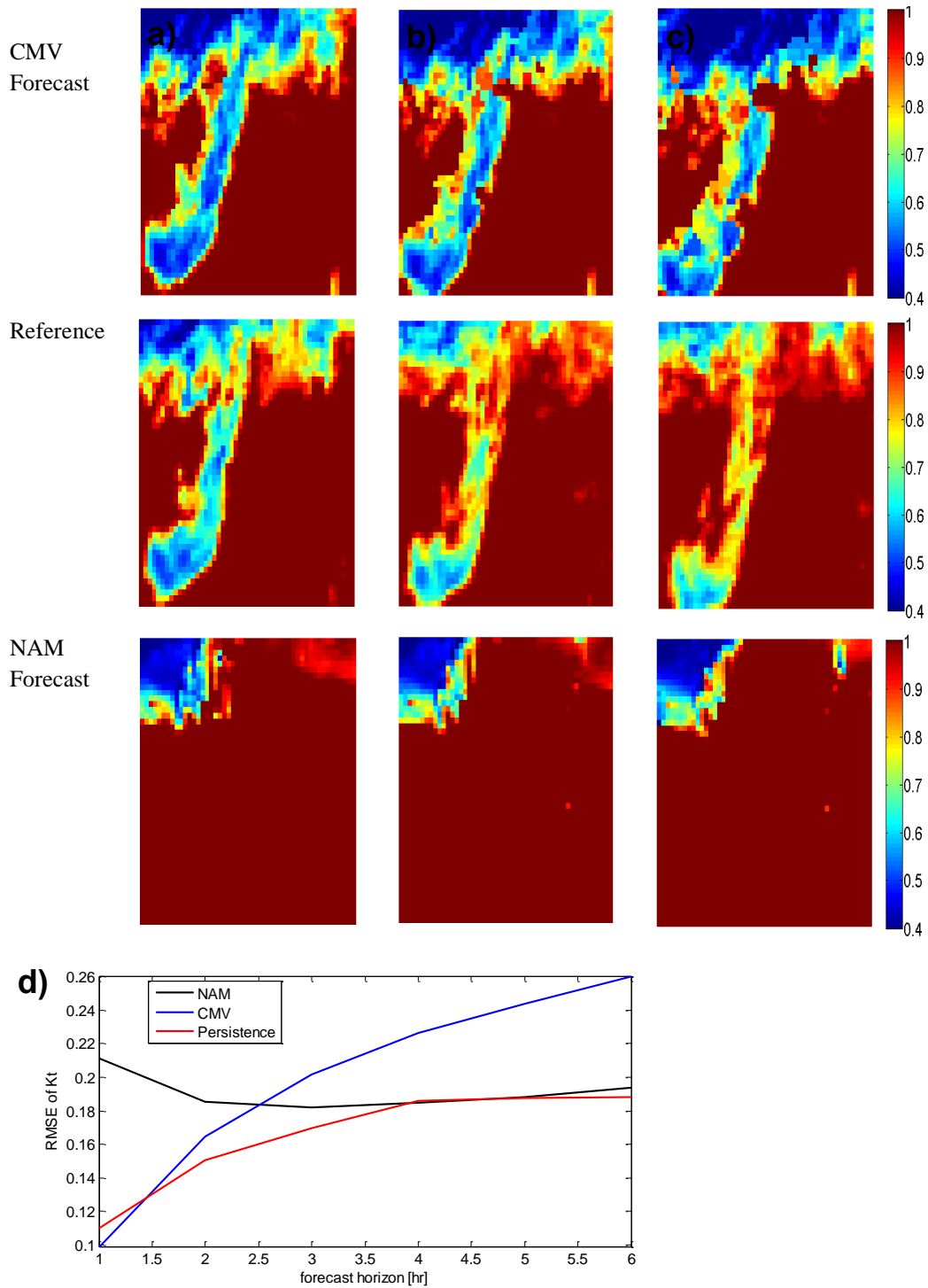
Figure 6.8: Images of kt at forecast horizons of (a) 1 hour (b) 2 hours (c) 3 hours for CMV and NAM forecasts and reference Kt on 27th May, 2012. The area shown is 234 x 470 km. In this case, the dissipation of clouds near the top and in the center and a phase-shift (i.e. change in cloud speed) of the north-south cloudy stripe led to a rapid increase in CMV forecast error and cross-over with the NAM forecast at 2.5 hours (d).

Table 6.2 shows a comparison of the performance of the multivariate regression and analog method in predicting CMV forecast error and RP. The analog method is shown to outperform the multivariate linear regression method with best combination of features (largest R^2) for both predicting CMV forecast error and RP_{NAM}. The average rRMSE of predicting CMV forecast error is 14.4% and 23.9% for analog and regression method respectively, while the average rRMSE of predicting RP_{NAM} is 27.6% and 31.7%. The errors of both methods are found to be independent of the forecast horizon. Satellite image Uniformity is observed to be the most important feature for predicting the RP_{NAM}. The impact of other features is small and varies by forecast horizon. The performance of both models could be improved if more training data were available.

Table 6.2: Predicted CMV forecast error and RP with analog and multivariate linear regression model. Training data is from April 24 - May 24 and test data is May 25 – June 25, 2012.

	Predicted CMV forecast error		Predicted RP_{NAM}	
Forecast hour	rRMSE$_{analog}$	rRMSE$_{regression}$	rRMSE$_{analog}$	rRMSE$_{regression}$
1	16.5%	21.4%	31.2%	37.3%
2	13.0%	28.6%	27.2%	30.6%
3	13.5%	19.7%	20.3%	27.2%
4	15.6%	23.5%	29.6%	31.0%
5	14.0%	23.2%	30.3%	35.1%
6	13.9%	27.1%	27.1%	29.0%

6.5 Conclusion

A technique for characterizing and predicting hours-ahead forecast error using cloud patterns with multivariate linear regression and the analog model is presented. Solar forecasts based on CMVs are computed using the cross-correlation algorithm on pixel scale with PATMOS-x processed GOES-15 data for one month. On average, the

RMSE is found to increase with increasing forecast horizon for both persistence and CMV forecast and becoming large than NAM forecast error (i.e. with a cross-over time) at between 6 hours and 1 day.

The dependence of forecast RMSE on different cloud patterns is analyzed using multivariate linear regression and analog models. Entropy is shown to contribute the most to the forecast RMSE. The rRMSE of predicting CMV forecast error is found to be 14.4% for the analog model and 23.9% for the regression model. This indicates that the CMV model errors show a non-linear dependence on the selected features that can be better captured by the analog method versus a linear regression.

The relative performance of NAM and satellite CMV models is also analyzed. Relative performance decreases from about 3.5 at the 1st forecast hour to 1.5 at the 6th forecast hour, on average. The cross-over time (when numerical weather prediction forecasts become more accurate than CMV) is found to be more than 6 hours, on average, but as early as 2 hours under certain conditions. The rRMSE of predicting RP for is shown to be 27.6% and 31.7% for analog and regression method respectively. Consequently, both models have skill in informing forecasters at what forecast horizon a numerical weather prediction forecast should start to be used instead of a satellite forecast.

With only two months of available PATMOS-x data, the model performance is limited by the data size being used to train the models. Especially for the analog method a larger training data set increases the chance to find similar analogs and improve model accuracy.

7 Conclusion

The use of sky imagery to assess the solar resource for solar energy applications shows much potential for augmenting the spatial and temporal resolution provided by satellite and numerical forecasting methods. Only short deterministic forecast horizons up to 15 mins are feasible using a single sky imager at our site due to low clouds and large cloud variability at the fine spatial scale studied. Capturing these features deterministically is nearly impossible with satellites or numerical weather prediction. Based on imagery alone, the optical depth (not shown in this dissertation), percentage cloud cover, cloud stability, cloud pattern, and velocity of clouds can be approximated at sub-kilometer resolution. This information regarding the properties of the cloud field above a solar power plant should therefore serve as a valuable input for short-term irradiance forecasting.

Cloud detection using ground-based imagery provides high spatial and temporal resolution of cloud cover monitoring and solar irradiance forecasting. The clear sky library method is illustrated to be a simple and powerful technique to extract cloud automatically from sky image. However, color information at pixel scale is shown to be insufficient to detect clouds reliably, especially image region proximate to the solar disk. Saturation due to the forward scattering by the sun, noise, and change in illumination, greatly affect cloud segmentation by means of thresholding spectral information at pixel scale. These errors motivated the author to study more sophisticated image segmentation form the computer vision community. Before the author applying the method in Section 4.3, cluster analysis of long-term motion trajectories based on optical flow was attempted to segment clouds from sky images. However, the optical flow results tend to blur the

motion boundaries and hence limit the skill of the segmentation methods based solely on motion trajectories. Therefore, in this dissertation, motion information is integrated with the spatial affinity matrix based on a state-of-the-art contour detection method instead. The contour detection method is found to increase the chance of cloud being detected correctly by clustering similar spatial image structure. Lastly, a random walk algorithm is implemented with the cloud decision from the CSL method as seeds to produce the final segmentation and is shown to outperform the classical CSL method.

The analysis of image-based performance for cloud detection and versus actual images demonstrated forecast skill. However, CCM forecasts for the USI exhibited a larger cloud matching error than image persistence forecast for forecast horizon of 5 min in 11 of 22 days partially because the spatial homogeneity of the cloud motion assumption is not appropriate given cloud deformation, topographically induced wind speed variations, and the changing optical perspective. It motivated the study of estimating cloud motion using variational optical flow. The VOF forecast was demonstrated to be superior to CCM forecast. In addition, the VOF analysis confirmed that unstable clouds make accurate cloud motion forecasts impossible. Such conditions need to be identified to quantify forecast confidence and can be done by estimating cloud stability, which were successfully quantified using point trajectory lengths by tracking points in an image sequence to form trajectories.

All aforementioned results are image-based forecasts and the introduction of other variables in our forecast procedures (i.e. cloud base height and cloud optical depth), which are necessary to produce a site-specific irradiance forecast, causes larger errors for deterministic forecast. Cloud optical depth remains very difficult to obtain from sky

images due to perspective issues (side versus bottom of cloud) and the importance of three-dimensional radiative transfer. Currently, the assignment of clear sky indices to clear sky, thin, and thick clouds was based upon persistence forecasts applied to external data such as pyranometers and PV arrays. Cloud height is currently obtained from the nearby METAR stations of the forecast site and cloud height discrepancy leads to inaccurate cloud shadow location with the ray tracing method. Cloud based height is able to obtained at the forecast site with a stereographic method with two or more sky imagers. A technique to estimate cloud based height with a sky imager and multiple pyranometers is under development. After all, this dissertation shows the ability of advanced computer vision and image-processing techniques to extract cloud characteristics and the potential of utilizing these techniques to extract cloud based height and cloud optical depth and improve solar forecasting is enormous.

References

Allmen, M., Kegelmeyer, W., 1996. The computation of cloud-base height from paired whole-sky imaging cameras. Journal of Atmospheric and Oceanic Technology 13, 97–113.

Arbelaez, P. (2006, June). Boundary extraction in natural images using ultrametric contour maps. In Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on (pp. 182-182). IEEE.

Arbeláez, P., & Cohen, L. (2008, June). Constrained image segmentation from hierarchical boundaries. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (pp. 1-8). IEEE.

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour detection and hierarchical image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 33(5), 898-916.

Arbelaez, P., Pont-Tuset, J., Barron, J., Marques, F., & Malik, J. (2014, June). Multiscale combinatorial grouping. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on (pp. 328-335). IEEE.

Beaubien, M., Freedman, J., 2001. Comparison of Two Imager-based Methods for Determination of Winds Aloft. American Meteorological Society Conf.

Bedka, K. M., & Mecikalski, J. R. (2005). Application of satellite-derived atmospheric motion vectors for estimating mesoscale flows. Journal of Applied Meteorology, 44(11), 1761-1772.

Benjamin, S., Brown , J., Olson, J., Wilczak, J., Banta, R., DiMego, G., Weng, F., 2010. Improvements in NOAA modeling and data assimilation to improve guidance for the renewable energy industry. American Meteorological Society Conf.

Bernecker, D., Riess, C., Angelopoulou, E., & Hornegger, J. (2014). Continuous short-term irradiance forecasts using sky images. Solar Energy, 110, 303-315.

Black, M. J., & Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer Vision and Image Understanding, 63(1), 75-104.

Bosch, J. L., and J. Kleissl. "Cloud motion vectors from a network of ground sensors in a solar power plant." Solar Energy 95 (2013): 13-20.

Bosch, J. L., Zheng, Y., & Kleissl, J. (2013). Deriving cloud velocity from an array of solar radiation measurements. Solar Energy, 87, 196-203.

Boucher, O., and Randall, D., (2013). Clouds and aerosols. In Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment

Report of the Intergovernmental Panel on Climate Change (pp. 571-657). Cambridge University Press

Brox, T. (2005). From pixels to regions: partial differential equations in image analysis (Doctoral dissertation).

Brox, T., & Malik, J. (2010). Object segmentation by long term analysis of point trajectories. In Computer Vision–ECCV 2010 (pp. 282-295). Springer Berlin Heidelberg.

Cano D., Monget J., Aubuisson M., Guillard H., Regas N., Wald L., 1986. A method for the determination of global solar radiation from meteorological satellite data. Solar Energy 37, 31–39.

Cazorla, A., Husillos, C., Antón, M., & Alados-Arboledas, L. (2015). Multi-exposure adaptive threshold technique for cloud detection with sky imagers.Solar Energy, 114, 268-277.

Cazorla, A., Olmo, F., Alados-Arboledas, L., 2008. Development of a sky imager for cloud cover assessment. J. Opt. Soc. Am. A 25, 29–39.

Cazorla, A., Shields, J., Karr, M., Olmo, F., Burden, A., Alados-Arboledas, L. 2009. Technical Note: Determination of aerosol optical properties by a calibrated sky imager, Atmos. Chem. Phys., 9, 6417–6427.

Chen, C., Duan, S., Cai, T., Liu, B., November 2011. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Solar Energy 85 (11), 2856–2870.

Chow, C. W., Belongie, S., & Kleissl, J. (2015a). Sky image sequence segmentation for solar forecasting. In preparation.

Chow, C. W., Belongie, S., & Kleissl, J. (2015b). Cloud motion and stability estimation for intra-hour solar forecasting. Solar Energy, 115, 645-655.

Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., & Washom, B. (2011). Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy, 85(11), 2881-2893.

Chu, Y., Pedro, H. T., Nonnenmacher, L., Inman, R. H., Liao, Z., & Coimbra, C. F. (2014). A Smart Image-Based Cloud Detection System for Intrahour Solar Irradiance Forecasts. Journal of Atmospheric and Oceanic Technology, 31(9), 1995-2007.

Corpetti, T., Heitz, D., Arroyo, G., Memin, E., & Santa-Cruz, A. (2006). Fluid experimental flow estimation based on an optical-flow scheme. Experiments in fluids, 40(1), 80-97.

Corpetti, T., Mémin, É., & Pérez, P. (2002). Dense estimation of fluid flows.Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(3), 365-380.

DelleMonache, L., Nipen, T., Liu, Y., Roux, G., Stull, R., 2011: Kalman filter and analog schemes to post-process numerical weather predictions. Accepted to appear on Monthly Weather Review

Dollár, P., & Zitnick, C. L. (2013, December). Structured forests for fast edge detection. In Computer Vision (ICCV), 2013 IEEE International Conference on(pp. 1841-1848). IEEE.

Eber, K., & Corbus, D. (2013). Hawaii Solar Integration Study: Executive Summary.

Ela, E., Diakov, V., Ibanez, E., & Heaney, M. (2013). Impacts of Variability and Uncertainty in Solar Photovoltaic Generation at Multiple Timescales (No. NREL/TP-5500-58274). National Renewable Energy Laboratory (NREL), Golden, CO.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The shuttle radar topography mission. Reviews of Geophysics 45 (2).

Feister U., Shields, J., 2005. Cloud and radiance measurements with the VIS/NIR Daylight Whole Sky Imager at Lindenberg (Germany). Meteorol. Z., 14, 627-639.

Fujita, T., 1968: Present status of cloud velocity computations from ATS-1 and ATS-3. COSPAR Space Res., 9, 557-570

Fung, V., Bosch, J. L., Roberts, S. W., & Kleissl, J. (2013). Cloud speed sensor. Atmospheric Measurement Techniques Discussions, 6(5), 9037-9059.

Galasso, F., Cipolla, R., & Schiele, B. (2013). Video segmentation with superpixels. In Computer Vision–ACCV 2012 (pp. 760-774). Springer Berlin Heidelberg.

Ghonima, M. S., Urquhart, B., Chow, C. W., Shields, J. E., Cazorla, A., Kleissl, J., 2012. A method for cloud detection and opacity classification based on ground based sky imagery. Atmospheric Measurement Techniques 5, 2881–2892.

Gohari, M. I., Urquhart, B., Yang, H., Kurtz, B., Nguyen, D., Chow, C. W., ... & Kleissl, J. (2014). Comparison of solar power output forecasting performance of the total sky imager and the University of California, San Diego sky imager.Energy Procedia, 49, 2340-2350.

Grady, L. (2006). Random walks for image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(11), 1768-1783.

Hamill, T. M., & Nehrkorn, T. (1993). A short-term cloud forecast scheme using cross correlations. Weather and Forecasting, 8(4), 401-411.

Hammer, A., Heinemann, D., Hoyer, C., Kuhlemann, R., Lorenz, E., Müller, R. and Beyer, H. G. (2003): Solar Energy Assessment Using Remote Sensing Technologies. Remote Sensing of Environment 86, 423–432.

Haralock, R. M., & Shapiro, L. G. (1991). Computer and robot vision. Addison-Wesley Longman Publishing Co., Inc..

Héas, P., & Mémin, E. (2008). Three-dimensional motion estimation of atmospheric layers from image sequences. Geoscience and Remote Sensing, IEEE Transactions on, 46(8), 2385-2396.

Héas, P., Memin, E., Papadakis, N., & Szantai, A. (2007). Layered estimation of atmospheric mesoscale dynamics from satellite imagery. Geoscience and Remote Sensing, IEEE Transactions on, 45(12), 4087-4104.

Heinle, A., Macke, A., Srivastav, A., 2010. Automatic cloud classification of whole sky images. Atmos. Meas. Tech., 3, 557–567.

Heitz, D., Mémin, E., & Schnörr, C. (2010). Variational fluid flow measurements from image sequences: synopsis and perspectives. Experiments in fluids,48(3), 369-393.

Helman, U., Loutan, C., Rosenblum, G., Rothleder, M., Xie, J., & Zhou, H. (2010). Integration of renewable resources: Operational requirements and generation fleet capability at 20% rps. California Independent System Operator.

Hoff, T. E., & Perez, R. (2012). Modeling PV fleet output variability. Solar Energy, 86(8), 2177-2189.

Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. Artificial Intelligence, 17(1), 185-203.

Huang, H., Xu, J., Peng, Z., Yoo, S., Yu, D., Huang, D., & Qin, H. (2013, October). Cloud motion estimation for short term solar irradiation prediction. InSmart Grid Communications (SmartGridComm), 2013 IEEE International Conference on (pp. 696-701). IEEE.

Huo, J., & Lu, D. (2009). Cloud determination of all-sky images under low-visibility conditions. Journal of Atmospheric and Oceanic Technology, 26(10), 2172-2181.

Huo, J., & Lu, D. (2010). Preliminary retrieval of aerosol optical depth from allsky images. Adv. Atmos. Sci. 27 (2), 421–426.

Huo, J., & Lu, D. (2012). Comparison of Cloud Cover from All-Sky Imager and Meteorological Observer. Journal of Atmospheric and Oceanic Technology, 29(8), 1093-1101.

Ineichen, P., 2006. Comparison of eight clear sky broadband models against 16 independent data banks. Solar Energy 80, 468– 478.

Ineichen, P., Perez R., 2002. A new airmass independent formulation for the Linke turbidity coefficient. Solar Energy, 73(3), 151–157.

Jacobs, N., Islam, M. T., & Workman, S. (2013). Cloud Motion as a Calibration Cue. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on (pp. 1344-1351). IEEE.

Johnson, R., Hering W., Shields, J., 1989. Automated visibility and cloud cover measurements with a solid-state imaging system. Tech. Rep., University of California, San Diego, Scripps Institution of Oceanography, Marine Physical Laboratory, SIO Ref. 89-7, GL-TR-89-0061, 128 pp.

Johnson, R., Shields, J., Koehler, T., 1991. Analysis and Interpretation of Simultaneous Multi-Station Whole Sky Imagery. Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, SIO 91-3, PL-TR-91-2214

Kalisch, J., Macke, A., 2008. Estimation of the total cloud cover with high temporal resolution and parameterization of short-term fluctuations of sea surface insolation. Meteorol. Z., 17, 603-611.

Kassianov, E., Long, C., Christy, J., 2005. Cloud-base-height estimation from paired ground-based hemispherical observations. J. Appl. Meteor., 44, 1221–1233.

Lara-Fanego, V., Ruiz-Arias, JA., Pozo-Vazquez, D., Santos-Alamillos, F., Tovar-Pescador, J., 2011. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). doi:10.1016/j.solener.2011.02.014, Solar Energy.

Lave, M., & Kleissl, J. (2013). Cloud speed impact on solar variability scaling–Application to the wavelet variability model. Solar Energy, 91, 11-21.

Lave, M., Kleissl, J., Arias-Castro, A., 2011. High-frequency irradiance fluctuations and geographic smoothing. Submitted to Solar Energy.

Linke, F., 1922. Transmissions-Koeffizient und Tru¨bungsfaktor. Beitr. Phys. fr. Atmos. 10, 91–103.

Lipperheide, M., Bosch, J. L., & Kleissl, J. (2015). Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. Solar Energy, 112, 232-238.

Liu, C. (2009). Beyond pixels: exploring new representations and applications for motion analysis (Doctoral dissertation, Massachusetts Institute of Technology).

Liu, S., Wang, C., Xiao, B., Zhang, Z., & Cao, X. (2013). Tensor Ensemble of Ground-Based Cloud Sequences: Its Modeling, Classification, and Synthesis.IEEE Geoscience and Remote Sensing Letters, 10, 1190-1194.

Long, C. J., Sabburg, J., Calbó, J., Pagès, D., 2006. Retrieving cloud characteristics from ground-based daytime color all-sky images. J. Atmos. Ocean. Technol. 23, 633–652.

Lorenz, E., Hammer, A., & Heinemann, D. (2004). Short term forecasting of solar radiation based on satellite data. In Proc. ISES Europe Solar Congress EUROSUN2004, Freiburg, Germany.

Lorenz, E., Heinemann, D., Wickramarathne, H., Beyer, H., Bofinger, S., 2007. Forecast of ensemble power production by grid-connected pv systems. In: Proc. of the 22th European PV Conference. pp. 3–9.

Lorenz, E., Hurka, J., Heinemann, D., Beyer, H., 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems.IEEE Journal of Selected Topics in Applied Earth Observa-tions and Remote Sensing 2 (1), 2–10.

Lucas, B. D., & Kanade, T. (1981, August). An iterative image registration technique with an application to stereo vision. In IJCAI (Vol. 81, pp. 674-679).

Luoma, J., Kleissl, J., & Murray, K. (2012). Optimal inverter sizing considering cloud enhancement. Solar energy, 86(1), 421-429.

Marquez R., H.T.C. Pedro and C.F.M. Coimbra (2012), Hybrid ANN/Satellite Imaging Forecasting Method for Global Horizontal Irradiance," submitted to Solar Energy.

Marquez, R., & Coimbra, C. F. (2013). Intra-hour DNI forecasting based on cloud tracking image analysis. Solar Energy, 91, 327-336.

Marquez, R., Pedro, H. T. C., Coimbra, C. F. M., June 2013. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to anns. Solar Energy 92, 176–188.

Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 26(5), 530-549.

Martins, F., Pereira, E., Abreu, S., 2007. Satellite-derived solar resource maps for Brazil under SWERA project, Solar Energy, 81, 517-528, ISSN 0038-092X, DOI: 10.1016/j.solener.2006.07.009.

Mathiesen, P., Kleissl, J., 2011. Evaluation of numerical weatherprediction for intra-day solar forecasting in the CONUS. Sol. Energy85 (5), 967–977.

McCartney, E., 1976. Optics of the Atmosphere: Scattering by Molecules and Particles. John Wiley & Sons, pp. 216.

Mecikalski, J. R. 2002. Using cloud-motion winds to understand kinematic processes of deep convection: Anvil-level outflow and momentum transport. Proc. Sixth Int. Winds Workshop, Madison, WI, WMO, 97–104.

Menzel, W. P. (2001). Cloud tracking with satellite imagery: From the pioneering work of Ted Fujita to the present. Bulletin of the American Meteorological Society, 82(1), 33-47.

Mills, A. (2010). Understanding variability and uncertainty of photovoltaics for integration with the electric power system. Lawrence Berkeley National Laboratory.

Nguyen, D. A., & Kleissl, J. (2014). Stereographic methods for cloud base height determination using two sky imagers. Solar Energy, 107, 495-509.

Nieman, S. J., W. P. Menzel, C. M. Hayden, D. Gray, S. T. Wanzong, C. S. Velden, and J. Daniels, 1997: Fully automated cloud-drift winds in NESDIS operations. Bull. Amer. Meteor Soc., 78, 1121-1133.

Norris, J. R., & Slingo, A. (2009). 2 Trends in Observed Cloudiness and Earth's Radiation Budget What Do We Not Know and What Do We Need to Know?.

Paoli, C., Voyant, C., Muselli, M., Nivet, M.-L., December 2010. Forecasting of preprocessed daily solar radiation time series using neural networks. Solar Energy 84 (12), 2146–2160.

Pavolonis, Michael J. and Heidinger, Andrew K.. 2004: Daytime cloud overlap detection from AVHRR and VIIRS. Journal of Applied Meteorology, Volume 43, Issue 5, 2004, pp.762-778.

Pavolonis, Michael J.; Heidinger, Andrew K. and Uttal, Taneil. 2005: Daytime global cloud typing from AVHRR and VIIRS: Algorithm description, validation, and comparisons. Journal of Applied Meteorology, Volume 44, Issue 6, 2005, pp.804-826.

Pereira, E., Martins, F., Abreu, S., Couto, P., Stuhlmann, R., Colle, S., 2000. Effects of burning of biomass on satellite estimations of solar irradiation in Brazil, Solar Energy, 68, 91-107, ISSN 0038-092X, DOI: 10.1016/S0038-092X(99)00044-4.

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F., 2002. A new operational model for satellite-derived irradiances: description and validation. Solar Energy 73, 307-317.

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr, K., Renné, D., & Hoff, T. E. (2010). Validation of short and medium term operational solar radiation forecasts in the US. Solar Energy, 84(12), 2161-2172.

Perez, R., M. Beauharnois, K. Hemker Jr., S. Kivalov, E. Lorenz, S. Pelland, J. Schlemmer, and G. Van Knowe, (2011): Evaluation of Numerical Weather Prediction Solar Irradiance Forecasts in the US. Proc. Solar 2011, American Solar Energy Society's Annual Conference.

Pfister, G., McKenzie, R. L., Liley, J. B., Thomas, A., Forgan, B. W., & Long, C. N. (2003). Cloud coverage based on all-sky imaging and its impact on surface solar irradiance. Journal of Applied Meteorology, 42(10), 1421-1434.

Quesada-Ruiz, S., Chu, Y., Tovar-Pescador, J., Pedro, H. T. C., & Coimbra, C. F. M. (2014). Cloud-tracking methodology for intra-hour DNI forecasting. Solar Energy, 102, 267-275.

Rabin, R. M., S. F.Corfidi, J. C.Brunner, and C. E.Hane.2004. Detecting winds aloft from water vapor satellite imagery in the vicinity of storms. Weather 59:251–257.

Remund, J., Perez, R., Lorenz, E., 2008. Comparison of Solar Radiation Forecasts for the USA. Proc. of the 23rd European PV Conference, 1.9-4.9, Valencia, Spain.

Ren, X. and Bo, L. (2012). Discriminatively trained sparse code gradients for contour detection. In Advances in neural information processing systems(pp. 584-592).

Rigollier, C., Lefèvre, M., Wald, L., 2004. The method Heliosat-2 for deriving shortwave solar radiation from satellite images, Solar Energy, 77, 159-169, ISSN 0038-092X, DOI: 10.1016/j.solener.2004.04.017.

Schmetz, J., 1989. Towards a surface radiation climatology: retrieval of downward irradiances from satellites. Atmos. Res. 23, 287–321.

Schroedter-Homscheidt, M., Hoyer-Klick, C., Rikos, E., Tselepsis, S., Pulvermüller, B., 2009. Nowcasting and forecasting of solar irradiance for energy electricity generation, SolarPACES Conf.

Seiz, G., Shields, J., Feister, U., Baltsavias, E., Gruen, A., 2007. Cloud mapping with ground-based photogrammetric cameras. International Journal of Remote Sensing 28 (9) 2001–2032.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(8), 888-905.

Shields, J. E., Karr, M. E., Johnson, R. W., & Burden, A. R. (2013). Day/night whole sky imagers for 24-h cloud and sky assessment: history and overview.Applied Optics, 52(8), 1605-1616.

Shields, J., Johnson, R., Karr, M., Wertz, J., 1998. Automated Day/Night Whole Sky Imagers for Field Assessment of Cloud Cover Distributions and Radiance Distributions. Tenth Symposium on Meteorological Observations and Instrumentation, American Meteorological Society, pp165 – 170.

Shields, J., Karr, M., Burden, A., Johnson, R., Hodgkiss, W., 2007. Continuing Support of Cloud Free Line of Sight Determination including Whole Sky Imaging of Clouds, Final Report for ONR Contract N00014-01-D-0043 DO #13, Marine Physical

Laboratory, Scripps Institution of Oceanography, University of California San Diego, Technical Note 273.

Shields, J., Karr, M., Burden, A., Johnson, R., Mikuls, V., Streeter, J., Hodgkiss, W., 2009. Research toward Multi-Site Characterization of Sky Obscuration by Clouds, Final Report for Grant N00244-07-1-009, Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, Technical Note 274.

Shields, J., Karr, M., Tooman, T., Sowle, D., Moore, S., 1998. The whole sky imager—a year of progress. presented at the Eighth Atmospheric Radiation Measurement (ARM) Science Team Meeting, Tucson, Arizona, March 23–27.

Slingo, A. (1990). Sensitivity of the Earth's radiation budget to changes in low louds. Nature, 343(6253), 49-51.

Sundaram, N., Brox, T., & Keutzer, K. (2010). Dense point trajectories by GPU-accelerated large displacement optical flow. In Computer Vision–ECCV 2010 (pp. 438-451). Springer Berlin Heidelberg

Szeliski, R. (2010). Computer vision: algorithms and applications. Springer.

Tapakis, R., & Charalambides, A. G. (2013). Equipment and methodologies for cloud detection and classification: A review. Solar Energy, 95, 392-430.

Urquhart, B., Ghonima, M., Nguyen, D., Kurtz, B., Chow, C. W., Kleissl, J. (2013). Sky-Imaging Systems for Short-Term Forecasting. In J. Kleissl, Solar Energy Forecasting and Resource Assessment (pp. 195-232). Academic Press.

Urquhart, B., Kurtz, B., Dahlin, E., Ghonima, M., Shields, J. E., & Kleissl, J. (2014). Development of a sky imaging system for short-term solar power forecasting. Atmospheric Measurement Techniques Discussions, 7(5), 4859-4907.

Velden, C. S., C. M. Hayden, S. J. Nieman, W. P. Menzel, S. Wanzong, and J. S. Goerss. 1997. Upper-tropospheric winds derived from geostationary satellite water vapor observations. Bull. Amer. Meteor. Soc. 78:173–195.

Velden, C. S., T. L. Olander, and S. Wanzong. 1998. The impact of multispectral GOES-8 wind information on Atlantic tropical cyclone track forecasts in 1995. Part I: Dataset methodology, description, and case analysis. Mon. Wea. Rev. 126:1202–1218.

Wacker, S., Gröbner, J., Zysset, C., Diener, L., Tzoumanikas, P., Kazantzidis, A., ... & Kämpfer, N. (2015). Cloud observations in Switzerland using hemispherical sky cameras. Journal of Geophysical Research: Atmospheres.

West, S. R., Rowe, D., Sayeef, S., & Berry, A. (2014). Short-term irradiance forecasting using skycams: Motivation and development. Solar Energy, 110, 188-207.

WRF, 2010.Weather Research and Forecasting (WRF) Model.<http://www.wrf-model.org>.

Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C. W., Ghonima, M., & Kleissl, J. (2014). Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. Solar Energy, 103, 502-524.

Zelenka, A., Perez, R., Seals, R., Renné, D., 1999. Effective accuracy of satellite-derived irradiance. Theor. Appl. Climatol. 62, 199–207.