Spatiotemporal Analysis of Irradiance Data using Kriging

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Engineering Sciences (Mechanical Engineering)

by

Seyed Mohammad Jamaly

Committee in charge:

Professor Jan Kleissl, Chair
Professor Ery Arias-Castro
Professor Renkun Chen
Professor Carlos Coimbra
Professor Daniel Rudnick

2017
The dissertation of Seyed Mohammad Jamaly is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017
# TABLE OF CONTENTS

Signature Page .................................................................................................................. iii  
Table of Contents................................................................................................................ iv  
List of Abbreviations .......................................................................................................... vii  
List of Symbols .................................................................................................................... ix  
List of Figures ....................................................................................................................... xii  
List of Tables ....................................................................................................................... xv  
Acknowledgements ............................................................................................................ xvi  
Vita........................................................................................................................................ xvii  
Abstract of the Dissertation ............................................................................................... xviii  
Chapter 1 Introduction ....................................................................................................... 1  

Chapter 2 Analysis, Quality Control, and Validation of the Irradiance and PV Power  
Output Datasets in California ............................................................................................. 10  
2.1. Datasets & Data Quality Control .............................................................................. 10  
  2.1.1 SAW modeled irradiance data ........................................................................... 10  
  2.1.2 CSI power output data ......................................................................................... 11  
  2.1.3 CIMIS & ISIS ground measured irradiance data ................................................ 13  
2.2 SolarAnywhere (SAW) Validation ............................................................................ 14  
  2.2.1 SAW validation against ground measured irradiation ...................................... 15  
     2.2.1.1 Error metrics ................................................................................................. 16  
     2.2.1.2 Comparison of California-wide GHI averages across the year .................. 17  
  2.2.2 SAW validation against power output of PV systems .................................... 20  
     2.2.2.1 Clear sky index from CSI PV power output ............................................... 20  
     2.2.2.2 Comparison of California-wide $kt$ averages across the year .................... 24  
2.3 Performance Model .................................................................................................... 25  
  2.3.1 Methodology of the conversion model ............................................................... 25  
  2.3.2 Calculating temperature and MPP efficiencies ............................................... 27  
  2.3.3 Model validation .................................................................................................. 28  
2.4 Conclusions ................................................................................................................ 30  
Acknowledgments ........................................................................................................... 33  

Chapter 3 Ramp Rates Analysis of Distributed PV Systems in California ..................... 34  
3.1 Introduction & Backgrounds ....................................................................................... 34  
3.2 Methodology ............................................................................................................... 36  
3.3 Absolute Ramp Rates ................................................................................................. 37
LIST OF ABBREVIATIONS

AST Anisotropic spatiotemporal
CCC Cross-correlation coefficient
CCM Cross-correlation method
CGILS CFMIP-GCSS intercomparison of Large-Eddy and Single-Column models) straticumulus cloud over-land case
CI Confidence interval
CIMIS California irrigation management information system
CMV Cloud motion vector
CSA Cross-spectral analysis
CSI California solar initiative rebate program
DG Distributed generation
DHI Diffuse horizontal irradiance
DNI Direct normal irradiance
DWR Department of water resources
ERCOT Electricity reliability council of Texas
GHI Global horizontal irradiance
GI Isotropic spatiotemporal
GOES Geostationary operational environmental satellite
IST Isotropic spatiotemporal
IOU California’s electric investor-owned utilities
ISIS NOAA integrated surface irradiance study network
LES Large eddy simulation
LOSO Leave-one-site-out
LOSOE LOSO using the entire time series
LOSOP LOSO using past data
LOT Leave-one-time-step-out
LOTIE LOTO using the entire time series
LOTOP LOTO using past data
LWP Liquid water path
MAE Mean absolute error
MBE Mean bias error
MDS Multidimensional scaling
MOS Modeled output statistics
MPP Maximum power point
nRMSE Normalized root mean squared error
NWP Numerical weather prediction
OK Ordinary Kriging
PBI Performance-based-incentive
PGE Pacific Gas & Electric
PTC Performance test condition
PV Photovoltaic
QC Quality control
RICO Rain in cumulus over the ocean
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>Root mean squared error</td>
</tr>
<tr>
<td>RSR</td>
<td>Rotating shadowband radiometers</td>
</tr>
<tr>
<td>SAW</td>
<td>SolarAnyWhere</td>
</tr>
<tr>
<td>SCE</td>
<td>Southern California Edison</td>
</tr>
<tr>
<td>SDG&amp;E</td>
<td>San Diego Gas &amp; Electric</td>
</tr>
<tr>
<td>SMUD</td>
<td>Sacramento municipality utility district</td>
</tr>
<tr>
<td>SP</td>
<td>Spatial</td>
</tr>
<tr>
<td>ST</td>
<td>Spatiotemporal</td>
</tr>
<tr>
<td>STC</td>
<td>Standard test condition</td>
</tr>
<tr>
<td>SZA</td>
<td>Solar zenith angle</td>
</tr>
<tr>
<td>ToD</td>
<td>Time of day</td>
</tr>
<tr>
<td>WLS</td>
<td>Weighted least squares</td>
</tr>
</tbody>
</table>
### LIST OF SYMBOLS

**Roman Symbols**

- \( a \)  
  A coefficient of the parametric semivariogram function
- \( c \)  
  A coefficient of the parametric semivariogram function
- \( C \)  
  Covariance function
- \( C_f \)  
  Calibration factor in performance model
- \( C_{GI} \)  
  A correction factor for gi in performance model
- \( C_S \)  
  Spatial covariance function
- \( C_T \)  
  Temporal covariance function
- \( D \)  
  Displacement vector at a given location
- \( day_{cl} \)  
  Clearest day in a period
- \( d_{lim} \)  
  Distance limit between pairs of sites in CSA method
- \( d_{max} \)  
  Maximum distance between pairs of sites (domain size)
- \( dn \)  
  Distance between a given location \((x_n, y_n)\) and a site \((i, j)\)
- \( e \)  
  Error index in CSA method
- \( G1 \)  
  First pass in multi-pass successive corrections scheme
- \( G2 \)  
  Second pass in multi-pass successive corrections scheme
- \( G3 \)  
  Third pass in multi-pass successive corrections scheme
- \( GHI_{CIMIS} \)  
  Measured GHI at a given CIMIS station
- \( GHI_{CS} \)  
  GHI in clear sky conditions
- \( GHI_{SAW} \)  
  SAW estimated GHI of the pixel in which the PV system located
- \( GI_{CS} \)  
  Plane-of-array global irradiance (GI) in clear sky conditions
- \( GI_{max, ToD} \)  
  GI expected for clear condition at a given ToD
- \( h_1 \)  
  Distance component in along-wind direction
- \( h_2 \)  
  Distance component in cross-wind direction
- \( h_c \)  
  Spatial decorrelation distance
- \( h(l) \)  
  Number of bins in empirical semivariogram
- \( I_{S=0} \)  
  Dirac delta function
- \( kt \)  
  Clear sky index
- \( kt_{CSI} \)  
  \( kt \) for measured CSI power output
- \( kt_{SAW} \)  
  \( kt \) for modeled SolarAnywhere (SAW) ghi data
- \( kt_{SMUD} \)  
  \( kt \) for measured SMUD GHI data
- \( kt_b \)  
  \( kt \) data of the pixels within a given box
- \( k_t \)  
  \( kt \) time series at a given site
- \( kW_{AC} \)  
  AC rating power output of a given PV system
- \( kW_{DC} \)  
  DC rating power output of a given PV system
- \( L \)  
  Total number of bins of distances in the domain
- \( m \)  
  Number of time steps of time series
- \( max_t \)  
  Maximum time lags between irradiance at site pairs
- \( max_{QC} \)  
  Maximum value of the cross-correlation of the pair of the sites
- \( mean_{QC} \)  
  Average value of the cross-correlation of the pair of the sites
- \( M(\theta) \)  
  Time delay moment
- \( MBE \)  
  Average mean bias error (MBE)
- \( MBE_{clear} \)  
  Mean bias error (MBE) for the times with clear sky conditions
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBE&lt;sub&gt;MF&lt;/sub&gt;</td>
<td>Mean bias error (MBE) at a given ToD and month</td>
</tr>
<tr>
<td>MBE&lt;sub&gt;year&lt;/sub&gt;</td>
<td>Mean bias error (MBE) for the whole year data</td>
</tr>
<tr>
<td>n</td>
<td>Number of sites (samples)</td>
</tr>
<tr>
<td>N</td>
<td>Number of segments (periods) in a year</td>
</tr>
<tr>
<td>nRMSE</td>
<td>Normalized root mean squared error</td>
</tr>
<tr>
<td>N(h(l),u)</td>
<td>Number of pairs in each bins and time lag of the semivariogram</td>
</tr>
<tr>
<td>P&lt;sub&gt;CSI&lt;/sub&gt;</td>
<td>Measured CSI PV power output</td>
</tr>
<tr>
<td>P&lt;sub&gt;CSI,t&lt;/sub&gt;</td>
<td>P&lt;sub&gt;CSI&lt;/sub&gt; in a given segment (period)</td>
</tr>
<tr>
<td>P&lt;sub&gt;max,day&lt;/sub&gt;</td>
<td>PV output expected for a day with clear condition</td>
</tr>
<tr>
<td>P&lt;sub&gt;max,ToD&lt;/sub&gt;</td>
<td>PV output expected for clear condition at a given ToD</td>
</tr>
<tr>
<td>P&lt;sub&gt;model&lt;/sub&gt;</td>
<td>Modeled PV output in performance model</td>
</tr>
<tr>
<td>P&lt;sub&gt;SAW&lt;/sub&gt;</td>
<td>Power output estimation of SAW irradiance by performance model</td>
</tr>
<tr>
<td>P&lt;sub&gt;SAW,AC&lt;/sub&gt;</td>
<td>AC power output of the PV system in performance model</td>
</tr>
<tr>
<td>P&lt;sub&gt;SAW,AC,i&lt;/sub&gt;</td>
<td>Initial estimate for P&lt;sub&gt;SAW,AC&lt;/sub&gt;</td>
</tr>
<tr>
<td>P&lt;sub&gt;SAW,AC,i2&lt;/sub&gt;</td>
<td>Second estimate for P&lt;sub&gt;SAW,AC&lt;/sub&gt; (for Second step in performance model)</td>
</tr>
<tr>
<td>P&lt;sub&gt;SAW,DC&lt;/sub&gt;</td>
<td>DC power output of the PV system in performance model</td>
</tr>
<tr>
<td>P&lt;sub&gt;SAW,DC,i&lt;/sub&gt;</td>
<td>Initial estimate for P&lt;sub&gt;SAW,DC&lt;/sub&gt;</td>
</tr>
<tr>
<td>pf</td>
<td>Power factor in performance model</td>
</tr>
<tr>
<td>R</td>
<td>Radius of influence in objective analysis</td>
</tr>
<tr>
<td>r&lt;sub&gt;e&lt;/sub&gt;</td>
<td>Equivalent radius of the drop size distribution</td>
</tr>
<tr>
<td>r&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Distance between i&lt;sup&gt;th&lt;/sup&gt; pair of sites</td>
</tr>
<tr>
<td>r&lt;sub&gt;QC&lt;/sub&gt;</td>
<td>Cross-correlation QC ratio</td>
</tr>
<tr>
<td>rMBE</td>
<td>Relative MBE</td>
</tr>
<tr>
<td>rMBE&lt;sub&gt;clear&lt;/sub&gt;</td>
<td>Relative MBE for the times with clear sky conditions</td>
</tr>
<tr>
<td>rMBE&lt;sub&gt;MT&lt;/sub&gt;</td>
<td>Relative MBE at a given ToD and month</td>
</tr>
<tr>
<td>SR</td>
<td>Skill ratio</td>
</tr>
<tr>
<td>std</td>
<td>Standard deviation of kt data</td>
</tr>
<tr>
<td>t&lt;sub&gt;dim&lt;/sub&gt;</td>
<td>Length of time series</td>
</tr>
<tr>
<td>T&lt;sub&gt;cell&lt;/sub&gt;</td>
<td>PV cell temperature</td>
</tr>
<tr>
<td>U&lt;sub&gt;X&lt;/sub&gt;</td>
<td>Velocity component in X direction in CCM method</td>
</tr>
<tr>
<td>u</td>
<td>Time lag in parametric covariance function</td>
</tr>
<tr>
<td>u&lt;sub&gt;c&lt;/sub&gt;</td>
<td>Temporal decorrelation length</td>
</tr>
<tr>
<td>v</td>
<td>Cloud motion speed</td>
</tr>
<tr>
<td>U&lt;sub&gt;r&lt;/sub&gt;</td>
<td>Velocity component in Y direction in CCM method</td>
</tr>
<tr>
<td>V&lt;sub&gt;var(θ)&lt;/sub&gt;</td>
<td>Variation moment of the CSA method</td>
</tr>
<tr>
<td>var&lt;sub&gt;s&lt;/sub&gt;</td>
<td>Variation ratio of the site</td>
</tr>
<tr>
<td>var&lt;sub&gt;b&lt;/sub&gt;</td>
<td>Variation ratio of a given box</td>
</tr>
<tr>
<td>W</td>
<td>Standard Cressman weighting function</td>
</tr>
<tr>
<td>w&lt;sub&gt;l,u&lt;/sub&gt;</td>
<td>Weights function in weighted least squares (WLS) method</td>
</tr>
<tr>
<td>X&lt;sub&gt;dim&lt;/sub&gt;</td>
<td>Dimension size in X direction</td>
</tr>
<tr>
<td>X&lt;sub&gt;radius&lt;/sub&gt;</td>
<td>Searching radius in X direction in CCM method</td>
</tr>
<tr>
<td>Y&lt;sub&gt;dim&lt;/sub&gt;</td>
<td>Dimension size in X direction</td>
</tr>
<tr>
<td>Y&lt;sub&gt;radius&lt;/sub&gt;</td>
<td>Searching radius in Y direction in CCM method</td>
</tr>
<tr>
<td>Z&lt;sub&gt;0&lt;/sub&gt;</td>
<td>Irradiance estimate at an arbitrary location (x&lt;sub&gt;0&lt;/sub&gt;) and time (t&lt;sub&gt;0&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Z&lt;sub&gt;α/2&lt;/sub&gt;</td>
<td>Confidence level coefficient</td>
</tr>
</tbody>
</table>
Observed irradiance at location \((x_i)\) and time \((t_j)\)

**Greek Symbols**

- **\(\alpha\)** A coefficient of the parametric semivariogram function
- **\(\alpha_1\)** An MPP coefficient in performance model
- **\(\alpha_{1,i}\)** Initial estimate for \(\alpha_1\)
- **\(\alpha_2\)** An MPP coefficient in performance model
- **\(\alpha_{2,i}\)** Initial estimate for \(\alpha_2\)
- **\(\alpha_3\)** An MPP coefficient in performance model
- **\(\alpha_{3,i}\)** Initial estimate for \(\alpha_3\)
- **\(\alpha_T\)** Temperature coefficient in performance model
- **\(\beta\)** Separable factor in the spatiotemporal covariance function
- **\(\gamma\)** Parametric semivariogram function
- **\(\gamma_{ik}^{jl}\)** \(\gamma\) for \(|i-k|\) spatial and \(|j-l|\) temporal lag respectively
- **\(\hat{\gamma}\)** Empirical semivariogram function
- **\(\gamma_{AI}\)** Anisotropic semivariogram function
- **\(\gamma_{FS}\)** Isotropic semivariogram function
- **\(\gamma_{LGR}\)** Lagrangian semivariogram function
- **\(\Gamma\)** Spatiotemporal semivariogram matrix
- **\(\delta\)** A coefficient of the parametric semivariogram function
- **\(\Delta\)** An angle correction factor in power output model
- **\(\Delta T_i\)** Time delay for \(i^{th}\) pair of sites
- **\(\zeta\)** A coefficient of the parametric semivariogram function
- **\(\eta_{AC}\)** Inverter efficiency in performance model
- **\(\eta_{AC,i}\)** Initial estimate for \(\eta_{AC}\)
- **\(\eta_{MPP}\)** MPP efficiency in performance model
- **\(\eta_{MPP,i}\)** Initial estimate for \(\eta_{MPP}\)
- **\(\eta_{Temp}\)** Temperature efficiency in performance model
- **\(\eta_{Tot}\)** Total solar irradiance to power conversion efficiency
- **\(\theta\)** Angle between sites
- **\(\theta_m\)** Cloud movement angle
- **\(\lambda_{ij}\)** Weights of the Kriging method
- **\(\mu\)** Lagrange multiplier
- **\(\nu_S\)** Nugget effects of the parametric semivariogram function
- **\(\rho\)** Parameter for convex combination model of anisotropic semivariogram function
- **\(\sigma\)** Standard deviation
- **\(\sigma^2\)** Variance of the spatiotemporal process
- **\(\sigma_{OK}^2\)** Kriging variance
- **\(\phi\)** General form of functions for stationary non-separable parametric covariance function
- **\(\omega\)** General form of functions for stationary non-separable parametric covariance function
LIST OF FIGURES

Figure 2.1: Map of 192 PV systems in SDG&E, SCE, and PGE territories....................... 12

Figure 2.2: Map of the ISIS and 52 CIMIS stations in California................................. 14

Figure 2.3: rMBE by month and time-of-day (rMBE_{MT}) for CIMIS versus SAW for (a) all data and (b) data in clear sky conditions in 2010 (averaged over 52 stations)............ 19

Figure 2.4: Flowchart showing the steps for calculating $kt_{CSI}$ for a given segment........ 23

Figure 2.5: Modeled SolarAnywhere versus measured power output; 30 min power output (for SZA<75°) in 2010 averaged over (a) SDG&E (45 PV sites), (b) SCE (81 PV sites), (c) PGE (66 PV sites), and (d) combined...................................................... 29

Figure 3.1: Largest absolute ramp magnitude versus ramp time interval (from 15-min upto 5-hours) for aggregate (a) normalized output ($P_{CSI}/kW_{AC}$) and (b) clear sky GHI ($GHI_{CS}/1000 \text{ W m}^{-2}$) from 192 PV sites in California...................................................... 38

Figure 3.2: Cumulative distribution of absolute value of 1-hour ramp rates of aggregate absolute and weather-induced 15-min output (both normalized by kW_{AC}) and clear sky GHI ($GHI_{CS}/1000 \text{ W m}^{-2}$) from all 192 PV sites in California...................................................... 39

Figure 3.3: Month of occurrence and direction of large absolute ramps; histogram of the largest 1-hour ramp rates of aggregate (a) absolute 15-min output ($P_{CSI}/kW_{AC}$) and (b) clear sky GHI ($GHI_{CS}/1000 \text{ W m}^{-2}$) from all 192 PV sites in California................................. 40

Figure 3.4: Four days with largest absolute ramps in SDG&E territory; normalized aggregate 15-minute PV output from all 45 PV sites (bars) for the days with the largest 1-hour ramp rates in 2010...................................................... 41

Figure 3.5: GOES images for the days with the largest ramp in SDG&E territory; GOES satellite images at 15 minute resolution on Jan. 22, 2010................................. 42

Figure 3.6: Largest weather-induced ramps; same as Figure 3.1 but for weather-induced normalized ramps (normalized by kW_{AC}) in California...................................................... 43

Figure 3.7: The day with largest weather-induced ramp in California, Jan. 19, 2010; normalized aggregate 15-minute PV output from all 192 PV sites (bars) minus 30 day average diurnal power output for the day with the largest 1-hour ramp rates in 2010.... 45

Figure 3.8: GOES images for the day with the largest weather-induced ramp in California; GOES satellite images at 15 minute resolution on Jan. 19, 2010.................. 46

Figure 3.9: GOES images for a day with marine layer breakup (second largest absolute ramp) in SDG&E territory............................................................................... 47
Figure 4.1: $kt$ contour of the simulated dataset obtained from RICO simulation at different time steps; (a) 08:20:00, (b) 08:22:00, (c) 08:24:00, and (d) 08:26:00. .......................... 68

Figure 4.2: Time series of the domain averaged cloud cover fraction (solid line) and variation ratio. .......................................................... 69

Figure 4.3: Calculation of the time delay variation moment (Eq. 4.4) over all angles by considering pairs of sites within the distance limit using all points in the domain (16384 sites) and 625 sites with the highest variability. ........................................... 70

Figure 4.4: Cloud velocity field calculation by modified CCM using boxes with 5 x 5 pixels through the whole domain at 08:27:40. .......................................................... 72

Figure 4.5: Time series of CCM cloud (a) speed and (b) direction by using boxes with 5 x 5 pixels through the whole domain for original CCM (dashed line) and modified CCM (solid line). .......................................................... 73

Figure 4.6: $kt$ contour of the simulated dataset obtained from CGILS simulation at different time steps during the cloud breakup period; (a) 13:10:00, (b) 13:11:00, (c) 13:12:00, and (d) 13:13:00. .......................................................... 76

Figure 4.7: Same as Figure 4.6 but for CGILS simulation at different time steps during the overcast period; (a) 11:10:00, (b) 11:11:00, (c) 11:12:00, and (d) 11:13:00. ..................... 77

Figure 4.8: Same as Figure 4.4 but at (a) 13:38:10 during the cloud breakup period and (b) 09:38:10 during the overcast period. .......................................................... 78

Figure 4.9: Time series of cloud (a) speed and (b) direction by CSA using 20 min subintervals during the cloud breakup period through the whole domain. .................. 83

Figure 4.10: Same as Figure 4.9 but for the overcast period. ......................................... 84

Figure 4.11: Cloud velocity estimation of CGILS simulation during cloud (i) breakup and (ii) overcast periods using CSA at different subinterval lengths for the whole domain and subdomain; (a) cloud speed, (b) cloud direction, and (c) error index. ........................................... 85

Figure 5.1: $kt$ contour of the simulated dataset obtained from RICO simulation at different time steps; (a) 08:23:00, (b) 08:25:00, (c) 08:29:00, and (d) 08:31:00. ............... 100

Figure 5.2: Same as Figure 5.1 but for the simulated dataset obtained from CGILS simulation at different time steps during the overcast period; (a) 11:10:00, (b) 11:11:00, (c) 11:12:00, and (d) 11:13:00. ........................................................................... 102

Figure 5.3: Same as Figure 5.1 but for the simulated dataset obtained from CGILS simulation at different time steps during the cloud breakup period; (a) 13:10:00, (b) 13:11:00, (c) 13:12:00, and (d) 13:13:00. ........................................................................... 103
Figure 5.4: $kt$ contour of the 1 min SMUD grid irradiance data on Jan. 24, 2012 at (a) 08:09:00, (b) 08:11:00, (c) 08:13:00, and (d) 08:15:00 PT. .......................................................... 105

Figure 5.5: $kt$ contour of the PV power output data in San Diego Gas and Electric (SDG&E) territory on Jan. 22, 2010 at (a) 10:00:00, (b) 10:15:00, (c) 10:30:00, and (d) 10:45:00 PT.......................................................... 106

Figure 5.6: The correlated zone calculated from the RICO simulated data for the isotropic case and cross-direction in the anisotropic case for non-separable (solid) and separable (dashed) cases. .......................................................... 111

Figure 5.7: The correlated zone calculated from the RICO simulated data for the different directions in the anisotropic case. ......................................................................................... 112

Figure 5.8: The correlated zone calculated from the RICO simulated data for the along-wind direction ($a = 0^\circ$ in Figure 5.7) in the anisotropic case.......................................................... 112

Figure 5.9: Application of the purely spatial and spatiotemporal ordinary Kriging to model $kt$ at one site in the whole domain for the RICO simulated dataset using (a) leave-one-site-out (LOSO) (b) leave-one-time-step-out (LOTO) evaluation methods........ 114

Figure 5.10: Same as Figure 5.9 but for $kt$ estimation at one site in the domain for the CGILS simulated dataset during overcast period; (a) LOSO and (B) LOTO evaluation methods.......................................................... 116

Figure 5.11: Same as Figure 5.9 but for $kt$ estimation at one site in the domain for the CGILS simulated dataset during the cloud breakup period; (a) LOSO and (B) LOTO evaluation methods.......................................................... 117

Figure 5.12: Same as Figure 5.9 but for 1 min $kt_{SMUD}$ estimation at one site in the SMUD grid network on Jan. 24, 2012; (a) LOSO and (b) LOTO evaluation methods ................. 120

Figure 5.13: Same as Figure 5.9 but for 15 min $kt_{CSI}$ estimation at one PV site in SDG&E territory on Jan. 22, 2010; (a) LOSO and (B) LOTO evaluation methods.................. 122

Figure 5.14: Same as Figure 5.9 but for 30 min $kt_{SAW}$ estimation at one PV site in SDG&E territory on Jan. 22, 2010; (a) LOSO and (B) LOTO evaluation methods...... 123
LIST OF TABLES

Table 2.1: Statistics of PV systems in SDG&E, SCE, and PGE territories ................. 13
Table 2.2: Statistics of 30 min $kt_{SAW}$ versus $kt_{CSI}$ (averaged over 192 PV sites) .......... 25
Table 3.1: Comparison of the largest ramp rates for different utility territories ............. 49
Table 4.1: Validation of cloud velocity for the RICO simulation by CSA .................... 71
Table 4.2: CCM cloud velocity of RICO simulation using boxes with 5 x 5 pixels through the domain at different horizons $H$. ............................................................. 74
Table 4.3: Cloud velocity calculation of CGILS simulation during cloud breakup and overcast periods by CSA (similar to Table 4.1) and CCM (similar to Table 4.2). .......... 79
Table 5.1: Evaluation statistics for the spatial (SP), isotropic spatiotemporal (IST), and anisotropic spatiotemporal (AST) Kriging methods for the 49 (30) sites selected in the whole domain (subdomain) of the RICO simulation ........................................ 115
Table 5.2: Same as Table 5.1 but for 49 sites in the whole domain of the CGILS simulation during the cloud breakup and overcast periods .......................................................... 118
Table 5.3: Same as Table 5.1 but for the 1, 5, and 15 min $kt_{SMUD}$ using 58 SMUD sites on Jan. 24, 2012 as well as averaged over all the days with partly cloudy or overcast conditions (27 days in 2012) ................................................................. 121
Table 5.4: Same as Table 5.1 but for the 15 min $kt_{CSI}$ (using 45 CSI sites) and 30 min $kt_{SAW}$ (using 1 km resolution data) in SDG&E territory on Jan. 22, 2010 as well as averaged over 10 days in which the largest ramp rates occurred ........................................... 122
ACKNOWLEDGEMENTS

I would like to acknowledge Professor Jan Kleissl for his support as my advisor and the chair of my PhD committee. Through multiple drafts and many long nights, his guidance has proved to be invaluable. Also, I appreciate all the help and guidance from co-workers in UCSD Solar Resource Assessment and Forecasting (SRAF) lab.

Chapter 2, in part, uses published material from Jamaly, M., Bosch, J.L., and Kleissl, J., 2013, “Aggregate Ramp Rates of Distributed Photovoltaic Systems in San Diego County”, IEEE Transactions on Sustainable Energy. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been previously published as Jamaly, M., Bosch, J.L., and Kleissl, J., 2013, “Aggregate Ramp Rates of Distributed Photovoltaic Systems in San Diego County”, IEEE Transactions on Sustainable Energy. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in Jamaly, M., and Kleissl, J., 2017a, “Robust Cloud Motion Estimation by Spatiotemporal Correlation Analysis of Irradiance Data”, Solar Energy. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, has been previously published as Jamaly, M., and Kleissl, J., 2017b, “Spatiotemporal Interpolation and Forecast of Irradiance Data Using Kriging”, Solar Energy. The dissertation author was the primary investigator and author of this paper.

I would like to thank each of my previous co-authors for their contributions and for their permission to reproduce these works in this dissertation.
VITA

2005 Bachelor of Science, Sharif University of Technology, Tehran, Iran
2007 Master of Science, Sharif University of Technology, Tehran, Iran
2017 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS


FIELDS OF STUDY

Major Field: Engineering Sciences (Mechanical Engineering), Fluid Mechanics, Mathematics, Atmospheric Sciences.
ABSTRACT OF THE DISSERTATION

Spatiotemporal Analysis of Irradiance Data using Kriging

by

Seyed Mohammad Jamaly

Doctor of Philosophy in Engineering Sciences (Mechanical Engineering)

University of California, San Diego, 2017

Professor Jan Kleissl, Chair

Solar power variability is a concern to grid operators as unanticipated changes in photovoltaic (PV) plant power output can strain the electric grid. The main cause of solar variability is clouds passing over PV modules. However, geographic diversity across a region leads to a reduction in cloud-induced variability, but the reduction depends on cloud speed. To illustrate the magnitude of solar variability, irradiance and PV power output datasets are first evaluated, validated and applied to detect the largest aggregate ramp rates in California.
Afterwards, spatiotemporal correlations of irradiance data are analyzed and cloud motion is estimated using two different methods; the cross-correlation method (CCM) applied to two or a few consecutive time steps and cross-spectral analysis (CSA) where the cloud speed and direction are estimated by cross-spectral analysis of a longer timeseries. CSA is modified to estimate the cloud motion direction as the case with least variation for all the velocities in the cloud motion direction. To ensure reliable cloud motion estimation, quality control (QC) is added to the CSA and CCM. The results show 33% (52°) and 21% (6°) improvement in the cloud motion speed (direction) estimation using the modified CSA and CCM over the original methods (without QC), respectively.

Spatial and spatiotemporal ordinary Kriging methods are applied to model irradiation at an arbitrary point. The correlations among the irradiances at observed locations are modeled by general parametric covariance functions. Besides the isotropic covariance function (which is independent of direction), a new non-separable anisotropic parametric covariance function is proposed to model the transient clouds. Also, a new approach is proposed to estimate the spatial and temporal decorrelation distances analytically using the applied parametric covariance functions, which reduced the size of the computations without loss in accuracy (parameter shrinkage). Results confirm that the non-separable anisotropic parametric covariance function is most accurate with an average normalized root mean squared error (nRMSE) of 7.92% representing a 66% relative improvement over the persistence model.

The results confirm the accuracy and reliability of the Kriging method for estimating irradiation at an arbitrary point even in more challenging real applications where cloud motion is unknown.
Chapter 1 Introduction

As demands for integration of large amounts of photovoltaic (PV) power plants into the electricity grid have increased recently, fully resolved (time steps on the order of seconds and spatial resolution on the order of 10 m) spatiotemporal irradiance data is needed. Simulations of solar power output for distribution feeder power flow studies (Nguyen and Kleissl, 2015) and short-term forecasting of power output from large power plants (Lipperheide et al., 2015) are some applications of such fully resolved irradiance data.

Ground measurements of solar irradiance are sparse and continuous high quality measurements require more effort in maintenance and data quality control than common meteorological state variables (Vignola et al., 2013). On the other hand, temporal downscaling and spatial interpolation is usually required for satellite-derived irradiance data with coarse temporal resolution at 15 to 30 min and large pixel size (1 km or more). Therefore, an interpolation technique is required to provide such spatially and temporally resolved irradiance data at unobserved locations. While linear interpolation techniques may be appropriate to estimate the average annual solar resource at unobserved locations, they reduce the solar irradiation variance at unobserved locations and do not preserve correlation properties.

As an alternative, a stochastic interpolation method (i.e., Kriging method) can be applied to high fidelity solar resource modeling at unobserved locations. For timeseries (with $m$ time steps) of observed irradiances at $n$ sites, $Z_{ij} = Z(x_i, t_j)$, the Kriging method can be applied to obtain irradiance at an arbitrary location ($x_0$) and time ($t_0$).
\[ Z(x_0, t_0) = \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij} Z_{ij}, \]  

Eq. 1.1

where \( \lambda_{ij} \) are weights which is estimated through the method by considering the correlation between the observed data (the process to obtain these weights are described in Section 5.1).

Kriging was first developed as a geostatistical interpolation method for spatial statistics (Krige, 1951). The Kriging method is superior to deterministic interpolation techniques (which use only mathematical functions; e.g. Widen (2015)) since both analytical and statistical methods are applied to predict unknown values based on correlations in the irradiance data. Kriging methods have been shown to be the best linear unbiased prediction method in many fields of study. In the Kriging method, the mean and correlation properties of the irradiance data are preserved and an estimation of the error of the process (Kriging variance) is provided as a basis for stochastic simulation. The Kriging method can be applied to both ground measured as well as satellite derived solar irradiance data for spatial interpolation, temporal downscaling, or forecast of the irradiance data. A literature review of state-of-the-art spatiotemporal Kriging methods is provided in Perez et al. (2016). According to Perez et al. (2016), various forms of the spatiotemporal Kriging are used in solar engineering: simple Kriging assumes that the mean of the process across space and time is known. Otherwise, ordinary Kriging (assuming unknown but constant mean) and universal Kriging (assuming the unknown mean is a known function of co-variates, e.g., latitude, longitude) can be used. In this study, the spatial and spatiotemporal ordinary Kriging method is applied to estimate irradiation at an arbitrary point.
In the Kriging method, irradiance correlations are modeled by a parametric spatiotemporal covariance function and parameters are calculated empirically (Zimmerman and Stein, 2010; Zimmerman, 1989) based on some simplifications and assumptions including stationarity, separability, and isotropy. In a temporal (or spatial) stationary process, the mean and other statistical properties of solar radiation is constant over time (or space) and the covariance function is a function of time lags (or distance between locations) only; spatiotemporal stationarity of solar radiation is achieved if the process is stationary both spatially and temporally. In separable covariance function, spatial solar radiation variation is independent of its temporal variation. Gneiting (2002) and Gneiting et al. (2007) demonstrated special requirements for covariance functions and showed that the geostatistical covariance functions cannot be considered separable (especially for meteorological data such as wind velocity fields or solar irradiance data). In isotropic covariance functions, the covariance in solar radiation does not depend on direction.

Since the main source of spatiotemporal variability in irradiance data are transient clouds, to increase the performance of the Kriging method, anisotropic covariance functions are required to model transient clouds in the domain of investigation. Gneiting et al. (2007) considered a physically motivated directional dependence of the covariance (Lagrangian covariance function): introducing anisotropy according to the velocity vector of the flow, the covariance is a function of \((x-vt)\) where \(x\), \(v\), and \(t\) represent space, velocity vector, and time, respectively. Adjusting an isotropic covariance function by a Lagrangian covariance function has been shown to improve forecast of meteorological data including wind velocity (Gneiting et al., 2007) and solar irradiance data (Aryaputra
et al., 2015). Schlather (2010) developed a general form of the Lagrangian covariance by considering a variable velocity vector with a multivariate normal distribution. The Lagrangian covariance function is applied in many solar irradiance studies to account for anisotropy due to cloud motion (Lonij et al., 2013; Inoue et al., 2012; Shinozaki et al., 2014). Yang et al. (2013), on the other hand, achieved spatial stationarity and isotropy through deformations of the geographical space based on the two-step method developed by Sampson and Guttorp (1992).

A challenge in applying the Kriging method for solar forecasting or interpolation is to reduce the size of the computations by quantifying spatial and temporal decorrelation distances. It has been widely demonstrated that the correlation decreases in distance and at some point irradiance time series at two sites will be uncorrelated (Perez et al., 2012; Lave and Kleissl, 2013; Arias Castro et al., 2014; Lonij et al., 2013; Perez and Fthenakis, 2015; Lohmann et al., 2016; Elsinga et al., 2013 and Elsinga et al., 2014). This distance is called spatial decorrelation distance (or effectiveness range or correlation length). A similar phenomenon is expected to occur in time (temporal decorrelation). Quantifying the spatial and temporal correlated zone is helpful for reducing the computational cost, since for a specific point and time only data within the correlated zone are needed in the Kriging model. This is critical when the size of the problem (number of sites and/or time steps) is relatively large. The decorrelation also specifies a limit for interpolation and downscaling. State-of-the-art estimation of the spatial decorrelation length is reviewed by Perez et al. (2016) who mentioned that research on temporal decorrelation in solar energy literature is lacking.
As a practical example, Yang et al. (2014) applied parameter shrinkage by computing spatial decorrelation (threshold distance) using correlations from all directions. They presented circumstantial evidence on temporal decorrelation, hypothesizing that the temporal decorrelation can be quantified through the standard deviation of time lag distribution at the threshold distance. Their study showed improvement in forecasting by applying data within the correlated zone (parameter shrinkage). In fact, their proposed correlated zone is a rectangular box in time-space coordinates, which has been obtained based on the purely spatial decorrelation length (threshold distance). However, more accurate analysis is needed to quantify the spatial and temporal correlated zone, especially for anisotropic cases.

In this study, a promising method is proposed to estimate spatial and temporal decorrelation analytically using the applied parametric covariance function. In this method, the decorrelation lengths are calculated at desired spatial and/or temporal lags. For the anisotropic Kriging method, to distinguish between along-wind and cross-wind decorrelation zones, separate decorrelation lengths are calculated in the cloud motion direction as well as other directions, including the cross-wind direction.

Gneiting et al. (2007) and Aryaputera et al. (2015) introduced a very simple Lagrangian covariance function. However, a more advanced Lagrangian covariance function is required for more accurate results using anisotropic spatiotemporal (AST) Kriging method. On the other hand, the Lagrangian covariance function proposed by Schlather (2010) is too complicated to be applied to real data.

Lonij et al. (2013) considered a covariance function which is a product of a Lagrangian covariance function (in the form of an exponential function) and a purely
temporal one (in the form of a powered exponential function). Inoue et al. (2012) and Shinozaki et al. (2014) considered a very similar covariance function as a product of the Lagrangian covariance function (in the form of the covariance function proposed by Schlather, 2010 and assuming constant cloud motion speed) and a purely temporal covariance function (in the form of an exponential function).

However, as reported by Shinozaki et al. (2014), their AST Kriging method led to worse results than the spatial Kriging method (using purely spatial covariance function) on some days. The main issue with the models proposed by Lonij et al. (2013), Inoue et al. (2012) and Shinozaki et al. (2014) is that the AST covariance function is reduced to a separable covariance function when the Lagrangian term in the covariance function is negligible, which occurs if the cloud motion is negligible or it varies rapidly so that the main direction is not detectable and, therefore, the anisotropic covariance function is practically reduced to an isotropic one. In these situations, the anisotropic model proposed by Inoue et al. (2012) and Shinozaki et al. (2014) underperforms presumably because a separable covariance function is a questionable assumption.

In this study, a new non-separable anisotropic covariance function, which is not reduced to a separable one for isotropic case (i.e., if the Lagrangian term is negligible), is proposed based on an advanced form of the Lagrangian covariance function developed by Schlather (2010). The proposed advanced Lagrangian covariance function has the capability of considering complex cases with different weather conditions. The proposed AST Kriging method assumes constant cloud motion speed during the interval under investigation and, therefore, it leads to accurate results for the cases in which the cloud motion is either steady or at most slowly varying in time. Moreover, for the cases with
unsteady cloud motions, it ensures that the AST Kriging method is similar or more accurate than the spatial (SP) and isotropic spatiotemporal (IST) Kriging method due to the fact that the proposed anisotropic covariance function is never reduced to a separable one. For unsteady cloud speeds and/or long periods of time (eg., a few days), the accuracy of the AST Kriging method can be further improved by splitting the whole time series into shorter time interval. Therefore, the proposed AST Kriging method is suitable for spatiotemporal interpolation and forecast of irradiance data in all weather conditions including clear-sky condition as well as overcast and partly-cloudy conditions with either steady or unsteady cloud motions.

In general, for the cases with unknown cloud motion speed and direction, the application of Lagrangian covariance functions is sensitive to accurate estimation of the cloud motion vector which is an emerging research area (Fung et al., 2014; Bosch et al., 2013; Bosch and Kleissl, 2013). In this study, the cloud motion is estimated using two different methods; the cross-correlation method (CCM) and cross-spectral analysis (CSA). The cloud motion estimation methodology is described in Chapter 4.

In CSA, the cloud speed and direction are estimated by cross-spectral analysis of the irradiance data at some given locations (sites) through the domain (Inoue et al., 2012; Shinozaki et al., 2014). The CSA method suggested by Inoue et al. (2012) and Shinozaki et al. (2014) is restricted by the spatial arrangement of the sites such that the cloud direction may be inaccurate if there are only a few distinct angles between the pairs of the chosen sites. To remove the restriction, a new CSA approach for cloud motion direction is proposed by considering the case with least variation for all the velocities in the cloud motion direction.
In CCM, the cloud motion is estimated by comparing correlation between irradiance data at two or more time steps (Hamill and Nehrkorn, 1993). The CCM suggested by Hamill and Nehrkorn (1993) is generalized for cloud movement estimation using unstructured ground measured data. Moreover, to compare the consistency of the method when applied to different scales, CCM is applied by considering the whole domain as well as smaller subdomains. Also, to ensure reliable cloud motion estimation, quality control (QC) is added to the CSA and CCM analyses including removing conditions with low variability and less correlated sites.

The rest of this dissertation structured as follow. In Chapter 2, the irradiance and PV power output datasets in California (which will be applied in the spatiotemporal Kriging method) are analyzed, quality controlled, and validated. To illustrate the need for accurate spatiotemporal analysis for solar resource assessment and forecast, in Chapter 3, the available irradiance and power output datasets are applied to detect the largest aggregate ramp rates in 3 territories in California in 2010. The cloud motion estimation methodology and results are described in Chapter 4. To estimate cloud motion, the CCM and CSA analyses have been performed using two spatially and temporally resolved irradiance simulated dataset generated from large eddy simulation (LES).

In Chapter 5, the spatiotemporal Kriging method has been performed using the simulated irradiance datasets as well as the real irradiance and output power data in California. The procedure of performing the spatial and spatiotemporal Kriging methods is described in Section 5.1. Validation methods as well as Kriging methods results are presented in Sections 5.2 & 5.3. For the real data in California (Sacramento and San Diego areas), with unknown atmospheric velocities, the cloud motion had to be estimated
during the process using Cross-Correlation Method (CCM) and the days in which the highest ramp rates occurred (which are detected through the ramp rates analysis in Chapter 3) is applied as the worst case scenario for performance of the spatiotemporal Kriging method. Finally, conclusions are made and some relevant future works are suggested in Chapter 6.
Chapter 2 Analysis, Quality Control, and Validation of the Irradiance and PV Power Output Datasets in California

The proposed spatiotemporal Kriging method will be applied on real irradiance and PV power output data in California. In this regard, the applied irradiance and PV power output datasets are analyzed and quality controlled (Section 2.1). Also, the satellite-derived solar irradiation estimate has been compared to solar irradiation measurements at 53 ground stations (Section 2.2.1) as well as measured power output at the 192 PV systems throughout California (Section 2.2.2). Moreover, a method is developed to concert the irradiance and PV power output data into clear sky index ($kt$), which is required for the application of spatiotemporal Kriging method in Chapter 5. In some cases (see Chapter 3), for comparison, irradiance data needs to be converted to power output. To convert irradiation data to power output, a performance model has been developed. The methodology of the conversion model as well as validation of the method (comparing to the measured power output of 192 PV systems) is described in Section 2.3.

2.1. Datasets & Data Quality Control

2.1.1 SAW modeled irradiance data

Modeled global horizontal irradiation (GHI) and direct normal irradiation (DNI) are provided by Clean Power Research’s commercially available SolarAnywhere (SAW) derived from Geostationary Operational Environmental Satellite (GOES) visible imagery at 1 km spatial and 30 minutes temporal resolution for California (SAW, 2016). To obtain GHI, a cloud index is calculated from the reflectance in each pixel measured by the satellite. Instantaneous GHI for each pixel is then calculated by using the cloud index to
decrease the irradiation calculated using a clear sky model that considers local and seasonal effects of turbidity (Perez et al., 2002; Perez et al. 2010).

2.1.2 CSI power output data

Moreover, power output of distributed PV systems in California is considered. The California Solar Initiative (CSI) rebate program requires a performance-based-incentive (PBI) payout for systems larger than 30 kW and makes it optional for smaller systems (CSI, 2011). This requires metering and monthly submission of 15 minute energy output to the payout administrator. We have obtained the 2009 & 2010 CSI measured output ($P_{CSI}$)- quality controlled for system performance (Itron and KEMA, 2010)- for 194, 385, and 403 PV power plants in three California's Electric Investor-Owned Utilities (IOUs); San Diego Gas & Electric (SDG&E), Southern California Edison (SCE), and Pacific Gas & Electric (PGE) territories, respectively.

The CSI database also includes street address and PV system specifications including DC Rating ($kW_{DC}$) at standard test condition (STC), AC Rating ($kW_{AC}$) at performance test condition (PTC, typically 14% less than STC), module and inverter models, inverter maximum efficiency, panel azimuth and tilt angles, and tracking type. The STC rating is obtained under idealized, controlled conditions of 1000 W m$^{-2}$ plane-of-array irradiance and cell temperature at $25^\circ$C while the PTC is developed in an attempt to simulate more realistic conditions at 1000 W m$^{-2}$ plane-of-array irradiance with panel temperature derived from ambient air temperature at $20^\circ$C and 1 m s$^{-1}$ wind speed. Given the rapid increase in solar DG in coastal urban centers in California which are included in our study (e.g. Los Angeles, San Francisco, and San Diego) this dataset presents
important information to PV system owners whether existing solar resource datasets can provide accurate estimates of irradiation (and as a result, power) in these areas.

Quality control Luoma and Kleissl (2012a) was used to exclude all CSI sites with at least one of the following characteristics not representative of irradiance: PV systems with hourly averaged (versus 15 min) data, more than 70% missing data (mostly because they were installed during 2010), significant noise or large spikes in power due to recording issues, decrease in power due to soiling, significant clipping of power due to undersized inverters, less than 5 distinct power output for the whole year, or plants divided into sub-arrays with different panel tilt and azimuth angles. Therefore, a final set of 192 PV systems are analyzed (Figure 2.1 & Table 2.1). To avoid errors due to sensor cosine response and shading by nearby obstructions (not considered by SAW), only data for solar zenith angles less than 75° are considered.

Figure 2.1: Map of 192 PV systems in SDG&E, SCE, and PGE territories.
Table 2.1: Statistics of PV systems in SDG&E, SCE, and PGE territories

<table>
<thead>
<tr>
<th>IOU</th>
<th>No. of PV systems</th>
<th>Total PTC rated capacity (MW)</th>
<th>Mean PTC rated (kW)</th>
<th>Median PTC rated (kW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDG&amp;E</td>
<td>45</td>
<td>4.73</td>
<td>105.1</td>
<td>46.43</td>
</tr>
<tr>
<td>SCE</td>
<td>81</td>
<td>17.48</td>
<td>215.8</td>
<td>192.9</td>
</tr>
<tr>
<td>PGE</td>
<td>66</td>
<td>16.29</td>
<td>229.4</td>
<td>165.8</td>
</tr>
</tbody>
</table>

2.1.3 CIMIS & ISIS ground measured irradiance data

For comparison, ground measured irradiation data are analyzed; the California Irrigation Management Information System (CIMIS) with 124 active weather stations (CIMIS, 2011) and the NOAA Integrated Surface Irradiance Study (ISIS) network with one station in Hanford, CA (ISIS, 2012). CIMIS is operated by the Department of Water Resources (DWR). Each CIMIS station is equipped with a Li-Cor LI200S photodiode pyranometer, accurate under typical conditions to ±5% (CIMIS, 2011). GHI is reported as an hourly average of 60 measurements within the hour. Also, ambient air temperature and wind speed are provided at each CIMIS station, which will be used in proposed performance model as described in Section 2.3.

CIMIS provides an initial QC assessment (Meek and Hatfield, 1994; Eching and Moellenberndt, 1998) issuing flags that allow the user to remove any data that appears erroneous. The best 52 CIMIS stations are considered in this study according to further quality control in Quality control Luoma and Kleissl (2012b) and Section 2.1. The ISIS site in Hanford, CA (with 3min irradiance data) is of special interest as data quality is expected to be higher and since SAW was calibrated and validated at this site before (SoDa, 2012). The same quality control is applied to the ISIS data and all flagged ISIS data (as described on NOAA website) are not considered (ISIS, 2012). Only data with
solar zenith angle less than 75° are considered to avoid error in sensor cosine response and shading. Figure 2.2 shows the location of the best 52 CIMIS stations and the ISIS site in California.

Unfortunately, however, the CIMIS data could not be applied in an operational environment, because station data are only downloaded once per day at midnight. Nevertheless, the data can elucidate whether ground measurement networks, such as those installed in Sacramento Municipality Utility Distract (SMUD) and SDG&E territories, are beneficial in tracking PV output.

![Figure 2.2: Map of the ISIS and 52 CIMIS stations in California.](image)

### 2.2 SolarAnywhere (SAW) Validation

Reliability of SAW data, which will be used in the spatiotemporal Kriging method, has been validated against ground measured solar irradiation (Section 2.2.1) and power output at the 192 PV systems throughout California (Section 2.2.2). Moreover, a general algorithm is developed to convert PV power output data into the clear sky index ($kt$). Since the spatiotemporal Kriging method works best with normalized irradiances
that produce a stationary signal, $kt$ estimation is required for detrending solar irradiance time series in Chapter 5.

The SAW enhanced resolution satellite-derived irradiation with 30-min temporal and 1 km spatial resolution is applied in this study. Perez et al. (2010) conducted a validation of an earlier version of the SAW algorithm against high quality ground measurements sites across the US (but outside California); the SAW dataset was found to have mean bias errors (MBE) between -5 and 15 W m$^{-2}$ and root mean square errors (RMSE, based on hourly averages) ranging from 73-118 W m$^{-2}$. The SMUD grid GHI data was compared with SAW data by Bing et al. (2012).

2.2.1 SAW validation against ground measured irradiation

The measured GHI of the ISIS in Hanford, CA and the 52 CIMIS stations are compared with the SAW GHI data of the corresponding pixel in which the stations are located. The analysis is conducted for yearly data in 2009 and 2010. For the validation, clear sky conditions are considered separately to evaluate differences between modeled and measured GHI caused by atmospheric composition or aerosol optical depth and not by cloud cover or cloud optical depth. For clear skies the SAW irradiation is essentially calculated from the Ineichen clear sky model with climatological (monthly) turbidity from NREL’s METSTAT database, which is a source of error given variable actual aerosol optical depth in the atmosphere. Clear conditions are assumed to exist if $0.85 < kt < 1.1$ and $std[kt(t-1h:t+1h)] < 0.03$, where $kt$ is clear sky index (the ratio of GHI to the GHI in clear sky condition) and $std$ is the standard deviation, and $t$ is time. These expressions filter for large $kt$ and low variability which is characteristic for clear
conditions. This criterion had to be met by both CIMIS (or ISIS) and SAW so that only simultaneous and collocated clear data was considered. For reference, we compute clear conditions from the Ineichen model with the SoDA turbidity (see Perez et al. (2002), Ineichen (2006), and SoDa (2012) for more details).

2.2.1.1 Error metrics

For each analysis, Mean Bias Error (MBE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are calculated. For example, for comparing SAW ($GHI_{SAW}$) against CIMIS GHI ($GHI_{CIMIS}$), MBE, MAE, and RMSE are calculated as

$$\left\{ \begin{array}{l}
MBE = \frac{1}{n} \sum_{s=1}^{n} (GHI_{SAW} - GHI_{CIMIS}) \\
MAE = \frac{1}{n} \sum_{s=1}^{n} abs(GHI_{SAW} - GHI_{CIMIS}) \\
RMSE = \frac{1}{n} \sqrt{\sum_{s=1}^{n} (GHI_{SAW} - GHI_{CIMIS})^2}
\end{array} \right. \quad \text{Eq. 2.1}$$

where $n$ is the number of samples. MBE describes persistent differences between $GHI_{SAW}$ and $GHI_{CIMIS}$. MAE and RMSE describe random differences between $GHI_{SAW}$ and $GHI_{CIMIS}$. Also, the relative MBE (rMBE), relative MAE (rMAE), and normalized RMSE (nRMSE) are calculated as

$$\left\{ \begin{array}{l}
rMBE = \frac{MBE}{mean(P_{CSI})} * 100\% \\
rMAE = \frac{MAE}{mean(P_{CSI})} * 100\% \\
nRMSE = \frac{RMSE}{mean(P_{CSI})} * 100\%
\end{array} \right. \quad \text{Eq. 2.2}$$
The confidence Interval (CI) is calculated to determine the significance level of the difference between $GHI_{SAW}$ and $GHI_{CIMIS}$ as

$$CI = \overline{MBE} \pm Z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right),$$

where $\overline{MBE}$ is the average $GHI_{SAW}$ bias error over all CIMIS stations ($GHI_{CIMIS}$), $Z_{\alpha/2}$ is the confidence level coefficient ($Z_{\alpha/2} = 1.96$ for 95% confidence level or $\alpha=0.05$), $\sigma$ is the standard deviation of MBE for the CIMIS stations, and $n$ is the total number of the stations.

To illustrate diurnal or seasonal patterns in the data, bias errors are averaged for each time of day (ToD) and each month separately to yield $MBE_{MT}$ (Nottrott and Kleissl, 2010) as the difference between $GHI_{SAW}(m, ToD)$ and $PGHI_{CIMIS}(m, ToD)$ for $m = 1, \ldots, 12$ (months) and $ToD = 1, \ldots, 24$ (60 min segments).

$$r_{MBE_{MT}(m, ToD)} = \frac{MBE_{MT}(m, ToD)}{mean[P_{CSI}(m, ToD)]} * 100.$$  

2.2.1.2 Comparison of California-wide GHI averages across the year

The daily average of CIMIS, SAW, and clear sky GHI (based on the Ineichen model with Linke Turbidity from the SoDa database) averaged over 52 stations is computed for the year 2010. Generally, SolarAnywhere overestimates CIMIS irradiation data by 17 to 20 W m$^{-2}$ or 3 to 5% throughout the year. The bias is somewhat larger for the clear periods at 21 to 31 W m$^{-2}$. During clear periods, SAW is larger than the clear sky model, while CIMIS measurements are smaller than the clear sky model.
There is excellent agreement between SAW and ISIS with seasonal biases of less than 2%. The only exceptions are July and August when consistent over- and underestimates of 2 to 5% are observed. Similar results are observed in clear conditions.

$rMBE_{\text{year}}$ (for the whole data) and $rMBE_{\text{clear}}$ (for the times with clear condition) between SAW modeled GHI and CIMIS and ISIS measured GHI is calculated. To exclude outliers, the CIMIS stations with $rMBE_{\text{year}}$ or $rMBE_{\text{clear}}$ out of the range of 0.25-0.75 quantiles are excluded. Therefore, 52 CIMIS stations, along with the ISIS station, are considered in this study.

Averaged 2010 $rMBE_{\text{MT}}$ of all 52 CIMIS stations is shown in Figure 2.3 (for both the whole dataset and the clear sky conditions). Overall bias error (SAW overestimates) of 18 W m$^{-2}$ or 3.7% are observed. MBE is 24 W m$^{-2}$ or 3.2% in clear conditions. The biases are largest in June and to a lesser extent in May and July. For the rest of the year biases are less than 2% during midday, but larger in the (less important) morning and evenings. Clear sky biases are largest from February until June. The same analysis performed for 2009. Although MBEs are larger in 2009, the same trends are observed for both years.

The coastal stations and the inland stations are also analyzed separately. The $rMBE$ is larger at coastal stations with 5% on average. Largest $rMBE$ is still observed in May and June, but the difference to the other months is less pronounced than in the California-wide data. The inland stations are more numerous and consequently the magnitude and trends are very similar to the overall data with an $rMBE$ of 3.3%. $MBE_{\text{year}}$, $rMBE_{\text{year}}$, $MBE_{\text{clear}}$ and $rMBE_{\text{clear}}$ (averaged over the 52 high quality CIMIS stations) are 18.07 W m$^{-2}$, 3.74%, 24.38 W m$^{-2}$, and 3.86% with the corresponding confidence intervals of +
4.15 W m\(^{-2}\), +0.87\%, 4.92 W m\(^{-2}\), and 0.68\%, respectively. Also, rMBE\(_{MT}\) at Hanford, CA is qualitatively consistent with the overall trends. There is no bias on average over the year, but SAW overpredicts in June and July by 3 to 5%.

![Figure 2.3](image)

Figure 2.3: rMBE by month and time-of-day (rMBE\(_{MT}\)) for CIMIS versus SAW for (a) all data and (b) data in clear sky conditions in 2010 (averaged over 52 stations). The caption indicates annual mean SAW and CIMIS GHI, 0.25 and 0.75 quantiles of rMBE\(_{MT}\), annual mean rMBE, rMAE, and nRMSE. All relative errors are obtained from hourly data by dividing annual MBE, MAE, and RMSE by mean (GHI\(_{CIMIS}\)).
2.2.2 SAW validation against power output of PV systems

SAW data has been compared against power output at the 192 PV systems throughout California. Since PV power output is a function of many parameters not related to irradiation, a normalized metric that quantifies the local solar resource is desirable. For this reason, clear sky index (kt) is used here to intercompare irradiation (SAW) and power (CSI) data. kt is defined as

\[ kt = \frac{GHI}{GHI_{CSI}} \]

where GHI is the Global Horizontal Irradiance and GHI_{CSI} is the GHI in clear sky conditions. According to the site longitude and latitude, 1-min GHI_{CSI} is calculated based on the Ineichen model with Linke Turbidity from the SoDa database (see Perez et al. (2002), Ineichen (2006), and SoDa (2012) for more details). GHI_{CSI} is averaged over the corresponding CSI and SAW time intervals (15-min and 30-min respectively).

2.2.2.1 Clear sky index from CSI PV power output

A complex procedure was designed to convert CSI PV power output to clear sky index. In order to calculate kt from the CSI power output (kt_{CSI} see flowchart in Figure 2.4), the year is split into N segments. For example, for a segment length L=30-days, N=12 or a segment length L=7-days, N=52. For each segment \((n=1,\ldots, N)\), the CSI measured power output \(P_{CSI,n}\) is extracted from the respective year-long dataset (\(P_{CSI}\)). Then, power output representative for clear conditions \((P_{\text{max,ToD}}\) for use in the denominator of Eq. 2.9 later) is obtained by finding the maximum of \(P_{CSI,n}\) at each time of day (ToD) \((P_{\text{max,ToD}} = \max\{ P_{CSI,n} | t = \text{ToD} \})\). The ‘n’ index is dropped hereon forward.
We assume that there is at least one clear period at each ToD through all days in the segment.

The tradeoff for the choice of segment length $L$ is as follows: Shorter segment lengths mean that $P_{\text{max}, \text{ToD}}$ is more specific to the day under investigation since it is selected from the power output of neighboring days with similar atmospheric composition and sun angles. On the other hand if the segment is too short, then no clear segment may have existed for each ToD and $k_{\text{CSI}}$ would be too large. Therefore $k_{\text{CSI}}$ will not become more accurate as the segment length is shortened beyond some threshold.

Due to mostly clear weather conditions in California, segment lengths as short as 7 days generally yield a clear event for every ToD. Therefore, only the segment length $L = 7$ days ($N=52$) is considered in this study.

To remove outliers from $P_{\text{max}, \text{ToD}}$ (see Eq. 2.7) the clear day output power needs to be modeled as

$$P_{\text{model}} = kW_{AC} \cos(SZA + \Delta),$$  \hspace{1cm} \text{Eq. 2.6}

where $SZA$ is the solar zenith angle on the day with the largest aggregate power output ($\text{day}_{cl}$, the ‘clearest day’), and $\Delta$ is an angle correction which is determined as an offset to $SZA$ from the time offset between solar noon and the time of the peak of daily power (such an offset would be caused by an azimuth different from south). Data that fall outside the range

$$0.9 < P_{\text{max,ToD}} / P_{\text{model}} < 1.1$$  \hspace{1cm} \text{Eq. 2.7}

are excluded. The ratio $P_{\text{max,ToD}} / P_{\text{model}}$ may be less than 0.9 due to local topographic shading. If the ratio is larger than 1.1 (due to cloud enhancement, see Luoma et al.
(2012c) for more details), the second largest power output at that ToD is chosen and the procedure is repeated until the criteria in Eq. 2.7 is satisfied.

Now $k_{CSI}$ could be computed as the ratio between $P_{CSI}$ and PV output expected for clear condition ($P_{\text{max,ToD}}$), but that would ignore effects of increasing (fall) and decreasing (spring) SZAs during the days in the segment. Consequently, clear sky plane-of-array Global Irradiance ($GI_{CS}$) is calculated from the Page model using $GHI_{CS}$ along with tilt and azimuth angles of the PV panel (Page, 2003); the Boland function is used for diffuse fraction (Boland et al., 2008). Then, a correction factor $C_{GI}$ is calculated using $GI_{\text{max,ToD}} = GI_{CS} \left( \text{day} = \text{day}_{cl} \right)$ as

$$ C_{GI}(k) = GI_{CS}(k) / GI_{\text{max,ToD}}(ToD), $$

Eq. 2.8

where $GI_{\text{max,ToD}}$ is used for the corresponding ToD. Then, PV output expected for clear condition is calculated as

$$ P_{\text{max,day}}(k) = P_{\text{max,ToD}}(ToD)C_{GI}(k), $$

Eq. 2.9

and $k_{CSI}$ can be obtained from

$$ k_{CSI}(k) = \frac{P_{CSI}(k)}{P_{\text{max,day}}(k)}. $$

Eq. 2.10

The procedure is self-calibrating by comparing actual power output against maximum power output at the same time of day during the surrounding days in each period. The method to determine the maximum power output corrects for unrealistic cloud enhancement and persistent shading effects and differences in expected clear sky output due to solar geometry. The same procedure is repeated for other segments to calculate $k_{CSI}$ for the whole year.
Figure 2.4: Flowchart showing the steps for calculating $kt_{CSI}$ for a given segment.

The procedure has the following shortcomings:
- PV power conversion efficiency is assumed to be constant, i.e. changes in PV power conversion efficiency with panel temperature are not accounted for. Result: CSI clear sky index is overestimated on cool days and overestimated in high irradiance conditions.

- Variability in clear sky atmospheric transmissivity (due to aerosols, water vapor, etc.) is not accounted for. Typically the maximum output in the surrounding 7 day period would not occur on a ‘normal’ clear day, but a clear day that also has the cleanest atmosphere. Result: CSI clear sky index is underestimated. This issue will be mitigated by calibration. Calibrating is performed to force \( k_{SAW} \) and \( k_{CSI} \) to be (on average) consistent in clear conditions and then focus on quantifying differences in cloudy conditions.

### 2.2.2.2 Comparison of California-wide \( kt \) averages across the year

The average SAW and CSI clear sky indices are in good agreement for all three IOUs with small bias and random errors. Generally, \( k_{SAW} \) overestimates \( k_{CSI} \) by 2 to 3%. As expected the scatter is larger in cloudy conditions, but still surprisingly small with rMAEs between 4.22 and 5.81% for the different IOUs. The smaller geographic diversity of the SDG&E sites causes the rMAE to be larger.

rMBE, rMAE, and nRMSE between \( k_{SAW} \) and \( k_{CSI} \) for the whole year, clear conditions, and cloudy conditions (the whole data excluding the clear conditions) are shown in Table 2.2. The statistics are presented for the calculated \( k_{SAW} \) and \( k_{CSI} \) both with and without calibration. After calibration, \( k_{SAW} \) overestimates \( k_{CSI} \) by 4.06% in non-
clear conditions and 2.45% overall. rMAE is only 4.22% overall and 5.83% in non-clear conditions. Table 2.2 shows that the calibration reduces the errors only by about 2%.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>rMBE (%)</th>
<th>rMAE (%)</th>
<th>nRMSE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All data</td>
<td>4.54</td>
<td>5.46</td>
<td>7.22</td>
</tr>
<tr>
<td>All data, calibrated</td>
<td>2.45</td>
<td>4.22</td>
<td>6.12</td>
</tr>
<tr>
<td>Clear conditions</td>
<td>2.01</td>
<td>2.3</td>
<td>2.94</td>
</tr>
<tr>
<td>Clear conditions calibrated</td>
<td>-0.14</td>
<td>1.71</td>
<td>2.17</td>
</tr>
<tr>
<td>Cloudy conditions</td>
<td>6.06</td>
<td>7.24</td>
<td>8.91</td>
</tr>
<tr>
<td>Cloudy conditions calibrated</td>
<td>4.06</td>
<td>5.83</td>
<td>7.59</td>
</tr>
</tbody>
</table>

2.3 Performance Model

In some cases including ramp rate analysis (Chapter 3), for comparison and validation against measured power output data, irradiance data is needed to be converted to PV power output data. Moreover, there is a need for estimating power output at any location through the spatiotemporal Kriging (virtual PV sites at the interpolated points).

To convert SAW irradiation data for each site to CSI measured power output, a performance model has been developed. The inputs to the model are the GHI and DNI (from SAW dataset) as well as tilt and azimuth angles of the PV panel. Also, to predict PV cell temperature, ambient air temperature and wind speed are obtained from measurements at the closest CIMIS station.

2.3.1 Methodology of the conversion model

Using the GHI, DNI and tilt and azimuth angles of the PV panel, plane-of-array Global Irradiance (GI) is calculated using the Page model (Page, 2003). To predict cell temperature ($T_{cell}$), ambient air temperature and wind speed are obtained from
measurements at the closest CIMIS station. Then, $T_{cell}$ is calculated using a 1D transient heat transfer model (Davis et al., 2001; Jones and Underwood, 2001). The temperature efficiency correction is then calculated as

$$\eta_{Temp} = 1 - \alpha_T (T_{cell} - 25^\circ C), \quad \text{Eq. 2.11}$$

where $\alpha_T$ is the temperature coefficient and calculated such that the highest linear correlation between modeled and measured AC power output are obtained (see Section 2.3.2). The DC power output of the PV system is estimated as

$$P_{SAW, DC} = kW_{DC} \cdot \eta_{Temp} \frac{GI}{1000 \, \text{Wm}^{-2}}. \quad \text{Eq. 2.12}$$

Inverter efficiency is modeled as (Luoma et al., 2012c)

$$\eta_{AC} = \frac{pf}{0.007 + 1.009 pf + (0.0975 pf)^2}, \quad \text{Eq. 2.13}$$

where the power factor is $pf = P_{SAW, DC} / kW_{AC}$.

Next maximum power point (MPP) efficiency is considered as (Williams et al., 2003)

$$\eta_{MPP} = a_1 + a_2 GI + a_3 \ln(GI). \quad \text{Eq. 2.14}$$

The MPP efficiency is applied to correct typically observed deviations in modeled output from measurements across different irradiation values using empirically obtained coefficients $a_1$, $a_2$ and $a_3$ (Beyer et al., 2004). In this study, these coefficients are calculated such that the highest nonlinear correlation between modeled and measured AC power output are obtained (see Section 2.3.2). Then, the AC power output of the PV system is estimated as
$P_{\text{SAW,AC}} = \eta_{\text{AC}} \cdot \eta_{\text{MPP}} \cdot P_{\text{SAW,DC}} f_c,$ \hspace{1cm} \text{Eq. 2.15}

where $C_f = \text{avg}(P_{\text{CSI}})/\text{avg}(P_{\text{SAW,AC}})$ accounts for the line losses and soiling by calibrating the modeled performance by the ratio of the annual average CSI measured to modeled power output.

The calibration guarantees that modeled averaged annual performance based on SAW irradiation is consistent (or without bias) with observed performance. Such a ‘modeled output statistics’ (MOS) correction would typically also be applied in operational forecasting of PV power output. The total solar irradiance to power conversion efficiency ($\eta_{\text{Tot}}$) is equal to

$$\eta_{\text{Tot}} = \eta_{\text{Temp}} \cdot \eta_{\text{AC}} \cdot \eta_{\text{MPP}} \cdot C_f.$$ \hspace{1cm} \text{Eq. 2.16}

### 2.3.2 Calculating temperature and MPP efficiencies

For each PV system, the temperature (linear regression) and MPP (nonlinear regression) efficiencies are calculated such that the highest correlation between modeled and measured AC power output are obtained (in 2 steps).

First, $P_{\text{SAW,DC,i}}$ and, then, $\eta_{\text{AC,i}}$ are calculated using Eqs. 2.12 & 2.13 by assuming the temperature coefficient ($\alpha_T$ in Eq. 2.11) $\alpha_T = 5 \times 10^{-3}$ K$^{-1}$ (this is a widely used value for $\alpha_T$ for silicon PV which is applied as an initial guess). The initial performance AC output is

$$P_{\text{SAW,AC,i}} = \frac{kWDC}{1000 \text{ Wm}^{-2}} \left[ -0.005 K^{-1}(T_{\text{cell}} - 25^\circ \text{C}) \right] \eta_{\text{AC,i}} \left[ a_{1,i} + a_{2,i} GI + a_{3,i} \ln(GI) \right]$$ \hspace{1cm} \text{Eq. 2.17}
where $P_{\text{SAW,AC},i}$ is an initial guess for AC performance output. The initial guess for MPP efficiencies $a_{1,i}$, $a_{2,i}$ and $a_{3,i}$ are obtained using nonlinear fit between $P_{\text{SAW,AC},i}$ and $P_{\text{CSI}}$.

Then, by applying the obtained values for the initial MPP efficiencies $a_{1,i}$, $a_{2,i}$ and $a_{3,i}$, the performance AC output at the second step is

$$P_{\text{SAW,AC},2} = kW_{\text{DC}} \frac{GI}{1000 \text{ Wm}^{-2}} \left[ 1 - \alpha_T (T_{\text{cell}} - 25^\circ C) \right] \eta_{\text{AC},i} \eta_{\text{MPP},i}, \quad \text{Eq. 2.18}$$

where $\eta_{\text{MPP},i} = a_{1,i} + a_{2,i} \ln(GI) + a_{3,i} \ln(GI)$. $\alpha_T$ is obtained using linear correlation between $P_{\text{SAW,AC},i2}$ and $P_{\text{CSI}}$.

Afterwards, by applying the obtained $\alpha_T$, the temperature efficiency, DC power output, inverter efficiency, MPP efficiency, and AC power output are calculated using Eqs. 2.11-2.15. Note that the calibration factor ($C_f$ in Eq. 2.15) is essentially the product of the constant coefficients in the two correlation steps above.

### 2.3.3 Model validation

The model was compared to measured power output from 192 PV systems over a year (Figure 2.5). The average SAW and CSI power outputs are in good agreement for all three IOUs: They are essentially unbiased on average due to the calibration ($C_f$ in Eq. 2.15) with small random errors. Typical differences (as measured by the rMAE and nRMSE) between the calibrated modeled and measured performance are 4 to 9% for 30-min averaged data. The average calibration factor of 0.91, $C_f$ in Eq. 2.15, confirms that the performance model generally overestimates before calibration, likely since line losses and soiling of the PV systems that are not considered. The average temperature
The temperature coefficient (used in Eq. 2.11) was found to be $\alpha_T = 5.5 \times 10^{-3} \text{ K}^{-1}$ (Temperature coefficients are expected to be between $3 \times 10^{-3} \text{ K}^{-1}$ and $5 \times 10^{-3} \text{ K}^{-1}$).

Figure 2.5: Modeled SolarAnywhere versus measured power output; 30 min power output (for SZA<75°) in 2010 averaged over (a) SDG&E (45 PV sites), (b) SCE (81 PV sites), (c) PGE (66 PV sites), and (d) combined. The caption indicates the calibration factor (Eq. 2.15), the correlation coefficient ($\rho$), rMBE, rMAE, nRMSE between $P_{\text{SAWAC}}$ and $P_{\text{CSI}}$. 
2.4 Conclusions

The irradiance and PV power output datasets in California are analyzed and quality controlled in Section 2.1. Then, in Section 2.2, the datasets are applied and validations of the new state-of-the-art solar resource model for California (SolarAnywhere, SAW) were conducted using ground measured solar irradiation (Section 2.2.1) and power output at the 192 PV systems throughout California (Section 2.2.2). Overall SAW is the most accurate publicly available solar resource dataset.

Section 2.2.1 showed that SAW is unbiased compared to the Hanford ISIS data. SAW overestimates the measured GHI at CIMIS stations by $18.07 \pm 4.15$ W m$^{-2}$ or $3.7\% \pm 0.9\%$ (95% confidence interval), on average. SAW is also biased large in clear conditions compared to the Ineichen / SoDa clear sky model and the CIMIS measurements. For completeness, rMAEs and RMSEs are calculated as well for hourly data and normalized by average irradiation. Typical rMAEs and RMSEs were 9% and 13% respectively. Overall the SAW solar resource data are very accurate both in bias and random error.

Despite careful quality control by the authors, CIMIS stations have inferior sensors and are generally less well maintained than high quality solar resource sites such as ISIS. That may suggest that the differences between SAW & CIMIS are at least partially related to CIMIS measurement errors. Especially soiling of the ground sites due to infrequent cleaning likely explains some of the bias. However, persistent trends over the year likely indicate some underlying bias in SAW. Also for PV performance applications one could argue that soiling of PV panels will be even larger than for CIMIS.
sensors, so CIMIS measurements may be more reflective of performance of solar power plants. From our analysis the following recommendations emerge:

- The relative mean bias error ($rMBE_{MT}$, averaged over all 52 stations) is largest in May through July. This is the most significant finding as it holds both for clear data and all data, 2009 and 2010, and for CIMIS and ISIS. The cause of this difference, however, is unclear.

- The SoDa turbidity climatology appears to be more accurate than the METSTAT turbidity database used in SAW, but the average observed CIMIS clear sky data lie in between.

- $rMBE_{MT}$ of the coastal stations is slightly larger than for inland stations. For the coastal stations the largest differences occurred in both morning (Jun.-Nov.) and evening (Mar.-Jul.). The morning differences could be related to marine layer clouds, while the evening differences show correlation to the clear sky model. However, overall the coastal differences are much smaller than those found for the National Solar Radiation Database (Nottrott and Kleissl, 2010) indicating an improvement in SAW compared to previous versions.

The clear sky index ($kt$) is used to intercompare irradiation (SAW) and power (CSI) data in Section 2.2.2. Clear sky indices were calculated from both datasets ($kt_{SAW}$ and $kt_{CSI}$), where a general algorithm is developed to convert CSI PV power output data into the clear sky index. $kt_{SAW}$ was found to overestimate $kt_{CSI}$ by 2.11% $\pm$ 0.2% (95% confidence interval) throughout the year. The relative biases are typically smaller during midday, but independent of month. Since the main differences between $kt_{SAW}$ & $kt_{CSI}$
occurred in non-clear conditions (relative mean bias errors of 4.06%), the smaller conversion efficiency of PV systems at lower irradiation could contribute to the differences. On the other hand, higher efficiency is expected at lower irradiation since the temperature is lower. This suggests that PV systems don’t operate at maximum power point (MPP) condition at lower irradiation.

Also, despite careful quality control by the authors, some of the PV systems are not well maintained. Consequently, some of the differences between $kt_{SAW}$ & $kt_{CSI}$ are related to the reduced power output caused by soiling or topographic shading. However, The small relative Mean Absolute Error (rMAE) of SAW (4.22% overall) indicates its suitability as input for PV variability studies from scales of a feeder to a balancing area. The other limitation is the missing/ or unreliable measured power output data is one of the shortcoming in this analysis. Although a quality control was used to exclude the PV systems with missing/ or unreliable data, it might be still possible to have less accurate data at some time for some sites (especially at high solar zenith angles).

A performance model has been developed in Section 2.3 to concert irradiance to PV power output. The performance model was compared to measured power output from 192 PV systems over a year and it was shown that the average SAW and CSI power outputs are in good agreement in California. Dependencies of the error in modeled performance on cell temperature, ambient temperature, wind speed, zenith angle, and inverter efficiency were examined. However, no trend in the error was found indicating that a more complex model for these variables would not necessarily improve the agreement. The remaining random errors (RMSE) could be related to the temperature
model and stem from the irradiation input data due to location errors in the SAW cloud fields (satellite navigation errors and cloud-to-shadow parallax).

The bias error between modeled and measured power output was found to be larger in summer (up to 5%) while SAW+ performance model underestimate the measured data in the other months. An average PV efficiency derate ($\eta_{\text{Tot}}$) of 79% (loss of 21%) describes losses due to panel temperature, AC conversion, MPP tracking, and annual calibration. The average calibration factor of 0.91 confirms that the SAW+ performance model generally overestimates, likely due to line losses and soiling of the PV systems that are not considered in Eq. 2.15. The bias error between modeled and measured power output was found to be less than 3% during middays while SAW+ performance underestimates the measured data in winter and spring times. SAW+ performance provides a validated means to simulate real world system performance and integrate PV simulations into grid simulations or other tools. SAW+ performance does not, however, account for losses due to complete system outages due to maintenance, etc.

Acknowledgments

Chapter 2, in part, uses published material from Jamaly, M., Bosch, J.L., and Kleissl, J., 2013, “Aggregate Ramp Rates of Distributed Photovoltaic Systems in San Diego County”, IEEE Transactions on Sustainable Energy. The dissertation author was the primary investigator and author of this paper.
Chapter 3 Ramp Rates Analysis of Distributed PV Systems in California

The available irradiance (SAW and CIMIS) and power output (CSI) datasets (as described in Section 2.1) are applied to detect the largest aggregate ramp rates in 3 territories in California in 2010. The goal of the ramp rate analysis is to quantify the largest aggregate ramp rates and evaluate how much on-line metering and telemetry of PV systems is necessary to track output of distributed generation for resource-adequacy applications. This study illustrates the need for accurate spatiotemporal analysis for solar resource assessment and forecast.

Moreover, the ramp rates analysis reveals the days with partly cloudy or overcast conditions (the days in which the highest ramp rates occurred). The spatiotemporal Kriging method is going to be applied on the ground measured data in California; SMUD irradiance data and CSI power output data (See Chapter 5). To demonstrate the capability of the method, spatiotemporal Kriging method will be applied on days with partly cloudy or overcast conditions, which is detected in this study as the days in which the highest ramp rates occurred. These days with the highest variabilities are chosen as the worst case scenario for the performance of the spatiotemporal Kriging method in Chapter 5.

3.1 Introduction & Backgrounds

Integration of large amounts of photovoltaic (PV) into the electricity grid poses technical challenges due to the variable solar resource. Solar distributed generation (DG) is often behind the meter and consequently invisible to grid operators. Variability is a concern for both wind and solar energy.
Wan (2011) analyzed power ramping behavior of wind power plants in the Electricity Reliability Council of Texas (ERCOT). Examining over four years of data, large ramps exceeding 25% of total capacity were observed almost once every other day. The ability to forecast such ramps can mitigate some of the economic and reliability issues (Bacher et al., 2009). Murata et al. (2009) and Wiemken et al. (2001) have analyzed output from fleets of solar DG. Murata et al. (2009) analyzed 1-min data from 52 PV systems spread across Japan using a fluctuation index, which is the maximum difference in aggregated power output as a function of time interval. Over 1-min, sites more than about 50-100 km apart were uncorrelated. For times greater than 10-min, however, sites within 1000 km were not independent, though some of the dependence may be due to diurnal solar cycles. Analyzing output from 100 PV sites spread throughout Germany Wiemken et al. (2001) found that 5-min fluctuations of ±5% of aggregate power output at rated capacity are virtually nonexistent.

The ability to understand actual variability of solar DG will allow grid operators to better accommodate the variable electricity generation for resource adequacy considerations that inform scheduling and dispatching of power. From a system operator standpoint, it is especially important to understand when aggregate power output is subject to large ramp rates. If in a future with high PV penetration all PV power systems were to strongly increase or decrease power production simultaneously, it may lead to additional cost or challenges for the system operator to ensure that sufficient flexibility and reserves are available for reliable operations.
3.2 Methodology

To demonstrate single and aggregate solar variability in California, aggregate ramp rates of 192 distributed photovoltaic (PV) systems installed in California were analyzed and compared to modeled power calculated from satellite irradiances. Irradiance measured at 39 ground stations was considered as well. The objective of the ramp rate analysis is two-fold. Firstly, knowledge about the largest possible aggregate ramp rates and underlying meteorological conditions is useful for system operators to plan for worst-cases. Secondly, under extreme ramp rate conditions, knowing the PV output in real time would be most valuable since regulation up or regulation down capacity may have to be quickly procured. The ability of CIMIS and SAW modeled PV performance to match actual output is therefore of interest.

At each PV system, the CSI measured power output \( P_{CSI} \) is compared with the measured GHI at the closest CIMIS \( GHI_{CIMIS} \) station as well as the SAW estimated GHI \( GHI_{SAW} \) and power output \( P_{SAW} \) of the pixel in which the PV system is located. \( P_{SAW} \) is obtained by converting irradiation data into power output using the performance model. The largest absolute and weather-induced aggregate ramp rates are analyzed in SDG&E, SCE, and PGE territories (each territory separately as well as all 3 territories together). The analysis is conducted for January 1st to December 31st, 2010. To avoid errors due to sensor cosine response and shading by nearby obstructions (not considered by SAW), only data for solar zenith angles less than 75° are considered. Performance when the solar zenith angle is less than 75° for a flat plate system is less than 26% of rated capacity so hourly change rates are likely to be substantially less during those periods’.
3.3 Absolute Ramp Rates

To calculate the largest absolute ramp rates, on each day, the aggregate PV power output is calculated at each time step; PV sites with any missing data on that day are completely excluded. Differences in the aggregate PV power output are calculated for different ramp duration intervals; 15-min through 5-hour in 15-min increments. To facilitate scaling the results to future PV penetration scenarios (assuming a similar geographic diversity), the aggregate power outputs are normalized by the aggregate (PTC) $kW_{AC}$ capacity of the PV systems for each time period.

To consider the effects of diurnal cycles, 1-min GHI in clear sky conditions ($GHI_{CS}$) at each PV site is calculated as well. Then, $GHI_{CS}$ is averaged over the CSI time interval (15-min). The aggregate $GHI_{CS}$ is calculated at each time step and differences in the aggregate $GHI_{CS}$ are also calculated for different ramp duration intervals; 15-min through 5-hour in 15-min increments.

The largest step sizes in the absolute aggregate PV power output (normalized by $kW_{AC}$) and the aggregate $GHI_{CS}$ (normalized by 1000 W m$^{-2}$) are detected over the year for different intervals (Figure 3.1). As expected, the maximum ramp magnitude increases with the ramp interval. For the absolute ramp rates, the increase is near linear up to about 120 minutes at about 0.46% of PTC per minute. The maximum ramp magnitude approaches 90% for 5 hour ramps reflective of the diurnal cycle (e.g. from zero output at 0700 to near maximum output at 1200) on a clear day.
Figure 3.1: Largest absolute ramp magnitude versus ramp time interval (from 15-min upto 5-hours) for aggregate (a) normalized output ($P_{CSI/kW_{AC}}$) and (b) clear sky GHI ($GHI_{CS}/1000$ W m$^{-2}$) from 192 PV sites in California.

1-hour ramps have a special significance as most energy exchange between electric balancing areas is currently scheduled over hourly intervals. The distribution of hourly ramp rates in normalized aggregate measured PV power output along with 1-hour ramp rates in the aggregate clear sky GHI ($GHI_{CS}/1000$ W m$^{-2}$) are presented in Figure 3.2. The clear sky rate is of interest because it simulates the output ramps that would be experienced if there were no clouds or fog and the weather was always clear. This is the precisely predictable rate that is not governed by weather. The distribution of hourly absolute ramp rates in the aggregate PV output (Figure 3.2) shows that ramps over 23% h$^{-1}$ of PTC capacity are rare, occurring only for 150 hours of the year. For smaller ramps, the distribution decreases linearly.
Figure 3.2: Cumulative distribution of absolute value of 1-hour ramp rates of aggregate absolute and weather-induced 15-min output (both normalized by kW_{AC}) and clear sky GHI (GHI_{CS}/1000 W m^{-2}) from all 192 PV sites in California. The ramps are zero for the remaining hours up to 8760 h, because these are night time conditions or missing data.

Figure 3.3 shows a histogram (by month) of the 1-hour absolute ramp rates of aggregate normalized power output (normalized by kW_{AC}) and 1-hour ramp rates in the aggregate clear sky GHI, which are larger than 23% of PV capacity and 22% of maximum GHI_{CS} respectively. The largest absolute ramps in March and April can be explained by clear sky ramps.
Figure 3.3: Month of occurrence and direction of large absolute ramps; histogram of the largest 1-hour ramp rates of aggregate (a) absolute 15-min output \( P_{CSI}/kW_{AC} \) and (b) clear sky GHI \( (GHI_{CS}/1000 \text{ W m}^{-2}) \) from all 192 PV sites in California. The black lines show percentage of the measured aggregate 15-minute PV output (in Figure a) and the aggregate 15-min clear sky GHI (in Figure b) from all 192 PV sites normalized to the maximum month (May).

The analysis revealed that the largest absolute ramp rates in California (considering all the three territories together) are solely a result of diurnal cycles (confirmed using GOES images), and are therefore predictable. However, this is not the case for smaller region like SDG&E territory. Figure 3.4 shows daily profiles of the normalized aggregate CSI measured and SAW modeled power outputs along with the normalized aggregate CIMIS GHI for the days when the four largest absolute ramps were observed in SDG&E territory. The largest ramp caused a change of 60% of PTC capacity within one hour. SAW estimates tracked the CSI power output typically within 1-12% rMAE at 30-min resolution. The 15 minute consecutive GOES images for the time period with the largest absolute ramp in SDG&E territory (Figure 3.4a) are illustrated in Figure 3.5.
Figure 3.4: Four days with largest absolute ramps in SDG&E territory; normalized aggregate 15-minute PV output from all 45 PV sites (bars) for the days with the largest 1-hour ramp rates in 2010 (magenta bars show the timing of the large ramp). Normalized (to a maximum of 1) aggregate 30-minute performance output (red) and GHI (green) obtained from SolarAnywhere at each pixel and normalized aggregate hourly measured GHI of 5 weather stations (black) are also shown. The caption indicates the daily available GHI from SolarAnywhere averaged over all 45 PV sites, the CSI ramp magnitudes (normalized by both PV PTC capacity and daily maximum output). Relative (divided by annual average CSI output) mean absolute error (MAE), mean bias error (MBE), and RMSE between aggregate SAW performance and CSI outputs are also presented. (a) Jan. 22, 2010, (b) Sep. 18, 2010, (c) Jan. 4, 2010, (d) Aug. 12, 2010.
Figure 3.5: GOES images for the days with the largest ramp in SDG&E territory; GOES satellite images at 15 minute resolution on Jan. 22, 2010 (Figure 3.4a). The circles represent 45 PV systems in SDG&E territory shown in Figure 2.1. The area of the circles is proportional to the power rating of the PV system and the largest system is 501 kW. The color bar shows the ratio of 15-min averaged output to annual maximum output at that time of day (ToD).
3.4 Weather-Induced Ramp Rates

Weather-induced ramp rates with reference to a 30-day average of power output, on the other hand, are helpful to detect unexpected variations. These unexpected variations are more likely caused by weather than the sun’s movement through the sky. First the average aggregate PV power output of the previous 30 days at a given time of day (ToD), corrected for differences in aggregate PV capacity, is subtracted from the aggregate PV power output at that ToD. Then, the differences in the resulting timeseries constitute weather-induced ramp rates, which are calculated for different ramp duration intervals; 15-min through 5-hour in 15-min increments.

Similar to Figure 3.1, the largest step sizes in the weather-induced aggregate ramp rates (normalized by $kW_{AC}$) in California are presented in Figure 3.6. The maximum ramp magnitude increases with the ramp interval and approaches 50% for 5 hour ramps.

Figure 3.6: Largest weather-induced ramps; same as Figure 3.1 but for weather-induced normalized ramps (normalized by $kW_{AC}$) in California.
The distribution of 1-hour weather-induced ramp rates in the aggregate PV output shows that ramps over 9% h\(^{-1}\) of PTC capacity are rare. The large weather-induced ramps were most likely in the winter months (primarily December and January) when they occurred about once per day. Presumably, this is related to overcast conditions (large morning down or evening up-ramp compared to the 30 day – mostly clear - average) or when storm systems moving into (large down ramp) or out of the area (large up-ramp). Large ramps are anti-correlated with the average output over a month.

Similar to Figure 3.3, Figure 3.7 shows the daily profile for the day when the largest weather-induced ramp (with reference to 30-days average of the aggregate output) was observed in California. SAW estimates tracked the CSI power output typically within 6-15% rMAE at 30-min resolution. The 15 minute consecutive GOES images for the time period with the largest weather-induced ramp (Figure 3.8) confirm that a large cloud band that is parallel to the cost covers most PV sites south of the bay area around noon, but rapidly clears the Southern California coastal area starting at 1300h. This day also caused the largest weather-induced ramp in SCE territory with a 29.8% magnitude. The greater geographic diversity of sites serves to reduce the ramp magnitude for the whole state by almost one-third. In SDG&E territory, January 22 was both the day with the largest weather-induced and largest absolute ramp (Figure 3.4a) and January 4\(^{th}\) was also in the top 4 for both (Figure 3.4c).
Figure 3.7: The day with largest weather-induced ramp in California, Jan. 19, 2010; normalized aggregate 15-minute PV output from all 192 PV sites (bars) minus 30 day average diurnal power output for the day with the largest 1-hour ramp rates in 2010 (magenta bars show the timing of the large ramp). Normalized (to a maximum of 1) aggregate 30-minute performance output (red) and GHI (green) obtained from SolarAnywhere at each pixel and normalized aggregate hourly measured GHI of 39 weather stations (black) are also shown (all minus 30 day average diurnal values). The caption indicates the daily available GHI from SolarAnywhere averaged over all 192 PV sites, the CSI ramp magnitudes (normalized by both PV PTC capacity and daily maximum output). Relative (divided by annual average CSI output) mean absolute error (MAE), mean bias error (MBE), and RMSE between aggregate SAW performance and CSI outputs are also presented. The (b) graphs also show the 30 day average diurnal power output (yellow lines) and are not normalized to 1.
Figure 3.8: GOES images for the day with the largest weather-induced ramp in California; GOES satellite images at 15 minute resolution on Jan. 19, 2010 (Figure 3.7). The circles represent 192 PV systems shown in Figure 2.1. The area of the circles is proportional to the power rating of the PV system and the largest system is 1000 kW. The color bar shows the ratio of 15-min averaged output to annual maximum output at that time of day (ToD). The largest aggregated 1 hour weather-induced ramp was 20% of PV capacity and occurred from 1330 to 1430 PST.
3.5 Marine Layer Breakup

While it did not cause the largest ramp, marine layer breakup caused most of the large ramps occurring during May through November and two of the four largest ramps (Sep 18 and Aug 12) in SDG&E territory. The 15 minute consecutive GOES images for the time period with the second largest absolute ramp in SDG&E territory (Figure 3.4b, as an example of marine layer cloud breakup) are illustrated in Figure 3.9. A large morning up-ramp (44% of installed PV capacity per hour) occurred due to marine layer retreat that occurs over 2 hours (8am-10am).

Figure 3.9: GOES images for a day with marine layer breakup (second largest absolute ramp) in SDG&E territory; same as Figure 3.5 but for Sep. 18, 2010 (Figure 3.4b). The largest aggregated 1 hour absolute ramp for this period was 44% of PV capacity and occurred from 830 to 930 PST.
3.6 Conclusions

Aggregate ramp rates of 192 distributed photovoltaic (PV) systems installed in California were analyzed and compared to modeled power calculated from satellite (SAW) and measured (CIMIS) irradiances. Over one year the largest hourly aggregate absolute ramp in California (considering all three territories) was a 30% increase and hourly ramps over 23% occurred only about once per day. By investigating ramp rates of aggregate irradiance at clear sky conditions, the largest absolute ramp rates were found to be predominantly related to the rising and setting sun, and are therefore predictable. Weather-induced ramp rates with reference to a 30-day average of diurnal power output, on the other hand, can elucidate unexpected variations. The largest weather-induced hourly ramp in California with reference to the 30-day average was 20% and occurred due to a widespread decrease in cloud cover on an afternoon. In a very high PV penetration scenario, if such ramps hit the operator unprepared, they may indeed cause reliability challenges and additional costs for the system operator. Other weather-induced ramps were 16% or less per hour.

In SDG&E territory, the largest hourly absolute and weather-induced ramps were 60% and 55% of PTC capacity respectively. However, many of the largest ramp rates are caused by summer marine layer breakup when cloud evaporation coincides with an increase in solar altitude nearly every morning (e.g. the day with the second and fourth largest absolute ramp). During the winter months, the ramp rates are mainly caused by the winter frontal storm systems; when fast-moving storm systems move into the area (creating a large down ramp) or out of the area (creating a large up-ramp).
This analysis was focused on distributed generation systems that are relatively well distributed across the state. Groups of larger but less geographically diverse systems may experience larger weather induced ramps. Table 3.1 compares the largest absolute and weather-induced ramp rates for different utility territories within California.

Table 3.1: Comparison of the largest ramp rates for different utility territories.

<table>
<thead>
<tr>
<th>Ramp Type</th>
<th>SDG&amp;E</th>
<th>SCE</th>
<th>PG&amp;E</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Largest absolute ramp</td>
<td>60.4%</td>
<td>31.4%</td>
<td>30.2%</td>
<td>30.9%</td>
</tr>
<tr>
<td>Largest weather induced ramp</td>
<td>55.5%</td>
<td>29.8%</td>
<td>28.1%</td>
<td>20.3%</td>
</tr>
</tbody>
</table>

**Acknowledgments**

Chapter 3, in full, has been previously published as Jamaly, M., Bosch, J.L., and Kleissl, J., 2013, “Aggregate Ramp Rates of Distributed Photovoltaic Systems in San Diego County”, IEEE Transactions on Sustainable Energy. The dissertation author was the primary investigator and author of this paper.
Chapter 4 Robust Cloud Motion Estimation by Spatiotemporal Correlation Analysis of Irradiance Data

The power output from solar photovoltaic (PV) power plants is usually more variable than conventional power generation sources. Variability is the main challenge for integration of large amounts of PV power plants into the electricity grid (Marcos et al., 2011). The ability to forecast actual variability of solar distributed generation (DG) will allow grid operators to better accommodate the variable electricity generation for resource adequacy considerations, such as scheduling and dispatching of power.

Besides predictable solar variability according to diurnal and annual irradiance patterns, the main source of spatiotemporal variability in the solar resource is transient clouds and that variability is related to the cloud optical depth and speed. Cloud motion is the main input to most short-term solar variability and forecast models (Arias-Castro et al., 2014; Hoff and Perez, 2010; Lave and Kleissl, 2013; Chow et al., 2011; Marquez and Coimbra, 2013; Perez et al., 2010; Yang et al., 2014a; Yang et al., 2014b; Lorenzo et al., 2014). Therefore, cloud motion estimation has been extensively investigated recently (Bosch et al., 2013; Bosch and Kleissl, 2013; Fung et al., 2013; Huang et al., 2013; Quesada-Ruiz et al., 2014; Chow et al., 2015). Accurate cloud motion vectors are critical for solar forecast, interpolation, and variability analyses. In spatiotemporal Kriging method (Chapter 5), for the cases where the cloud motion is unknown, the cloud motion should be estimated within the Kriging process (which is necessary for anisotropic Kriging method).

Vega-Riveros and Jabbour (1989) reviewed various techniques related to the motion analysis and detection. Motion analysis methods are either based on the direct
numerical solution of the optical flow constraint equation (method of differentials) or correspondence-based approaches, where image features are identified and tracked to measure their displacement. These measurements are then used to calculate the displacement of the object as a whole. Estimating cloud motion for sky imaging and satellite data by solving the optical flow equation incurs less computational expense. However, it has many restrictions. Therefore, most of the methods for estimation of the Cloud Motion Vectors (CMVs) are developed using correspondence-based approaches.

In general, CMVs are obtained by first locating salient image features such as brightness gradients, corners, cloud edges, or brightness temperature gradients (Bedka and Mecikalski, 2005; Menzel, 2001). Then, assuming the features do not change significantly over a short interval, CMVs are computed by tracking the features in successive images.

CMVs have been obtained using sky imaging devices (Marquez et al., 2013) for very short-term solar forecasts up to 20 min ahead. Moreover, CMVs have been estimated from satellite imagery (Perez and Hoff, 2013; Menzel, 2001; Hammer et al., 1999; Leese et al., 1971). Escrig et al. (2013) applied multispectral tests and binary cross-correlations for cloud motion estimation using geostationary satellite imagery. They applied coherence and quality control tests to the resulting motion vectors and proposed new thresholds for infrared and visible tests. Fuh and Maragos (1991) developed a model for estimating the displacement field in spatiotemporal image sequences that allows for affine cloud shape deformations. The model is based on the block matching method (which is based on the same principal as the cross-correlation method presented later) and parameters were found using a least-squares algorithm. Post-smoothing the velocity field
via spatiotemporal vector median filtering almost always improves the performance of the matching algorithm. However, block matching has a higher computational complexity.

Farnebäck (2003) developed a method for motion estimation based on a two-frame algorithm. The first step is to approximate each neighborhood of both frames by quadratic polynomials. Then, a method to estimate displacement fields from the polynomial expansion coefficients was derived. The main weakness of the algorithm is the assumption of a slowly varying displacement field, causing discontinuities to be smoothed out. Hammer et al. (1999) developed a statistical method based on conditional probabilities to compute CMVs and predict solar radiation up to 2 h ahead. Lorenz et al. (2004) used a similar method (applying extrapolation of motion assuming persistence of cloud speed, size, and shape) to obtain solar radiation forecasts up to 6 hours ahead. For longer forecast time horizons, non-linearities in atmospheric motion and cloud formation and evaporation cause Numerical Weather Prediction (NWP) models to outperform satellite-based CMV forecasts (Perez et al., 2010). Arking et al. (1978) applied Fourier phase difference technique which allows motion estimates to be made for individual spatial frequencies related to cloud pattern dimensions. However in the presence of mixtures of motions, changes in cloud shape and edge effects, the cross-correlation scheme yields a more reliable estimate of cloud motion than the phase difference technique.

Since CMV estimation by either sky imaging, satellite data, or NWP lack granularity and computational efficiency, local ground measurements of cloud speed are advantageous for short-term solar variability and solar forecasting (Bosch et al., 2013).
Bosch and Kleissl (2013) showed that cloud motion can be detected from spatio-temporal irradiance or power measurements across a utility-scale PV plant from the timing of cloud arrival at three different points.

Prior methods using ground data were predicated upon sparse data. The analysis in this study is motivated by the increased availability of dense PV power output observations in urban areas with spatial resolution on the order of 100s of meters. Actual PV power output can be converted to clear sky index (see e.g. Engerer and Mills, 2014) and then cloud motion could be estimated just like if the PV system was an irradiance sensor. Therefore the success of two algorithms in detecting cloud motion is estimated from simulated dense ground data: cross-spectral analysis (CSA) and the cross-correlation method (CCM). In CSA, the cloud speed and direction are estimated by cross-spectral analysis of the irradiance data at some given locations (sites) through the domain (Inoue et al., 2012; Shinozaki et al., 2014). The CSA method suggested by Inoue et al. (2012) and Shinozaki et al. (2014) is restricted by the spatial arrangement of the sites such that the cloud direction may be inaccurate if there are only a few distinct angles between the pairs of the chosen sites. To remove the restriction, a new CSA approach for cloud motion direction is proposed by selecting the direction with least variation for all the velocities in the cloud motion direction.

In CCM, the cloud motion is estimated by comparing correlation between spatial irradiance data at two or more time steps (Hamill and Nehrkorn, 1993). The CCM suggested by Hamill and Nehrkorn (1993) is generalized for cloud movement estimation using unstructured ground measured data. Moreover, to compare the consistency of the method when applied to different scales, CCM is applied by considering the whole
domain as well as smaller subdomains. Also, to ensure reliable cloud motion estimation, quality control (QC) is added to the CSA and CCM analyses including removing conditions with low variability and less correlated sites.

The algorithms are tested only on simulated ground data, which is advantageous because the true cloud speed is known. In real datasets the true cloud speed is unknown and such data suffer from spatial heterogeneity in surface and atmospheric conditions that manifests in spatial differences in cloud motion vectors. Such heterogeneities can be avoided in a simulated dataset and the cloud motion estimation results are therefore expected to be more generalizable. In Chapter 5 the CSA and CCM methods are applied to real data for spatiotemporal interpolation or forecast of solar irradiance.

4.1 Dataset

The analysis has been performed using two spatially and temporally resolved simulated irradiance datasets generated from large eddy simulation (LES). LES is a three-dimensional computational fluid mechanics technique that numerically integrates the Navier-Stokes equations. The momentum, temperature, and moisture transport is simulated at each grid point. High spatial and temporal resolution allows simulating the large turbulent motions in the atmospheric boundary layer explicitly and LES therefore produces more accurate wind, temperature, moisture, and cloud fields than other techniques. Periodic boundary conditions in the horizontal directions are used to represent an infinitely long, homogeneous domain that allows atmospheric turbulence to develop in a realistic manner. LES is forced by a geostrophic wind at the top of the domain. Surface fluxes of heat and water largely determine the relative humidity in the
boundary layer and whether clouds will form. We apply the well-validated UCLA-LES using the same settings as Ghonima et al. (2016). Simulated datasets are considered since LES wind vectors at the average cloud height can be considered as the reference cloud motion.

4.1.1 RICO simulation

In the first simulation, a spatial domain of 2540 m x 2540 m (128 x 128 grid points) with boundary and initial conditions from the rain in cumulus over the ocean (RICO) field study (vanZanten et al., 2011) centered at 18.0° N, 61.8° W is setup. The simulation is performed up to 4000 m height resolved by 100 grid points. The precipitating RICO case with boundary layer moisture in the initial profile equal to 12.35 g/kg is simulated. Following 4 hours of spinup, 10 sec liquid water path (LWP) aggregated from cloud base to cloud top is output over a 30 min interval. Also, a representative wind speed vector is output at each time step; the two velocity components are $u(x,y,z_c,t)$ and $v(x,y,z_c,t)$, where $z_c$ is average cloud height. The wind velocity is considered as the reference cloud motion and compared against estimated cloud motion in Section 4.3.

4.1.2 CGILS simulation

The second simulated dataset consists of simulated LWP obtained by the CGILS (CFMIP-GCSS Intercomparison of Large-Eddy and Single-Column Models) stratocumulus cloud over-land case (see Zhang et al., 2009) with Bowen ratio of 0.1 (moist surface) near the California Coast (35° N, 125° W). The spatial domain covers 2400 m x 2400 m resolved by 96 x 96 grid points in the horizontal and extends to 965 m height in
the vertical using 193 grid points. 10 sec resolution LWP is output for 24 hours (8641 time steps) starting at midnight. Also, a representative wind speed vector at the average cloud height is output every 1200 sec. Since the cloud deck is overcast till noon and breaks up in the afternoon, the cloud speed is estimated separately during 0800-1200 h and 1200 - 1500 h. The CGILS dataset presents more challenging conditions for cloud motion estimation than RICO: (i) Overcast stratocumulus (CGILS) lack the sharp transition between cloud boundaries and clear sky that are common for cumulus; (ii) Cloud motion is unsteady for CGILS while it is steady for RICO. Unsteady cloud motion is a challenge for the CSA method, which relies on temporal averaging.

4.1.3 Converting LWP to clear sky index

The 1D shortwave radiation scheme proposed by Slingo (1989) is applied to calculate total transmission coefficient (equivalent to clear sky index $k_t$, the ratio of global irradiation to the global irradiation in clear sky condition) using the LWP and the equivalent radius of the drop size distribution ($r_e$). According to Hess et al. (1998), the typical effective radius for cumulus clouds is 5.8 μm (clean) and 4.0 μm (polluted), so 5 μm is chosen for the first simulation. The effective radius was set to be 12.7 μm for the stratocumulus cloud in the second simulation. Note that the Slingo model with 4 spectral bands is applied for simplicity but the full 24 bands model can be used as well. The inverse relation between LWP and $k_t$ was confirmed visually.
4.2 Cloud Speed Methodology

4.2.1 Cross-spectral analysis (CSA)

In CSA, the cloud speed and direction are estimated by analyzing the time series of the irradiance data at the given sites through the domain (Inoue et al., 2012 and Shinozaki et al., 2014).

4.2.1.1 Classical CSA definition

The cloud speed and direction are estimated by CSA of the observed \( kt \) at some given locations. As suggested by Inoue et al. (2012) and Shinozaki et al. (2014), for each pair of sites \( i \) with \( i = 1, \ldots, n \), the time difference \( \Delta T_i \) that maximizes the delayed (cross) correlation between the two \( kt \) time series is obtained. Then, the time delay moment \( M(\theta) \) for all the sites with angle \( \theta \) is calculated using the delay times for all pairs of sites as

\[
M(\theta) = \sum_{\theta-30<\theta_i<\theta+30} \frac{\Delta T_i}{r_i},
\]

where \( r_i \) is the distance between the \( i^{th} \) pair of sites with angle \( \theta_i \). The movement angle \( \theta_m \) is calculated such that the moment \( M(\theta) \) of the time delay is maximized. After that, for each pair of sites, the velocity vector in the \( \theta_m \) direction \( (V_i) \) is calculated as

\[
V_i = \frac{r_i \cos(\theta_i - \theta_m)}{\Delta T_i},
\]

and the cloud speed is chosen as the median of the \( V_i \) of all pairs of sites.

However, there may be no local maximum in the moment and then it is difficult to estimate the movement speed and direction. If there is no clear pattern in the delay time, the moment would approach 0. As suggested by Shinozaki et al. (2014), an error index \( e \) quantifies the confidence.
\[ e = \frac{V(60\%) - V(40\%)}{V(50\%)} \times 100, \]  
Eq. 4.3

where \( V(50\%) \) is the median of \( V_i \) and \( V(40\%) \) and \( V(60\%) \) are the 40\% and 60\% percentiles of \( V_i \) respectively. Shinozaki et al. (2014) mentioned that when a clear movement trend is observed, the error index \( e \) was usually less than 40\%. An unclear movement trend was correlated with error index \( e \) of about 40\% to 70\%; when there was no movement trend, the error index \( e \) was greater than 70\%.

### 4.2.1.2 Robust estimation of the cloud motion

The approach suggested by Inoue et al. (2012) and Shinozaki et al. (2014) is restricted by the spatial arrangement of the sites. This is because the movement angle \( \theta_m \) is the average angle between all sites. In the worst case scenario all sites are in a line with angles equal to 0 or 180, so the moment has nonzero values only at these two points which is not necessarily the cloud motion direction. Therefore the number and range of possible values of \( \theta_m \) are limited by the diversity of angles between the sites.

Instead of the time delay moment in the original CSA (Eq. 4.1), we propose a new approach for cloud motion direction. For a given movement angle \( \theta_s \), the velocity detected by each pair of sites is given by Eq. 4.2. Ideally, the velocity components in the given movement direction for all pairs of the sites should be equal when \( \theta_s = \theta_m \). So, the best candidate for movement direction is the case with least variation for all the velocities. For a given \( \theta_s \), the variation for all the velocities is calculated as

\[ V_{var}(\theta) = \text{var} \left( r_i \cos(\theta_i - \theta) / \Delta T_i \right), \]  
Eq. 4.4

where the sample variance is calculated as
\[ \text{\textit{var}} = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z})^2, \tag{Eq. 4.5} \]

in which \( Z_i = [r_i \cos(\theta_i - \theta_s)] / \Delta T_i \) and \( \bar{Z} = [\sum_{i=1}^{n} Z_i] / n \).

The time delay variation moment \( V_{\text{var}}(\theta_s) \) is calculated for all angles (\( \theta = 1^\circ, \ldots, 360^\circ \)) and the angle for which the minimum occurs is the cloud movement direction (\( \theta_m \)). Also, the cloud movement speed is considered as the median of velocities at this direction (Eq. 4.2).

To get a realistic time delay \( \Delta T_i \) for each pair of sites (which is applied in Eq. 4.4), a maximum limit for allowable time lag (\( \text{\text{max}}_{\text{lag}} \)) is needed. \( \text{\text{max}}_{\text{lag}} \) is critical since a small \( \text{\text{max}}_{\text{lag}} \) could be shorter than the physical cloud transit time while a large \( \text{\text{max}}_{\text{lag}} \) causes the overlapping length of the lagged time series to become too small and, therefore, the obtained \( \Delta T_i \) is not reliable. Since irradiance signals from clear to cloudy (and vice versa) are similar there is a risk of spurious correlation peaks from two different clouds passing two sites. Therefore, different maximum allowable time lags are applied to achieve robust results; \( \text{\text{max}}_{\text{lag}} = [1, 2, 3, 4, 5, 6] \cdot t_{\text{dim}} / 6 \), where \( t_{\text{dim}} \) is length of time series. Cloud movement direction, speed and error index (Eqs. 4, 2 and 3 respectively) are obtained for each \( \text{\text{max}}_{\text{lag}} \) and the best results are considered as the case with the smallest error index. A problem that is unique to this simulated dataset with periodic boundary conditions is that a site is both upwind and downwind of other sites as the motion across the domain boundary is considered. Such correlation peaks associated with motion across domain boundaries are discarded.
4.2.1.3 Excluding sites and data

Quality control (QC) improves the robustness and accuracy of CSA as some pairs of sites have low quality data. Including these low quality data may lead to inaccurate and unreliable cloud motion estimation. Therefore, a new QC approach is developed and applied in the analysis as follows. These steps are critical for reliable results. The specific thresholds in the QC are all chosen empirically, but appear to yield good results for the wide range of cloud conditions simulated here.

A. Remove conditions with low variability

When the variability in $kt$ is small such as in clear conditions, the signal is insufficient to detect cloud direction and speed. Therefore, for each site with given $kt$ time series ($kt_s$), the variation ratio of the site ($\text{var}_s$) is calculated as

$$\text{var}_s = 1 - \frac{\text{mean}(Kt_s)}{\text{max}(Kt_s)}$$

Eq. 4.6

Also, for each site, the cloud cover fraction is calculated as the fraction of time steps with $kt_s < 0.85$. It is empirically found that sites with low variation ratio ($\text{var}_s < 0.1$) or low cloud cover fraction (less than 0.1) lead to inaccurate cloud motion estimation and, hence, these sites are removed from the analysis. In the RICO simulation with total 128 x 128 = 16,384 sites (pixels), 11,103 sites (68%) with low variation ratio or low cloud cover fraction are removed from the analysis, while for the CGILS simulation 5,571 (60%) and 6,318 (69%) out of 9,216 sites are removed during cloud breakup and overcast periods respectively.
B. Remove site pairs that are too distant

The cross-correlation is applied on the remaining pairs of sites. Note that for a given maximum time lag, \( \text{max}_{\text{lag}} \), there is a limit on the cloud speed that can be detected between each pair of sites: the cloud speed for a given pair of sites with distance \( r_i \) can only take on the values of \( r_i / \text{max}_{\text{lag}} \), 2 \( r_i / \text{max}_{\text{lag}} \), \( r_i / \text{max}_{\text{lag}} \), \..., and \( r_i \). If the pair of sites is too distant and/or the length of time series (and therefore \( \text{max}_{\text{lag}} \)) is relatively short (\( r_i / \text{max}_{\text{lag}} >> 1 \)), the velocity between the pair of sites may be overestimated. To ensure reliable velocity field estimation, a limit on distance between pairs of sites should be applied (\( d_{\text{lim}} \)). Without loss of generality and to be able to detect velocity values down to 1 m/s, the limit is applied such that \( r_i / \text{max}_{\text{lag}} = 1 \) m/s. Also, since pairs of sites with distance greater than half of the domain size were empirically found to be less correlated, the \( d_{\text{lim}} \) is set to be less than or equal to half the domain size \( d_{\text{max}} / 2 \). Therefore, for a given \( \text{max}_{\text{lag}} \), \( d_{\text{lim}} \) is calculated as

\[
d_{\text{lim}} = \min \left( \frac{d_{\text{max}}}{2}, \text{max}_{\text{lag}} \times 1 \text{ m/s} \right), \tag{4.7}
\]

where \( d_{\text{max}} \) is the domain size. To improve accuracy of the results and robustness of the method, site pairs with distance greater than \( d_{\text{lim}} \) are not considered in the CSA. For instance, for the case with \( \text{max}_{\text{lag}} = t_{\text{dim}} \), 5.1 \( \times \) 10^6 (or 36%) of the remaining 1.4 \( \times \) 10^7 pairs of sites, 2.4 \( \times \) 10^6 (or 37%) of the remaining 6.6 \( \times \) 10^6 pairs of sites, and 1.5 \( \times \) 10^6 (or 37%) of the remaining 4.2 \( \times \) 10^6 pairs of sites are removed from the analysis in RICO simulation, CGILS simulation during cloud breakup, and overcast periods respectively.
C. Remove less correlated sites

For each pair of sites, maximum and average values of the cross-correlation of the pair \( (\text{max}_{QC} \text{ and } \text{mean}_{QC}) \) are calculated. The cross-correlation QC ratio \( (r_{QC}) \) is then

\[
r_{QC} = 1 - \frac{\text{mean}_{QC}}{\text{max}_{QC}}.
\]  

Eq. 4.8

\( \text{max}_{QC} \) shows the degree of similarity between the two time series. \( r_{QC} \) represents the level of the certainty of the cross-correlation between the site pair; large \( r_{QC} \) means that there is a large maximum cross-correlation relative to the average and, therefore, the cross-correlation result is more reliable. To ensure high quality results, the site pairs with \( \text{max}_{QC} < 0.8 \) or \( r_{QC} < 0.8 \) are removed from the analysis. These empirically-derived thresholds are found to be sufficient for accurate results in all cases. Therefore, for \( \text{max}_{\text{lag}} = t_{dim} \) case, \( 1.1 \times 10^6 \) (out of the remaining \( 8.8 \times 10^6 \) pairs of sites), \( 5.0 \times 10^5 \) (out of the remaining \( 4.2 \times 10^6 \) pairs of sites), and \( 3.2 \times 10^5 \) (out of the remaining \( 2.7 \times 10^6 \) pairs of sites) pairs of sites are removed from the analysis in RICO simulation, CGILS simulation during cloud breakup and overcast periods, respectively.

4.2.1.4 Reduce the density of sites

To reduce computational costs, it will be investigated whether choosing less points with the highest local variabilities are sufficient to achieve the same results. This capability of the method would be advantageous in practice since the site density is usually lower than in the simulated case. Therefore, the whole domain is split into boxes with \( 5 \times 5 \) pixels and at each box the point with the highest variation ratio (as defined in Eq. 4.6) is chosen (625 and 361 sites for the RICO and CGILS simulations respectively).
Geographically restricting the selection of points ensures that a diverse set of pairs with many distances and directions is available.

4.2.2 Cross-correlation method (CCM)

Hamill and Nehrkorn (1993) proposed calculating the cloud velocity and direction of motion through the cross-correlation method (CCM) applied to two consecutive time steps. Hamill and Nehrkorn (1993) applied CCM on satellite data to forecast cloud movement. Here, the method is generalized for cloud movement estimation using unstructured ground measured data by interpolating $kt$ data on a regular structured grid.

The CCM finds the position that best matches each given subset of pixels at the previous time step given the data at the current time step. The CCM yields a cloud vector (direction and speed) with the largest cross-correlation coefficient (CCC) that quantifies the quality of the match. Besides CCC, there are other cost functions including sum of the absolute value of difference (SAVD) and sum of squared difference (SSD); however, CCC is the most robust (Yang et al., 2012). In general, as mentioned by Chow et al. (2011), the CCM obtains accurate motion in heterogeneous areas with high contrast of pixel values, such as clouds with a sharp boundary in a clear sky. Quality control is essential to detect such cases as described below.

4.2.2.1 Calculating velocity components in different domains

The cloud velocity is estimated using data of the whole domain ($X_{dim}$ and $Y_{dim}$ are dimension in $X$ and $Y$ directions, respectively) and smaller rectangular subsets (boxes). The box sizes and searching radii ($X_{radius}$ and $Y_{radius}$) can be different in $X$ and $Y$ directions:
I. Whole domain: CCM is applied on the whole domain by setting the search radii in $X$ and $Y$ directions to 5 pixels. To avoid the box from crossing the domain boundary, box dimensions are equal to $X_{dim}-2X_{radius}$ and $Y_{dim}-2Y_{radius}$. For each velocity component ($U_X$ and $U_Y$ calculated separately), the velocity representative of the whole domain is obtained by finding the maximum correlated frame and calculating the displacement vector at the center of the box.

II. Smaller Boxes (block matching algorithm): the domain is split into boxes (e.g. 5 x 5 pixels) and for each box, ($U_X$, $U_Y$) are obtained by finding the maximum correlated frame and the displacement vector in ($X$, $Y$) direction at the center of each box. Smaller boxes result in a more granular velocity field in the domain.

The exhaustive (full) search method, which is the most computationally expensive block matching algorithm (Yang et al., 2012), is considered in this study. Alternatively, a three-step search (Li et al., 1994) can be used to overcome the cost of the exhaustive search. However, the accuracy is limited by the reduced number of matching candidates. Note that the CCM is very sensitive to the box size and search radius. A small box size can result in insufficient pixels to distinguish the displacement maximum likelihood estimate. On the other hand, a large box size lead to insufficient number of vectors to describe the variation in cloud motion (Yang et al., 2012). Therefore, these parameters are chosen based on the expected range of the physical characteristics of the problem (typical wind speeds, etc.). Also, when there is more than one maximum correlated displacement (e.g. as is the case for periodic boundary conditions in the simulated
dataset), the one with minimum displacement (equals to minimum velocity magnitude) is chosen.

4.2.2.2 Improving cloud speed resolution

Note that the resolution of the velocity vector that can be detected from two subsequent images in X and Y directions are $\Delta X/\Delta t$ and $\Delta Y/\Delta t$ respectively, where $\Delta X$ and $\Delta Y$ are pixel spacing in X and Y directions and $\Delta t$ is the time step. Therefore, valid values for the velocities are $U_X = \{\Delta X/\Delta t, 2 \times \Delta X/\Delta t, \ldots, X_{radius} \times \Delta X/\Delta t\}$ and $U_Y = \{\Delta Y/\Delta t, 2 \times \Delta Y/\Delta t, \ldots, Y_{radius} \times \Delta Y/\Delta t\}$. Setting $X_{radius}$ and $Y_{radius}$ equal to 5 pixels ensure sufficient higher limit. However, if velocities are small such that, e.g. $U_X \Delta t < \Delta X$ or $U_Y \Delta t < \Delta Y$, the estimated velocity components $U_X$ or $U_Y$ will be equal to zero. Moreover, finer resolution (increment) of the velocity vectors are desirable. Computing velocities on additional images beyond two consecutive time steps improve the resolution and the lowest detectable cloud speed. In all cases and at each time step, the CCM results for horizons = 1, 2, 3, 4, and 5 (equal to $\Delta t = 10, 20, 30, 40, \text{ and } 50 \text{ s}$) are calculated and the one with the largest CCC is considered as the estimated cloud velocity. The cloud motion is considered to be zero if the averaged velocity vector equals zero up to horizon = 5.

4.2.2.3 QC

For local consistency as suggested by Hamill and Nehrkorn (1993), the vectors which differ from the local average (within a given distance) by more than a threshold are removed. Also, since the CCM for boxes with low variation, cloud fraction, and/or information content give unreliable and meaningless results, different QC parameters which were derived empirically are considered: (i) the variation ratio of each box ($var_b$)
computed per Eq. 4.6 must be greater than 0.1, (ii) the box cloud cover fraction (the fraction of pixels in the box with \(kt_b < 0.85\)) must be greater than 0.1. (iii) maximum CCC of the box \(maxQC\) must be greater than 0.8, (iv) cross-correlation QC ratio \(r_{QC}\) per Eq. 4.8 must be greater than 0.8. These QC parameters with the specified thresholds along with the local consistency are applied to the CCM for more accurate results (modified CCM).

4.2.2.4 Obtaining Velocity Field by Objective Analysis

After calculating the cloud velocity at the boxes in the domain and applying the QC, an objective analysis is applied to smooth the velocity field and produce a continuous cloud motion pattern. Therefore, each velocity vector is calculated using a multi-pass successive corrections scheme (3 passes) as suggested by Hamill and Nehrkorn (1993). The first estimate is

\[
G1(i, j) = \frac{\sum_{g=1}^{nodes} W^2 D(x_g, y_g)}{\sum_{g=1}^{nodes} W}, \quad \text{Eq. 4.9}
\]

where \(D\) is the displacement vector at location \((x_g, y_g)\). \(W\) is the standard Cressman weighting function as

\[
W_n = \begin{cases} 
\frac{R^2 - d_g^2}{R^2 + d_g^2} & \text{for } d_g < R, \\
0 & \text{for } d_g > R
\end{cases}, \quad \text{Eq. 4.10}
\]

where \(d_g\) is the distance between location \((x_g, y_g)\) and pixel \((i, j)\) and \(R\) is the radius of influence.

The second pass is calculated as
\[ G_2(i, j) = G_1(i, j) + \sum_{g=1}^{\text{nodes}} \left\{ W^2 \left[ D(x_g, y_g) - G_1(x_g, y_g) \right] \right\} \sum_{g=1}^{\text{nodes}} W. \]  

Eq. 4.11

And the third pass (the estimated velocity field) is obtained in the same form as Eq. 4.11 with \( G_3 \) and \( G_2 \) replacing \( G_2 \) and \( G_1 \), respectively. This 3 step scheme is applied for \( U_X \) and \( U_Y \) components separately.

4.3. Results

CCM and CSA methods are first validated on a moving 1D sinusoidal wave (not shown) and then using a 1D transect of the data in the along-wind and cross-wind directions (also not shown). Afterwards, the methods are applied to the 2D simulated data representing the simulated cumulus (RICO case) and stratocumulus cloud (CGILS case).

4.3.1 RICO simulation

The 1800 sec time series of all points in the domain (128x128 grid points) is analyzed. The reference cloud motion vector is obtained through the wind speeds at the average cloud height of 1031 m as 2.61 m/s at 180° (using the standard reference angle as the X axis and the positive angles calculated counter clockwise) in the whole domain. Figure 4.1 shows \( kt \) contours at selected time steps which illustrates cloud movement in the domain. While cloud motion is visually discernible to be along the negative X axis, motion detection is complicated by clouds changing shape and evaporating / forming in time.
Figure 4.1: $kt$ contour of the simulated dataset obtained from RICO simulation at different time steps; (a) 08:20:00, (b) 08:22:00, (c) 08:24:00, and (d) 08:26:00.

Figure 4.2 shows time series of the domain averaged variation ratio and cloud cover fraction. As described in Section 4.2, to ensure reliable cloud motion estimation, the points with low variation ratio or low cloud cover fraction are excluded from analysis.
4.3.1.1 CSA results of the RICO simulation

The CSA is applied on the selected 16,384 (7.8 x 10⁶ pairs after QC) and the 625 sites with highest variation ratio (5.8 x 10⁴ pairs after QC) to estimate the cloud motion. Figure 4.3 shows the time delay variation (Eq. 4.4) as a function of angle for $\text{max}_\text{lag} = t_{\text{dim}}/6$ after application of the distance limit (Eq. 4.7). Figure 4.3 confirms that choosing a subset of points (625 sites) is sufficient to obtain representative results.

The original CSA (Inoue et al. (2012), not shown) using all the 16,384 sites yields a cloud motion of 1.34 m/s at 111° with error index $e = 21\%$. The 33% improvement in speed and 52° absolute error reduction in cloud motion direction estimation as well as the reduction in the error index (from 21% to 8%) confirms that the proposed CSA is more accurate and robust than the original CSA and, therefore, the new robust CSA will be applied hereinafter.
Figure 4.3: Calculation of the time delay variation moment (Eq. 4.4) over all angles by considering pairs of sites within the distance limit using all points in the domain (16384 sites) and 625 sites with the highest variability. The caption indicates the calculated cloud direction, speed, and error index using all points as well as applied maximum time lag and maximum allowable distance between pairs of sites. The reference velocity at the average cloud height was 2.61 m/s at 180°.

Table 4.1 compares the results of cloud estimation of RICO simulation using the CSA by considering different sites. The results using all the points in the domain (16,384 points) are presented. Then, the QC steps described in Section 4.2.1.3 are applied one at a time and the accuracy of cloud motion vectors are quantified. Also, results using the sites with the highest variability (after QC) are presented. The CSA (after QC as described in Section 4.2.1.3) estimates the cloud motion in the whole domain with average relative mean bias error between the reference value and calculated velocity (rMBE) and difference between the reference cloud direction and calculated cloud motion direction (absolute error) equal to 15.7% and 17° respectively. All QC steps need to be applied to ensure accurate results. Also the average error index equals to 10% confirming the reliability of the method.
Table 4.1: Validation of cloud velocity for the RICO simulation by CSA. The effect of the QC steps described in Section 3.1.3 is quantified for the dataset with all 16,384 points: removing sites with low variability (step 1), removing sites with low variability and site pairs that are too distant (step 2), removing sites with low variability, site pairs that are too distant, and less correlated sites (step 3). 11,103 sites (68%) with low variation ratio or low cloud cover fraction, $5.1 \times 10^6$ (or 36%) of the remaining $1.4 \times 10^7$ pairs of sites (for the case with $\text{max}_{\text{lag}} = \text{tdim}$), and $1.1 \times 10^6$ (or 12.5%) of the remaining $8.8 \times 10^6$ pairs of sites (for the case with $\text{max}_{\text{lag}} = \text{tdim}$) are removed from the analysis in steps 1, 2, and 3, respectively. Also, results using only the 625 the sites with the highest variability (after QC) are presented. rMBE indicates the relative mean bias error between the reference value and CCM velocity and the absolute error indicates the difference between the reference cloud direction and calculated cloud direction. The average reference cloud motion in the whole domain is 2.61 m/s at $180^\circ$. The error index ($e$) quantifies the confidence of the results (Eq. 4.3).

<table>
<thead>
<tr>
<th>Sites</th>
<th>Velocity</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speed (m/s)</td>
<td>rMBE (%)</td>
</tr>
<tr>
<td>All Points (16,384)</td>
<td>0.7</td>
<td>73.2</td>
</tr>
<tr>
<td>All Points, QC Step 1</td>
<td>1.2</td>
<td>54.0</td>
</tr>
<tr>
<td>All Points, QC Step 2</td>
<td>1.4</td>
<td>46.4</td>
</tr>
<tr>
<td>All Points, QC step 3</td>
<td>2.2</td>
<td>15.7</td>
</tr>
<tr>
<td>625 Sites</td>
<td>2.2</td>
<td>15.7</td>
</tr>
</tbody>
</table>

4.3.1.2 CCM results of the RICO simulation

The cloud velocity is estimated through the CCM applied to all grid points by considering the whole domain as a box. For more spatially granular estimation of the cloud motion, boxes with 5 x 5 pixels are chosen to calculate the velocity field using the objective analysis. QC criteria as described in Section 4.2.2.3 are applied to both the whole domain and the 5 x 5 boxes (modified CCM).

Figure 4.4 shows the detailed modified CCM calculated cloud motion field at a single time step using boxes with 5 x 5 pixels through the whole domain. At this time step, the best result is achieved at horizon = 30 s. The cloud motion vector field is homogeneous.
Figure 4.4: Cloud velocity field calculation by modified CCM using boxes with 5 x 5 pixels through the whole domain at 08:27:40. The caption indicates the average calculated cloud speed.

Figure 4.5 shows time series of the calculated cloud velocity by applying the local consistency and without QC criteria mentioned in Section 4.2.2.3 (original CCM) as well as after applying local consistency and QC (modified CCM). The improvements from applying QC are 21% in the speed and a 6° in the direction. Also the temporal variability in cloud motion vector is reduced. Therefore, the more robust and accurate modified CCM is applied hereinafter.
Figure 4.5: Time series of CCM cloud (a) speed and (b) direction by using boxes with 5 x 5 pixels through the whole domain for original CCM (dashed line) and modified CCM (solid line). The dotted line shows the reference averaged wind (a) speed and (b) direction at average cloud height. The caption indicates the time-averaged cloud speed and direction.
To confirm consistency of the results and investigate the dependence on horizon, Table 4.2 shows the modified CCM results of the RICO simulation using the entire domain as a box as well as boxes with 5 x 5 pixels through the whole domain for horizons up to 50 s. In general, for this case with almost constant velocity over time, the CCM results using boxes with 5 x 5 pixels are in good agreement with actual cloud speed and direction with average rMBE and absolute error in the whole domain equal to 3.4% and 2° respectively. For horizons less than 30 s the cloud speed errors are larger, but speed and direction errors converge for horizons of 30 s or more.

Table 4.2: CCM cloud velocity of RICO simulation using boxes with 5 x 5 pixels through the domain at different horizons $H$. rMBE indicates the relative mean bias error between the reference value and calculated velocity and the absolute error indicates the difference between the reference cloud direction and calculated cloud direction. The average reference cloud motion in the whole domain is 2.61 m/s at 180°.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Resolution</th>
<th>Velocity</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Speed (m/s)</td>
<td>rMBE (%)</td>
</tr>
<tr>
<td>$H = 1$ (Δ$t = 10$ s)</td>
<td>1 Box</td>
<td>2.00</td>
<td>23.4</td>
</tr>
<tr>
<td></td>
<td>5x5 Boxes</td>
<td>2.42</td>
<td>7.3</td>
</tr>
<tr>
<td>$H = 2$ (Δ$t = 20$ s)</td>
<td>1 Box</td>
<td>2.01</td>
<td>23.0</td>
</tr>
<tr>
<td></td>
<td>5x5 Boxes</td>
<td>2.49</td>
<td>4.6</td>
</tr>
<tr>
<td>$H = 3$ (Δ$t = 30$ s)</td>
<td>1 Box</td>
<td>2.18</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>5x5 Boxes</td>
<td>2.55</td>
<td>2.3</td>
</tr>
<tr>
<td>$H = 4$ (Δ$t = 40$ s)</td>
<td>1 Box</td>
<td>2.37</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>5x5 Boxes</td>
<td>2.51</td>
<td>3.8</td>
</tr>
<tr>
<td>$H = 5$ (Δ$t = 50$ s)</td>
<td>1 Box</td>
<td>2.33</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>5x5 Boxes</td>
<td>2.63</td>
<td>-0.8</td>
</tr>
</tbody>
</table>

The CCM that uses the entire domain as a box estimates actual cloud motion direction perfectly (180°). For cloud speed, however, the results at horizon = 1 are not that accurate (average rMBE equals to 23.4%). This confirms that, due to physical restriction, the CCM by choosing the entire domain as a box just gives a rough estimate.
and CCM method using boxes with 5 x 5 pixels through the domain should be applied to obtain more detailed results. Also, in such a case, considering CCM method at larger horizons provide more accurate results as suggested by Table 4.2, which illustrates the importance of calculating the cloud velocity using larger horizons. The optimum horizon is not generalizable, but depends on the cloud speed, time step, and station distance for a particular setup.

4.3.2 CGILS simulation

The CGILS case is split into the overcast and breakup sections of 10800 sec and 14400 sec duration with 10 sec resolution using all 96x96 equally spaced grid points in the domain (Figures 4.6 & 4.7). While the reference cloud motion direction is constant at 280° over the CGILS simulation, the speed varies: during the overcast period the reference speed decreases from 9.9 to 7.5 m/s with an average of 8.6 m/s; during the breakup period the speed is more constant with 20 min means varying from 5.1 to 5.8 m/s with an average of 5.3 m/s.
Figure 4.6: $kt$ contour of the simulated dataset obtained from CGILS simulation at different time steps during the cloud breakup period; (a) 13:10:00, (b) 13:11:00, (c) 13:12:00, and (d) 13:13:00. The black dashed box represents the location of a subdomain which is used for further analysis in Section 4.3.2.2.
Figure 4.7: Same as Figure 4.6 but for CGILS simulation at different time steps during the overcast period: (a) 11:10:00, (b) 11:11:00, (c) 11:12:00, and (d) 11:13:00. The black dashed box represents the location of a subdomain which is used for further analysis.

### 4.3.2.1 CSA and CCM results of the CGILS simulation

CCM and CSA are applied to estimate the cloud motion in CGILS simulation during the cloud breakup and overcast periods (Table 4.3). In CSA, 361 sites with highest variability within the whole domain are applied. CCM is applied by considering the entire domain as a box as well as using boxes with 5 x 5 pixels.

Similar to Figure 4.5, Figure 4.8 shows the detailed CCM calculated cloud motion field at a time step during each period using boxes with 5 x 5 pixels through the whole domain. The best cloud motion is estimated at horizon = 40 s and 10 s for the breakup
(Figure 4.8a) and overcast (Figure 4.8b) periods respectively. While the CGILS cloud motion is not constant, the applied QC ensures local consistency and smoothness of the cloud motion field through the objective analysis.

Figure 4.8: Same as Figure 4.4 but at (a) 13:38:10 during the cloud breakup period and (b) 09:38:10 during the overcast period.
Table 4.3: Cloud velocity calculation of CGILS simulation during cloud breakup and overcast periods by CSA (similar to Table 4.1) and CCM (similar to Table 4.2). The averaged cloud cover fraction and variation ratio (Eq. 4.6) are, respectively, equal to 0.71 (0.67) and 0.37 (0.34) in the whole domain (subdomain) during the cloud breakup period while these values are equal to 1 (1) and 0.36 (0.16) in the whole domain (subdomain) during the overcast period respectively. Also, the average reference cloud motion in the whole domain is 5.32 m/s at 279° and 8.58 m/s at 281° during cloud breakup and overcast periods respectively.

<table>
<thead>
<tr>
<th>Time</th>
<th>Method</th>
<th>Domain</th>
<th>Res.</th>
<th>Speed (m/s)</th>
<th>rMBE (%)</th>
<th>e (%)</th>
<th>Angle (°)</th>
<th>Absolute Error (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200 - 1500 h</td>
<td>CSA</td>
<td>Whole Domain</td>
<td>361 Sites</td>
<td>0.29</td>
<td>94.5</td>
<td>82</td>
<td>334</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sub-domain</td>
<td>30 Sites</td>
<td>4.71</td>
<td>11.7</td>
<td>3</td>
<td>289</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>CCM</td>
<td>Whole Domain</td>
<td>1 Box</td>
<td>5.12</td>
<td>3.8</td>
<td>-</td>
<td>270</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5x5 Boxes</td>
<td>5.01</td>
<td>5.8</td>
<td>-</td>
<td>276</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sub-domain</td>
<td>1 Box</td>
<td>5.18</td>
<td>2.6</td>
<td>-</td>
<td>275</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5x5 Boxes</td>
<td>5.26</td>
<td>1.1</td>
<td>-</td>
<td>276</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>0800 - 1200 h</td>
<td>CSA</td>
<td>Whole Domain</td>
<td>361 Sites</td>
<td>1.50</td>
<td>82.5</td>
<td>19</td>
<td>353</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sub-domain</td>
<td>30 Sites</td>
<td>2.50</td>
<td>70.9</td>
<td>40</td>
<td>345</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>CCM</td>
<td>Whole Domain</td>
<td>1 Box</td>
<td>8.69</td>
<td>-1.3</td>
<td>-</td>
<td>287</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5x5 Boxes</td>
<td>8.53</td>
<td>0.6</td>
<td>-</td>
<td>281</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sub-domain</td>
<td>1 Box</td>
<td>8.48</td>
<td>1.2</td>
<td>-</td>
<td>283</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5x5 Boxes</td>
<td>8.61</td>
<td>-0.3</td>
<td>-</td>
<td>279</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3 confirms that, during the cloud breakup (with average cloud cover fraction equal to 0.71 in the whole domain), CCM results are in agreement with the average reference cloud speed (5.32 m/s) and direction (279°), with average rMBE of speed and absolute error of direction equal to 4.8% and 6° respectively. On the other hand, the CSA results are not accurate during cloud breakup; the average rMBE and absolute error are equal to 94.5% and 55° for the cloud motion speed and direction respectively. At least the average error index of 82% confirms that the results are not reliable. In the overcast period (with cloud cover fraction equal to 1), CCM results are accurate with average rMBE and absolute error equal to -0.3% and 3° respectively. In this
period, the CSA results are again not accurate; the average rMBE and absolute error equal to 82.5% and 72° for the cloud motion speed and direction respectively and the average error index equals to 19%.

4.3.2.2 CCM results of a subdomain of the CGILS simulation

The quality control ensures reliable results as the points with low information content (including low variability and low cloud cover fraction) are removed from the analysis. Given the spatial homogeneity in the domain and boundary conditions, the cloud velocity field should be consistent in the whole domain. To evaluate the performance of the CCM in a smaller region and with less available data, subdomains with 25 x 80 (the black box in Figure 4.6) and 15 x 90 grid points (the black box in Figure 4.7) are considered during cloud breakup and overcast periods respectively. While the locations and sizes of the subdomains were chosen arbitrarily they were aligned with the cloud motion direction to resemble an idealized 1-dimensional setup.

Boxes with 5 x 5 pixels as well as a box covering the entire subdomain are considered in CCM. Similar to the analysis for the whole domain, the cloud fields in the subdomains are estimated using CCM up to horizon = 50 sec and the best CCM results are presented in Table 4.3. The cloud motion speed and direction is estimated using CCM by applying 5 x 5 boxes in the subdomains with average rMBE (absolute error) equal to 5.8% (3°) and -0.3% (2°) during the cloud breakup and overcast periods respectively. The results confirm the accuracy of the CCM method in the subdomain and the consistency with the results for the whole domain.
4.3.2.3 CSA results of a subdomain of the CGILS simulation

CSA results are not accurate in the whole domain and, therefore, it should be tested if the method performs better in the subdomains. The subdomains described in Section 4.3.2 are considered for the CGILS simulation during the cloud breakup and overcast periods. 30 sites with highest variability through the subdomains are applied in CSA.

Table 4.3 shows the cloud estimation results for the subdomains using CSA. During the cloud breakup, in comparison with the results for the whole domain, the CSA method provides more accurate estimates of the cloud motion in the subdomain with average rMBE and absolute error equal to 11.7% and 10° for the cloud motion speed and direction respectively and average error index equals to 3%. In the overcast period, similar to the results for the whole domain, the CSA results are not accurate with the average rMBE and absolute error equal to 70.9% and 64° for the cloud motion speed and direction respectively and the average error index equals to 40%.

4.3.2.3 Analysis of subintervals of the CGILS simulation

The results of the CGILS simulation during the cloud breakup show that CSA is unable to detect the cloud motion pattern in the whole domain while this pattern is detected in the smaller subdomain. Moreover, during the overcast period with cloud cover fraction equals to 1, results illustrate that CSA is unable to detect the cloud motion pattern, even in the smaller subdomain. This is related to the fact that, in CSA, a constant velocity vector (as the average of the velocity field over the whole period under investigation) is assumed and temporal changes in cloud speed and/or direction degrade
the accuracy of the CSA method. For high cloud cover fractions, the probability of the temporal changes in cloud motion increases and CSA is unable to detect the cloud motion pattern in these situations.

To understand the maximum length of subintervals for the CSA, the entire time series in each period (during cloud breakup and overcast) is split into subintervals and CSA is applied to estimate the cloud motion in each subinterval. The lengths of subintervals are varied from 5 to 120 min in 5 min increments. Figures 4.9 & 4.10 show the estimated cloud motion speed and direction using CSA in 20 min subintervals. Also, average relative mean absolute error (rMAE) between the reference cloud speed and the estimated cloud speed using CSA at each subinterval and mean absolute error (MAE) between the reference cloud direction and cloud direction estimated by CSA at each subinterval are calculated. Using the 20 min subintervals, the CSA estimates the cloud motion in the cloud breakup (overcast) period with average rMAE and rMBE of speed equal to 15.8% (22.2%) and 15.8% (20.7%) respectively as well as MAE and absolute error of direction equal to 7° (23°) and 5° (23°) respectively. While this is a large improvement over using CSA for the entire time series; the CCM results are still more accurate in both cases.
Figure 4.9: Time series of cloud (a) speed and (b) direction by CSA using 20 min subintervals during the cloud breakup period through the whole domain. The CCM results and reference values are illustrated as well. The caption indicates the average relative mean absolute error (rMAE) between cloud speed calculated by CSA and the reference value (in a) and the mean absolute error (MAE) between cloud direction calculated by CSA and the reference cloud direction.
Figure 4.10: Same as Figure 4.9 but for the overcast period.

Figure 4.11 shows the cloud motion estimation results for CGILS simulation during the cloud breakup and overcast conditions. For each subinterval length, the cloud motion speed and direction are calculated in each subinterval and the averaged values over all the subintervals (entire timeseries) are presented. The analysis is performed in the whole domains as well as the subdomains in each period.
Figure 4.11: Cloud velocity estimation of CGILS simulation during cloud (i) breakup and (ii) overcast periods using CSA at different subinterval lengths for the whole domain and subdomain; (a) cloud speed, (b) cloud direction, and (c) error index. For each case, the values are averaged over all the subintervals and the reference values are illustrated as well.
The results confirm that during the cloud breakup period, the cloud motion speed can be estimated with average rMBE less than 20% by choosing subintervals up to 120 min. However, during the overcast period, choosing subintervals larger than 30 min will lead to inaccurate (rMBE > 27%) and unreliable results in both the whole domain and the subdomain.

4.4 Conclusions

The cloud motion has been estimated by CSA and CCM using two spatially and temporally resolved simulated irradiance datasets generated from large eddy simulation (LES). CCM estimates cloud motion by comparing correlation between $k_t$ data at two or more time steps. CCM is applied by considering the whole domain as a box. For spatially resolved estimation of the cloud motion the domain is subdivided into boxes with 5 x 5 pixels and the velocity field is post-processed using objective analysis. Quality control (QC) is performed to remove data with low information content (including low variability and low cloud cover fraction), which ensures reliable results. The results show 33% (52°) and 21% (6°) improvement in the cloud motion speed (direction) estimation using the modified CSA and CCM over the original methods (without QC), respectively. Moreover, estimation of the cloud cover fraction and variation ratio may help to improve cloud motion estimation which, in turn, will increase the accuracy of solar forecasting algorithms.

In general, CCM results are accurate in all cases and the CCM method using boxes with 5 x 5 pixels in the domain provides more reliable results for the cloud motion speed. In this study, CCM results are calculated up to a horizon = 50 sec, which was
empirically found to be sufficient to obtain the best results. However, in general, the best forecast horizon is mainly a function of cloud speed and distance between the sites. In real situations with unknown cloud speed, the maximum horizon can be adjusted or even detected automatically according to the specifications of the problem. Specifically the variation ratio and cloud cover fraction, where the former depends on how frequently the clouds are passing while the latter represents the fraction of the area shaded by clouds in the domain, affect the optimum horizon.

CSA estimates the cloud speed and direction by cross-spectral analysis of $kt$ data at all grid points in the domain. Computational costs can be reduced by choosing sites with the highest local variability (625 and 361 sites in the whole domain of RICO and CGILS, respectively) and it was shown that this does not negatively impact the accuracy of the results. Since there are usually less sites available in reality, the number of sites was further reduced (not shown) to as few as 36 (25) sites for RICO (CGILS) and the accuracy of the results was maintained. CSA results are reliable for the cases with low cloud cover fractions; for the RICO case (for both the whole domain and the subdomain) and for the subdomain in the CGILS simulation during the cloud breakup. Since the entire time series is considered simultaneously in the CSA approach, a constant velocity vector (as the average of the velocity field over the whole period under investigation) is assumed, i.e., the cloud motion is assumed to be spatially homogeneous and steady or at most slowly varying in time, such that it could be assumed to be piecewise steady. Therefore, temporal changes in cloud speed and/or direction degrade the accuracy of the CSA method. Moreover, in the cases with relatively high cloud cover fractions, the problem becomes more complicated and the probability of temporal changes in cloud
motion increases. Therefore, CSA is unable to detect the cloud motion pattern for the whole time series in these situations, however, CSA estimates cloud motion more accurately in shorter time intervals (in which the cloud motion can be assumed to be spatially homogeneous and steady). To obtain acceptable results in CSA, the maximum length of the subintervals are observed to be 120 and 30 min for CGILS simulation during the cloud breakup and overcast periods respectively.

Although the results shown here are obtained using the simulated data where the cloud motion is either constant (RICO) or slightly decreasing (CGILS), the proposed modified CSA and CCM can be applied for a wide range of cloud motion direction and/or speed as shown in Chapter 5.

**Acknowledgements**

Chapter 4, in full, is a reprint of the material as it appears in Jamaly, M., and Kleissl, J., 2017a, “Robust Cloud Motion Estimation by Spatiotemporal Correlation Analysis of Irradiance Data”, Solar Energy. The dissertation author was the primary investigator and author of this paper.
Chapter 5 Spatiotemporal Interpolation and Forecast of Irradiance Data Using Kriging

The Kriging method obtains irradiance at an arbitrary location and time by considering the correlations between observed data. The semivariogram function, an important function in the ordinary Kriging method, is a function describing the degree of spatiotemporal dependence of solar radiation. The semivariogram is defined as the variance of the difference between solar irradiation at two given points in space-time coordinates. For an (intrinsically) stationary process, the semivariogram is linearly related to the (parametric) covariance function. Temporal stationarity (and isotropy) is usually achieved by detrending solar irradiance time series usually through the clear-sky index (the ratio of irradiation to the irradiation in clear sky condition) or clearness index (the ratio of irradiation to the extraterrestrial irradiation) as suggested by Inman et al. (2013). The non-dimensional clear-sky index \( (k_t) \) is applied in this study. Spatial stationarity is usually valid for small domains. However, the assumption of spatial isotropy has been shown to be incorrect as the covariance along the direction of the cloud motion is different than perpendicular to it (Hinkelman, 2013; Perez et al., 2012; Lave and Kleissl, 2013; Arias Castro et al., 2014) and, therefore, the anisotropic Kriging method is considered as well.

In this study, the analysis has been performed and the spatial and spatiotemporal Kriging methods are first validated by using two spatially and temporally resolved artificial irradiance datasets generated from large eddy simulation. Then, the spatiotemporal Kriging method is applied on real irradiance and output power data in
California (Sacramento and San Diego areas) where the cloud motion had to be estimated during the process using Cross-Correlation Method (CCM).

5.1 Procedure of Performing Empirical Ordinary Kriging

The spatiotemporal Kriging method is applied in 3 steps; empirical variogram calculation (using available data), parametric variogram fitting (by choosing appropriate covariance/variogram functions), and performing Ordinary Kriging (using the parametric variogram).

5.1.1 Empirical variogram calculation

The empirical semivariogram is obtained using the observed irradiances. For the spatial problem, the empirical semivariogram is calculated as

\[ \hat{\gamma}(h(l)) = \frac{1}{2N(h(l))} \sum_{(i,j) \in N(h(l))} [Z(x_i) - Z(x_j)]^2, \]

Eq. 5.1

where \( h(l) \) is number of bins (the distances are divided into \( L \) bins, \( l=0, \ldots, L \) and \( N(h(l)) \) is the number of pairs in each bins. For the spatio-temporal problem, the empirical semivariogram is calculated as

\[ \hat{\gamma}(h(l), u) = \frac{1}{2N(h(l), u)} \sum_{(i,j,m,n) \in N(h(l), u)} [Z(x_i, t_m) - Z(x_j, t_n)]^2, \]

Eq. 5.2

where \( N(h(l), u) = \{ \forall (i,j,m,n): |m - n| = u, \ |x_i - x_j| \in Bin(h(l)) \} \) in which \( h(l) \) is number of bins (the distances are divided into \( L \) bins, \( l=0, \ldots, L \) and \( N(h(l), u) \) is the number of pairs in each bin and time lags \( u=0, 1, \ldots, \text{max}_t \) where \( \text{max}_t \) represents the maximum time lag.
It is common to set the maximum distance bin, \( h(L) \) equal to half of the maximum distance between the sites. Note that an accurate variogram calculation (a good estimation for the number of bins) is critical to obtain a realistic empirical variogram, which is critical for the robust parametric variogram fitting (Gribov et al., 2001 and Genton, 1998).

5.1.2 Parametric variogram fitting

After calculating the empirical variogram, it is important to model it using a parametric function which will be used in the Kriging method. The most common covariance (and semivariogram) functions are in the form of powered exponential, Whittle-Matern, and Cauchy functions. Gneiting (2002) suggested the general form of stationary non-separable isotropic parametric covariance function as

\[
\mathcal{C}(h, u) = \frac{\sigma^2}{\psi(|u|^2)^{\delta+1}} \varphi\left(\frac{|h|^2}{\psi(|u|^2)}\right). \tag{5.3}
\]

According to the model suggested by Gneiting (2002), powered exponential and Cauchy functions are selected for the spatial \((C_S)\) and temporal \((C_T)\) correlation functions, respectively. Nugget effects (discontinuity at origin due to measurement error) are added to the spatial function as well \((\nu_S)\). Assuming an intrinsically stationary process, the correlation functions are related to the semivariogram as

\[
\gamma(h, u) = \sigma^2[C(0,0) - C(h, u)], \tag{5.4}
\]

where \(\sigma^2\) is the variance of the process that can be estimated using the whole dataset. Therefore, the general parametric isotropic semivariogram is modeled as
\[ \gamma(h, u) = \sigma^2 \left( 1 - (1 - \nu_S) \frac{\exp \left( \frac{-c|h|^{2\zeta}}{(1 + a|u|^{2\alpha})^{\beta \zeta}} \right)}{(1 + a|u|^{2\alpha})^{\delta + \beta}} \right) \]

\text{Eq. 5.5}

\[ - \nu_S I_{S=0} \frac{1}{(1 + a|u|^{2\alpha})^{\delta + \beta}} \]

where \(0 \leq \nu_S \leq 1, \ c > 0, \ a > 0, \ 0 < \zeta \leq 1, \ 0 < \alpha \leq 1, \ 0 \leq \beta \leq 1\) and \(0 \leq \delta\) and

\[ I_{S=0} = \begin{cases} 1 & \text{if } h = 0, \\ 0 & \text{otherwise.} \end{cases} \quad \text{Eq. 5.6} \]

Note that \(\beta\) is the separable factor (the covariance function is separable if \(\beta = 0\)).

The coefficients of the parametric function \((\nu_S, c, \zeta, a, \alpha, \delta \text{ and } \beta)\) are calculated by applying the weighted least squares (WLS) method and minimizing the difference between the empirical and parametric semivariograms. WLS is applied to the isotropic semivariograms for purely spatial \((u = 0)\), purely temporal \((h = 0)\), and the general cases (using all data) consequently; that is to minimize \(S(\theta_1), S(\theta_2), \text{ and } S(\theta_3)\)

\[
\begin{align*}
S(\theta_1) &= \sum_{l=0}^{L} \sum_{u=0}^{\max_t} w_{l,u}(\gamma(h(l),0; \theta_1) - \hat{\gamma}(h(l),0))^2 \\
S(\theta_2) &= \sum_{u=0}^{\max_t} w_{0,u}(\gamma(h(0),0; \theta_2) - \hat{\gamma}(h(0),0))^2, \\
S(\theta_3) &= \sum_{l=0}^{L} \sum_{u=0}^{\max_t} w_{l,u}(\gamma(h(l),u; \theta_3) - \hat{\gamma}(h(l),u))^2,
\end{align*}
\text{Eq. 5.7}
\]

where \(\hat{\gamma}\) is the obtained empirical semivariogram, \(\theta_1 = (\nu_S, c, \zeta), \theta_2 = (a, \alpha, \delta \text{ and } \beta), \theta_3 = (\beta)\) and \(w_{l,u}\) is the weights function. According to the method proposed by Gribov et al. (2001), two iterations are considered to obtain more accurate results. For the first iteration, the weights function is
\[ w_{l,u} = \frac{N(h(l),u)}{\sum_{k=1}^{L} \sum_{t=0}^{\text{max} t} N(h(k),t)}. \]  
\text{Eq. 5.8}

For the second iteration, the weight function proposed by Cressie (1993) is applied as

\[ w_{l,u} = \frac{N(h(l),u)}{\gamma^2(h(l),u; \theta)} \cdot \frac{1}{\sum_{k=1}^{L} \sum_{t=0}^{\text{max} t} \frac{N(h(k),t)}{\gamma^2(h(k),t; \theta)}}. \]  
\text{Eq. 5.9}

To consider the cloud motion effect, the anisotropic semivariogram should be applied to the problem, where space is always combined with time through the cloud motion vector. Gneiting et al. (2007) considered a Lagrangian covariance function as convex combination with an isotropic term. The proposed anisotropic semivariogram function is

\[ \gamma_{AI}(h,u) = (1 - \rho)\gamma_{FS}(h,u) + \rho \gamma_{LGR}(h,u), \]  
\text{Eq. 5.10}

where \( \gamma_{FS} \) is the isotropic variogram (Eq. 5.5) and \( 0 \leq \rho \leq 1 \).

Schlather (2010) developed a more advanced form of the Lagrangian covariance based on a variable velocity vector with a multivariate normal distribution. According to his model, for constant cloud motion, the motion invariant covariance function is in the form of

\[ C(h,u) = \varphi \left( (|h_1 - vu|^2 + |h_2|^2)^{1/2} \right), \]  
\text{Eq. 5.11}

where \( v \) is the cloud motion speed and \( h_1 \) and \( h_2 \) are the distance components in along and cross-wind directions respectively.

To avoid underperformance of the anisotropic model on days with variable cloud motion as reported by Shinozaki et al. (2014), in this study a non-separable anisotropic
covariance function is proposed which is not reduced to a separable one for the case that the Lagrangian term is negligible. In this approach, the covariance function in Eq. 5.11 is replaced by the purely spatial term, $\varphi(|h|)$, in Eq. 5.3. Therefore, similar to Eq. 5.5, the general parametric anisotropic semivariogram is modeled as

$$
\gamma(h,u) = \sigma^2 \left\{ 1 - (1 - \nu_s) \frac{\exp \left( -c \frac{|h_1 - vu|^2 + |h_2|^2}{(1 + a|u|)^{2\alpha}} \right)}{(1 + a|u|^{2\alpha})^{\delta+\beta}} 
- \nu_s l_{s=0} \frac{\exp \left( -c |vu|^{2\zeta} \right)}{(1 + a|u|^{2\alpha})^{\delta+\beta}} \right\}.
$$

Eq. 5.12

To calculate the coefficients of the general parametric anisotropic semivariogram function, a procedure similar to Eq. 5.7 is performed. However, WLS is applied separately to the anisotropic semivariograms for purely spatial case ($u = 0$), the anisotropic semivariograms along $h_1 = vu$ line ($|h_1 - vu| = 0$), and the general anisotropic semivariograms (using all data).

Note that if the cloud motion is unknown, to apply the anisotropic spatiotemporal Kriging method, it should be estimated during the process. The cloud motion can be estimated by using either cross-spectral analysis (CSA) or cross-correlation method (CCM) as described in Chapter 4. In CSA, the cloud motion speed and direction are estimated by quantifying the maximum delayed correlation between the irradiance data at all the observed locations. As proposed in Section 4.2.1, variations in velocity between all pairs of the sites are calculated and the cloud motion direction is detected by considering the case with the least variation. The cloud motion speed is estimated as the median of the speed of all the all pairs of the sites in the cloud motion direction. In CCM
(Section 4.2.2), on the other hand, the cloud motion is estimated by interpolating the irradiance data on a regular structured grid (using the spatial Kriging method) and comparing correlation between the interpolated data at two time steps. The whole domain is split into boxes (e.g., 5x5 pixels) and the cross correlation finds the position that best matches each given box at the previous time step given the data at the current time step. As suggested in Section 4.2, to ensure reliable cloud motion estimation, quality control (QC) is added to the CSA and CCM analyses including removing conditions with low variability and less correlated sites. According to higher order of accuracy, CCM is applied to estimate unknown cloud motion in this study.

5.1.3 Ordinary Kriging

Ordinary Kriging assumes unknown constant mean over the search neighborhood and is more accurate than simple Kriging which assumes a constant known mean over the whole domain. For spatial Kriging, an estimate at location \( x_0 \), \( Z_0^* \) is obtained as

\[
Z_0^* = \sum_{i=1}^{n} \lambda_i Z_i,
\]

with condition that

\[
\sum_{i=1}^{n} \lambda_i = 1,
\]

where \( n \) is the number of the sites and \( Z_i \) are the observed irradiances. The weights are obtained such as to minimize the residual variance

\[
\min E [(Z_0^* - Z_0)^2],
\]

which leads to
\[
\begin{bmatrix}
0 & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\
\gamma_{21} & 0 & \cdots & \gamma_{2n} & 1 \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
\gamma_{n1} & \gamma_{n2} & \cdots & 0 & 1 \\
1 & 1 & \cdots & 1 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_n \\
\mu \\
\end{bmatrix}
= 
\begin{bmatrix}
\gamma_{10} \\
\gamma_{20} \\
\vdots \\
\gamma_{n0} \\
1 \\
\end{bmatrix},
\]
Eq. 5.16

where \( \mu \) is Lagrange multiplier and \( \gamma_{ij} = \gamma(x_i - x_j) \) is calculated using the parametric semivariogram.

The ordinary Kriging method can be generalized for the spatiotemporal process. In this case, an estimate at location \( x_0 \) and time \( t_0 \), \( Z_0^* \) is obtained as

\[
Z^*(x_0, t_0) = \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij} Z_{ij},
\]
Eq. 5.17

with condition that

\[
\sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij} = 1,
\]
Eq. 5.18

where \( n \) is the number of the sites, \( m \) is the number of time steps and \( Z_0 = Z(x_0, t_j) \) are the observed irradiances. The weights are obtained such as to minimize the residual variance

\[
\min \ E [(Z_{00}^* - Z_{00})^2],
\]
Eq. 5.19

which leads to
where Γ is the spatio-temporal semivariogram matrix and obtained using the parametric spatiotemporal semivariogram as

\[
\begin{pmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn}
\end{pmatrix}^{11} \quad \begin{pmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn}
\end{pmatrix}^{12} \quad \begin{pmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn}
\end{pmatrix}^{1m}
\]

\[
\begin{pmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn}
\end{pmatrix}^{m1} \quad \begin{pmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn}
\end{pmatrix}^{m2} \quad \begin{pmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn}
\end{pmatrix}^{mm}
\]

\[
\begin{pmatrix}
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \cdots & 1
\end{pmatrix}^{11} \quad \begin{pmatrix}
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \cdots & 1
\end{pmatrix}^{12} \quad \begin{pmatrix}
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \cdots & 1
\end{pmatrix}^{1m}
\]

Also, Kriging variance \( \sigma_{\text{OK}}^2 \) which provides an estimation of the error of the ordinary Kriging process is calculated as (Burrough and McDonnell, 1998)

\[
\sigma_{\text{OK}}^2 = \mu + \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij} \gamma_{i0}^{j0}. \quad \text{Eq. 5.22}
\]

### 5.1.4 Spatial & Temporal decorrelation

A promising method is to estimate spatial and temporal decorrelation analytically using the applied parametric covariance (or semivariogram) function. In this method, the
decorrelation lengths can be calculated at desired spatial/ or temporal lags. Moreover, if a Lagrangian covariance function exists in the parametric covariance function, the cloud motion effect can be considered by estimating along-wind decorrelation and decorrelation in other directions (including the cross-wind direction). In general, the spatial decorrelation distance at given time $u^*$ is obtained by finding the point at which the semivariogram function reaches 95% of its final value (semivariogram sill). For example, for the purely spatial case ($u^*=0$), the purely spatial semivariogram function is obtained by setting time lag equals to zero in Eqs. 5.5 or 5.12.

$$
\gamma(h, 0) = \sigma^2 \left[ 1 - (1 - \nu_s) \exp \left(-c|h|^2\zeta\right) - \nu_s I_{S=0} \right], \quad \text{Eq. 5.23}
$$

and the spatial decorrelation distance ($h_c$) is equal to

$$
\gamma(h_c, 0) = 0.95 \gamma(h_\infty, 0) \rightarrow h_c = \left[ \frac{3 + \ln(1 - \nu_s)}{c} \right]^{1/2\zeta}. \quad \text{Eq. 5.24}
$$

In a similar way, the analytical temporal decorrelation can be obtained using either isotropic or anisotropic semivariogram functions. For example, in the general parametric isotropic semivariogram function (Eq. 5.5), the purely temporal case is obtained by setting distance equals to zero ($h^*=0$)

$$
\gamma(0, u) = \sigma^2 \left[ 1 - \frac{1}{(1 + a|u|^{2\alpha})^{\delta+\beta}} \right], \quad \text{Eq. 5.25}
$$

and the temporal decorrelation ($u_c$) is equal to

$$
\gamma(0, u_c) = 0.95 \gamma(0, u_\infty) \rightarrow u_c = \left[ \frac{20^{1/(\delta+\beta)} - 1}{a} \right]^{\left(\frac{1}{2\alpha}\right)}. \quad \text{Eq. 5.26}
$$
5.2 Data & Validation

5.2.1 Simulated data with known atmospheric velocities

Unless the skies are clear, the application of Lagrangian covariance functions is sensitive to accurate estimation of the cloud motion vector which may be problematic for any real dataset that is typically poorly resolved in time and/or space. Therefore, validating the Kriging method with a simulated dataset with known atmospheric velocities can yield better information about the accuracy of different Kriging methodologies, especially for anisotropic spatiotemporal Kriging method with the Lagrangian covariance function. Two spatially and temporally resolved simulated irradiance datasets generated from large eddy simulation (LES) are applied for the spatiotemporal analysis as described in Section 4.1. To achieve temporal stationarity (Perez et al., 2016), the 1D shortwave radiation scheme proposed by Slingo (1989) is applied to convert the LWP to clear sky index ($kt$) as described in Section 4.1.

5.2.1.1 RICO simulation

The RICO simulation is simulated within a spatial domain of 2540m x 2540m (128x128 equally spaced grid points) centered at 18.0°, -61.8°. 1800 sec time series at all points in the domain are chosen. The reference cloud motion is obtained through the velocity at the cloud average height of 1031 m as 2.61 m/s at 180°. Figure 5.1 shows $kt$ contour at selected time steps.

For evaluation of the Kriging method, a subset of 49 sites is selected in the whole domain (black circles in Figure 5.1) and the Kriging method is applied only on these sites. To ensure the selected sites cover the whole domain, the domain is split into 5x5
boxes and one site is selected per box. Also, as the Kriging method performs worse for sites with higher variability, the site with the highest local variability (within the given box) is selected at each box. Therefore, the results reported here can be considered a worst case scenario for the performance of the Kriging method.

Figure 5.1: $kt$ contour of the simulated dataset obtained from RICO simulation at different time steps; (a) 08:23:00, (b) 08:25:00, (c) 08:29:00, and (d) 08:31:00. The black dashed box represents the location of a subdomain which is used for further analysis. The black (red) circles represent the locations of the sites with the highest variabilities which are used for the Kriging methods in the whole domain (subdomain).
To test if the spatial and isotropic Kriging methods yield consistent results for less data and in smaller regions, a subdomain with 60x25 grid points is considered (the black dashed box in Figure 5.1) and Kriging is applied to a subset of 30 sites with the highest variability (red circles in Figure 5.1). Since the cloud motion is almost constant over time in the whole domain, the performance of the anisotropic Kriging method should be consistent as well.

5.2.1.2 CGILS simulation

In the CGILS simulation, 10s LWP is output for 24 hours (8641 time steps) near the California coast (35° N, 125° W) in a 2400 m x 2400 m spatial domain resolved by 96 x 96 grid points in the horizontal and extends to 965 m height in the vertical using 193 grid points. The CGILS case is split into the overcast and breakup sections of 10800 sec and 14400 sec duration (Figures 5.2 & 5.3). The reference cloud motion vector is not constant; the average cloud motion is 5.32 m/s at 279° and 8.58 m/s at 281° during cloud breakup and in overcast periods respectively. For evaluation of the Kriging method, similar to the RICO simulation, 49 sites with the highest local variabilities are selected in the whole domain in each section (black circles in Figures 5.2 & 5.3).
Figure 5.2: Same as Figure 5.1 but for the simulated dataset obtained from CGILS simulation at different time steps during the overcast period; (a) 11:10:00, (b) 11:11:00, (c) 11:12:00, and (d) 11:13:00.
Figure 5.3: Same as Figure 5.1 but for the simulated dataset obtained from CGILS simulation at different time steps during the cloud breakup period; (a) 13:10:00, (b) 13:11:00, (c) 13:12:00, and (d) 13:13:00.

5.2.2 Real irradiance data with unknown cloud motion

5.2.2.1 SMUD irradiance data

A spatially and temporally resolved irradiance dataset is available from real gridded global horizontal irradiance measurements in Sacramento, California.
Sacramento Municipality Utility District (SMUD) deployed a network of 73 solar monitoring devices including 66 GHI sensors and 7 Rotating Shadowband Radiometers (RSR) capable of measuring direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI). The network covers most of SMUD’s 2330 km$^2$ service territory with 5 km spacing. 1 min averages are stored and 5 min and 15 min averaged $kt_{\text{SMUD}}$ are also considered here. After further quality control, 1 min irradiance data of 2012 has been applied for 58 sites with maximum distance of 54.2 km. To quantify the accuracy of the Kriging method, the 27 days with partly cloudy or overcast conditions during 1000 – 1500 PT in 2012 are chosen. The days are chosen according to cloud cover fraction (the fraction of time steps with $kt < 0.85$), as introduced in Section 4.2. The cloud cover fraction is greater than 0.4 during the chosen 27 days. The time-of-day limit avoids errors due to sensor cosine response and shading by nearby obstructions. Due to missing data, the number of sites varies on each selected day with an average of 43 over the 27 days. These days with the highest variabilities are chosen as the worst case scenario for the performance of the Kriging method. Figure 5.4 shows 1 min $kt_{\text{SMUD}}$ maps at selected time steps.
Figure 5.4: $kt$ contour of the 1 min SMUD grid irradiance data on Jan. 24, 2012 at (a) 08:09:00, (b) 08:11:00, (c) 08:13:00, and (d) 08:15:00 PT. The black circles represent the location of the SMUD sites which are applied for the Kriging method (total 43 sites on this day).

5.2.2.2 CSI power output data

15 minute power output of distributed PV systems provided by California Solar Initiative (CSI) rebate program, $P_{CSI}$, in San Diego Gas & Electric (SDG&E) territory in 2010 has been applied. However, as mentioned by Killinger et al. (2017), the PV power output are subject to significant errors and uncertainty which needs a quality control.
After the quality control provided in Section 2.1.2, a final set of 27 PV sites with maximum distance of 58 km are chosen as shown by black circles in Figure 5.5.

![Figure 5.5](image-url)

Figure 5.5: $kt$ contour of the PV power output data in San Diego Gas and Electric (SDG&E) territory on Jan. 22, 2010 at (a) 10:00:00, (b) 10:15:00, (c) 10:30:00, and (d) 10:45:00 PT. The black circles represent the location of the PV sites which are applied for the Kriging method (27 sites on this day).

The power output data is converted to the clear sky index ($kt_{CSI}$) through a complex procedure as described in Section 2.2.2.1. The procedure is self-calibrating by comparing actual power output against maximum power output (representative for clear
conditions) at each time of day during the surrounding days in a given period (e.g. a week or a month). The method to determine the maximum power output \( P_{\text{max,day}} \) corrects for unrealistic cloud enhancement and persistent shading effects and differences in expected clear sky output due to solar geometry. \( k_t \) is estimated by the ratio of \( P_{\text{CSI}}/P_{\text{max,day}} \). To remove the seasonal trend, the power ratio is then corrected by clear sky plane-of-array global irradiance. \( k_t \) is shown in Figure 5.5 as well.

The spatiotemporal Kriging method is applied to the days with the largest ramp rates in the SDG&E territory (10 days in 2010 as determined in Chapter 3). Lohmann et al. (2016) mentioned that mixed sky conditions (rapid succession of clear and overcast conditions) are the most potentially problematic in short term power fluctuation, which makes \( k_t \) interpolation and forecast more complicated. Therefore, these days are chosen as the worst case scenario for the performance of the spatiotemporal Kriging method. Figure 5.5 shows \( k_t \) contours at selected time steps on a day with the largest ramp rate in 2010.

### 5.2.2.3 SAW modeled irradiance data

For comparison, the SAW enhanced resolution satellite-derived irradiation with 30 min temporal and 1 km spatial resolutions is considered in SDG&E territory (see Section 2.1.1). The clear sky index \( (k_{\text{SAW}}) \) is obtained by comparing the SAW modeled GHI and GHI in clear sky conditions (Section 2.3).

### 5.2.3 Validation

The implementation of the spatiotemporal Kriging method is first validated by 1D sinusoidal wave and then using a 1D transect of the data in the along-wind and cross-
wind directions (not shown). Afterwards, the method is applied to the 2D simulated data (RICO and CGILS simulations) as well as the real irradiance and output power data in California. For evaluation of the spatial Kriging method, the cross validation is applied using different methods including the leave one site out (LOSO) and leave one time step out (LOTO) methods. LOSO and LOTO methods are applied using both the entire time series (LOSOE and LOTOE respectively) as well as past data (LOSOP and LOTOFP respectively). The accuracy of the methods is quantified through the normalized root mean squared error between the calculated quantity and true value (nRMSE) and skill ratio (SR) which is defined as 1-\(\text{nRMSE}/\text{nRMSE}_{\text{ref}}\); where \(\text{nRMSE}_{\text{ref}}\) is the persistence model for the LOSOP and LOTOFP (\(kt\) at next time is assumed to be equal to \(kt\) at current time step) while it is the linear interpolation model for LOSOE and LOTOE (\(kt\) at current time is assumed to be equal to the average of \(kt\) at previous and next time steps).

### 5.2.3.1 LOSO method using the entire timeseries (LOSOE)

First, the cross validation is applied using the leave-one-out method (Inoue et al., 2012; Shinozaki et al., 2014): at each site the observed value is removed from the training data and only the observed \(kt\) at all other locations are used in the ordinary Kriging method. The procedure is performed for all time steps at each site. The same procedure is applied using the spatiotemporal Kriging method. So, for a given site \((i=i_{j})\) and time step \((j=j_{1})\), \(kt(i_{j_{1}})\) is estimated by applying \(kt\) at all other locations and all the time steps, \(j=1,...,m\) where \(m\) is the number of time steps. LOSOE is helpful for evaluation of the spatial and spatiotemporal Kriging methods for interpolation at unobserved locations.
5.2.3.2 LOSO method using only past data (LOSOP)

To evaluate forecast capability of the spatiotemporal Kriging method at unobserved locations, for a given time step \((j=j_i)\), only \(k_t\) at all other locations and at all the previous time steps \(j=1,\ldots,j_i\) are considered through the LOSO method.

5.2.3.3 LOTO method using the entire timeseries (LOTOE)

The leave one site out method (LOSO) shows the capability of the spatiotemporal Kriging method for spatial interpolation and temporal downscaling (or forecast), without using information at the given location. However, in practical applications that require interpolation or downscaling, irradiance data are usually available at the given location and at other time steps. So, for further analysis, only the observed value at the given site and time step is removed from the training data (leave one time step out method). So, for a given site \((i=i_i)\) and time step \((j=j_i)\), \(k_t(i_i,j_i)\) is estimated by applying the whole dataset except for \(k_t(i_i,j_i)\). LOTOE illustrates the performance of the spatiotemporal Kriging method for interpolation (and downscaling) at observed location.

5.2.3.4 LOTO method using only past data (LOTOP)

Also, to evaluate forecast capability of the spatiotemporal Kriging method at observed locations and similar to the LOSOP method, only \(k_t\) at all the previous time steps \(j=1,\ldots,j_i\) are considered through the LOTO method.
5.3 Results

5.3.1 Parameter shrinkage

The performance of the spatial and spatiotemporal Kriging methods is evaluated using both the whole dataset and data within the obtained corresponding correlated zone (parameter shrinkage). Quantifying the spatial and temporal correlated zone is helpful for reducing the computational cost, since for a specific point and time only data within the correlated zone are needed in the Kriging model. This is critical when the size of the problem (number of sites and/or time steps) is large. The spatial and temporal decorrelation distances have been estimated analytically using the applied parametric semivariogram function (Eqs. 5.5 & 5.12) at all spatial and temporal lags by finding the point at which the semivariogram function reaches the semivariogram sill.

Figure 5.6 shows the correlated zone for the isotropic case as well as the anisotropic case in the cross-wind direction (setting $h_l=0$ in the general anisotropic parametric semivariogram function, Eq. 5.12) using the RICO simulated data. These are the cases where the correlated zones are fully symmetric. The rhombus correlated zones for the cases that the semivariogram functions are assumed separable are shown as well (dashed lines).
Figure 5.6: The correlated zone calculated from the RICO simulated data for the isotropic case and cross-direction in the anisotropic case for non-separable (solid) and separable (dashed) cases. The caption indicates the purely spatial and temporal decorrelation lengths.

For the anisotropic analysis, the correlated zone changes dramatically according to the cloud motion effect as depicted in Figure 5.7 using the RICO simulated data.
Figure 5.7: The correlated zone calculated from the RICO simulated data for the different directions in the anisotropic case. The legend represents the angle, $a$, with the cloud motion direction shown by the yellow dashed line.

Figure 5.8 shows the correlated zone in the along-wind direction (the case with $a = 0^\circ$ in Figure 5.7) for the RICO simulated data.

Figure 5.8: The correlated zone calculated from the RICO simulated data for the along-wind direction ($a = 0^\circ$ in Figure 5.7) in the anisotropic case. The dashed line, $h = 2.42 \ t$, represents the cloud motion direction where the average cloud motion speed equals to 2.42 m/s in this case.
5.3.2 Simulated data

The spatial and spatiotemporal Kriging methods are applied to the RICO and CGILS simulated data. First, as a benchmark, the spatial Kriging method is performed by setting the time lag \( u = 0 \) (purely spatial case) in the general semivariogram functions (Eqs. 5.5 or 5.12). Then, the process is assumed to be stationary and isotropic and spatiotemporal Kriging is performed by using the general isotropic semivariogram function (Eq. 5.5). To consider the cloud motion effect, the spatiotemporal anisotropic Kriging is performed by using the general anisotropic semivariogram function (Eq. 5.12) and applying the given averaged cloud motion vector. For evaluating the performance of the Kriging methods, all the cross validation methods are applied using both the whole dataset and only data within the correlated zone (parameter shrinkage). nRMSE and skill ratio (SR) for the validation methods are calculated. The Kriging variance \( (\sigma_{OK}^2) \) which provides an estimation of the error of the ordinary Kriging process (Eq. 5.22) is provided as well.

5.3.2.1 RICO simulation

Figure 5.9 compares results of the \( kt \) data obtained by SP Kriging as well as IST and AST Kriging methods at one site in RICO simulation.
Figure 5.9: Application of the purely spatial and spatiotemporal ordinary Kriging to model \(kt\) at one site in the whole domain for the RICO simulated dataset using (a) leave-one-site-out (LOSO) (b) leave-one-time-step-out (LOTO) evaluation methods. The true \(kt\) is shown as well. The caption indicates the nRMSE, skill ratio (SR), and Kriging variance (\(\sigma_{OK}^2\)) of the spatial and spatiotemporal models for the site that is shown.

The LOSOE, LOSOP, LOTOE, and LOTOP validation methods are applied to all sites in the domain. Table 5.1 shows statistics for the whole domain and the subdomain as well. The results of the RICO simulation confirm that the spatial and spatiotemporal Kriging methods are consistent in the whole domain and the subdomain. However, in comparison with the whole domain, the results are more accurate in the subdomains, especially for LOSO methods. Also, the anisotropic spatiotemporal Kriging method (with averaged nRMSE, skill ratio, and \(\sigma_{OK}^2\) equal to 7.96%, 0.61, and 0.19 respectively) are more accurate than the isotropic spatiotemporal Kriging (11.00%, 0.46, and 0.35) and spatial Kriging (14.98%, 0.27, and 0.45 respectively) methods. The Kriging variance (\(\sigma_{OK}^2\)), as calculated in Eq. 5.22, provides an estimation of the error of the process and it trends against nRMSE and SR.
Table 5.1: Evaluation statistics for the spatial (SP), isotropic spatiotemporal (IST), and anisotropic spatiotemporal (AST) Kriging methods for the 49 (30) sites selected in the whole domain (subdomain) of the RICO simulation. Different validation methods are applied including leave one site out (LOSO) method using the entire time series (LOSOE) and past data (LOSOP) as well as leave one time step out (LOTO) method using the entire time series (LOTOE) and past data (LOTOP). Each value represents the average over all sites.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Validation Method</th>
<th>Kriging Method</th>
<th>Entire Data</th>
<th>Parameter Shrinkage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>nRMSE (%)</td>
<td>SR (-)</td>
</tr>
<tr>
<td>Whole</td>
<td>LOSOE</td>
<td>SP</td>
<td>15.59</td>
<td>0.15</td>
</tr>
<tr>
<td>Domain</td>
<td></td>
<td>IST</td>
<td>15.28</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>11.17</td>
<td>0.40</td>
</tr>
<tr>
<td>LOSOP</td>
<td></td>
<td>IST</td>
<td>15.34</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>13.36</td>
<td>0.27</td>
</tr>
<tr>
<td>LOTOE</td>
<td></td>
<td>IST</td>
<td>4.95</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>2.84</td>
<td>0.84</td>
</tr>
<tr>
<td>LOTOP</td>
<td></td>
<td>IST</td>
<td>8.67</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>6.28</td>
<td>0.65</td>
</tr>
<tr>
<td>Sub-domain</td>
<td>LOSOE</td>
<td>SP</td>
<td>14.37</td>
<td>0.39</td>
</tr>
<tr>
<td>Domain</td>
<td></td>
<td>IST</td>
<td>14.21</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>8.16</td>
<td>0.66</td>
</tr>
<tr>
<td>LOSOP</td>
<td></td>
<td>IST</td>
<td>14.50</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>10.00</td>
<td>0.58</td>
</tr>
<tr>
<td>LOTOE</td>
<td></td>
<td>IST</td>
<td>5.80</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>4.72</td>
<td>0.80</td>
</tr>
<tr>
<td>LOTOP</td>
<td></td>
<td>IST</td>
<td>9.30</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>7.14</td>
<td>0.70</td>
</tr>
</tbody>
</table>

The parameter shrinkage (considering only data within the correlated zone) causes a small improvement (1.11%) for the spatiotemporal methods while the accuracy is slightly reduced for the spatial cases (0.33%). However, parameter shrinkage is of great interest as it reduces the computational costs drastically, even with the same order of accuracy. The parameter shrinkage, on average, reduced the number of applied sites and time steps in the calculations by 83% and 91% respectively. Srinivasan et al. (2010) mentioned \(O(n^4)\) computational complexity of the spatial Kriging method for interpolation at \(O(n)\) points, where \(n\) is the number of the applied sites (for
spatiotemporal Kriging $n$ is equivalent to the total number of the applied sites multiplied by total time steps). Therefore, on average, the parameter shrinkage reduced the computational cost of spatial and spatiotemporal Kriging by 99.61% and 99.99% respectively.

5.3.2.2 CGILS simulation

Figures 5.10 & 5.11 compare results of the $kt$ data obtained by spatial Kriging as well as isotropic and anisotropic spatiotemporal Kriging methods at one site during the overcast and cloud breakup periods, respectively.

Figure 5.10: Same as Figure 5.9 but for $kt$ estimation at one site in the domain for the CGILS simulated dataset during overcast period; (a) LOSO and (B) LOTO evaluation methods.
Figure 5.11: Same as Figure 5.9 but for \( k_t \) estimation at one site in the domain for the CGILS simulated dataset during the cloud breakup period; (a) LOSO and (B) LOTO evaluation methods.

Table 5.2 shows statistics for the CGILS simulated \( k_t \) during the cloud breakup and overcast periods. The errors are higher during the cloud breakup period (where the variability is relatively high as confirmed in Section 4.3.2) while in the overcast period (with lower relative variability) the rRMSE errors are only 1% to 7%. On average, the results of CGILS simulation using the anisotropic spatiotemporal Kriging method (with averaged nRMSE, skill ratio, and \( \sigma_{OK}^2 \) equal to 6.64%, 0.63, and 0.26 respectively) are more accurate than the isotropic spatiotemporal Kriging (10.35%, 0.37, and 0.39 respectively) and spatial Kriging (14.21%, 0.15, and 0.45 respectively) methods. Similar to the RICO simulation results, there is a large improvement by applying the anisotropic method in the cloud breakup (with 30.48% and 50.32% improvement in averaged RMSE over the isotropic spatiotemporal and spatial Kriging methods respectively) and overcast (with 51.51% and 62.51% improvement in averaged RMSE over the isotropic...
spatiotemporal and spatial Kriging methods respectively) periods. Also, similar to the RICO simulation case, the parameter shrinkage has a minor effect on the Kriging accuracy with 0.06% improvement for the spatiotemporal methods and 5.56% worsening for the spatial cases.

Table 5.2: Same as Table 5.1 but for 49 sites in the whole domain of the CGILS simulation during the cloud breakup and overcast periods.

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Validation Method</th>
<th>Kriging Method</th>
<th>Entire Data nRMSE (%)</th>
<th>SR (-)</th>
<th>( \sigma_{O^2}^E ) (-)</th>
<th>Parameter Shrinkage nRMS E (%)</th>
<th>SR (-)</th>
<th>( \sigma_{O^2}^E ) (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800 - 1200 h</td>
<td>LOSOE</td>
<td>SP</td>
<td>6.83</td>
<td>0.22</td>
<td>0.71</td>
<td>8.25</td>
<td>0.06</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IST</td>
<td>6.41</td>
<td>0.27</td>
<td>0.71</td>
<td>6.40</td>
<td>0.27</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>2.49</td>
<td>0.71</td>
<td>0.33</td>
<td>2.49</td>
<td>0.71</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>LOSOP</td>
<td>IST</td>
<td>6.49</td>
<td>0.26</td>
<td>0.71</td>
<td>6.49</td>
<td>0.26</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>3.18</td>
<td>0.63</td>
<td>0.40</td>
<td>3.18</td>
<td>0.63</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>LOTOE</td>
<td>IST</td>
<td>3.41</td>
<td>0.61</td>
<td>0.29</td>
<td>3.41</td>
<td>0.61</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>1.92</td>
<td>0.78</td>
<td>0.26</td>
<td>1.92</td>
<td>0.78</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>LOTOP</td>
<td>IST</td>
<td>4.81</td>
<td>0.45</td>
<td>0.39</td>
<td>4.81</td>
<td>0.45</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>2.64</td>
<td>0.70</td>
<td>0.33</td>
<td>2.64</td>
<td>0.70</td>
<td>0.33</td>
</tr>
<tr>
<td>1200 - 1500 h</td>
<td>LOSOE</td>
<td>SP</td>
<td>21.58</td>
<td>0.08</td>
<td>0.49</td>
<td>21.74</td>
<td>0.07</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IST</td>
<td>21.21</td>
<td>0.10</td>
<td>0.49</td>
<td>21.18</td>
<td>0.10</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>12.58</td>
<td>0.46</td>
<td>0.21</td>
<td>12.58</td>
<td>0.46</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>LOSOP</td>
<td>IST</td>
<td>21.32</td>
<td>0.09</td>
<td>0.49</td>
<td>21.27</td>
<td>0.10</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>13.91</td>
<td>0.40</td>
<td>0.25</td>
<td>13.90</td>
<td>0.40</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>LOTOE</td>
<td>IST</td>
<td>7.21</td>
<td>0.69</td>
<td>0.03</td>
<td>7.21</td>
<td>0.69</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>6.89</td>
<td>0.70</td>
<td>0.13</td>
<td>6.89</td>
<td>0.70</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>LOTOP</td>
<td>IST</td>
<td>11.94</td>
<td>0.49</td>
<td>0.06</td>
<td>11.94</td>
<td>0.49</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>9.50</td>
<td>0.59</td>
<td>0.17</td>
<td>9.50</td>
<td>0.59</td>
<td>0.17</td>
</tr>
</tbody>
</table>

5.3.3 Real data

The spatial and spatiotemporal Kriging methods are applied on real irradiance and output power data in California where the cloud motion is unknown. To apply the AST Kriging method the cloud motion vector is required, which is estimated using CCM. For each day under investigation, the average of the velocity field over the period under
investigation is applied. To obtain data on a structured grid, which is required for CCM, \( k_{t_{\text{SMUD}}} \) and \( k_{t_{\text{CSI}}} \) are interpolated on a regular grid using the spatial Kriging method and the CCM is performed on the interpolated (structured) data.

The averaged estimated cross-wind (along-wind) spatial decorrelation lengths equal to 8.36 (20.72) and 17.16 (39.41) km for the SMUD and CSI data respectively while the averaged temporal decorrelation lengths equal to 23 and 120 min for the SMUD and CSI data respectively. For reducing the size of the computations and since it has been observed that the performance of the methods are comparable, only the data within the obtained correlated zones are presented for each method (parameter shrinkage). All the cross validation methods are applied and the performances of the Kriging methods are quantified by nRMSE and skill ratio (SR).

5.3.3.1 SMUD grid network

The spatial and spatiotemporal Kriging methods are applied on the 1, 5, and 15 min \( k_{t_{\text{SMUD}}} \) on the selected 27 days in 2012 with partly cloudy or overcast conditions. Figure 5.12 compares results of \( k_{t_{\text{SMUD}}} \) obtained by SP, IST, and AST Kriging methods at one site and one day.
Figure 5.12: Same as Figure 5.9 but for 1 min $k_{SMUD}$ estimation at one site in the SMUD grid network on Jan. 24, 2012; (a) LOSO and (b) LOTO evaluation methods

Table 5.3 shows nRMSE and skill ratio between calculated quantities and true values on Jan. 24, 2012 as well as the averaged statistics over all the days with partly cloudy or overcast conditions (27 days in 2012). For 1 min averaged data, the spatial Kriging method (with averaged nRMSE and skill ratio equal to 18.36% and -0.07 respectively) underperforms with respect to the persistence method. However, the performance of the spatial Kriging method increased for 5 and 15 min averaged data (with averaged nRMSE and skill ratio equal to 12.12% and 0.69 respectively). On the other hand, the spatiotemporal Kriging method outperform with respect to the both spatial Kriging and persistence methods in all cases while the anisotropic method, with averaged nRMSE (skill ratio) equal to 11.11% (0.43), 11.99% (0.68), and 8.99% (0.82) for 1, 5, and 15 min data respectively, is always more accurate than the isotropic one, with averaged nRMSE (skill ratio) equal to 13.57% (0.27), 12.84% (0.65), and 9.39% (0.81) for 1, 5, and 15 min data respectively.
Table 5.3: Same as Table 5.1 but for the 1, 5, and 15 min $kt_{SMUD}$ using 58 SMUD sites on Jan. 24, 2012 as well as averaged over all the days with partly cloudy or overcast conditions (27 days in 2012). The AST Kriging method is applied by using the averaged estimated cloud velocity.

<table>
<thead>
<tr>
<th>Day</th>
<th>Validation Method</th>
<th>Kriging Method</th>
<th>1 min Data nRMSE (%)</th>
<th>SR (-)</th>
<th>5 min Data nRMSE (%)</th>
<th>SR (-)</th>
<th>15 min Data nRMSE (%)</th>
<th>SR (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 24, 2012</td>
<td>LOSOE</td>
<td>SP</td>
<td>16.31</td>
<td>0.12</td>
<td>10.23</td>
<td>0.68</td>
<td>7.70</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IST</td>
<td>14.84</td>
<td>0.22</td>
<td>9.47</td>
<td>0.71</td>
<td>7.05</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>10.67</td>
<td>0.45</td>
<td>9.30</td>
<td>0.72</td>
<td>6.34</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>LOSOP</td>
<td>IST</td>
<td>15.03</td>
<td>0.21</td>
<td>9.32</td>
<td>0.72</td>
<td>7.35</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>12.55</td>
<td>0.35</td>
<td>9.30</td>
<td>0.72</td>
<td>7.05</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>LOTOE</td>
<td>IST</td>
<td>7.35</td>
<td>0.62</td>
<td>9.21</td>
<td>0.72</td>
<td>6.33</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>6.44</td>
<td>0.67</td>
<td>8.74</td>
<td>0.74</td>
<td>5.80</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>LOTOP</td>
<td>IST</td>
<td>13.17</td>
<td>0.50</td>
<td>9.42</td>
<td>0.72</td>
<td>6.93</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>8.76</td>
<td>0.54</td>
<td>9.27</td>
<td>0.72</td>
<td>6.36</td>
<td>0.83</td>
</tr>
<tr>
<td>All cloudy days in 2012</td>
<td>LOSOE</td>
<td>SP</td>
<td>18.36</td>
<td>-0.07</td>
<td>14.53</td>
<td>0.59</td>
<td>9.71</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IST</td>
<td>17.08</td>
<td>0.02</td>
<td>13.47</td>
<td>0.63</td>
<td>9.48</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>13.49</td>
<td>0.28</td>
<td>11.68</td>
<td>0.69</td>
<td>8.89</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>LOSOP</td>
<td>IST</td>
<td>17.33</td>
<td>0.01</td>
<td>13.89</td>
<td>0.61</td>
<td>9.69</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>15.20</td>
<td>0.14</td>
<td>12.85</td>
<td>0.65</td>
<td>9.34</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>LOTOE</td>
<td>IST</td>
<td>8.97</td>
<td>0.62</td>
<td>11.40</td>
<td>0.70</td>
<td>8.78</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>6.50</td>
<td>0.74</td>
<td>10.96</td>
<td>0.72</td>
<td>8.63</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>LOTOP</td>
<td>IST</td>
<td>10.91</td>
<td>0.46</td>
<td>12.61</td>
<td>0.66</td>
<td>9.62</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>9.27</td>
<td>0.57</td>
<td>12.47</td>
<td>0.67</td>
<td>9.10</td>
<td>0.81</td>
</tr>
</tbody>
</table>

5.3.3.2 SDG&E territory

The spatial and spatiotemporal Kriging methods are applied on the $kt_{CSI}$ data obtained from CSI PV power output data on the 10 days with the largest ramp rates in SDG&E territory in 2010. Figure 5.13 compares results of the $kt_{CSI}$ obtained by spatial Kriging as well as isotropic and anisotropic spatiotemporal Kriging methods at one site on Jan. 22, 2010 when the largest ramp rates occurred. Table 5.4 shows nRMSE and skill ratio on Jan. 22, 2010 as well as the averaged statistics over all 10 days.
Figure 5.13: Same as Figure 5.9 but for 15 min $kt_{\text{CSI}}$ estimation at one PV site in SDG&E territory on Jan. 22, 2010; (a) LOSO and (B) LOTO evaluation methods.

Table 5.4: Same as Table 5.1 but for the 15 min $kt_{\text{CSI}}$ (using 45 CSI sites) and 30 min $kt_{\text{SAW}}$ (using 1 km resolution data) in SDG&E territory on Jan. 22, 2010 as well as averaged over 10 days in which the largest ramp rates occurred in SDG&E territory in 2010. The AST Kriging method is applied by using the averaged estimated cloud velocity.

<table>
<thead>
<tr>
<th>Day with the largest ramp rates in 2010</th>
<th>Validation Method</th>
<th>Kriging Method</th>
<th>CSI Data</th>
<th>SAW Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>nRMSE</td>
<td>SR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(%)</td>
<td>(-)</td>
</tr>
<tr>
<td>Jan. 22, 2010</td>
<td>LOSOE</td>
<td>SP</td>
<td>12.39</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IST</td>
<td>11.57</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>9.51</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>LOSOP</td>
<td>IST</td>
<td>12.21</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>10.13</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>LOTOE</td>
<td>IST</td>
<td>9.62</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>7.78</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>LOTOP</td>
<td>IST</td>
<td>11.01</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>8.71</td>
<td>0.69</td>
</tr>
<tr>
<td>10 days with the largest ramp rates in 2010</td>
<td>LOSOE</td>
<td>SP</td>
<td>14.86</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IST</td>
<td>14.06</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>11.87</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>LOSOP</td>
<td>IST</td>
<td>14.34</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>13.04</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>LOTOE</td>
<td>IST</td>
<td>11.16</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>8.93</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>LOTOP</td>
<td>IST</td>
<td>12.55</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AST</td>
<td>10.08</td>
<td>0.75</td>
</tr>
</tbody>
</table>
The satellite-derived $kt_{SAW}$ is applied to show the capability of the spatiotemporal Kriging method in estimation of irradiation at an arbitrary point without nearby ground measured stations. For comparison, Figure 5.14 shows results of the $kt_{SAW}$ obtained by spatial Kriging as well as isotropic and anisotropic spatiotemporal Kriging methods at the same site as in Figure 5.13 on Jan. 22, 2010. Table 5.4 also provides the evaluation statistics for the case where the spatial and spatiotemporal Kriging methods are applied using $kt_{SAW}$. For all the evaluations, the statistics are provided in comparison with actual modeled $kt_{SAW}$ at the given sites.

Figure 5.14: Same as Figure 5.9 but for 30 min $kt_{SAW}$ estimation at one PV site in SDG&E territory on Jan. 22, 2010; (a) LOSO and (B) LOTO evaluation methods.

The results of CSI and SAW analysis in SDG&E territory also confirm that the anisotropic spatiotemporal Kriging method, with averaged nRMSE (skill ratio) equal to 10.98% (0.73) and 9.81% (0.72) for $kt_{CSI}$ and $kt_{SAW}$ respectively, are more accurate than the isotropic spatiotemporal Kriging (with averaged nRMSE (skill ratio) equal to 13.02% (0.67) and 10.11% (0.71) for $kt_{CSI}$ and $kt_{SAW}$ respectively) and spatial Kriging (with
averaged nRMSE (skill ratio) equal to 14.86% (0.63) and 10.57 (0.70) for \( k_{t_{\text{CSI}}} \) and \( k_{t_{\text{SAW}}} \) respectively) methods.

5.4 Conclusions

The spatiotemporal ordinary Kriging method was applied to estimate irradiation at an arbitrary point and time. The empirical semivariogram was obtained using observed data. The main issue in the current anisotropic approach as proposed by Inoue et al. (2012) and Shinozaki et al. (2014) was the underperformance of the method in some situations related to the fact that the AST covariance function was reduced to a separable covariance function for the cases that the Lagrangian term in the covariance function was negligible so that the main direction was not detectable and, therefore, the anisotropic covariance function was practically reduced to an isotropic one. Therefore, to fit the empirical semivariogram, a general stationary non-separable anisotropic parametric semivariogram function was proposed using the general non-separable function proposed by Gneiting (2002) and the Lagrangian function suggested by Schlather (2010). The coefficients of the parametric function were calculated by WLS.

The spatial and spatiotemporal Kriging methods were validated with two spatially and temporally resolved simulated irradiance datasets generated from LES with known atmospheric velocities; including RICO as well as CGILS simulation during cloud breakup and overcast periods. To consider the cloud motion effect, an AST Kriging was performed using the given averaged cloud motion vector. Results of the RICO simulation confirm that the spatial and spatiotemporal Kriging methods are consistent in the whole domain and the subdomain. For the CGILS simulation, the performance of the methods
(quantified by nRMSE and SR) was lower during the cloud breakup period where the variability was relatively high, while in the overcast period (with lower relative variability) the results were almost in perfect agreement.

Then, the spatial and spatiotemporal Kriging methods have been applied on real irradiance and output power data in California. To apply the anisotropic spatiotemporal Kriging method, the cloud motion is required which is estimated using cross-correlation method (CCM). For SMUD data at high temporal resolution (1 min), the spatial Kriging method shows improvement over the persistence model only for 5 and 15 min averaged data. Comparable CSI and SAW results in coastal Southern California demonstrates the capability of the spatiotemporal Kriging method in estimation of irradiation at an arbitrary point without nearby ground measured stations.

The results confirmed that the anisotropic spatiotemporal Kriging method, with averaged nRMSE (skill ratio) over all datasets equal to 7.92% (0.66), are always more accurate than isotropic spatiotemporal Kriging (11.33% (0.54)) and spatial Kriging (13.93% (0.40)). In general, application of the Lagrangian covariance function in the anisotropic spatiotemporal Kriging method is sensitive to accurate estimation of the cloud motion. Therefore, obtaining large skill ratios for SMUD, CSI, and SAW data, where cloud motion is unknown, confirms the accuracy and reliability of the method even in these more challenging real applications.

The results for the cases where parameter shrinkage (considering data within the obtained corresponding correlated zone) was applied show a small improvement for the spatiotemporal methods while the accuracy is slightly reduced for the spatial cases. However, even with the same order of accuracy, parameter shrinkage is of great interest
as it reduces the computational costs drastically (99.86% on average) since only data within the correlated zone are needed in the Kriging method.

**Acknowledgements**

Chapter 5, in full, has been previously published as Jamaly, M., and Kleissl, J., 2017b, “Spatiotemporal Interpolation and Forecast of Irradiance Data Using Kriging”, Solar Energy. The dissertation author was the primary investigator and author of this paper.
Chapter 6 Conclusions & Future Work

In this study, the applied irradiance and PV power output datasets in California are first analyzed, quality controlled, and validated (Sections 2.1 & 2.2). A performance model has been developed to convert irradiation data to power output as well (Section 2.3). Then, to illustrate the need for accurate spatiotemporal analysis for solar resource assessment and forecast, the available irradiance and power output datasets are applied to detect the largest aggregate ramp rates in California in 2010 (Chapter 3). Over one year the largest hourly aggregate ramp in California was a 30% increase based on the Performance Test Conditions (PTC) rating but ramps over 23% of PTC occurred only about once per day.

In Chapter 4, the cloud motion has been estimated by CSA and CCM using two spatially and temporally resolved simulated irradiance datasets generated from large eddy simulation (LES). In general, CCM results are accurate for all the different cloud cover fractions with average relative mean bias error (rMBE) of cloud speed and absolute error of cloud direction equal to 3% and 3°, respectively. CSA estimates the cloud speed and direction by cross-spectral analysis of \( k_t \) data at all grid points in the domain. For low cloud cover fractions, CSA estimates the cloud motion speed and direction with average rMBE and absolute error equal to 10% and 11°, respectively. For the cases where the cloud motion varies rapidly in time, CSA is unable to detect the cloud motion pattern for the whole time series while the results are more accurate in shorter time intervals (in which the cloud motion can be assumed to be spatially homogeneous and steady). However, splitting the whole time series into shorter time interval reduces errors to 15% and 16° respectively. The maximum lengths of the subintervals are observed to be 120
and 30 min for CGILS simulation during the cloud breakup and overcast periods respectively.

In Chapter 5, the spatiotemporal Kriging method has been performed using RICO and CGILS simulated irradiance datasets as well as real irradiance and output power data in California (with unknown atmospheric velocities). For the real data in California (Sacramento and San Diego areas), the cloud motion had to be estimated during the process using Cross-Correlation Method (CCM) and the days in which the highest ramp rates occurred (which are detected through the ramp rate analysis in Chapter 3) is applied as the worst case scenario for performance of the spatiotemporal Kriging method. Also, the Lagrangian covariance function is added to consider the anisotropic behavior related to the cloud motion effect. Results confirm that the anisotropic model is most accurate with an average normalized root mean squared error (nRMSE) of 7.92% representing a 66% improvement over the persistence model.

In general, application of the Lagrangian covariance function in the anisotropic spatiotemporal Kriging method is sensitive to accurate estimation of the cloud motion. Therefore, obtaining large skill ratios for SMUD, CSI, and SAW data, where cloud motion is unknown, confirms the accuracy and reliability of the method even in these more challenging real applications. Moreover, comparable CSI and SAW results in coastal Southern California demonstrates the capability of the spatiotemporal Kriging method in estimation of irradiation at an arbitrary point without nearby ground measured stations. Also, the spatial and temporal decorrelation had been estimated analytically using the applied parametric covariance function. The results for the cases where parameter shrinkage (considering data within the obtained corresponding correlated zone)
are applied show a small improvement for the spatiotemporal methods while the accuracy is slightly reduced for the spatial cases. However, even with the same order of accuracy, parameter shrinkage is of great interest as it reduces the computational costs drastically (99.86% on average).

After all, this dissertation shows the ability of robust cloud motion estimation using observed irradiance data at given locations as well as application of the spatiotemporal Kriging method for interpolation and forecast of irradiance data. However, there are some suggestions for improving the proposed methods as follow.

I. CCM results are calculated up to a horizon = 50 sec, which was empirically found to be sufficient to obtain the best results. However, in general, the best forecast horizon is mainly a function of cloud speed and distance between the sites. In real situations with unknown cloud speed, the maximum horizon can be adjusted or even detected automatically according to the specifications of the problem. In specific the variation ratio and cloud cover fraction, where the former depends on how frequently the clouds are passing while the latter represents the fraction of the area shaded by clouds in the domain, affect the optimum horizon. By analyzing these variables for different datasets and in different situations, a general relationship between the maximum horizon and these variables may be obtained.

II. As mentioned before, for high cloud cover fractions and unsteady cloud speed, CSA results are not reliable for the whole time series. In these cases, splitting the whole time series into shorter time interval are required to obtain acceptable
results. For generalization of CSA application, it would be helpful to detect the
maximum length of time series or allowable variation in cloud motion for which
acceptable results are achieved. This can be performed by considering some
parameters including the domain size as well as the spatial and temporal
resolutions of the \(kt\) data. Therefore, for a given dataset, the maximum allowable
length of time series can be quantified by estimating a threshold for cloud cover
fraction and variation ratio of \(kt\) time series. Datasets with different resolutions in
different weather conditions are required to obtain a general relation between
these parameters.

III. Note that spatiotemporal Kriging method can be applied either on \(kt\) data or its
ramp rate. Lonij et al. (2013) showed that the ramp rate covariance function is the
second partial derivative in time of the covariance function of the \(kt\) data. So, this
derivative can be derived from the parametric covariance function. It would be
interesting to perform the spatiotemporal Kriging on both \(kt\) and ramp rates data
for validation and/or parameter identification as well as more accurate estimation
of the temporal decorrelation length.

IV. As mentioned, the spatiotemporal Kriging method can be applied to both
ground measured as well as satellite derived solar irradiance data (as applied in
Chapter 5). For spatial interpolation, it is mostly advantageous for ground data
where available data are typically unstructured and sparse. However, when it
comes to temporal downscaling or forecasting, the method would be promising
for both satellite and ground datasets or their combination (as in the so-called co-
Kriging method; Perez et al., 2016). Gutierrez-Corea et al. (2014) concluded that when the simulation domain has adequate ground station density (as for SDG&E territory in Section 5.3.3.2), merging ground measured and satellite derived GHI data will improve the accuracy. Therefore, an interesting study would be to merge the SAW and CSI data in Section 5.3.3.2 and applying co-Kriging method.

V. In this study, the spatiotemporal Kriging method is performed on the physical space and the Lagrangian covariance function is applied to account for anisotropy occurred by cloud motion. Yang et al. (2013), on the other hand, achieved spatial stationarity and isotropy through deformations of the geographical space based on the two-step method developed by Sampson and Guttorp (1992). The two–step method estimates such covariance structures using a non–parametric approach by introducing spatial dispersion (which is a dissimilarity measure). The first step is called multidimensional scaling (MDS) and the second step is thin plate spline (TPS) mapping. It would be helpful to apply the two–step method and compare the results with the proposed anisotropic spatiotemporal Kriging method.

Applying Kriging method on the transformed dispersion space has the advantage that the process is no longer dependent on the cloud motion field which improves the methods robustness. However, in this approach, the cloud velocity direction and its effect on the problem cannot be quantified. Also, the space transform may be complicated and even impossible if there is large number of stations and/or an ill-conditioned problem.
References


