Journal of Spatial Science, 2015
http://dx.doi.org/10.1080/14498596.2015.1051147

Taylor & Francis
Taylor & Francis Group

Optimal solar radiation sensor network design using spatial and geostatistical analyses

Ramesh S.V. Teegavarapu[a*], Chandra S. Pathak[b], John R. Mecikalski[c] and Jayanthi Srikishen[d]

[a]*Department of Civil Engineering, Florida Atlantic University,Boca Raton, FL, USA;* [b]*Hydrology, Hydraulics and Coastal Community of Practice, United States Army Corps of Engineers, Washington, DC, USA;* [c]*Atmospheric Sciences Department, University of Alabama in Huntsville, 320 Sparkmen Drive, Huntsville, Alabama, USA;* [d]*Universities Space Research Association, Huntsville, Alabama, 320 Sparkman Drive, Huntsville, AL, USA*

A methodology for optimal ground-based sensor network design for an evapotranspiration (ET) estimation method which uses solar radiation as the only parameter is developed and evaluated in this study. The methodology employs geospatial analyses and a geostatistical approach, and data from ground-based sensors and satellite-based estimates of solar insolation (i.e. total amount of solar radiation energy received on a given surface area during a given time) considering the spatial variability of the data. The applicability of the methodology is demonstrated by using Geostationary Operational Environmental Satellite (GOES)-estimated and 29 ground sensor-based observed solar insolation data in the South Florida region of the USA. Results indicate that the optimal design of network depends on the spatial variability of insolation, analysis block size defined based on region-specific radiation characteristics, and the standard error used as a metric of network estimation accuracy.

Keywords: solar radiation; optimal sensor network; evapotranspiration; geospatial analysis; geostatistics; standard error

1. Introduction

Evapotranspiration (ET) is one of the major components of the hydrological cycle. Estimation of this hydro-meteorological parameter is critical for hydrological modelling and management of surface and groundwater resources. The most common meteorological parameters that are essential for ET estimation are air temperature, humidity, solar radiation, barometric pressure and wind speed. Potential ET (PET) and reference ET (RET) are the two ET-related parameters that are needed to estimate ET, which is the rate that water loss to the atmosphere occurs from well-watered soil and plant surfaces. Methods available for ET estimation are generally classified as: (1) temperature-based methods (Blaney–Criddle method; Doorenbos & Pruitt 1977; Abtew & Melesse 2013; Hargreaves–Samani method, 1985); (2) radiation-based methods (Abtew method, 1996); (3) the mass transfer method (Abtew & Melesse 2013); (4) energy balance methods and their variants (such as Penman 1948, corrected Penman, and Penman–Monteith; Monteith 1973). ET is a multi-dimensional process occurring in space and time and

*Corresponding author. Email: rteegava@fau.edu

solar radiation (insolation) flux is the largest determinant of temporal variation in ET flux. An important parameter for estimation of ET (PET and RET) is net radiation (R_{net}). One component of R_{net} is incoming (downwelling) solar radiation, which can be estimated as solar insolation from a geostationary satellite data or ground-based sensors referred to as pyranometers. Figure 1 illustrates solar insolation in the context of a simple energy balance model.

The main focus of this study is the development of a methodology for design of an optimal solar radiation sensor network that can help estimation of ET in South Florida, USA, using a conceptually simple solar radiation-based ET estimation method (Abtew 1996). This method uses solar radiation as the only input variable and can be used for estimation of daily wetland evapotranspiration or shallow open water evaporation or potential evapotranspiration. Abtew and Melesse (2013) point out that in many regions of the world where solar radiation explains most of the variation in evaporation and evapotranspiration, a simple method such as the one suggested by Abtew (1996) is adequate for estimation of ET. Several studies conducted all over the world (Abtew & Melesse 2013; Abtew, Irizarry-Ortiz, Lyon, & Obeysekera, 2002; Enku *et al.* 2011; Zhai *et al.* 2009; Delclaux & Coudrain 2005; Oudin *et al.* 2005; Shoemaker & Sumner 2006; Xu & Singh 2000; Jacobs *et al.* 2002) have confirmed the application and benefits of this method over other methods. Even though ET is influenced by several meteorological parameters, the availability of a conceptually simple method (Abtew 1996) to accurately estimate the same using solar radiation as the sole parameter in the study region forms the motivation for development of an optimal solar radiation sensor-based monitoring network.

The most desirable ET datasets for the purposes of distributed hydrological modelling and water budget and resource analysis are spatially continuous gridded data, rather than point values derived from the traditional field weather station networks. Estimation of ET using satellite-based solar radiation with the help of methods (e.g. Abtew 1996) that primarily use this parameter can provide such datasets. A number of methods currently exist in the literature for estimating solar insolation using geostationary operational environmental satellite (GOES) visible channel observations. The methods ranged from statistical-empirical relationships (Tarpley 1979) to physical models of varying complexity (Gautier *et al.* 1984;

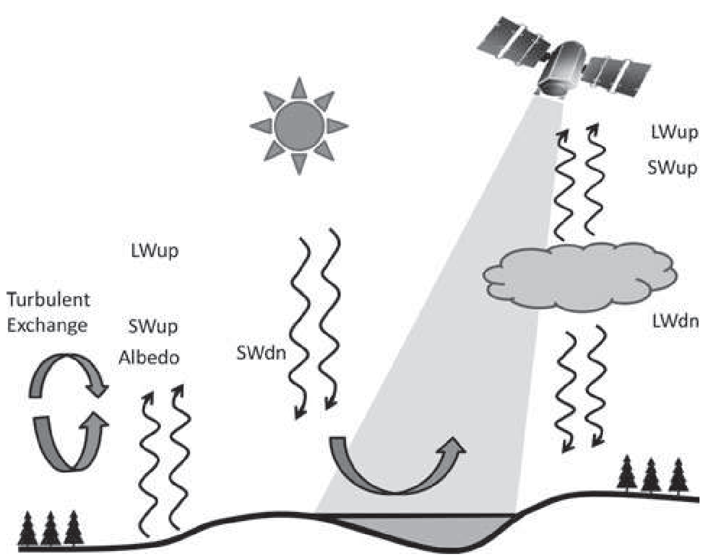

Figure 1. Schematic of simplified energy balance model. Here, ‘up’ and ‘dn’ relate to the upward and downward components, respectively. ‘SW’ is shortwave radiation, and ‘LW’ is longwave radiation.

Moser & Raschke 1984; Pinker & Ewing 1985; Dedieu *et al.* 1987; Darnell *et al.* 1988; Frouin & Chertock 1992; Pinker & Laszlo 1992; Weymouth & LeMarshall 1999; Paech *et al.* 2009). Schmetz (1989) and Pinker *et al.* (1995) confirm the utility of satellite-estimated solar insolation methods in producing accurate estimates of insolation for different sky conditions.

Development of a methodology to design an optimal solar radiation sensor-based monitoring network requires assessment of the available ground sensor- and/or satellite-based solar radiation data. In the first phase of this study spatial variability solar insolation based on GOES-estimated datasets at a specific spatial resolution using geospatial analysis is evaluated. This task is expected to help in development of spatial evapotranspiration estimates in the next phase. The initial phase will also help understand the variability in land-surface features and land-use across the region of interest that contribute to regional variations in solar insolation due to the heterogeneity in heating (from both sensible and latent contributions) that in turn cause (a) meso-scale circulations and (b) local enhancements in convective cloud formation and maintenance. It is expected that some sub-regions may experience spatially uniform insolation due to prevalence of land-surface features such as large lakes and non-varying types of land-use and ecosystem-related vegetation. The results from the initial phase of spatial analysis are used in a geostatistical framework to determine an optimal sensor network to measure and characterise the variable solar radiation across the region. A properly sized and optimised network can result in significantly increased efficiency in measurement of spatially varying solar insolation.

2. Monitoring network design

Design of optimal monitoring networks requires an evaluation of heterogeneity of the solar insolation based on observations at sampling locations in a region. Fortin and Dale (2005) indicate that any optimal spatial sampling scheme requires a careful balance between sampling locations that are too close to one another, thus not providing enough new information (data highly auto-correlated), and sampling locations that are too sparse, so that processes at other spatial scales introduce too much variability (Haining 1990). Similarities exist between the designs of solar radiation sensor and rain gauge networks as they measure a spatially varying meteorological parameter. Existing or variants of approaches available for rain gauge network design can be adopted for solar radiation network design. Several approaches exist in the context of rain gauge monitoring network design, and they rely on conceptually simple methods based on variance of rainfall in space. The variance of rainfall is calculated based on the existing number of rain gauges in the region. Rakhecha and Singh (2009) provide a review of several methods of rain gauge network design, by Rycroft (1949), Ganguli *et al.* (1951) and Ahuja (1960). Rycroft (1949) used variance in space and allowable variance in estimate of mean rainfall to determine the optimum number of rain gauges, while Ganguli *et al.* (1951) used the mean monthly coefficient of variation and a pre-defined tolerance value and the existing gauges to obtain the optimum number of rain gauges. A review of recent literature related to optimal design of monitoring networks points to several different currently available methods for a variety of applications and they include an information-theoretic approach for transportation applications (Xing *et al.* 2013) and wireless sensor networks (Larish & Riley 2011), an entropy and multi-objective-based approach (Mogheir *et al.* 2013) and fuzzy theory and multiple criteria analysis (Chang & Lin 2014) for water quality monitoring. In the current study a geostatistical approach is used for the design of a solar radiation monitoring network.

Methodology

The methodology adopted in the current study is shown as a series of tasks to be executed in two parallel tracks, as shown in Figure 2.

As indicated in Figure 2, initially the solar insolation data from the GOES satellite are validated using pyranometer data (i.e. ground-based sensors). Stratification of the region into an array of analysis blocks of specific spatial resolution is essential to determine an optimal network of sensors. Methods for partitioning of spatial data such as k-means clustering (Han & Kamber 2006; Teegavarapu 2014) and variance quad-tree (Minasny, McBratney, & Walvoort, 2007; Fortin & Dale 2005) can be used for determination of analysis block size. However,

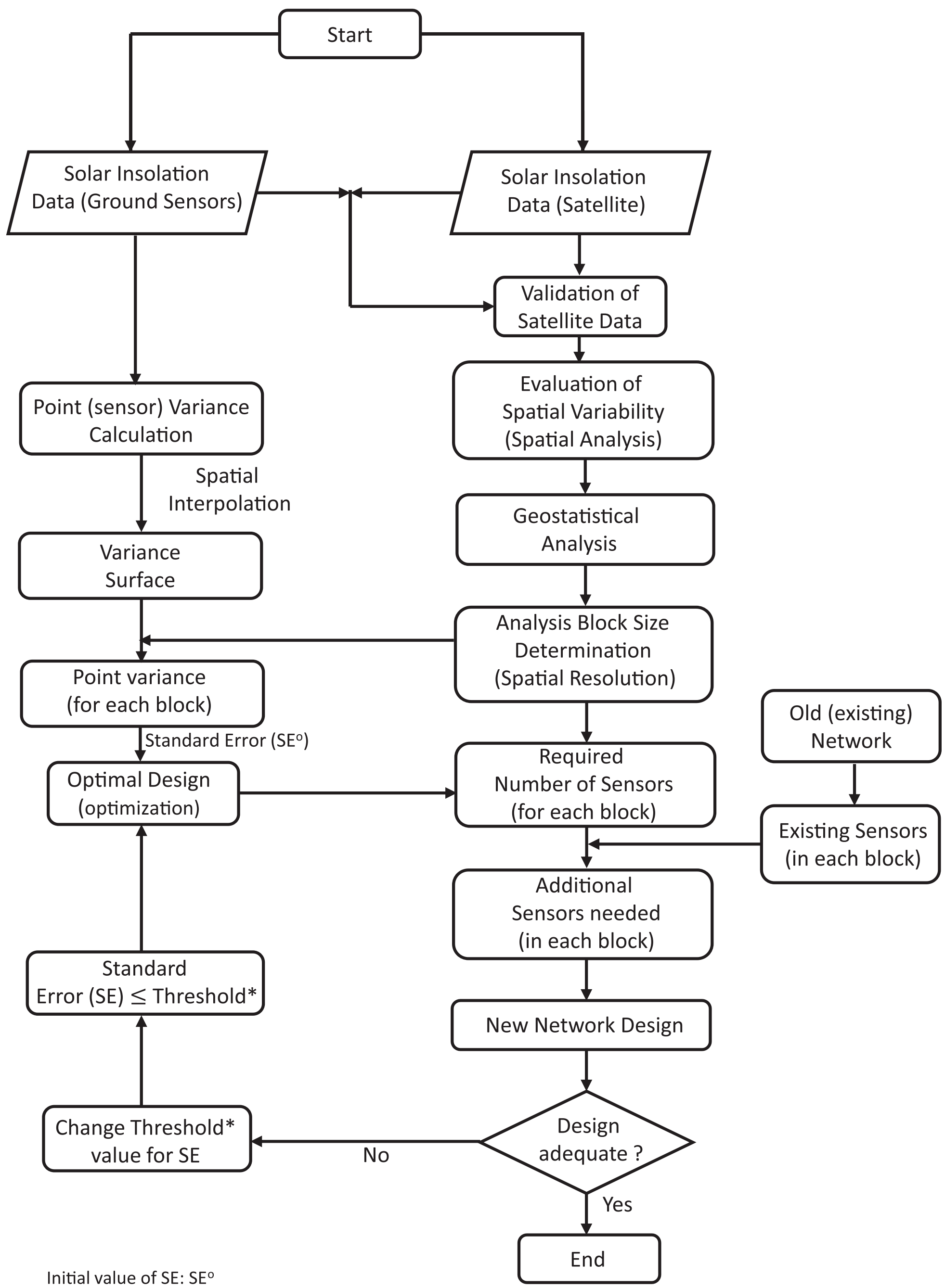

Figure 2. Schematic of steps utilised in optimal design of a solar insolation network.

the size of the block is determined using the range parameter of an empirical semivariogram fitted to the solar radiation data in the current study. Once the spatial partitioning task is completed, GOES insolation data are analysed for spatial variability in the region and geostatistical analysis using ordinary kriging is completed to determine the appropriate size of the analysis block. In a series of parallel tasks, point variance is estimated from the ground sensor data measurements and an interpolated surface of this variance is created covering the entire region using any spatial interpolation method. Point variance is calculated using temporal observations of insolation from a single point thereby assessing the seasonal variation. Multiple spatial interpolation methods are then evaluated using different performance measures before the best method is selected for creation of an authentic surface of point variance of insolation. For a selected analysis block size, the optimal number of sensors needed is determined with the help of any unconstrained nonlinear optimisation method using information about point variance for that particular block and a pre-specified standard error. In each analysis block, after accounting for already existing sensors, the number of additional sensors is determined. A few steps shown in Figure 2 are executed in an iterative way until an implementable network design is achieved considering sensor placement and monetary constraints. Even though the methodology developed in this study is mainly aimed at enhancing the existing monitoring network, it can be applied to regions with no existing network of sensors.

3. Geospatial analysis

This section briefly describes geospatial data processing tasks carried out in this study using several tools available under ArcGIS including the geostatistical toolbox. Initially detrending of insolation data sets was performed for purposes of identification of spatial correlations through a (semi) variogram analysis approach. Detrending specifically will remove the *means* of the data, and any trend in insolation not yet addressed in the calibration work of Paech *et al.* (2009) to better isolate the true variability within the data. A ‘parametric’ form of detrending in each analysis block is used to remove trends, which is later explained in Section 9.

4. Geostatistical analysis

The degree of spatial dependence is generally expressed as a semivariogram in geostatistical analysis (O’Sullivan & Unwin 2010; Teegavarapu 2007). A general expression (O’Sullivan & Unwin 2010) used to estimate the semivariogram is given by Equation 1

$$\gamma(d) = \frac{1}{2n(d)} \sum_{dij=d} (\theta_i - \theta_j)^2 \tag{1}$$

where $\gamma(d)$ is the semivariance which is defined over observations θ_i and θ_j, which are lagged successively by distance d, and $n(d)$ is the number of pairs of points at separation distance d. Depending on the shape of the semivariogram cloud several mathematical models are possible, including linear, spherical, circular, exponential and Gaussian. Three semivariogram models, namely spherical, exponential and Gaussian, defined by Equations, 2, 3 and 4 are evaluated in this study.

$$\gamma(h)_1 = C_o + C_1\left[\frac{1.5h}{R} - 0.5\left(\frac{h}{R}\right)^3\right] \tag{2}$$

$$\gamma(h)_2 = C_o + C_1\left[1 - \exp\left(-\frac{3h}{R}\right)\right] \tag{3}$$

$$\gamma(h)_3 = C_o + C_1\left[1 - \exp\left(-\frac{3h^2}{R^2}\right)\right] \tag{4}$$

The parameters C_o, h and R are referred to as the nugget (or nugget effect), distance and range parameters of a semivariogram respectively. The value at zero separation distance is referred to as the nugget. The summation of C_o and C_1 is referred to as the sill and the semivariance at range R is equal to the sill value. The range parameter (R) defines the distance at which no spatial correlation exists or semivariance is constant. This parameter can be used to define the analysis block size as recommended by Pathak and Vieux (2007). A nonlinear least squares fitting method available through the '*nls*' function in the proprietary statistical package S-plus (S-Plus 2007) is used to obtain the values of C_o and C_1 (SFWMD 2008). Preliminary analyses (SFWMD 2008) of insolation data indicated that the exponential semivariogram model is the best among the three models evaluated to characterise the solar insolation in the current study region.

5. Analysis of point variance solar insolation from ground sensors

Spatial interpolation using point variance values based on observations at ground-based sensor locations is adopted to spatially characterise the solar insolation over a region. A surface (continuous field) from the point variances of solar insolation values is generated. Two different interpolation methods (one deterministic and another stochastic) are investigated: (1) the inverse distance weighting method (IDWM) and (2) ordinary kriging (O'Sullivan & Unwin 2010; Teegavarapu 2007). The IDWM (Teegavarapu & Chandramouli 2005) for spatial interpolation uses distance as weight for a weighted estimate at a point in space. The estimate of an observation, θ_m, at a point in space, using the observed values at other sensors, is given by Equation 5:

$$\theta_m = \frac{\sum_{i=1}^{n} \theta_i d_{mi}^{-k}}{\sum_{i=1}^{n} d_{mi}^{-k}}, \tag{5}$$

where again θ_m is the estimate of the observation at a point in space m; n is the number of sensors; θ_i is the observation at sensor i, d_{mi} is the distance from the location of sensor i to the observation point m; and k is referred to as friction distance that ranges from 1 to 6. A value of 2 is chosen in this study for the friction distance (i.e. k), which is the most commonly adopted value (O'Sullivan & Unwin 2010). The number of nearest neighbours used for interpolation is another parameter that is critical for the success of the method. The number of neighbours is selected by a trial and error process and the number that results in the lowest root mean squared error (RMSE) based on observed and estimated values of variance is selected.

Ordinary kriging (Isaaks & Srivastava 1989; Webster & Oliver 2001; Teegavarapu 2007) is a stochastic interpolation method based on scalar measurements at different points in space. Surface interpolation using kriging depends on the selected semivariogram model and the same fitted with a mathematical function or model. Depending on the shape of the empirical semivariogram, kriging weights (Webster & Oliver 2001) are derived and the estimate at any point in space is obtained by the weighted sum of observations at all other locations. The estimation of a value θ_{m} at a location is given by Equation 6.

$$\theta_m = \sum_{i=1}^{n} \delta_i \theta_i \tag{6}$$

The variable θ_i is the value of observation at location i and δ_i is a weight associated with the observation.

6. Sensor network design

The design of the sensor network is aimed at capturing and characterising the spatial variability of the solar radiation across the region. The design of the network depends on several factors, including (1) placement of sensors to maximise the information obtained from the sensors, (2) the existing network of sensors and (3) the monetary cost involved in the purchase and placement of the sensors. The standard error (SE) of the mean is used as a metric of accuracy of the monitoring network, to identify the optimal number of ground sensors in this study. Multiple studies (e.g. Olea 1984; Spruill & Candela 1990; Bhat *et al.* 2015) in the past have used the *SE* as an accuracy metric for monitoring network design. Olea (1984) indicated that the average standard error and the maximum standard error of estimation over the sampling domain can be used as global indices of sampling efficiency. The required number of sensors in a specific analysis block can be obtained by improving the accuracy of the mean solar insolation measurement. An exponential semivariogram model is found to be appropriate based on preliminary analysis of insolation data conducted (SFWMD 2008) using three different semivariogram models discussed in the Section 4. The semivariogram model expressed in terms of the range, R, distance parameter, h, and point variance, σ_o^2, is given by Equation 7.

$$\gamma(h) = \sigma_o^2\left(1 - e^{\frac{-3h}{R}}\right) \tag{7}$$

The variable R is the range in the same units as the distance, h. A relationship is derived between correlation represented by the semivariogram and the same correlation represented by the correlogram. The correlation coefficient, ρ_h, between observations at two insolation measurement locations at a separation distance, h, is defined as (Equations 8–11),

$$\rho_h = \frac{\sigma_h}{\sigma_o^2} \tag{8}$$

$$\gamma_h = \sigma_o^2 - \sigma_h \tag{9}$$

$$\gamma_h = \sigma_o^2(1 - \rho_h) \tag{10}$$

$$\rho_h = e^{\left(\frac{-3h}{R}\right)} \tag{11}$$

An exponential semivariogram, used to model the covariance structure of the solar insolation from satellite data, provides the range parameter that indicates the distance at which no spatial correlation exists. The effective number (or minimum number) of sensors is obtained from the correlation distance from satellite data and point variance information from ground-based sensor data for each analysis block. Haan (2002) and Matalas and Langbein (1962) indicate that information contained in data from n monitoring stations in a region having an average inter-station correlation of ρ_h (i.e. correlation coefficient among n stations) is equivalent to the information contained in n' uncorrelated stations (sensors) in the region. The relationship between n' and n can be established and is given by Equation 12 (Haan 2002).

$$n' = \frac{n}{1 + \rho_h(n-1)} \tag{12}$$

Equation 12 suggests that as n increases, n' approaches $1/\rho_h$. Hence, the effective number of sensors n' can be related to separation distance h. The *SE* of the mean is calculated for all the analysis blocks using Equation 13.

$$SE = \sqrt{\frac{\sigma^2}{n'}} \tag{13}$$

A scaled or normalised *SE* referred to as *SE*' is obtained by Equation 14. The *SE* is calculated based on effective number of sensors (n').

$$SE' = \frac{\sqrt{\frac{\sigma^2}{n'}}}{SE_{n'=1}} \tag{14}$$

7. Optimal network design

The optimal network design is based on the *SE* and the number of sensors in each analysis block. The network size is optimal when an optimisation formulation is solved using constraints related to block size and cost associated with installation and removal of sensors.

Variable density analysis block approach

In the variable density analysis block approach, the point variance (σ^2) for each block is used to identify the number of sensors (n) that have a correlation coefficient (ρ) to achieve a desired constant level of accuracy defined by the magnitude of SE. The relationship linking σ^2, n, ρ and SE is given by Equation 15.

$$n_k(SE)^2 = (1 + \rho_k(n_k - 1))\sigma_k^2 \quad \forall k \tag{15}$$

The variable k identifies each analysis block. Equation 15 is solved for n_k for different values of SE for each analysis block by using a nonlinear root-finding method (Brent 1973). The value obtained for n is rounded off to the nearest integer to obtain the number of sensors. The approach provides the number of sensors needed within each block. Further study using land-cover data and other information would be needed to assess exact placements of sensors within a given block.

8. Case study area description and data

The region of interest in South Florida for the current study is shown in Figure 3[!–insert–]; its boundary is defined and managed by a state agency, the South Florida Water Management District (SFWMD). The state water agency SFWMD is referred to as 'District' and is responsible for the collection, validation and archiving of the hydrological and meteorological data (such as barometric pressure, solar radiation, air temperature, relative humidity and wind speed) at stations that form the District's meteorological monitoring network. The District's broad objective is to optimise the regional hydro-meteorological monitoring network to cater to the needs of different water resources and environmental management projects in the region. Locations of existing solar radiation ground-based sensors in the SFWMD region are shown in Figure 3 and the details of these sensors are provided in Table 1.

The solar insolation data used in this study are derived from NOAA (National Oceanographic and Atmospheric Administration) GOES observations that cover the state of Florida, as described by Paech *et al.* (2009). GOES data are obtained from the GOES data archive at the Space Science and Engineering Center at the University of Wisconsin-Madison. For the Paech *et al.* (2009) study, which produced an extended GOES insolation data record from 1995–2004, over 102,000 individual GOES images were processed using the model of Gautier *et al.* (1980) to produce half-hourly and daily-integrated solar insolation throughout the state of Florida at a 2 km spatial resolution. In 2005 and 2006, the University of Alabama in Huntsville (UAH), the University of New Hampshire (UNH), all water management districts (WMDs) in the state of Florida, and the United States Geological Survey (USGS) took part in the creation of a decade-long (1995–2004) statewide daily ET datasets at 2×2 km resolution (SFWMD 2008). Key inputs into this ET estimation methodology included GOES satellite estimated incoming solar insolation (Gautier *et al.* 1980; Diak *et al.* 1996; Otkin *et al.* 2005), as well as ancillary weather (e.g. wind, temperature, humidity) and land-surface information. This ongoing project provides a critical database for estimation of ET statewide and this effort will continue towards the creation of a multi-decadal dataset. Work also included calibrating the GOES solar insolation data (Paech *et al.* 2008) with ground sensors. The efforts of Paech *et al.* (2008) found that calibration of the GOES insolation reduced errors to 1.7 MJ m^{-2} day^{-1} (10 percent), and also removed temporal-, seasonal- and satellite sensor-related biases. Also, coefficient of determination (R^2) values based on satellite and ground (pyranometer)-based values reached values closer to 0.90 following further calibration activities to remove month-to-month and cloudiness-related error biases.

Solar radiation amounts vary geographically within central and South Florida.

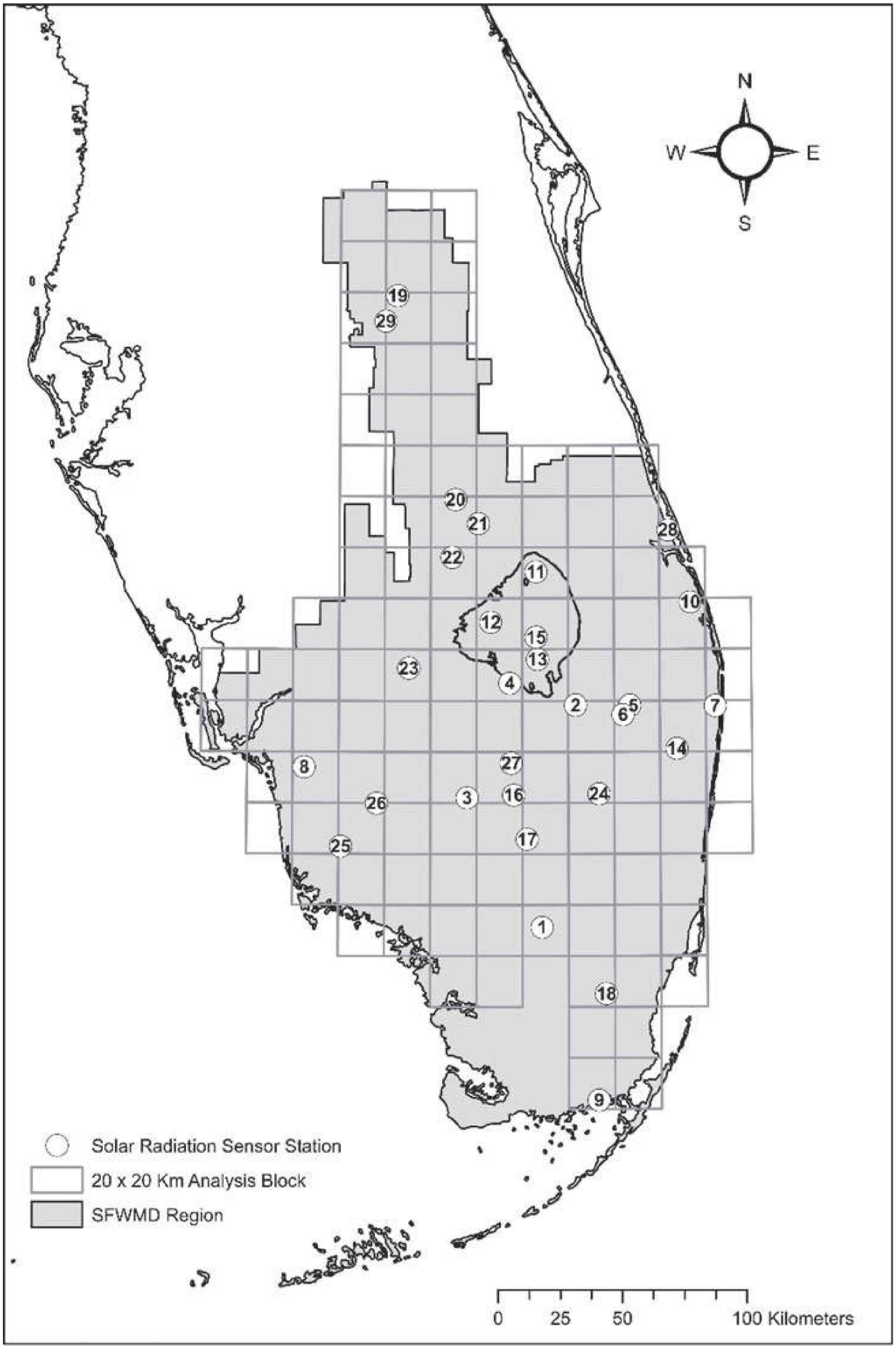

Figure 3. Location of 29 solar radiation sensor stations in the SFWMD region with overlay of 20×20 km grid. Each 20×20 km region represents an analysis 'block' as described in the text.

Radiation characteristics and patterns on land surrounding Lake Okeechobee and ocean are different from those of central overland mass. In addition, spatial variation in radiation amounts for shorter durations, such as one day, is significantly greater than for monthly, seasonal and annual radiation. Therefore, in this study the spatially varying sensor density is considered based on varying solar radiation conditions during dry and wet periods. In addition, solar radiation estimates of local patterns are used for determining optimum sensor placement as opposed to laying out the sensor in an evenly spaced grid. The GOES satellite-based daily solar radiation data from 1995 to 2004 available from University of Alabama in Huntsville (UAH) are used in this study. Daily solar radiation data from 29 ground-based LI-COR model LI-200S pyranometers (LI-COR 2015; Pathak 2008) installed in the SFWMD region are used for the study. Each of these sensors is calibrated against an Eppley precision pyranometer under daylight conditions (Abtew & Melesse 2013; LI-COR 2015). The typical error noted in measurements under these conditions is ± 5 percent (Kinsman, Kite, & Mtundu, 1994; Pathak 2008; Abtew & Melesse 2013). A sensitivity of 0.2 kilowatts per meters squared per millivolt (kW $m^{-2}mV^{-1}$) is noted for these sensors (Pathak 2008).

9. Results and analysis

Solar insolation data are processed from the 2×2 km resolution data grids and overlaid onto 20×20 km analysis blocks covering the region defined as District (refer to Figure 3). A total of 119 non-intersecting blocks cover the District region. Each 20×20 km analysis block contains approximately 100 insolation pixels at 2-km resolution (Figure 3). The insolation data are divided by seasons across South Florida. The climatology of South Florida demands that the data set be subdivided into cool, warm and transitional seasons. After initial analysis (SFWMD 2008), the datasets are split into the following time periods: (a) dry season: November–March, (b) wet season: April–October, (c) a set of completely (100 percent cloud cover) cloudy days, and (d) a set of completely clear days. Transitional time periods are not considered to add significance to these analyses. For sets (c) and (d), considerable effort is needed to identify days with these characteristics, given the relative rarity of clear and cloudy days across the entire District, especially during the wet season. Using 2×2 km gridded insolation data, surfaces of mean, standard deviation, and coefficients of variation are developed. To characterise the covariance structure of insolation across the region, definition of analysis blocks large enough to encompass the range of existing spatial correlation but sufficiently small to capture climatological gradients near coastal areas and inland water features such as Lake Okeechobee is essential. Characterisation of these gradients is very important given the driving forces for cumulus cloud convection across South Florida. In particular, time-dependent changes in cumulus clouds due to land, lake and sea breeze circulations are considered. Variations of mean and standard deviation of insolation for the wet season are shown in Figure 4 and Figure 5 respectively. These values are derived from measures of temporal variability at each 2 km grid, based on daily values. Figures 4 and 5 also shows layers of 119 non-intersecting 20×20 km analysis blocks over the SFWMD region.

It can be seen from Figure 4 that the mean values of insolation decrease slightly from north to south, reaching maximum values in the east (and over approximately Lake Okeechobee). Values are mainly range from 18.6 to 19.6 MJ m^{-2} day^{-1} (one MJ, or megajoule, is 10^6 joules). in the north and west, yet locally decrease to near 18 MJ m^{-2} day^{-1} in the far southeast. These values are close to 8 MJ m^{-2} day^{-1} and are below the maximum values that would be expected under clear-sky conditions, yet this is not unexpected

Table 1. Summary statistics for ground sensors and satellite-based datasets during wet season.

		Sensor-based data			Satellite-based data			
Sensor station*	Sensor station name	Average MJ/day/m^2	Variance $(MJ/day/m^2)^2$	Standard deviation MJ/day/m^2	Average MJ/day/m^2	Variance $(MJ/day/m^2)^2$	Standard deviation MJ/day/m^2	Correlation coefficient
1	3AS3WX	18.908	28.335	5.323	18.837	26.269	5.125	0.940
2	BELLEGL	18.813	33.134	5.756	18.639	29.413	5.423	0.916
3	BIGCYSIR	18.252	28.793	5.366	18.790	29.683	5.448	0.931
4	CFSW	19.429	34.017	5.832	19.001	31.480	5.611	0.925
5	ENR105	19.075	37.418	6.117	18.889	38.331	6.191	0.939
6	ENR308	18.729	30.984	5.566	18.658	31.248	5.590	0.939
7	FHHSWX	19.672	40.325	6.350	19.280	35.881	5.990	0.957
8	FPWX	18.097	31.016	5.569	18.635	27.586	5.252	0.914
9	JBTS	20.253	42.524	6.521	19.776	35.105	5.925	0.923
10	JDWX	18.424	32.614	5.711	18.885	31.707	5.631	0.947
11	L001	18.116	40.167	6.338	19.095	35.115	5.926	0.928
12	L005	18.747	41.209	6.419	19.628	36.205	6.017	0.916
13	L006	19.119	37.680	6.138	19.920	36.386	6.032	0.942
14	LOXWS	18.613	33.181	5.760	18.382	32.374	5.690	0.909
15	LZ40	17.990	34.160	5.845	19.599	36.570	6.047	0.931
16	ROTNWX	19.043	30.272	5.502	18.637	27.302	5.225	0.936
17	S140W	16.745	27.211	5.216	18.155	29.424	5.424	0.900
18	S331W	17.924	28.891	5.375	18.601	31.055	5.573	0.928
19	S61W	18.403	35.803	5.984	18.457	30.869	5.556	0.915
20	S65CW	19.345	29.920	5.470	18.959	30.964	5.565	0.935
21	S65DWX	19.880	27.999	5.291	18.906	26.179	5.117	0.945
22	S75WX	18.839	27.100	5.206	18.765	24.558	4.956	0.948
23	S78W	18.768	30.005	5.478	18.819	29.484	5.430	0.906
24	S7WX	17.465	23.228	4.820	18.592	27.084	5.204	0.933
25	SGGEWX	17.899	22.523	4.746	18.316	20.211	4.496	0.947
26	SILVER	20.033	27.007	5.197	18.788	24.871	4.987	0.956
27	STA5WX	18.374	25.355	5.035	18.404	20.246	4.500	0.935
28	SVWX	19.125	31.965	5.654	19.193	31.814	5.640	0.944
29	WRWX	19.538	29.144	5.399	19.115	28.979	5.383	0.944

*Numbers match with those shown in Figure 3.

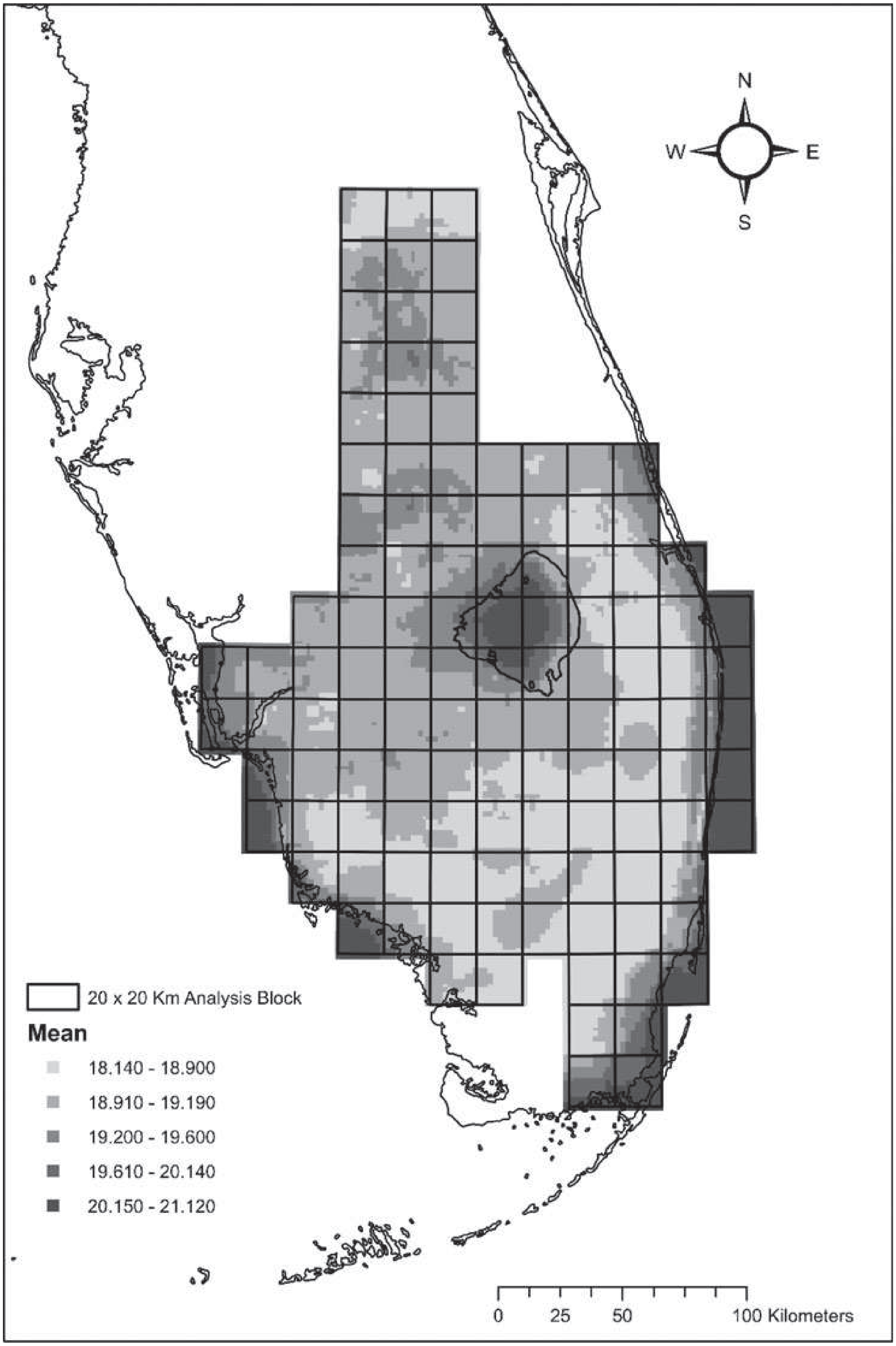

Figure 4. Mean solar insolation map based on 2-km grids with an overlay of 20-km analysis blocks for the wet season (April–October) based on data from years 1995–2004 in MJ m^{-2} day^{-1}.

given the high frequency of occurrence of convective and other clouds across this portion of Florida during the wet season. Values over the ocean and Lake Okeechobee are much higher than or equal to 20 MJ m^{-2} day^{-1}. The standard deviation map of insolation (Figure 5) shows several features: First, values are lowest over land, ~5.2–5.8 MJ m^{-2} day^{-1}. Second, higher values up to near 6 MJ m^{-2} day^{-1} are found near the coasts, which could be caused by a high variability in convective clouds (varying largely from day to day, specifically either mostly clear or mostly cloudy), and is especially exemplified near lakes and the ocean. Third, the analysis of coefficient of variance (as shown in Figure 6) indicated gradients in insolation along the east coast, with values exceeding 0.3 in these areas (SFWMD 2008). Lastly, the variability in the entire wet season of the daily solar insolation dataset and the histogram for all 10,186 days (April–October, 1995–2004) are determined. Here, the data are binned about the mean (19.13 MJ m^{-2} day^{-1}). The highest frequency of days with a given value occur at ~1.0 MJ m^{-2} day^{-1} above and below 19.13 MJ m^{-2} day^{-1}, which suggests a delineation between cloudier and clearer days, and land and water regions, across the region. From a meso-scale weather perspective, this would denote days with either a low or high coverage of convective storms, which is not atypical of tropical (or subtropical) weather regimes (Byers & Rodebush 1948; Riehl 1954; Heymsfield et al. 1996).

Range parameter assessment: relationships to land-surface features

The covariance structure of the solar insolation data evaluated indicates that there is no uniform analysis block size that is identifiable for the region under consideration. Therefore, two analysis size blocks, 40 × 40 km and 20 × 20 km, are identified for developing the optimum design of sensor network. A 'parametric' form of detrending (Lloyd 2007) in each analysis block is intended to remove non-stationarity. This type of detrending by use of regression models is recommended in several reference texts (e.g. Llyod 2007; Haan 2002; Isaaks & Srivastava 1989; Webster & Oliver 2001). Non-stationarity of mean insolation, referred to here as a trend, can affect the successful identification of the covariance structure. Without detrending, correlation lengths of the insolation magnitudes can become spuriously large (Wilks 2006; Gringarten & Deutsch 2001). Prior to detrending, days (a) in which large tropical storms and hurricanes affect the state of Florida, (b) of consistent cloudiness given the known deficiencies in the GOES insolation data in these cases, and (c) with missing GOES data and low data quality indices (Paech *et al.* 2009) are omitted from analysis. Through least squares, detrending is accomplished within each analysis block by fitting a linear regression equation to the insolation data in each time period. The detrended data are then used to compute empirical semivariograms using the 'Geostatistical wizard' within ArcGIS, as well as for the covariance analysis.

The semivariogram analysis is carried out for each analysis block using the daily insolation data, over subsets of the 10-year long data set (e.g. wet and dry seasons, cloudy days). The empirical semivariograms are fitted with data from the wet season in each year and in each analysis block; the wet season possesses the largest fraction of incoming insolation that can be highly variable due to small-scale convective cloudiness (of the order of 1–4 km). As there is a relatively weak correlation in space for insolation, especially when small-scale (2–25 km) convective clouds dominate the cloud climatology (and hence the insolation variability), the mean empirical semivariogram ordinates are selected. Subsequently, an exponential semivariogram model is fitted to empirical semivariograms of mean detrended pixel values in each block. The 'upper quartile' pixel values are used, but did not provide

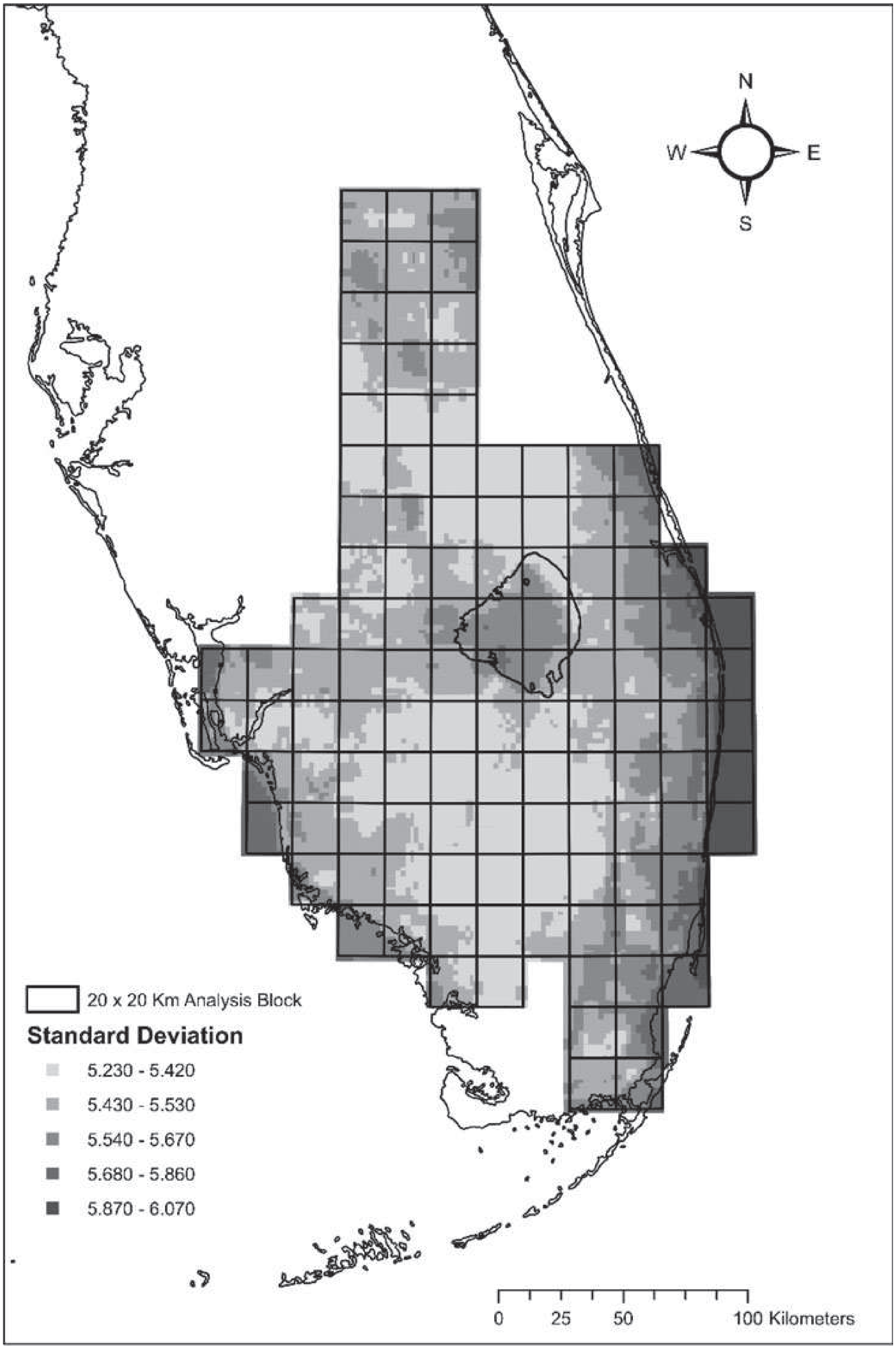

Figure 5. Map of standard deviation of solar insolation map based on 2-km grids with an overlay of 20-km analysis blocks for the wet season (April–October) based on data from years 1995–2004 in MJ m^{-2} day^{-1}.

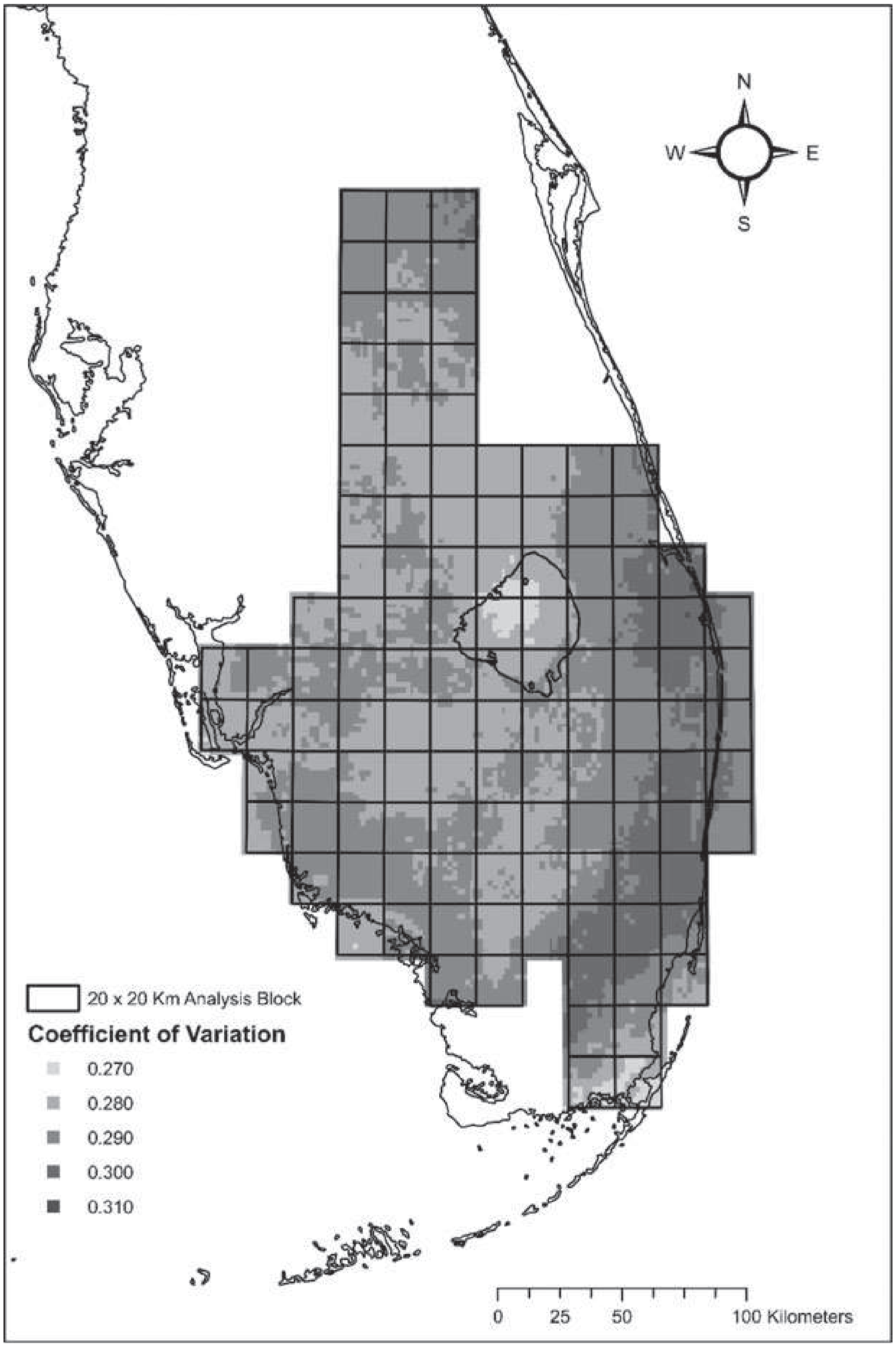

Figure 6. Coefficient of variation of solar insolation based on 2-km grids with overlay of 20-km analysis blocks for the wet season (April–October) based on data from years 1995–2004.

useful results because they changed the data's distribution away from 'normal'. A robust nonlinear estimator is used to obtain the optimal range and sill parameters of the two-parameter model. The nugget variance of insolation is found to be close to zero in almost all blocks. This is because the sub-grid variability (within the 2×2 km data resolution) is not considered in this analysis. The range values are found to be directionally independent (i.e. no dependence on direction within a block is presumed or made) and hence isotropic kriging is considered. Analysis of solar radiation data also confirmed the hypothesis that variations in land-surface features and land-use across the SFWMD region contribute to regional variations in solar insolation due to the heterogeneity in heating. Due to this the range parameter increases in regions with more spatially uniform insolation. The existence of wet lands, large lakes and continuous fields of sugarcane in the region contribute to some of the variations in observed insolation.

Figures 4–6 clearly show low insolation variability over the Lake Okeechobee. In contrast, locations where differential surface heating is dominated by small-scale (20 to ~100 km) surface characteristics, solar insolation would be more variable from day to day, and even from hour to hour (Campbell & Norman 1998, Chapter 10). An example of this would be shorelines where sea breeze circulations cause convective clouds to form (Atkinson 1981), and in areas containing a large lake such as Lake Okeechobee in the study region. A tendency for smaller ranges along the eastern and southeastern coastal boundary, from the Atlantic coast to approximately 40–50 km inland, is noted. Similar patterns are seen for regions on the southwest Gulf of Mexico coast. Over the far north, range values are highly variable, differing by nearly 15 km. This high block-to-block variability may be caused by the small lakes mixed with natural vegetation in this portion of the SFWMD region. The Lake Okeechobee area is mostly dominated by range parameters of approximately 19 km.

In summary, from the analysis of upper quartile and mean detrended GOES insolation data for the wet season, it can be concluded that: (1) wet season insolation had lower range values than those from the dry season; therefore, wet season insolation data are used for the network design analysis; this is due to the higher frequency of occurrence of small cumulus clouds; (2) mean detrended data provide reasonable and more stable (i.e. consistent across years) range values, and these are used in the following analysis; (3) range values are highly variable in time; therefore, annual mean insolation data for the wet season, and for each year, are used for the following analysis; and (4) range values are highly variable in space for 20-km, 40-km, 60-km and 80-km block sizes. Therefore, based on these results, a combination of 20- and 40-km blocks across the District are chosen for the detailed semivariogram analysis. These analysis blocks are shown in Figure 7. A sensitivity analysis between the block size and correlation lengths is performed. The computed correlation lengths (i.e. range parameter values from semivariograms) defined the optimal distances between the ground sensors for the network design, which are determined as a function of location across the District. The range parameter values for 40-km and 20-km analysis blocks are provided in Tables 2a–2c and Tables 3a and 3b, respectively.

Ground-based sensor data analysis

Solar insolation data collected from 29 ground sensors in the SFWMD region shown in Figure 3 are used for the analysis. Details of these sensors, data collection methods and frequency of measurements are provided by Pathak (2008). An initial assessment of available solar insolation data from the SFWMD revealed several limitations of ground sensor data due to observational and systematic errors. The suspicious data flagged by the SFWMD are removed prior to the subsequent outlier identification analyses. The sensor-based data are screened for outliers using a Z-score method (Shiffler 1988). Observations having a Z-score equal to or greater than 3 are identified as outliers and are removed. Anomalous observations are also identified and eliminated. The satellite-sensor

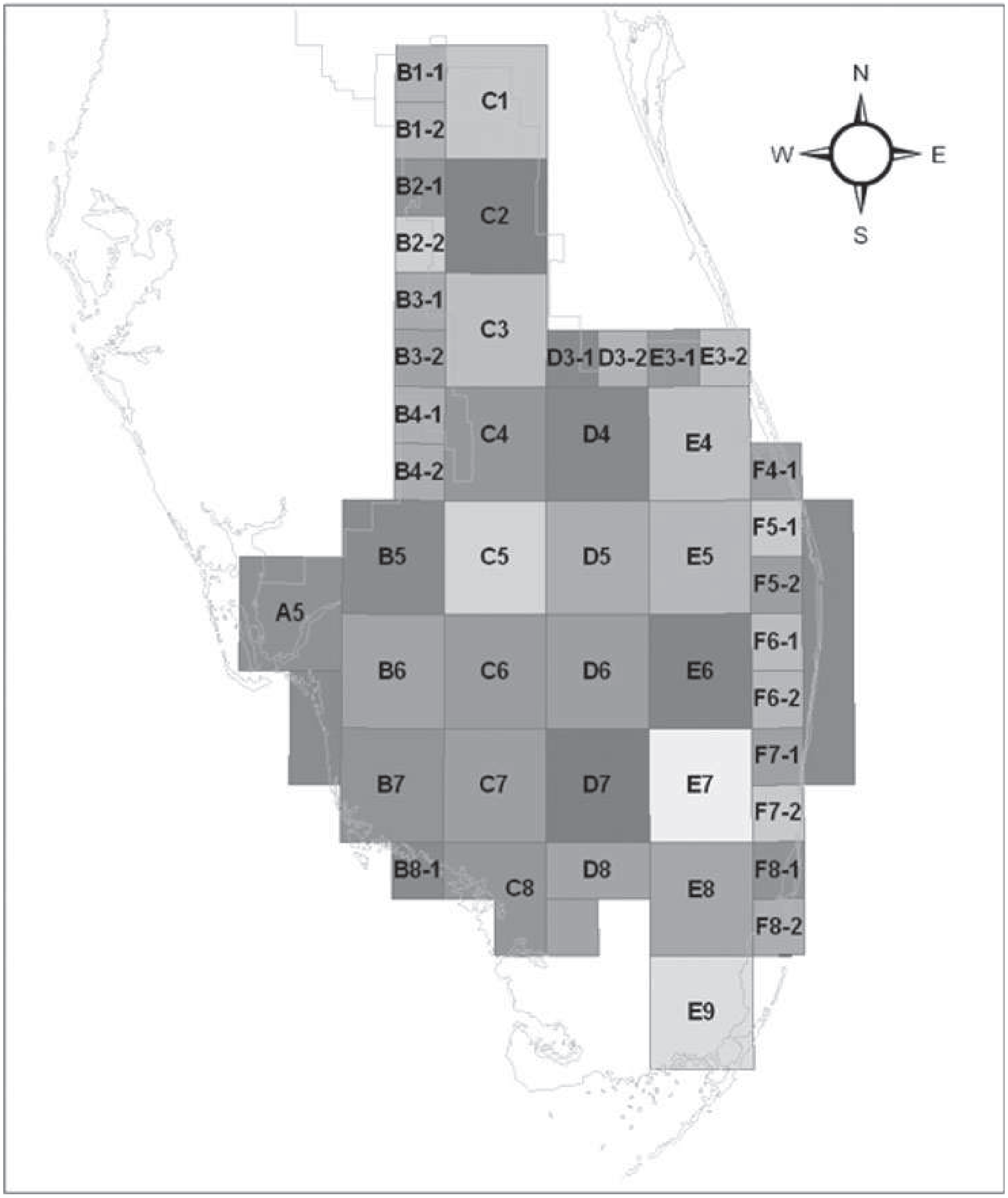

Figure 7. Analysis blocks of different sizes (40 km and 20 km) used for this study in the SFWMD region.

observation pairs are eliminated if the sensori or satellite-based data on any given day are identified as an outlier or an anomaly. Several outliers identified were already flagged as suspicious observed sensor data by the SFWMD. Summary statistics of processed insolation data for each of these sensors is provided in Table 1. The correlation coefficient values provided in Table 1 indicate a good agreement between satellite- and sensor-based radiation datasets. In order to spatially characterise the solar insolation data, the analysis of variance of the point measurements is needed. The point variance values are

Table 2a. Range parameter (in km) values for different 40-km analysis blocks based on 10-year solar insolation data. See Figure 7 for locations of blocks listed.

	40 × 40 km analysis block								
Year	B7	E8	C4	C5	A5	C3	C8	E4	E9
1995	37.36	25.25	37.27	33.95	29.39	10.29	38.35	37.35	17.83
1996	37.36	18.94	37.27	38.36	31.72	39.46	16.83	37.35	18.91
1997	37.36	38.43	17.19	36.61	16.62	25.24	17.83	37.35	21.79
1998	32.60	10.75	19.62	38.36	39.66	11.49	35.52	26.20	26.54
1999	33.09	28.18	15.55	10.44	33.06	23.06	38.35	12.53	30.19
2000	27.81	32.34	37.27	10.44	29.67	19.67	29.37	37.35	19.13
2001	29.65	14.11	26.73	37.58	26.48	39.46	12.06	37.35	21.54
2002	29.15	26.07	37.27	26.27	24.21	12.45	20.69	34.42	18.47
2003	31.64	38.43	37.27	23.39	16.87	14.97	13.88	35.07	38.09
2004	37.36	15.92	14.95	11.37	26.54	23.52	19.66	26.08	30.74
Average	33.34	24.84	28.04	26.68	27.42	21.96	24.26	32.11	24.33

obtained from 29 ground sensors for the wet periods in each of the years from 1995 to 2004. Two different interpolation methods, (1) the inverse distance weighting method (IDWM) and (2) ordinary kriging, are used to obtain an interpolated surface from the point variances of solar insolation. Twenty-five ground sensors are used for developing the surface, and four sensors (viz., sensor numbers 12, 18, 20, 26) are used for validation purposes. Ordinary kriging requires that the observations follow a Gaussian distribution. The normality of the point variance data from the ground sensors is evaluated using a normal probability plot (Mage 1982) and the Kolmogorov and Smirnov (KS) test (Sheskin 2003). An exponential semi-variogram model is selected in the case of kriging after several variogram models are evaluated. The IDWM is implemented using four nearest neighbours (i.e. sensor sites). The two interpolation methods are evaluated using two error measures (viz., mean absolute error and root mean squared error) based on observed and estimated values of solar radiation at validation sensor locations. The IDWM method

Table 2b. Range parameter (in km) values for different 40-km analysis blocks based on 10-year solar insolation data.

	40 × 40 km analysis block						
Year	B5	B6	C1	C2	C6	C7	D4
1995	28.89	13.27	7.51	37.27	13.90	21.61	23.04
1996	12.26	30.05	16.30	26.53	15.05	12.00	25.24
1998	23.54	9.58	24.18	12.97	13.93	14.35	23.75
1999	13.20	7.87	12.05	16.12	39.45	24.48	25.47
2000	38.47	29.16	8.55	31.85	10.43	11.69	24.33
2001	11.05	10.49	9.34	12.33	14.56	9.69	27.16
2002	19.93	15.63	14.33	13.87	10.12	21.19	14.96
2003	38.47	39.56	10.48	10.31	32.75	16.09	20.00
2004	10.51	10.17	11.91	37.27	20.40	24.86	13.70
Average	21.81	18.42	12.74	22.06	18.95	17.33	21.96

Table 2c. Range parameter (in km) values for different 40-km analysis blocks based on 10-year solar insolation data.

	40 × 40 km analysis blocks						
Year	D5	D6	D7	D8	E5	E6	E7
1995	36.35	14.88	18.69	19.34	27.59	9.07	13.19
1996	28.97	23.32	10.64	22.26	12.07	21.83	12.04
1998	29.25	29.69	11.24	15.78	15.93	39.54	27.19
1999	32.15	10.28	15.22	32.05	22.80	18.87	19.48
2000	38.34	7.75	13.90	38.34	25.09	19.40	11.21
2001	22.19	11.22	37.25	16.28	13.64	24.78	37.34
2002	29.36	9.07	18.99	19.36	11.59	15.88	12.24
2003	20.47	12.56	9.73	17.85	13.36	22.21	20.54
2004	24.69	9.62	37.25	26.28	25.44	23.76	22.46
Average	29.09	14.27	19.21	23.06	18.61	21.7	19.52

provided 30 percent and 17.5 percent lower values of MAE and RMSE values respectively compared to those based on application of kriging. The correlation coefficients based on observed and estimated data at four locations are 0.9 and 0.6 for IDWM and kriging respectively.

The range parameter from an exponential semivariogram used to model the covariance structure of the solar insolation from satellite data is obtained for each block. The effective number of sensors is then obtained from the correlation distance from satellite data and point variance from sensor data for each analysis block. The variations in normalised *SE* in relation to number of sensors are evident from plots shown in Figure 8a and Figure 8b. The exponential decay curves defining the relationship between *SE* values and number of sensors in all the 40-km analysis blocks are also shown in these plots. It can be observed that as the number of sensors increases the value of *SE* decreases and remains more or less constant after a specific number for each analysis block. This information is used to estimate the number of sensors needed in each block.

Table 3a. Range parameter (in km) values for different 20-km analysis blocks based on 10-year solar insolation data. See Figure 7 for locations of blocks listed.

	20 × 20 km analysis block										
Year	B1-1	B1-2	B2-1	B2-2	B3-1	B3-2	B4-1	B4-2	B8-1	D3-1	D3-2
1995	9.35	13.79	6.48	4.33	19.77	5.43	6.74	10.50	18.09	17.52	3.24
1996	8.59	14.85	8.91	7.70	19.77	5.19	11.05	6.48	16.12	6.10	6.37
1997	11.88	5.05	17.58	17.58	11.48	6.93	17.58	9.13	19.76	7.93	17.54
1998	5.67	6.46	17.58	4.80	3.64	8.54	10.88	14.57	19.76	6.40	5.91
1999	17.58	4.84	5.25	5.70	4.66	5.69	5.36	4.96	12.11	14.78	7.27
2000	4.04	9.93	6.90	17.58	17.52	7.62	5.51	10.00	13.43	17.52	4.69
2001	12.97	10.11	6.40	17.58	7.72	5.64	5.72	11.19	19.76	8.45	8.52
2002	9.75	8.16	5.52	15.18	12.04	7.24	6.74	13.14	17.98	9.77	16.55
2003	11.64	18.67	4.38	7.96	9.62	8.61	9.12	7.36	16.27	6.18	7.24
2004	17.58	6.40	10.59	11.00	9.39	7.64	11.84	17.58	10.53	4.81	5.00
Average	10.90	9.83	8.96	10.94	11.56	6.85	9.05	10.49	16.38	9.94	8.23

Table 3b. Range parameter (in km) values for different 20-km analysis blocks based on 10-year solar insolation data.

	20 × 20 km analysis block										
Year	E3-1	E3-2	F4-1	F5-1	F5-2	F6-1	F6-2	F7-1	F7-2	F8-1	E3-1
1995	11.96	5.19	5.33	9.34	9.74	12.84	16.12	10.88	4.70	4.57	17.61
1996	6.92	17.59	8.68	13.92	12.05	11.18	14.59	15.36	8.70	8.39	11.24
1997	5.22	17.59	9.31	18.71	5.65	13.61	15.02	7.97	17.61	10.91	17.61
1998	16.81	12.17	17.62	11.20	8.53	10.82	5.88	7.32	9.23	11.85	10.55
1999	5.31	7.89	17.62	18.71	17.62	7.55	5.71	7.11	17.61	6.52	17.61
2000	7.74	7.33	4.18	9.34	12.31	10.45	9.16	17.61	17.61	6.69	17.61
2001	17.57	9.17	15.98	9.29	4.72	12.74	10.85	7.81	14.83	13.53	17.61
2002	5.14	5.20	7.41	6.72	7.22	14.73	14.06	5.04	17.61	6.89	6.19
2003	5.74	7.31	17.62	11.83	13.39	7.33	6.56	7.58	13.53	7.21	17.61
2004	9.49	10.36	6.78	11.43	12.66	6.10	6.38	9.90	17.61	12.01	10.20
Average	9.19	9.98	11.05	12.05	10.39	10.73	10.43	9.66	13.90	8.86	14.38

Optimal network design

The optimal network design is conditioned on a specific value of achievable *SE* to obtain the number of sensors in each block. Based on the covariance structure of the solar insolation data, it is evident that there is no uniform analysis block size that is identifiable for the region under consideration. Two analysis size blocks, 40 × 40 km and 20 × 20 km as shown in Figure 7[!−insert−], are identified for use of an optimum design of the sensor network. The network designed is optimal considering the stipulations imposed on the block size and other constraints. In this study, a uniform value of *SE* is adopted to obtain the optimum number of sensors for each block. Once the optimum number is obtained for each block, this number is compared with the existing number of sensors in each analysis block. If the existing number of sensors in a block is higher than the

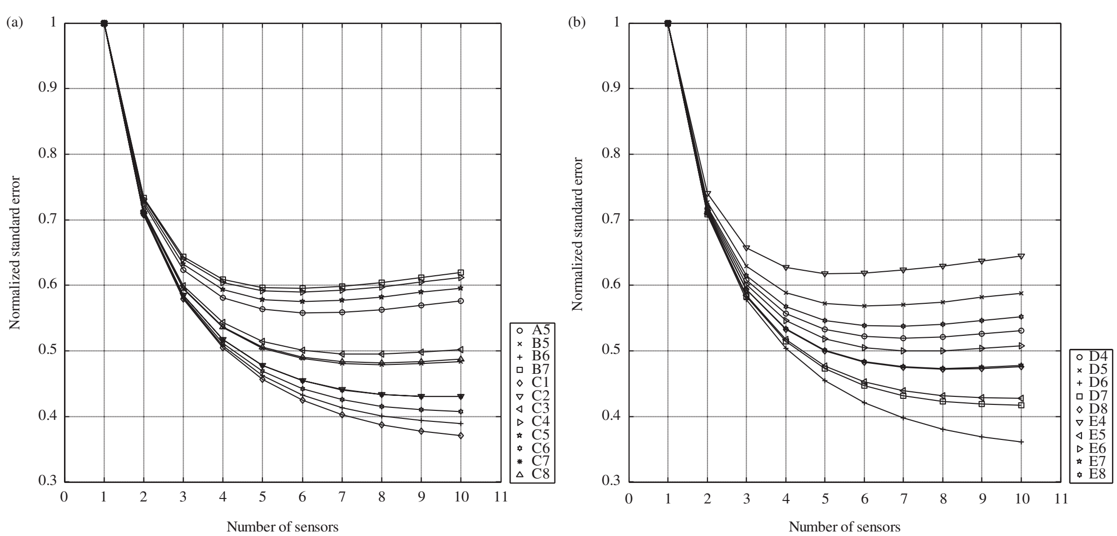

Figure 8. Variation of normalised standard error based on the number of sensors in all the 40-km analysis blocks.

Table 4a. Variation of standard error (*SE*) with the number of sensors for each 40-km analysis block. See Figure 7 for locations of blocks listed.

Number of sensors	Standard error (MJ day^{-1} m^{-2})										
	A5	B5	B6	B7	C1	C2	C3	C4	C5	C6	C7
1	5.448	5.541	5.418	5.010	5.784	5.659	5.560	5.493	5.684	5.425	5.250
2	3.944	3.946	3.834	3.674	4.092	4.012	3.968	4.019	4.137	3.842	3.722
3	3.399	3.297	3.145	3.225	3.349	3.311	3.333	3.517	3.595	3.159	3.072
4	3.166	2.968	2.753	3.051	2.920	2.930	3.021	3.319	3.373	2.778	2.718
5	3.069	2.794	2.506	2.990	2.643	2.706	2.861	3.247	3.288	2.545	2.510
6	3.041	2.705	2.346	2.983	2.456	2.573	2.786	3.237	3.269	2.399	2.386
7	3.045	2.664	2.241	3.001	2.328	2.495	2.756	3.254	3.281	2.308	2.314
8	3.067	2.653	2.172	3.029	2.239	2.453	2.752	3.284	3.308	2.253	2.274
9	3.103	2.664	2.133	3.067	2.182	2.439	2.771	3.325	3.349	2.226	2.261
10	3.137	2.681	2.109	3.103	2.143	2.437	2.793	3.363	3.387	2.226	2.259

optimal number, then a recommendation is made to remove the additional number of sensors. Additional sensors are recommended for any block if the existing number of sensors in that block is less than the optimal.

Ground sensor network determination

For recommending an optimal sensor network to the District, several values of *SE* are evaluated for each block. *SE* values for different numbers of sensors in each 40-km and 20-km analysis block are provided in Tables 4a and 4b and Tables 5a and 5b respectively. The total number of new sensors required for different values of *SE* is given in Table 6. The number of sensors required in each analysis block for a specific value of *SE* is provided in Table 7. Considering the cost associated with installation and long-term maintenance of new sensors and consultation with District staff, the proposed ground sensor network with *SE* of 5.0 MJ day^{-1} m^{-2} that would require 19 additional new sensors is identified. The selection of this network is based on two facts: (a) the standard deviation of satellite solar insolation data from wet seasons over 10 years varied between 5.23 and 6.07 MJ day^{-1} m^{-2} (Figure 5), which is in the range of the

Table 4b. Variation of standard error (*SE*) with the number of sensors for each 40-km analysis block.

Number of sensors	Standard error (MJ day^{-1} m^{-2})											
	C8	D4	D5	D6	D7	D8	E4	E5	E6	E7	E8	E9
1	5.272	5.874	5.975	5.453	5.310	5.346	5.846	5.877	5.526	5.350	5.411	6.011
2	3.755	4.212	4.340	3.857	3.762	3.804	4.328	4.166	3.948	3.807	3.896	4.292
3	3.140	3.570	3.759	3.155	3.098	3.172	3.842	3.437	3.322	3.176	3.328	3.609
4	2.831	3.270	3.517	2.746	2.731	2.848	3.665	3.039	3.017	2.853	3.073	3.276
5	2.667	3.128	3.421	2.479	2.510	2.673	3.612	2.804	2.863	2.680	2.958	3.107
6	2.585	3.069	3.396	2.296	2.375	2.582	3.616	2.663	2.792	2.590	2.916	3.028
7	2.548	3.053	3.407	2.168	2.293	2.538	3.643	2.581	2.766	2.547	2.912	2.998
8	2.539	3.062	3.433	2.076	2.246	2.523	3.680	2.535	2.765	2.533	2.927	2.996
9	2.551	3.090	3.475	2.015	2.225	2.531	3.727	2.519	2.785	2.542	2.958	3.017
10	2.568	3.120	3.514	1.971	2.217	2.546	3.769	2.515	2.808	2.556	2.989	3.042

Table 5a. Variation of standard error (*SE*) with the number of sensors for each 20-km analysis block.

Number of sensors	Standard error (MJ day^{-1} m^{-2})										
	B1-1	B1-2	B2-1	B2-2	B3-1	B3-2	B4-1	B4-2	B8-1	D3-1	D3-2
1	5.741	5.756	5.574	5.596	5.618	5.594	5.588	5.622	5.146	5.559	5.701
2	4.098	4.089	3.945	3.979	4.008	3.960	3.960	4.007	3.789	3.942	4.035
3	3.443	3.395	3.236	3.314	3.363	3.254	3.267	3.355	3.342	3.256	3.312
4	3.120	3.031	2.834	2.970	3.045	2.857	2.888	3.028	3.173	2.885	2.902
5	2.956	2.827	2.582	2.782	2.881	2.613	2.663	2.858	3.117	2.667	2.647
6	2.882	2.718	2.420	2.686	2.806	2.460	2.531	2.777	3.119	2.543	2.484
7	2.847	2.656	2.311	2.634	2.770	2.360	2.449	2.736	3.135	2.466	2.375
8	2.844	2.631	2.242	2.615	2.765	2.299	2.406	2.728	3.166	2.427	2.307
9	2.867	2.636	2.205	2.625	2.786	2.270	2.393	2.747	3.210	2.419	2.272
10	2.886	2.641	2.179	2.635	2.804	2.251	2.386	2.762	3.243	2.415	2.247

selected *SE* (5.5 MJ day^{-1} m^{-2}) of the proposed network; and (b) the selected proposed network would require three additional sensors compared to the 16 new sensors required for the proposed network with the *SE* of 5.5 MJ day^{-1} m^{-2} that is considered to be an optimal network. Therefore, the recommended optimal ground sensor network would need a total of 19 new sensors with the *SE* of 5.0 MJ day^{-1} m^{-2}. Figure 9 and Figure 10 show excess and insufficient sensor distributions for the SFWMD region for the *SE* of 5.0 MJ day^{-1} m^{-2}. Table 8a and Table 8b provide the details of the required, existing and excess number of sensors in each block.

Data accuracy validation and sensor placement

Sensors may be added or removed from an analysis block based on the optimal number specified by the geo-statistics-based methodology. The accuracy of the sensor network can be analysed when a selected number of sensors are removed from the analysis blocks. Analysis block D5 is selected to demonstrate the effect of withholding the sensors. Four ground sensors are located in the analysis block of D5 and are shown in Figure 3. The approach for accuracy assessment is to calculate the average value of solar insolation estimated by the satellite observations in the analysis block, and then compare them with

Table 5b. Variation of standard error (*SE*) with the number of sensors for each 20-km analysis block.

Number of sensors	Standard error (MJ day^{-1} m^{-2})										
	E3-1	E3-2	F4-1	F5-1	F5-2	F6-1	F6-2	F7-1	F7-2	F8-1	F8-2
1	5.769	5.759	5.743	5.801	5.969	5.880	5.715	5.559	5.500	5.532	5.552
2	4.086	4.087	4.079	4.150	4.264	4.201	4.070	3.940	4.029	3.920	4.099
3	3.360	3.382	3.386	3.503	3.587	3.538	3.402	3.249	3.532	3.231	3.627
4	2.956	3.005	3.021	3.193	3.257	3.215	3.064	2.871	3.337	2.853	3.452
5	2.710	2.789	2.817	3.040	3.091	3.054	2.885	2.647	3.268	2.627	3.397
6	2.558	2.668	2.707	2.977	3.018	2.984	2.798	2.514	3.264	2.494	3.402
7	2.460	2.596	2.644	2.950	2.985	2.953	2.753	2.432	3.277	2.410	3.421
8	2.403	2.562	2.618	2.954	2.983	2.952	2.741	2.387	3.308	2.364	3.455
9	2.378	2.559	2.622	2.982	3.009	2.979	2.757	2.374	3.354	2.349	3.504
10	2.362	2.559	2.627	3.005	3.030	3.000	2.771	2.366	3.388	2.340	3.539

Table 6. Variations in the number of sensors required and additional needed (new) for different values of standard error.

Standard error (MJ day^{-1} m^{-2})	Number of sensors		
	Required	Available	New
3.5	124	29	95
3.8	97	29	68
4.5	79	29	50
5.0	48	29	19
5.5	45	29	16

insolation values obtained by averaging the available ground sensor observations. This process involves progressively eliminating one ground sensor at a time and then evaluating the accuracy of the remaining sensors in characterising the variability of the solar insolation data. Summary statistics are calculated considering decreasing numbers of sensors (Table 9). The results show minimal deviations in summary statistics of the insolation data from those from satellite-based data in the block when two sensors are eliminated.

The sensor assignment and placement for a new network will depend on the existing number and required sensors and the practical considerations and logistics related to main-

Table 7. Number of sensors required in different analysis blocks for a specific standard error.

	Standard error (MJ day^{-1} m^{-2})					Standard error (MJ day^{-1} m^{-2})				
Block	3.5	3.8	4.5	5	5.5	Block	3.5	3.8	4.5	5	5.5
A5	2	2	1	1	1	B1-1	2	2	1	1	1
B5	2	2	1	1	1	B1-2	2	2	1	1	1
B6	2	2	1	1	1	B2-1	2	2	2	1	1
B7	2	2	1	1	1	B2-2	3	2	2	1	1
C1	2	2	1	1	1	B3-1	3	2	2	1	1
C2	2	2	1	1	1	B3-2	3	2	2	1	1
C3	2	2	1	1	1	B4-1	3	2	2	1	1
C4	2	2	1	1	1	B4-2	3	2	2	1	1
C5	2	2	1	1	1	B8-1	3	2	2	1	1
C6	3	2	2	1	1	D3-1	3	2	2	1	1
C7	3	2	2	1	1	D3-2	3	2	2	1	1
C8	3	2	2	1	1	E3-1	3	2	2	1	1
D4	3	2	2	1	1	E3-2	3	2	2	1	1
D5	3	2	2	1	1	F4-1	3	2	2	1	1
D6	3	2	2	1	1	F5-1	3	2	2	1	1
D7	3	2	2	1	1	F5-2	3	2	2	1	1
D8	3	2	2	1	1	F6-1	3	2	2	1	1
E4	3	3	2	1	1	F6-2	3	2	2	1	1
E5	3	3	2	1	1	F7-1	3	2	2	1	1
E6	3	3	2	1	1	F7-2	3	2	2	1	1
E7	3	3	2	1	1	F8-1	3	2	2	1	1
E8	3	3	2	2	1	F8-2	3	3	2	2	1
E9	4	3	2	2	1	Total	63	45	42	23	22
Total	61	52	37	25	23						

Table 8a. 40-km analysis blocks with insufficient and excess number of sensors for $SE = 5.0\,\mathrm{MJ\ day^{-1}\ m^{-2}}$.

Block	Required sensors	Existing sensors	Existing−required	Excess	Insufficient
A5	1	0	−1	0	1
B5	1	0	−1	0	1
B6	1	1	0	0	0
B7	1	2	1	1	0
C1	1	0	−1	0	1
C2	1	2	1	1	0
C3	1	0	−1	0	1
C4	1	2	1	1	0
C5	1	1	0	0	0
C6	1	1	0	0	0
C7	1	0	−1	0	1
C8	1	0	−1	0	1
D4	1	2	1	1	0
D5	1	4	3	3	0
D6	1	2	1	1	0
D7	1	1	0	0	0
D8	1	1	0	0	0
E4	1	1	0	0	0
E5	1	0	−1	0	1
E6	1	3	2	2	0
E7	1	0	−1	0	1
E8	2	1	−1	0	1
E9	2	1	−1	0	1
Total	25	25	0	10	10

tenance and installation options. The required sensor density is initially determined by using a specific value of *SE* in 40-km and 20-km analysis blocks. Additional sensors are proposed to meet the required optimal sensor density. These sensors are expected to be installed within an analysis block with a consideration given to the inter-sensor separation distance computed from the range parameter in the semivariogram analysis. The inter-sensor optimal separation distances are identified to be in between 32 and 40 km for 40-km analysis blocks and 16 and 20 km for 20-km analysis blocks respectively. The excess sensors from one or more analysis blocks can be utilised in blocks where the number of required sensors is greater than the existing number. The locations of the sensors in blocks near the coast (20-km blocks) need to be decided based on the field conditions.

10. General remarks

The focus of the current study is only on the design of an optimal solar radiation measurement sensor network. The placement of the proposed sensors within different analysis blocks of fixed spatial resolution will require an additional study that will address the cost associated with the installation and/or removal (if required) of sensors and other logistical issues. The methodology developed in the current study can also be applied to regions where there are no existing ground-based sensors. In the current study, data from an existing network of ground-based sensors are used to validate the satellite-based data. However, there is no need for ground-based sensors to obtain the design of the network as long as the satellite-based data are assured to be of good quality without any errors or anomalies. The establishment of spatially homogeneous areas for the definition of

Table 8b. 20-km analysis blocks with insufficient and excess number of sensors for $SE = 5.0\,\mathrm{MJ\ day^{-1}\ m^{-2}}$.

Block	Required sensors	Existing sensors	Existing−required	Excess	Insufficient
B1-1	1	0	−1	0	1
B1-2	1	0	−1	0	1
B2-1	1	0	−1	0	1
B2-2	1	0	−1	0	1
B3-1	1	0	−1	0	1
B3-2	1	0	−1	0	1
B4-1	1	0	−1	0	1
B4-2	1	0	−1	0	1
B8-1	1	0	−1	0	1
D3-1	1	0	−1	0	1
D3-2	1	0	−1	0	1
E3-1	1	0	−1	0	1
E3-2	1	0	−1	0	1
F4-1	1	1	0	0	0
F5-1	1	1	0	0	0
F5-2	1	1	0	0	0
F6-1	1	1	0	0	0
F6-2	1	0	−1	0	1
F7-1	1	0	−1	0	1
F7-2	1	0	−1	0	1
F8-1	1	0	−1	0	1
F8-2	2	0	−2	0	2
Total	23	4	−19	0	19

analysis blocks is carried out by evaluating the mean, standard deviation and coefficient of variation of solar insolation data in this study. Methods for identification of homogeneous areas using these summary statistics of data, similarity of probability distributions of observations at several sites in a region, and cluster analysis were recommended by Haan (2002) and Hosking, Wallis, and Wood (1985). The summary statistics-based method used for definition of analysis blocks (i.e. homogeneous areas) in the current study is conceptually simple and involves an iterative process. However, use of the variance quad-tree or cluster approach is recommended for definition of block sizes to avoid any element of subjectivity in the delineation process of homogeneous areas.

Table 9. Summary statistics of sensor data in the analysis block D5 with different number of sensors removed.

Summary statistic	Satellite	All four sensors	Three sensors[1]	Two sensors[2]
Average	19.422	18.721	18.713	18.383
Median	19.743	18.943	18.936	18.727
Standard deviation	3.058	3.147	3.365	4.007
Minimum	5.658	6.912	5.501	5.227
Maximum	25.600	25.661	26.064	27.950

[1] Sensor L005 not included.

11. Conclusions

A methodology to design an optimal sensor network to characterise the spatial variability of solar insolation useful for estimation of evapotranspiration (ET) in a region is proposed and evaluated in this study. Application of the methodology to upgrade an existing network of solar radiation sensors (i.e. pyranometers) in a region of South Florida, USA, is reported in this paper. Geostationary operational environmental satellite (GOES) satellite and ground sensor network-based data are used in this study for the design of the network. The optimal network is expected to improve the estimation of ET using a simple solar-radiation-based ET estimation method in the region. An array of analysis blocks with two different fixed spatial resolutions (20 and 40 km) is defined based on the evaluation of spatial variability of solar insolation data in the study region. Geospatial and geostatistical analyses are used to assess the solar insolation within each analysis block and to obtain an optimal number of sensors. Results from the analyses conducted in this study indicate that the number of sensors required in each analysis block depends on the standard error (*SE*) set as a criterion for network measurement accuracy. An optimal sensor network that is expected to provide a standard error (*SE*) of 5.0 MJ day^{-1} m^{-2} is selected to demonstrate the utility of the proposed methodology in this study. A separate study needs to be carried out to clearly define the implementation strategies for the recommended network developed in this study. The methodology proposed and evaluated in this study is generic and can used for design of an optimal monitoring network for any hydro-meteorological variable.

Acknowledgements

The authors sincerely thank the three anonymous reviewers and the associate editor for providing several constructive comments that have led to substantial improvement of the manuscript. The study reported in the paper was supported by South Florida Water Management District (SFWMD), Florida, USA.

References

Abtew, W. (1996) Evapotranspiration measurements and modeling for three wetland systems in south Florida. *Water Resources Bulletin*, vol. 32, no. 3, pp. 465–473.

Abtew, W., & Melesse, A. (2013) *Evaporation and Evapotranspiration: Measurements and Estimations*, Springer, Dordrecht.

Abtew, W., Irizarry-Ortiz, M., Lyon, D., & Obeysekera, J. (2002) Historical potential evapotranspiration estimation for sites in South Florida, Technical Notes. South Florida Water Management District, West Palm Beach, Florida.

Ahuja, P.R. (1960) Planning for precipitation network for water resources development in India, WMO Flood Control Series. Report No. 15, 106–112.

Atkinson, B.W. (1981) *Meso-scale Atmospheric Circulations*, Academic Press, London.

Bhat, S., Motz, L.H., Pathak, C., & Kuebler, L. (2015) Geostatistics-based groundwater-level monitoring network design and its application to the Upper Floridan aquifer, USA. *Environmental Monitoring and Assessment*, vol. 187, no. 1, pp. 4183 doi:10.1007/s10661-014-4183-x.

Brent, R. (1973) *Algorithms for Minimization without Derivatives*, Prentice-Hall, Upper Saddle River, NJ.

Byers, H.R., & Rodebush, H.R. (1948) Cause of thunderstorms of the Florida peninsula. *Journal of Atmospheric Science*, vol. 5, pp. 275–280.

Campbell, G.S., & Norman, J.M. (1998) *An Introduction to Environmental Biophysics*, Springer, Dordrecht.

Chang, C.L., & Lin, Y.T. (2014) A water quality monitoring network design using fuzzy theory and multiple criteria analysis. *Environmental Monitoring Assessment*, vol. 186, no. 10, pp. 6459–6469.

Darnell, W.L., Staylor, W.F., Gupta, S.K., & Denn, F.M. (1988) Estimation of surface insolation using sun-synchronous satellite data. *Journal of Climate*, vol. 1, pp. 820–835.

Delclaux, F., & Coudrain, A. (2005) Optimal evaporation models for simulation of large lake levels: application to Lake Titicaca, South America. *Geophysical Research Abstracts*, vol. 7, pp. 53–65.

Dedieu, G., Deschamps, P.Y., & Kerr, Y.H. (1987) Satellite estimates of solar irradiance at the surface of the earth and of surface albedo using a physical model applied to Meteosat data.

Journal of Climate and Applied Meteorology, vol. 26, pp. 79–87.
Diak, G.R., Bland, W.L., & Mecikalski, J.R. (1996) A note on first estimates of surface insolation from GOES-8 visible satellite data. *Agricultural Forest Meteorology*, vol. 82, pp. 219–226.
Doorenbos, J., & Pruitt, W.O. (1977) Guidelines for predicting crop water requirements, FAO irrigation and drainage paper 24.FAO, Rome.
Enku, T., van der Tol, C., Gieske, A.S.M., & Rientjes, T.H.M. (2011) Evapotranspiration modeling using remote sensing and empirical models in the Fogera floodplain, Ethiopia. *In:* Melesse, A., ed. *Nile River Basin: Hydrology, Climate and Water Use*, Springer, Dordrecht, pp. 163, 170.
Fortin, M.J., & Dale, M. (2005) *Spatial Analysis: A Guide for Ecologists*, Cambridge University Press, Cambridge.
Frouin, R., & Chertock, B. (1992) A technique for global monitoring of net solar irradiance at the ocean surface. Part I: model. *Journal of Applied Metrology*, vol. 31, pp. 1056–1066.
Ganguli, M.K., Rangarajan, R., & Panchang, G.M. (1951) Accuracy of mean rainfall estimates-Data of Damodar catchment. *Irrigation and Power Journal*, vol. 8, pp. 278–284.
Gautier, C., Diak, G.R., & Masse, S. (1984) An investigation of the effects of spatially averaging satellite brightness measurements on the calculation of insolation. *Journal of Climate and Applied Meteorology*, vol. 23, pp. 1380–1386.
Gringarten, E., & Deutsch, C.V. (2001) Teacher's aide: Variogram interpretation and modeling. *Mathematical Geology*, vol. 33, no. 4, pp. 507–534.
Haan, C.T. (2002) *Statistical Methods in Hydrology*, Iowa Press, Ames, IA.
Han, J., & Kamber, M. (2006) *Data Mining: Concepts and Techniques*, Morgan Kaufmann, Waltham.
Haining, R.P. (1990) *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge.
Hargreaves, G.H., & Samani, Z.A. (1985) Reference crop evapotranspiration from temperature. *Applied Engineering Agriculture*, vol. 1, no. 12, pp. 96–99.
Heymesfield, G.M., Sheperd, J.M., Bidwell, S.W., Boncyk, W.C., Caylor, I.J., Ameen, A., & Olson, W.S. (1996) Structure of Florida thunderstorms using high-altitude aircraft radiometer and radar observations. *Journal of Applied Meteorology*, vol. 35, pp. 1736–1762.
Hosking, J.R., Wallis, J.R., & Wood, E.F. (1985) An appraisal of the regional flood frequency procedure in the UK Flood Studies Report, *Hydrological Sciences Journal*, vol. 30, no. 1, 85–109.
Isaaks, H.E., & Srivastava, R.M. (1989) *An Introduction to Applied Geostatisitics*, Oxford University Press, New York.
Jacobs, J.M., Myers, D.A., Andersoin, M.C., & Diak, G.R. (2002) GOES surface insolation to estimate wetland evapotranspiration. *Journal of Hydrology*, vol. 266, pp. 53–65.
Kinsman, G., Kite, J., & Mtundu, D. (1994) Guidelines for the collection of hydrologic and meteorological data volume I: field applications, South Florida Water Management District, West Palm Beach, USA.
Larish, B., & Riley, G. (2011) An information theoretic approach for determining the minimum number of sensors in a wireless sensor network. *Information Theory Workshop, IEEE*, pp. 316–320 doi:10.1109/ITW.2011.6089445.
LI-COR, Inc (2015) *LS-200S Pyranometer Instruction Manual*, Campbell Scientific, Logan, UT.
Lloyd, D.C. (2007) *Local Models for Spatial Analysis*, CRC Press, Boca Raton, FL.
Mage, D.T. (1982) An objective graphical method for testing normal distributional assumptions using probability plots. *The American Statistician*, vol. 36, no. 2, pp. 116–120.
Matalas, N.C., & Langbein, W.B. (1962) Information content of the mean. *Journal of Geophysics Research*, vol. 67, no. 9, pp. 3441–3448.
Minasny, B., McBratney, A.B., & Walvoort, D.J.J. (2007) The variance quadtree algorithm: use for spatial sampling design. *Computers and Geosciences*, vol. 33, no. 3, pp. 383–392.
Mogheir, Y., de Lima, J.L.M., & Singh, V.P. (2013) Entropy and multi-objective based approach for groundwater quality monitoring network assessment and redesign. *Water Resources Management*, vol. 23, no. 8, pp. 1603–1620.
Monteith, J.L. (1973) *Principles of Environmental Physics*, Edward Arnold, London.
Moser, W., & Raschke, E. (1984) Incident solar radiation over Europe from METEOSAT data. *Journal of Climate and Applied Meteorology*, vol. 23, pp. 166–170.
Olea, R.A. (1984) Sampling design optimization for spatial functions. *Mathematical Geology*, vol. 16, no. 4, pp. 369–392.
O'Sullivan, D., & Unwin, D.J. (2010) *Geographical Information Analysis*, John Wiley, Hoboken, NJ.

Otkin, J., Anderson, M.A., Mecikalski, J.R., & Diak, G.R. (2005) Validation of GOES-based insolation estimates using data from the United States Climate Reference Network. *Journal of Hydrometeorology*, vol. 6, pp. 475–640.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andreassian, V., Anctil, F., & Loumagne, C. (2005) Which potential evapotranspiration input for a lumped model part 2-towards a simple and efficient potential evapotranspiration model for rainfall-runoff modeling. *Journal of Hydrology*, vol. 303, pp. 290–306.

Paech, S.J., Mecikalski, J.R., Sumner, D.M., Pathak, C.S., Wu, Q., Islam, S., & Sangoyomi, T. (2009) Satellite-based solar insolation estimates in support of a long-term climatology of potential and reference evapotranspiration estimates over Florida. *Journal of the American Water Resources Association (JAWRA)*, vol. 45, no. 6, pp. 1328–1342.

Pathak, C.S., ed. (2008) Hydrologic monitoring network of South Florida Water Management District – Technical Report, South Florida Water Management District, West Palm Beach, Florida.

Pathak, C.S., & Vieux, B.E. (2007) Geo-spatial analysis of NEXRAD rainfall data for central and south Florida, American Society of Civil Engineers /Environmental and Water Resources Institute, World Water & Environmental Congress 2007, May 15–19, 2007.Tampa, FL, USA.

Penman, H.L. (1948) Natural evaporation from open water, bare soil and grass. *Proceedings of Royal Society, London, Series A, Mathematical and Physical Sciences*, vol. 193, no. 1032, pp. 120–145.

Pinker, R.T., & Ewing, J.A. (1985) Modeling surface solar radiation: model formulation and validation. *Journal of Climate and Applied Meteorology*, vol. 24, pp. 389–401.

Pinker, R.T., Frouin, R., & Li, Z. (1995) A review of satellite methods to derive surface shortwave irradiance. *Remote Sensing of Environment*, vol. 51, 105–124.

Pinker, R.T., & Laszlo, I. (1992) Modeling surface solar irradiance for satellite applications on global scale. *Journal of Applied Metrology*, vol. 31, pp. 194–211.

Rakhecha, P.R., & Singh, V.P. (2009) *Applied Hydrometeorology*, Springer, New Delhi.

Riehl, H. (1954) *Tropical Meteorology*, McGraw-Hill, New York.

Rycroft, H.B. (1949) Random sampling of rainfall, *Journal of South African Forestry Association*, vol. 18, no. 1, 71–81.

Schiffler, R.E. (1988) Maximum Z-Score and outliers. *The American Statistician*, vol. 42, no. 1, pp. 79–80.

Schmetz, J. (1989) Towards a surface radiation climatology: Retrieval of downward irradiances from satellites. *Atmospheric Research*, vol. 23, no. 3–4, 287–321.

SFWMD (2008) Evapotranspiration (ET) Network design study - Part I - Solar radiation ground sensor network design, Final Draft Report, South Florida Water Management District, West Palm Beach, Florida.

Sheskin, D.J. (2003) *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, FL.

Shoemaker, W.B., & Sumner, D.M. (2006) Alternate corrections for estimating actual wetland evapotranspiration from potential evapotranspiration. *Wetlands*, vol. 26, no. 2, pp. 528–543.

S-Plus (2007) *User's Guide*, Insightful Corporation, Ishington, Seattle.

Spruill, T.B., & Candela, L. (1990) Two approaches to design of monitoring networks. *Groundwater*, vol. 28, no. 3, pp. 430–442.

Tarpley, J.D. (1979) Estimating incident solar radiation at the surface from geostationary satellite data. *Journal of Applied Meteorology*, vol. 18, pp. 1172–1181.

Teegavarapu, R.S.V. (2007) Use of universal function approximation in variance-dependent interpolation technique: an application in hydrology. *Journal of Hydrology*, vol. 332, pp. 16–29.

Teegavarapu, R.S.V. (2014) Missing precipitation data estimation using optimal proximity metric-based imputation, nearest-neighbour classification and cluster-based interpolation methods. *Hydrological Sciences Journal*, vol. 59, no. 11, pp. 2009–2026 http://dx.doi.org/10.1080/02626667.2013.862334

Teegavarapu, R.S.V., & Chandramouli, V. (2005) Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, vol. 312, pp. 191–206.

Webster, R.A., & Oliver, M.A. (2001) *Geostatistics for Environmental Scientists*, John Wiley, Chichester.

Weymouth, G., & Le Marshall, J. (1999) An operational system to estimate global solar exposure over the Australian region from satellite observations. *Australian Meteorological Magazine*, vol. 48, pp. 181–195.

Wilks, D.S. (2006). 2nd edn. *Statistical Methods in the Atmospheric Sciences*, Academic Press, St Louis, MO.

Xing, T., Xuesong, Z., & Taylor, J. (2013) Designing heterogeneous sensor networks for estimating and predicting path travel time dynamics: an information-theoretic modeling approach. *Transportation Research Part B: Methodological*, vol. 57, pp. 66–90.

Xu, C.Y., & Singh, V.P. (2000) Evaluation and generalization of radiation-based methods for calculating evaporation. *Hydrological Processes*, vol. 14, pp. 339–349.

Zhai, L., Feng, Q., Li, Q., & Xu, C. (2009) Comparison and modification of equations for calculating evapotranspiration (ET) with data from Gansu Province, Northwest China. *Irrigation and Drainage* doi:10.1002/ird.502.