UNIVERSIDAD POLITÉCNICA DE MADRID

**Escuela Técnica Superior de Ingeniería Agronómica,
Alimentaria y de Biosistemas.**

**Assessment of reference evapotranspiration and global solar
radiation in Yucatán Peninsula Mexico.**

TESIS DOCTORAL

VICTOR HUGO QUEJ CHI

Ingeniero Agrónomo

Madrid 2017

**DEPARTAMENTO DE PRODUCCIÓN AGRARIA
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA
ALIMENTARIA Y DE BIOSISTEMAS**

UNIVERSIDAD POLITÉCNICA DE MADRID

**Assessment of reference evapotranspiration and global solar
radiation in Yucatán Peninsula Mexico.**

TESIS DOCTORAL

**VICTOR HUGO QUEJ CHI
Ingeniero Agrónomo**

**Director de Tesis
Javier Almorox Alonso
Doctor Ingeniero Agrónomo**

Madrid 2017

UNIVERSIDAD POLITÉCNICA DE MADRID

Tribunal nombrado por el Mgfco. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día..... de..... de 2017.

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Suplente:

Suplente:

Realizado el acto de defensa y lectura de la Tesis el día..... de..... de 2017, en la E.T.S.I. Agronómica, Alimentaria y de Biosistemas.

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

A Moisés Abiel Quej,

A Olga González,

A mi madre María Dolores Chi,

A mi padre y hermanos.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología de México (CONACYT) – Fundación Pablo García del Estado de Campeche por el apoyo económico brindado.

Al Colegio de Postgraduados de México (COLPOS) por autorizar el permiso para ingresar al Subprograma de Formación de Profesores Investigadores (SFPI).

Al Campus Campeche del Colegio de Postgraduados por la oportunidad que me brindaron para continuar mis estudios.

Agradezco en forma muy valiosa a mi director de Tesis, Dr. Javier Almorox Alonso, por sus enseñanzas, por su tiempo y sus valiosas aportaciones para el desarrollo de mi Tesis y artículos científicos.

A los doctores, Mirzakhayot Ibrakhimov, Laurel Saito, Rubén Moratíel y Richard Sneyder por sus valiosas sugerencias y aportaciones.

A la Universidad Politécnica de Madrid (UPM), por aceptarme como estudiante y por las facilidades que me otorgaron para el uso de la biblioteca digital, software e instalaciones físicas.

De manera especial agradezco a mi esposa Olga González por su paciencia y por estar siempre mi lado, a mis padres María Dolores y Víctor, a mis hermanos Wilbert, Elvira y Fredy.

ABSTRACT

In the Yucatan Peninsula Mexico, the irrigated agriculture plays a key role in the crop production. In 2014, irrigated lands accounted for about 88500 ha of the total cultivable area. The development of irrigation systems that make efficient and accurate use of water are essential for the sustainability of crop production systems. Thus, the reference evapotranspiration (ET_0) is one of the most important factors related to irrigation system design, water management under irrigated and rainfed production. Also, the precise knowledge of daily global solar radiation (H) becomes very important in ET_0 process.

To address this challenge, the potential of various empirical models and soft computing (SC) techniques named Support vector machine (SVM), Artificial neural network (ANN), and Adaptive neuro fuzzy inference system (ANFIS) were evaluated for estimating ET_0 and H under the Yucatan Peninsula environment.

In the first part of this thesis, seven temperature based (TET) models and the standardize reference evapotranspiration equation for short canopies (ET_0) method were compared. Using only temperature data, FAO-Penman Monteith Temperature (PMT) model was used to estimate daily values of ET_0 . Also, the ability and precision of SVM, ANFIS and ANN techniques were examined for estimating daily ET_0 using measured meteorological variables. Three different combinations of minimum air temperature, maximum air temperature, rainfall, relative humidity and extraterrestrial radiation as input were investigated with air temperatures and extraterrestrial radiation as the base data set. In a second part, twelve existing empirical models on meteorological parameters –based, four existing day of the year –based (DYB) models, and three SC techniques (i.e., ANFIS, ANN, and SVM) were assessed for H predicting by using measured meteorological variables. In addition, two new models were proposed for H modelling: (1) a model on meteorological parameters based and (2) a model on DYB. A qualitative analysis was performed on the database to find Incorrect or missing weather observations, mainly associated with the malfunction of measuring instruments, and to find weather observations affected by weather systems. The performance of the models in this thesis were evaluated using six different standard statistical measures: root mean squared error (RMSE), mean

bias error (MBE), mean percentage error (MPE), mean absolute percentage error (MAPE), mean bias error (MAE) and coefficient of determination (R^2).

Results for ET_0 modelling showed that the non-calibrated PMT expression using temperatures alone produced the best results. The others seven temperature-based models with and without calibration had poorer performance. The Hargreaves-Samani calibrated and Camargo calibrated models exhibited the best performance of the seven temperature-based models, but neither did as well as the PMT model.

For the SC techniques approach, the results indicate that the SVM technique performed better than ANFIS and ANN approaches. Further, the influence of relative humidity and rainfall on the performance of models were investigated. The analysis revealed that the inclusion of the relative humidity data into the models significantly improves the accuracy of the ET_0 estimates.

As for the H modelling, the main findings were: According to the comparisons between empirical models, it was found that the newly developed empirical model which requires temperature, precipitation and relative humidity input variables obtained the best accuracy. However, if only temperature data are available, the Bristow and Campbell model can be used with good performance. Regarding comparisons between DYB models, it was found that the new proposed model in this thesis estimates daily global solar radiation better than other DYB models. Furthermore, a seasonal analysis shows that the DYB model has good performance in all seasons, including in the rainy season. Finally, among the SC techniques for H modelling, the SVM technique found to give the most accurate results in all evaluated scenarios.

RESUMEN

En la península de Yucatán México, la agricultura de regadío desempeña un papel clave en la producción de cultivos. En 2014, las tierras de regadío representaron alrededor de 88500 hectáreas de la superficie total cultivable. El desarrollo de sistemas de riego que hagan uso eficiente y preciso del agua, son esenciales para la sostenibilidad de los sistemas de producción de cultivos. Así, la evapotranspiración de referencia (ET_0) es uno de los factores más importantes relacionados con el diseño del sistema de riego, la gestión del agua bajo producción de regadío y de secano. Además, el conocimiento preciso de la radiación solar global diaria (H) llegar ser muy importante en el proceso de la ET .

Para hacer frente a este reto, el potencial de varios modelos empíricos y técnicas de computación flexible (SC) denominadas Máquinas de soporte vectorial (SVM), Redes neuronales artificiales (ANN) y Sistemas adaptativos de inferencia neurodifusa (ANFIS) se evaluaron para la estimación de la ET_0 y H bajo las condiciones ambientales de la Península de Yucatán.

En la primera parte de esta tesis, se compararon siete modelos basados en temperatura (TET) y el método de la ecuación estandarizada de la evapotranspiración de referencia (ET_0) para gramíneas. Usando solo datos de temperatura, el modelo FAO-Penman Monteith Temperatura (PMT) se utilizó para estimar valores de diarios de ET_0 . También, la utilidad y precisión de técnicas SVM, ANFIS y ANN fueron examinadas para la estimación diaria de la ET_0 usando variables meteorológicas. Se investigaron tres diferentes combinaciones de temperatura mínima y máxima del aire, lluvia, humedad relativa y radiación solar extraterrestre como datos de entrada, y como datos base la temperatura del aire y radiación solar extraterrestre. En una segunda parte, doce modelos empíricos existentes basados en parámetros meteorológicos, cuatro modelos existentes basados en el día del año (DYB), y tres técnicas basadas en SC (p.ej.: ANFIS, ANN, y SVM) fueron evaluadas para la predicción de H mediante el uso de variables meteorológicas. En adición, dos nuevos modelos fueron propuestos para el modelado de H : (1) un modelo basado en parámetros meteorológicos y (2) un modelo basado en DYB. Se realizó un análisis cualitativo en la base de datos para encontrar las observaciones meteorológicas incorrectas o faltantes, principalmente asociados

con el mal funcionamiento de los instrumentos de medición, y para encontrar observaciones meteorológicas afectadas por perturbaciones meteorológicas. La validación estadística de los modelos en esta tesis se evaluó utilizando seis diferentes medidas estadísticas estándar: raíz del error cuadrático medio (RMSE), error de la desviación media (MBE), error porcentual medio (MPE), error absoluto porcentual medio (MAPE), error absoluto medio (MAE) y coeficiente de determinación (R^2).

Los resultados de la modelización de la ET_0 mostraron que la expresión PMT no calibrada, que sólo emplea los datos de temperaturas, ofreció los mejores resultados. Los otros siete modelos basados en temperatura, tanto con, como sin calibración, obtuvieron un menor rendimiento. Los modelos calibrados de Hargreaves-Samani y Camargo mostraron el mejor resultado de los siete modelos basados en temperatura, pero obtuvieron unos resultados peores a los del modelo PMT.

Para el enfoque de técnicas SC, los resultados indican que la técnica SVM ofreció mejores resultados que las técnicas ANFIS y ANN. Además, se investigó la influencia de la humedad relativa y la lluvia sobre el rendimiento de los modelos. El análisis reveló que la inclusión de la humedad relativa en los modelos mejora significativamente la precisión de las estimaciones de la ET_0 .

En cuanto al modelado de H, los principales hallazgos fueron: De acuerdo a las comparaciones entre los modelos empíricos, se encontró que el modelo empírico recién desarrollado el cual utiliza variables de entrada como la temperatura, precipitación y humedad relativa obtuvo la mejor precisión. Sin embargo, si solo se dispone de datos de temperatura, el modelo de Bristow y Campbell puede ser usado con buenos resultados. En cuanto a las comparaciones entre los modelos DYB, se encontró que el nuevo modelo propuesto en esta tesis estima la radiación solar global diaria mejor que los otros modelos DYB. Además, un análisis estacional muestra que el modelo DYB tiene un buen funcionamiento en todas las estaciones del año, incluyendo la temporada de lluvias. Finalmente, entre las técnicas de SC para el modelado de H, la técnica SVM obtuvo resultados más precisos en todos los escenarios evaluados.

CONTENTS

AGRADECIMIENTOS	I
ABSTRACT	II
RESUMEN	IV
LIST OF TABLES	VIII
LIST OF FIGURES	IX
LIST OF SIMBOLS, ACRONYMS AND ABBREVIATIONS	X
1. INTRODUCTION	1
1.1. Problem Definitions	1
1.2. Research Objectives	3
2. LITERATURE REVIEW	4
2.1. Evapotranspiration	4
2.1.1. Concepts	4
2.1.2. Modelling of reference evapotranspiration.	6
2.2. Solar radiation	9
2.2.1. Components of solar radiation	9
2.2.2. Estimation of extraterrestrial radiation	10
2.2.3. Modelling of Solar radiation	11
2.3. Soft-Computing Techniques	14
2.3.1. Artificial Neural Network (ANN)	14
2.3.2. Support Vector Machines (SVM)	17
2.3.3. Adaptive Neuro-Fuzzy Inference System (ANFIS)	22
3. MATERIALS AND METHODS	27
3.1. Study Area and data	27
3.2. Weather data quality assessment.	28
3.3. Performance evaluation of empirical and Soft-computing models.	30
3.4. Temperature – based Methods for estimating daily reference evapotranspiration.	32
3.4.1. Temperature based methods (TET)	33
3.4.2. Penman – Monteith method using only temperature data (PMT)	35

3.4.3.	Models performance evaluation	37
3.5.	Empirical models for estimating daily global solar radiation	37
3.5.1.	Temperature-based	38
3.5.2.	Meteorological parameter-based	41
3.5.3.	Calibration and performance evaluations.....	43
3.6.	Estimating daily global solar radiation by day of the year.....	44
3.6.1.	Day of the Year Based Models.	45
3.6.2.	Calibration and performance evaluations.....	47
3.7.	Soft-computing methods to estimate Daily global solar radiation and reference evapotranspiration	47
3.7.1.	Data pre-process and software	47
3.7.2.	SVM Model architecture.....	48
3.7.3.	ANFIS Model architecture	49
3.7.4.	ANN Model architecture	50
3.7.5.	Estimation of daily reference evapotranspiration	51
3.7.6.	Estimation of daily global solar radiation	53
3.7.7.	Models performance evaluation	56
4.	RESULTS AND DISCUSSIONS.....	57
4.1.	Temperature – based Methods for estimating daily reference evapotranspiration.	57
4.2.	Empirical models for estimating daily global solar radiation	62
4.3.	Estimating daily global solar radiation by day of the year.....	70
4.3.1.	Overall model performance.....	70
4.3.2.	Seasonal analysis of the proposed model	75
4.4.	Soft – computing models.....	76
4.4.1.	Estimation of daily reference evapotranspiration	76
4.4.2.	Estimation of daily global solar radiation.....	80
5.	CONCLUSIONS.....	85
	REFERENCES.....	89
	APPENDIX A.	103
	APPENDIX B	104

LIST OF TABLES

Table 2.1. Different kernel functions.....	21
Table 2.2. The basic MF's.	23
Table 3.1. Geographical location of the meteorological stations.....	27
Table 3.2. Models and original coefficients used in this study for the estimation of evapotranspiration in Temperature based methods.	35
Table 3.3. A list of models used for estimation of daily global solar radiation...	43
Table 3.4. Features of the ANN models.	52
Table 3.5. Optimal SVM – rbf parameters obtained by the GA-based.	53
Table 3.6. Features of the ANN models.	54
Table 3.7. Optimum SVM parameters obtained by the GA-based.....	55
Table 4.1. Statistical indicators for the ET ₀ estimation in four study stations (the model with the best-fit is shown in bold). The R ² forced to the origin, the MAE, RMSE values expressed in mm day ⁻¹ , and MAPE expressed in %. a and b are the regression coefficients used for calibrated methods.	57
Table 4.2. Error statistics of the thirteen empirical models for the estimation of daily global solar radiation in the six sites of Yucatán, Peninsula.....	63
Table 4.3. Calibrated coefficients for each empirical model.....	66
Table 4.4. Performance metrics of the five empirical models in six selected cities in Yucatán Peninsula, Mexico. The model with the best performance for each location is shown in italics.	70
Table 4.5. Regression coefficients for the selected models in six cities of Yucatán, Peninsula, Mexico.....	71
Table 4.6. RMSE, MBE and MAPE values obtained with Model 5 during the standard meteorological season.	75
Table 4.7. Error statistics during testing phase, for five locations in Yucatán, Peninsula, Mexico.....	77
Table 4.8. The R ² , RMSE and MAE statistics for each model in testing phases, for six locations in Yucatán, Peninsula, Mexico.	82

LIST OF FIGURES

Figure 2.1. A three –layer feed-forward neural network structure.	15
Figure 2.2. Information processing in a neural network unit.	16
Figure 2.3. Illustration of nonlinear SVR with Vapnik’s ξ —insensitivity loss function.	21
Figure 2.4. Basic structure of an ANFIS model.....	23
Figure 2.5. The ANFIS learning algorithm.	26
Figure 3.1. Distribution of the selected stations in Yucatán Peninsula, Mexico.	28
Figure 3.2. Typical variation of daily global solar radiation in Yucatán Peninsula (triangles), and curves generated by the single Gaussian function and sum of two Gaussian correlation formulas.	46
Figure 4.1. The MAE and R^2 comparisons with the PM estimated ET_0 by model.	60
Figure 4.2. Scatter plots of the measured and estimated solar radiation in six sites in Yucatán, Peninsula, Mexico for model M5.....	69
Figure 4.3. Comparison between daily measured global solar radiation data (measured and daily mean of long-term measured data) and values estimated by proposed Model 5 for six locations on the Yucatán Peninsula.	74
Figure 4.4. The FAO56-PM ET_0 and estimated ET_0 values obtained by the best model in each scenario during testing phase for Calakmul, Campeche and Efraín Hernández stations.....	79
Figure 4.5. The FAO56-PM ET_0 and estimated ET_0 values obtained by the best model in each scenario during testing phase for Mérida and Tantakín stations.	80
Figure 4.6. Regression analysis plot for the SVM – rbf 1 model between measured and estimated of daily global solar radiation.	84

LIST OF SIMBOLS, ACRONYMS AND ABBREVIATIONS

$\{a, b, c\}$	ANFIS premise parameters or bell – shaped parameters
$\{\gamma, r, d\}$	SVM kernel parameters
$\{\varepsilon, C\}$	SVM parameters to be optimized
$\{p_i, q_i, r_i\}$	ANFIS consequent parameters
$[C, D]$	Predefined boundary for the normalized data
a - g	Regression coefficients
A_i, B_i	ANFIS linguistic labels
ASCE	American Society of Civil Engineers
AND/OR	Fuzzy rule operator
ANFIS	Adaptive Neuro – Fuzzy Inference system
ANN	Artificial Neural Network
A_w	Tropical savanna climate
b	Bias values
BP	Back – propagation algorithm
C	SVM cost factor or empirical error
CV	Cross validation
D	Daily averaged saturation deficit (kPa)
DM	Data mining
DYB	Day of the year – based
D_t	Function correcting the effect of site differences in day length
e	A vector of network errors
e_a	Actual vapor pressure (kPa)
$e_s(T_{\max})$	Saturation vapor pressure at temperature $T_{\max}$ (kPa)
$e_s(T_{\min})$	Saturation vapor pressure at temperature $T_{\min}$ (kPa)
e_s	Saturation vapor pressure (kPa)
$e_s - e_a$	Vapor pressure deficit (kPa)
E_0	Eccentricity correction factor of the Earth's orbit
ELM	Extreme Learning Machine
ET	Evapotranspiration (mm day^{-1})
ET_0	Reference evapotranspiration (FAO 56 PM) (mm day^{-1})
ET_c	Crop evapotranspiration (mm day^{-1})

ET_{CA}	Camargo evapotranspiration mm day^{-1})
ET_{HA}	Hamon evapotranspiration (mm day^{-1})
ET_{HS}	Hargreaves-Samani evapotranspiration (mm day^{-1})
ET_{MA}	Malmström evapotranspiration (mm day^{-1})
ET_{OU}	Oudin evapotranspiration (mm day^{-1})
ET_{PA}	Papadakis evapotranspiration (mm day^{-1})
ET_{TH}	Thornthwaite evapotranspiration (mm day^{-1})
EWRI	Environmental & Water Resources Institute
f	ANN transfer function
f_i	ANFIS linear function
$f(T_{avg})$	Function based on the daily mean temperature ($^{\circ}\text{C}$)
FAO56-PM	FAO-56 Penman-Monteith evapotranspiration (mm day^{-1})
FAO	Food and Agriculture Organization
FFA	Fire Fly Algorithm
FFNN	Feed – forward neural networks
FIS	Fuzzy inference system
FL	Fuzzy logic
G	Soil heat flux density ($\text{MJ m}^{-2} \text{day}^{-1}$)
GA	Genetic Algorithm
GDA	Gradient descent algorithm
GEP	Gene Expression Programming
GP	Genetic Programming
H	Global solar radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
HS	Hargreaves – Samani ET_0 formulation
H_b	Direct beam radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
H_d	Diffuse radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
H_{day}	Half-day length (radians)
H_0	Extraterrestrial radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
i	Monthly heat index
I_{sc}	Solar constant ($118.108 \text{ MJ m}^{-2} \text{day}^{-1}$)
I	Annual heat index
ICSKY	Corrected clear-sky solar irradiation
Id	Identity matrix

INIFAP	Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias
IQR	Interquartile – range
I_s	Clear-day solar radiation ($\text{MJ m}^{-2} \text{ day}^{-1}$)
J	Jacobian matrix
$j-1, j, j+1$	Previous, current and next day
K	SVM kernel function
K_c	Crop coefficient
K_{HS}	Empirical radiation adjustment coefficient $^{\circ}\text{C}^{-0.5}$
K_T	Clearness index
LD	Longest day of year (h)
L_{ε}	SVM loss function
LM	Levenberg – Marquardt
Logsig	Logistic sigmoid transfer function
LSA	Least square algorithm
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MBE	Mean bias error
MF	Membership function
MM	Minimum and maximum normalization technique
MLP	Multi – Layer Perceptron
MPE	Mean percentage error
n	Number of observation of a data set
N	Maximum number of sunshine hours of the month (h day^{-1})
N'	Number of data sets in a SVM regression
NT	Number of training data
N_d	Number of days in the month (Day)
n_{day}	Day of year, starting 1 January
O_i	Measured response variable
O_m	Average of measured values
P	Precipitation (mm)
P_o	Standard pressure (kPa)
P_a	Atmospheric pressure (kPa)

P_i	Predicted response variable
P_m	Average of predicted values
PCHIP	Piecewise Cubic Hermite Interpolating Polynomial
PET	Potential evapotranspiration (mm day^{-1})
PMT	Penman-Monteith temperature equation (mm day^{-1})
PMT_{rad}	Radiation term of the PMT
PMT_{aero}	Aerodynamic term of the PMT
Pureline	Linear transfer function
Q1	The 25 th percentile
Q3	The 75 th percentile
$R(C)$	SVM regularized risk function
R^2	Coefficient of Determination
RBF	Radial basis function
RH	Daily average relative humidity (%)
RMSE	Root mean square error
R_n	Net radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
RP	Resilient propagation
R_s	Solar radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
R_{ns}	Net shortwave radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
R_{nl}	Net longwave radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
RT	Transformed rainfall, using $P>0$, $RT=1$; $P=0$, $RT=0$
SIAP	Servicio de Información Agroalimentaria de México
SMN	Servicio Meteorológico Nacional
SMO	Sequential minimal optimization
SC	Soft computing
SSE	Sum – squared error
SVM	Support Vector Machines
SVM - rbf	Support Vector machine radial basis
SVR	Support vector regression
$^{\circ}C$	Transmissibility coefficient for the Mahmood-Hubbard model
T	Temperature
T_{ef}	Effective temperature ($^{\circ}C$)
TET	Temperature based ET model

$T_{\max}$	Maximum air temperature (°C)
T_{mean}	Mean temperature (°C)
$T_{\min}$	Minimum air temperature (°C)
u_2	Mean wind speed at 2 m height (m s^{-1})
w	ANN synaptic weight
$\overline{w_0}$	ANFIS normalized firing strengths
w_0	ANFIS firing strengths
WT	Wavelet transform
X'	Normalized data
$X_{\min}$	Minimum value to be normalized
$X_{\max}$	Maximum value to be normalized
z	Site elevation (m)

Greek letters

α, α^*	SVM Lagrange multipliers
δ	Solar declination (degrees)
Δ	Slope of the saturation vapor pressure ($\text{kPa } ^\circ\text{C}^{-1}$)
ΔT	Daily temperature difference (°C)
λ	Latent heat of vaporization (MJ kg^{-1})
ε	SVM insensitive region
ξ, ξ^*	SVM slack variables
$\mu(x)$	ANFIS membership
π	Number Pi
ρ	Psychrometric constant ($\text{kPa } ^\circ\text{C}^{-1}$)
τ_a	Transmittance affected by atmospheric aerosols and ozone
$\tau_{c f}$	Atmospheric transmittance of the cloud-free atmosphere
τ_o	Transmittance of clean dry air
τ_v	Transmittance affected by atmospheric water vapor
Γ	Day angle (radians)
σ	Stefan – Boltzmann constant
φ	SVM nonlinear mapping function
ϕ	Site latitude (degrees)
ω	SVM weight vector
$\ \omega\ ^2$	SVM norm flatness
ω_s	Sunset hour angle (degrees)

1. INTRODUCTION

1.1. Problem Definitions

In the Yucatán Peninsula, irrigated agriculture plays a central role in the production of food and raw materials. Most of this production would not be possible without irrigation. In 2014, the SIAP (Servicio de Información Agroalimentaria de México) estimated that in the Yucatán Peninsula, the agricultural irrigated land was 88877 ha. In precision agriculture, deciding when and how much water to apply to a field has a significant impact on the total amount of water used by the crop water use efficiency and irrigation efficiency. Proper irrigation management is a key to ensure healthy, high quality crops while protecting valuable water resources. Thus, the appropriate knowledge of evapotranspiration (ET) is necessary for the crop water requirements estimations, management of the water resources, and in the irrigation scheduling. To determine crop irrigation requirements, it is necessary to estimate ET by on site measurements or by using climatological data. On site measurements are very costly and are mostly employed to calibrate ET methods using climatological data. The crop water use is directly related to ET, and this can be determined by multiplying the reference evapotranspiration (ET_0) by a crop coefficient (K_c) value to arrive at a crop evapotranspiration (ET_c) estimate. The resulting ET_c is used to help and irrigation manager schedule when an irrigation should occur and how much water should be put back into the soil. Numerous equations, classified as temperature-based, radiation-based, pan evaporation-based and combination-type, have been developed for ET_0 estimating. They vary in terms of data requirement and accuracy. Recently, the adapted FAO-56 Penman-Monteith equation (FAO56-PM) has been recommended as the standard equation for estimating ET_0 and calibrations other ET_0 equations (Allen *et al.*, 1998; Baba *et al.*, 2013). The FAO56-PM equation requires measurements of air temperature, relative humidity, solar radiation and wind speed. However, parameters such as solar radiation, relative humidity and wind speed are not easily available in most of the weather stations, usually due to the high cost of installation and maintenance of measuring equipment. In fact, in the Yucatán Peninsula, the ratio

of stations measuring solar radiation and wind speed to those measuring temperature and precipitation is 1:160. Moreover, most of these stations have erroneous measurements or missing data due to the lack of maintenance or calibration of sensors (Hunt *et al.*, 1998; Iziomon and Mayer, 2002). In this context, empirical equations that using less climatological data must be used. Nevertheless, these empirical equations are not adequate for all locations. Local calibrations are always necessary to get obtain reliable and good estimates of the crop water requirements. In addition, the majority of empirical equations are based on air temperature. The different methods of estimating ET_0 approaches reviewed generally performed better when solar radiation was included as input variable. However, this data is often not available. Thus, is necessary a method have accurate and reliable solar radiation data, especially in a rainy and humid area as is the Yucatán Peninsula. This method should be based on data of air temperature, rainfall and relative humidity. These parameters can be easily acquired by producers or agronomists. Otherwise, using the day of the year as a single input parameter. Moreover, the knowledge of solar radiation has other applications, including crop growth and yield simulation, and scientific applications (e.g., photovoltaic systems).

1.2. Research Objectives

Apply seven temperature based (TET) models for estimating daily reference evapotranspiration.

Test the Penman –Monteith temperature equation (PMT) for modelling daily reference evapotranspiration.

Investigate the capability of three soft-computing techniques for modelling daily reference evapotranspiration.

Evaluate the accuracy and applicability of eight empirical models to estimate daily global solar radiation.

Develop a new model that utilizes temperature, rainfall and mean relative humidity for estimating daily global solar radiation.

Examine four days of the year based-models (DYB) for their suitability for estimating daily global solar radiation.

Proposed a new day of the year based –model for generating daily global solar radiation data.

Evaluate the accuracy and performance of three Soft-computing techniques using measured meteorological variables for estimating daily global solar radiation.

These objectives are examined at several specific sites in Yucatán, Peninsula, Mexico, and are assumed to be applicable to at least similar agro–environmental conditions.

2. LITERATURE REVIEW

2.1. Evapotranspiration

2.1.1. Concepts

2.1.1.1. Evapotranspiration

Evapotranspiration (ET) is the combination of two separate processes: evaporation and transpiration. Evaporation is a physical process that involves the conversion of liquid water into water vapor into the atmosphere. Evaporation of water into the atmosphere occurs on the surface of rivers, lakes, soils and vegetation. Transpiration is basically a process of evaporation. The transpiration is a physical process that involves the flow of liquid water from the soil (root zone) to the surface of leaves/ branches and trunk; and conversion of liquid water from the plant tissue into water vapors into the atmosphere. The water evaporates from the leaves and plant tissue, and the resultant water vapor diffuses into the atmosphere through the stomates. An energy gradient is created during the evaporation of water, which causes the water movement into and out of the plant stomates. In the majority of green plants, stomates remain open during the day and stay closed during the night. If the soil is too dry, the stomates will remain closed during the day in order to slow down the transpiration.

2.1.1.2. Potential evapotranspiration

Potential evapotranspiration (PET) is a water loss from the soil surface completely covered by vegetation. Meteorological processes determine the evapotranspiration of a crop (Goyal and Harmsen, 2013). The closing of stomates and reduction in transpiration are usually important only under drought or under stress conditions of a plant. The evapotranspiration depends on three factors: (1)

Vegetation, (2) Water availability in the soil and (3) Behavior of stomates. Vegetation affects the ET in various forms. It affects the ability of soil surface to reflect light. The vegetation changes the amount of absorbed energy by the soil surface. Soil properties, including soil moisture, also affect the amount of energy that flows through the soil.

2.1.1.3. Reference evapotranspiration

The term reference evapotranspiration (ET_0) was introduced by the United Nations Food and Agriculture Organization (FAO) as a methodology for calculating crop evapotranspiration (ET_c) (Doorenbos and Pruitt, 1977). ET_0 is "the rate at which water, if available, would be removed from the soil and plant surface of a specific crop, arbitrarily called a reference crop" (Jensen and Allen, 1990). The reference crop is typically grass or alfalfa under well-watered conditions (height 0.12 m, surface resistance 70 s m^{-1} and albedo 0.23) (Allen *et al.*, 1998).

Thus, due to its comprehensive theoretical base, the Penman- Monteith (FAO56-PM) equation (equation 2.1) (Allen *et al.*, 1998) is recommended by the United Nations Food and Agriculture (FAO) as the sole method to calculate ET_0 and for evaluating other ET_0 calculation method as well. The FAO approach to calculating ET_0 using the PM equation was published in the FAO irrigation and drainage paper number 56 (FAO-56).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \rho \frac{900}{T_{mean} + 273} u_2 (e_s - e_a)}{\Delta + \rho(1 + 0.34u_2)} \quad (2.1)$$

where ET_0 is reference evapotranspiration (mm day^{-1}), Δ is the slope of the saturation vapor pressure ($\text{kPa } ^\circ\text{C}^{-1}$), ρ is the psychrometric constant ($\text{kPa } ^\circ\text{C}^{-1}$), R_n is net radiation at the crop surface ($\text{MJ m}^{-2} \text{ day}^{-1}$), G is soil heat flux density ($\text{MJ m}^{-2} \text{ day}^{-1}$), as the magnitude of the day soil flux is small, may be ignored, T_{mean} is mean daily air temperature ($^\circ\text{C}$), u_2 is average wind speed at 2 m height

(m s^{-1}), e_s is saturation vapor pressure (kPa), e_a actual vapor pressure (kPa), $e_s - e_a$ is saturation vapor pressure deficit (kPa). The computation of all data required for calculating ET_0 followed the method and procedure given in chapter 3 of FAO-56 (Allen *et al.*, 1998).

2.1.2. Modelling of reference evapotranspiration.

A lysimeter can be used to measure ET_c from a well-watered agricultural crop directly and accurately, however its wide application is restricted by costly and time-consuming (Wang *et al.*, 2014). The most common approach used for estimating ET_c is the crop coefficient (K_c) approach, which consists of multiplying ET_0 with the K_c (Allen *et al.*, 1998). Recently, the adapted FAO-56 Penman-Monteith equation (FAO56 –PM) has been recommended as the standard equation for estimating ET_0 and calibrations other ET_0 equations (Allen *et al.*, 1998; Kisi, 2013). The FAO56 –PM can be used in a great variety of climate conditions, at different time steps and need no local calibration because of its physical basis. ET_0 is considered as an incidental, nonlinear, complex and unsteady process so it is difficult to derive an accurate formula to represent all the physical processes involved. The main shortcoming of FAO56 –PM methodology is that it requires weather variables that are commonly lacking, especially in developing countries where reliable meteorological data of solar radiation, relative humidity and wind speed are limited (Gocic and Trajkovic, 2010; Tabari *et al.*, 2013). In these situations, alternative equations with fewer meteorological parameters requirements are the only alternative. So, Numerous equations, classified as temperature-based (Thornthwaite, 1948; Blaney and Criddle, 1950; Hamon, 1963; Papadakis, 1966; Malmström, 1969; Hargreaves and Samani, 1982; Camargo *et al.*, 1999; Hargreaves and Allen, 2003; Oudin *et al.*, 2005), radiation-based (Jensen and Haise, 1963; McGuinness and Bordne, 1972; Jones and Ritchie, 1990; Irmak *et al.*, 2003), and combination-type (Romanenko, 1961; Brockamp and Wenner, 1963; Mahringer, 1970) have been developed for estimating ET_0 , they vary in terms of data requirement and accuracy. However, in most cases, only the maximum and minimum air temperature are available (Allen *et al.*, 1998). For this reason, Temperature-

based evapotranspiration (TET) methods are particularly interesting. Air temperature is one of the most widely meteorological parameters and those data are readily available from most of agroclimatic weather station (Mendicino and Senatore, 2013). Therefore, TET methods and temperature databases are critically important for ET estimation in poor areas of the world.

Allen *et al.* (1998) indicated that, when solar radiation, relative humidity and/or wind speed data are missing, ET_0 should be estimated using Hargreaves – Samani (HS) ET_0 formulation (Hargreaves and Samani, 1985). The HS ET_0 method was tested extensively by the ASCE-EWRI (Allen *et al.*, 2006), and they found that it often provided good estimates of ET_0 in arid environments. However, since it does not account of wind speed and humidity, it is known to somewhat in accurate in climates with high humidity and extremely high or low wind speeds. As pointed out by Almorox *et al.* (2015) and Almorox and Grieser (2015) equations like the HS ET_0 equation often require empirical calibration coefficients to account for climate variation.

Considerable information on TET models for estimating ET_0 is reported in the worldwide literature. Many studies have evaluated for the estimation of the FAO ET_0 using only minimum and maximum temperature data (PMT). Annandale *et al.* (2002) in South Africa, Trajkovic (2005) in Serbia, Popova *et al.* (2006) in south Bulgaria, Jabloun and Sahli (2008) in Tunisian, Cai *et al.* (2009) in the North of China, López-Moreno *et al.* (2009) in Spanish Pyrinee, Martinez and Thepadia (2010) in Florida (USA), Gocic and Trajkovic (2010) in Davis (California, USA), Kra (2010) in four countries of west Africa, Raziei and Pereira (2003) in Iran, Todorovic *et al.* (2013) in 16 mediterranean countries, Vangelis *et al.* (2013) in Greece, Pandey and Pandey (2016) in India and Ren *et al.* (2016) in inner Mongolia (China), among others. In this study, we used Hargreaves and Samani formulation HS (Hargreaves and Samani, 1982; Hargreaves and Samani 1985; Hargreaves and Allen 2003) and the PMT formulation to estimate ET_0 because they require only air temperature data. The PMT formulation produces acceptable results using only temperature data. It is based on the Penman Monteith equation and retains many features of the combination model of Penman, considering a combination of net radiation and aerodynamic principles (Pereira *et al.*, 2015).

The performance of the HS and PMT equations depend on the range of climate, and some previous studies have been reported better performance of PMT model over HS formulation. Vangelis *et al.* (2013) observed that PMT and HS formulation both provide better accuracy when compared to Thornthwaite and Blaney-Criddle models in arid and semiarid areas. Todorovic *et al.* (2013) showed that, in Mediterranean moist sub-humid areas, the best performance was obtained by PMT method. The same results were obtained by López-Moreno *et al.* (2009) in moist sub-humid areas in Spain and Trajkovic (2005) in Serbia. Pandey and Pandey (2016) reported HS method had larger overestimation compared to PMT against standard FAO-56 in humid areas of India. Both approaches were reported as not appropriate to detect trends in PM when solar radiation and wind speed are the major driving variables.

Moreover, evapotranspiration processes can also be modeled with high accuracy using soft – computing techniques such as Artificial Neural Network (ANNs), Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Machines (SVM) and Gene Expression Programming (GEP), Genetic Programming (GP), Extreme Learning Machine (ELM). Soft computing methods is an alternative and emerging method, can be used as an innovative approach because they offer benefits such as no required knowledge of internal variables, simpler solutions for multi – variable problems and accurate calculation (Gocić *et al.*, 2015). In several studies, the accuracy of these techniques has been improved by using algorithms [e.g. Fire Fly Algorithm (FFA) and Genetic Algorithm (GE)] to tune the model's parameter and use of wavelet transform (WT) techniques to transform data series into sub-series. Tabari *et al.* (2012) examined the accuracy of SVM and ANFIS techniques in estimating ET_0 in a semi-arid environment in Iran and compared them with empirical equations. They found that the SVM and ANFIS models provide further accuracy compared to the empirical equations. Baba *et al.* (2013) compared ANFIS and ANN techniques for ET_0 modelling in two weather stations in South Korea. Their results indicated that the two techniques performed quite well in modelling ET_0 . Citakoglu *et al.* (2014) applied ANFIS and ANN to estimate ET_0 in Turkey and found that the both techniques can be employed successfully to predict ET_0 . Shiri *et al.* (2014) compared in Iran the performance of ANFIS, GEP, ANN and SVM for ET_0

modelling in 29 weather stations. The results showed the GEP based models obtained better performances than the others. Wen *et al.* (2015) assessed the potential of SVM to model ET_0 in an extremely arid region of China and compared it with ANN technique and three empirical models. They found that SVM technique was the best among these models. Gocić *et al.* (2015) evaluated the capability of GP, ANN, SVM –FFA, and SVM –WT techniques for ET_0 estimation in 12 locations in Serbia. Their results indicated that the SVM –WT model is more accurate in predicting ET_0 compared to the other models. Falamarzi *et al.* (2014) employed ANN and ANN –WT to predict ET_0 in Redesdale, Australia. The results showed that the ANN –WT technique were more accurate than the ANN approach. Also, it has been shown that computational models have much better performances than the empirical models.

2.2. Solar radiation

2.2.1. Components of solar radiation

Solar radiation consists of electromagnetic radiation emitted by the sun in spectral regions ranging from X –rays to radio waves. For terrestrial applications of renewable energy utilizing solar radiation usually based on radiation, or photons, referred to as “optical solar radiation”, with a spectral range of about 300–4000 nm, from the ultra violet wavelengths to the near and mid infrared wavelengths that Earth’s atmosphere allows to reach the ground (Badescu, 2014).

Solar radiation from the sun after traveling in space enters the atmosphere at the space-atmosphere interface, where the ionization layer of the atmosphere ends. Afterwards, a certain amount of solar radiation of photons is absorbed by the atmosphere, clouds, and particles in the atmosphere, a certain amount is reflected back into the space, and a certain amount is absorbed by the earth’s surface. The earth’s surface also reflects a certain amount of energy by radiation at different wavelengths due to the earth’s surface temperature.

After the solar radiation arrives at the earth’s atmosphere, it is partially scattered and partially absorbed. The scattered radiation is called diffuse

radiation (H_d). On the other hand, the solar radiation that travels in a straight line from the sun to the earth's surface is called direct radiation or direct beam radiation (H_b) (Sen, 2008). Then, the sum of the direct and diffuse solar radiation is termed global solar radiation (H). In a simple equation, the three types of radiation are expressed as follows:

$$H = H_b + H_d \quad (2.2)$$

Global solar radiation at a location is roughly proportional to direct solar radiation, and varies with the geometry of the receiving surface. The other components, such as diffuse radiation, vary slightly from slope to slope within a small area and the variations can be linked to slope gradient.

2.2.2. Estimation of extraterrestrial radiation.

The amount of solar radiation received on a unit horizontal surface area per unit time at the top of the atmosphere is known as the extraterrestrial radiation (H_0), and is calculated from a solar constant, an eccentricity correction factor of the Earth's orbit (E_0), site latitude (ϕ), day of the year (n_{day}), and solar angle (δ), using standard geometric procedures according to equations 2.3 to 2.7 (Spencer, 1971; Almorox *et al*, 2011):

$$H_0 = (1/\pi)I_{sc}E_0(\cos \phi \cos \delta \sin \omega_s + (\pi/180) \sin \phi \sin \delta \omega_s) \quad (2.3)$$

$$E_0 = 1.00011 + 0.034221 \cos \Gamma + 0.00128 \sin \Gamma + 0.000719 \cos(2\Gamma) + 0.000077 \sin(2\Gamma) \quad (2.4)$$

$$\delta = (180/\pi) \cdot (0.006918 - 0.399912 \cos \Gamma + 0.070257 \sin \Gamma - 0.006758 \cos 2\Gamma + 0.000907 \sin 2\Gamma - 0.002697 \cos 3\Gamma + 0.00148 \sin 3\Gamma) \quad (2.5)$$

$$\Gamma = 2\pi(n_{\text{day}} - 1)/365 \quad (2.6)$$

$$\omega_s = \cos^{-1} [(-\sin \phi \sin \delta)/(\cos \phi \cos \delta)] \quad (2.7)$$

2.2.3. Modelling of Solar radiation

Global solar radiation (H) is an important variable for a wide range of applications in areas such as meteorology, climatology, hydrology, crop simulation models and estimation of crop evapotranspiration (ET_c). H is the most important source of renewable energy on the planet. Reliable estimation of H is of fundamental importance for applications such as efficient determination of irrigation water needs and potential yield of crops (El-Sebaili *et al.*, 2010; Almorox *et al.*, 2011; Boukelia *et al.*, 2014)

Researchers have developed a large number of methods to estimate H from other available meteorological variables. These methods are based on empirical modelling related to variables such as temperature (Hargreaves and Samani, 1982; Bristow and Campbell, 1984; Donatelli and Campbell, 1998; Goodin *et al.*, 1999; Mahmood and Hubbard, 2002; Benghanem and Mellit, 2014), sunshine duration (Angstrom, 1924; Prescott, 1940; Al-Mostafa *et al.*, 2014; Despotovic *et al.*, 2015; Lockart *et al.*, 2015; Teke *et al.*, 2015), rainfall (McCaskill, 1990; Liu and Scott, 2001; Chen *et al.*, 2006; Wu *et al.*, 2007; Kirmani *et al.*, 2015), cloudiness (Supit and van Kappel, 1998; Gu *et al.*, 2001; Badescu, 2002), elevation (Allen, 1995; Winslow *et al.*, 2001; Annandale *et al.*, 2002), humidity (Gu *et al.*, 2001; Chen *et al.*, 2006; Almorox *et al.*, 2011; Li *et al.*, 2013; Antonanzas-Torres *et al.*, 2013; Li *et al.*, 2015) and latitude (Toğrul and Onat, 1999; Skeiker, 2006). Several studies have shown that the best performing models use sunshine duration data, followed by those utilizing cloudiness and then temperature data (Bakirci, 2009; Chen and Li, 2013). However, sunshine and cloudiness data are not available at most meteorological stations. Thus, if temperature-based models can be used with sufficient reliability, solar radiation estimates could be available at more locations provided the empirical parameters are calibrated for each specific location (Almorox *et al.*, 2011). Also, several studies have demonstrated that adding rainfall data to temperature-based models results in increased performance (Liu & Scott, 2001; Chen *et al.*, 2006; Wu *et al.*, 2007; Woli and Paz, 2012). In addition to predicting global solar radiation as accurately as possible, such empirical approaches should also have uncomplicated functional forms with limited and readily available inputs.

Recently, some studies have been aimed at developing and establishing simple models to estimate global solar radiation without meteorological data using the day of the year as the sole independent parameter, hereinafter referred as day of the year –based (DYB) models (Khorasanizadeh *et al.*, 2014).

Al-Salaymeh (2006) proposed four correlation formulas of DYB models to predict daily global solar radiation on horizontal surfaces of Amman City in Jordan, with the best results obtained from a sine wave model. Bulut and Büyükalaca (2007) developed a DYB model for simulating daily global solar radiation in Istanbul, Turkey by using long-term data with a sine wave formulation. The model was tested for 68 locations in Turkey and the results had good fit for the measured data. Kaplanis and Kaplani (2007) developed a cosine wave DYB model to estimate daily global solar radiation in six climate zones of Greece. They found that the model predicts values of daily global solar radiation for those climate zones with high accuracy, with correlation coefficients for all cases being higher than 0.996. Li *et al.* (2010) proposed a new DYB equation using a hybrid sine and cosine wave formula to estimate daily global solar radiation in China. The predictions were compared with three DYB existing models. Statistical results indicated that the new method provides better estimation and has good adaptability to highly variable weather conditions. Results from this model were site dependent with a mean correlation coefficient of 0.937. This model can be used for estimating daily values of global solar radiation with higher accuracy. Zang *et al.* (2012) developed a new hybrid sine and cosine wave DYB formulation that performed best for estimation of global solar radiation for six climatic zones of China. Khorasanizadeh and Mohammadi (2013) conducted a study to estimate daily global solar radiation using DYB models in four cities situated in sunny regions of Iran and found that the hybrid sine and cosine wave and the 4th order polynomial models had better performance. The authors conclude that because DYB models are not dependent on any meteorological data, they can be utilized to estimate daily global solar radiation in regions where meteorological data do not exist. Khorasanizadeh *et al.* (2014) conducted a study in Birjand, Iran to compare DYB models and empirical temperature and sunshine duration based equations in the estimation of solar radiation. In general, their results revealed that DYB models can be easily utilized for estimating solar radiation in the study area and its neighboring regions with similar climates.

Solar radiation can be estimated by using soft – computing techniques. These set of techniques are within the framework of artificial intelligence that has received much attention for dealing with practical problems (Gopalakrishnan *et al.*, 2011). Soft – computing includes artificial neural networks (ANN), genetic algorithms (GAs), fuzzy logic (FL), adaptive neuro fuzzy inference system (ANFIS), support vector machines (SVM) and data mining (DM). These methods offer advantages over conventional modelling, including the ability to handle large amounts of noisy data from dynamic and nonlinear systems, especially when the underlying physical processes are not fully understood (He *et al.*, 2014). Several soft – computing techniques have been used in recent years to estimate global solar radiation, where ANFIS and ANN are the most popular. Mohammadi *et al.* (2015) applied ANFIS and SVM techniques to predict global solar radiation based on air temperatures in Bandar Abbas, located in the southern of Iran. The results showed that the SVM models outperform the ANFIS. Chen and Li (2014) investigated the ability of different SVM models in global solar radiation modelling for 15 synoptic stations with different climate conditions located in China. The results conclude that the SVM models could be used successfully in modelling global solar radiation. Piri *et al.* (2015) performed a comparative investigation between four sunshine durations based on empirical models and SVM models to estimate global solar radiation in two cities (Zahedan and Bojnurd) of Iran. Their results indicated that all SVM models outperform the empirical models. Olatomiwa *et al.* (2015b) compared the accuracy of SVM with ANFIS models and one empirical model for global solar radiation modelling in a semi-arid environment in Nigeria. They found that the SVM models gave better performance to ANFIS and the empirical model. Ramedani *et al.* (2014) presented a radial basis SVM (SVM - rbf) model to predict global solar radiation over Teheran, Iran. The SVM - rbf prediction was compared with ANFIS and ANN models. The results demonstrated that the SVM - rbf have higher performance over the other models. Olatomiwa *et al.* (2015a) developed a novel method using meteorological data of three different regions of Nigeria, by hybridizing the Support Vector Machines (SVMs) with Firefly Algorithm (FFA) to predict the monthly average global solar radiation using sunshine hours, maximum temperature and minimum temperature as inputs. The results indicate that the hybrid model can be used as an efficient machine learning technique for accurate

estimation of global solar radiation. Şenkal (2015) modeled global solar radiation using measured data of five stations in Turkey. He used an ANN technique with resilient propagation (RP) in the training phase with geographical and meteorological data as the input variables. The ANN technique had better results compared to other algorithms and input variables. Kumar *et al.* (2015) compared several models based on regression and ANN models to predict global solar radiation. In general, the ANN models had better results than the regression techniques. Sumithira *et al.* (2012) conducted a comparative study between ANFIS and other soft computing models techniques for estimating global solar radiation in Tamilnadu, India. The ANFIS model had promising results when compared to other models in the literature.

2.3. Soft-Computing Techniques

Soft – computing techniques offered many popular data-driven models which have been used extensively in the past couple of decades in different aspects, including solar radiation and evapotranspiration modelling. Over the last few decades, Soft – computing based techniques have been introduced and widely applied in hydrological studies as powerful alternative modelling tools, such as ANN, ANFIS, and SVM models.

2.3.1. Artificial Neural Network (ANN)

The story of ANNs started in the early 1940s when McCulloch and Pitts (McCulloch and Pitts, 1943) developed the first computational representation of a neuron. Later, Rosenblatt (1962) proposed the idea of perceptron to solve problems (linearly separable classes) in the area of character recognition, who used a single layer feed-forward networks of McCulloch– Pitts. The applications of ANNs are based on their ability to mimic the human mental and neural structure to construct a good approximation of functional relationships between past and future values of a time series. The supervised one is the most commonly used ANNs, in which the input is presented to the network along with the desired output, and the weights are adjusted so that the network attempts to produce the

desired output. An ANN is capable of identifying complex nonlinear relationships between input and output datasets.

To solve non-linearly separable problems, additional layer(s) of neurons placed between the input layer and the output layer are needed leading to “Multi –Layer Perceptron” (MLP) architecture. These intermediate layers do not interact with the external environment, they are called hidden layers. The MLP is the most common architecture of feed-forward neural networks (FFNN). A three –layer feed –forward neural network (one input layer, one hidden layer, and one output layer) is the most commonly used topology in hydrology (Yadav and Chandel, 2014; Qazi *et al.*, 2015; Rezrazi *et al.*, 2015), as shown in figure 2.1. This topology has proved its ability in modelling many real –world functional problems.

Figure 2.1. A three –layer feed-forward neural network structure.

In the MLP each of the various inputs to the network is multiplied by a connection weight. These products are simply summed and pass through a transfer function or activation function (f) to generate a result, as described in equation 2.8.

$$y_j = f\left(\sum_{i=1}^n w_{ji} x_i\right) \quad (2.8)$$

where y is a result, x is an input signal, w is the synaptic weight in a connection

and f is the transfer function that can be a simple threshold function or a sigmoidal, hyperbolic tangent or radial basis function. Figure 2.2 illustrates this procedure. The activation function acts as a squashing function, such that the output of a neuron in a neural network is between certain values (usually 0 and 1, or -1 and 1).

Figure 2.2. Information processing in a neural network unit.

One popular activation function used by the hidden neurons is the basis logistic sigmoid function (Logsig) shown in equation 2.9; and that by the output neurons a linear function (purelin) shown in equation 2.10. The logsig function possesses the distinctive properties of continuity and differentiability on $(-\infty, \infty)$.

$$f(w) = 1/(1 + e^{-w}) \quad (2.9)$$

where w is the weighted sum of the input.

$$f(x) = x \quad (2.10)$$

and x is the input to the output layer.

The procedure for updating the synaptic weights is called back-propagation (BP). BP refers to the way error computed at the output side is propagated backward from the output to the hidden layer(s), and finally to the input layer (Esmaeelzadeh *et al.*, 2014). The error is minimized across many training cycles

called epochs. During each cycle, the network reaches a specified level of accuracy. Generally, the error estimator used here is the sum of the squared error (SSE). In conjunction with the BP procedure, the following algorithms can be used as a second training algorithm: Gradient Descent [Gradient Descent back-propagation algorithm (traingd), Gradient Descent with Momentum (traingdm), Resilience back-propagation (trainrp)], Conjugate Gradient algorithms [Scaled conjugate Gradient (trainscg), Conjugate Gradient back-propagation with Fletcher-Reeves Updates (traincgf), Conjugate Gradient back-propagation with Polak-Ribere Updates (traincgp)], and Quasi-Newton algorithms [Broyden-Fletcher-Goldfarb-Shanno (trainbfg), Levenberg–Marquardt back-propagation (trainlm)]. Selection of an appropriate training algorithm, transfer function and number of neurons in the hidden layer are fundamental characteristics of the ANN model. Each training algorithm has its own characteristics that must be adjusted according to a particular model.

2.3.2. Support Vector Machines (SVM)

A support vector machine (SVM) was introduced by Vapnik (2013) and is a supervised learning technique from the field of machine learning theory and structural risk minimization, applicable to both classification and regression. SVMs are essentially a sub-discipline of machine learning. SVMs are derived from the concept of structural risk minimization hypothesis to minimize both empirical risk and the confidence interval of the learning machine, which in turn helps to produce good generalization capability.

In addition to their solid mathematical foundation in statistical learning theory, SVMs have demonstrated highly competitive performance in numerous real-world applications. Initially developed for solving classification problems, SVM techniques can also be successfully applied in regression problems, usually called SVR (support vector regression). A regression is estimated by using SVM for a data set $\{(x_i, y_i)\}_{i=1}^{N'}$ where x_i is the input vector, y_i is the output value and N' is the total number of data sets by mapping x into a feature space via a nonlinear function $\varphi(x)$, and then finding a regression function as follows:

$$f(x) = \omega \cdot \varphi(x) + b \quad (2.11)$$

which can best approximate the actual output y with an error tolerance ε , where ω and b are regression function parameters known as weight vector and bias value, respectively. φ is known as a nonlinear mapping function.

The coefficients b and ω are calculated by minimizing the following regularized risk function

$$R(C) = C \frac{1}{n} \sum_{i=1}^n L_{\varepsilon}(f(x_i), y_i) \frac{1}{2} \|\omega\|^2 \quad (2.12)$$

where the term $\frac{1}{2} \|\omega\|^2$ improves the generalization of the SVM by regulating the degree of model complexity, which denotes the Euclidean norm. C is a positive trade-off parameter that determines the degree of the empirical error in the optimization problem that is chosen by the user. The most important difference with respect to classic regression is the use of a novel loss function (ε). This is the Vapnik's linear loss function with ε -insensitivity (tube size of SVM) zone defined as:

$$L_{\varepsilon}(f(x_i), y_i) = \begin{cases} 0 & \text{for } |f(x_i) - y_i| \leq \varepsilon \\ |f(x) - y| - \varepsilon & \text{otherwise} \end{cases} \quad (2.13)$$

Thus, the loss is equal to 0 if the difference between the predicted and the measured value is less than ε . If the predicted value is within the tube, the loss error is zero. For all other predicted points outside the tube, the loss equals the magnitude of the difference between the predicted value and the radius ε of the tube. To avoid outliers, are introduces ξ and ξ^* called slack variables for measurements "above" and "below" of the ε tube respectively. Both slack variables are positive values.

Thus, minimizing the risk as follow:

minimize

$$R(\xi, \xi^*, \omega, b) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2.14)$$

$$\text{subjected to } \begin{cases} y_i - \omega\phi(x_i) - b_i \leq \varepsilon + \xi_i \\ \omega\phi(x_i) + b_i - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Where the $C \sum_{i=1}^n (\xi_i + \xi_i^*)$ controls the degree of empirical risk. Figure 2.3 illustrates the concept of SVR.

To solve the optimization problem, Lagrange multipliers α and α^* are added to the condition equations, and the equation can be written as its dual form:

$$\begin{aligned} R(\alpha, \alpha^*) = & \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i - \alpha_i^*) \\ & - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) \end{aligned} \quad (2.15)$$

with constrains:

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad 0 \leq \alpha_i \leq C, \quad 0 \leq \alpha_i^* \leq C \quad i = 1, 2, \dots, n$$

where α_i and α_i^* are Lagrange multipliers to be solved, and $K(x_i, x_j)$ is called kernel function and is calculated by $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ on the feature space. The kernel allows SVMs to form nonlinear boundaries; in other words, it provides the SVM the ability to model complicated separating hyperplanes.

After calculating Lagrange multipliers, find an optimal desired weights vector of the regression hyperplane as follow:

$$\omega = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(x) \quad (2.16)$$

and the equation (2.11) it can be rewritten as follow:

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \cdot K(x_i, x_j) + b \quad (2.17)$$

where n is the number of support vectors, $(\alpha_i - \alpha_i^*)$ are their Lagrange multipliers, the term $K(x_i, x_j)$ is the kernel function in the input space and the bias b is calculated from training samples.

In general, mathematically, a basic function for the statistical learning process in SVM is

$$y = f(x) = \sum_{i=1}^n \alpha_i \cdot \varphi(x) = w\varphi(x) \quad (2.18)$$

where the output is a linearly weighted sum of n . The nonlinear transformation is carried out by $\varphi(x)$.

The decision function of SVM is represented as

$$y = f(x) = \left\{ \sum_{i=1}^{NT} \alpha_i K(x_i, x) \right\} - b \quad (2.19)$$

where K is the kernel function, α_i and b are parameters, NT is the number of training data, x_i are vectors used in training process, and x is the independent vector. The parameter α_i and b are derived by maximizing their objective function. A suitable choice of kernel allows the data to become separable in the feature space despite being non-separable in the original input space. The four basic kernel functions are showed in table 2.1.

Table 2.1. Different kernel functions

Name of kernel function	Equation
Radial basis function(RBF)	$K(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2), \gamma > 0$ (2.20)
Polynomial	$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \gamma > 0$ (2.21)
Linear	$K(x_i, x_j) = x_i \cdot x_j$ (2.22)
Sigmoid	$K(x_i, x_j) = \tanh(\gamma x_i \cdot x_j + r)$ (2.23)

x_i and x_j are vectors in the input space, d , r , and γ are the kernel parameters.

Figure 2.3. Illustration of nonlinear SVR with Vapnik's ξ —insensitivity loss function.

2.3.3. Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS, proposed by Prof. J.S. Roger Jang (Jang, 1993) of National Tsing Hua University, is a hybrid model composed of a fuzzy and artificial neural network, where the nodes in the different layers of a feed-forward network handle fuzzy parameters. This is equivalent to fuzzy inference systems (FIS) with distributed parameters. At its core, the technique splits the representation of prior knowledge into subsets in order to reduce the search space, and used the back propagation algorithm to adjust the fuzzy parameters. The resulting system is an adaptive neural network functionally equivalent to a first-order Takagi –Sugeno inference system, where the input-output relationship is linear.

In first –order Sugeno system, a typical rule set with two fuzzy IF/THEN rules can be expressed as:

- Rule 1. If x is A_1 and y is B_1 , then $f_1 = p_1 x + q_1 y + r_1$ (2.24)

- Rule 2. If x is A_2 and y is B_2 , then $f_2 = p_2 x + q_2 y + r_2$ (2.25)

where x and y are the crisp inputs to the node i , A_i and B_i are the fuzzy sets in the antecedent, f_i is the output within the fuzzy region specified by the fuzzy rule; p_i , q_i and r_i are the design parameters that are determined during the training process.

ANFIS architecture consists of five layers, namely: fuzzy layer, product layer, normalized layer, de-fuzzy layer and total output layer. The ANFIS structure for two inputs, one output and two rules is shown in figure 2.4. Each layer performs a particular task in the fuzzy inference system. For identification, the adaptive nodes are represented by squares, and fixes nodes are represented by circles.

Figure 2.4. Basic structure of an ANFIS model.

Layer 1. (Fuzzy layer): Each node i in this layer (denoted with a square) represents a node function:

$$O_{1,i} = \mu_{A_i}(x), \text{ for } i = 1, 2, \dots, n \quad O_{1,i} = \mu_{B_{i-2}}(y), \text{ for } i = 1, 2, \dots, n \quad (2.26)$$

where x (or y) is the input to node i , and A_i (or B_{i-2}) is the linguistic labels (small, large, etc.) characterized by appropriate membership functions (MF's) $\mu_{A_i}(x)$ and $\mu_{B_{i-2}}(y)$. The MF's can take any shape or function such as triangular, trapezoidal, Gaussian, tor bell-shaped (table 2.2). The parameters of the MF's in this layer are named as premises.

Table 2.2. The basic MF's.

Name of MF's	Equation
Triangular MF	$\mu_{A_i}(x) = \max \left\{ \min \left(\frac{x-a}{b-a}, \frac{c-x}{c-b} \right), 0 \right\}$ (2.27)
Trapezoidal MF	$\mu_{A_i}(x) = \max \left\{ \min \left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right), 0 \right\}$ (2.28)
Gaussian MF	$\mu_{A_i}(x) = e^{\frac{-(x-c)^2}{2\sigma^2}}$ (2.29)
Bell-Shaped MF	$\mu_{A_i}(x) = \frac{1}{1 + \left \frac{x-c}{a} \right ^{2b}}$ (2.30)

$\{a, b, c, d\}$ is the parameter set that changes the shapes of the MFs with maximum 1 and minimum 0.

Layer 2. (Product layer): This layer has the circle nodes represented with Π in figure 2.4. In this layer, the rule operator (AND/OR) is applied to get one output that represents the results of the antecedent for a fuzzy rule that multiplies the incoming signals such as:

$$\text{Output}_{2,i} = w_{oi} = \mu_{Ai}(x) * \mu_{Bi}(y), \quad \text{for } i = 1, 2. \quad (2.31)$$

The output signal w_{oi} represents the firing strength of the rule.

Layer 3. (Normalized layer): The nodes in this layer denoted with N and they calculate the ratio of the i th rule's firing strength to the sum of firing strengths of all rules by:

$$\text{Output}_{3,i} = \bar{w}_0 = \frac{w_{oi}}{w_{o1} + w_{o2}}, \quad \text{for } i = 1, 2. \quad (2.32)$$

The quantity $\bar{w}_0$ is known as the normalized firing strengths.

Layer 4. (De-fuzzy layer): The nodes in this layer are denoted with a square and they calculate the weighted output of each linear function as follows:

$$\text{Output}_{4,i} = \bar{w}_{0i} \cdot f_i = \bar{w}_{0i} (p_i x + q_i y + r_i), \quad \text{for } i = 1, 2. \quad (2.33)$$

where $\bar{w}_0$ is the output of layer 3, and $\{p_i, q_i, r_i\}$, are the coefficients of a linear combination in the Sugeno inference system. These parameters of this layer are referred as the consequent parameters.

Layer 5. (Total output layer): The single node denoted with an Σ computes overall output as follows:

$$\text{Output}_{5,i} = \sum \bar{w}_{0i} \cdot f_i = \frac{\sum_i w_{0i} \cdot f_i}{\sum_i w_{0i}} = f_{\text{out}} = \text{Estimated overall output} \quad (2.34)$$

ANFIS uses a hybrid learning algorithm (figure 2.5) for estimation of the premise and consequent parameters. The hybrid learning algorithm procedure estimates the consequent parameters in a forward pass and the premise parameters in a backward pass. In the forward phase, the information propagates forward until layer 4, where the consequent parameters are optimized by the least square regression algorithm (LSA). In the backward phase, the error signals propagate backwards and the premise parameters are updated by gradient descent algorithm (GDA)(Jang *et al.*, 1997). This error measure is usually defined by the sum of the squared difference between measured and modeled values and is minimized to a desired value.

The final overall output in the figure 2.4 can be rewritten as:

$$f_{\text{out}} = (\bar{w}_{01} \cdot x)p_1 + (\bar{w}_{01} \cdot y)q_1 y + (\bar{w}_{01})r_1 + (\bar{w}_{02} \cdot x)p_2 + (\bar{w}_{02} \cdot y)q_2 y + (\bar{w}_{02})r_2 \quad (2.35)$$

where x and y are the input parameters of the model, $\bar{w}_{01}$, $\bar{w}_{02}$ are the normalized firing strengths of fuzzy rules and $(p_1; q_1; r_1; p_2; q_2; r_2)$ are the consequent parameters.

Figure 2.5. The ANFIS learning algorithm.

3. MATERIALS AND METHODS

3.1. Study Area and data

All case studies were conducted in the Yucatán Peninsula which is located between 19°40' and 21°37'N, and 87°30' and 90°26'W, in east Mexico. The region covers an area of 142,210 km² and is surrounded by the Caribbean Sea and the Gulf of Mexico. The climate of the Yucatán Peninsula is classified as tropical savanna (Aw) according to the Köppen system (Köppen, 1936), with a rainy summer and dry winter. Annual mean temperature ranges from 25.8 to 26.3 °C and precipitation ranges from a minimum in the northwest of 600 mm/year to higher quantities toward the southeast (1400 mm/year). The majority of the Yucatán Peninsula lies at 50 m below mean sea level. The weather data [minimum air temperature (°C), maximum air temperature (°C), mean wind speed (m s⁻¹), mean relative humidity (%), rainfall (mm) and global solar radiation (MJ m⁻² day⁻¹)] were provided by The Mexican National Meteorological Service (SMN; *Servicio Meteorológico Nacional*) and *Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias* (INIFAP). Weather data were collected at 15 and 10 minute intervals and transformed to daily averaged time series. Table 3.1 and figure 3.1 shows the geographical locations of the automatic weather stations used in this study.

Table 3.1. Geographical location of the meteorological stations

Station name	Longitude (°W)	Latitude (°N)	Elevation (m)	Time Period
Calakmul	-89.8925	18.365	28	2003 – 2014
Campeche	-90.5072	19.8361	11	2001 – 2014
Cancún	-86.7758	21.075	8	2000 – 2013
Celestún	-90.3831	20.858	10	2000 – 2014
Chetumal	-88.3278	18.5005	14	2000 – 2013
Efraín Hernández	-88.7020	18.1935	90	2006 – 2014
Mérida	-89.6517	20.9463	18	2000 – 2006
Tantakín	-89.0472	20.0303	30	2003 – 2011

Figure 3.1. Distribution of the selected stations in Yucatán Peninsula, Mexico.

3.2. Weather data quality assessment.

The database provided by the SMN and INIFAP was analyzed to find Incorrect or missing weather observations, mainly associated with the malfunction of instruments, and to find weather observations affected by weather systems. To overcome these problems, the following procedure was implemented:

1. To find and add missing dates in the database, a macro was written in Microsoft Excel software.
2. To identify erroneous or outlier data of air temperature, relative humidity, wind speed and precipitation, two methods were used, which may be distinguished as follows: (a) analysis of standardized residuals and (b) adjusted functional boxplots technique.

- a) The standardized residuals are the residuals divided by the estimates of their standard errors. Standardizing is a method for transforming data so that its mean is zero and standard deviation is one. If the distribution of the residuals is approximately normal, then 95% of the standardized residuals should fall between -2 and +2. If the residuals fall outside of + or – 2, then they could be considered outliers (Norusis, 2008).
- b) The adjusted functional boxplot is constructed by drawing a box between the upper and lower quartiles with a solid line drawn across the box to locate the median. The following quantities (called fences) are needed for identifying extreme values as following:

Near outlier limits

- Lower inner fence: $Q1 - 1.5 \times IQR$
- Upper inner fence: $Q3 + 1.5 \times IQR$

Far outlier limits

- Lower outer fence: $Q1 - 3 \times IQR$
- Upper outer fence: $Q3 + 3 \times IQR$

where $Q1$ and $Q3$ are defined as the 25th and 75th percentiles and IQR is named the interquartile-range ($IQR=Q3-Q1$), This is a robust range for interpretation because the 50% central region is not affected by outliers or extreme values, and gives a less biased visualization of the curves' spread (Sun and Genton, 2012). A point beyond an inner fence on either side is considered a mid-outlier. A point beyond an outer fence is considered an extreme outlier.

3. In the case of solar radiation data to find erroneous values, the daily sky clearness index (K_T) was applied as an indicator of sky condition (transparency or cloudiness). K_T is calculated as the ratio of measured daily global solar radiation intensity to the daily extraterrestrial solar radiation on a horizontal surface (Badescu, 2014). The upper and lower limits for K_T

represent a clear sky day and completely cloudy day, respectively. Values of 0.015 and 1.00 for lower and upper threshold respectively are recommended by Jiang (2009) and Khorasanizadeh and Mohammadi (2013). For this reason, data were eliminated if the values of the daily sky clearness index were outside the range of $0.015 < K_T < 1.00$ (unitless).

4. Verify subjectively in the database if these outliers or erroneous values are not related a weather system and remove them otherwise.
5. If there were more than 5 consecutive days of incorrect values in a month, data for the entire month were deleted.
6. To fill in missing values, a cubic interpolation was conducted using the Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) method (Fritsch and Carlson, 1980; Kahaner *et al.*, 1989). The PCHIP is a non – linear interpolation method that interpolates between two data points using less oscillation that ensures the original shape of the peak is preserved with only a minimal degree of curvature existing between data points, resulting in an interpolated peak that retains the original maxima and minima of the data points. PCHIP was carried out using the Matlab V. R215b Software.

3.3. Performance evaluation of empirical and Soft-computing models.

The performance of different models was assessed using statistical parameters that compare the calculated and measured data: mean percentage error (MPE; Equation 3.1), mean absolute percentage error (MAPE; Equation 3.2), root mean square error (RMSE; Equation 3.3), mean absolute error (MAE; Equation 3.4), mean bias error (MBE; Equation 3.5), and coefficient of determination (R^2 ; Equation 3.6):

$$MPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{O_i - P_i}{O_i} \right) \times 100 \quad (3.1)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{O_i - P_i}{O_i} \right| \times 100 \quad (3.2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (3.3)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (3.4)$$

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^n (O_i - P_i) \quad (3.5)$$

$$R^2 = \frac{[\sum_{i=1}^n (O_i - O_m)(P_i - P_m)]^2}{\sum_{i=1}^n (O_i - O_m)^2 \sum_{i=1}^n (P_i - P_m)^2} \quad (3.6)$$

where n is the number of comparisons, O_i is some measured values, P_i is the corresponding model prediction, O_m and P_m are the average values of O_i and P_i .

The MPE is an overall measure of forecast bias and is a computed average of percentage errors. The MAPE is an overall measure of forecast accuracy and is computed from the absolute differences between a series of estimated and measured data. The RMSE provides information on the short-term performance of the model by allowing a term-by-term comparison of the deviation between the estimated values around the measured data. The MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. Smaller values of RMSE and MAE imply a higher accuracy in the modelling. The MBE is an indicator for the average deviation of the estimated values from the observed data and describes systematic error or bias, with positive values of MBE indicating over-estimates, and negative values indicating under-estimates by the model. R^2 is the square of the coefficient of determination and indicates how well model estimates represent trends in the observed data. A high value of R^2 is

desirable as this shows a lower unexplained variation. R^2 is a statistic that gives some information about the goodness-of-fit of a model. In regression, the R^2 is a statistical measure of how well the regression line approximates the real data points. An R^2 of 1.0 indicates that the regression line perfectly fits the data.

These statistical indicators are commonly used statistical tests and provide reasonable criteria to evaluate model performance (Besharat *et al.*, 2013; Teke *et al.*, 2015).

3.4. Temperature – based Methods for estimating daily reference evapotranspiration.

The assessment of TET models for determining ET_0 has caught the attention of numerous studies worldwide. In this study, use Hargreaves and Samani formulation HS (Hargreaves and Samani, 1982; Hargreaves and Samani, 1985; Hargreaves and Allen, 2003) and the PMT formulation for ET_0 modelling. PMT equation, can produce acceptable results using only air temperature data, is based on the FAO56-PM formulation and retains many of the basis and philosophy of the combination model of Penman, considering a combination of net radiation and aerodynamic principles (Pereira *et al.*, 2015).

This study was carried out using daily weather data set from four meteorological stations (Campeche, Efraín Hernández, Mérida and Tatankín) located in the Yucatán Peninsula, Mexico (Figure. 3.1).

The objectives of this study are to: 1) apply seven TET methods for estimating ET_0 in four cities of the Yucatán Peninsula in a tropical sub-humid climate, 2) assess the performance of the TET models when compared with the PM-FAO model, and 3) test the PMT model and evaluating the advantage of adopting the PMT model in a scenario of limited meteorological data.

3.4.1. Temperature based methods (TET)

Seven TET methods were considered in this study (Table 3.2). These methods were chosen as commonly used methods that rely on air temperature data (Almorox *et al.*, 2015).

- **Thornthwaite (1948) method**

Thornthwaite's model (Thornthwaite, 1948) is a widely used formulation for estimating ET:

$$ET_{TH} = 16(10 T_{mean}/I)^a Nd/(30 Nd) N/12 \text{ mm day}^{-1} \quad (3.7)$$

$$I = \sum I_j ; j = 1, \dots, 12$$

$$i = (T_{meanj}/5)^{1.514} \quad (3.8)$$

$$a = 6.7512 \times 10^{-7} I^3 - 7.711 \times 10^{-5} I^2 + 1.7921 \times 10^{-2} I + 0.49239 \quad (3.9)$$

ET_{TH} is zero for negative temperatures; for mean temperatures between 26.5°C and 38°C (above 38°C the unadjusted potential evapotranspiration becomes asymptotic at 185 mm/month, Thornthwaite, 1948), the ET_{TH} is expressed by (If, $T_{mean} > 26.5^\circ\text{C}$):

$$ET_{TH} = (-13.861666 + 1.07466 (T_{mean}) - 0.014333 (T_{mean})^2) Nd/30 N/12 \text{ mm day}^{-1} \quad (3.10)$$

- **Hamon (1963) method**

Hamon developed a modified form (Hamon, 1963):

$$ET_{HA} = 0.1651 N/12 [216.7 e^{(T_{mean})/(T_{mean}+273.3)}] \quad (3.11)$$

(If $T_{mean} < 0$ $ET_{HA} = 0$)

where ET_{HA} is the daily evapotranspiration (mm day^{-1})

- **Papadakis (1966) method**

The Papadakis model (Papadakis, 1966) depends on the difference in the saturated vapor pressure above the water body at maximum and minimum air temperatures.

As saturation vapor pressure is related to air temperature, it can be calculated from the air temperature.

$$ET_{PA} = 5.625/Nd [e^{\circ}(T_{max}) - e^{\circ}(T_{min}-2)] \quad (3.12)$$

where: ET_{PA} is the evapotranspiration (mm day^{-1})

- **Malmström (1969) method**

The Malmström formula to estimate ET is (Malmström, 1969):

$$ET_{MA} = 4.09/Nd [e^{\circ}(T_{mean})] \quad (3.13)$$

where: ET_{MA} is the potential evapotranspiration mm day^{-1} .

- **Hargreaves and Samani (1982, 1985) method**

The ET_{HS} model (Hargreaves and Samani) is given as

$$ET_{HS} = 0.0023 \times 0.408 \times H_0 (T_{mean}+17.8) (T_{max}-T_{min})^{0.5} \quad (3.14)$$

where: ET_{HS} is the reference evapotranspiration estimated (mm day^{-1})

- **Camargo (1999) method:**

In this model the average temperature (T_{mean}) was replaced, in the Thornthwaite model, by the effective temperature (T_{ef}), given by (Camargo *et al.*, 1999):

$$T_{ef} = 0.36 (3T_{max}-T_{min}) \quad (3.15)$$

- **Oudin (2005) method**

For input into hydrological models, Oudin *et al.* (2005) derived a new formulation calibrated on catchments in USA, France and Australia:

$$ET_{OU} = 0.408 H_0 [0.01(T_{mean}+5)] \text{ if } T_{mean} > 5^{\circ}\text{C}; \text{ otherwise } ET_{OU} = 0 \quad (3.16)$$

where ET_{OU} is the evapotranspiration (mm day^{-1}).

Table 3.2. Models and original coefficients used in this study for the estimation of evapotranspiration in Temperature based methods. .

ET model	Sym.	Variables	Original coefficients	Equation (mm day^{-1})
(Thornthwaite, 1948)	TH	T_{mean}, N	$K_{TH} = 16$	$ET_{TH} = K_{TH} \times (10 \times T_{mean} / I)^a \times N / 360$
(Hamon, 1963)	HA	T_{mean}, N	$K_{HA} = 0.1651$	$ET_{HA} = K_{HA} \times N / 12 \times [216.7 \times e^{\circ(T_{mean})} / (T_{mean} + 273.3)]$
(Papadakis, 1966)	PA	T_{max}, T_{min}	$K_{PA} = 5.625$	$ET_{PA} = K_{PA} / N \times [e^{\circ(T_{max})} - e^{\circ(T_{min}-2)}]$
(Malmström, 1969)	MA	T_{mean}	$K_{MA} = 4.09$	$ET_{MA} = K_{MA} / N \times [e^{\circ(T_{mean})}]$
(Hargreaves and Samani 1982, 1985)	HS	$T_{max}, T_{min}, T_{mean}, H_0$	$K_{HS} = 0.0023$	$ET_{HS} = K_{HS} \times 0.408 H_0 (T_{mean} + 17.8) (T_{max} - T_{min})^{0.5}$
(Camargo et al., 1999)	CA	$T_{max}, T_{min}, T_{mean}, N$	$K_{CA1} = 16; K_{CA2} = 0.36$	$ET_{CA} = K_{CA1} \times (10 \times (K_{CA2} \times (3T_{max} - T_{min})) / I)^a \times N / 360$
(Oudin et al., 2005)	OU	T_{mean}, H_0	$K_{OU} = 0.01$	$ET_{OU} = [0.408 H_0] [K_{OU} (T_{mean} + 5)] \text{ if } T_{mean} > 5^{\circ}\text{C} (0 \text{ otherwise})$

The study was followed by the parametric calibration of the seven equations for the four stations studied. Table 3.2 shows the expressions used in each method and their original coefficients. Locally calibration using climatological data is important in decreasing the errors of the evapotranspiration estimates. In the results section (Table 4.1) shows the coefficients (a , b) used for the calibrated methods in the different locations studied.

3.4.2. Penman – Monteith method using only temperature data (PMT).

The limited availability of meteorological data (i.e., air humidity, wind speed and global solar radiation) restricts the use of the PM method in many locations. In these cases, the use of a temperature-based model (e.g. PMT) is a viable alternative. In the PMT method, the global solar radiation and air humidity are estimated by as a function of air temperature.

The complete expression of the PMT equation when PM is applied using only measured temperature data (mm day^{-1}) is disaggregated in several components. Equation 3.17 estimates the net solar shortwave radiation and Equation 3.18 estimates the net longwave radiation. Equation 3.19 estimates the radiation term of the PM equation, and equation 3.20 estimates the aerodynamic contribution to ET_0 . The equations assume a mean daily wind speed $u_2=2.0 \text{ ms}^{-1}$. A different value may be needed in climates with different wind speeds. The PMT_{rad} , PMT_{aero} , and PMT have the unit mm day^{-1} .

$$R_{ns} = 0.77 H_0^{0.17} (T_{\max} - T_{\min})^{0.5} \quad (3.17)$$

$$R_{nl} = 1.35 \left(\frac{0.17 (T_{\max} - T_{\min})^{0.5}}{0.75 + 2 \times 10^{-5}} - 0.35 \right) \sigma \left(0.34 - 0.14 \left(0.6108 \exp \left(\frac{17.27 T_{\min}}{T_{\min} + 237.3} \right)^{0.5} \right) \right) \left(\frac{(T_{\max} + 273.15)^4 + (T_{\min} + 273.15)^4}{2} \right) \quad (3.18)$$

$$\text{PMT}_{\text{rad}} = \left(\frac{0.408 \Delta}{\Delta + \rho(1 + 0.34 u_2)} \right) (R_{ns} - R_{nl} - G) \quad (3.19)$$

$$\text{PMT}_{\text{aero}} = \frac{\rho \frac{900 \cdot u_2}{T_{\text{mean}} + 273} \left(\frac{e_{s(T_{\max})} + e_{s(T_{\min})}}{2} - e_{s(T_{\min})} \right)}{\Delta + \rho(1 + 0.34 u_2)} \quad (3.20)$$

$$\text{PMT} = \text{PMT}_{\text{rad}} + \text{PMT}_{\text{aero}} \quad (3.21)$$

where, PMT is the reference crop evapotranspiration for short crop (grass) PMT_{rad} is the radiation term of the PMT corresponding to the incoming net short radiation component and the outgoing net long wave radiation component (mm day^{-1}) and PMT_{aero} is the aerodynamic component (mm day^{-1}).

Note that the Hargreaves and Samani (1982) equation was employed to estimate the incoming global solar radiation (H) in equation 2.1. The equation for H is expressed as:

$$H = H_0 \times [k_{\text{HS}} \times (T_{\max} - T_{\min})^{0.5}] \quad (3.22)$$

where H_0 is the extraterrestrial radiation and k_{HS} is a coefficient that needs to be adjusted locally for better accuracy (Hargreaves and Allen, 2003). In this study, the k_{HS} value 0.17 was fixed in the calculation.

The relative humidity is calculated from the e_a value which assumes that the daily T_{dew} is close to the daily minimum air temperature (T_{min}). For our stations with a moist sub-humid climate, the estimated T_{dew} value from T_{min} does not require a correction for aridity (Raziei and Pereira 2013; Todorovic *et al.*, 2013).

Finally, wind speed data is one of the least available data among the ones needed. In locations where no wind data are accessible, the average value $u_2 = 2 \text{ m s}^{-1}$ (Allen *et al.*, 1998) was adopted. Impacts of strong windy conditions could be overcome by employing a mesoclimatic wind speed average instead of the default value (Ren *et al.*, 2016, Almorox *et al.*, 2017).

3.4.3. Models performance evaluation

The accuracy and suitability of the models were assessed using the following statistical parameters: the determination coefficient (equation 3.6), mean absolute error (equation 3.4) mm day^{-1} , mean absolute percentage error (equation 3.2) %, and root mean square error (equation 3.3) mm day^{-1} .

3.5. Empirical models for estimating daily global solar radiation

Empirical models employ relationships founds between atmospheric transmissivity and other metrological variables. In this study, daily climatic data from six automated weather station (Calakmul, Campeche, Celestún, Efraín Hernández, Mérida and Tatankín) were used to evaluate and establish the models.

A number of methods that use empirical relationships to H modelling from commonly measured meteorological variables have been developed in the past. Eight of these empirical models that estimate H from temperature, these models

hereinafter referred to as “*temperature-based models*”, three that use temperature and rainfall data to estimate H , and one that utilizes temperature, rainfall and relative humidity data to estimate H were selected for testing, the last three models known as “*meteorological parameters-based models*”. Selection of the models considered availability of records, extensiveness of use, previous performance, simplicity, and the current state of model development. In addition, a new model was developed that requires rainfall data, temperature and relative humidity.

3.5.1. Temperature-based

Temperature-based models for estimating global solar radiation utilize the fact that thermal amplitude directly affects atmospheric transmissivity. These models assume that a high or low transmissivity is due to an increase or decrease in air temperature. The temperature increases due to clear sky and high shortwave radiation, and decreases due to increase of transmissibility.

Hargreaves and Samani (1982) (T1) estimated H using H_0 and daily air temperature extremes as an indicator of cloudiness and atmospheric transmittance according to equation 3.21:

$$H = H_0 [a(T_{\max} - T_{\min})^{0.5}] \quad (3.21)$$

Bristow and Campbell (1984) (T2) proposed a method for estimating solar radiation from air temperature measurements (equation 3.22). They developed an empirical relationship to express the daily total atmospheric transmittance as a function of daily range of the air temperature (equation 3.23):

$$H = H_0 a[1 - \exp(-b\Delta T^c)] \quad (3.22)$$

$$\Delta T (^{\circ}\text{C}) = T_{\max(j)} - [(T_{\min(j)} + T_{\min(j+1)})/2] \quad (3.23)$$

(Donatelli and Campbell, 1998) (T3) improved Bristow and Campbell's (1984) model (Bristow and Campbell, 1984) by introducing a correction factor for seasonality effects occurring in middle latitude areas ($f(T_{\text{mean}})$; Equations 3.24-3.27):

$$H = H_0 a \left[1 - \exp(-b \cdot f(T_{\text{mean}}) \Delta T^2 \exp(T_{\text{min}(j)}/c)) \right] \quad (3.24)$$

$$\Delta T(^{\circ}\text{C}) = T_{\text{max}(j)} - [(T_{\text{min}(j)} + T_{\text{min}(j+1)})/2] \quad (3.25)$$

$$f(T_{\text{avg}}) = 0.017 \exp(\exp(-0.053 T_{\text{mean}(j)})) \quad (3.26)$$

$$T_{\text{avg}(j)} (^{\circ}\text{C}) = (T_{\text{max}(j)} + T_{\text{min}(j)})/2 \quad (3.27)$$

Goodin *et al.* (1999) (T4) proposed a modified version of the Bristow-Campbell equation by including daily H_0 as an extra term in the equation that acts as a scaling factor, thereby allowing ΔT to accommodate a greater range of solar radiation values (equation 3.28). The modification allows application of the model at sites that are distant from the calibration site.

$$H = H_0 a \left[1 - \exp \left(-b \left(\Delta T(^{\circ}\text{C})^c / H_0 \right) \right) \right] \quad (3.28)$$

Winslow *et al.* (2001) (T5) introduced a method that was developed to be globally applicable in the prediction of H , using saturation vapor pressure that was based on minimum and maximum temperature.

$$H = H_0 \tau_{cf} D_t [1 - a(e_s(T_{\text{min}}))/(e_s(T_{\text{max}}))] \quad (3.29)$$

$$\tau_{cf} = (\tau_o \tau_a \tau_v)^{P_a/P_o} \quad (3.30)$$

τ_{cf} is calculated using Equations 3.31- 3.33 from the site's latitude (ϕ), elevation (z) and mean annual temperature (T_{mean}). τ_o is the transmittance of clean dry air; τ_a represents the transmittance affected by atmospheric aerosols and ozone (the

absorption of radiation by aerosols is extremely variable and inherently unpredictable from limited data set; therefore, τ_a is often set at 1.0); τ_v is the transmittance affected by atmospheric water vapor, and P_a/P_o is a correction for site elevation z (m).

$$\frac{P_a}{P_o} = [1 - (2.2569 \times 10^{-5})z]^{5.2553} \quad (3.31)$$

$$\tau_o = 0.947 - (1.033 \times 10^{-5})(|\phi|^{2.22}) \quad (3.32)$$

for $|\phi| \leq 80^\circ$

And; $\tau_o = 0.774$ for $|\phi| \geq 80^\circ$

$$\tau_v = 0.9636 - 9.092 \times 10^{-5}[(T_{\text{mean}} + 30)^{1.8232}] \quad (3.33)$$

Function D_l corrects for errors introduced by site differences in day length, which causes a difference between the time of T_{max} , where the relative humidity reaches its minimum, and sunset. The day-length correction is approximated by:

$$D_l = \left[1 - (H_{\text{day}} - (\pi/4))^2 / 2H_{\text{day}}^2\right]^{-1} \quad (3.34)$$

Annandale *et al.* (2002) (T6) presents a model based on Hargreaves and Samani (Hargreaves and Samani, 1982) incorporating the effect of the site altitude (equation 3.35):

$$H = H_o[a(1 + 2.7 \times 10^{-5}z)(T_{\text{max}} - T_{\text{min}})^{0.5}] \quad (3.35)$$

Mahmood and Hubbard (2002) (T7) proposed a model that requires measured daily range of air temperature and estimated daily clear sky solar radiation (*ICSKY*) that is calculated from day of the year and maximum day length for the year at a given latitude.

$$H = a(T_{\text{max}} - T_{\text{min}})^b \text{ICSKY}^c \quad (3.36)$$

$$ICSKY = I_s \text{ } ^\circ\text{C} \quad (3.37)$$

$$^\circ\text{C} = 0.8 + 0.12(|182 - n_{\text{day}}/183|)^{1.5} \quad (3.38)$$

$$I_s = 0.04188 \{ A + B \sin[(2\pi((n_{\text{day}} + 10.5)/365) - (\pi/2))] \} \quad (3.39)$$

$$A = (0.29 \cos \phi + 0.52) \{ \sin \phi (46.355 \text{ LD} - 574.388) + (816.41 \cos \phi \sin[(\text{LD}\pi/24)]) \} \quad (3.40)$$

$$B = (0.29 \cos \phi + 0.52) \left\{ \sin \phi (574.3885 - 1.509 \text{ LD}) - (29.59 \cos \phi \sin \left[\left(\frac{\text{LD}\pi}{24} \right) \right] \right\} \quad (3.41)$$

$$\text{LD} = 0.267 \sin^{-1} \{ 0.5 + (0.007895/\cos \phi) + (0.2168875 \tan \phi) \}^{0.5} \quad (3.42)$$

Mahmood and Hubbard (2002) proposed to reduce their model's local bias by accounting for local scale advection and frontal movements with the following linear regression (equation 3.43):

$$H = \frac{[a(T_{\text{max}} - T_{\text{min}})^b \text{ICSKY}^c] - 2.4999}{0.8023} \quad (3.43)$$

Almorox *et al.* (2011) (T8) proposed a model, which is a function of extraterrestrial solar radiation, difference in maximum and minimum air temperature, saturation vapor pressure at temperature T_{min} , and saturation vapor pressure at temperature T_{max} :

$$H = a H_0 (T_{\text{max}} - T_{\text{min}})^b [1 - \exp(-c (e_s(T_{\text{min}})/(e_s(T_{\text{max}}))^d))] \quad (3.44)$$

3.5.2. Meteorological parameter-based

Accuracy of models that use temperature data can be improved by adding other variables such as rainfall or humidity. It has been found that in some cases lower

levels of solar radiation are associated with rainy days and high humidity (Liu and Scott, 2001; Li *et al.*, 2013). These conditions apply to the case in the Yucatán, Peninsula where very low solar radiation levels are associated with rainy days.

Hunt *et al.* (1998) (M1) proposed a model for estimating the daily global radiation on a horizontal surface based on maximum and minimum temperature and daily precipitation (equation 3.45):

$$H = a H_0 (T_{\max} - T_{\min})^{0.5} + b T_{\max} + c P + d P^2 + e \quad (3.45)$$

Liu and Scott (2001) (M2) suggested a relationship for estimating daily global radiation as a function of daily H_0 and transformed rainfall data RT . Rain data were transformed to calculate RT as follows: if $P > 0$, $R = 1$; $P = 0$, $R = 0$, with subscripts $j-1$, j and $j+1$ referring to the previous, current and next days (equation 3.46):

$$H = H_0 a [1 - \exp(-b \Delta T^c)] (1 + d RT_{j-1} + e RT_j + f RT_{j+1}) + g \quad (3.46)$$

Chen *et al.* (2006) (M3) estimated H using a model based on temperature, total daily rainfall and daily average saturation deficit D .

$$H = H_0 (a (T_{\max} - T_{\min})^{0.5} + b) + c P + d D + e \quad (3.47)$$

Wu *et al.* (2007) (M4) introduced a model for estimating the daily global radiation based on maximum and minimum temperature, daily average temperature, H_0 and RT .

$$H = H_0 (a + b (T_{\max} - T_{\min})^{0.5} + c T_{\text{avg}} + d RT_j) \quad (3.48)$$

A new empirical model (M5) utilizing maximum and minimum temperature, transformed rainfall and average relative humidity was developed for the current project. Assuming that low solar radiation values are related to rainfall events and

high relative humidity, the estimation of H can be done by the following expression (equation 3.49):

$$H = H_0((T_{\max} - T_{\min})^a + b)(1 + cRH) + dRT_j \quad (3.49)$$

3.5.3. Calibration and performance evaluations

Performance of different models to estimate the global solar radiation was assessed using statistical parameters that compare the calculated and measured daily solar radiation data: mean percentage error (equation 3.1) %, mean absolute percentage error (equation 3.2) %, root mean square error (equation 3.3) $\text{MJm}^{-2} \text{day}^{-1}$, mean absolute error (equation 3.4) $\text{MJm}^{-2} \text{day}^{-1}$, mean bias error (equation 3.5) $\text{MJm}^{-2} \text{day}^{-1}$, and coefficient of determination (equation 3.6):

Model coefficients of the table 3.3 were fit in each case by using nonlinear regression according to the Marquardt – Levenberg method in Statgraphics Centurion (v.16.2) software. Also, the new model was developed using non-linear regression techniques.

Table 3.3. A list of models used for estimation of daily global solar radiation.

Author Model	Model ID	Model requirements	Equation
Hargreaves and Samani (1982)	T1	$T_{\min}, T_{\max}, H_0$	$H = H_0[a(T_{\max} - T_{\min})^{0.5}]$
Bristow and Campbell (1984)	T2	$T_{\min}, T_{\max}, H_0$	$H = H_0 a[1 - \exp(-b \Delta T^c)]$
Donatelli and Campbell (1998)	T3	$T_{\min}, T_{\max}, H_0$	$H = H_0 a [1 - \exp(-b \cdot f(T_{\text{avg}}) \Delta T^2 \exp(T_{\min(j)}/c))]$
Goodin et al. (1999)	T4	$T_{\min}, T_{\max}, H_0$	$H = H_0 a[1 - \exp(-b(\Delta T(^{\circ} \text{C})^c / H_0))]$
Winslow et al. (2001)	T5	$T_{\min}, T_{\max}, T_{\text{mean}}, z, \phi, H_0, H_{\text{day}}$	$H = H_0 \tau_c f D \cdot [1 - a(e_s(T_{\min})) / (e_s(T_{\max}))]$
Annandale et al. (2002)	T6	$T_{\min}, T_{\max}, z, H_0$	$H = [a(1 + 2.7 \times 10^{-5} z)(T_{\max} - T_{\min})^{0.5}]$

Mahmood and Hubbard (2002)	T7	$T_{min}, T_{max}, LD, \phi, n_{day}$	$H = a(T_{max} - T_{min})^b ICSKY^c$
Almorox et al. (2011)	T8	T_{min}, T_{max}, H_0	$H = a H_0 (T_{max} - T_{min})^b [1 - \exp(-c (e_s(T_{min}) / (e_s(T_{max})))^d)]$
Hunt et al. (1998)	M1	T_{min}, T_{max}, P, H_0	$H = a H_0 (T_{max} - T_{min})^{0.5} + b T_{max} + c P + d P^2 + e$
Liu and Scott (2001)	M2	T_{min}, T_{max}, P, H_0	$H = H_0 a [1 - \exp(-b \Delta T^c)] (1 + d R T_{j-1} + e R T_j + f R T_{j+1}) + g$
Chen et al. (2006)	M3	$T_{min}, T_{max}, D, P, H_0$	$H = H_0 (a(T_{max} - T_{min})^{0.5} + b) + c P + d D + e$
Wu et al. (2007)	M4	T_{min}, T_{max}, P, H_0	$H = H_0 (a + b(T_{max} - T_{min})^{0.5} + c T_{avg} + d R T_j)$
Proposed Model	M5	$T_{min}, T_{max}, RH, P, H_0$	$H = H_0 ((T_{max} - T_{min})^a + b) (1 + c RH) + d R T_j$

Where T_{max} , T_{min} , T_{mean} are daily maximum, minimum and average air temperature ($^{\circ}\text{C}$) respectively, H_0 extraterrestrial radiation ($\text{MJm}^{-2} \text{day}^{-1}$), ΔT daily temperature range, τ_{cf} accounts for atmospheric transmittance, D corrects for error introduced by site differences in day length, $e_s(T)$ saturation vapor pressure at temperature T , H_{day} half day length (rad), z elevation above the sea level (m), ϕ latitude (deg), LD Longest day of year (h), n_{day} day of year, $ICSKY$ corrected clear sky solar irradiation, P daily total precipitation (mm), $R T_{j-1}$, $R T_j$ and $R T_{j+1}$ previous, current and next days of the transformed rainfall, RH relative humidity (%), D daily average saturation deficit (kPa).

3.6. Estimating daily global solar radiation by day of the year

H is a yearly quasi-periodic phenomenon due to seasonal effects. It is convenient to model daily values by the day of the year (n_{day}) (Li *et al.*, 2010). If there is limited availability of weather stations in a particular region, the use of a *DYB* model may be useful for H estimating.

The present study aimed to examine and evaluate four existing *DYB* models for their suitability for H estimating in the following weather stations: Calakmul, Cancún, Celestún, Chetumal, Mérida and Campeche. If none of these models are suitable, a proposed new *DYB* model is also tested for H generating.

3.6.1. Day of the Year Based Models.

Bulut and Büyükalaca (2007) presented a simple model using a sine wave formula to estimate global solar radiation (equation 3.50). The model was based on a trigonometric function, which has only one independent parameter.

$$H = a + b \left| \sin \left[\frac{\pi}{365} (n_{\text{day}} + 5) \right] \right|^{1.5} \quad (3.50)$$

Kaplanis and Kaplani (2007) developed a cosine wave equation (equation 3.51) to estimate daily global solar radiation over the six climatic zones in Greece.

$$H = a + b \cos \left(\frac{2\pi}{364} n_{\text{day}} + c \right) \quad (3.51)$$

Al-Salaymeh (2006) proposed a sine wave model (equation 3.52) to estimate daily global solar radiation in Amman, Jordan.

$$H = a + b \sin \left(\frac{2\pi}{c} n_{\text{day}} + d \right) \quad (3.52)$$

Li et al. (2010) proposed a sine and cosine wave equation (equation 3.53) for 79 meteorological stations across China. The model was applied by Zang *et al.* (2012) and Khorasanizadeh and Mohammadi (2013)

$$H = a + b \sin \left(\frac{2\pi c}{365} n_{\text{day}} + d \right) + e \cos \left(\frac{2\pi f}{365} n_{\text{day}} + g \right) \quad (3.53)$$

Al-Salaymeh (2006) used a single Gaussian function to predict daily global solar radiation in Amman city (Jordan). This function generates a graph characteristically symmetrical and bell-shaped, where its parameters define the position and shape of the curve. For the Yucatán Peninsula, the bell curve was not adjusted to the weather conditions during the rainy season. However, by making use of a sum of two Gaussian functions, the curve fit well (figure 3.2). Thus, a new DYB model based on a sum of two Gaussian correlation formulas

(equation 3.54) is proposed to estimate daily global solar radiation on a horizontal surface. The model was developed using non-linear regression techniques.

$$H = a + b \exp \left[-0.5 \left(\frac{n_{\text{day}} - c}{d} \right)^2 \right] + e \exp \left[-0.5 \left(\frac{n_{\text{day}} - f}{g} \right)^2 \right] \quad (3.54)$$

where H is the daily global solar radiation on a horizontal surface [$\text{MJm}^{-2} \text{day}^{-1}$]; n_{day} is the day of the year starting from January 1. For the 1st of January, $n_{\text{day}}=1$, and for 31st of December, $n_{\text{day}}=365$. a , b , c , d , e , f and g are empirical coefficients that should be determined for each site.

Figure 3.2. Typical variation of daily global solar radiation in Yucatán Peninsula (triangles), and curves generated by the single Gaussian function and sum of two Gaussian correlation formulas.

3.6.2. Calibration and performance evaluations

Model performance was evaluated with metrics that included mean percentage error (equation 3.1) %, mean absolute percentage error (equation 3.2) %, root mean squared error (equation 3.3) $\text{MJm}^{-2} \text{ day}^{-1}$, mean absolute error (equation 3.4) $\text{MJm}^{-2} \text{ day}^{-1}$, mean bias error (equation 3.5) $\text{MJm}^{-2} \text{ day}^{-1}$, and coefficient of determination (equation 3.6). These performance metrics were calculated for each model for each automatic weather station. These metrics are the most widely used by researchers to evaluate the performance of global solar radiation models (Teke *et al.*, 2015).

Model coefficients of equations 3.50 – 3.54 were fit in each case by using nonlinear regression according to the Marquardt-Levenberg method in CurveExpert software (<http://www.curveexpert.net/>). Also, the new model was developed using non-linear regression techniques.

3.7. Soft-computing methods to estimate Daily global solar radiation and reference evapotranspiration

In this section, the ability and accuracy of three soft – computing techniques (SVM, ANFIS and ANN) were assessed for H and ET_0 modelling.

3.7.1. Data pre-process and software

In order to preclude the possibility of overfitting, all data in the measured dataset were normalized to the range from zero to one. This normalization is performed in order to minimize noise, highlight important relationships, detect trends and flatten the distribution of the variable (Rezrazi *et al.*, 2015). In this study, the minimum and maximum (MM) technique was used (Patro and Sahu, 2015; Bharat *et al.*, 2016). MM normalization is a simple technique which provides linear transformation on the original range of the data. As per Min-Max normalization technique,

MM normalization transforms a value X to X' which fits in the range $[C, D]$.

$$X' = \left(\frac{X - X_{\min}}{X_{\max} - X_{\min}} \right) * (D - C) + C \quad (3.55)$$

where,

X' = the normalized data with $[C, D]$ predefined boundary.

X = the input data before normalized.

$X_{\min}$ = the minimum value of the input/output vector.

$X_{\max}$ = the maximum value of the input/output vector.

Moreover, to ensure the representativeness of the dataset, the database was split into two subsets randomly, using 70% for training, and the remaining 30% was used to validate the model. The training dataset was used to train all the models, while validation dataset was used to verify the performance of the trained models.

To bring out computer simulation of ANFIS and ANN techniques, two script files written in MATLAB 2015b software version were employed (Appendix B). The SVM technique was applied using an open source software package named LIBSVM 3.2 (Chang and Lin, 2001) that included the sequential minimal optimization (SMO) algorithm that was implemented using R (RDevelopment, 2012; Meyer and Wien, 2014) (Appendix B).

3.7.2. SVM Model architecture.

For H and ET_0 modelling, the SVMs were trained using the LIBSVM software with ε – SV regression. There are four major kernels predominantly used in SVMs such as linear, polynomial, radial basis and sigmoid. In this study, the radial basis function (RBF) kernel was chosen as the best after a process of trial and error experiments. The SVM hypothesis suggested that the performance of SVM depends on the slack parameter (ε) the cost factor (C) and kernel parameter (γ) while using RBF as the kernel function. The C assigns a penalty for the number of vector falling between the two hyperplanes in the hypothesis. The C suggests that if the data are of good quality, the distance between the two hyperplanes is narrower. If the data is noisy it is preferable to have a smaller value of C which

will not penalize the vectors. So it is important to find the optimum C value for SVM modelling. In this study, these parameters (ε , C , γ) were optimized by using the genetic algorithm (GA) with five-fold cross-validation (CV), varying the ε values between $\varepsilon = 0.002$ to $\varepsilon = 2$, and the varying the C parameters between $C = 0.0001$ to $C = 10$. In addition, the parameter γ of the kernel function is adjusted via GA, varying the γ values between $\gamma = 0.0001$ to $\gamma = 2$. GA is a search-based optimization technique based on the principles of genetics and natural selection. GA combines selection operators, generation, crossover and mutation for identifying the best solution for the optimization problem (Antonanzas-Torres *et al.*, 2015; Zhang *et al.*, 2015).

3.7.3. ANFIS Model architecture

The ANFIS technique was used for H and ET_0 modelling, in which a set of parameters in ANFIS were identified through a hybrid learning rule that combined the back-propagation gradient descent and a least square method. The choice of appropriate membership function (MF's) and a given number of fuzzy IF-THEN rules for each variable is essential to achieve good estimates during training. In the present study, the Bell-shaped function (Equation 2.30) better represented the variability of the data points and was used for the MF's. The bell-shaped function was chosen by an arbitrary process of testing the four types of MF's cited in Table 2.2. The "IF" part of fuzzy rules is related to input variables to MF's and is called premise or antecedent; the "THEN" part, which is related to the output variable for the MF's, is called the conclusion or consequent part of the rule (Cobaner, 2011). The number of fuzzy IF-THEN rules in a system depends on the number of MF's and the number of input variables. So, two IF-THEN rules were employed for each input variable, since the best results were obtained with this value as determined by iterative processes. The grid partition method on the data was used to generate the Takagi-Sugeno fuzzy inference system (FIS) structure (Cobaner, 2011; Shiri *et al.*, 2012). In the grid partition method, the input is divided into different spaces using the axis – paralleled method in which each input represents a fuzzy MF.

3.7.4. ANN Model architecture

This investigation used a three-layer or FFNN for H and ET₀ simulation (Yadav and Chandel, 2014; Qazi *et al.*, 2015; Rezrazi *et al.*, 2015), where the first layer is the input layer representing input variables, the second layer is the hidden layer, and the third layer is the output layer. This topology has proved its ability in modelling many real-world functional problems (Ata, 2015; Piotrowski *et al.*, 2015; Antonopoulos and Antonopoulos, 2017). The selection of hidden neurons is the tricky part in ANN modelling, as it relates to the complexity of the system being modeled. In this study, the optimum numbers of neurons in the hidden layer was determined by a simple trial and errors process. A range of 2 to 80 neurons were evaluated until a minimum acceptable error was achieved between the predicted and observed output. The next step was to choose the transfer functions for the hidden and output layers. In this study, the logistic sigmoid (Logsig) equation (2.9) was used as transfer function in the hidden layer and linear transfer function (Pureline) equation (2.10) was applied in the output layer (Rezrazi *et al.*, 2015). The neural networks were trained using Backpropagation (BP) with Levenberg-Marquardt (LM) algorithm, which is considered to be one of the faster and more accurate algorithm. It combines the speed of the Newton algorithm with the stability of the steepest decent method (Wilamowski *et al.*, 1999). The LM algorithm uses Newton's method to calculate Jacobian matrices without computing the hessian matrices. This makes the LM algorithm have a faster convergence with minimal error. When the performance function has the form of a sum of square (as is typical in training FFNN), The LM algorithm can be expressed as, where the Hessian matrix can be approximated by the following expression:

$$\text{Hessian} = J^T J \quad (3.57)$$

and the gradient can be computed as

$$\text{gradient} = J^T e \quad (3.58)$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. This,

the LM algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$X_{k+1} = X_k - [J^T J + \mu_k \text{Id}]^{-1} J^T e \quad (3.59)$$

where μ_k are non-negative scalars and Id is the identity matrix. When the scalar μ is large, this becomes gradient descent with a small step size, while for small μ the algorithm becomes Gauss – Newton. Thus, μ is decreased after each successful step and is increased only when a tentative step would increase the performance function. In this way, the performance function is always reduced at each interaction of the algorithm. For the neural network mapping problem, the Jacobian matrix can be computed through a standard backpropagation technique that is much less complex than computing the Hessian matrix.

3.7.5. Estimation of daily reference evapotranspiration

This study is performed using weather data from five meteorological stations (Calakmul, Campeche, Efraín Hernández, Mérida and Tatankín). These data have been used for training and testing the models. The main aim of this study is to investigate the capability of three soft – computing techniques namely ANFIS, ANN and SVM – rbf for ET_0 modelling. To achieve this objective, three input combinations of the daily minimum air temperature ($^{\circ}\text{C}$), maximum air temperature ($^{\circ}\text{C}$), mean relative humidity (%), extraterrestrial radiation ($\text{MJ m}^{-2} \text{ day}^{-1}$), and rainy days as binary number [(RT), $P>0$, $RT=1$; $P=0$, $RT=0$], were used as input to the ANFIS, ANN and SVM – rbf models to estimate ET_0 given by the FAO56-PM equation. The H_0 was calculated as a function of the day of year, site latitude and solar angle, according to the equation proposed by Allen *et al.* (1998).

Three scenarios were evaluated with $T_{\max}$, $T_{\min}$, and H_0 as the base data set:

- (1) Relative humidity-based: $T_{\min}$, $T_{\max}$, RH and H_0 (SVM – rbf 1, ANFIS 1 and ANN 1)
- (2) Rainfall-based: $T_{\min}$, $T_{\max}$, RT and H_0 (SVM – rbf 2, ANFIS 2 and ANN 2)

(3) Temperature-based: $T_{\min}$, $T_{\max}$ and H_0 (SVM – rbf 3, ANFIS 3 and ANN 3)

Table 3.4. Shows the ANN features for each model and station. Also, the optimal SVM parameters (C , γ , and ε) obtained by the GA-based are presented in table 3.5. These parameters were used to train the SVM – rbf models.

Table 3.4. Features of the ANN models.

Station/Model ID	Activation functions		Number of hidden layer units
	Hidden layer	Output layer	
Calakmul			
ANN 1	Logsig	Pureline	23
ANN 2	Logsig	Pureline	53
ANN 3	Logsig	Pureline	35
Campeche			
ANN 1	Logsig	Pureline	10
ANN 2	Logsig	Pureline	22
ANN 3	Logsig	Pureline	23
Efraín Hernández			
ANN 1	Logsig	Pureline	30
ANN 2	Logsig	Pureline	46
ANN 3	Logsig	Pureline	30
Mérida			
ANN 1	Logsig	Pureline	10
ANN 2	Logsig	Pureline	46
ANN 3	Logsig	Pureline	21
Tantakín			
ANN 1	Logsig	Pureline	3
ANN 2	Logsig	Pureline	11
ANN 3	Logsig	Pureline	22

Table 3.5. Optimal SVM – rbf parameters obtained by the GA-based.

Station/Model ID	Optimum values		
	C	γ	ε
Calakmul			
SVM – rbf 1	1.957	0.196	0.318
SVM – rbf 2	2.170	0.452	0.305
SVM – rbf 3	1.135	0.668	0.350
Campeche			
SVM – rbf 1	4.774	0.334	0.299
SVM – rbf 2	2.911	0.114	0.420
SVM – rbf 3	5.412	0.136	0.472
Efraín Hernández			
SVM – rbf 1	1.488	0.381	0.244
SVM – rbf 2	3.617	0.422	0.356
SVM – rbf 3	1.963	0.551	0.299
Mérida			
SVM – rbf 1	4.585	0.108	0.291
SVM – rbf 2	2.263	0.227	0.413
SVM – rbf 3	7.250	0.598	0.402
Tantakín			
SVM – rbf 1	9.315	0.148	0.252
SVM – rbf 2	5.835	0.110	0.404
SVM – rbf 3	5.966	0.200	0.509

3.7.6. Estimation of daily global solar radiation

In the Yucatán Peninsula, several studies have been carried out to estimate solar radiation by using empirical or semi-empirical equations based on available meteorological parameters. However, there have been no studies that have estimated global solar radiation with soft computing techniques and where the precipitation factor in humid climates may play an important role in the modulation of the solar radiation produced by abundant cloud cover. In the present work, ANFIS, SVM and ANN techniques are used to predict daily global solar radiation in six cities of the Yucatán Peninsula (Calakmul, Campeche, Celestún, Efraín

Hernández, Mérida and Tatankín).

For all three techniques, the H variable was defined as the dependent variable. The daily minimum temperature ($T_{\min}$), daily maximum temperature ($T_{\max}$), rainy days [(RT) defined as a binary variable (i.e., rainfall>0, RT=1; rainfall=0, RT=0)], and daily extraterrestrial radiation (H_0), were used as input to the ANFIS, ANN and SVM models for H modelling. The H_0 value was calculated as a function of the day of year, site latitude and solar angle according to the equation proposed by Allen et al. (1998).

Two scenarios were evaluated in this study by considering the actual $T_{\min}$, $T_{\max}$, RT and H_0 : (1) $T_{\min}$, $T_{\max}$, W , and H_0 were used for the SVM 1, ANFIS 1 and ANN 1 models; (2) $T_{\min}$, $T_{\max}$ and H_0 were used for the SVM 2, ANFIS 2 and ANN 2 models. The rainfall parameter (RT) was included in scenario 1, but not in scenario 2 in order to evaluate its effect on the models. The H measured was used as target values for all models.

Table 3.6 shows the ANN features for each model and station. Also, the optimum values of C , γ and ε , are given in the table 3.7 for all meteorological stations.

Table 3.6. Features of the ANN models.

Station/Model ID	Activation functions		Number of hidden layer units
	Hidden layer	Output layer	
Campeche			
ANN 1	Logsig	Pureline	13
ANN 2	Logsig	Pureline	12
Calakmul			
ANN 1	Logsig	Pureline	37
ANN 2	Logsig	Pureline	26
Celestún			
ANN 1	Logsig	Pureline	27
ANN 2	Logsig	Pureline	52
Efraín Hernández			

ANN 1	Logsig	Pureline	19
ANN 2	Logsig	Pureline	24
Tantakín			
ANN 1	Logsig	Pureline	18
ANN 2	Logsig	Pureline	32

Table 3.7. Optimum SVM parameters obtained by the GA-based.

Station/Model ID	Optimum values		
	C	γ	ε
Campeche			
SVM – rbf 1	1.423	0.532	0.407
SVM – rbf 2	2.779	0.147	0.495
Calakmul			
SVM – rbf 1	1.162	0.358	0.342
SVM – rbf 2	2.293	0.465	0.474
Celestún			
SVM – rbf 1	1.084	0.409	0.533
SVM – rbf 2	3.662	0.518	0.519
Efraín Hernández			
SVM – rbf 1	3.675	0.235	0.287
SVM – rbf 2	3.295	0.466	0.512
Mérida			
SVM – rbf 1	6.514	0.046	0.437
SVM – rbf 2	1.308	0.702	0.423
Tantakín			
SVM – rbf 1	9.747	0.132	0.453
SVM – rbf 2	5.826	0.128	0.275

3.7.7. Models performance evaluation

The performance of all Soft-computing models were evaluated using the followings statistical test: root mean square error (equation 3.3), mean absolute error (equation 3.4) and coefficient of determination (equation 3.6).

4. RESULTS AND DISCUSSIONS.

4.1. Temperature – based Methods for estimating daily reference evapotranspiration.

In this study, FAO56-PM, PMT and seven (calibrated and uncalibrated) versions of TET models are calculated for the Campeche, Efraín Hernández, Mérida and Tantakín weather stations. Table 4.1 shows the statistical summary (for each station) of the comparison between the seven TET (calibrated/uncalibrated), the PMT models; and the standardized reference FAO56-PM method. All the methods were calibrated for each location by nonlinear regression techniques with respect to FAO56 –PM.

Table 4.1. Statistical indicators for the ET_0 estimation in four study stations (the model with the best-fit is shown in bold). The R^2 forced to the origin, the MAE, RMSE values expressed in mm day^{-1} , and MAPE expressed in %. a and b are the regression coefficients used for calibrated methods.

Model	R^2	MAE (mm day^{-1})	MAPE (%)	RMSE (mm day^{-1})	a	b
Location: Campeche						
TH	0.52	0.80	22.94	1.02		
TH calibrated	0.52	0.69	19.10	0.89	14.37	
HA	0.55	0.66	18.39	0.85		
HA calibrated	0.55	0.66	18.49	0.85	0.166	
PA	0.43	1.02	27.77	1.26		
PA calibrated	0.43	0.76	19.51	0.96	4.75	
MA	0.47	0.85	26.11	1.07		
MA calibrated	0.47	0.74	20.95	0.92	3.65	
HS	0.69	0.68	19.55	0.86		
HS calibrated	0.69	0.55	14.77	0.70	0.00207	
CA	0.61	0.86	23.46	1.05		
CA calibrated	0.61	0.62	16.44	0.80	40.3	0.219

OU	0.58	0.61	17.82	0.84	
OU calibrated	0.58	0.61	17.08	0.82	0.01
PMT	0.80	0.46	11.88	0.58	
Location: Efraín Hernández					
TH	0.41	0.88	30.52	1.10	
TH calibrated	0.41	0.69	22.54	0.86	13.58
HA	0.43	0.70	25.59	0.90	
HA calibrated	0.43	0.68	23.07	0.85	0.153
PA	0.47	1.14	36.14	1.37	
PA calibrated	0.47	0.65	18.94	0.82	4.37
MA	0.37	1.00	36.96	1.23	
MA calibrated	0.37	0.74	24.76	0.89	3.35
HS	0.67	0.99	33.02	1.15	
HS calibrated	0.67	0.51	16.34	0.65	0.00184
CA	0.62	0.98	31.74	1.17	
CA calibrated	0.62	0.56	16.95	0.69	39.26 0.214
OU	0.45	0.76	28.72	1.04	
OU calibrated	0.45	0.66	22.47	0.83	0.01
PMT	0.66	0.51	16.35	0.65	
Location: Mérida					
TH	0.47	0.90	23.40	1.11	
TH calibrated	0.47	0.84	20.99	1.05	14.88
HA	0.49	0.82	19.63	1.05	
HA calibrated	0.49	0.81	20.40	1.03	0.174
PA	0.60	1.04	26.96	1.27	
PA calibrated	0.60	0.74	17.91	0.91	4.75
MA	0.42	0.94	25.57	1.15	
MA calibrated	0.42	0.88	22.41	1.10	3.81
HS	0.69	0.67	17.59	0.87	
HS calibrated	0.69	0.62	15.52	0.80	0.00215
CA	0.68	0.86	22.57	1.06	
CA calibrated	0.68	0.64	16.00	0.81	40.46 0.219
OU	0.50	0.80	19.42	1.02	
OU calibrated	0.50	0.79	19.77	1.02	0.01
PMT	0.70	0.64	16.43	0.81	
Location: Tantakín					

TH	0.38	1.00	28.71	1.21		
TH calibrated	0.38	0.87	23.48	1.10	14.29	
HA	0.41	0.84	23.45	1.07		
HA calibrated	0.41	0.84	23.19	1.07	0.164	
PA	0.57	2.15	60.00	2.35		
PA calibrated	0.57	0.71	18.91	0.91	3.74	
MA	0.34	1.05	32.57	1.27		
MA calibrated	0.34	0.89	25.16	1.13	3.6	
HS	0.65	1.36	39.19	1.52		
HS calibrated	0.65	0.63	16.95	0.83	0.00177	
CA	0.66	1.71	47.79	1.89		
CA calibrated	0.66	0.63	16.71	0.82	37.64	0.207
OU	0.45	0.84	24.21	1.07		
OU calibrated	0.45	0.79	21.71	1.03	0.01	
PMT	0.87	0.62	18.12	0.77		
<hr/>						
All locations						
TH	0.44	0.89	26.39	1.11		
TH calibrated	0.44	0.77	21.53	0.97		
HA	0.47	0.76	21.76	0.97		
HA calibrated	0.47	0.75	21.29	0.95		
PA	0.52	1.34	37.72	1.56		
PA calibrated	0.52	0.71	18.82	0.90		
MA	0.40	0.96	30.30	1.18		
MA calibrated	0.40	0.81	23.32	1.01		
HS	0.68	0.92	27.34	1.10		
HS calibrated	0.68	0.58	15.90	0.74		
CA	0.64	1.10	31.39	1.29		
CA calibrated	0.64	0.61	16.53	0.78		
OU	0.50	0.75	22.54	0.99		
OU calibrated	0.50	0.71	20.26	0.92		
PMT	0.76	0.56	15.70	0.70		

Analysis of the statistics of the eight methods shows that the proposed PMT has the best performance (Table 4.1); followed by the HS model. The PA and MA

models had the worst fit. In general, the proposed PMT model provided estimates of ET_0 that were in good agreement with PM ET_0 estimates. In terms of R^2 values, the stations with the best and worst PMT estimates are Tantakín (0.876) (in the middle part) and Efraín Hernández (0.66) (in the southern part of Yucatán Peninsula).

The R^2 , MAE, MAPE and RMSE values reveal large differences between the TET models. In general, the TET calibrated models showed little improvement over uncalibrated models. Considering the joint analysis of the four seasons considered, the scatter of the R^2 obtained values from 0.34 (MA model) to 0.87 (PMT model). Regarding MAE values, these varying from 1.34 mm day^{-1} (PA) to $0.750 \text{ mm day}^{-1}$ (OU) in the uncalibrated models; and from 0.81 mm day^{-1} (MA) to 0.58 mm day^{-1} (HS) for the calibrated models (Figure 4.1). In overall, the model PA and MA provides the worst results among all methods, and the HS model was ranked the top among the TET models. In figure 4.1 the values of R^2 and MAE are plotted to facilitate comparison. Clearly, the PMT showed the best performance (low MAE and High R^2), but the calibrated HS and calibrated CA performed nearly as well.

Figure 4.1. The MAE and R^2 comparisons with the PM estimated ET_0 by model.

Under conditions of missing or reliable meteorological data of solar radiation, relative humidity, and wind speed, the use of TET models is necessary for ET_0 estimation. In the literature, preferences have been directed toward the

Hargreaves–Samani method (Pereira *et al.*, 2015). In general, the TET models performed better when calibrated for local climatic conditions (Almorox and Grieser, 2015). In the four stations HS calibrated formulation, the HS coefficient differs from 0.00177 (Tantakín) to 0.00215 (Mérida), against an original value of 0.0023. The performance HS model clearly improved with calibration.

At stations where no wind or global solar radiation data (or no sunshine data) are available, the use of PMT methodology appears to be well suited. The RMSE values obtained in this study are similar to those achieved by other researchers using PMT equation in semi-arid and sub-humid climatic zones. Jabloun and Sahli (2008) reported RMSE 0.41-0.80 mm day⁻¹ for eight locations of Tunisia. Jabloun and Sahli (2008) obtained better performance in PMT model than Hargreaves method although they did not calibrated the Hargreaves method. Todorovic *et al.* (2013) used data from 577 weather station around Mediterranean areas and they found that the PMT method produce better results than HS uncalibrated, with RMSE smaller than 0.52 mm day⁻¹; Martinez and Thepadia (2009) evaluated monthly temperature data from Florida and they found values of RMSE for PMT method ranged from 0.40 to 1.26 mm day⁻¹ with a mean of 0.70 mm day⁻¹. These authors showed values ranged from 0.29 to 1.31 mm day⁻¹ for Hargreaves equation. In Serbia, Trajkovic (2005) reported that the PMT method had lower RMSE (0.27 mm day⁻¹) than HS method (0.36 mm day⁻¹) for monthly estimates. Raziei and Pereira (2013) for regions humid and moist sub-humid of Iran found RMSE=0.3 mm day⁻¹ for PMT and 0.31 mm day⁻¹ for a calibrated HS method using monthly averages. Ren *et al.* (2016) showed values of RMSE between 0.44-0.58 mm day⁻¹ and 0.61-0.46 mm day⁻¹ for PMT and Hargreaves methods in Moist sub humid region of Inner Mongolia China.

This study shows that the best estimation of PM ET₀ is given by PMT formulation. The PMT produced smaller ET₀ errors and improved over the calibrated HS method. Results confirm that the PMT model gives acceptable results and, with the current software, this formulation can easily be used. The applied PMT is not a perfect alternate for PM ET₀ calculations, but more accurate ET₀ estimations are obtained using this model in the Yucatán Peninsula rather than with any other calibrated TET formulations. The PMT calculation process has the advantage of following the PM philosophy and it can be easily implemented by using new generation computers. In locations where the air

humidity, solar radiation and wind speed play a major role affecting PM, the use of PMT or other TET models might be inappropriate (Raziei and Pereira, 2013; Ren *et al.*, 2016).

4.2. Empirical models for estimating daily global solar radiation

In this analysis, thirteen empirical models were evaluated for their suitability for estimating global solar radiation at six sites on the Yucatán Peninsula. The empirical formulations were suitable and fairly straightforward tools to use with commonly available meteorological variables. However, the coefficients of the empirical formulas require calibration to achieve a good fit.

All tested models performed well in estimating global solar radiation (table 4.2). Overall, the newly developed model M5 performed best for 5 of the 6 sites where RT and RH data were included. For this model, R^2 was between 0.706 and 0.620 for the six sites, RMSE values were between 2.3 to 2.94 MJ m⁻² day⁻¹, and MBE varied between -0.054 and 0.303 MJ m⁻² day⁻¹ (table 4.2). Negative values of MBE indicate that the model slightly overestimates H. The relationship between the measured and estimated solar radiation for six weather stations using the new model M5 is shown in figure 4.2.

Model M3 had lower performance compared to the M5 model due to the incorporation of D and RT values. Models, M4 and M2, used RT and three RT consecutive days, respectively, had similar performance, and Model M1, which used P values, was the weakest when compared with all models using RT data. The models that use air temperature as the sole parameter had ranges for R^2 , RMSE and MBE of 0.47 to 0.631, 2.7 to 3.46 MJm⁻² day⁻¹, and -0.35 to 0.64 MJ m⁻² day⁻¹, respectively. Of these temperature-based models, Models T2 and T8 performed best, whereas Models T5 and T7 had the worst performance. The models based on Bristow and Campbell equation (T2, T3, and T4) performed slightly better than those based on the Hargreaves –Samani model (T1). The correction for altitude applied in the model T6 had no effect model performance.

In general, temperature-based models were less accurate compared to models that used additional meteorological data. Model T2 is the best performing temperature-based model for estimating global solar radiation; the equations for this model could be used when measured data exist for calibration of coefficients. Model T1 is simple to use and could also be used to estimate solar global radiation with relatively high accuracy. These models are recommended for use to estimate the daily global radiation when only temperature data are available and model coefficients cannot be determined directly from available data or estimation.

When more meteorological data are available, the estimation of solar radiation using temperature, precipitation and relative humidity could be used to estimate daily solar radiation with higher accuracy compared to the solely temperature-based models. However, such models are more complex and require more data. The newly developed equation (model M5) in this study can be useful to provide global solar radiation data in the Yucatán Peninsula with sufficient accuracy even if information on relative humidity is lacking. The validity of the M5 global solar radiation regression relation should be tested and calibrated on other sites (see Table 4.3).

Table 4.2. Error statistics of the thirteen empirical models for the estimation of daily global solar radiation in the six sites of Yucatán, Peninsula.

Station/Model ID		Errors statistics				
Calakmul	R ²	RMSE	MBE	MAE	MPE	MAPE
T1	0.603	2.522	-0.351	1.948	-6.850	16.078
T2	0.607	2.513	0.427	1.940	-0.439	14.869
T3	0.575	2.580	0.024	1.968	-3.205	15.599
T4	0.562	2.639	0.203	2.067	-2.934	16.234
T5	0.437	3.241	-0.185	2.398	-4.744	19.179
T6	0.603	2.672	0.861	2.102	2.135	15.578
T7	0.589	2.538	-0.029	1.991	-4.405	15.884
T8	0.599	2.516	0.266	1.947	-1.724	15.127
M1	0.637	2.381	0.118	1.827	-2.449	14.289
M2	0.636	2.382	0.015	1.824	-3.218	14.345

M3	0.660	2.306	0.092	1.769	-2.548	13.830
M4	0.650	2.337	0.011	1.787	-3.007	13.993
M5	0.661	2.300	-0.054	1.755	-3.506	13.847
Campeche	R²	RMSE	MBE	MAE	MPE	MAPE
T1	0.542	3.222	0.184	2.427	-3.299	15.070
T2	0.547	3.179	0.080	2.376	-4.154	15.133
T3	0.538	3.213	0.077	2.415	-3.989	15.165
T4	0.541	3.212	-0.001	2.413	-5.071	15.405
T5	0.530	3.240	0.109	2.437	-4.239	15.692
T6	0.542	3.222	0.184	2.427	-3.299	15.070
T7	0.507	3.311	-0.001	2.512	-5.054	15.976
T8	0.549	3.199	0.449	2.451	-2.242	15.183
M1	0.594	3.011	0.130	2.300	-3.274	14.001
M2	0.620	2.908	0.041	2.179	-3.839	13.819
M3	0.611	2.943	0.020	2.228	-3.753	13.881
M4	0.615	2.928	0.002	2.195	-4.179	13.972
M5	0.620	2.907	0.006	2.180	-4.000	13.844
Celestún	R²	RMSE	MBE	MAE	MPE	MAPE
T1	0.492	3.604	0.114	2.747	-2.344	15.412
T2	0.499	3.393	0.292	2.588	-2.484	15.070
T3	0.486	3.454	0.122	2.626	-2.942	15.094
T4	0.506	3.353	0.187	2.571	-3.313	14.827
T5	0.486	3.477	0.216	2.640	-2.735	15.485
T6	0.492	3.604	0.109	2.747	-2.372	15.415
T7	0.470	3.460	-0.001	2.648	-4.443	15.517
T8	0.505	3.382	0.337	2.601	-2.071	14.972
M1	0.570	3.118	-0.034	2.399	-3.542	13.555
M2	0.605	2.988	-0.071	2.231	-3.750	13.078
M3	0.609	2.971	0.038	2.265	-2.986	13.021
M4	0.599	3.013	0.027	2.257	-3.180	13.149
M5	0.620	2.946	0.047	2.218	-2.704	12.864
Efraín Hernández	R²	RMSE	MBE	MAE	MPE	MAPE
T1	0.570	3.465	-0.145	2.747	-8.912	20.576
T2	0.606	3.260	0.035	2.568	-5.465	17.823
T3	0.626	3.178	0.006	2.488	-4.796	17.084
T4	0.608	3.258	-0.078	2.567	-6.078	17.747
T5	0.530	3.608	0.168	2.893	-7.531	21.401
T6	0.570	3.465	-0.145	2.747	-8.912	20.576
T7	0.578	3.392	-0.094	2.698	-7.470	19.093

T8	0.625	3.179	-0.018	2.489	-5.282	17.150
M1	0.656	3.042	0.001	2.412	-4.843	16.530
M2	0.683	2.922	-0.027	2.266	-4.913	15.716
M3	0.685	3.040	-0.543	2.302	-6.845	15.396
M4	0.666	3.002	0.012	2.356	-4.880	16.399
M5	0.689	2.904	0.061	2.277	-4.447	15.904
Mérida	R²	RMSE	MBE	MAE	MPE	MAPE
T1	0.624	3.011	0.636	2.366	-0.098	14.096
T2	0.630	2.980	0.584	2.339	-0.320	13.934
T3	0.629	2.939	0.261	2.278	-1.755	13.729
T4	0.630	2.949	-0.284	2.255	-5.352	14.126
T5	0.614	2.983	0.089	2.285	-3.281	14.344
T6	0.624	3.009	0.627	2.363	-0.147	14.090
T7	0.598	3.047	-0.028	2.365	-4.277	14.802
T8	0.631	2.926	0.229	2.266	-2.278	13.837
M1	0.664	2.782	0.008	2.151	-3.106	13.105
M2	0.695	2.669	0.303	2.052	-1.334	12.436
M3	0.692	2.665	-0.029	2.055	-3.192	12.604
M4	0.690	2.673	0.007	2.035	-3.111	12.613
M5	0.706	2.604	-0.014	1.981	-3.175	12.327
Tantakín	R²	RMSE	MBE	MAE	MPE	MAPE
T1	0.622	2.806	0.198	2.205	-2.710	14.102
T2	0.631	2.763	0.118	2.167	-2.768	13.743
T3	0.629	2.776	0.053	2.148	-2.615	13.499
T4	0.618	2.831	0.120	2.205	-3.166	14.029
T5	0.622	2.798	0.051	2.191	-3.039	13.936
T6	0.622	2.805	0.183	2.203	-2.793	14.104
T7	0.602	2.867	0.000	2.269	-3.677	14.423
T8	0.630	2.768	0.182	2.167	-2.384	13.669
M1	0.658	2.656	0.000	2.087	-3.052	13.090
M2	0.657	2.660	0.048	2.059	-2.835	13.027
M3	0.673	2.598	0.061	2.035	-2.770	12.880
M4	0.661	2.653	0.041	2.074	-2.680	13.046
M5	0.674	2.597	0.017	2.028	-2.856	12.895

Table 4.3. Calibrated coefficients for each empirical model.

Station/Model							
ID	Coefficients						
Calakmul	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
T1	0.120						
T2	0.740	0.077	0.930				
T3	0.480	0.400	40.300				
T4	0.490	1.160	1.610				
T5	1.102						
T6	0.110						
T7	0.084	0.613	1.163				
T8	0.140	0.840	0.580	0.420			
M1	-4.050	0.108	0.163	-0.132	0.001		
M2	1.134	0.109	0.634	0.003	-0.012	-0.012	-1.450
M3	0.095	0.015	-0.045	2.290	0.530		
M4	-0.099	0.121	0.004	-0.049			
M5	0.155	-0.869	-0.004	-1.149			
Campeche	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
T1	0.176						
T2	0.840	0.160	0.830				
T3	0.643	1.076	809.095				
T4	0.660	2.130	1.540				
T5	0.656						
T6	0.176						
T7	0.204	0.398	1.183				
T8	0.277	0.547	1.089	0.377			
M1	2.500	0.160	-0.008	-0.151	0.000		
M2	0.723	0.138	1.193	-0.024	-0.114	-0.018	-1.113
M3	0.114	0.174	-0.114	1.523	0.497		
M4	0.193	0.117	0.001	-0.083			
M5	0.146	-0.706	-0.002	-2.612			
Celestún	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
T1	0.170						
T2	0.820	0.330	0.630				
T3	0.636	1.910	126.200				

T4	0.690	7.270	1.060				
T5	0.553						
T6	0.200						
T7	0.497	0.285	1.011				
T8	0.460	0.170	1.520	-0.760			
M1	5.900	0.152	-0.014	-0.411	0.004		
M2	0.736	0.358	0.695	-0.043	-0.173	-0.004	1.350
M3	0.092	0.204	-0.207	2.720	2.438		
M4	0.378	0.101	-0.002	-0.120			
M5	0.124	-0.496	-0.003	-3.558			

Efraín

Hernández	a	b	c	d	e	f	g
T1	0.163						
T2	0.712	0.023	1.731				
T3	0.686	0.389	45.018				
T4	0.652	0.426	2.117				
T5	0.780						
T6	0.163						
T7	0.096	0.729	1.139				
T8	-0.253	1.449	-0.131	1.091			
M1	-6.629	0.156	0.253	-0.214	0.002		
M2	0.674	0.022	1.837	0.003	-0.115	-0.051	0.817
M3	0.135	-0.149	-0.021	9.461	2.759		
M4	-0.244	0.197	0.006	-0.081			
M5	0.244	-0.840	-0.005	-1.747			

Mérida	a	b	c	d	e	f	g
T1	0.160						
T2	0.740	0.096	1.100				
T3	0.630	0.690	121.700				
T4	0.640	0.650	2.000				
T5	0.730						
T6	0.160						
T7	0.172	0.457	1.172				
T8	0.250	0.330	0.600	-2.960			
M1	-0.095	0.158	0.038	-0.130	0.000		
M2	0.770	0.150	1.000	-0.005	-0.123	-0.025	-1.650

M3	0.096	0.187	-0.059	3.100	-0.842		
M4	0.088	0.125	0.003	-0.083			
M5	0.156	-0.660	-0.004	-1.979			
Tantakín	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
T1	0.140						
T2	0.800	0.062	1.080				
T3	0.634	0.296	46.430				
T4	0.620	0.530	1.860				
T5	0.931						
T6	0.140						
T7	0.125	0.566	1.133				
T8	0.177	0.740	0.693	0.283			
M1	-4.890	0.122	0.227	-0.179	0.001		
M2	0.670	0.039	1.429	-0.007	-0.083	-0.014	0.150
M3	0.125	-0.020	-0.065	2.210	1.280		
M4	-0.116	0.150	0.004	-0.054			
M5	0.184	-0.936	-0.003	-1.354

Figure 4.2. Scatter plots of the measured and estimated solar radiation in six sites in Yucatán, Peninsula, Mexico for model M5.

4.3. Estimating daily global solar radiation by day of the year

4.3.1. Overall model performance

Table 4.4 shows the performance metrics of the different models in each location. As can be seen, among the DYB – models, Model 5 had the best performance for all locations according to overall performance metrics (i.e., average performance metrics for all locations: $R^2=0.868$, $RMSE=1.191 \text{ MJ m}^{-2}\text{day}^{-1}$, $MBE=0.006 \text{ MJ m}^{-2}\text{day}^{-1}$, $MAE=0.928 \text{ MJ m}^{-2}\text{day}^{-1}$, $MPE=-0.385 \%$ and $MAPE=5.09 \%$). Model 4 had the second best overall performance ($R^2=0.850$, $RMSE=1.273 \text{ MJ m}^{-2}\text{day}^{-1}$, $MBE=0.005 \text{ MJ m}^{-2}\text{day}^{-1}$, $MAE=0.982 \text{ MJ m}^{-2}\text{day}^{-1}$, $MPE=-0.429 \%$ and $MAPE=5.377 \%$). Models 1 and 2 had the worst performance when assessed by the statistical indices (R^2 , $RMSE$, MBE , MAE , MPE and $MAPE$). In terms of R^2 and $RMSE$ indices, the locations with the best and worst values are Campeche (0.892 and $1.036 \text{ MJ m}^{-2}\text{day}^{-1}$, respectively) and Mérida (0.831 - $1.449 \text{ MJ m}^{-2}\text{day}^{-1}$, respectively). The nonlinear regression coefficients of the models are presented in table 4.5.

Table 4.4. Performance metrics of the five empirical models in six selected cities in Yucatán Peninsula, Mexico. The model with the best performance for each location is shown in italics.

Station/Model		Errors statistics				
Calakmul	R^2	RMSE	MBE	MAE	MPE	MAPE
1	0.633	1.607	0.000	1.274	-1.293	9.019
2	0.713	1.423	0.000	1.140	-1.066	8.196
3	0.700	1.454	0.012	1.194	-0.932	8.789
4	0.856	1.007	0.000	0.802	-0.511	5.861
5	<i>0.863</i>	<i>0.984</i>	<i>0.000</i>	<i>0.777</i>	<i>-0.509</i>	<i>5.660</i>
Cancún	R^2	RMSE	MBE	MAE	MPE	MAPE
1	0.778	1.757	0.038	1.359	-0.628	7.150
2	0.795	2.051	-1.144	1.573	-7.286	9.174
3	0.824	1.560	0.006	1.225	-0.600	6.605

4	0.843	1.475	0.034	1.099	-0.335	5.822
5	0.880	1.292	0.000	1.003	-0.469	5.366
Celestún	R²	RMSE	MBE	MAE	MPE	MAPE
1	0.697	1.816	0.000	1.462	-0.798	7.074
2	0.772	1.578	0.000	1.278	-0.642	6.319
3	0.802	1.470	0.003	1.184	-0.453	5.855
4	0.869	1.195	0.000	0.932	-0.335	4.626
5	0.884	1.122	0.000	0.877	-0.308	4.345
Chetumal	R²	RMSE	MBE	MAE	MPE	MAPE
1	0.553	2.197	0.000	1.772	-1.214	8.874
2	0.682	1.853	0.000	1.502	-0.906	7.685
3	0.703	1.793	0.009	1.450	-0.699	7.387
4	0.823	1.383	0.000	1.094	-0.476	5.643
5	0.849	1.278	0.000	1.005	-0.426	5.165
Mérida	R²	RMSE	MBE	MAE	MPE	MAPE
1	0.714	1.946	0.000	1.511	-1.070	8.011
2	0.767	1.756	0.000	1.339	-0.918	7.137
3	0.784	1.693	0.011	1.301	-0.652	6.965
4	0.820	1.545	0.000	1.178	-0.644	6.264
5	0.831	1.499	0.038	1.159	-0.350	6.167
Campeche	R²	RMSE	MBE	MAE	MPE	MAPE
1	0.761	1.573	0.037	1.212	-0.188	5.978
2	0.785	1.459	0.000	1.148	-0.601	5.866
3	0.827	1.310	0.001	1.001	-0.412	5.096
4	0.892	1.036	0.000	0.785	-0.274	4.049
5	0.904	0.975	0.000	0.751	-0.250	3.857

Table 4.5. Regression coefficients for the selected models in six cities of Yucatán, Peninsula, Mexico.

Station/Model		Coefficients					
Calakmul	a	b	c	d	e	f	g
1	10.565	6.235					
2	14.041	3.167	-15.217				
3	16.620	163.039	-176.730				

4	14.442	1.632	-2.206	42.746	-2.872	1.137	0.173
5	10.269	7.633	109.048	41.191	5.355	226.989	43.932
Cancún	a	b	c	d	e	f	g
1	13.766	9.867					
2	20.450	4.370	-9.080				
3	23.290	166.670	166.260				
4	2.229	27.692	0.258	12.199	-20.202	-0.436	9.684
5	13.344	8.625	97.143	40.092	9.407	211.286	55.796
Celestún	a	b	c	d	e	f	g
1	16.207	8.133					
2	20.743	4.097	3.615				
3	24.108	160.381	189.166				
4	20.838	1.505	2.042	-1.396	4.008	1.022	3.569
5	15.321	8.901	99.822	43.203	7.515	217.625	54.421
Chetumal	a	b	c	d	e	f	g
1	15.903	7.217					
2	19.927	3.839	-2.538				
3	22.990	154.560	196.260				
4	20.108	-1.827	2.098	-4.658	-3.650	1.040	6.811
5	15.158	9.502	101.605	41.624	6.601	222.054	45.593
Mérida	a	b	c	d	e	f	g
1	14.237	9.072					
2	19.296	4.503	-9.015				
3	23.040	162.540	172.197				
4	19.230	-1.081	2.131	-4.781	-4.403	0.972	6.786
5	13.719	8.396	99.369	42.394	8.296	210.814	55.348
Campeche	a	b	c	d	e	f	g
1	15.040	9.050					
2	20.123	3.939	-9.050				
3	23.431	165.999	186.465				
4	20.351	-1.674	1.990	-4.439	-3.908	1.075	6.456
5	14.533	8.421	100.561	43.291	7.720	220.373	58.924

Figure 4.3 shows the long term daily measured global solar radiation data, the measured mean daily values and Model 5 predictions for the six locations on the

Yucatán, Peninsula. It is evident that the deviation between the measured and calculated values is very small. As expected, the maximum global solar radiation occurs in all cases in months with maximum daylight hours (i.e., May, June, July and August). In contrast, the solar radiation is lower in December and January. During winter, the daily global solar radiation reaches the Earth surface at a greater angle, which significantly reduces the energy per unit area and the energy distribution on a horizontal surface. During summer months, the sun's path at noon is near the zenith and thus, the solar radiation is greater.

Due to the seasonal variation of the daily global solar radiation throughout the year, the largest amounts of solar radiation reach most of the Yucatán Peninsula from April to September, with maximum values in April and August. The daily global solar radiation in June is lower compared to that in April and May likely due to the rainy season that occurs between June and October. During the rainy season, the global solar radiation is strongly modulated by clouds and absorption of the atmospheric water vapor (Galindo *et al.*, 2009; Stephens *et al.*, 2012).

Figure 4.3. Comparison between daily measured global solar radiation data (measured and daily mean of long-term measured data) and values estimated by proposed Model 5 for six locations on the Yucatán Peninsula.

4.3.2. Seasonal analysis of the proposed model

To investigate the performance of the best-performing model (Model 5) in different seasons, the standard meteorological season scheme of Trenberth (1983) was applied (i.e., winter is defined as December, January, and February; spring is March, April, and May; summer is June, July and August; and autumn is September, October and November). Trenberth (1983) analyzed the first harmonic of solar radiation at the top of the atmosphere and its relationship to surface temperature to show that use of the standard meteorological season scheme is more appropriate than the astronomical season scheme over the continental regions of the Northern Hemisphere. The RMSE, MBE and MAPE indicators are used to assess statistical model errors during seasons. These statistical errors are presented in table 4.6. In general, the model performed best in the winter season, with RMSE values that range between $0.795 - 1.113 \text{ MJ m}^{-2}\text{day}^{-1}$. Model performance was lowest during summer with RMSE values ranging between $1.154 - 1.944 \text{ MJ m}^{-2}\text{day}^{-1}$. This low performance is likely related to low values of solar radiation caused by precipitation events during the summer season. Reduced model performance is also observed during the autumn season, mainly during the months of September and October. Moreover, the model tends to slightly overestimate values of solar radiation throughout autumn with a range in MBE of $1.624 - 3.173 \text{ MJ m}^{-2}\text{day}^{-1}$. The MAPE measures the size of the error in percentage terms, with preferred values between $\pm 10\%$ (Li *et al.*, 2010; Khorasanizadeh and Mohammadi, 2013). In general, seasonal MAPE values are in the desired range for all locations. During the summer and autumn, the MAPE values are slightly larger than during the spring. In general, Model 5 estimates solar radiation with good performance, but performance varies by season.

Table 4.6. RMSE, MBE and MAPE values obtained with Model 5 during the standard meteorological season.

Location	Winter	Spring	Summer	Autumn
Calakmul				
RMSE	0.795	0.984	1.154	0.958
MBE	0.032	0.064	-0.002	1.667

MAPE	5.610	4.886	5.854	6.194
Cancún				
RMSE	0.948	1.153	1.683	1.264
MBE	-0.001	-0.047	0.030	2.731
MAPE	5.110	4.111	6.055	6.193
Celestún				
RMSE	0.867	0.888	1.457	1.168
MBE	0.022	-0.092	0.077	2.336
MAPE	4.145	2.889	5.247	5.102
Chetumal				
RMSE	0.994	1.270	1.609	1.282
MBE	0.196	0.194	0.167	2.857
MAPE	4.919	4.246	6.153	5.989
Mérida				
RMSE	1.113	1.363	1.944	1.426
MBE	0.054	-0.025	0.023	3.173
MAPE	5.934	5.001	6.974	6.586
Campeche				
RMSE	0.872	0.792	1.236	0.938
MBE	0.016	-0.069	0.037	1.624
MAPE	4.541	2.849	4.277	3.776

4.4. Soft – computing models.

4.4.1. Estimation of daily reference evapotranspiration

In the present study, the ability of three soft-computing techniques was assessed for ET_0 modelling. Measured daily climatic data in five weather stations were used as inputs (T_{max} , T_{min} , RH, RT, H_0) and ET_0 as the output. Table 4.7 gives the R^2 , RMSE and MAE values for ANFIS, ANN and SVM – rbf models during test phase. The first scenario, when relative humidity data is available, the SVM – rbf 1 model had the best performance in four of five evaluated stations (Mérida, Efraín Hernández, Tantakín and Campeche), with an RMSE of 0.371 – 0.479 mm day⁻¹, MAE of 0.285 – 0.372 mm day⁻¹ and R^2 of 0.937 - 0.862. For the Calakmul station, the ANN 1 model outperforms the SVM – rbf 1 and ANFIS 1 models. The ANN 1 and ANFIS 1 models can be ranked as the second and third respectively. For the second scenario, the rainfall effect on the performance models was investigated, since it is usually measured and can be used to improve ET_0 estimation. The SVM – rbf 2 models outperform the ANN 2 and ANFIS 2

models for all locations with an RMSE of 0.494 – 0.700 mm day⁻¹, MAE of 0.379 – 0.534 mm day⁻¹ and R² of 0.834 - 0.758. In the other hand, the performance of ANFIS 2 models is similar to the performance of ANN 2 models. In the third scenario, where only temperature data are available, the SVM 3 model provide the best performance in all stations with RMSE values of 0.548 – 0.731 mm day⁻¹, MAE of 0.418 – 0.563 mm day⁻¹ and R² of 0.779 - 0.738. Secondly, the ANN 3 model provides more accurate estimates than the ANFIS 3 model in most stations. The FAO56-PM ET₀ values and the estimates obtained by the best models for each scenario are plotted in figure 4.4 and 4.5. These plots clearly reveal that the input variable RH is the most influential for ET₀ estimations.

In general, during the test phase, it should be noticed that adding RH data into model SVM – rbf 1 significantly increases the estimation accuracy of 32.73 and 32.71% reduction in MAE and RMSE values respectively and an increase in R² of 14.5%. Therefore, by adding RT data into model SVM – rbf 2, a little increase is obtained in accuracy of 9.06 and 8.87% reduction in MAE and RMSE values respectively and increases in R² of 4.4%.

Table 4.7. Error statistics during testing phase, for five locations in Yucatán, Peninsula, Mexico.

Station/model ID	Error statistics		
Calakmul	R²	RMSE	MAE
ANN 1	0.890	0.432	0.325
SVM 1	0.862	0.447	0.340
ANFIS 1	0.869	0.473	0.355
SVM 2	0.767	0.569	0.435
ANN 2	0.748	0.648	0.488
ANFIS 2	0.759	0.707	0.538
SVM 3	0.746	0.607	0.461
ANN 3	0.713	0.690	0.531
ANFIS 3	0.695	0.765	0.580
Campeche	R²	RMSE	MAE
SVM 1	0.862	0.479	0.372
ANN 1	0.831	0.542	0.399
ANFIS 1	0.807	0.569	0.422

SVM 2	0.793	0.584	0.455
ANFIS 2	0.765	0.659	0.529
ANN 2	0.768	0.662	0.534
SVM 3	0.745	0.652	0.508
ANFIS 3	0.683	0.749	0.598
ANN 3	0.676	0.757	0.605
Efraín Hernández	R²	RMSE	MAE
SVM 1	0.864	0.411	0.315
ANN 1	0.852	0.421	0.329
ANFIS 1	0.846	0.451	0.338
SVM 2	0.805	0.494	0.379
ANFIS 2	0.801	0.512	0.385
ANN 2	0.809	0.516	0.393
SVM 3	0.774	0.548	0.418
ANN 3	0.747	0.564	0.435
ANFIS 3	0.738	0.583	0.450
Mérida	R²	RMSE	MAE
SVM 1	0.937	0.371	0.285
ANN 1	0.924	0.383	0.292
ANFIS 1	0.921	0.391	0.300
SVM 2	0.834	0.580	0.450
ANN 2	0.831	0.575	0.451
ANFIS 2	0.828	0.611	0.467
SVM 3	0.779	0.678	0.521
ANN 3	0.755	0.695	0.523
ANFIS 3	0.749	0.705	0.537
Tantakín	R²	RMSE	MAE
SVM 1	0.901	0.437	0.333
ANFIS 1	0.861	0.485	0.357
ANN 1	0.870	0.508	0.393
SVM 2	0.758	0.700	0.534
ANN 2	0.706	0.727	0.576
ANFIS 2	0.736	0.877	0.730
SVM 3	0.738	0.731	0.563
ANN 3	0.691	0.793	0.648
ANFIS 3	0.695	0.952	0.785

Figure 4.4. The FAO56-PM ET_0 and estimated ET_0 values obtained by the best model in each scenario during testing phase for Calakmul, Campeche and Efraín Hernández stations.

Figure 4.5. The FAO56-PM ET₀ and estimated ET₀ values obtained by the best model in each scenario during testing phase for Mérida and Tantakín stations.

4.4.2. Estimation of daily global solar radiation

SVM – rbf, ANFIS and ANN computing techniques were used for H prediction in six meteorological stations located in the Yucatán Peninsula, Mexico.

To training all ANN models, one input layer with three and four neurons for ANN 1 and ANN 2 were used respectively and 2 to 80 neurons were employed for the hidden layer. The optimum ANFIS architecture to carry out the analysis consists of 8 and 16 fuzzy rules for ANFIS 1 and ANFIS 2 models respectively, the Bell – Shaped function was chosen as the best MF. In the training phase, a hybrid

learning algorithm and Sugeno fuzzy model have been used. The optimum values of C , γ and ε are given in the table 3.6 for all meteorological stations.

The statistical performance measures for each model are presented in Table 4.8 for testing data sets. In the first scenario, when rainfall data are included, the SVM1 model achieved the best performance for all locations in the testing phase according to mean performance statistics ($R^2=0.689$, $RMSE= 2.678 \text{ MJ m}^{-2} \text{ day}^{-1}$ and $MAE= 1.973 \text{ MJ m}^{-2} \text{ day}^{-1}$), followed by ANN1 ($R^2=0.652$, $RMSE=2.775 \text{ MJ m}^{-2} \text{ day}^{-1}$ and $MAE=2.150 \text{ MJ m}^{-2} \text{ day}^{-1}$). The ANFIS1 model had a similar performance to the ANN1 ($R^2=0.645$, $RMSE=2.801 \text{ MJ m}^{-2} \text{ day}^{-1}$, and $MAE=2.153 \text{ MJ m}^{-2} \text{ day}^{-1}$).

In the second scenario, the SVM2 model with inputs of $T_{\min}$, $T_{\max}$, and H_0 performed the best for all locations according to overall mean errors ($R^2=0.624$, $RMSE=2.877 \text{ MJ m}^{-2} \text{ day}^{-1}$, $MAE=2.203 \text{ MJ m}^{-2} \text{ day}^{-1}$), and the ANN2 and ANFIS2 had similar performance measures ($R^2=0.596$, $RMSE=3.023 \text{ MJ m}^{-2} \text{ day}^{-1}$, $MAE=2.352 \text{ MJ m}^{-2} \text{ day}^{-1}$ for ANN2 and $R^2=0.587$, $RMSE=3.052 \text{ MJ m}^{-2} \text{ day}^{-1}$, $MAE=2.365 \text{ MJ m}^{-2} \text{ day}^{-1}$ for ANFIS2). In all cases, the first scenario models performed better than the second scenario models.

Estimated daily solar global radiation values by the SMV1 models for the testing phase are plotted against the measured values for the six meteorological stations in figure 4.6

The MAE and RMSE values found in this study are similar to those found by other authors (Tabari *et al.*, 2012; Mohammadi *et al.*, 2015). The results demonstrate the ability of SVM1 models to adapt to existing conditions in Yucatán.

Thus, according to the statistical indicators, the SVM approach performed better than the ANN and ANFIS models. In addition, the rainfall input parameter had a positive effect on model performance in estimating daily solar radiation for this tropical area.

Table 4.8. The R², RMSE and MAE statistics for each model in testing phases, for six locations in Yucatán, Peninsula, Mexico.

Station/Model ID	Error statistics		
Campeche	R²	RMSE	MAE
SVM – rbf 1	0.663	2.786	2.049
SVM – rbf 2	0.562	3.171	2.326
ANFIS 1	0.652	2.769	2.057
ANFIS 2	0.539	3.188	2.364
ANN 1	0.655	2.758	2.064
ANN 2	0.535	3.206	2.419
Calakmul	R²	RMSE	MAE
SVM – rbf 1	0.698	2.142	1.668
SVM – rbf 2	0.643	2.330	1.809
ANFIS 1	0.636	2.423	1.888
ANFIS 2	0.597	2.549	1.999
ANN 1	0.644	2.406	1.871
ANN 2	0.623	2.477	1.952
Celestún	R²	RMSE	MAE
SVM – rbf 1	0.635	2.827	2.164
SVM – rbf 2	0.517	3.320	2.493
ANFIS 1	0.513	3.381	2.715
ANFIS 2	0.483	3.668	3.017
ANN 1	0.526	3.344	2.706
ANN 2	0.497	3.614	2.982
Efraín Hernández	R²	RMSE	MAE
SVM – rbf 1	0.732	2.697	2.103
SVM – rbf 2	0.701	2.972	2.350
ANFIS 1	0.718	2.884	2.200
ANFIS 2	0.672	3.081	2.399
ANN 1	0.727	2.824	2.189
ANN 2	0.671	3.105	2.422
Mérida	R²	RMSE	MAE

SVM – rbf 1	0.704	2.535	1.933
SVM – rbf 2	0.653	2.914	2.232
ANFIS 1	0.689	2.561	1.935
ANFIS 2	0.617	2.849	2.150
ANN 1	0.697	2.526	1.952
ANN 2	0.625	2.818	2.113
Tantakín	R²	RMSE	MAE
SVM – rbf 1	0.700	2.482	1.918
SVM – rbf 2	0.667	2.556	2.010
ANFIS 1	0.660	2.788	2.121
ANFIS 2	0.616	2.976	2.259
ANN 1	0.661	2.795	2.116
ANN 2	0.626	2.921	2.222
Overall average	R²	RMSE	MAE
SVM – rbf 1	0.689	2.578	1.973
SVM – rbf 2	0.624	2.877	2.203
ANFIS 1	0.645	2.801	2.153
ANFIS 2	0.587	3.052	2.365
ANN 1	0.652	2.775	2.150
ANN 2	0.596	3.023	2.352

Figure 4.6. Regression analysis plot for the SVM – rbf 1 model between measured and estimated of daily global solar radiation during test phase.

5. CONCLUSIONS

This study assesses the potential of empirical equations and soft – computing techniques (ANN, ANFIS and SVM – rbf) for ET_0 and H prediction in Yucatán Peninsula Mexico. In the first part of the present thesis, the meteorological – based and soft – computing based models were evaluated for ET_0 prediction. In the second part, the meteorological – based, day of the year-based and soft – computing- based models were used for H modelling. The main conclusion for each part can be summed up as follows:

Conclusions for “Meteorological based and Soft –computing based models to estimate daily reference evapotranspiration”.

The performance of TET and PMT temperature based models to estimate daily ET_0 was assessed in four station in Yucatán Peninsula (Mexico).

1. The results showed that PMT estimates fit better to FAO56-PM than the calibrated HS estimates and six others calibrated TET models
2. The HS calibrated and PMT models showed the best performance with RMSE values of 0.74 and 0.70 mm day⁻¹, respectively
3. The statistical indicators for the PMT model indicate that the method can be employed for estimating daily ET_0 values with a high accuracy in regions where the required meteorological variables for PM calculations are missing. It is highly recommended to approximate the missing meteorological variables for use the PMT method for ET_0 estimation. In this study, these missing meteorological variables were approximate according the following criteria: a) use the common default value of 2 m s⁻¹ for wind speed; b) use the uncalibrated values for solar global radiation from the HS equation; and c) approximate T_{dew} as equal to T_{min} .

4. When only temperature data are available, the PMT approach is a viable option for agroclimatic and hydrologic applications in Yucatán Peninsula Mexico and in other areas worldwide.
5. The PMT model showed superiority over the HS and HS calibrated models in the Yucatán Peninsula. Before using the HS or PMT models, it is highly recommended to calibrate and evaluate the TET models. This study will make a considerable contribution to the evapotranspiration scientific studies and engineering applications in Mexico and in other areas worldwide.

The ability of three soft – computing techniques was assessed for ET_0 modelling.

6. As a result, the SVM – rbf soft computing approach for ET_0 modelling has demonstrated superiority to ANN and ANFIS techniques for all evaluated scenarios.
7. It also was shown that adding rainfall data as binary number into a soft – computing model slightly increased the model performance. However, if the relative humidity values are included into a model, estimates are improved significantly. This suggests that if relative humidity data are available, the SVM - rbf model using relative humidity, temperature and rainfall data should be used to obtain better results.

Conclusions for “*Meteorological – based, day of the year – based and Soft – computing- based models to estimate daily global solar radiation*”.

Twelve existing empirical models were calibrated and one new empirical model termed M5 was developed for H prediction.

8. When temperature, rainfall and relative humidity measurements are available, it is recommended to use Model M5 (proposed model) due to its simplicity, robustness and high performance.

9. If only temperature and rainfall data are available, the Liu and Scott model that includes current rainfall data is a better choice, and in cases when only temperature data are available, the Bristow and Campbell model provides good results.

Four DYB models from the literature and a newly developed model that estimate daily global solar radiation were evaluated.

10. The overall result of this investigation indicated that the new model, in which a sum of two Gaussian correlation formulas was used, performed best for all evaluated locations.
11. According to seasonal analysis, despite the existence of rain events or persistent cloud cover during the summer and autumn, daily solar radiation was estimated with acceptable accuracy with only day of the year as an input parameter.

The ability of three soft – computing techniques was assessed for H modelling.

12. It is concluded that, among the examined Soft – computing modelling techniques, the statistical analysis revealed that the SVM – *rbf* models for all evaluated scenarios shows superiority over ANN and ANFIS models.
13. The results also showed that the ANFIS and ANN techniques got a similar performance.
14. In addition, the soft – computing models using rainy – day information significantly increase the estimation than the ones based on temperatures and extraterrestrial radiation.

General conclusions

15. The choice of the best method to predict the global solar radiation or reference evapotranspiration does not only depend on its accuracy, but also on the availability of the input parameters and simplicity of implementation.
16. Soft Computing models can adequately estimate daily values of solar radiation and reference evapotranspiration, if only temperature data are available, the Bristow and Campbell and PMT methods are recommended to modelling solar radiation and reference evapotranspiration respectively. However, the use of Soft computing techniques improves the results and reduces errors, but nevertheless, they have the disadvantage of requiring statistical software and knowledge suitable for their use.

REFERENCES

- Al-Mostafa, Z. A., Maghrabi, A. H., & Al-Shehri, S. M. (2014). Sunshine-based global radiation models: A review and case study. *Energy Conversion and Management*, 84, 209–216. <https://doi.org/http://dx.doi.org/10.1016/j.enconman.2014.04.021>
- Al-Salaymeh, A. (2006). modeling of global daily solar radiation on horizontal surfaces for Amman city. *Emirates Journal for Engineering Research*, 11 (1)(2006) 49, 56.
- Allen, R. G. (1995). Evaluation of procedures for estimating mean monthly solar radiation from air temperature.
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *FAO, Rome*, 300, 6541. article.
- Allen, R., Pruitt, W., Wright, J., & Howell, T. (2006). A recommendation on standardized surface resistance for hourly calculation of reference ET o by the FAO56 Penman-Monteith method. *Agricultural Water*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S037837740500154X>
- Almorox, J., & Grieser, J. (2015). Calibration of the Hargreaves–Samani method for the calculation of reference evapotranspiration in different Köppen climate classes. *Hydrology Research*, nh2015091. article. <https://doi.org/10.2166/nh.2015.091>
- Almorox, J., Hontoria, C., & Benito, M. (2011). Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain). *Applied Energy*, 88(5), 1703–1709. <https://doi.org/http://dx.doi.org/10.1016/j.apenergy.2010.11.003>
- Almorox, J., Quej, V. H., & Martí, P. (2015). Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. *Journal of Hydrology*, 528, 514–522. <https://doi.org/10.1016/j.jhydrol.2015.06.057>
- Angstrom, A. (1924). Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and

- atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society*, 50(210), 121–126.
- Annandale, J., Jovanovic, N., Benade, N., & Allen, R. (2002). Software for missing data error analysis of Penman-Monteith reference evapotranspiration. *Irrigation Science*, 21(2), 57–67.
- Antonanzas-Torres, F., Sanz-Garcia, A., Martínez-de-Pisón, F. J., & Perpiñán-Lamigueiro, O. (2013). Evaluation and improvement of empirical models of global solar irradiation: Case study northern Spain. *Renewable Energy*, 60, 604–614.
- Antonanzas-Torres, F., Urraca, R., Antonanzas, J., Fernandez-Ceniceros, J., & Martinez-De-Pison, F. J. (2015). Generation of daily global solar irradiation with support vector machines for regression. *Energy Conversion and Management*, 96, 277–286.
<https://doi.org/10.1016/j.enconman.2015.02.086>
- Antonopoulos, V. Z., & Antonopoulos, A. V. (2017). Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Computers and Electronics in Agriculture*, 132, 86–96.
<https://doi.org/10.1016/j.compag.2016.11.011>
- Ata, R. (2015). Artificial neural networks applications in wind energy systems: a review. *Renewable and Sustainable Energy Reviews*, 49, 534–562.
<https://doi.org/10.1016/j.rser.2015.04.166>
- Baba, A. P.-A., Shiri, J., Kisi, O., Fard, A. F., Kim, S., & Amini, R. (2013). Estimating daily reference evapotranspiration using available and estimated climatic data by adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN). *Hydrology Research*, 44(1), 131–146.
- Badescu, V. (2002). A new kind of cloudy sky model to compute instantaneous values of diffuse and global solar irradiance. *Theoretical and Applied Climatology*, 72(1–2), 127–136.
- Badescu, V. (2014). *Modeling solar radiation at the earth's surface*. Springer.
- Bakirci, K. (2009). Models of solar radiation with hours of bright sunshine: A review. *Renewable and Sustainable Energy Reviews*, 13(9), 2580–2588.
<https://doi.org/http://dx.doi.org/10.1016/j.rser.2009.07.011>
- Benghanem, M., & Mellit, A. (2014). A simplified calibrated model for estimating

- daily global solar radiation in Madinah, Saudi Arabia. *Theoretical and Applied Climatology*, 115(1–2), 197–205.
- Besharat, F., Dehghan, A. A., & Faghih, A. R. (2013). Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, 21, 798–821.
- Bharat, V., Shelale, B., Khandelwal, K., & Navsare, S. (2016). A Review Paper on Data Mining Techniques. *International Journal of Engineering Science*, 6268.
- Blaney, H. F., & Criddle, W. D. (1950). Determining water needs from climatological data. *USDA Soil Conservation Service.SOS–TP, USA*, 8–9.
- Boukelia, T. E., Mecibah, M.-S., & Meriche, I. E. (2014). General models for estimation of the monthly mean daily diffuse solar radiation (Case study: Algeria). *Energy Conversion and Management*, 81, 211–219.
<https://doi.org/http://dx.doi.org/10.1016/j.enconman.2014.02.035>
- Bristow, K. L., & Campbell, G. S. (1984). On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology*, 31(2), 159–166.
[https://doi.org/http://dx.doi.org/10.1016/0168-1923\(84\)90017-0](https://doi.org/http://dx.doi.org/10.1016/0168-1923(84)90017-0)
- Brockamp, B., & Wenner, H. (1963). Verdunstungsmessungen auf den Steiner See bei Münster. *Dt Gewässerkundl Mitt*, 7, 149–154.
- Bulut, H., & Büyükalaca, O. (2007). Simple model for the generation of daily global solar-radiation data in Turkey. *Applied Energy*, 84(5), 477–491.
- Cai, J., Liu, Y., Xu, D., & Paredes, P. (2009). Simulation of the soil water balance of wheat using daily weather forecast messages to estimate the reference evapotranspiration. *Hydrology and Earth*. Retrieved from <http://www.repository.utl.pt/handle/10400.5/4862>
- Camargo, A. P., Marin, F. R., Sentelhas, P. C., & Picini, A. G. (1999). Adjust of the Thornthwaite's method to estimate the potential evapotranspiration for arid and superhumid climates, based on daily temperature amplitude. *Rev.Bras.Agronometeorol*, 7(2), 251–257. JOUR.
- Chang, C., & Lin, C. (2001). {LIBSVM}: a Library for Support Vector Machines (Version 2.3).
- Chen, J.-L., & Li, G.-S. (2014). Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables.

- Theoretical and Applied Climatology*, 115(3–4), 627–638.
- Chen, J., & Li, G. (2013). Estimation of monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China. *International Journal of Climatology*, 33(2), 487–498.
- Chen, R., Kang, E., Lu, S., Yang, J., Ji, X., Zhang, Z., & Zhang, J. (2006). New methods to estimate global radiation based on meteorological data in China. *Energy Conversion and Management*, 47(18–19), 2991–2998. <https://doi.org/http://dx.doi.org/10.1016/j.enconman.2006.03.025>
- Citakoglu, H., Cobaner, M., Haktanir, T., & Kisi, O. (2014). Estimation of Monthly Mean Reference Evapotranspiration in Turkey. *Water Resources Management*, 28(1), 99–113. <https://doi.org/10.1007/s11269-013-0474-1>
- Cobaner, M. (2011). Evapotranspiration estimation by two different neuro-fuzzy inference systems. *Journal of Hydrology*, 398(3–4), 292–302. <https://doi.org/10.1016/j.jhydrol.2010.12.030>
- Despotovic, M., Nedic, V., & Despotovic, D. (2015). Review and statistical analysis of different global solar radiation sunshine models. *And Sustainable Energy* Retrieved from <http://www.sciencedirect.com/science/article/pii/S1364032115008953>
- Donatelli, M., & Campbell, G. S. (1998). A simple model to estimate global solar radiation. In *Proceedings of the 5th European society of agronomy congress, Nitra, Slovak Republic* (Vol. 2, pp. 133–134).
- Doorenbos, J., & Pruitt, W. (n.d.). 0.(1977). *Guidelines for Predicting Crop Water Requirements*.
- El-Sebaili, A. A., Al-Hazmi, F. S., Al-Ghamdi, A. A., & Yaghmour, S. J. (2010). Global, direct and diffuse solar radiation on horizontal and tilted surfaces in Jeddah, Saudi Arabia. *Applied Energy*, 87(2), 568–576. <https://doi.org/http://dx.doi.org/10.1016/j.apenergy.2009.06.032>
- Esmaeelzadeh, S. R., Adib, A., & Alahdin, S. (2014). Long-term streamflow forecasts by Adaptive Neuro-Fuzzy Inference System using satellite images and K-fold cross-validation (Case study: Dez, Iran). *KSCE Journal of Civil Engineering*, 1–9.
- Falamarzi, Y., Palizdan, N., Huang, Y. F., & Lee, T. S. (2014). Estimating evapotranspiration from temperature and wind speed data using artificial and wavelet neural networks (WNNs). *Agricultural Water Management*, 140, 26–

36. <https://doi.org/10.1016/j.agwat.2014.03.014>
- Fritsch, F. N., & Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2), 238–246.
- Galindo, I., Castro, S., & Valdés, M. (2009). Satellite derived solar irradiance over Mexico. *Atmósfera*, 4(3).
- Gocić, M., Motamedi, S., Shamshirband, S., Petković, D., Ch, S., Hashim, R., & Arif, M. (2015). Soft computing approaches for forecasting reference evapotranspiration. *Computers and Electronics in Agriculture*, 113, 164–173. <https://doi.org/10.1016/j.compag.2015.02.010>
- Gocic, M., & Trajkovic, S. (2010). Software for estimating reference evapotranspiration using limited weather data. *Computers and Electronics in Agriculture*, 71(2), 158–162. article.
- Goodin, D. G., Hutchinson, J. M. S., Vanderlip, R. L., & Knapp, M. C. (1999). Estimating solar irradiance for crop modeling using daily air temperature data. *Agronomy Journal*, 91(5), 845–851.
- Gopalakrishnan, K., Khaitan, S. K., & Kalogirou, S. (2011). *Soft computing in green and renewable energy systems* (Vol. 269). Springer.
- Goyal, M. R., & Harmsen, E. W. (2013). *Evapotranspiration: Principles and Applications for Water Management*. CRC Press.
- Gu, L., Fuentes, J. D., Garstang, M., da Silva, J. T., Heitz, R., Sigler, J., & Shugart, H. H. (2001). Cloud modulation of surface solar irradiance at a pasture site in southern Brazil. *Agricultural and Forest Meteorology*, 106(2), 117–129.
- Hamon, W. R. (1963). Computation of direct runoff amounts from storm rainfall. *Int.Assoc.Sci.Hydrol.Publ*, 63, 52–62. JOUR.
- Hargreaves, G.H., Samani, Z. . (1982). Estimating potential evapotranspiration. *Journal of Irrigation and Drainage Engineering-ASCE*, 108(3), 225–230. article.
- Hargreaves, G. H., & Allen, R. G. (2003). History and evaluation of Hargreaves evapotranspiration equation. *Journal of Irrigation and Drainage Engineering-Asce*, 129(1), 53–63. article. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2003\)129:1\(53\)](https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53))
- Hargreaves, G. H., & Samani, Z. A. (1982). Estimating potential evapotranspiration. *Journal of the Irrigation and Drainage Division*, 108(3),

- 225–230. article.
- Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from ambient air temperature. *American Society of Agricultural Engineers (Microfiche collection)(USA)*. No. Fiche No. 85-2517. article.
- He, Z., Wen, X., Liu, H., & Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, 509, 379–386.
- Hunt, L. A., Kuchar, L., & Swanton, C. J. (1998). Estimation of solar radiation for use in crop modelling. *Agricultural and Forest Meteorology*, 91(3–4), 293–300. [https://doi.org/http://dx.doi.org/10.1016/S0168-1923\(98\)00055-0](https://doi.org/http://dx.doi.org/10.1016/S0168-1923(98)00055-0)
- Irmak, S., Irmak, A., Allen, R., & Jones, J. (2003). Solar and net radiation-based equations to estimate reference evapotranspiration in humid climates. *And Drainage Engineering*. Retrieved from [http://ascelibrary.org/doi/abs/10.1061/\(ASCE\)0733-9437\(2003\)129:5\(336](http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9437(2003)129:5(336)
- Iziomon, M. G., & Mayer, H. (2002). Assessment of some global solar radiation parameterizations. *Journal of Atmospheric and Solar-Terrestrial Physics*, 64(15), 1631–1643. [https://doi.org/http://dx.doi.org/10.1016/S1364-6826\(02\)00131-1](https://doi.org/http://dx.doi.org/10.1016/S1364-6826(02)00131-1)
- Jabloun, M. de, & Sahli, A. (2008). Evaluation of FAO-56 methodology for estimating reference evapotranspiration using limited climatic data: Application to Tunisia. *Agricultural Water Management*, 95(6), 707–715. article.
- Jang, J.-S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3), 665–685. <https://doi.org/10.1109/21.256541>
- Jang, J.-S. R., Sun, C.-T., & Mizutani, E. (1997). *Neuro-fuzzy and. Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall, New Jersey.
- Jensen, M. E., Burman, R. D., & Allen, R. G. (1990). Evapotranspiration and irrigation water requirements: a manual. *ASCE Manuals and Reports on Engineering Practice (USA).no.70*.
- Jensen, M. E., & Haise, H. R. (1963). Estimating evapotranspiration from solar radiation. *Proceedings of the American Society of Civil Engineers, Journal of*

- the Irrigation and Drainage Division*, 89, 15–41.
- Jiang, Y. (2009). Estimation of monthly mean daily diffuse radiation in China. *Applied Energy*, 86(9), 1458–1464.
- Jones, J. W., & Ritchie, J. T. (1990). Crop growth models. *IN: Management of Farm Irrigation Systems. American Society of Agricultural Engineers, St. Joseph, MI. 1990. p 63-89, 10 Fig, 2 Tab, 72 Ref.*
- Kahaner, D., Moler, C., & Nash, S. (1989). Numerical methods and software. *Englewood Cliffs: Prentice Hall*, 1989, 1.
- Kaplanis, S., & Kaplani, E. (2007). A model to predict expected mean and stochastic hourly global solar radiation I (h; nj) values. *Renewable Energy*, 32(8), 1414–1425.
- Khorasanizadeh, H., & Mohammadi, K. (2013). Prediction of daily global solar radiation by day of the year in four cities located in the sunny regions of Iran. *Energy Conversion and Management*, 76, 385–392. <https://doi.org/http://dx.doi.org/10.1016/j.enconman.2013.07.073>
- Khorasanizadeh, H., Mohammadi, K., & Jalilvand, M. (2014). A statistical comparative study to demonstrate the merit of day of the year-based models for estimation of horizontal global solar radiation. *Energy Conversion and Management*, 87(0), 37–47. <https://doi.org/http://dx.doi.org/10.1016/j.enconman.2014.06.086>
- Kirmani, S., Jamil, M., & Rizwan, M. (2015). Empirical correlation of estimating global solar radiation using meteorological parameters. *International Journal of Sustainable Energy*, 34(5), 327–339. <https://doi.org/10.1080/14786451.2013.826222>
- Kisi, O. (2013). Least squares support vector machine for modeling daily reference evapotranspiration. *Irrigation Science*, 31(4), 611–619.
- Köppen, W. (1936). Das geographische System der Klimate In: W. Köppen and G. Geiger, ed. *Handbuch Der Klimatologie (Handbuch Der Klimatologie, vol. 1: C. Gebr, Borntraeger)*.
- Kra, E. (2010). An empirical simplification of the temperature Penman–Monteith model for the tropics. *Journal of Agricultural Science*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.665.6511&rep=rep1&type=pdf>
- Kumar, R., Aggarwal, R. K., & Sharma, J. D. (2015). Comparison of regression

- and artificial neural network models for estimation of global solar radiations. *Renewable and Sustainable Energy Reviews*, 52, 1294–1299. <https://doi.org/http://dx.doi.org/10.1016/j.rser.2015.08.021>
- Li, H., Cao, F., Bu, X., & Zhao, L. (2015). Models for calculating daily global solar radiation from air temperature in humid regions—A case study. *Environmental Progress & Sustainable Energy*, 34(2), 595–599.
- Li, H., Ma, W., Lian, Y., & Wang, X. (2010). Estimating daily global solar radiation by day of year in China. *Applied Energy*, 87(10), 3011–3017.
- Li, H. S., Ma, W. B., Bu, X. B., Lian, Y. W., & Wang, X. L. (2013). A Multiple Linear Regression Model for Estimating Global Solar Radiation in Guangzhou, China. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 35(4), 321–327. <https://doi.org/10.1080/15567036.2010.499422>
- Liu, D. L., & Scott, B. J. (2001). Estimation of solar radiation in Australia from rainfall and temperature observations. *Agricultural and Forest Meteorology*, 106(1), 41–59. [https://doi.org/http://dx.doi.org/10.1016/S0168-1923\(00\)00173-8](https://doi.org/http://dx.doi.org/10.1016/S0168-1923(00)00173-8)
- Lockart, N., Kavetski, D., & Franks, S. W. (2015). A new stochastic model for simulating daily solar radiation from sunshine hours. *International Journal of Climatology*, 35(6), 1090–1106.
- López-Moreno, J. I., Hess, T. M., & White, A. S. M. (2009). Estimation of Reference Evapotranspiration in a Mountainous Mediterranean Site Using the Penman-Monteith Equation With Limited Meteorological Data. *Pirineos JACA*, 164, 7–31. article.
- Mahmood, R., & Hubbard, K. G. (2002). Effect of time of temperature observation and estimation of daily solar radiation for the Northern Great Plains, USA. *Agronomy Journal*, 94(4), 723–733.
- Mahringer, W. (1970). Verdunstungsstudien am neusiedler See. *Archiv Für Meteorologie, Geophysik Und Bioklimatologie, Serie B*, 18(1), 1–20.
- Malmström, V. H. (1969). A New Approach To The Classification Of Climate. *Journal of Geography*, 68(June 2015), 351–357. JOUR. <https://doi.org/10.1080/00221346908981131>
- Martinez, C., & Thepadia, M. (2009). Estimating reference evapotranspiration with minimum data in Florida. *Journal of Irrigation and Drainage*. Retrieved from [http://ascelibrary.org/doi/abs/10.1061/\(ASCE\)IR.1943-4774.0000214](http://ascelibrary.org/doi/abs/10.1061/(ASCE)IR.1943-4774.0000214)

- McCaskill, M. R. (1990). An efficient method for generation of full climatological records from daily rainfall. *Crop and Pasture Science*, 41(3), 595–602.
- McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*. Retrieved from <http://link.springer.com/article/10.1007/bf02478259>
- McGuinness, J. L., & Bordne, E. F. (1972). *A comparison of lysimeter-derived potential evapotranspiration with computed values*. US Dept. of Agriculture.
- Mendicino, G., & Senatore, A. (2013). Regionalization of the Hargreaves Coefficient for the Assessment of Distributed Reference Evapotranspiration in Southern Italy. *Journal of Irrigation and Drainage Engineering*, 139(5), 349–362. article. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000547](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000547)
- Meyer, D., & Wien, F. H. T. (2014). Support vector machines. *The Interface to Libsvm in Package e1071*.
- Mohammadi, K., Shamshirband, S., Anisi, M. H., Alam, K. A., & Petković, D. (2015). Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Conversion and Management*, 91, 433–441. <https://doi.org/http://dx.doi.org/10.1016/j.enconman.2014.12.015>
- Norusis, M. J. (2008). SPSS 16.0 Guide to Data Analysis Value Package (includes SPSS 16.0 Student Version for Windows).
- Olatomiwa, L., Mekhilef, S., Shamshirband, S., Mohammadi, K., Petković, D., & Sudheer, C. (2015a). A support vector machine–firefly algorithm-based model for global solar radiation prediction. *Solar Energy*, 115, 632–644. <https://doi.org/http://dx.doi.org/10.1016/j.solener.2015.03.015>
- Olatomiwa, L., Mekhilef, S., Shamshirband, S., & Petkovic, D. (2015b). Potential of support vector regression for solar radiation prediction in Nigeria. *Natural Hazards*, 77(2), 1055–1068.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, 303(1), 290–306. JOUR.
- Pandey, P., & Pandey, V. (2016). Evaluation of temperature-based Penman–Monteith (TPM) model under the humid environment. *Modeling Earth Systems and Environment*. Retrieved from

- <http://link.springer.com/article/10.1007/s40808-016-0204-9>
- Papadakis, J. (1966). Climates of the world and their agricultural potentialities. *Climates of the World and Their Agricultural Potentialities*. JOUR.
- Patro, S., & Sahu, K. K. (2015). Normalization: A Preprocessing Stage. *arXiv Preprint arXiv:1503.06462*.
- Pereira, L. S., Allen, R. G., Smith, M., & Raes, D. (2015). Crop evapotranspiration estimation with FAO56: Past and future. *Agricultural Water Management*. article. <https://doi.org/10.1016/j.agwat.2014.07.031>
- Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., & Osuch, M. (2015). Comparing various artificial neural network types for water temperature prediction in rivers. *Journal of Hydrology*, 529, 302–315. <https://doi.org/10.1016/j.jhydrol.2015.07.044>
- Piri, J., Shamshirband, S., Petković, D., Tong, C. W., & ur Rehman, M. H. (2015). Prediction of the solar radiation on the Earth using support vector regression technique. *Infrared Physics & Technology*, 68, 179–185. <https://doi.org/http://dx.doi.org/10.1016/j.infrared.2014.12.006>
- Popova, Z., Kercheva, M., & Pereira, L. S. (2006). Validation of the FAO methodology for computing ETo with limited data. Application to south Bulgaria. *Irrigation and Drainage*, 55(2), 201–215. article. <https://doi.org/10.1002/ird.228>
- Prescott, J. A. (1940). Evaporation from a water surface in relation to solar radiation. *Transactions of the Royal Society of South Australia*, 64(1940), 114–118.
- Qazi, A., Fayaz, H., Wadi, A., Raj, R. G., Rahim, N. A., & Khan, W. A. (2015). The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. *Journal of Cleaner Production*, 104, 1–12. <https://doi.org/http://dx.doi.org.access.biblio.colpos.mx/10.1016/j.jclepro.2015.04.041>
- Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S., & Khoshnevisan, B. (2014). Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable and Sustainable Energy Reviews*, 39, 1005–1011. <https://doi.org/http://dx.doi.org/10.1016/j.rser.2014.07.108>

- Raziei, T., & Pereira, L. S. (2013). Estimation of ETo with Hargreaves-Samani and FAO-PM temperature methods for a wide range of climates in Iran. *Agricultural Water Management*, 121, 1–18. article. <https://doi.org/10.1016/j.agwat.2012.12.019>
- RDevelopment, C. (2012). TEAM 2009: R: A Language and Environment for Statistical Computing. Vienna, Austria. Internet: <http://www.R-Project.org>.
- Ren, X., Qu, Z., Martins, D. S., Paredes, P., & Pereira, L. S. (2016). Daily Reference Evapotranspiration for Hyper-Arid to Moist Sub-Humid Climates in Inner Mongolia, China: I. Assessing Temperature Methods and Spatial Variability. *Water Resources Management*, 30(11), 3769–3791. JOUR. <https://doi.org/10.1007/s11269-016-1384-9>
- Rezrazi, A., Hanini, S., & Laidi, M. (2015). An optimisation methodology of artificial neural network models for predicting solar radiation: a case study. *Theoretical and Applied Climatology*, 123(3–4), 1–15. <https://doi.org/10.1007/s00704-015-1398-x>
- Romanenko, V. A. (1961). Computation of the autumn soil moisture using a universal relationship for a large area. *Proceedings Ukrainian Hydrometeorological Research Institute (Kiev)*, 3.
- Rosenblatt, F. (1962). Principles of neurodynamics. Retrieved from <http://www.citeulike.org/group/780/article/507136>
- Sen, Z. (2008). *Solar energy fundamentals and modeling techniques: atmosphere, environment, climate change and renewable energy*. Springer Science & Business Media.
- Şenkal, O. (2015). Solar radiation and precipitable water modeling for Turkey using artificial neural networks. *Meteorology and Atmospheric Physics*, 1–8.
- Shiri, J., Kisi, O., Landaras, G., Javier Lopez, J., Nazemi, A. H., & Stuyt, L. C. P. M. (2012). Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). *Journal of Hydrology*, 414, 302–316. <https://doi.org/10.1016/j.jhydrol.2011.11.004>
- Shiri, J., Nazemi, A. H., Sadraddini, A. A., Landaras, G., Kisi, O., Fakheri Fard, A., & Marti, P. (2014). Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Computers and Electronics in Agriculture*, 108, 230–241. <https://doi.org/10.1016/j.compag.2014.08.007>

- Skeiker, K. (2006). Correlation of global solar radiation with common geographical and meteorological parameters for Damascus province, Syria. *Energy Conversion and Management*, 47(4), 331–345. <https://doi.org/http://dx.doi.org/10.1016/j.enconman.2005.04.012>
- Spencer, J. W. (1971). Fourier series representation of the position of the sun. *Search*, 2(5), 172. JOUR. Retrieved from <http://www.mail-archive.com/sundial@uni-koeln.de/msg01050.html>
- Stephens, G. L., Wild, M., Jr, P. W. S., L'Ecuyer, T., Kato, S., & Henderson, D. S. (2012). The global character of the flux of downward longwave radiation. *Journal of Climate*, 25(7), 2329–2340.
- Sumithira, T. R., Kumar, A. N., & Rameshkumar, R. (2012). An adaptive neuro-fuzzy inference system (ANFIS) based Prediction of Solar Radiation. *J.of App.Sci.Res*, 8(1), 346–351.
- Sun, Y., & Genton, M. G. (2012). Functional boxplots. *Journal of Computational and Graphical Statistics*.
- Supit, I., & van Kappel, R. R. (1998). A simple method to estimate global radiation. *Solar Energy*, 63(3), 147–160. [https://doi.org/http://dx.doi.org/10.1016/S0038-092X\(98\)00068-1](https://doi.org/http://dx.doi.org/10.1016/S0038-092X(98)00068-1)
- Tabari, H., Grismer, M. E., & Trajkovic, S. (2013). Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrigation Science*, 31(2), 107–117.
- Tabari, H., Kisi, O., Ezani, A., & Hosseinzadeh Talaei, P. (2012). SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology*, 444–445, 78–89. <https://doi.org/10.1016/j.jhydrol.2012.04.007>
- Teke, A., Yıldırım, H. B., & Çelik, Ö. (2015). Evaluation and performance comparison of different models for the estimation of solar radiation. *Renewable and Sustainable Energy Reviews*, 50, 1097–1107.
- Thornthwaite, C. W. (1948). An Approach toward a Rational Classification of Climate. *Geographical Review*, 38(1), 55. JOUR. <https://doi.org/10.2307/210739>
- Todorovic, M., Karic, B., & Pereira, L. S. (2013). Reference evapotranspiration estimate with limited weather data across a range of Mediterranean climates.

- Journal of Hydrology*, 481, 166–176. article.
<https://doi.org/10.1016/j.jhydrol.2012.12.034>
- Toğrul, I. T., & Onat, E. (1999). A study for estimating solar radiation in Elaziğ using geographical and meteorological data. *Energy Conversion and Management*, 40(14), 1577–1584.
[https://doi.org/http://dx.doi.org/10.1016/S0196-8904\(99\)00035-7](https://doi.org/http://dx.doi.org/10.1016/S0196-8904(99)00035-7)
- Trajkovic, S. (2005). Temperature-based approaches for estimating reference evapotranspiration. *Journal of Irrigation and Drainage Engineering*. Retrieved from [http://ascelibrary.org/doi/abs/10.1061/\(ASCE\)0733-9437\(2005\)131:4\(316\)](http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9437(2005)131:4(316))
- Trenberth, K. E. (1983). What are the seasons? *Bulletin of the American Meteorological Society*, 64(11), 1276–1282.
- Vangelis, H., Tigkas, D., & Tsakiris, G. (2013). The effect of PET method on Reconnaissance Drought Index (RDI) calculation. *Journal of Arid Environments*, 88, 130–140. article.
<https://doi.org/10.1016/j.jaridenv.2012.07.020>
- Wang, Z., Wu, P., Zhao, X., Cao, X., & Gao, Y. (2014). GANN models for reference evapotranspiration estimation developed with weather data from different climatic regions. *Theoretical and Applied Climatology*, 116(3–4), 481–489. <https://doi.org/10.1007/s00704-013-0967-0>
- Wen, X., Si, J., He, Z., Wu, J., Shao, H., & Yu, H. (2015). Support-Vector-Machine-Based Models for Modeling Daily Reference Evapotranspiration With Limited Climatic Data in Extreme Arid Regions. *Water Resources Management*, 29(9), 3195–3209. <https://doi.org/10.1007/s11269-015-0990-2>
- Wilamowski, B. M., Chen, Y., & Malinowski, A. (1999). Efficient algorithm for training neural networks with one hidden layer. In *IJCNN. Proc. 1999 International Joint Conference* (Vol. 3, pp. 1725–1728). Citeseer.
- Winslow, J. C., Hunt, E. R., & Piper, S. C. (2001). A globally applicable model of daily solar irradiance estimated from air temperature and precipitation data. *Ecological Modelling*, 143(3), 227–243.
- Woli, P., & Paz, J. O. (2012). Evaluation of various methods for estimating global solar radiation in the southeastern United States. *Journal of Applied Meteorology and Climatology*, 51(5), 972–985.

- Wu, G., Liu, Y., & Wang, T. (2007). Methods and strategy for modeling daily global solar radiation with measured meteorological data – A case study in Nanchang station, China. *Energy Conversion and Management*, 48(9), 2447–2452.
<https://doi.org/http://dx.doi.org/10.1016/j.enconman.2007.04.011>
- Yadav, A. K., & Chandel, S. S. (2014). Solar radiation prediction using Artificial Neural Network techniques: A review. *Renewable and Sustainable Energy Reviews*, 33, 772–781.
<https://doi.org/http://dx.doi.org.access.biblio.colpos.mx/10.1016/j.rser.2013.08.055>
- Zang, H., Xu, Q., & Bian, H. (2012). Generation of typical solar radiation data for different climates of China. *Energy*, 38(1), 236–248.
<https://doi.org/http://dx.doi.org/10.1016/j.energy.2011.12.008>
- Zhang, L., Juyang, L. E. I., Qilin, Z., & Yudong, W. (2015). Using Genetic Algorithm to Optimize Parameters of Support Vector Machine and Its Application in Material Fatigue Life Prediction. *Advances in Natural Science*, 8(1).

APPENDIX A.

LIST OF PUBLICATIONS

Quej, V.H., Almorox, J., Arnaldo, J.A., Saito, L., (2017). ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *J. Atmos. Solar-Terrestrial Phys.* 155, 62–70. doi:10.1016/j.jastp.2017.02.002

Quej, V. H., Almorox, J., Ibrakhimov, M., & Saito, L. (2017). Estimating daily global solar radiation by day of the year in six cities located in the Yucatán Peninsula, Mexico. *Journal of Cleaner Production*, 141, 75-82.

Quej, V. H., Almorox, J., Ibrakhimov, M., & Saito, L. (2016). Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico. *Energy Conversion and Management*, 110, 448-456.

Almorox, J., Quej, V. H., & Martí, P. (2015). Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. *Journal of Hydrology*, 528, 514-522.

Almorox, J., Senatore, A., Quej, V. H., & Mendicino, G. (2016). Worldwide assessment of the Penman–Monteith temperature approach for the estimation of monthly reference evapotranspiration. *Theoretical and Applied Climatology*, 1-11.

APPENDIX B**CODES TO SIMULATE THE SOFT – COMPUTING MODELS.****MATLAB Code of Artificial Neural Networks Estimation.**

%For random weight initialization problem, the next line sets the same seed every time, so the rand() sequence is always identical.

```
RandStream.setGlobalStream (RandStream ('mrg32k3a','Seed', 1234));
```

```
% Set size of the hidden layers
```

```
n = 00;
```

```
% Create Feedforward neural network
```

```
net = feedforwardnet(n);
```

```
%Set training algorithm
```

```
net.trainFcn = 'trainlm';
```

```
% Set performance function for feedforward networks.
```

```
net.performFcn = 'mse';
```

```
% Divide targets into three sets using blocks of indices.
```

```
%[trainInd,valInd,testInd] = divideblock(00,0.60, 0.20, 0.20);
```

```
net.divideFcn = 'divideblock';
```

```
% Set training parameters
```

```
net.trainParam.max_fail = 200;
```

```
net.performParam.regularization = 0.01;
```

```
net.performParam.normalization = 'none';
```

```
net.trainParam.mu_max=1e200;
```

```
net.trainParam.goal = 0.01;
```

```
net.trainParam.epochs = 250;
```

```

%net.trainParam.mc = 0.5;
%net.trainParam.lr = 0.01;

% Set the transfer function in the hidden layer and the output layer.
net.layers{1}.transferFcn = 'logsig';
net.layers{2}.transferFcn = 'purelin';

% Trains the network net according to net.trainFcn and net.trainParam, that
returns newly trained network (net) and Training record (tr).

[net,tr] = train(net,input,target);
plotperform(tr);

% Simulate the network
training = net(input);
view(net);

% Verify the network's performance
MSE_Train = perform(net,training,target);
RMSE_Train = sqrt(MSE_Train);

% Using the trained network with other values.
predicted=sim(net,test);

% Verify the accuracy of the estimated data.
RMSE_Val=norm(measured-predicted)/sqrt(length(measured));
MAE_Val=sum(abs(measured-predicted) *1/(length(measured)));

```

MATLAB code for numerically simulating a Sugeno-type Fuzzy Inference System structure from data using grid partition.

```

%Input: trnData (Train), chkData(test).
%Set the membership function parameters.
%Set the numbers of membership function.

```

```

numMFs = [2 2 2];
%Set the type of membership function.
mfType = char('gbellmf','gbellmf','gbellmf');
initFIS=genfis1(trnData,numMFs,mfType);
%Opens the rule viewer for the fuzzy inference system.
ruleview(initFIS)
%Plot the membership functions.
[x,mf]=plotmf(initFIS,'input',1)
plot(x,mf)
%Set training parameters.
numEpochs=200;
%Set training method (1- hibrid or 0-back propagation)
optMethod = 1;
%Training routine for Sugeno-type fuzzy inference system.
[fis,trnError,stepsize,chkfis,chkErr] = anfis(trnData,initFIS,numEpochs,
NaN,chkData,optMethod);

%Perform fuzzy inference calculations.
trnOut=evalfis(trnData,fis);
trnRMSE=norm(trnOut-trnData)/sqrt(length(trnOut));

%Evaluate fuzzy inference system using test data.
chkOut=evalfis(chkData,chkfis);
chkRMSE=norm(chkOut-chkData)/sqrt(length(chkOut));

```

Code for simulating Support Vector Machines (SVM) by using R interface to libsvm in package e1071.

```

#Reading data from file.
data<-read.table("file.txt")
#Split data in two subsets for training and test.
index <- 1:nrow(data)
testindex <- sample(index, trunc(length(index)/3))
testset <- (data[testindex,1:5])

```

```

trainset <- (data[-testindex,1:5])
#Training the support vector regression with radial basis kernel, the parameters
cost, gamma and epsilon are computed using the genetic algorithm.
svm.model <- svm(V1 ~ ., data = trainset, type= "eps-regression", kernel=
"radial", cost = 0.000, gamma = 0.000, epsilon = 0.000, scale=TRUE)
#using the trained data to make predictions.
svm.pred <- predict(svm.model, testset[,-1])
#Test with train data.
svm.pred2 <- predict(svm.model, trainset[,-1])
# Validation accuracy.
#RMSE.
rmse <- function(error)
{
  sqrt(mean(error^2))
}
error <- (svm.pred - testset[,1])
error2<-(svm.pred2 - trainset[,1])
RMSEmodel <- rmse(error)
RMSEtrain <-rmse(error2)
#MAE.
mae <- function(error2)
{
  mean(abs(error2))
}
error3 <- svm.pred - testset[,1]
MAEmodel <- mae(error3)
error4 <- svm.pred2 - trainset[,1]
MAEtrain <- mae(error4)

```