

Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations

Graduate Studies

12-2013

Modeling and forecasting evapotranspiration for better management of irrigation command areas

Roula Bachour
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>

 Part of the [Civil and Environmental Engineering Commons](#)

Recommended Citation

Bachour, Roula, "Modeling and forecasting evapotranspiration for better management of irrigation command areas" (2013). *All Graduate Theses and Dissertations*. 2077.
<https://digitalcommons.usu.edu/etd/2077>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



EVAPOTRANSPIRATION MODELING AND FORECASTING FOR EFFICIENT
MANAGEMENT OF IRRIGATION COMMAND AREAS

By

Roula Bachour

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Irrigation Engineering

Approved:

Dr. Wynn R. Walker
Major Professor

Dr. L. Niel Allen
Committee Member

Dr. Adele Cutler
Committee Member

Dr. Mac McKee
Committee Member

Dr. Gilberto Urroz
Committee Member

Dr. Mark McLellan
Vice President Research and
Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2013

Copyright © Roula Bachour 2013

All Rights Reserved

ABSTRACT

Evapotranspiration Modeling and Forecasting for Efficient
Management of Irrigation Command Areas

by

Roula Bachour, Doctor of Philosophy
Utah State University, 2013

Major Professor: Dr. Wynn R. Walker
Department: Civil and Environmental Engineering

It has become very crucial to manage water resources to meet the needs of the growing population. In irrigation command areas, and in order to build a better plan to manage service delivery from canals and reservoirs, it is important to build appropriate knowledge of water needs on a field basis. There is often a lag between the order and delivery of water to the field. Knowledge of the crop water requirement at the field level helps the decision maker to make the right choices leading to more efficient handling of the available water. The purpose of this study was to develop methodologies and tools that allow better management of irrigation water and water delivery systems, such as machine learning models that can be used as tools for decision support systems of water management. To achieve better modeling and prediction, wavelet decompositions were explored for their ability to give information about time and frequency changes in the data. Remote sensing approaches were also used for their ability to quantify water requirements at the spatial level. Therefore, this dissertation explored the use of the

above-mentioned data tools and techniques to address water management problems. The framework of this dissertation consisted of three components that provide tools to support irrigation system operational decisions. In general, the results for each of the methods developed were satisfactory, relevant, and encouraging. They provided significant potential for improving decision making for real-time applications in irrigation command areas and better management of the water resources.

(114 pages)

PUBLIC ABSTRACT

Evapotranspiration Modeling and Forecasting for Efficient

Management of Irrigation Command Areas

by

Roula Bachour, Doctor of Philosophy
Utah State University, 2013

Major Professor: Dr. Wynn R. Walker
Department: Civil and Environmental Engineering

It has become very crucial to manage water resources to meet the needs of the growing population. In irrigation command areas, and in order to build a better plan to manage service delivery from canals and reservoirs, it is important to build appropriate knowledge of water needs on a field basis. There is often a lag between the order and delivery of water to the field. Knowledge of the crop water requirement at the field level helps the decision maker to make the right choices leading to more efficient handling of the available water. The purpose of this study was to develop methodologies and tools that allow better management of irrigation water and water delivery systems, such as machine learning models that can be used as tools for decision support systems of water management. To achieve better modeling and prediction, wavelet decompositions were explored for their ability to give information about time and frequency changes in the data. Remote sensing approaches were also used for their ability to quantify water requirements at the spatial level. Therefore, this dissertation explored the use of the

above-mentioned data tools and techniques to address water management problems. The framework of this dissertation consisted of three components that provide tools to support irrigation system operational decisions. In general, the results for each of the methods developed were satisfactory, relevant, and encouraging. They provided significant potential for improving decision making for real-time applications in irrigation command areas and better management of the water resources.

Roula Bachour

I dedicate my dissertation to my sweet mother Mariam and my supporting father Mikhael

ACKNOWLEDGMENTS

I never imagined a day would come when I would meet the author of those irrigation books I studied with, and here I am, feeling a great honor to be one of your students, Dr. Wynn Walker. I would like to express my deep gratitude and thanks to you for being my advisor. I feel very fortunate to have met you and worked with you. I thank you for your guidance and support, and I appreciate your remarks about the practical utility of my research.

I am also grateful for Dr. Mac McKee who provided me this wonderful opportunity to work with him at the Utah Water Research Laboratory. Dr. Mac, you gave me the best supervision, encouragement, help, and moral support, which made this research possible. I appreciate and acknowledge the opportunity you gave me to work in a professional and very friendly environment.

I am thankful to my committee members, Dr. L. Niel Allen, Dr. Adele Cutler, and Dr. Gilberto Urroz, for their time, assistance, and valuable comments to improve the quality of my research. I sincerely appreciate Dr. David Stevens for his intelligent remarks, and his help answering my statistics questions.

I extend my thanks to Dr. Inga Maslova and Dr. Andres Ticlavilca for their time, supervised teaching, professionalism, and valuable inputs that improved this work a lot; I really enjoyed working with both of you. And to my friend, colleague, and fieldwork partner Dr. Alfonso Torres, I was very happy to work with you during my doctoral studies; many thanks for your help, for your smart valuable comments, and for your availability to answer my questions, especially in MATLAB. I also would like to express

gratitude to Dr. Bethany Neilson mainly for moral support; thank you for all your advice and your ability to listen to my complaints.

Many thanks go to my friends in Lebanon who never gave up on me, kept in touch, gave me their support, and always made sure to draw a smile on my face when I needed it. And to my friends in Logan I say, you guys were my family during the last few years, thank you for keeping me sane, thank you for caring and sharing the laughs and the tears. Living in Logan would definitely have been impossible without you, "my Tribe."

There are no words that could describe my thanks and gratitude for my mentor, advisor, or as I always like to call him my second father, Dr. Musa Nimah. Without you, I would not be the person I am now. I am thankful for all your teachings, your continuous scientific and moral support, and I am mainly thankful for the opportunity you opened for me to be at USU.

I am indebted for my parents and family's unconditional support and encouragement. And to my fiancé, Carlos, I would like to say thank you for holding my hand throughout those years, thank you for being there to listen to me, especially when whatever I said didn't make any sense, and thank you for your patience, love, and support.

Roula Bachour

CONTENTS

	Page
ABSTRACT.....	iii
PUBLIC ABSTRACT.....	v
ACKNOWLEDGMENTS.....	viii
LIST OF TABLES.....	xiii
LIST OF FIGURES.....	xiv
CHAPTER	
1. INTRODUCTION.....	1
1.1 General Introduction.....	1
1.2 Purpose and Objectives.....	3
1.3 Dissertation Organization.....	5
References.....	6
2. ASSESSMENT OF REFERENCE EVAPOTRANSPIRATION BY THE HARGREAVES METHOD IN THE BEKAA VALLEY, LEBANON	7
Abstract.....	7
2.1 Introduction.....	8
2.2 Methodology.....	11
2.2.1 Site Description.....	11
2.2.2 Estimating Reference Evapotranspiration.....	13
2.2.2.1 FAO-56 Penman-Monteith.....	13
2.2.2.2 Hargreaves Method.....	13
2.2.2.3 Calibrated Hargreaves Equation.....	14
2.2.3 Evaluation Procedure.....	15
2.3 Results and Discussion.....	15
2.3.1 Evaluation of the Hargreaves Method.....	15
2.3.2 Local Calibration.....	16
2.4 Conclusions.....	21

References.....	21
3. WAVELET-MULTIVARIATE RELEVANCE VECTOR MACHINE HYBRID MODEL FOR FORECASTING DAILY EVAPOTRANSPIRATION	25
Abstract.....	25
3.1 Introduction.....	26
3.2 Methodology.....	30
3.2.1 Data Collection and Description.....	30
3.2.2 Wavelet Multiresolution Analysis.....	31
3.2.3 Multivariate Relevance Vector Machine.....	33
3.2.4 Performance Estimation Criteria.....	37
3.3 Results and Discussion.....	39
3.3.1 Wavelet Decomposition Selection.....	39
3.3.2 Forecasting Model Selection.....	42
3.3.4 Model Robustness.....	48
3.3.5 Forecasting ET Using 10-days of Forecasted Temperature.....	49
3.4 Conclusions and Future Work.....	52
References.....	53
4. SPATIAL DISTRIBUTION OF EVAPOTRANSPIRATION USING REMOTE SENSING AND RELEVANCE VECTOR MACHINE.....	57
Abstract.....	57
4.1 Introduction.....	58
4.2 Models Background.....	61
4.2.1 METRIC Algorithm.....	61
4.2.2 Relevance Vector Machine.....	63
4.3 Methodology.....	65
4.3.1 Study Area.....	65
4.3.2 Satellite Data.....	66
4.3.3 Image Processing.....	68
4.3.4 Model Selection.....	69
4.3.5 Model Performance.....	70
4.4 Results and Discussion.....	72

4.4.1 Spatial Distribution of ET by METRIC.....	72
4.4.2 System Efficiency.....	74
4.4.3 RVM Model Results.....	75
4.5 Conclusions.....	79
References.....	80
5. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS.....	84
5.1 Summary and Conclusions.....	84
5.2 Recommendations and Future Work.....	86
APPENDIX.....	89
CURRICULUM VITAE.....	96

LIST OF TABLES

Table	Page
2.1 Monthly Averages of Weather Variables at Terbol Station During 1993-2010.....	12
2.2 Comparison of ET Estimated by FAO-PM Equation and the Different Forms of HG Equations for Three Timescales.....	18
3.1 Wavelet MODWT-Based Power Distribution by Level.....	40
3.2 Models Inputs and Average Statistics of the Testing Data Set for the 16 Days of Forecasted ET.....	44
3.3 Statistics of the Selected Models (M1, M2, and M5) for All the Forecasted Days.....	45
3.4 Statistical Results of M2, M5, and 9-Years Average as Compared to the Observed Data for the Two Years of Test Dataset.....	47
3.5 Statistics of the Models for the 16 Days with the Additional 10-Days Forecasted T_{\min} and T_{\max} as Input.....	51
4.1 Landsat Satellite Images Used for ET Estimation	67
4.2 Canal B Diversions and System Efficiency.....	74
4.3 Performance Criteria Results for RVM Model in the Training and Testing Phase	76
4.4 Comparison Between ET from METRIC and ET from RVM Model for Selected Days	77

LIST OF FIGURES

Figure		Page
2.1	Location of Terbol weather station in the Bekaa Valley, Lebanon.....	12
2.2	Comparison of ET _{PM} with original HG for (a) daily, (b) weekly, and (c) monthly time steps for 16 years	16
2.3	Comparison of ET _{PM} with modified HG (HG _{mod1})for (a) daily, (b) weekly, and (c) monthly time steps for 16 years	18
2.4	Comparison of ET _{PM} with wind-modified HG (HG _{mod2}) for (a) daily, (b) weekly, and (c) monthly time steps for 16 years	20
3.1	Design 1 of Wavelet-MRA decomposition	41
3.2	Design 2 of Wavelet MRA decomposition.....	43
3.3	Left: Predicted ET vs, observed ET times series for two-years of unseen testing data for Model 2 for selected days. Right: Plots of predicted ET vs. observed ET for the same time period.....	47
3.4	Left: Predicted ET vs, observed ET times series for two-years of unseen testing data for Model 5 for selected days. Right: Plots of predicted ET vs. observed ET for the same time period.....	48
3.5	Boxplots for the results of bootstrapping analysis (1000 times) for models (M1, M2 and M5) for selected days.....	50
3.6	Scatter plots of forecasted ET using modified models: M1* (a), M2* (b) and M5* (c) compared to measured ET.....	52
4.1	Location of Canal B irrigation area in the Lower Sevier River Basin....	67
4.2	Flow diagram of the RVM model inputs and outputs used in this study.....	70
4.3	Spatial distribution of daily ET from METRIC for selected days over the Canal B area.....	73
4.4	Modeled vs. calculated spatial ET for the study area for the unseen data.....	76

Figure		Page
4.5	Residual error maps for the study area for selected days	77
4.6	Histograms of residuals errors for the selected days.....	78
4.7	ET_METRIC vs. ET_RVM for a test image.....	79

CHAPTER 1

INTRODUCTION

1.1 General Introduction

The steadily increasing world population will place a growing burden on global food security, especially in regions where water supplies may not be sufficient for both food production and potable water. Even more striking is the growing dependence on irrigated production to provide sufficient food supply. Competition for water, high pumping costs, complexities of water storage and delivery, and concerns for the environment are among the factors that drive an interest in improving the water use efficiency and the operation of large irrigation systems. For purposes of timely and efficient water application, agricultural managers have long relied on evapotranspiration (ET) measurements or estimations. Therefore, an accurate assessment of ET is prerequisite to improving water management practices.

ET is one of the main components of the hydrological cycle. It is a complex process driven mainly by weather parameters. The FAO Penman-Monteith (PM) model has become the generally accepted standard for calculating ET (Allen et al. 2006). Inputs required for the PM computations include several climatic variables such as air temperature, relative humidity, wind speed, and solar radiation that are not always available or reliable. In such cases, the equation developed by Hargreaves and Samani (1985) can be used with reasonable results, and it only requires measured daily air temperature data and computed extraterrestrial radiation. This equation, when calibrated for a specific area, is a useful tool for estimating ET with limited weather data.

Accurate estimation of ET is very important, but the crucial element for management of irrigation systems remains in providing a reliable short-term forecast of ET. Data-driven models based on the machine learning approach have been widely used in hydrological modeling for forecasting purposes. Data-driven models have the capability to learn from previous experiences, and thus offer a potentially powerful tool in capturing the behavior of real systems by relating inputs and outputs. They are robust and are capable of making reasonable prediction using historical data (Gill et al. 2006; Kaheil et al. 2008).

ET is characterized by high non-linearity and non-stationarity (Hernandez et al. 2011), which makes forecasting daily ET difficult. Recently, wavelet transform has become a useful technique for analyzing variations, periodicities and trends in hydrological time series (Labat et al. 2005). Wavelet transforms, which can produce a good local representation of a signal in both time and frequency domains, provides considerable information about the structure of the physical process to be modeled (Li et al. 1999). The wavelet transform can be combined with machine learning, the multivariate relevance vector machine (MVRVM), and the resulting hybrid models can provide a new alternative to the evapotranspiration estimation and forecasting problem, and provide accurate reliable forecasts.

Another critical element in agricultural water management is the capability to develop spatially distributed estimations of ET rates over large irrigated areas. Methods for quantifying ET based on meteorological point measurements are still unable to provide spatial ET at large scales, mostly because of the heterogeneity of the land surface and the dynamic nature of the heat transfer processes (Kaheil et al. 2008). In the recent

years, remote sensing has become a very useful tool for estimating ET at various temporal and spatial scales. Algorithms for mapping ET using energy balance approaches have demonstrated accuracy in determining the spatial distribution of actual ET (Bastiaanssen et al. 2008). For irrigation command area applications, medium spatial resolution images (<100 m) are available from satellites and are used in those algorithms.

Nonetheless, the estimation is restricted to the day when the geospatial information is obtained, which can be up to 16 days for a medium spatial resolution satellite like Landsat. Therefore, daily spatial ET is not available due to temporal resolution of satellites and/or gaps in image acquisition due to cloud cover. Also, without information of precise daily water crop demand, there is a continuous challenge for the implementation of better water distribution and management policies in the irrigation system. As an alternative, data from remote sensing ET algorithms could be used for developing machine learning models, such as relevance vector machines (RVM), that can predict daily ET on spatial level using information that can be available on daily basis. This is the general objective of this work and the information generated by the technology proposed herein should be beneficial to both farmers and operators of irrigation supply/distribution systems.

1.2 Purpose and Objectives

The purpose of this study is to develop methodologies and tools that promote better management of irrigation water and water delivery systems. The work reported here is directed towards forecasted crop water demands. Using machine learning tools in which time and frequency decompositions of the ET time series are applied to Landsat

data, results in models that provide real-time information for use in improved irrigation scheduling at the farm level and enhanced operations of canal deliveries.

The specific objectives of this research were to:

- Assess the performance of the Hargreaves (HG) equation for estimating ET in the semiarid conditions at daily, weekly, and monthly timescales.
- Develop an adjusted form of the HG equation calibrated to semiarid conditions in order to improve the estimation of daily, weekly, and monthly ET. And, to evaluate the performance of the adjusted HG equation if combined with additional weather variables.
- Present a methodology to decompose ET time series in the frequency domain using wavelet analysis.
- Develop data-driven modeling techniques coupled with the wavelet-based multiresolution analysis (MRA) to build a hybrid model that will generate short-term daily ET forecasts. And, to evaluate these hybrid models for accuracy and robustness.
- Develop spatially distributed estimation of ET rates over large irrigated areas using remotely sensed data from Landsat satellite imagery and scientifically accepted algorithms for spatial estimation of ET on a field-by-field basis. And, to evaluate the efficiency of irrigation supply systems in the irrigation command area using spatial estimation of ET.
- Develop a data-driven model to provide a spatial distribution of ET that can be applied even on the days when Landsat is not passing over the area.

1.3 Dissertation Organization

The dissertation has three main components in paper format. In Chapter 2, an assessment of the performance of the Hargreaves (HG) equation for estimating evapotranspiration (ET) in the semiarid conditions of the Bekaa Valley of Lebanon at daily, weekly, and monthly timescales has already been published. This work included an adjusted form of the HG equation calibrated to the study area.

Chapter 3 discusses a methodology combining a wavelet multiresolution analysis (MRA) with a statistical machine learning algorithm, the multivariate relevance vector machine (MVRVM), in order to develop a hybrid model that can forecast daily ET up to 16 days ahead. The models are compared to MVRVM model and evaluated for accuracy and robustness. The addition of 10-days forecasted air temperature to the forecasting models is also investigated in this chapter.

Chapter 4 discusses the application of the remote sensing algorithm METRIC (Mapping Evapotranspiration at High Resolution with Internalized Calibration) to map ET at spatial level with high resolution. It also covers the evaluation of the efficiency of water delivery and distribution systems. This chapter also presents the methodology to develop a relevance vector machine that can provide spatial distribution of ET, it presents the basic details of the RVM and the methodology used to develop the model.

Chapter 5 provides a summary of this work, draws the major conclusions that follow, and presents recommendations for further research.

The structure of this dissertation is based on the multiple-paper format. As a result, some redundancies and repetition of parts of the material presented occur, especially the description of the study area and of the data-driven algorithm used.

References

- Allen, R. G., Pruitt, W. O., Wright, J. L., Howell, T. A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J. B., Smith, M., Raes, D., Perrier, A., Alves, I., Walter, I., Elliot, R. (2006). "A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO 56 Penman-Monteith method." *Agric. Water Manage.*, 81, 1-22.
- Bastiaanssen, W. G. M., Allen, R. G., Pelgrum, H., Teixeira, A. H., Soppe, R. W. O., and Thoreson, B. P. (2008). "Thermal-infrared technology for local and regional scale irrigation analyses in horticultural systems." *Acta Horticulturae*, 792 (Mildura Special Issue, editors I. Goodwin and M.G. O'Connell), ISHS.
- Gill, M. K., Asefa, T., Kemblowski, M. W., and McKee, M. (2006). "Soil moisture prediction using support vector machines." *Am. Water Resour. Assoc.*, 42, 1033-1046.
- Hargreaves, G. H., and Samani, Z.A. (1985). "Reference crop evapotranspiration from temperature." *Appl. Eng. Agric.*, 1(2), 96-99.
- Hernandez, S., Morales, L., and Sallis, P. (2011). "Estimation of reference evapotranspiration using limited climatic data and Bayesian model averaging." *ems*, pp.59-63, 2011 UK Sim 5th European Symposium on Computer Modeling and Simulation.
- Kaheil, Y. H., Rosero, E., Gill, M. K., McKee, M., and Bastidas, L. A. (2008). "Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines." *Geoscience and Remote Sensing, IEEE Transactions on*, 46(9), 2692-2707.
- Labat, D., Ronchail, J., and Guyot, J. L. (2005). "Recent advances in wavelet analyses: Part 2-Amazon, Parana, Orinoco and Congo discharges time scale variability." *J. Hydrology*, 314(1-4), 289-311.
- Li, X., Ding, J., and Li, H. (1999). "Combing neural network models based on wavelet transform." *J. Hydraulic*, 2, 1-4.

CHAPTER 2
ASSESSMENT OF REFERENCE EVAPOTRANSPIRATION BY THE
HARGREAVES METHOD IN THE BEKAA VALLEY, LEBANON[†]

Abstract

Evapotranspiration (ET) is an important component of the hydrologic cycle, especially for irrigated agriculture. Direct methods of estimating reference ET are difficult or require many weather variables that are not always available at all weather stations. The Hargreaves Equation (HG) requires only measured daily air temperature data and computed extraterrestrial radiation for ET estimates. Unless it is regionally calibrated, however, HG often tends to systematically overestimate or underestimate ET. This equation was evaluated under semiarid conditions in the Bekaa Valley of Lebanon using 18 years of complete daily climatic data from the Terbol weather station. HG results were compared to ET estimates obtained from the FAO56 Penman Monteith equation (PM), which was used as a standard. The original HG equation overestimated ET by 23, 17, and 12% for daily, weekly, and monthly ET as compared to PM. The results of a simple linear regression applied to obtain the calibrated HG coefficients for all three time steps showed that the calibrated equation improved the accuracy of the estimation to 3, 2, and 1% difference from ET computed by the PM method, with root mean square error (RMSE) of 0.48, 0.33, and 0.25 mm d⁻¹ for daily, weekly, and monthly ET, respectively. Additional improvement in HG estimation accuracy was achieved by

[†] Reprinted from Journal of Irrigation and Drainage Engineering, ASCE, 10.1061/(ASCE)IR.1943-4774.0000646 (Vol. 139, No. 11, November 1, 2013), Roula Bachour, Wynn R. Walker, Alfonso F. Torres-Rua and Mac McKee, Assessment of Reference Evapotranspiration by the Hargreaves Method for the Bekaa Valley, Lebanon. Copyright (2013), with permission from ASCE

adding the wind speed using a backward variable selection method. This method resulted in only a slight improvement, reaching less than a 1% difference for all timescales and RMSE of 0.46, 0.31, and 0.22 mm d⁻¹ for daily, weekly, and monthly ET, respectively. Thus, when only temperature data are available, the calibrated HG equation is recommended for use in the semiarid conditions of Lebanon, and when complete and reliable weather data exist, the use of the standard FAO56 PM equation is recommended.

2.1 Introduction

Evapotranspiration (ET) is an important component of the hydrologic cycle of agricultural systems, particularly of irrigated agriculture. The steadily increasing world population will place a growing burden on global food security, especially in regions where water supplies may not be sufficient for both food production and potable water. Even more striking is the growing dependence on irrigated production, which results in higher, more dependable yields. It is now reported that 19% of the global agricultural land is irrigated but is producing almost 40% of the total food supply (Food and Agriculture Organization [FAO] 2002). In the United States the irrigated fraction has now reached 18%, but this relatively small area generates 50% of the total value of U.S. cropland production (U.S. Department of Commerce [USDC] 1999).

One assessment indicates that only 10% of the needed increases in food production necessary to support projected population growth can come from expanded development in arable land (Schultz et al. 2005). At most, 13% will be contributed from new cropping systems and increased cropping intensity. The remaining must be met from yield increases and better water use efficiency (FAO 2003). Most studies conclude that much, if not most, of the water for expanded food production and urban requirements will

have to be met by non-technical means such as better and more efficient water management. Thus, in the arid and semiarid regions where irrigation is needed to achieve high crop yields, an accurate assessment of ET is prerequisite to water management practices aimed at improving conservation and yields or mitigating water overutilization and environmental degradation (Hargreaves and Allen 2003).

The FAO Penman-Monteith (PM) model has become the generally accepted standard for calculating ET (Allen et al. 1998, 2006). Input required for the PM computations include several climatic variables such as air temperature, relative humidity, wind speed, and solar radiation that are not always available, especially in developing countries. Where climatic data are limited, the equation developed by Hargreaves and Samani (1985) can be used with reasonable results (Allen et al. 1998). The Hargreaves Equation (HG) requires only measured daily air temperature data and computed extraterrestrial radiation (Droogers and Allen 2002). Hargreaves and Allen (2003) stated that the HG method produces best results for weekly or longer period estimates of ET because daily estimates are subject to errors caused by the influence of the temperature range, wind speed, and cloud cover. Samani (2004) suggested that the HG equation should not be overextended to different climatic conditions unless it has been calibrated to the specific area. Calibration is usually accomplished by comparison with the PM method.

Several attempts to assess the performance of the HG equation have been made. In some cases, the results indicated a close alignment with results given by PM (Droogers and Allen 2002; Hargreaves and Allen 2003; Trajkovic and Kolakovic 2009; Tabari 2010). A tendency to overestimate ET by the HG method was reported in inland areas

where temperature differences are high and wind speed is low, while underestimation was reported mainly in coastal areas (Droogers and Allen 2002; Vanderlinden et al. 2004; Gavilán et al. 2006; Jabloun and Sahli 2008; Martinez and Thepadia 2010; Azhar and Perera 2011; Mendicino and Senatore 2012; Tabari et al. 2013). However, it is possible to improve the accuracy of the HG equation either by adjusting the parameters to local conditions (Droogers and Allen 2002; Vanderlinden et al. 2004; Rahimi Khoob 2008; Zhai et al. 2010) or by adding other variables such as wind speed or elevation atmospheric pressure (Allen 1993, 1995; Jensen et al. 1997; Droogers and Allen 2002; Marinez-Cob and Tejero-Juste 2004; Fooladmand et al. 2008; Fooladmand 2011; Xystrakis and Matzarakis 2011; Ravazzani et al. 2012). Whenever complete reliable data are not available, the HG equation is used after local calibrations for a specific climate or region (Fooladmand et al. 2008; Martinez and Thepadia 2010).

In Lebanon, more than 70% of surface and groundwater is diverted for irrigated agriculture, making the accurate estimation of ET very important. The Bekaa Valley, with a semiarid climate, is the major agricultural region in Lebanon, and the evaluation of ET equations specific to this area is of major interest. Although few weather stations in the region have complete data for use in PM equation, many have air temperature and precipitation data. Therefore, the main objective of this paper is to develop an adjusted form of the HG equation that is calibrated to the Bekaa valley at daily, weekly, and monthly timescales. A second objective is to assess the performance of the adjusted HG equation while adding different weather variables in order to assess what additional weather data would improve the estimation of ET.

2.2 Methodology

2.2.1 Site Description

The International Center for Agricultural Research in the Dry Areas (ICARDA) research center in Terbol, Lebanon operates a weather station near the center of the Bekaa Valley (Fig. 2.1) in which daily solar radiation, minimum and maximum temperatures, wind speed, precipitation, and relative humidity are recorded. The weather station is located at 33°49' N latitude, 35° 59' E longitude and 890 m above sea level, in the middle of a 1.2 ha (110 x 110m) plot of well irrigated grass maintained at a height of 10–15 cm. It is also surrounded by irrigated agricultural crops. The temperature and humidity sensors, and the anemometer are placed at 1.5 and 2 m above the soil surface, respectively. The quality and integrity of the weather data were assessed before being used in the study following the procedures described in Allen (1996). Data for the days for which one or more weather variable were not available were excluded from the analysis. The Bekaa Valley is characterized by a semiarid climate, with cold winters, dry summers, and favorable growing conditions; it has an average annual precipitation of 500–600 mm. The average monthly weather data from 1993–2010 are presented in Table 2.1, which shows January as the coldest month, with a mean temperature of 5.2°C, and August as the warmest month, with a mean temperature of 23°C. The average wind speed throughout the year is 1.9 m s⁻¹, with 58.4 percent relative humidity, and 18.8 MJ m⁻² d⁻¹ average annual solar radiation.

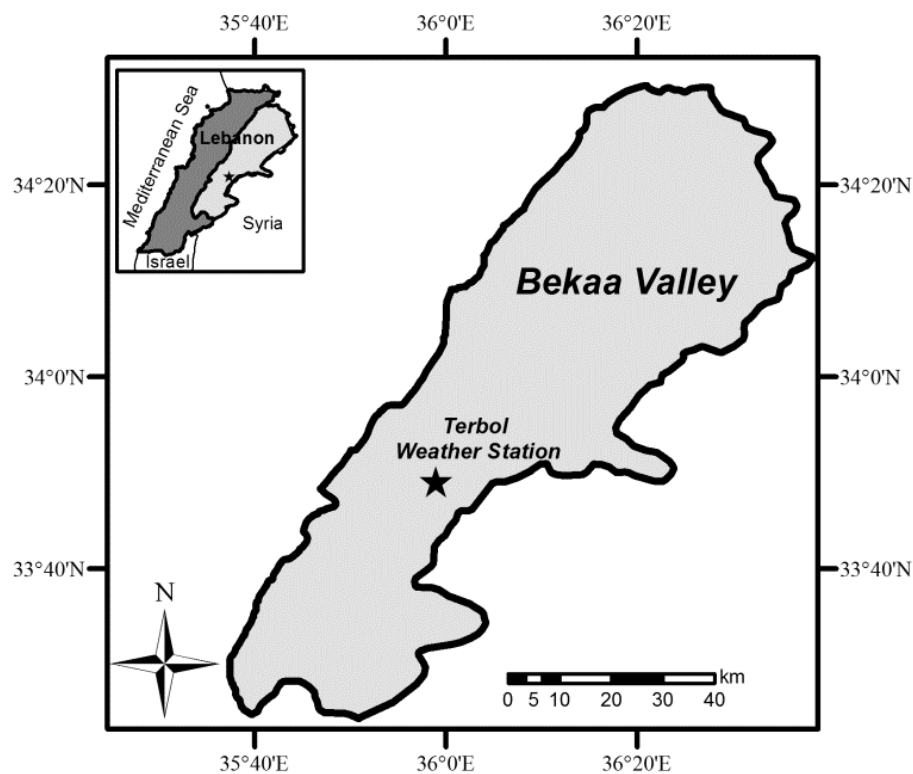


Fig. 2.1. Location of Terbol weather station in the Bekaa Valley, Lebanon

Table 2.1. Monthly Averages of Weather Variables at Terbol Station During 1993–2010

Months	T_{\max} (°C)	T_{\min} (°C)	T_{mean} (°C)	RH (%)	R_s (MJ/m ²)	U_2 (m/s)	P (mm)
January	11.5	-1.2	5.2	66.9	8.7	1.7	124
February	12.7	-0.4	6.1	65.0	11.5	2.1	114
March	16.6	1.3	8.9	60.0	16.5	2.2	71
April	21.3	3.8	12.6	57.3	21.3	2.1	26
May	27.3	6.8	17.0	53.4	26.6	1.9	5
June	31.4	9.1	20.3	51.3	29.4	2.2	0
July	33.8	11.5	22.7	52.2	28.9	2.4	0
August	34.3	11.8	23.0	53.2	26.3	2.1	0
September	31.1	9.8	20.4	55.7	21.8	2.0	2
October	26.1	7.1	16.6	57.4	15.4	1.6	20
November	18.7	2.6	10.6	61.7	10.9	1.5	61
December	13.8	0.0	6.9	66.8	8.1	1.5	87

Note: T_{\max} is the maximum air temperature (°C), T_{\min} is the minimum temperature (°C), T_{mean} is the mean monthly temperature (°C), RH is the average relative humidity (%), R_s is the solar radiation (MJ/m²), U_2 is the wind speed at 2m (m/s), and P is the monthly precipitation (mm)

2.2.2 Estimating Reference Evapotranspiration

2.2.2.1 FAO-56 Penman-Monteith

Daily values of reference crop ET were made from the Terbol weather data for the period 1993–2010 using the following PM equation as suggested by Allen et al. (1998):

$$ET_{PM} = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (2.1)$$

In Eq. (2.1), ET_{PM} = reference evapotranspiration (mm d^{-1}); Δ = slope vapor pressure curve ($\text{kPa } ^\circ\text{C}^{-1}$); R_n = net radiation at the crop surface ($\text{MJ m}^{-2} \text{d}^{-1}$); G = soil heat flux density [$\text{MJ m}^{-2} \text{d}^{-1}$]; γ = psychrometric constant ($\text{kPa } ^\circ\text{C}^{-1}$); T = mean daily air temperature at 2 m height ($^\circ\text{C}$); U_2 = wind speed at 2 m height (m s^{-1}); e_s = saturation vapor pressure (kPa); e_a = actual vapor pressure (kPa); $e_s - e_a$ = saturation vapor pressure deficit (kPa); and it is assumed that Eq. (1) is applied to a hypothetical crop with a height of 0.12 m, having a surface resistance of 70 s m^{-1} and an albedo $\alpha = 0.23$. Soil heat flux was ignored for daily and weekly time steps, while for longer time steps it was calculated according to Allen et al. (1998). The computations of all required data for calculating ET were accomplished using the method given in Chapter 3 of FAO paper 56 (Allen et al. 1998).

2.2.2.2 Hargreaves Method

The ET was also estimated using the HG equation according to Hargreaves and Samani (1985):

$$ET_{HG} = 0.0023R_a(T + 17.8)\sqrt{T_{\max} - T_{\min}} \quad (2.2)$$

with,

$$R_a = 37.6d_r[\omega_s \sin(\varphi) \sin(\delta) + \cos(\varphi) \cos(\delta) \sin(\omega_s)] \quad (2.3)$$

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi}{365} J\right) \quad (2.4)$$

$$\omega_s = ar \cos[-\tan(\varphi) \tan(\delta)] \quad (2.5)$$

$$\delta = 0.409 \sin\left(\frac{2\pi}{365} J - 1.39\right) \quad (2.6)$$

where: ET_{HG} = reference evapotranspiration (mm d^{-1}); R_a = extraterrestrial radiation (mm d^{-1}) [Eq. (2.3)]; T , T_{max} and T_{min} = mean, maximum and minimum temperature ($^{\circ}\text{C}$), respectively; d_r = inverse relative distance Earth-Sun [Eq. (2.4)]; ω_s = sunset angle (radians) [Eq. (2.5)]; φ = latitude (radians); δ = solar declination (radians) [Eq. (2.6)]; and J = the day of the year.

2.2.2.3 Calibrated Hargreaves Equation

To acquire an improved estimate of ET with limited weather data, the parameters of the HG equation can be calibrated to fit the local conditions. A simple linear regression was applied to fit the two empirical constants of the HG equation to ET values calculated using the PM method. The equation was calibrated by finding new values for the HG empirical constants that minimize the root mean square error (RMSE) between the ET values determined using the PM method and those obtained from this modified HG equation (HG_{mod}). This was done for ET estimations at daily, weekly, and monthly timescales.

The significance of all available weather parameters was also tested to determine if the modified HG equation could be further improved. This was done using a backward variable selection method where all the weather parameters are included in the model as

variables. These variables are used as potential predictors, and then the least significant variable with the highest P-value is dropped. This step is repeated successively until all of the remaining variables are statistically significant at the $\alpha = 0.05$ level.

2.2.3 Evaluation Procedures

All comparisons between the different forms of the equations were performed by simple linear regression $y = b_0 + b_1x$, where y is the dependent variable, ET_{PM} , x is the independent variable (ET by the different HG forms), b_0 is the intercept, and b_1 is the slope. The coefficient of determination, R^2 , and the RMSE were used for evaluating the different equations. These were computed as:

$$R^2 = \frac{\left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.7)$$

$$RMSE = \left[\frac{\sum_{i=1}^n (y_i - x_i)^2}{n} \right] \quad (2.8)$$

where, y_i = estimated ET by the PM method for day i (mm d^{-1}); x_i = estimated ET by the different types of the HG equation for day i (mm d^{-1}); \bar{x} and \bar{y} = average of x_i and y_i ; and n = total number of observations.

2.3 Results and Discussion

2.3.1 Evaluation of the Hargreaves Method

The original HG method [Eq. (2.2)] was compared to PM [Eq. (2.1)] for daily, weekly and monthly ET estimates. The HG method showed an overestimation of ET for

all time steps, as shown in Fig. 2.2. For daily ET, the overestimation is about 23% (Fig. 2.2[a]), which agrees with the findings of Vanderlinden et al. (2004), Jabloun and Sahli (2008), Martinez and Thepadia (2010), Tabari et al. (2013), and others who reported overestimation by the HG method in interior areas. Therefore, the HG method is not suggested for use in the semiarid conditions of the Bekaa Valley without being modified to fit the local conditions. The deviation from the results of the PM equation is reduced with longer time steps. The differences between the HG and PM results were 17% for weekly and 12% for monthly ET estimates (Fig. 2.2[b and c]), which confirms the recommendation made by Hargreaves and Allen (2003) to use the HG method for long time steps due to the fact that the daily fluctuation of temperature, ΔT , and other weather variables can change ET estimation significantly.

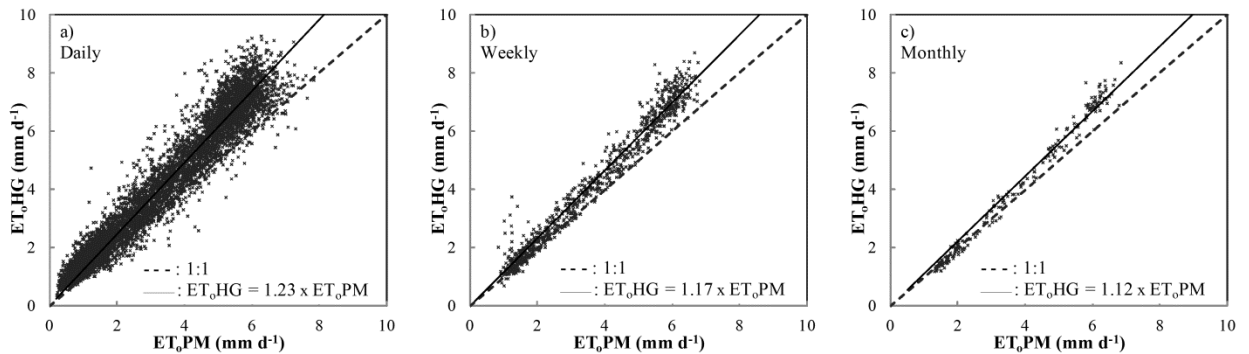


Fig. 2.2. Comparison of ET_{PM} with original HG for (a) daily; (b) weekly; (c) monthly time steps for 16 years

2.3.2 Local Calibration

The simplest modification to the Hargreaves equation is to calibrate the equation by finding values for the two constant parameters that minimize the RMSE between ET values determined using the PM method as recommended by Allen (1993), Droogers and Allen (2002), Hargreaves and Allen (2003), and Jabloun and Sahli (2008).

A linear regression was applied to the entire data set ($n = 6,223$) in order to get new coefficients for the HG equation. The resulting modified HG (HG_{mod1}) equation forms are as follows:

$$\text{For daily estimates: } HG_{\text{mod1}}^d = 0.0019R_a(T + 15.8)\sqrt{T_{\text{max}} - T_{\text{min}}} \quad (2.9)$$

$$\text{For weekly estimates: } HG_{\text{mod1}}^w = 0.0020R_a(T + 16.5)\sqrt{T_{\text{max}} - T_{\text{min}}} \quad (2.10)$$

$$\text{For monthly estimates: } HG_{\text{mod1}}^m = 0.0020R_a(T + 18.4)\sqrt{T_{\text{max}} - T_{\text{min}}} \quad (2.11)$$

Several studies have addressed the accuracy of the HG equation and have produced similar coefficient changes: Allen (1993) changed the coefficients to 0.0030 and 20 for Davis, California, while Droogers and Allen (2002) found 0.0025 and 16.8 to be good calibrating coefficients for the International Water Management Institute (IWMI) Climate Atlas monthly data grids. Gavilán et al. (2006) reported the coefficients 0.0021 and 0.0027 for the Andalusia (Spain) regional calibration. Martinez-Cob and Tejero-Juste (2004) recommended using the coefficient 0.0020 for non-windy locations instead of 0.0023, as originally proposed by Hargreaves and Samani (1985). ET results calculated by the modified HG equations were again compared to ET estimates by the PM method and are plotted in Fig. 2.3. An improvement in ET estimation is obvious in Fig. 2.3, which shows the relationship between modified HG and PM for daily, weekly, and monthly ET. Table 2.2 lists the statistics of the comparisons between the PM method and the different forms of the HG equation reported here. The agreement between PM and original HG increased from daily to weekly time steps, reaching a close correlation ($R^2 = 0.98$) and very low RMSE for monthly time steps (0.29 mm d^{-1}). The original HG equation was overestimating ET by 23, 17 and 12% for daily, weekly, and monthly ET, respectively. When the HG equation coefficients were modified [Eqs. (2.9),(2.10) and

(2.11)], the improvement in the estimates was significant. ET estimated by the modified HG equation (HG_{mod1}) was on average 2.6% lower than PM for daily estimation, with a RMSE of 0.48 mm d^{-1} , while for weekly averages it was 2.2% less, with a RMSE of 0.33 mm d^{-1} . Finally, for monthly ET, HG_{mod1} was 1.3% less than PM, with a RMSE of 0.25 mm d^{-1} . These results show the importance of adjusting empirical equations.

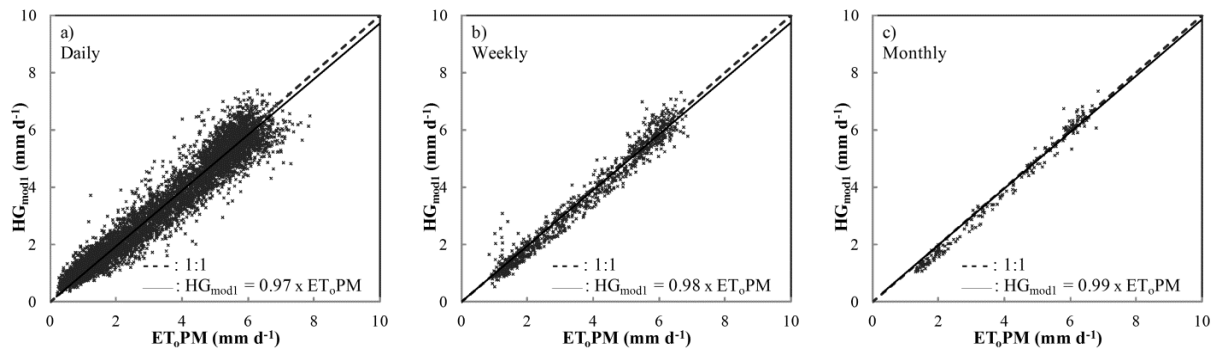


Fig. 3. Comparison of ET_{PM} with modified HG (HG_{mod1}) for (a) daily; (b) weekly; (c) monthly time steps for 16 years

After modifying the HG equation coefficients, and in order to determine whether a better estimation of daily and weekly ET could be achieved, a backward variable selection method was applied that identifies the important weather variables to be added to the modified equations for a more accurate ET estimate. All weather variables were

Table 2.2. Comparison of ET Estimated by FAO-PM Equation and the Different Forms of HG Equations for Three Timescales

Equation Form	Daily			Weekly			Monthly		
	Fit	RMSE (mm d^{-1})	R^2	Fit	RMSE (mm d^{-1})	R^2	Fit	RMSE (mm d^{-1})	R^2
ET_HG	$1.232 \times ET^{PM}$	0.600	0.930	$1.166 \times ET^{PM}$	0.387	0.969	$1.117 \times ET^{PM}$	0.295	0.982
HG_{mod1}	$0.974 \times ET^{PM}$	0.481	0.930	$0.978 \times ET^{PM}$	0.333	0.968	$0.987 \times ET^{PM}$	0.253	0.983
HG_{mod2}	$1.009 \times ET^{PM}$	0.462	0.937	$1.009 \times ET^{PM}$	0.312	0.972	$1.014 \times ET^{PM}$	0.221	0.987

Note: RMSE is the root mean square error; R^2 is the coefficient of determination

included in the model, and then the least significant variable was dropped. This step was repeated successively until all of the remaining variables were statistically significant at $\alpha = 0.05$. For ET estimation using the adjusted HG method, wind speed was selected by the model as the significant variable for all time steps. This selection agrees with the recommendations of Martinez-Cob and Tejero-Juste (2004) for semiarid regions, although Ravazzani et al. (2012) noted the importance of adding altitude to the calibrated HG equation. Since the wind speed variable was shown to be of more importance to the model, it was added to the modified HG equation forms, and the following equations ($HG_{\text{mod}2}$) were evaluated:

$$\text{Daily: } HG_{\text{mod}2}^d = 0.0019R_a(T + 15.8)\sqrt{T_{\text{max}} - T_{\text{min}}} + 0.0765U \quad (2.12)$$

$$\text{Weekly: } HG_{\text{mod}2}^w = 0.0020R_a(T + 16.5)\sqrt{T_{\text{max}} - T_{\text{min}}} + 0.0704U \quad (2.13)$$

$$\text{Monthly: } HG_{\text{mod}2}^m = 0.0020R_a(T + 18.4)\sqrt{T_{\text{max}} - T_{\text{min}}} + 0.0608U \quad (2.14)$$

Eqs. (2.12), (2.13) and (2.14) were used to estimate ET, and the results were compared to the standard ET_{PM} as plotted in Fig. 2.4. The estimation showed slight improvement over the results given by $HG_{\text{mod}1}$ when compared to the PM method, with a 1% overestimation on average for all time steps. The new form, $HG_{\text{mod}2}$, produced ET estimates 0.9% higher on average than ET_{PM} for both daily and weekly ET, and 1.4% higher for monthly ET values. Adding the wind speed (U) variable to the modified HG equation further improved the estimation compared to PM as presented in Table 2.2, showing performance similar to that cited by Droogers and Allen (2002), who added relative humidity to the equation when reliable data are available, or Ravazzani et al. (2012), who added altitude to get a better estimation by the HG method in different regions. The addition of the wind variable showed to be valuable for more accurate ET

estimation, but HG_{mod2} should be used only when reliable wind speed data is available. However, when complete weather data are available, the standard FAO56 PM remains the most reliable method for ET estimates as reported by Allen (1993), Hargreaves and Allen (2003), and other researchers who have attempted to modify HG for different regions. Hence, this study suggests the use of the modified HG (HG_{mod1}) form for daily, weekly, and monthly estimates in the Bekaa Valley when complete and reliable weather data are limited; otherwise, the standard FAO56 PM method is recommended.

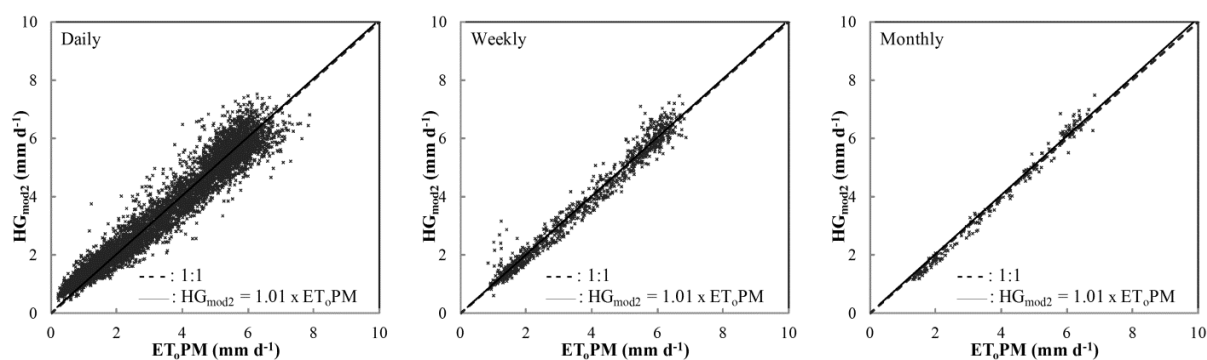


Fig. 2.4. Comparison of ET_{PM} with wind-modified HG (HG_{mod2}) for (a) daily; (b) weekly; (c) monthly time steps for 16 years

It is important to note that in this study, lysimeter data were not available to evaluate the results, and it was assumed that the FAO56 PM can be used as the sole standard for obtaining ET estimates, as suggested by Allen et al. (1998). However, the semiarid conditions in the Bekaa Valley of Lebanon differ from the conditions under which the PM equation was developed, where the ET estimates represent water use under well-watered conditions. Also, this study used data from only one weather station to develop the suggested equations that are generalized for the whole region. ET estimations using data from a set of recently installed regional agrometeorological stations that was

recently installed could be a subject for further research and validation of the suggested calibrations.

2.4 Conclusions

The comparisons of daily, weekly, and monthly estimates of ET showed that the original HG method overestimated ET compared to the PM method in the Bekaa Valley of Lebanon by 23, 17, and 12%, respectively. The original HG equation was modified, and better estimates were achieved when the coefficients of the HG equation were calibrated as suggested by Hargreaves and Allen (2003). The regional calibration of the original HG equation (HG_{mod1}) produced better results in this regard; daily, weekly and monthly ET differed, respectively, by 3, 2, and 1% from the PM method. This suggests that when only air temperature data are available, the calibrated HG equation should be used. When adding wind speed to the original HG equation, the estimation improved to reach an average difference of 1% from PM for all time steps, showing the importance of the wind speed variable in ET estimation for the Bekaa Valley. When reliable and complete weather records are available, the PM method remains the most reliable method to calculate reference evapotranspiration for daily estimates. However, the proposed calibrations of HG can be a feasible approach to be used for ET estimation in the Bekaa Valley when data on weather parameters are limited.

References

- Allen R. G. (1993). "Evaluation of a temperature difference method for computing grass reference evapotranspiration." *Report submitted to the Water Resources Development and Man Service, Land and Water Development Division, FAO, Rome.* 49 pp.
- Allen, R. G. (1995). "Evaluation procedures for estimating grass reference evapotranspiration using air temperature data only." *Report submitted to Water*

- Resources Development and Management Service, Land and Water Development Division, United Nations Food and Agriculture Service, Rome, Italy.*
- Allen R. G. (1996). "Assessing integrity of weather data for a reference evapotranspiration estimation." *J. Irrig. Drain. Eng. ASCE*, 122(2), 97-106.
- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). "Crop evapotranspiration: guidelines for computing crop water requirements." *FAO Irrigation and Drainage Paper No. 56*, FAO, Rome, 300.
- Allen, R. G., Pruitt, W. O., Wright, J. L., Howell, T. A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J. B., Smith, M., Raes, D., Perrier, A., Alves, I., Walter, I., and Elliot, R. (2006). "A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO 56 Penman-Monteith method." *Agric. Water Manage.*, 81(1-2), 1-22.
- Azhar, A., and Perera, B. (2011). "Evaluation of reference evapotranspiration estimation methods under southeast Australian conditions." *J. Irrig. Drain. Eng. ASCE*, 137(5), 268-279.
- Droogers, P., and Allen, R. G. (2002). "Estimating reference evapotranspiration under inaccurate data conditions." *Irrig. and Drain. Syst.*, 16, 33-45.
- FAO. (2002). "Assessment of the World Food Security Situation." Committee on World Food Security, 28th session. June 6-9, 2002. Rome.
- FAO. (2003). *World agriculture: Toward 2015/2030, An FAO perspective*. Earthscan Publications Ltd, London.
- Fooladmand, H. (2011). "Evaluation of some equations for estimating evapotranspiration in the south of Iran." *Archives of Agron. and Soil Sci.*, 57(7), 741-752.
- Fooladmand, H. R., Zandilak, H., and Ravanan, M. H. (2008). "Comparison of different types of Hargreaves equation for estimating monthly evapotranspiration in the south of Iran." *Archives of Agron. and Soil Sci.*, 54(3), 321-330.
- Gavilán, P., Lorite, I., Tornero, S., and Berengena, J. (2006). "Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment." *Agric. Water Manage.*, 81(3), 257-281.
- Hargreaves, G. H., and Allen, R. G., (2003). "History and evaluation of Hargreaves evapotranspiration equation." *J. Irrig. Drain. Eng. ASCE*, 129(1), 53-63.
- Hargreaves, G. H., and Samani, Z. A. (1985). "Reference crop evapotranspiration from temperature." *Appl. Eng. Agric.*, 1(2), 96-99.

- Jabloun, M., and Sahli, A. (2008). "Evaluation of FAO-56 methodology for estimating reference evapotranspiration using limited climatic data Application to Tunisia." *Agric. Water Manage.*, 95(6), 707-715.
- Jensen, D. T., Hargreaves, G. H., Temesgen, B., and Allen, R. G. (1997). "Computation of ET_0 under nonideal conditions." *J. Irrig. and Drain. Eng.*, 123(5), 394-400.
- Martínez-Cob, A., and Tejero-Juste, M. (2004). "A wind-based qualitative calibration of the Hargreaves ET_0 estimation equation in semiarid regions." *Agric. Water Manage.*, 64(3), 251-264.
- Martínez, C. J., and Thepadia, M. (2010). "Estimating reference evapotranspiration with minimum data in Florida." *J. Irrig. Drain. Eng. ASCE*, 136(7), 494-501.
- Mendicino, G., and Senatore, A. (2012). "Regionalization of the Hargreaves Coefficient for the assessment of distributed reference evapotranspiration in southern Italy." *J. Irrig. Drain Eng. ASCE*, 139(5), 349-362.
- Rahimi Khoob, A. (2008). "Comparative study of Hargreaves's and artificial neural network's methodologies in estimating reference evapotranspiration in a semiarid environment." *Irrig. Sci.*, 26(3), 253-259.
- Ravazzani, G., Corbari, C., Morella, S., Gianoli, P., and Mancini, M. (2012). "Modified Hargreaves-Samani equation for the assessment of reference evapotranspiration in Alpine River Basins." *J. Irrig. Drain. Eng. ASCE*, 138(7), 592-599.
- Samani, Z. (2004). "Discussion of "History and Evaluation of Hargreaves Evapotranspiration Equation"." *J. Irrig. Drain. Eng. ASCE*, 130(5), 447-448.
- Schultz, B., Thatte, C. D., and Labhsetwar, V. K. (2005). "Irrigation and Drainage – Main Contributors to Global Food Production." *Irrig. and Drain.*, 54(3), 263-278.
- Tabari, H. (2010). "Evaluation of reference crop evapotranspiration equations in various climates." *Water Resour. Manage.*, 24(10), 2311-2337.
- Tabari, H., Grismer, M., and Trajkovic, S. (2013). "Comparative analysis of 31 reference evapotranspiration methods under humid conditions." *Irrig. Sci.*, 31(2), 107-117.
- Trajkovic, S., and Kolakovic, S. (2009). "Evaluation of reference evapotranspiration equations under humid conditions." *Water Resour. Manage.*, 23(14), 3057-3067.
- USDC. (1999). *1997 Census of Agriculture*. US Department of Commerce, Washington, D. C.

- Vanderlinden, K., Giraldez, J. V., and Van Meirvenne, M. (2004). "Assessing reference evapotranspiration by the Hargreaves method in southern Spain." *J. Irrig. Drain. Eng. ASCE*, 130(3), 184-191.
- Xystrakis, F., and Matzarakis, A. (2011). "Evaluation of 13 empirical reference potential evapotranspiration equations on the Island of Crete in southern Greece." *J. Irrig. Drain. Eng. ASCE*, 137(4), 211-222.
- Zhai, L., Feng, Q., Li, Q., and Xu, C. (2010). "Comparison and modification of equations for calculating evapotranspiration (ET) with data from Gansu Province, Northwest China." *Irrig. and Drain.*, 59(4), 477-490.

CHAPTER 3
WAVELET-MULTIVARIATE RELEVANCE VECTOR MACHINE
HYBRID MODEL FOR FORECASTING DAILY
EVAPOTRANSPIRATION[‡]

Abstract

Evapotranspiration (ET) is one of the main components of the hydrological cycle. It is a complex process driven mainly by weather parameters, and as such, is characterized by high non-linearity and non-stationarity. This paper introduces a methodology that combines wavelet multiresolution analysis (MRA) with a machine learning algorithm, the multivariate relevance vector machine (MVRVM), in order to predict 16 days of future daily reference ET quantities. This methodology lays the ground for forecasting the spatial distribution of ET using Landsat satellite imagery, hence the choice of 16 days, which corresponds with the Landsat overpass cycle. An accurate prediction of daily ET is needed to improve the management of irrigation schedules as well as the operations of water supply facilities like canals and reservoirs. In this paper, various wavelet decompositions have been performed and combined with the MVRVM to develop the hybrid models to predict reference ET over a 16-day period. These models were compared to a MVRVM model, and models accuracy and robustness were evaluated. The addition of 10 days of forecasted air temperature as additional inputs to the forecasting models has also been investigated in this paper. The results of the wavelet-MVRVM hybrid modeling methodology presented in this study show that a

[‡] Co-authored by Inga Maslova, Andres M. Ticlavilca, Wynn R. Walker, and Mac McKee

reliable forecast of ET up to 16 days ahead is possible.

3.1 Introduction

Evapotranspiration (ET) is a complex process affected by several environmental factors and driven mainly by weather parameters. Numerous analytical methods have been proposed to estimate ET on the basis of measured weather parameters. The Food and Agricultural Organization of the United Nations (FAO) Penman-Monteith model has become a generally accepted standard for calculating ET (Allen et al. 1998, 2006). The relationship between ET and its driving factors is complicated and not easily modeled (Partal 2009; Torres et al. 2011). The non-stationarity nature of ET time series leads to difficulties in forecasting future values (Pandey et al. 2009; Hernandez et al. 2011).

Time series analysis techniques have been widely used for modeling and predicting different hydrological parameters including ET (Mariño et al. 1993; Cigizoglu 2003; Gorantiwar et al. 2011). Several researchers have found that the seasonal autoregressive integrated moving average (SARIMA) model provides good forecasts of monthly and weekly ET (Trajkovic 1998; Landeras et al. 2009). However, difficulties related to these techniques has motivated researchers to look for other modeling approaches including the use of data-driven tools or statistical learning machines, such as artificial neural networks (ANN), multiple regression methods, support vector machines (SVM), and relevance vector machines (RVM). For instance, Landeras et al. (2009) showed that ANN's have good performance for weekly ET forecasts. Trajkovic et al. (2003) developed radial basis ANN's for predicting ET using limited weather data. Kisi (2007) estimated daily ET using ANN methods and compared ANN test results to those of the Penman, Hargreaves and Turc empirical models. Although, ANN's have been used

extensively as a useful tool for prediction, they have difficulty dealing with non-stationary data (Cannas et al. 2006; Partal 2009). RVM's were used extensively by many researchers for modeling and forecasting hydrological parameters. Unfortunately, these models are limited to one step ahead of forecasting, i.e. single output. When predictions are to be made for multiple steps in time, multiple outputs, the Multivariate Relevance Vector Machine (MVRVM) reported by Thayanathan et al. (2008) has proven more effective (Ticlavilca et al. 2011; Ticlavilca and McKee 2011; Torres et al. 2011).

In addition to its ability to predict multiple outputs, the MVRVM is a Bayesian regression tool that has the same properties of the conventional RVM: high prediction accuracy, robustness, and estimation of uncertainty in the predictions. The MVRVM is an extension of the RVM algorithm developed by Tipping and Faul (2003) to produce multivariate outputs when given a set of inputs. Therefore, developing a model with all these properties provides a good forecasting tool to produce multiple predictions that are difficult or not practical to obtain from traditional modeling approaches. Thus, the MVRVM algorithm was used in this study to forecast daily ET values for multiple future time steps.

In the last decade, wavelet transformation has become a useful technique for analyzing variations, periodicities and trends in time series (Labat et al. 2005; Chou and Wang 2002; Küçük et al. 2009). Wavelet transforms, which can produce a good local representation of a signal in both time and frequency domains, provide considerable information about the structure of the physical process to be modeled (Li et al. 1999). Recently, there has been an increased interest in the use of wavelet analyses in a wide range of fields related to water resources. Labat (2005) reviewed the most recent wavelet

applications in the field of earth sciences and illustrated new wavelet analysis methods in the field of hydrology. Most of these studies demonstrated that the application of wavelets leads to several improvements in the analysis of global hydrological signal fluctuations and of their mutual time varying relationships.

Wavelets provide a formal method to break down a complex time series into simpler units to facilitate accurate prediction (Ahmad et al. 2005; Cobaner 2013). Wavelet transform analysis appears to be a more effective technique than the Fourier transform for studying non-stationary time series and is preferred to a windowed Fourier transform (Labat et al. 2005; Partal and Cigizoglu 2008). However, there have been very few applications of wavelet transform techniques to evapotranspiration modeling (Cobaner 2013; Kisi 2011).

Researchers have developed hybrid models that combine wavelet transforms with time series models and artificial intelligence algorithms. In such approaches, a wavelet transform is used first to decompose the time series into varying scales of temporal resolution. Then, a time series model or a regression model is applied. The results of such hybrid models show significant advantages over traditional time series analysis and prediction. For example, Kisi (2011) modeled daily ET using a wavelet regression model and resulted in better results than empirical models. Partal (2009) modeled ET using wavelet decomposition and neural networks showing improvement of ET modeling with the hybrid models. Wang and Luo (2007) combined the wavelet transformation and neural network techniques and were able to develop a wavelet-neural network hybrid model to forecast one day ahead of ET. Since the evapotranspiration process is characterized by high non-linearity and non-stationarity (Hernandez et al. 2011), hybrid

wavelet models provide a new alternative to the evapotranspiration estimation and forecasting problem.

None of the aforementioned studies have attempted to forecast ET multiple several days ahead using the wavelet hybrid models. This type of forecasts are very important for irrigation management, especially when they can be extrapolated spatially. The spatial distribution of ET is usually done using remote sensing algorithms and, for example, Landsat satellite imagery. The latter data set is only once available every 16 days over a specific area for a specific Landsat satellite. Hence, forecasting ET up to 16 days ahead could be beneficial to manage irrigation schedules at the farm level as well and operations of canal delivery and distribution systems. Therefore, the objective of this study was to develop a wavelet-MVRVM model to forecast daily ET simultaneously, up to 16 days ahead. Wavelet-based decompositions were performed and combined with the MVRVM. The performance and accuracy of these hybrid models are then compared to the performance of a MVRVM model.

The remainder of the paper describes the data used in this study (Section 3.2.1), the wavelet multiresolution analysis (MRA) is presented in Section 3.2.2, then a description of the MVRVM learning model in Section 3.2.3. Section 3.3 summarizes the selected wavelet-MVRVM hybrid models, and their forecasting results along with a discussion of these results. Finally the conclusions that can be drawn and future work are shown in Section 3.4.

3.2 Methodology

3.2.1 Data Collection and Description

The weather data for this study were taken from the meteorological station located in Delta, Utah. This station is a part of the Community Environmental Monitoring Program (CEMP) network of 29 monitoring stations located in the Western states. It is operated and monitored by the Desert Research Institute (DRI) of the Nevada System of Higher Education. The station is located at 39°21'11"N latitude, 112°34'42"W longitude and 1415 m above sea level. It records daily solar radiation, minimum and maximum temperatures, wind speed, precipitation and relative humidity. Records over the full period of January 2002 until June 2012 were used in this study. These data were available on the CEMP website (CEMP 2012).

Delta is characterized by a semiarid to arid climate, with an average annual precipitation of 200 mm. The average monthly weather data from 2002 until 2012 showed the coldest month to be January, with a minimum temperature of -7.5°C, and the hottest month July, with 34.3°C as average maximum temperature. The average wind speed throughout the year was 1.14 m s⁻¹, with 148 W m⁻² average solar radiation per day.

The daily values of reference crop evapotranspiration (ET) used in this study were calculated from the Delta weather data using FAO 56 Penman-Monteith (PM) equation as suggested by Allen et al. (1998):

$$ET = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (3.1)$$

where, ET = reference evapotranspiration (mm d^{-1}); Δ = slope vapor pressure curve ($\text{kPa } ^\circ\text{C}^{-1}$); R_n is the net radiation at the crop surface ($\text{MJ m}^{-2} \text{d}^{-1}$); G = soil heat flux density ($\text{MJ m}^{-2} \text{d}^{-1}$); γ is the psychrometric constant ($\text{kPa } ^\circ\text{C}^{-1}$); T = mean daily air temperature at 2 m height ($^\circ\text{C}$); U_2 = wind speed at 2 m height (m s^{-1}); e_s = saturation vapor pressure (kPa); e_a = actual vapor pressure (kPa); $(e_s - e_a)$ = saturation vapor pressure deficit (kPa). It is assumed that Eq. (3.1) is applied to a hypothetical crop with a height of 0.12 m, having a surface resistance of 70 s m^{-1} and an albedo $\alpha = 0.23$. Soil heat flux was ignored as suggested by Allen et al. (1998) for daily time steps. The computations of all required data for calculating ET were done using the method given in Chapter 3 of FAO paper 56 (Allen et al. 1998).

The data sample consisted of 11 years (2002–2012) of daily ET records. The first seven years (2002–2008) were used for training, two years (2009–2010) for calibration, and the remaining data have been used for testing (2011–2012). Only the growing season data (April 1 till October 31) were considered in this study.

3.2.2 Wavelet Multiresolution Analysis

Wavelet multiresolution analysis was used in this paper to study the ET characteristics in time and frequency domains. One of the advantages of wavelet-based techniques is the ability to deal with non-stationary data. It is an alternative to windowed Fourier transform that requires selection of a window where data are stationary and assumes that time series variability pattern stays the same over time, which is not the case with ET series analyzed in this paper.

There are two main types of discrete wavelet transforms: orthogonal, usually referred to as discrete wavelet transform (DWT) and non-orthogonal, also known as maximal overlap discrete wavelet transform (MODWT) (see Torrence and Compo 1998). DWT decomposes the original signal into components at dyadic frequencies. It uses the so-called mother and father wavelets to capture the detailed and smooth parts of a signal. Mother wavelets are used to extract the high-frequency components of the signal, father wavelets - the low-frequency components. The signal is then represented by its features, which are referred to as wavelet and scaling coefficients (Mallat 1989; Daubechies 1992). However, the DWT suffers from a lack of translation invariance. This means that circularly shifting a time series will not necessarily shift its DWT coefficients in a similar manner (Daubechies 1992; Lau and Weng 1995). So, the results of such decomposition depend on the starting point of the series. This transform is also limited to the series of a 2^j length, where $j = 1, 2, 3 \dots$. This problem is solved by means of a highly redundant non-orthogonal transform called maximal overlap discrete wavelet transform (MODWT) also known as non-decimated DWT. For this transform, an input time series of any length N results in the same number of wavelet and scaling coefficients at each resolution level. Therefore, the features of wavelet coefficients in the wavelet multiresolution analysis (MRA) are aligned with the original time series (Percival and Walden 2000).

In this study, a MODWT-based MRA was used to preprocess ET series and produce the approximately independent components. The latter were used for ET modeling. For MRA decomposition based on MODWT the wavelet filter selection is not critical (see Percival and Walden 2000). Therefore, the filter that produces least artifacts

at the beginning and the end of the series was used, i.e. Haar filter. Let $ET(t)$ denote the daily evapotranspiration computed using the PM equation. We can write it as:

$$ET(t) = \sum_{j=1}^J D_j(t) + S_J(t) \quad (3.2)$$

where, $D_j(t) = D_j(1), \dots, D_j(N)$ are vectors of length N , also known as j^{th} level detail, and $S_J(t) = S_J(1), \dots, S_J(N)$ is the smooth (or approximation) at level J . The decomposition in Eq. (3.2) is known as wavelet MRA. The details, $D_j(t)$, correspond to the component at time of approximately 2^j days, and capture a part of the record that corresponds to the frequencies in the range from $2^{-(j+1)}$ to 2^{-j} cycles per day. This range corresponds to physical scales between 2^j and 2^{j+1} days. The smooth, $S_J(t)$, captures the low frequency variations of the time series that correspond to time of approximately 2^J days, or to averages over intervals of 2^{J+1} days. For further details see Percival and Walden (2000).

3.2.3 Multivariate Relevance Vector Machine

A multivariate relevance vector machine (MVRVM) model was used in this study to model and forecast ET. The MVRVM is an extension of the Relevance Vector Machine (RVM) developed by Thayananthan et al. (2008) for multiple outputs of the machine learning model, which in this case are the ET forecasts for multiple days ahead. Tipping (2001) introduced the Relevance Vector Machine (RVM), as a general sparse Bayesian modeling approach for classification and regression. In RVM regression model, the weight of each input is governed by a set of hyperparameters that describe the posterior distribution of these weights. They are estimated iteratively during the machine learning training step. The value of most of the hyperparameters approaches infinity, and the corresponding weights become zero. The remaining non-zero weights are called the

relevance vectors. RVMs have good generalization performance and produce a sparsity representation of the nonlinear processes involved (Thayananthan et al. 2008).

For developing a MVRVM, a training data set, as input-target vector pairs $\{x_n, y_n\}_{n=1}^N$ is needed, where N is the number of observations, $x \in R^B$ is an input vector, and $y \in R^M$ is the multiple output vector. The model "learns" the dependence between input and output target with the purpose of making accurate predictions of the target vector y for previously unseen values of x :

$$y = W\Phi(x) + \varepsilon \quad (3.3)$$

Here, $\Phi(x) = [1, f(x, x_1), \dots, f(x, x_N)]^T$ is a set of $N + 1$ vectors of basis functions f , W is the $M \times P$ matrix of weights of these basis functions ($P = N + 1$), and ε is the noise vector assumed to be Gaussian with zero-mean and diagonal covariance matrix $D = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$. The kernel basis functions, f , considered in this paper were the Gaussian kernel function ($f(x, x_n) = \exp(-r^{-2}\|x - x_n\|^2)$), Laplace kernel function ($f(x, x_n) = \exp(-(r^{-2}\|x - x_n\|^2)^{1/2})$) and Cauchy kernel $f(x, x_n) = 1/(1 + r^{-2}\|x - x_n\|^2)$, where r is the kernel width parameter. These types of kernels have been used by many authors in hydrology applications (Kaheil et al. 2008; Ticlavilca et al. 2011; Torres-Rua et al. 2011).

Let $y = [\tau_1, \dots, \tau_m, \dots, \tau_M]^T$ and $W = [w_1, \dots, w_m, \dots, w_M]^T$. The multivariate Gaussian likelihood distribution for the target vector y can be written as:

$$P(\{y_n\}_{n=1}^N | W, D) = \prod_{n=1}^N N(y_n | W\Phi(x_n), D) = \prod_{m=1}^M N(\tau_m | w_m \Phi, \sigma_m^2) \quad (3.4)$$

where, $\Phi = [1, \Phi(x_1), \dots, \Phi(x_n)]$ is the design matrix.

In order to avoid over-fitting in the maximum likelihood estimation of W and σ^2 , Tipping (2001) proposed adding a Gaussian prior term for the weights of each basis function. The prior distribution over the weights is shown in Eq. (3.5):

$$p(W | A) = \prod_{m=1}^M \prod_{k=1}^P \mathcal{N}(w_{mk} | 0, \alpha_k^{-2}) = \prod_{m=1}^M \mathcal{N}(w_m | 0, A) \quad (3.5)$$

where, $A = \text{diag}(\alpha_1^{-2}, \dots, \alpha_P^{-2})$ with each α_k being an independent hyperparameter that determines the relevance of the associated basis function. This provides the sparsity of the model (Tipping and Faul 2003). w_{mk} is the element in m^{th} row and k^{th} column of the weight matrix W .

The posterior distribution of the model parameters is then given by the combination of the likelihood and prior distributions within Bayes' rule:

$$\begin{aligned} p(W | \{y\}_{n=1}^N, D, A) &\propto p(\{y\}_{n=1}^N | W, D) p(W | A) \\ &\propto \prod_{m=1}^M \mathcal{N}(w_m | \mu_m, \Sigma_m) \end{aligned} \quad (3.6)$$

The posterior distribution of the weights is Gaussian $\mathcal{N}(\mu_m, \Sigma_m)$ with mean $\mu_m = \sigma_m^{-2} \Sigma_m \Phi^T \tau_m$ and covariance $\Sigma_m = (A + \sigma_m^{-2} \Phi^T \Phi)^{-1}$.

Given the posterior of the weights, an optimal set of hyperparameters A^{opt} can be obtained by maximizing the data likelihood in Eq. (3.6) (Tipping and Faul 2003; Thayananthan et al. 2008). The data likelihood is marginalized as:

$$\begin{aligned} p(\{y_n\}_{n=1}^N | A, D) &= \int p(\{y_n\}_{n=1}^N | W, D) p(W | A) dW \\ &= \prod_{m=1}^M \int \mathcal{N}(\tau_m | w_m \Phi, \sigma_m^2) \mathcal{N}(w_m | 0, A) \\ &= \prod_{m=1}^M |C_m|^{-1/2} \exp\left(-\frac{1}{2} \tau_m^T C_m^{-1} \tau_m\right) \end{aligned} \quad (3.7)$$

where, $C_m = \sigma_m^2 I + \Phi A^{-1} \Phi^T$. And the optimal set of hyperparameters $\alpha^{opt} = \{\alpha_k^{opt}\}_{k=1}^P$ and noise $(\sigma^{opt})^2 = \{\sigma_m^{opt}\}_{m=1}^M$ are obtained using a bottom-up basis function selection approach described by Tipping and Faul (2003). During this optimization process, many elements of α go to infinity setting the corresponding posterior probability of the weight to zero. The few non-zero weights correspond to the so-called "relevance vectors" (RV) that are the sparse core of the RVM model (Tipping and Faul 2003). The optimal parameters are then used to obtain the optimal weight matrix with optimal covariance Σ^{opt} and mean μ^{opt} . The mathematical formulation, likelihood maximization, and optimization procedure of the RVM and MVRVM are discussed in detail in Tipping (2001), Tipping and Faul (2003), and Thayananthan et al. (2008).

Given a new input x^* , the predictive distribution for the corresponding target y^* can be computed as (Tipping 2001):

$$p(y^*/y, \alpha^{opt}, (\sigma^{opt})^2) = \int p(y^* | W, (\sigma^{opt})^2) p(W | y, \alpha^{opt}, (\sigma^{opt})^2) dW \quad (3.8)$$

$$\Rightarrow p(y^*/y, \alpha^{opt}, (\sigma^{opt})^2) = N(y^* | \mu^*, (\sigma^*)^2) \quad (3.9)$$

where, $\mu^* = [\mu_1^*, \dots, \mu_M^*]^T$ is the predictive mean, and $(\sigma^*)^2 = [(\sigma_1^*)^2, \dots, (\sigma_M^*)^2]^T$ is the predictive variance. This predictive variance is the sum of variances of two terms: the noise in the data and the uncertainty in the prediction of the weight parameters (Tipping 2004). The standard deviation σ^* of the predictive distribution was used to estimate the 95% Bayesian confidence interval as: $\hat{y} = \mu^* \pm 1.96\sigma^*$.

3.2.4 Performance Estimation Criteria

In this paper, the accuracy of the models has been estimated and model robustness is evaluated using a bootstrap approach. The accuracy was estimated using the root mean square error, $RMSE$, the coefficient of determination, R^2 , and the Nash-Sutcliffe efficiency coefficient, E . The $RMSE$ was calculated as shown in Eq. (3.10):

$$RMSE = \left[\sqrt{\frac{\sum_{t=1}^N (\hat{y}_t - y_t)^2}{N}} \right] \quad (3.10)$$

The coefficient of determination, R^2 , is computed as:

$$R^2 = \frac{\left[\sum_{t=1}^N (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}}) \right]^2}{\sum_{t=1}^N (y_t - \bar{y})^2 \sum_{t=1}^N (\hat{y}_t - \bar{\hat{y}})^2} \quad (3.11)$$

where, \hat{y}_t = predicted ET for day t (mm d⁻¹); y_t = calculated using PM equation for day t (mm d⁻¹); \bar{y} = mean of the observed ET; $\bar{\hat{y}}$ = mean of the estimated ET; and N = total number of observations.

The Nash-Sutcliffe efficiency, E , is a normalized statistic that determines a relative magnitude of the residual variance, “noise,” compared to the measured data variance, “information” (Nash and Sutcliffe 1970). It is recommended by the ASCE (1993) and Legates and McCabe (1999) as a measure of model performance. It indicates how well the plot of observed versus simulated data fits the 1:1 line and was computed as shown in Eq. (3.12):

$$E = 1 - \frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{\sum_{t=1}^N (y_t - \bar{y})^2} \quad (3.12)$$

It must be noted that an efficiency value of 1 ($E = I$) corresponds to a perfect match between modeled ET and observed data. An efficiency of 0 ($E = 0$) indicates that the model predictions are as accurate as the mean of the observed data. An efficiency less than zero ($E < 0$) occurs when the observed mean is a better predictor than the model, which indicates unacceptable performance.

In order to select the forecasting model, the main goal was to select the adequate number of inputs, or days to the past, the kernel width and the kernel type. The optimal values of these parameters were selected by trial and error procedure to obtain the best RMSE and E values.

The bootstrap method (Efron and Tibshirani 1993) was used in this study to guarantee good generalization ability and robustness of the machine learning model. The bootstrap data set was created by randomly sampling with replacement from the training data set. In the bootstrap estimation, the selection process was repeated 1,000 times to yield 1,000 bootstrap training data sets, which were treated as independent sets. For each of the bootstrap sets, the MVRVM was retrained and its performance was evaluated by calculating E for the two years of unseen testing data set. It is important to note that the kernel type, kernel width, and numbers of days of past time series data used as input to the MVRVM stayed the same, but the hyperparameters changed with each bootstrap sample. The bootstrap method provides implicit information on the uncertainty of the estimator (E) evaluated in the model. A more detailed description of the results is detailed in Section 3.3.4.

3.3 Results and Discussion

A classical time series modeling approach, the seasonal autoregressive integrated moving average (SARIMA) model was first used to forecast ET. But the results were converging the forecast to the mean value of the ET time series. The Nash-Sutcliffe coefficient of efficiency values for SARIMA models were negative indicating that the observed mean is a better predictor than the model, which is considered as unacceptable performance in hydrological modeling (Legates and McCabe 1999). This is due to the “backward looking” nature of the model (Meyler et al. 1998). Using high resolution daily data was also another limitation for the SARIMA model, since the season is 365 days which did not allow the model to converge making it not applicable for real time modeling and forecasting. The results of the SARIMA models are not shown in this paper. In this results and discussion section we only present the results of the wavelet-based MRA and the MVRVM models.

3.3.1 Wavelet Decomposition Selection

The properties of the ET time series, both physical and statistical, were examined in time and frequency domains. Table 3.1 showed the power spectrum results of the maximum overlap discrete wavelet transform (MODWT). From this table, it was obvious that the major changes in ET were the seasonal and annual components (Levels d7 through s8), while the short-term variations had the lowest percentage of ET changes. Based on the power spectrum in the wavelet domain, a wavelet multiresolution analysis (MRA) was performed on the ET time series. Two wavelet decomposition designs were considered to aid ET forecasting. In Design 1, three levels of MRA decomposition were used to isolate the short-term changes of ET. This design focused on the short-term

frequencies (up to 16 days). The records were then decomposed into four components, where the details capture high frequency changes of ET:

$$ET_{(1)}(t) = D_1(t) + D_2(t) + D_3(t) + S_3(t) \quad (3.13)$$

where, the MRA decomposition was performed for $J = 3$ levels, using MODWT, Haar filter, and the reflection boundary rule (see Fig. 3.1). In this model, the highest level of the detail captured the ET changes on scale of 8 to 16 days.

The second design of the wavelet-MRA analysis was based on the basic energy balance equation that defines ET:

$$ET(t) = R_n(t) + H(t) + G(t) \quad (3.14)$$

where, $G(t)$ = daily heat transfer from/to the ground; $H(t)$ = daily heat transfer to the atmosphere; and $R_n(t)$ = net daily solar radiation. The energy balance concept was used to

Table 3.1. Wavelet MODWT-Based Power Distribution by Level

Level	ET %	Physical scale in days (months, years)
d ₁	4.70	2 - 4
d ₂	4.41	4 - 8
d ₃	3.28	8 - 16
d ₄	2.44	16 - 32
d ₅	2.92	32 - 64
d ₆	6.89	64 - 128
d ₇	21.06	128 - 256 (4.3 m - 8.5 m)
d ₈	42.38	256 - 512 (8.5 m - 17.06 m)
s ₈	11.84	256 - 512 (8.5 m - 17.06 m)

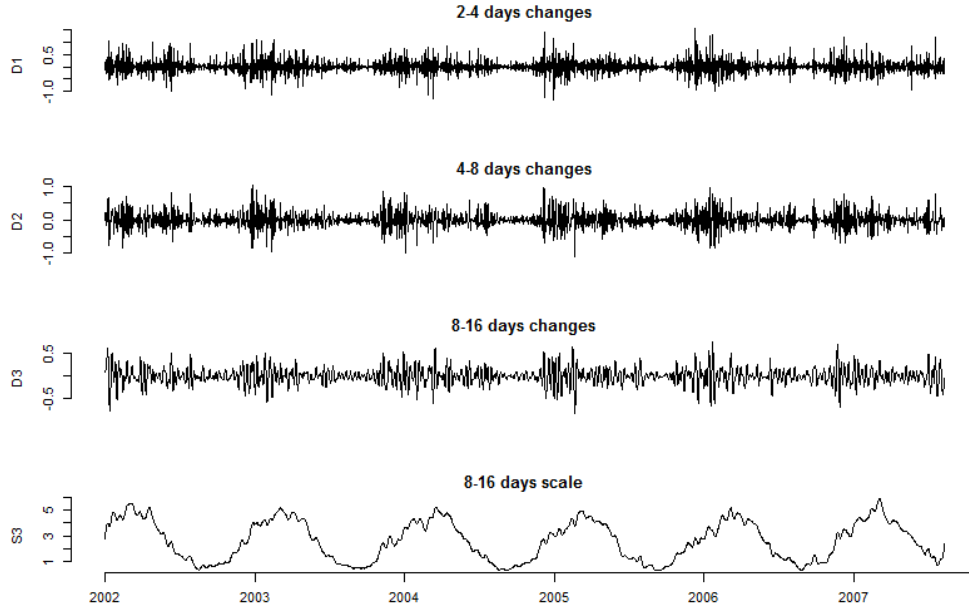


Fig. 3.1. Design 1 of Wavelet-MRA decomposition

extract meaningful components for building the forecasting model. Here, $G(t)$ corresponds to a long-term component that capture the variations of the low frequency scales, $H(t)$ contains the short-term variations, which capture the high frequency changes in ET, and $R_n(t)$ corresponds to the mid-frequency variations. Based on this, Design 2 used a wavelet-based MRA performed for $J = 8$ levels to breaks down the ET time series into three meaningful components, and was defined as:

$$ET_{(2)}(t) = D(t) + S(t) + A(t) \quad (3.15)$$

where, $D = D_1 + D_2 + D_3$, $S = D_4 + D_5 + D_6 + D_7$, $A = D_8 + S_8$. Here, D , S , and A are referred to as the daily (short-term variations), seasonal, and annual (long-term variations) components as shown in Fig. 3.2. The daily component, D , captured ET variability from two to 16 days, which accounted for 12.39% of total variability. The seasonal component, S , captured changes in ET on the scale from 16 days to 8.5 months, which amounted about 33.31% of total variability. The annual component, A , captured

the changes of ET on a scale of 8.5 to 17.06 months which contributed 54.22% of the total variability.

3.3.2 Forecasting Model Selection

Two hybrid models, each consisting of a multivariate relevance vector machine (MVRVM) and a wavelet model developed from one of the wavelet decomposition designs were developed. Five different forecasting models were evaluated (the inputs for each model are specified below):

- Model 1 (M1): ET time series were used as an input for the MVRVM, which served as control a model for comparison,
- Model 2 (M2): Each wavelet decomposition component of Design 1 defined in Eq. (3.13) was used separately as input. A MVRVM was built for each decomposition, the outputs were then added to get an estimated ET,
- Model 3 (M3): All the wavelet decompositions of Design 1 were used simultaneously as a multivariate input for the MVRVM to model and forecast ET,
- Model 4 (M4): A separate MVRVM model was built for each wavelet MRA component of Design 2 defined in Eq. (3.15). The resulting outputs were then added to get forecasted ET,
- Model 5 (M5): All the wavelet components of Design 2 were used at once as a multivariate input for the MVRVM to get ET as output.

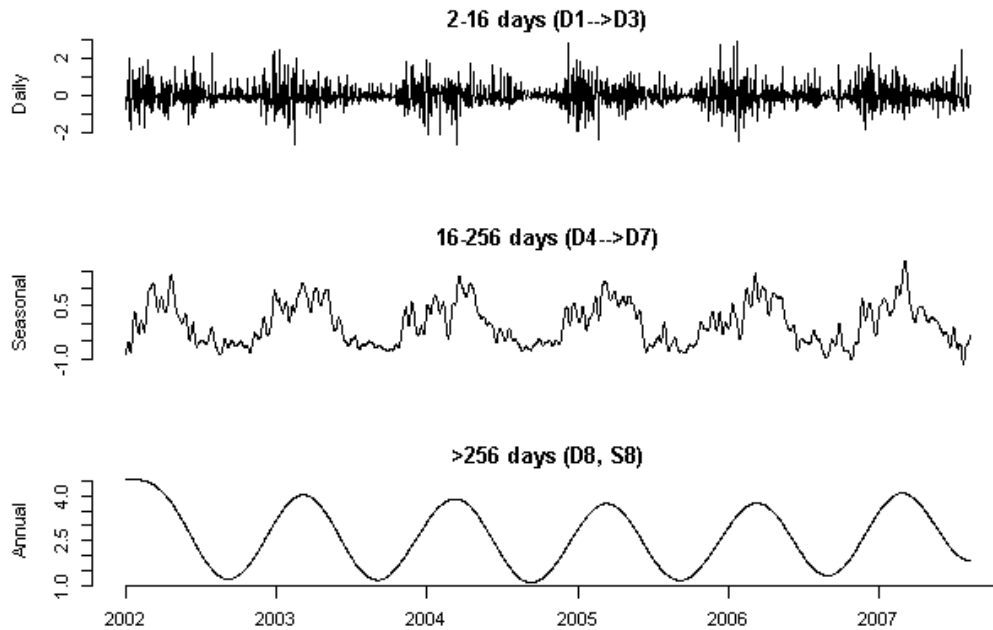


Fig. 3.2. Design 2 of Wavelet MRA decomposition

For the MVRVM application, 7 years of daily ET were used for training the machine. In this phase, several combinations of kernel type, kernel width, and number of inputs (or days in the past of the time series) were analyzed to get the optimal MVRVM parameters. The performance of the model, E and $RMSE$, for the different combinations was evaluated based on a 2-year calibration phase. The model that gave the highest E and lowest $RMSE$ for this calibration phase was selected, and then applied to 2-year unseen testing data.

The optimal kernel width, kernel type, and number of inputs (days to the past) of each model were presented in Table 3.2. This table also showed the performance (E , R^2 , and $RMSE$) of each model for the 2-year unseen testing data set. These statistics were the average values for the forecasted 16 days of ET for this testing phase. From the statistics presented in Table 3.2, it is obvious that all wavelet-MVRVM hybrid models outperformed the MVRVM model M1. M2 showed the best performance measured by E ,

R^2 and $RMSE$ ($E=0.604$, $R^2=0.619$ and $RMSE=0.766$ mm d⁻¹). The second best model was M5 with $E=0.602$, $R^2=0.616$ and $RMSE=0.767$ mm d⁻¹ with no significant difference from M2. However, it should be noted that M5 required only 9 days of the past records in order to forecast ET, while this number increased to 56 for M2. Requiring fewer number of inputs is an advantage of M5 over M2. Hence, less computational time is needed which is beneficial for real-time applications of the model, as well as when dealing with incomplete data. The overall performance of these hybrid models showed their ability to forecast daily ET for 16 days simultaneously. This is crucial for reliable irrigation systems management.

Next, a detailed analysis of the best two models, M2 and M5, was presented.

Table 3.3 included the statistics for M2 and M5 for all the 16 forecasted days. It indicated that in some cases, e.g. M2, E was higher when forecasting 16 days ahead (0.644) as compared to forecasting 7 days ahead (0.581). One would expect the efficiency to decline as we increased the forecasting horizon. This was not the case here due to the selection of the hyperparameters by the model that minimizes the average of E over all the outputs. Another reason is that the MVRVM forecasting is not iterative as it is in the time series regression models. The prediction errors do not accumulate.

Table 3.2. Models Inputs and Average Statistics of the Testing Data Set for the 16 Days of Forecasted ET

Model	Input variables	Maximum days to past	Kernel type	Kernel width	Statistic		
					E	R^2	$RMSE$ (mm d ⁻¹)
M1	ET	50	Gauss	10	0.561	0.580	0.805
M2	D ₁ , D ₂ , D ₃ , S ₃	56	Laplace/ Cauchy	14	0.604	0.619	0.766
M3	All Ds together	60	Cauchy	20	0.595	0.604	0.774
M4	D, S, A	70	Gauss/ Laplace	34	0.586	0.608	0.872
M5	All DSA together	9	Laplace	17	0.602	0.616	0.767

Figs. 3.3 and 3.4 showed the plots of the results of the best two hybrid models, M2 and M5, respectively. Four days (1, 6, 11, and 16 days ahead) were selected to display the results. The left panels in these figures provided a graph of the observed ET time series calculated by PM equation and predicted ET by the hybrid models, and the 95% confidence bounds for the two years of testing data set (2011–2012). The shaded areas in the left panels of these figures represented a 95% confidence interval of the forecasting models. M2 gave wider confidence interval bounds, hence the error over the prediction in M5 was smaller. This was due to the design of the wavelet MRA decompositions. M2 was decomposed using Design 1 (described in Section 3.3.1) which focused on the short- time variations. It focused on detailed modeling of the noisiest

Table 3.3. Statistics of the Selected Models (M1, M2, and M5) for All the Forecasted Days

Days ahead	Model								
	M1			M2			M5		
	E	R ²	RMSE (mm)	E	R ²	RMSE (mm)	E	R ²	RMSE (mm)
1	0.595	0.604	0.750	0.651	0.615	0.738	0.652	0.654	0.697
2	0.542	0.553	0.798	0.592	0.599	0.759	0.594	0.598	0.752
3	0.527	0.542	0.814	0.582	0.599	0.767	0.602	0.608	0.748
4	0.542	0.554	0.807	0.578	0.603	0.772	0.601	0.610	0.753
5	0.552	0.565	0.804	0.584	0.615	0.771	0.601	0.609	0.758
6	0.548	0.561	0.811	0.581	0.610	0.779	0.593	0.603	0.769
7	0.546	0.561	0.815	0.581	0.606	0.782	0.584	0.596	0.780
8	0.536	0.557	0.822	0.582	0.606	0.777	0.581	0.595	0.785
9	0.537	0.563	0.818	0.588	0.610	0.767	0.580	0.597	0.784
10	0.554	0.578	0.810	0.596	0.615	0.767	0.592	0.609	0.774
11	0.554	0.582	0.815	0.603	0.619	0.767	0.591	0.610	0.780
12	0.558	0.587	0.819	0.610	0.624	0.768	0.593	0.612	0.785
13	0.577	0.602	0.809	0.625	0.636	0.762	0.609	0.627	0.777
14	0.594	0.615	0.799	0.635	0.644	0.760	0.616	0.635	0.776
15	0.605	0.625	0.793	0.639	0.648	0.759	0.620	0.640	0.777
16	0.613	0.633	0.791	0.644	0.653	0.760	0.625	0.646	0.777
Average	0.561	0.580	0.805	0.604	0.619	0.766	0.602	0.616	0.767

12.39% of ET variability. The rest 87.61% of ET changes were modeled all together. On the other hand, M5 was decomposed using Design 2 in which the annual and seasonal components (D and S) were better forecasted due to the cyclic trends they present. M5 seemed to capture the variation in the ET time series better than M2. In the right panels of these figures, predicted ET was plotted against observed ET. These plots showed that the models underestimated ET values when observed ET was greater than about 6 mm d⁻¹. The underestimation was more pronounced when forecasting for more than one day ahead. This tendency has also been reported by Cobaner (2013) who used wavelet regression techniques for estimating ET. Since the MVRVM models used are regression models, they tend to underestimate larger values and overestimate low values. This was clear in those graphs, as the model was trying to regress towards the mean.

Another comparison between the best selected hybrid models (M2 and M5) was done to compare their forecasts for the two years of unseen data with the historical average of the ET time series in the area of study. The historical average was calculated for nine years of the available daily data, and had about 28% error rate compared to the observed data, with less than 10% error rate for the wavelet hybrid models M2 and M5 (Table 3.4). Which demonstrated the advantage of those hybrid models over traditional application for forecasting daily ET.

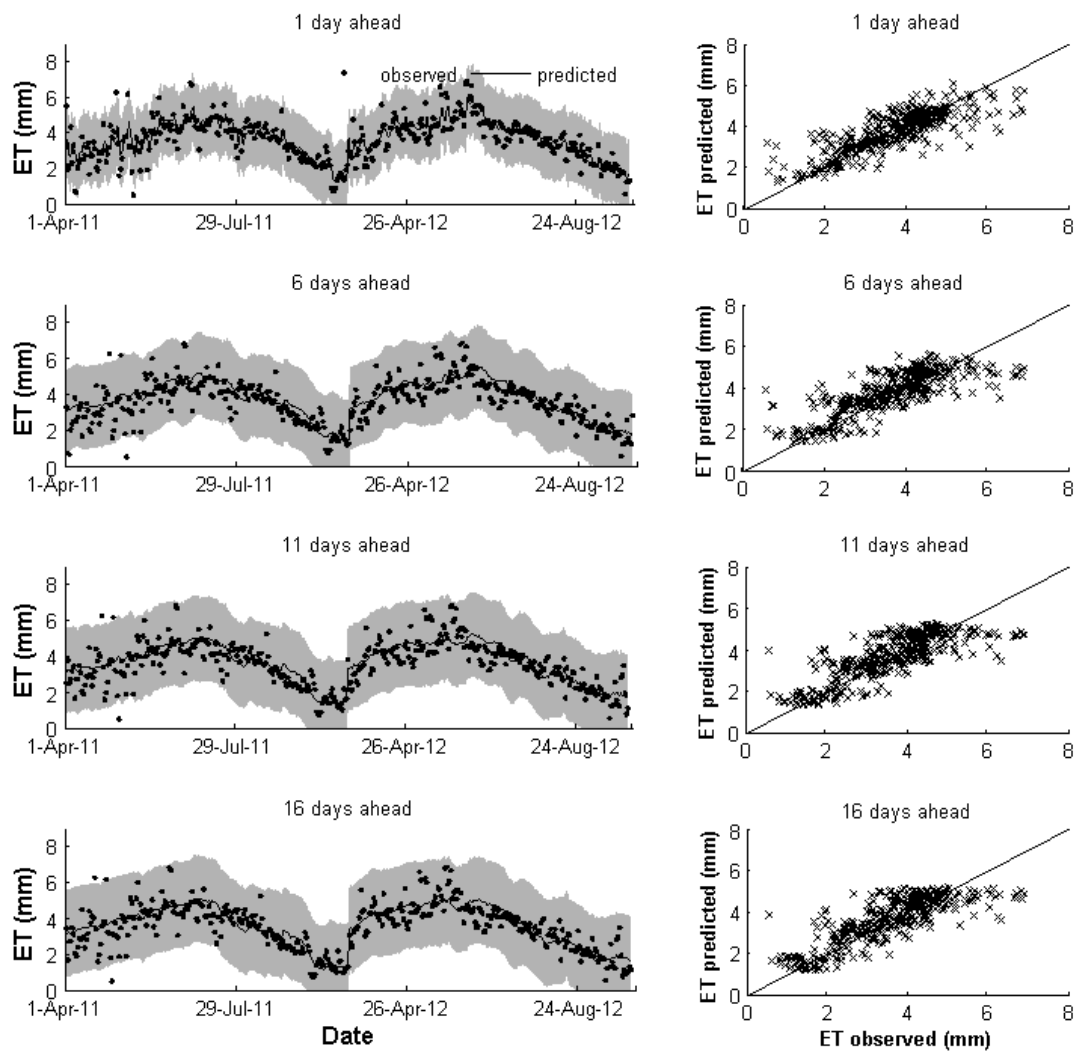


Fig. 3.3. Left: Predicted ET vs. observed ET time series for two-years of unseen testing data for Model 2 for selected days. Right: Plots of predicted ET vs. observed ET for the same time period

Table 3.4. Statistical Results of M2, M5, and 9-Years Average as Compared to the Observed Data for the Two Years of Test Dataset

Model	Coefficient of efficiency, CE	RMSE (mm/day)	Error rate %
M2	0.604	0.766	9.81
M5	0.602	0.767	9.83
9-yrs Avg	0.408	1.362	28.21

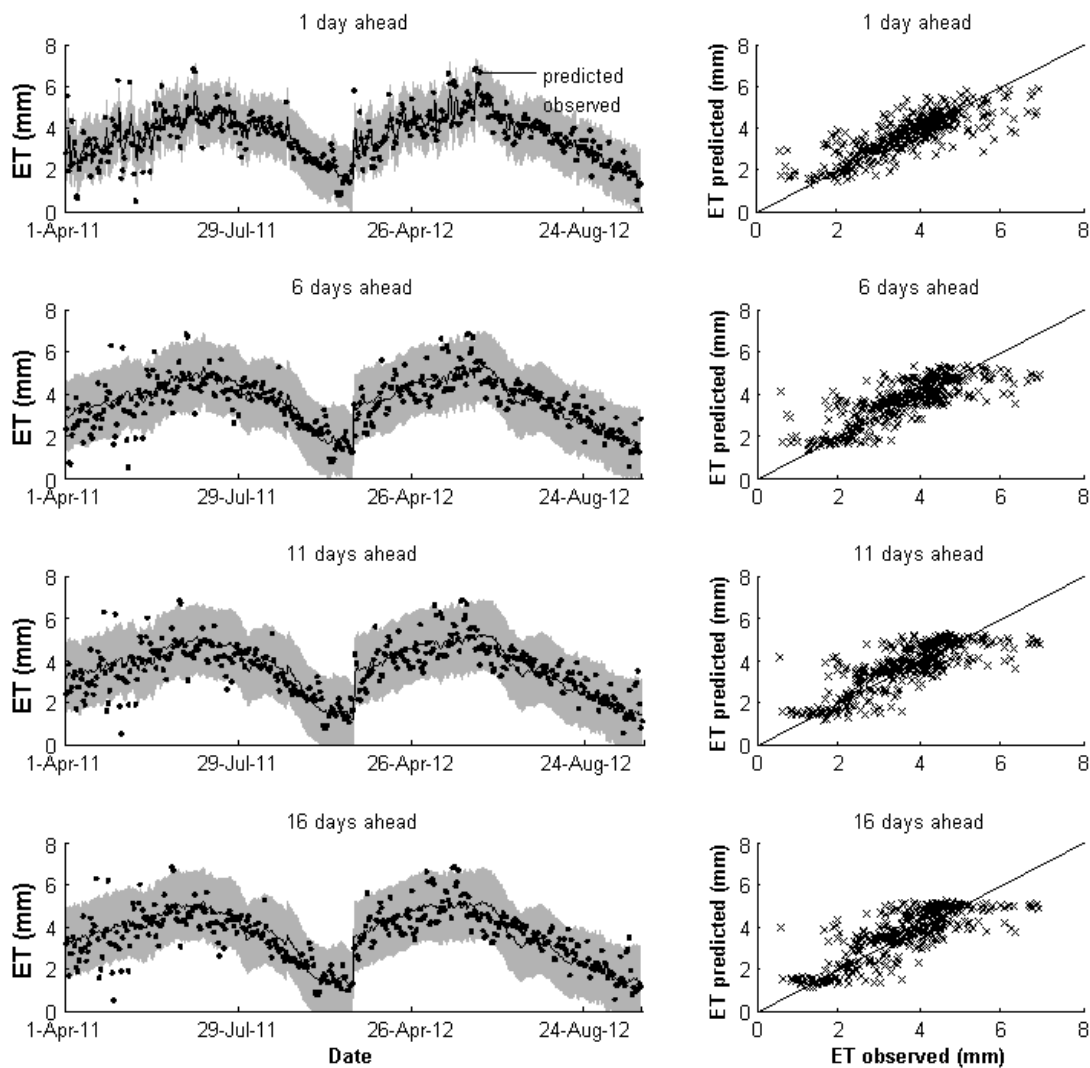


Fig. 3.4. Left: Predicted ET vs. observed ET time series for two-years of unseen testing data for Model 5 for selected days. Right: Plots of predicted ET vs. observed ET for the same time period

3.3.4 Model Robustness

Bootstrapping was performed to check for over-fitting and model generalization capability of the best two wavelet-MVRVM hybrid models, M2 and M5, and for M1 for comparison. Fig. 3.5 showed a boxplot of the results for 1000 bootstraps samples for which the Nash-Sutcliffe coefficient of efficiency, E , was computed (see Section 3.2.4 for details). No assumptions were made about the distribution of the data. Repeated samples

were drawn from the population with replacement. Since this type of samples are good approximation of the population, the bootstrap method provides a good approximation of the sampling distribution of the statistic E . For displaying purposes, only the results for four days (1, 6, 11, and 16 days ahead) were presented. On each box, the central mark was the median, the edges of the box were the 25th and 75th percentiles, the whiskers extended to the most extreme data points not considered outliers, and outliers were plotted individually. M5 showed a higher E for all the selected days, with no significant difference from M1 and M2. From these boxplots, it was noticed that the wavelet-MVRVM hybrid models M2 and M5 did not lose the robustness while improving the accuracy of the forecasts. The boxplots confirmed that the models were robust and can be used as ET forecasting models for real-time applications. The real-time application of ET forecasting models allows farmers to estimate the water demand for their fields and place the water orders. It will also allow the canal and reservoir operators to release the required amount of water needed for the command area.

3.3.5 Forecasting ET Using 10-days of Forecasted Temperature

In an attempt to improve the forecasting potential of the ET models, the possibility that additional variables serve as additional model inputs was considered. The National Oceanic and Atmospheric Administration (NOAA) provides daily forecast of minimum and maximum temperatures (T_{\min} and T_{\max}) up to 10 days ahead. In order to determine whether or not adding these data to the hybrid models that performed best (i.e. M2 and M5) would significantly improve the forecasts, the M2 and M5 analyses were repeated using additional temperature inputs .

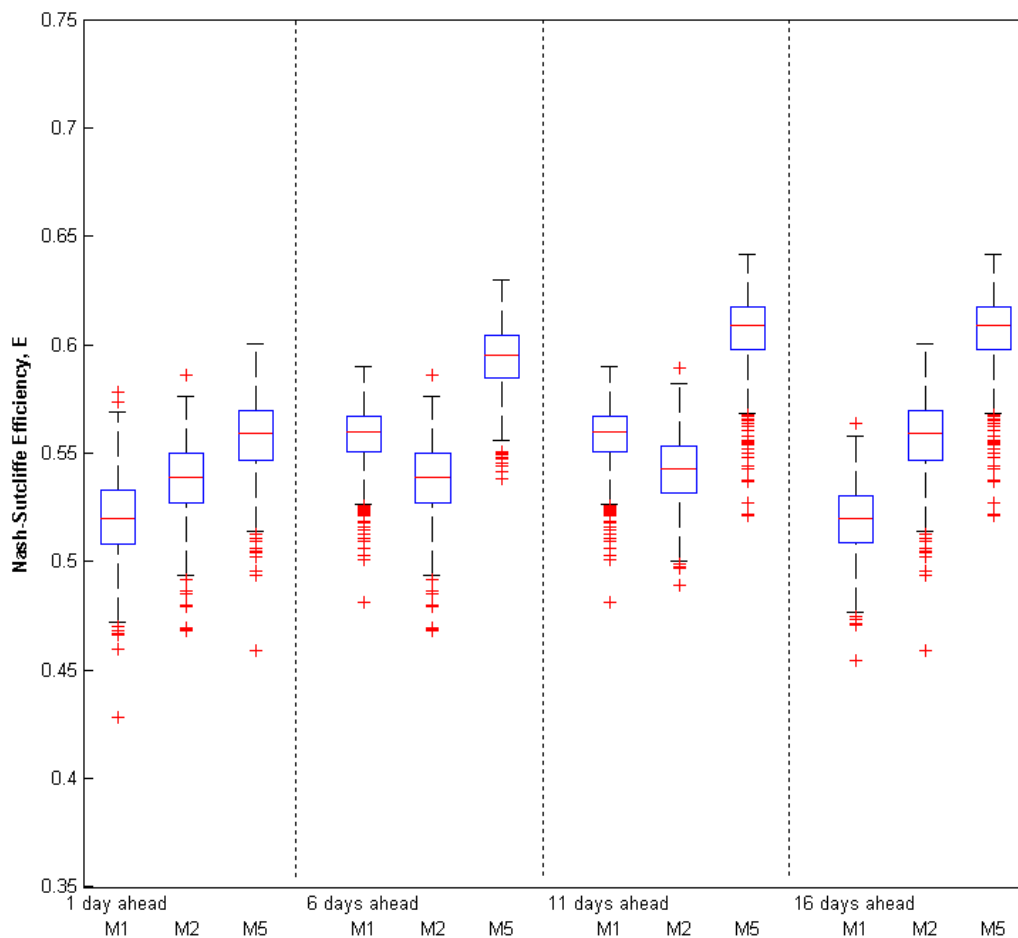


Fig. 3.5. Boxplots for the results of bootstrapping analysis (1000 times) for models M1, M2 and M5 for selected days

Table 3.5 showed the results for the modified models M1*, M2*, and M5*. It is clear from this table that adding the forecasted T_{\min} and T_{\max} improved the performance of the models, especially for forecasting the first 10 days for which the predicted temperature data were available. On average, the modified wavelet-MVRVM hybrid models M2* and M5*, respectively, outperformed again the MVRVM (M1*) model by 17.4% and 10.6% for E . Fig. 3.6 showed the scatter plots of the three modified models: M1*, M2*, and M5*. M2* showed improvement with underestimation of high ET values. The MVRVM uses Gaussian distribution which is not designed for capturing extremes.

Hence, the tendency to underestimate higher values of ET and overestimate lower values.

The introduction of wavelet decomposition helps overcoming this problem. M2* also

showed that the forecasts were more clustered around the 1:1 line. This highlights the

possible improvement of such hybrid models by adding information about the forecasted

future temperatures.

Table 3.5. Statistics of the Models for the 16 Days with the Additional 10-Days Forecasted T_{\min} and T_{\max} as Input

Days ahead	Model								
	M1*			M2*			M5*		
	E	R ²	RMSE (mm)	E	R ²	RMSE (mm)	E	R ²	RMSE (mm)
1	0.694	0.703	0.652	0.729	0.771	0.759	0.752	0.772	0.588
2	0.658	0.666	0.690	0.722	0.737	0.716	0.722	0.752	0.622
3	0.630	0.644	0.721	0.729	0.732	0.705	0.711	0.751	0.637
4	0.615	0.637	0.740	0.731	0.733	0.720	0.717	0.756	0.634
5	0.613	0.636	0.747	0.728	0.733	0.720	0.715	0.757	0.641
6	0.606	0.619	0.757	0.727	0.736	0.725	0.705	0.751	0.654
7	0.594	0.606	0.771	0.726	0.739	0.729	0.683	0.738	0.681
8	0.589	0.606	0.774	0.722	0.739	0.731	0.683	0.745	0.683
9	0.581	0.603	0.779	0.724	0.740	0.717	0.653	0.724	0.712
10	0.611	0.628	0.756	0.715	0.731	0.697	0.643	0.706	0.724
11	0.580	0.587	0.791	0.685	0.700	0.718	0.619	0.672	0.753
12	0.527	0.533	0.848	0.640	0.655	0.745	0.574	0.629	0.803
13	0.537	0.541	0.846	0.628	0.641	0.756	0.568	0.628	0.817
14	0.543	0.548	0.847	0.632	0.644	0.757	0.572	0.635	0.819
15	0.547	0.553	0.848	0.643	0.655	0.751	0.583	0.643	0.814
16	0.547	0.552	0.856	0.642	0.656	0.759	0.577	0.638	0.825
Average	0.592	0.604	0.776	0.695	0.709	0.732	0.655	0.706	0.713

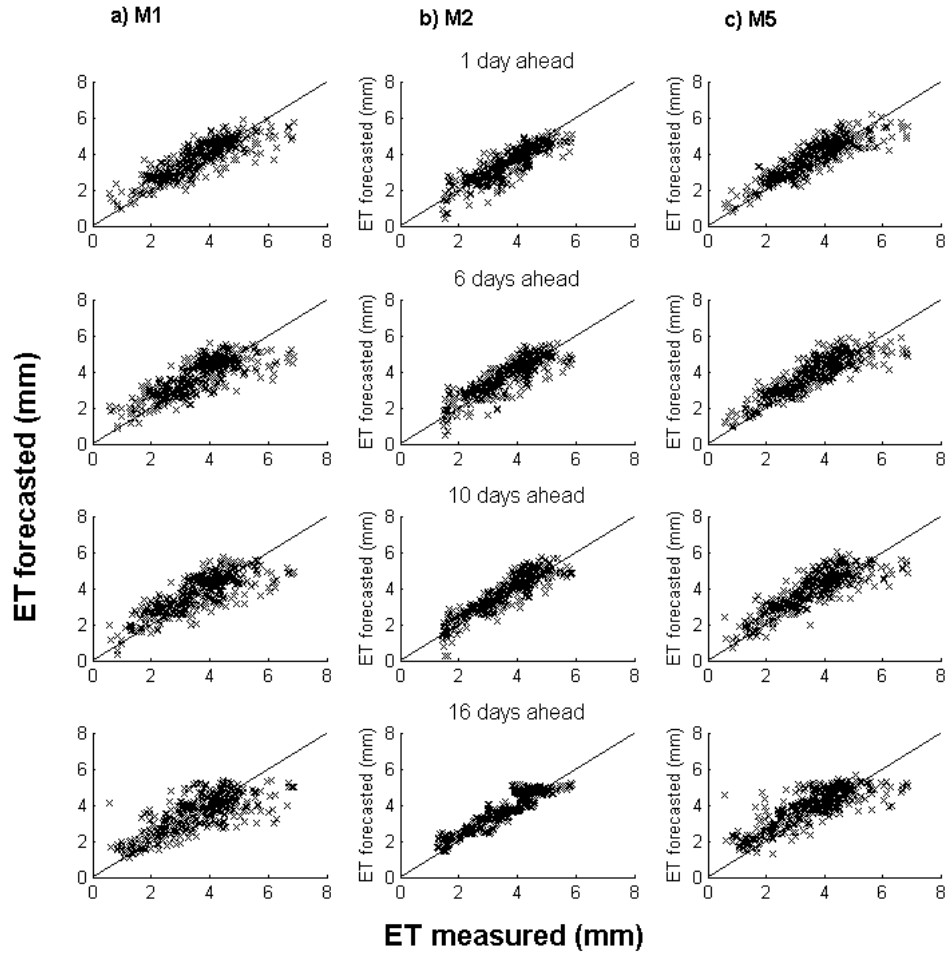


Fig. 3.6. Scatter plots of forecasted ET using modified models: M1* (a), M2* (b) and M5*(c) compared to measured ET

3.4 Conclusions and Future Work

In practice, predicting daily ET is difficult because it is characterized by high non-linearity and non-stationarity. Therefore, models that use components at different temporal resolutions provide a new alternative to ET forecasting problems. The potential of wavelet-MVRVM hybrid modeling for forecasting daily ET up to 16 days in advance was investigated in this paper. The study introduced the methodology of decomposing an ET time series using wavelet multiresolution analysis methods, and combining the wavelet decompositions with a MVRVM to develop an ET forecasting model. The

resulting models predicted daily ET up to 16 days ahead with good accuracy. Model that combined a wavelet daily, seasonal and annual multiresolution analysis components with MVRVM, performed the best in this study. It is recommended as a forecasting model for ET in the study area. The results showed that the wavelet-MVRVM hybrid models performed better than the MVRVM. The bootstrapping analysis showed robustness in forecasting of the wavelet-MVRVM hybrid models.

Inclusion of 10-days of forecasted minimum and maximum air temperatures as additional inputs to the models improved the performance for the first 10 days for which the weather forecast was available. Further research might be of interest to explore the potential of adding other weather variables, and their forecasts to the input vector.

The methodology presented in this study lays the grounds for further investigations and studies that could lead to forecast ET at the spatial level using remote sensing algorithms and the Landsat imagery.

References

- Ahmad, S., Popoola, A., and Ahmad, K. (2005). "Wavelet-based multiresolution forecasting." *UniS Technical Report*, June 2005.
- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). "Crop evapotranspiration: Guidelines for computing crop water requirements." *FAO Irrigation and Drainage Paper No. 56*, FAO, Rome, 300.
- Allen, R. G., Pruitt, W. O., Wright, J. L., Howell, T. A., Ventura, F., Snyder, R., Itenfisu, D., Stedudo, P., Berengena, J., Yrisarry, J. B., Smith, M., Raes, D., Perrier, A., Alves, I., Walter, I., and Elliot, R. (2006). "A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO 56 Penman-Monteith method." *Agric. Water Manage.*, 81, 1-22.
- ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models of the Watershed Management, Irrigation, Drainage Division (ASCE). (1993). "Criteria for evaluation of watershed models." *J. Irr. and Drain.Eng.*, 119(3), 429-442.

- Cannas, B., Fanni, A., See, L., and Sias, G. (2006). "Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning." *Phy. Chem. Earth.*, 31(18), 1164-1171.
- CEMP- Community Environmental Monitoring Program website, <http://www.cemp.dri.edu>, accessed in December 2012.
- Chou, C. M., and Wang, R. Y. (2002). "On-line estimation of unit hydrographs using the wavelet- based LMS algorithm." *Hydrological sciences journal*, 47(5), 721-738.
- Cigizoglu, H. K. (2003). "Incorporation of ARMA models into flow forecasting by artificial neural networks." *Environmetrics*, 14 (4), 417-427.
- Cobaner, M. (2013). "Reference evapotranspiration based on Class A pan evaporation via wavelet regression technique." *Irrig. Sci.*, 31 (2), 119-134.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
- Efron, B., and Tibshirani, R. (1993). *An introduction of the bootstrap, monographs on statistics and applied probability*. Vol. 57. CRC Press LLC, Boca Raton, Florida.
- Gorantiwar, S. D., Meshram, D. T., and Mittal, H. K. (2011). "Seasonal ARIMA model for generation and forecasting evapotranspiration of Solapur district of Maharashtra." *J. Agrometeorology*, 13 (2), 119-122.
- Hernandez, S., Morales, L., and Sallis, P. (2011). "Estimation of reference evapotranspiration using limited climatic data and Bayesian model averaging." *ems*, pp. 59-63, 2011 UK Sim 5th European Symposium on Computer Modeling and Simulation.
- Kaheil, Y. H., Rosero, E., Gill, M. K., McKee, M., and Bastidas, L. A. (2008). "Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines." *Trans. on Geoscience and Remote Sensing, IEEE*, 46(9), 2692-2707.
- Kisi, O. (2007). "Evapotranspiration modelling from climatic data using a neural computing technique." *Hydrol. Process.*, 21, 1925–1934.
- Kisi, O. (2011). "Evapotranspiration modeling using a wavelet regression model." *Irrig. Sci.*, 29, 241-252.
- Kisi, O. (2012). "Least squares support vector machine for modeling daily reference evapotranspiration." *Irrig. Sci.*, 31 (4), 611-619.
- Küçük, M., Kahya, E., Cengiz, T. M., and Karaca, M. (2009). "North Atlantic Oscillation influences on Turkish lake levels." *Hydrol. Process.*, 23(6), 893-906.

- Labat, D. (2005). "Recent advances in wavelet analyses: part 1. A review of concepts." *J. Hydrology*, 314 (1–4), 275-288.
- Labat, D., Ronchail, J., and Guyot, J.L. (2005). "Recent advances in wavelet analyses: part 2-Amazon, Parana, Orinoco and Congo discharges time scale variability." *J. Hydrology*, 314(1–4), 289-311.
- Landeras, G., Ortiz-Barredo, A., and Lopez, J. J. (2009). "Forecasting weekly evapotranspiration with ARIMA and artificial neural network models." *J. Irrig. Drain Eng. ASCE*, 135(3), 323-334.
- Lau, K. M., and Weng, H. (1995). "Climate signal detection using wavelet transform: how to make a time series sing." *Bull. Am. Meteorol. Soc.*, 76, 2391-2402.
- Legates, D. R., and McCabe, G. J. (1999). "Evaluating the use of "goodness-of-fit" measures in hydrological and hydroclimatic model validation." *Water Resour. Res.*, 35(1), 233-241.
- Li, X., Ding, J., and Li, H. (1999). "Combining neural network models based on wavelet transform." *J. Hydraulic*, 2, 1-4.
- Mallat, S. G. (1989). "A theory for multi resolution signal decomposition: the wavelet representation". *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7), 674-693.
- Mariño, M. A., Tracy, J. C., and Taghavi, S. A. (1993). "Forecasting of reference crop evapotranspiration." *Agric. Water Manage.*, 24, 163-187.
- Meyler, A., Kenny, G., and Quinn, T. (1998). "Forecasting Irish inflation using ARIMA models." Technical Paper 3/RT/98. *Economic Analysis, Research, and Publication Department*. Central Bank of Ireland.
- Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models." *I. J. Hydrol.*, 10, 282-290.
- Pandey, P. K., Pandey, V., Singh, R., and Bhakar, S. R. (2009). "Stochastic modelling of actual Black Gram evapotranspiration." *J. Water Resour. Protection*, 1, 448-455.
- Partal, T. (2009). "Modelling evapotranspiration using discrete wavelet transform and neural networks." *Hydrol. Process.*, 23(25), 3545-3555.
- Partal, T., and Cigizoglu, H. K. (2008). "Estimation and forecasting of daily suspended sediment data using wavelet-neural networks." *J. Hydrology*, 358, 317-331.
- Percival, D. B., and Walden, A. T. (2000). *Wavelet methods for time series analysis*. Cambridge University Press, Cambridge.

- Thayananthan, A., Navaratnam, R., Stenger, B., Torr, P. H., and Cipolla, R. (2008). "Pose estimation and tracking using multivariate regression." *Pattern Recognition Letters*, 29(9), 1302-1310.
- Ticlavilca, A. M., and McKee, M. (2011). "Multivariate bayesian regression approach to forecast releases from a system of multiple reservoirs." *Water Resour. Manage.*, 25, 523-543.
- Ticlavilca, A. M., McKee, M., and Walker, W. R. (2011). "Real-time forecasting of short-term irrigation canal demands using a robust multivariate Bayesian learning model." *Irrig. Sci.*, 31(2), 151-167.
- Tipping, M. E. (2001). "Sparse Bayesian learning and the relevance vector machine." *J. Mach. Learn.*, 1, 211-244.
- Tipping, M. E. (2004). "Bayesian inference: an introduction to principles and practice in machine learning." In: Bousquet, O., von Luxburg, U., and Ratsch, G. (eds.). *Advanced lectures on machine learning*. Springer, Berlin, pp. 41-62.
- Tipping, M. E., and Faul, A. C. (2003). "Fast marginal likelihood maximisation for sparse Bayesian models." In: Bishop, C. M., and Frey, B. J. (eds). *Proceedings of the ninth international workshop on artificial intelligence and statistics*, Key West, Florida, pp. 596-608.
- Torrence, C., and Compo, G. P. (1998). "A practical guide to wavelet analysis". *Bull. Am. Meteorol. Soc.*, 79, 61-78.
- Torres, A. F., Walker, W. R., and McKee, M. (2011). "Forecasting daily potential evapotranspiration using machine learning and limited climatic data." *Agricultural Water Manage.*, 98(4), 553-562.
- Trajkovic, S. (1998). "Comparison of prediction models of reference crop evapotranspiration." *Facta Universitatis, Series: Architecture and Civil Engineering*, 1(5), 617-626.
- Trajkovic, S., Todorovic, B., and Stankovic, M. (2003). "Forecasting of reference evapotranspiration by artificial neural networks." *J. Irri. and Drain. Eng.*, 129(6), 454-457.
- Wang, W. G., and Luo, Y. F. (2007). "Wavelet network model for reference crop evapotranspiration forecasting." *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR '07. (2)*, 751-755.

CHAPTER 4
SPATIAL DISTRIBUTION OF EVAPOTRANSPIRATION USING
REMOTE SENSING AND RELEVANCE VECTOR MACHINE^{††}

Abstract

With the development of surface energy balance analyses, remote sensing has become a spatially explicit and quantitative methodology for understanding of evapotranspiration (ET) - a critical requirement for planning and managing water resources. An algorithm titled "Mapping Evapotranspiration at high Resolution with Internalized Calibration (METRIC)" was used in this study to calculate actual ET from Landsat 5 Thematic Mapper images at 30 m spatial resolution. The analysis was applied to the Canal B service area of Delta Canal Company in Central Utah using data from the growing seasons of 2009–2011. The resulting ET maps were used to evaluate the irrigation efficiency of the system. The efficiency ranged from 65% to 78% throughout the growing season. A machine learning algorithm, the relevance vector machine (RVM) was then used to model ET spatially. The RVM was trained with a set of inputs of vegetation indices, land use, soil texture and weather data in order to model the ET using METRIC results as output. The developed RVM model provided accurate estimation of spatial ET based on a Nash-Sutcliffe coefficient (E) of 0.84 and a root mean squared error ($RMSE$) of 0.5 mm d^{-1} . This model lays the ground for the estimation of ET at spatial scale for the days when a Landsat image is not available. And could be used for

^{††} Co-authored by Wynn R. Walker, Mac McKee, Andres M. Ticlavilca, and Inga Maslova

forecasting daily spatial ET, if the vegetation indices model inputs are extrapolated in time, and the reference ET forecasted accurately.

4.1 Introduction

The world population is expected to be near 8 billion by 2025. Some researchers estimate that, to meet future food demand, at least another 2,000 km³ of water will be needed. Irrigated lands are a main sources of food and energy for populations around the globe, providing about 40% of the world's food from around 17% of the cultivated area (Thenkabail et al. 2009). These lands will be shaped increasingly by the effects of competition for water from other sectors, notably urban and rural domestic water supply and industrial needs. In order to cope with these changes, better water management practices and improved estimation of irrigation water demand are crucial.

Evapotranspiration (ET) being a main component of the hydrologic cycle, it must be estimated with the greatest precision possible. Direct measurement of actual ET is difficult and at the best, mostly provides point values at the weather station location. A spatially explicit and quantitative understanding of ET is critical for planning and managing water resources. The estimation of ET rates over large irrigated areas can provide beneficial information to farmers and to the operators of irrigation supply/distribution systems.

Methods for quantifying ET based on meteorological point measurements are still unable to provide spatial ET at large scales, mostly because of the heterogeneity of the land surface and the dynamic nature of the heat transfer processes (Kaheil et al. 2008).

The first attempts to determine spatial distributions of ET on a regional scale relied on geostatistical and interpolation procedures using data available from

meteorological stations (Mauser and Schädlich 1998). The complex spatial heterogeneity of the meteorological factors could not be considered in these conventional techniques. Therefore, they may result in inaccurate estimations over large areas. An alternative ET estimation approach is to combine remote sensing observations along with meteorological data to provide the spatial distribution of ET.

Remote sensing is known as a very useful tool for estimating ET at various temporal and spatial scales. It offers substantial potential in bridging the gap between point and larger-scale measurement of ET. In recent years, this approach has been demonstrated for determining the spatial distribution of ET at the particular time the remote sensor passed over (Anderson et al. 2007; Bastiaanssen et al. 2008; Allen et al. 2011). Mapping ET at a large spatial scale has been proven to be very useful for irrigation management practices. For instance, it has been used to monitor water rights (Allen et al. 2005), manage agricultural water management (Gowda et al. 2008; Anderson et al. 2012), as well as evaluating irrigation and distribution systems efficiency (Folhes et al. 2009).

Application of remote sensing algorithms solving the energy balance using high resolution satellite imagery has proven useful for establishing estimates of actual ET for large populations of fields and water users (Bastiaanssen et al. 1998; Tasumi et al. 2005; Neale et al. 2012). The use of a surface energy balance to determine ET has strong advantages, mainly that the specific vegetation type does not need to be known. And the energy balance approach can detect reduced ET caused by water shortage, salinity, or frost that may not correlate with vegetation amount. In addition, thermally based energy balance methods can detect evaporation from bare soil.

One of the widely used algorithms to map ET using satellite imagery is METRIC (Mapping Evapotranspiration at high Resolution with Internalized Calibration). It was developed by Allen et al. (2003) as an extended modification of SEBAL (Surface Energy Balance Algorithm for Land). This later was developed by Bastiaanssen (1995), and is a widely accepted algorithm that has been applied over many regions of the world. METRIC algorithm was used in this study to retrieve ET images using Landsat imagery to achieve high ET product resolution (30 m) that is useful for monitoring water consumption at field scales. Spatial ET estimates at field level or, in other words, the crop water demand can provide information about the efficiency of the irrigation and conveyance systems.

Nonetheless, the estimation has historically been restricted to the day when the geospatial information was obtained, corresponding to the 16-day over-pass day of the Landsat. So, daily spatial ET has not been available due to temporal resolution of satellites and/or gaps in image acquisition due to cloud cover. And without information of precise future daily water crop demand there is a continuous challenge for the implementation of better water distribution and management policies in the irrigation system.

In the recent decade, new tools have become available to perform analyses that exploit the statistical characteristics of the data. These models are known as statistical data-driven tools or machine learning, which have been used to estimate relationship among complex multidimensional non-linear variables. These machine learning models use the statistical properties of inputs and outputs of the process being studied to define relationships among them.

The Relevance Vector Machine (RVM) is one of such machine learning models. The RVM is a Bayesian regression algorithm developed by Tipping and Faul (2003) and

has been used in hydrology and other fields. It is a sparse model with high prediction accuracy, and robustness. For the purposes of this study, an RVM has been used to develop a model for estimating ET at the spatial level. The RVM was trained with a set of inputs to predict the ET from the METRIC output. These inputs were chosen in a way that they can be estimated or forecasted on daily basis whenever a Landsat image is not available.

The objectives of this paper were: 1) to apply the METRIC algorithm to get the actual ET spatially for the Canal B irrigation command area of Delta Canal Company in Central Utah; 2) to evaluate the efficiency of the irrigation supply system in the command area; and 3) to develop an RVM model that provides a spatial distribution of ET that can be applied even on the days when Landsat is not passing over the area.

4.2 Models Background

4.2.1 METRIC Algorithm

The METRIC algorithm is a satellite image processing model that estimates spatially distributed values of actual evapotranspiration ET as the residual of the energy balance (Allen et al. 2007):

$$LE = R_n - G - H \quad (4.1)$$

where LE = heat flux density (W m^{-2}); R_n = incoming radiation flux density (W m^{-2}); G = soil heat flux density (W m^{-2}); and H = sensible heat flux density (W m^{-2}).

The net incoming radiation flux, R_n , is calculated by solving the radiation balance as described by Allen et al. (2007). Soil heat flux, G , is the rate of heat conducted into soil and vegetation, and is estimated in METRIC from H , R_n , the surface temperature, T_s , and the Normalized Difference Vegetation Index, $NDVI$.

Sensible heat flux, H , is the convective heat loss from the surface to the air created by a near surface temperature gradients. It is estimated in METRIC using a one-dimensional, blended aerodynamic, temperature gradient based method:

$$H = \frac{\rho c_p dT}{r_{ah}} \quad (4.2)$$

where, ρ = air density (kg m^{-3}); c_p = air specific heat ($\text{J kg}^{-1} \text{K}^{-1}$); dT = temperature difference between two heights (near the surface and a height about 2 m) in a near surface blended layer (K); and r_{ah} = aerodynamic resistance to heat transport (s m^{-1}) between the two heights. dT is a linear function of T_s developed by using temperature values for the cold and the hot pixels, which provides internal and automatic calibration. In METRIC, cold and hot pixels should be located within 50 Km of the weather station. The cold pixels should represent a well-watered and fully vegetated areas of the image, representing a maximum or near maximum evaporative flux; and the hot pixel should be located in a dry and bare agricultural field where the evaporative flux is almost 0.

A special feature of METRIC is that its auto-calibration is made using ground-based reference ET as an index of evaporative behavior of well-irrigated full cover agricultural pixels, providing agreement with traditional methods for ET. Therefore, the METRIC approach assumes that ET for the entire area of interest varies in proportion to change in reference ET at the weather station. The main advantage of METRIC is the need for minimum ground data.

A practical limitation of METRIC is that the endpoint conditions used to calculate dT in Eq. (4.2), are selected from each image. This generally requires subjective user intervention by trained modelers to select the appropriate endpoint pixels. In this paper,

the selection of these so-called "hot" and "cold" pixels was done following the recommendations of Allen et al. (2013).

4.2.2 Relevance Vector Machine

Tipping (2001) introduced the Relevance Vector Machine (RVM), as a general sparse Bayesian modeling approach for classification and regression. In RVM regression models, the weight of each input is governed by a set of hyperparameters that describe posterior distribution of the weights, and are estimated iteratively during the machine learning training step. Most of those hyperparameters approach infinity setting the corresponding weights to zero. The remaining non-zero weights are called the relevance vectors. RVM's have good generalization performance with sparsity in the representation.

For developing an RVM, a training data set, as input-target vector pairs $\{x_i, y_i\}_{i=1}^N$ is needed, where, N is the number of observations. The model learns the dependence between inputs and output target with the purpose of making accurate predictions of y for previously unseen values of x :

$$y = w\Phi(x) + \varepsilon \quad (4.3)$$

where, w is a vector of weight parameters, $\Phi(x) = [1, f(x, x_1), \dots, f(x, x_N)]^T$ is a design matrix of $N + 1$ vectors of basis functions f , ε is the error usually assumed to be zero-mean Gaussian with variance σ^2 . The kernel basis functions, f , considered in this paper were the Gaussian kernel function ($f(x, x_i) = \exp(-r^{-2}\|x - x_i\|^2)$), Laplace kernel function ($f(x, x_i) = \exp(-(r^{-2}\|x - x_i\|^2)^{1/2})$) and Cauchy kernel $f(x, x_i) = 1/(1 + r^{-2}\|x - x_i\|^2)$, where r is the kernel width parameter. These types of kernels have been used by many authors in hydrology applications (Kaheil et al. 2008; Ticlavilca et al. 2011; Torres et al. 2011).

Then, a Gaussian likelihood distribution for the target vector can be written as:

$$p(y | w, \sigma^2) = (2\pi)^{-N/2} \sigma^{-N} \exp \left\{ -\frac{\|y - w\Phi\|^2}{2\sigma^2} \right\} \quad (4.4)$$

In order to avoid over-fitting in the estimation of the maximum likelihood of w and σ^2 , Tipping (2001) proposed adding a "prior term" to constrain the selection of parameters by defining an explicit zero-mean Gaussian prior probability distribution over them as shown in Eq. (4.5):

$$p(w | \alpha) = (2\pi)^{-M/2} \prod_{m=1}^M \alpha_m^{1/2} \exp \left(-\frac{\alpha_m w_m^2}{2} \right) \quad (4.5)$$

where, M is the number of independent hyperparameters $\alpha = (\alpha_1, \dots, \alpha_M)^T$. Each α is associated independently with every weight to moderate the strength of the prior and provide the sparsity of the model (Tipping 2001). The posterior distribution of the model parameters, which is given by the combination of the likelihood and prior distributions within Bayes' rule:

$$p(w | y, \alpha, \sigma^2) = \frac{p(y | w, \sigma^2) p(w | \alpha)}{p(y | \alpha, \sigma^2)} \quad (4.6)$$

The posterior distribution is Gaussian $N(w | \mu, \Sigma)$ with covariance

$\Sigma = (A + \sigma^{-2} \Phi^T \Phi)^{-1}$ and mean $\mu = \sigma^{-2} \Sigma \Phi^T y$; where, A is defined as $diag(\alpha_1, \dots, \alpha_M)$.

The optimal set of hyperparameters α^{opt} is obtained using a bottom-up basis function selection approach described by Tipping and Faul (2003). The marginal likelihood is then given by its logarithm $L(\alpha)$:

$$L(\alpha) = \log [p(y | \alpha, \sigma^2)] = \log \int_{-\infty}^{\infty} p(y | w, \sigma^2) p(w | \alpha) dw$$

$$= -\frac{1}{2} [N \log 2\pi + \log |C| + t^T C^{-1} y] \quad (4.7)$$

where, $C = \sigma^2 I + \Phi A^{-1} \Phi^T$. The optimal parameters are used to obtain the optimal weight matrix with optimal covariance Σ^{opt} and mean μ^{opt} .

Given a new input x^* , the predictive distribution for the corresponding target y^* can be computed as (Tipping 2001) :

$$p(y^*/y, \alpha^{opt}, (\sigma^{opt})^2) = \int p(y^*/w, (\sigma^{opt})^2) p(w/y, \alpha^{opt}, (\sigma^{opt})^2) dw \quad (4.8)$$

$$\Rightarrow p(y^*/y, \alpha^{opt}, (\sigma^{opt})^2) = N(y^*/\mu^*, (\sigma^*)^2) \quad (4.9)$$

where, μ^* is the predictive mean, $\mu^* = \mu^T \Phi(x^*)$; and $(\sigma^*)^2 = [(\sigma_1^*)^2, \dots, (\sigma_M^*)^2]^T$ is the predictive variance. This predictive variance is the sum of variances of two terms: the noise on the data and the uncertainty in the prediction of the weight parameters (Tipping 2004). The mathematical formulation, likelihood maximization, and optimization procedure of the RVM are discussed in detail in Tipping (2001), and Tipping and Faul (2003).

4.3 Methodology

4.3.1 Study Area

The Canal B irrigation command area of the Delta Canal Company lies in the Lower Sevier River Basin of Central Utah as shown in Fig. 4.1 was the study area for this research. This Sevier River Basin has seen significant canal lining and on-farm laser leveling for improved irrigation and delivery efficiencies. Canal automation was introduced in 1994 and resulted in substantial reduction in losses shortening of the

response time between farmer demands and system deliveries (Walker and Stringam 2000).

The study area consists of 10,425 irrigated hectares system and was organized to provide water from the upstream reservoir to an individual farm within 12 hours of an order by the irrigator. The weather data for this study were taken from the meteorological station located near the study area in Delta. This station is a part of the Community Environmental Monitoring Program (CEMP) network of 29 monitoring stations located in the Western states. It is operated and monitored by the Desert Research Institute (DRI) of the Nevada System of Higher Education. The station is located at 39°21'11"N latitude, 112°34'42"W longitude and 1,415 m above sea level, and it records hourly solar radiation, minimum and maximum temperatures, wind speed, precipitation and relative humidity. These data are available on the CEMP website (CEMP 2012).

The weather data for 2009–2011 period were used to calculate the reference evapotranspiration using the FAO 56 Penman-Monteith (PM) equation (Allen et al. 1998), for the dates corresponding to the Landsat scenes used in this study.

4.3.2 Satellite Data

The satellite images used in this study were from Landsat 5 Thematic Mapper (TM), which has a 16-day revisit time. Landsat 5 TM data were processed by the Level 1 Product Generation System at the United States Geological Survey (USGS) Center for Earth Resources Observation and Science (EROS). Fourteen cloud-free images, listed in Table 4.1, were acquired for a period of 3 years (2009–2011) covering crops growth stages for the study area in Delta, UT. The TM sensor onboard of Landsat 5 has six bands with 30 m spatial resolution in the shortwave, near infrared, and mid-infrared portions of

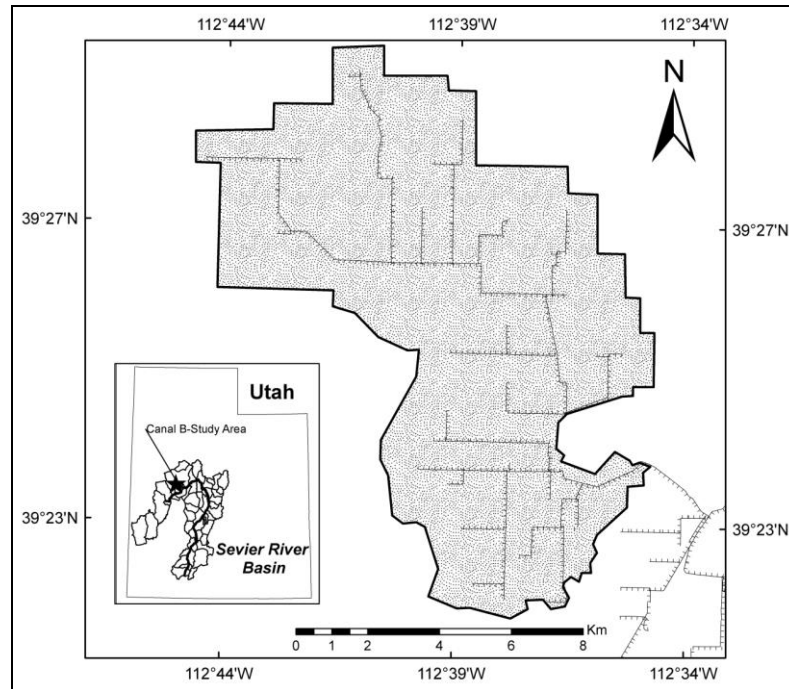


Fig. 4.1. Location of Canal B irrigation area in the Lower Sevier River Basin

the electromagnetic spectrum, while another band, the thermal band, has a spatial resolution of 120 m.

Table 4.1. Landsat Satellite Images Used for ET Estimation

Year	DOY	Satellite	Path/Row	Date
2009	149	Landsat 5 TM	38/33	29-May-09
	197	Landsat 5 TM	38/33	16-Jul-09
	213	Landsat 5 TM	38/33	1-Aug-09
	229	Landsat 5 TM	38/33	17-Aug-09
	261	Landsat 5 TM	38/33	18-Sep-09
2010	104	Landsat 5 TM	38/33	14-Apr-10
	136	Landsat 5 TM	38/33	16-May-10
	168	Landsat 5 TM	38/33	17-Jun-10
	200	Landsat 5 TM	38/33	19-Jul-10
	264	Landsat 5 TM	38/33	21-Sep-10
2011	107	Landsat 5 TM	38/33	17-Apr-11
	203	Landsat 5 TM	38/33	22-Jul-11
	251	Landsat 5 TM	38/33	8-Sep-11
	267	Landsat 5 TM	38/33	24-Sep-11

4.3.3 Image Processing

The acquired images from Landsat 5 downloaded as L1G data from USGS are not atmospherically corrected. Atmospheric corrections were made to these images to convert the radiance measurements at the top of the atmosphere to the surface-level reflectance. Surface emissivity, albedo, surface temperature and all the required inputs for METRIC were calculated using the procedures from Allen et al. (2012). Reflectance data in the visible and near-infrared were used to calculate two vegetation indices, the soil adjusted vegetation index, *SAVI*, and the normalized difference vegetation index, *NDVI*. The leaf area index, *LAI*, maps were also produced. All of these indices maps were used as inputs for the METRIC model. The following equations were used for the vegetation indices:

$$SAVI = (R_{NIR} - R_{Red})(1 + L)/(R_{NIR} + R_{Red} + L) \quad (4.10)$$

$$NDVI = (R_{NIR} - R_{Red})/(R_{NIR} + R_{Red}) \quad (4.11)$$

where, R_{NIR} , R_{Red} , = apparent reflectance values in the near-infrared (0.78-0.90 μm) and red (0.63-0.69 μm) wavebands respectively; L = calibration factor (Huete 1988).

The *LAI* is the ratio of the total area of all leaves on a plant to the ground area represented by the plant and ranges from 0 to 6. *LAI* is dimensionless ($\text{m}^2 \text{m}^{-2}$) and is an indicator of biomass and canopy resistance to vapor flux. It is computed using the following empirical equation from Allen et al. (2012) :

$$\begin{aligned} LAI &= 11 * SAVI^3; & \text{for } SAVI < 0.817 \\ LAI &= 6 ; & \text{for } SAVI > 0.817 \end{aligned} \quad (4.12)$$

All the image processing and mapping was done using ERDAS Imagine software. These resulting images, along with the weather data for the study area served as inputs for the energy balance model (METRIC) to derive actual ET maps at 30 m resolution.

4.3.4 Model Selection

A relevance vector machine (RVM) was developed using, information about land use, *NDVI*, *LAI*, crop classification and reference ET calculated from the weather station data in the study area as inputs. Urban areas, roads and all non-agricultural lands were excluded from the set of input using the land use map provided by the National Land Cover Database (NLCD 2006). A supervised crop classification was done in ERDAS Imagine for each of the three years of the study period using *NDVI* information. Three main crops were identified: alfalfa, corn, and small grains (barley, wheat). Fallow vegetation was also detected during the crop classification.

Each of the fourteen images consisted of about 120,000 pixels at 30 m spatial resolution. Information from each pixel (*NDVI*, *LAI*, crop class, reference ET) were treated as an independent set of inputs for the RVM model. The output of the model was the ET for each corresponding pixel as estimated by METRIC. The inputs and outputs for the RVM model are illustrated in the flow diagram (Fig. 4.2). In order to train the model, a simple random sample of 25,000 pixels from all the processed images was extracted. This number was chosen to balance computational time and model performance. All the remaining pixels were kept aside as test data to validate the performance of the machine. Simple random sampling has the advantage over other designs like stratified random sampling and systematic sampling, because it is easy to apply and provides satisfactory

results in evaluating the accuracy assessment of remotely sensed data (Congalton 1991; Kalkhan 2011).

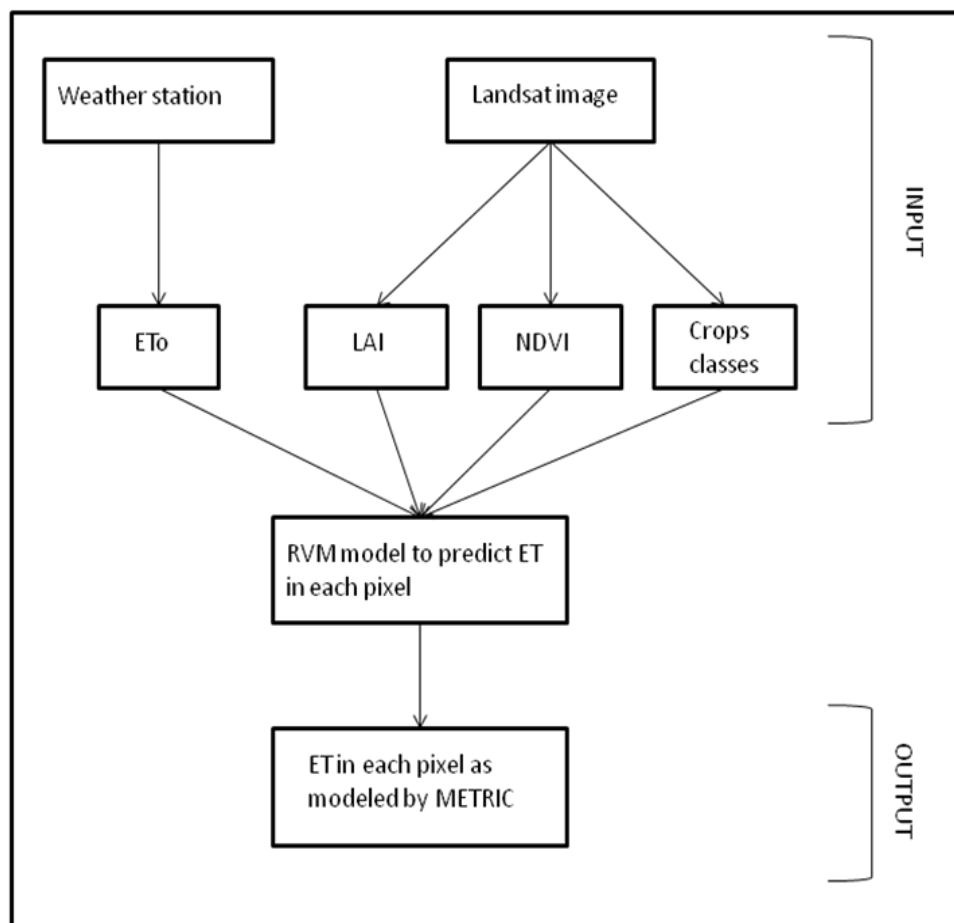


Fig. 4.2. Flow diagram of the RVM model inputs and outputs used in this study

4.3.5 Model Performance

For the purpose of testing the degree of association between the calculated and estimated ET, performance evaluation measures were used. To measure the magnitude of error, the root mean squared error, *RMSE*, was calculated [Eq. (4.13)]. Large values of *RMSE* mean that the difference between the actual measurements and the modeled values is large; hence, the model is not performing well. The *RMSE* has the same units as the data and therefore is easy to interpret.

$$RMSE = \left[\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N} \right] \quad (4.13)$$

The coefficient of determination, R^2 , was also used to evaluate to check the correlation between modeled and calculated ET [Eq.(4.14)]. It is a statistic that provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model.

$$R^2 = \frac{\left[\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) \right]^2}{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2} \quad (4.14)$$

The Nash-Sutcliffe efficiency, E , is a normalized statistic that determines the relative magnitude of the residual variance “noise” compared to the measured data variance “information” (Nash and Sutcliffe 1970). It is recommended by the ASCE (1993) and Legates and McCabe (1999) as a measure of model performance in hydrological modeling. It indicates how well the plot of observed versus simulated data fits the 1:1 line and is computed as shown in Eq. (4.15). It must be noted that an efficiency of 1 ($E = 1$) corresponds to a perfect match of modeled ET to the calculated data. An efficiency of 0 ($E = 0$) indicates that the model prediction is as accurate as the mean of the observed data, whereas an efficiency less than zero ($E < 0$) occurs when the observed mean is a better predictor than the model which indicates unacceptable performance.

$$E = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4.15)$$

The index of agreement, IA , was also used to check model performance [Eq. (4.16)]. The IA is calculated by comparing an observed group variance with an expected random variance. It ranges from zero (inferior model) to one (excellent model) (see Legates and McCabe 1999).

$$IA = 1.0 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2} \quad (4.16)$$

where, y_i = ET estimated for each pixel using METRIC (mm d^{-1}); \hat{y}_i = ET estimated by the RVM model (mm d^{-1}); \bar{y} = average of y_i ; $\bar{\hat{y}}$ = average of \hat{y}_i ; and N = total number of observations or pixels.

4.4 Results and Discussion

4.4.1 Spatial Distribution of ET by METRIC

METRIC was applied for all selected images from 2009 to 2011, and an actual daily ET map was produced for each of the Landsat scene listed in the Methods section. Each image was processed individually. Following the recommendations of Allen et al. (2013), a filter was applied to select several potential "hot" and "cold" pixels. Out of these potential pixels, one "hot" and one "cold" pixel were then manually selected to calibrate the model. The process was repeated until satisfactory results were obtained. An image for each month during the growing season was selected to show the change in ET across the irrigation season. The resulting daily ET maps at 30 m spatial resolution are presented in Fig. 4.3.

The values ranged between 0 and 8 mm, with patterns of ET linked to vegetation type. Lower values of ET ($<2\text{mm}$) corresponded to bare soils or fallow vegetation that

were mainly in the North-West side of the map, which is the farthest area from the canal diversion, and were not irrigated during the season. It is noticed that in May 16 image the fallow vegetation area was giving some values of ET, which represents the evaporation from the bare soil due to a rain event of 24 mm that preceded the day of acquiring the image. July was the month with higher ET mainly due to the full growth of the corn. The results indicated the potential of using METRIC as a tool for estimating irrigation water demand in order to support water management in this irrigation scheme, as applied by many researchers in many regions in the world (Allen et al. 2007; Folhes et al. 2009; Irmak et al. 2011; Singh et al. 2012). It is important to note the limitation of applying METRIC is that it is only applicable when a remotely sensed image is available for the area.

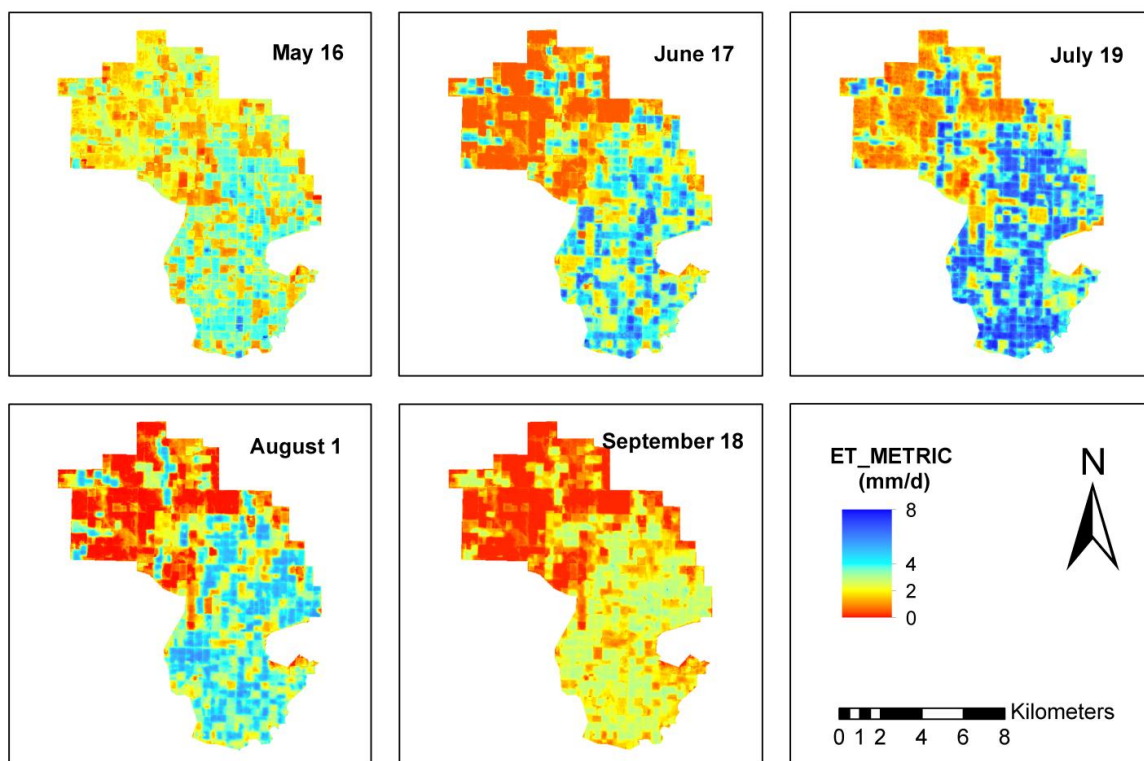


Fig.4.3. Spatial distribution of daily ET from METRIC for selected days over the Canal B area

4.4.2 System Efficiency

Mapping ET at spatial scale using METRIC was used to evaluate the irrigation system efficiency in the Canal B area. As mentioned earlier, the area of study is known to have advanced laser leveling performed on the lands. This gives the lands a low gradient for better irrigation application efficiency, i.e. very little tail water is generated. But the efficiency of the supply and distribution system needs to be evaluated for better management. The efficiency of the system was calculated as result of dividing the actual ET from METRIC for the area irrigated by Canal B for a specific day, by the canal diversions. The actual ET for the fallow areas were excluded from the calculations, since these fields were not irrigated. For each month during the irrigation season, one day was chosen to evaluate the system efficiency as shown in Table 4.2. The efficiency varied from 65% in June to 78% in July. Lower values are expected early in the season when the crop requirements are low, the root zone is shallow and the soil has higher intake rate. Therefore, low efficiency in the systems occurred in May (67.6%) and June (64.9%) when the irrigation of summer barley and corn starts. Better performance in the system was achieved in July, at the maximum crop water requirement and full vegetation cover. It is also noticed that the efficiency dropped again in September (65.1%). During this month, the farmers plough the lands again and plant the winter crops for the next season,

Table 4.2. Canal B Diversions and System Efficiency

Date	Canal B Diversions (m³ d⁻¹)	ET_METRIC (m³ d⁻¹)	System Efficiency (%)
May 16	248,196	167,741	67.6
Jun 17	314,440	204,106	64.9
July 19	317,192	247,842	78.1
August 1	284,726	202,559	71.1
Sept 18	190,594	124,068	65.1

which increases the intake rate of the soil, and the surface irrigation becomes less efficient.

4.4.3 RVM Model Results

METRIC has proven to be a reliable model to estimate spatial ET and was used widely for different management practices. However, the limitation of METRIC and other remotely sensed algorithms is their applicability only when a remotely sensed image is available, which is limited at the required high spatial resolution (French et al. 2012). Hence, an RVM model was developed in this study as an alternative model that can estimate spatial ET. The purpose of the RVM model was to model actual spatial ET as calculated by METRIC. Vegetation index (*NDVI*), reference ET, and *LAI* and the crop class were used as the RVM inputs. The RVM was trained and performance was tested based on selection of optimal kernel width and kernel type. The kernel types and width were discussed previously in Section 4.2.2. The highest *E* and least *RMSE* were obtained for a "Cauchy" kernel with a kernel width $r = 0.2205$. The performance criteria results for training and test data for the RVM model were shown in Table 4.3. The RVM demonstrated good performance with a training *RMSE* of 0.49 mm d^{-1} , R^2 of 0.84, *E* of 0.85 and *IA* of 0.95 in the training phase which indicated that observed data and modeled values were close. When tested on unseen data set or pixels, the model kept good performance with an *RMSE* = 0.5 mm d^{-1} , *E* = 0.84, *IA* = 0.81 and $R^2 = 0.82$ for the test data set. Fig. 4.4 shows a scatter plot of ET estimated by the RVM model and the ET calculated by METRIC. The points in the scatter plot all gathered around the 1:1 line indicating the ability of the model to predict ET. Some over or underestimation was also noticed for some pixels, which was explained by further analysis of the residuals.

Table 4.3. Performance Criteria Results for RVM Model in the Training and Testing Phase

Phase	Statistic	RVM
Training	RMSE (mm)	0.49
	E	0.85
	IA	0.82
	R ²	0.84
Test	RMSE (mm)	0.50
	E	0.84
	IA	0.81
	R ²	0.82

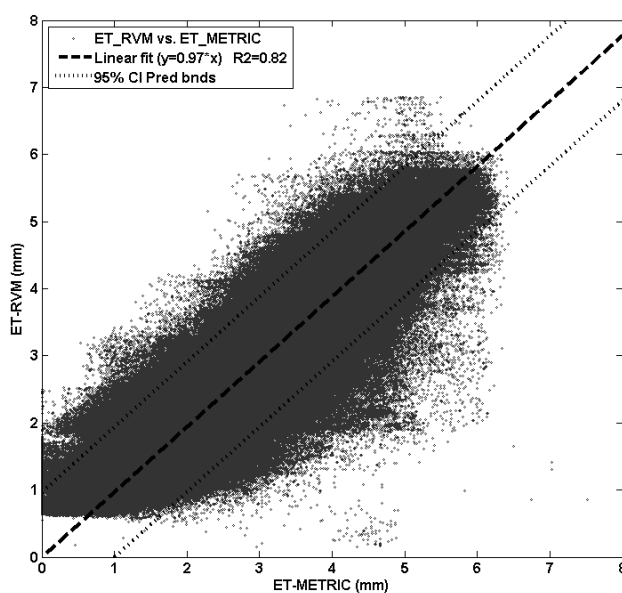


Fig.4.4. Modeled vs. calculated spatial ET for the study area for the unseen data

In order to display the RVM model performance over a growing season, an image was selected from each month of the growing season. The RVM was applied to those images after removing the pixels used in the training phase, and the results are summarized in Table 4.4. This table shows the total daily ET for the study area as calculated by METRIC and as modeled by the RVM model. The total residual errors ranged from $-2,232 \text{ m}^3 \text{ d}^{-1}$ to $8,026 \text{ m}^3 \text{ d}^{-1}$, with a maximum overestimation of 3.4% in the month of June.

Table 4.4. Comparison Between ET from METRIC and ET from RVM Model for Selected Days

Date	ET ($\text{m}^3 \text{d}^{-1}$)		Error	
	METRIC	RVM	$\text{m}^3 \text{day}^{-1}$	%
May 16	190,721	188,715	-2,006	-1.1
Jun 17	233,476	241,502	8,026	3.4
July 19	281,213	281,418	206	0.1
August 1	240,941	238,709	-2,232	-0.9
Sept 18	147,929	146,841	-1,088	-0.7

Further analysis of the residuals was conducted. Fig. 4.5 showed the maps of the residual errors over study area. Fig 4.6 illustrated the histogram of the residuals. These figures showed most of the errors falling within $\pm 0.6 \text{mm d}^{-1}$ of the ET calculated by METRIC. Some overestimation (red color) was noticed in the areas away from the delivery canal (North-West) that mainly corresponds to fallow vegetation that are not usually irrigated. These areas had very low ET and the model seemed to overestimate those values.

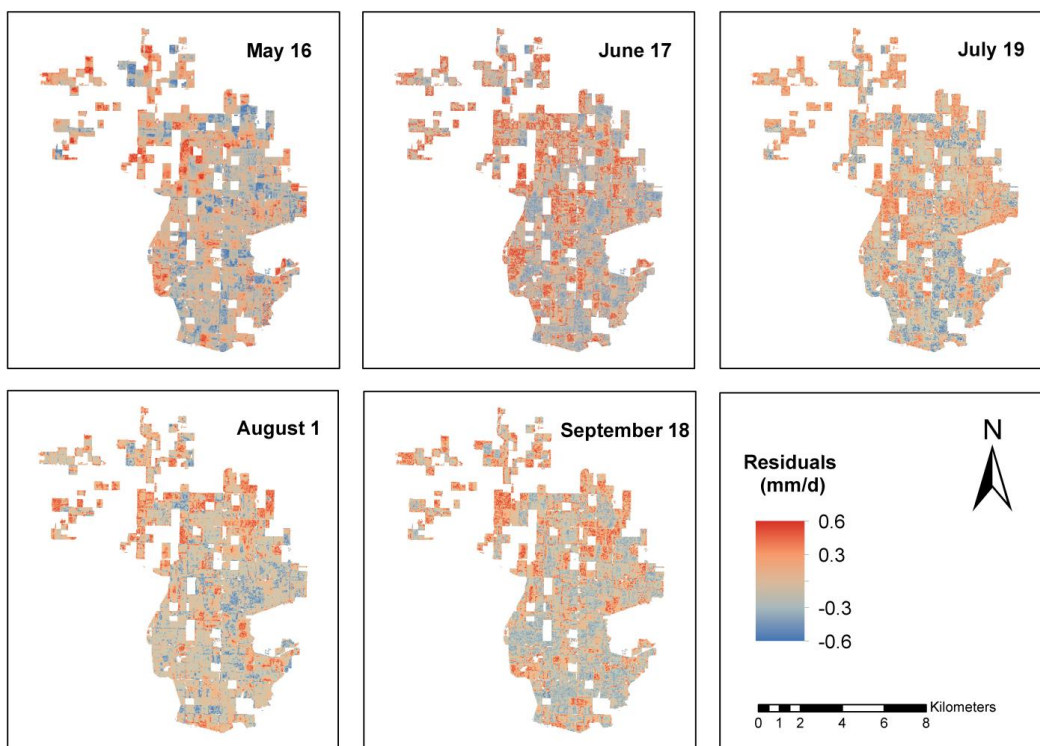


Fig. 4.5. Residual error maps for the study area for selected days

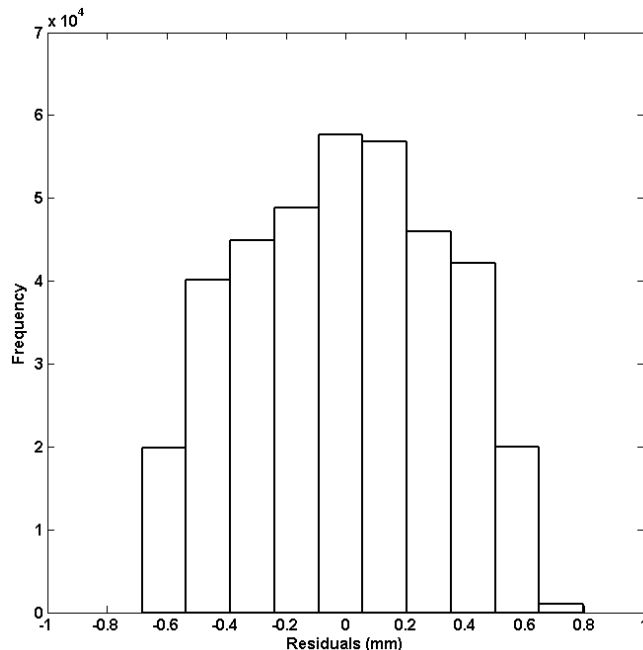


Fig. 4.6. Histogram of residual errors for the selected days

When developing the RVM model, one of the images was randomly selected and excluded from the training set. This image, from which no pixels were chosen to train the RVM, was used to test the practicability of the model. The overall Nash-Sutcliffe coefficient, E , was 0.83, $IA = 0.77$, $R^2 = 0.79$ and the $RMSE=0.67 \text{ mm d}^{-1}$, indicating good model performance. The total volume of water requirement estimated by METRIC for that day (May 29, 2009) was $192,961 \text{ m}^3 \text{ d}^{-1}$, while the RVM model estimated was $206,933 \text{ m}^3 \text{ d}^{-1}$ giving an over estimation of 7.2% for the overall area. Fig. 4.7 showed a close relation between ET map modeled by the RVM model and ET map calculated by METRIC. The RVM, being a regression model, tends to overestimate the low ET areas that corresponds usually to the fallow vegetation or bare soils.

The accuracy of the RVM modeling for images not included in the model training, makes it reliable for predicting ET at spatial scale using the input for the days when a Landsat image is not available. For that purpose, $NDVI$ and LAI can be

extrapolated in time using widely used temporal profile equations (Fischer 1994), and the reference ET could be forecasted accurately using machine learning techniques as presented in (Torres et al. 2011; Bachour et al. unpublished). The RVM model could be tested with real-time data without using a Landsat image to provide the model inputs. The output of this model could be utilized for development of decision-support systems that should be available to farmers and managers of reservoir releases and canal diversions.

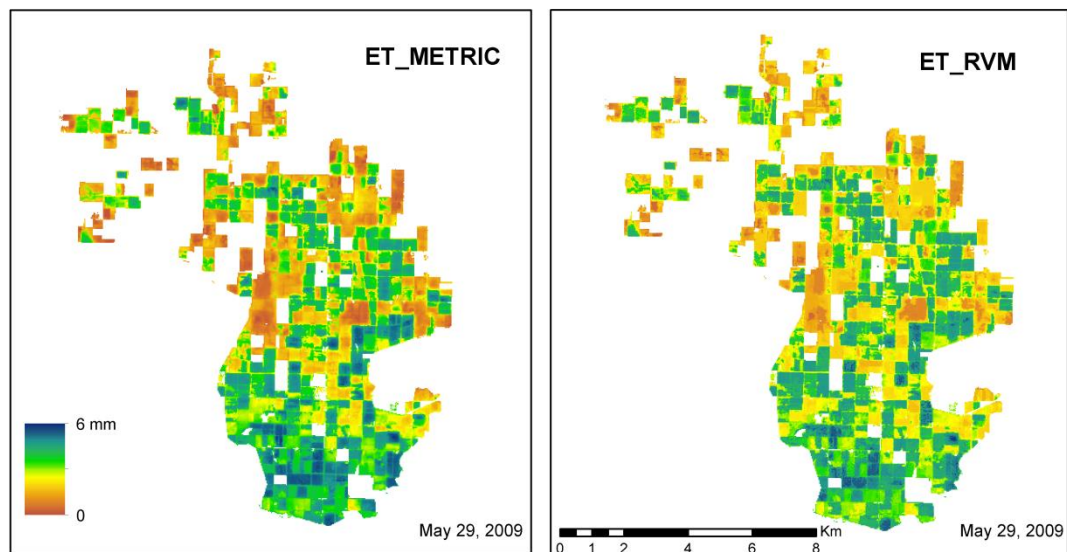


Fig. 4.7. ET_METRIC vs. ET_RVM for a test image

4.5 Conclusions

This study showed the application of METRIC algorithm in the Canal B irrigation command area of the Lower Sevier River Basin, UT. It provided the spatially distributed daily ET at high spatial resolution (30 m) over large irrigated areas using Landsat images and weather data. This is a very useful tool to monitor crop growth, and crop water demands and applications. The spatial distribution of ET helped evaluating the efficiency of the supply and distribution system, which can be used for better management of the canal diversions and irrigation applications on a field-by-field basis. Another important

element presented in this study, was the capability to produce spatially distributed estimations of ET using a data driven RVM model. The developed model used information about NDVI, LAI, crops classification and reference ET, and provided accurate estimation of spatial ET compared to METRIC.

This model lays the ground for the estimation of ET at spatial scale for the days when a Landsat image is not available or when there is cloud coverage. Such gaps in remote sensing information could reach a month or more when using Landsat. This methodology can be used for forecasting daily spatial ET, if the vegetation indices model inputs are extrapolated in time, and the reference ET forecasted accurately.

References

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). "Crop evapotranspiration: guidelines for computing crop water requirements." *FAO Irrigation and Drainage Paper No. 56*, FAO, Rome, 300.
- Allen, R. G., Morse, A., and Tasumi, M. (2003). "Application of SEBAL for western US water rights regulation and planning." *International Workshop on Remote Sensing of evapotranspiration for Large Regions. IEC Meeting of International Commission on Irrigation and Drainage (ICID) Proceedings*, p. 54.
- Allen, R. G., Tasumi, M., Morse, A., and Trezza, R. (2005). "A Landsat-based energy balance and evapotranspiration model in Western US water rights regulation and planning." *Irrig. and Drain. Syst.*, 19, 251-268.
- Allen, R. G., Tasumi, M., and Trezza, R. (2007). "Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)-Model." *J. Irr. Drain Eng.*, 133(4), 380-394.
- Allen, R. G., Irmak, A., Trezza, R., Hendrickx, J. M. H., Bastiaanssen, W. G. M., and Kjaersgaard, J. (2011). "Satellite-based ET estimation in agriculture using SEBAL and METRIC." *Hydrol. Process.*, 25, 4011-4027.
- Allen, R. G., Trezza, R., Tasumi, M., and Kjaersgaard, J. (2012). "Mapping evapotranspiration at high resolution using internalized calibration: application manual for Landsat satellite imagery." *University of Idaho*, Version 2.0.8, March 2012, Kimberly, Idaho.

- Allen, R. G., Burnett, B., Kramber, W., Huntington, J., Kjaersgaard, J., Kilic, A., Kelly, C. and Trezza, R. (2013). "Automated calibration of the METRIC-Landsat evapotranspiration process." *J. Am. Water Resour. Assoc. (JAWRA)*, 49(3),563-576.
- Anderson, M. C., Norman, J. M., Mecikalski, J. R., Otkin, J. P., and Kustas, W. P. (2007). "A climatological study of evapotranspiration and moisture stress across the continental U.S. based on thermal remote sensing: I. Model formulation." *J. Geophys. Res.*, 112, D10117.
- Anderson, M. C., Allen, R. G., Morse, A., and Kustas, W. P. (2012). "Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources." *Remote Sensing of Environ.*, 122, 50-65.
- ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models of the Watershed Management, Irrigation, Drainage Division (ASCE). (1993). "Criteria for evaluation of watershed models. *J. Irr. Drain. Eng.*, 119(3), 429-442.
- Bastiaanssen, W. G. M. (1995). "Regionalization of surface flux densities and moisture indicators in composite terrain: a remote sensing approach under clear skies in Mediterranean climates." *PhD Thesis*, University of Wageningen, Netherlands.
- Bastiaanssen, W. G. M., Menenti, M., Feddes, R. A., and Holtslag, A. A. M. (1998). "A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation." *J. Hydrol.*, 212, 198-212.
- Bastiaanssen, W. G. M., Allen, R. G., Pelgrum, H., Texeira, A. H., Soppe, R. W. O., and Thoreson, B. P. (2008). "Thermal-infrared technology for local and regional scale irrigation analyses in horticultural systems." *Acta Horticulturae*, 792 (Mildura Special Issue, editors I. Goodwin and M.G. O'Connell), ISHS.
- CEMP- Community Environmental Monitoring Program website, <http://www.cemp.dri.edu>, accessed in December 2012.
- Congalton, R. G. (1991). "A review of assessing accuracy of classifications of remotely sensed data." *Remote Sensing of Environ.*, 37, 35-46.
- Fischer, A. (1994). "A simple model for the temporal variations of NDVI at a regional scale over agricultural countries. Validation with ground radiometric measurements." *Int. J. Remote Sensing*, 15(7), 1421-1446.
- Folhes, M. T., Renno, C. D., and Soares, J. V. (2009). "Remote sensing for irrigation water management in the semi-arid Northeast of Brazil." *Agric. Water Manage.*, 969, 1398-1408.

- French, A. N., Hunsaker, D. J., and Clarke, T. R. (2012). "Forecasting Spatially Distributed Cotton Evapotranspiration by Assimilating Remotely Sensed and Ground-Based Observations." *J. Irrig. and Drain. Eng.*, 138(11), 984-992.
- Gowda, P. H., Chavez, J. L., Colaizzi, P. D., Evett, S. R., Howell, T. A., and Tolk, J. A. (2008). "ET mapping for agricultural water management: present status and challenges." *Irrig. Sci.*, 26, 223-237.
- Huete, A.R. (1988). "A soil-adjusted vegetation index (SAVI)". *Remote Sensing of Environ.*, 25, 53-70.
- Irmak, A., Ratcliffe, I., Ranade, P., Hubbard, K., Singh, R. K., Kamble, B., and Kjaersgaard, J. (2011). "Estimation of land surface evapotranspiration with a satellite remote sensing procedure." *Great Plains Res.*, 21, 73-88.
- Kaheil, Y. H., Rosero, E., Gill, M. K., McKee, M., and Bastidas, L. A. (2008). "Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines." *Trans. Geoscience and Remote Sensing, IEEE*, 46(9), 2692-2707.
- Kalkhan, M. A. (2011). *Geospatial information modeling and thematic mapping*. CRC Press, Taylor & Francis, Boca Raton, Florida.
- Legates, D. R., and McCabe, G. J. (1999). "Evaluating the use of "goodness-of-fit" measures in hydrological and hydroclimatic model validation." *Water Resour. Res.*, 35(1), 233-241.
- Mausser, W., and Schädlich, S. (1998). "Modelling the spatial distribution of evapotranspiration on different scales using remote sensing data." *J. Hydrol.*, 212, 250-267.
- Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models." *I. J. Hydrol.*, 10, 282-290.
- Neale, C. M. U., Geli H. M. E., Kustas, W. P., Alfieri, J. G., Gowda, P. H., Evett, S. R., Prueger, J. H., Hipps, L. E., Dulaney, W.P., Chávez, J. L., French, A. N., and Howell, T. A. (2012). "Soil water content estimation using a remote sensing based hybrid evapotranspiration modeling approach." *Advances in Water Resour.*, 50, 152-161.
- NLCD (2006) - National Land Cover Database website,
<http://www.mrlc.gov/nlcd2006.php>.
- Singh, R. K., Liu, S., Tieszen, L. L., Suyker, A. E. and Verma, S. B. (2012). "Estimating seasonal evapotranspiration from temporal satellite images." *Irrig. Sci.*, 30, 303-313.

- Tasumi, M., Trezza, R., Allen, R. G., and Wright, J. L. (2005). "Operational aspects of satellite-based energy balance models for irrigated crops in the semi-arid U.S." *Irrig. Drain. Syst.*, 19, 355-376.
- Thenkabail, P. S., Biradar, C. M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O. P., Turrall, H., Cai, X., Vithanage, J., Schull, M. A., and Dutta, R. (2009). "Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium." *Intl. J. Remote Sensing*, 30(14), 3679-3733.
- Ticlavilca, A. M., McKee, M., and Walker, W. R. (2011). "Real-time forecasting of short-term irrigation canal demands using a robust multivariate Bayesian learning model." *Irrig. Sci.*, 31(2), 151-167.
- Tipping, M. E. (2001). "Sparse Bayesian learning and the relevance vector machine." *J. Mach. Learn.*, 1, 211-244.
- Tipping, M. E. (2004). "Bayesian inference: an introduction to principles and practice in machine learning." In: Bousquet, O., von Luxburg, U., and Ratsch, G. (eds). *Advanced lectures on machine learning*. Springer, Berlin, pp. 41-62.
- Tipping, M. E., and Faul, A. C. (2003). "Fast marginal likelihood maximisation for sparse Bayesian models." In: Bishop, C. M., and Frey, B. J. (eds). *Proceedings of the ninth international workshop on artificial intelligence and statistics*, Key West, Florida, pp. 596-608.
- Torres, A. F., Walker, W. R., and McKee, M. (2011). "Forecasting daily potential evapotranspiration using machine learning and limited climatic data." *Agric. Water Manage.*, 98(4), 553-562.
- Walker, W. R., and Stringam, B. L. (2000). "Canal automation for water conservation and improved flexibility." In: *Proc., 4th Decennial Nat. Irrigation Symposium*.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Chapters 2 to 4 presented the body of the work and the main scientific results of the dissertation. In this chapter, a summary of the work is presented, important conclusions, and recommendations for future research are presented.

5.1 Summary and Conclusions

Efficient water management is becoming very critical in meeting the needs of the growing population. An effective water management approach requires that water users become involved in the decision-making and management process. In the management of irrigation command areas, and in order to build a better plan to manage service delivery from canals and reservoirs, it is important to build an appropriate knowledge base about water needs at the level of an individual field. There is often a lag between the order of water by a farmer and delivery of water to the field. Knowledge about the crop water requirement at the field level helps the system operator to make the right choices leading to more efficient handling of the available water.

With these goals in mind, this dissertation attempted to develop procedures which give information to farmers and irrigators about the crop water requirements of their fields. This information could help in the overall improvement of water management practices, both on-farm and system wide.

The methods used in this dissertation were directed toward designing machine learning models which can be used as tools for decision support systems for water management. They adapt to time-varying behavior and incrementally learn changes as

they come across more data in near real time. To achieve better modeling and prediction, wavelet decompositions were explored for their ability to give information about time and frequency changes in the data. Remote sensing approaches were also used for their ability to quantify water requirements at the spatial level. Therefore, this dissertation explored the use of the above-mentioned data, tools, and techniques to address water management problems. The framework of this dissertation consisted of three components that attempted to provide tools to support irrigation system operational decisions.

In Chapter 2, the Hargreaves (HG) equation to calculate ET was evaluated under semiarid conditions in the Bekaa Valley of Lebanon using 16 years of complete daily climatic data from the Terbol weather station. HG results were compared to ET estimates obtained from the FAO56 Penman Monteith equation (PM), which was used as a standard. A modification of the HG equation was then developed to achieve better accuracy of ET estimation using only temperature data. A slight additional improvement in HG estimation accuracy was attained by adding the wind speed variable. Therefore, it was recommended to use the calibrated HG equation when only temperature data are available, otherwise when complete and reliable weather data exist, the use of the standard PM equation is recommended.

Chapter 3 introduced a methodology to combine wavelet multiresolution analysis (MRA) with a multivariate relevance vector machine (MVRVM) to predict 16 consecutive days of daily ET. In this Chapter, different wavelet decompositions were performed and combined with the MVRVM to develop the hybrid prediction models. These performed better than MVRVM model with higher accuracy and robustness. This is a valuable element for improving the efficiency of irrigation water delivery systems. It

also helps decision-makers and operators to manage irrigation water delivery systems. Accurate forecast of daily ET is beneficial in estimating future crop water demands and responding with appropriate operational decisions for canal delivery/distribution systems. These forecasting models lay the ground for forecasting the spatial distribution of ET using remote sensing.

Chapter 4, Mapping Evapotranspiration at high Resolution with Internalized Calibration (METRIC) algorithm, was used in this study to calculate actual ET from Landsat 5 Thematic Mapper images at 30m spatial resolution in the Canal B irrigation command area of the Delta Canal Company in Central Utah for the 2009-2011 growing seasons. These ET maps were used to evaluate the efficiency of the irrigation supply system in the Canal B area. The efficiency ranged from 65% to 78% throughout the growing season. The second issue addressed in this Chapter was the ability to produce spatially distributed ET using data-driven modeling which can then be used when satellite imagery is not available. For this purpose, a machine learning algorithm, a relevance vector machine (RVM), was built and trained with a set of inputs of vegetation indices, land use, soil texture and weather data in order to predict the ET from METRIC as output. The resulting RVM model provided accurate estimation of spatial ET compared to METRIC.

5.2 Recommendations for Future Work

This dissertation was an effort to advance the use of wavelets, machine learning and remote sensing tools and their application in water resources management. In general, the results for each of the methods developed were satisfactory, relevant and encouraging. They provide significant potential for improving decision making for real-

time applications in river basin management, and can be extended to other fields and applications in water resource management.

Keeping in mind the concepts developed in this research and the results demonstrated, we recommend future research that will provide better strategies for improved irrigation water management, in particular, and water resources management in general.

The calibration procedure for the ET equation for semiarid conditions draws attention toward the use of this equation to estimate crop water requirements in similar conditions. Also, depending on data availability, it can be tested whether this calibration method can provide a good estimate at hourly time steps. This will be useful for purposes of applying remote sensing algorithms to calculate the spatial distribution of ET areas where complete reliable weather data are not available, such as the case of Lebanon.

The wavelet-MVRVM models presented in this study provided a reliable forecast of ET for 16 consecutive days ahead. The RVM model demonstrates a strong ability to distribute ET to the spatial level. This lays the grounds for further investigations that could lead to spatial forecasts of ET, given that the vegetation indices model inputs can be extrapolated in time, and the reference ET can be forecasted accurately. The spatial forecasting of ET will help decision-makers and water system managers to establish appropriate strategies in the operation of irrigation systems. Such spatial forecasts, if applied on a real-time basis, solve the problem of irrigation requirements estimation on large scales.

Including 10-days of forecasted minimum and maximum air temperatures as inputs to the wavelet-MVRVM models improved slightly the performance of these

models, especially for the first 10 days for which the weather forecast is available. This encourages the exploration of adding other weather variables in future research.

Remote sensing data is becoming more available. This provides more frequent information at different spatial resolutions. For example, the MODIS satellite provides daily images at 250m to 1km spatial resolution. Downscaling methods could be combined with the suggested methodologies to improve the spatial forecasting of ET.

Recently, the Landsat 8 satellite became operational at a 16-day revisit cycle. It operates in alternation with Landsat 7, which together allows users to acquire 30 m resolution images every 8 days. The hybrid modeling approach could be applied to provide more accurate 8-day ET forecasts.

All these tools should become operational and applied in real-time. They, or the data they produce, should be easily available for end-users (farmers and canal operators). A test phase for these tools should be done first. This will allow feedback from the end-users to the modeler to adapt the models in a comprehensible manner for them. It also allows refining of the models to adjust for real-time working conditions. It can also provide enough time for the farmers and operators to trust and adapt to the new sources of information.

APPENDIX

Permission Letters

From: Roula Bachour [mailto:bachour.roula@gmail.com]
Sent: Monday, July 15, 2013 9:19 PM
To: PERMISSIONS
Subject: Permission to use paper in dissertation

Dear editors,

I am requesting permissions to my paper in the ASCE Journal of Irrigation and Drainage Engineering, as a Chapter in my dissertation.

Here is the DOI of my paper and the reference:

DOI: 10.1061/(ASCE)IR.1943-4774.0000646 (Jun. 13, 2013).

Bachour, R., Walker, W., Torres-Rua, A., and McKee, M. (2013). (2013). "Assessment of Reference Evapotranspiration by the Hargreaves Method in the Bekaa Valley, Lebanon." *J. Irrig. Drain Eng.*, 10.1061/(ASCE)IR.1943-4774.0000646 (Jun. 13, 2013).
<http://ascelibrary.org/doi/abs/10.1061/%28ASCE%29IR.1943-4774.0000646>

Thank you for your help,

Roula Bachour

--

Roula Bachour

PhD Candidate, Irrigation Engineering
Civil and Environmental Engineering Department
Utah Water Research Laboratory
Utah State University
1600 East Canyon Rd
Logan, UT 84321

From: **PERMISSIONS** <permissions@asce.org>
to: Roula Bachour <bachour.roula@gmail.com>
date: Tue, Jul 16, 2013 at 4:39 AM
Subject: RE: Permission to use paper in dissertation
mailed-by: asce.org

Dear Roula Bachour:

Permission is given for you to reuse your paper "Assessment of Reference Evapotranspiration by the Hargreaves Method in the Bekaa Valley, Lebanon" in your dissertation. A full credit line must be added to the material being reprinted. For reuse in non-ASCE publications, add the words "With permission from ASCE" to your source

citation. For Intranet posting, add the following additional notice: "This material may be downloaded for personal use only. Any other use requires prior permission of the American Society of Civil Engineers.

Regards,

Joann Fogleson
American Society of Civil Engineers
1801 Alexander Bell Drive
Reston, VA 20191
[\(703\) 295-6278](tel:(703)295-6278)-FAX
PERMISSIONS@asce.org

A full credit line must be added to the material being reprinted. For reuse in non-ASCE publications, add the words "With permission from ASCE" to your source citation. For Intranet posting, add the following additional notice: "This material may be downloaded for personal use only. Any other use requires prior permission of the American Society of Civil Engineers.

Each license is unique, covering only the terms and conditions specified in it. Even if you have obtained a license for certain ASCE copyrighted content, you will need to obtain another license if you plan to reuse that content outside the terms of the existing license. For example: If you already have a license to reuse a figure in a journal, you still need a new license to use the same figure in a magazine. You need separate license for each edition. For more information please view Terms and Conditions at: <http://www.asce.org/Books-and-Journals/Permissions/Permission-Requests/Terms-and-Conditions/>

From: Roula Bachour [mailto:bachour.roula@gmail.com]
Sent: Monday, July 01, 2013 10:14 AM
To: Alfonso Torres-Rua
Subject: Permission to include papers in dissertation

Dear Alfonso,

I am in the process of preparing my dissertation. And I'm sending this email to request your permission to include in my dissertation a copy of the following paper in which you appear as a coauthor:

"ASSESSMENT OF REFERENCE EVAPOTRANSPIRATION BY THE HARGREAVES METHOD IN THE BEKAA VALLEY, LEBANON"

Thank you for your help,

Roula Bachour

Alfonso Torres <Alfonso.Torres@usu.edu> Monday, July 01, 2013 10:19 AM
to: Roula Bachour <bachour.roula@gmail.com>

Dear Roula,

You have my permission.

Alfonso

Alfonso Torres-Rua, Ph.D.
Research Engineer
Utah Water Research Laboratory
Utah State University
1600 East Canyon Rd
Logan, UT 84321
Phone: [+1 \(435\) 890-0196](tel:+14358900196)

From: Roula Bachour [mailto:bachour.roula@gmail.com]
Sent: Monday, July 01, 2013 10:15 AM
To: Inga Maslova <maslova@american.edu>
Subject: Permission to include papers in dissertation

Dear Inga,

I am in the process of preparing my dissertation. And I'm sending this email to request your permission to include in my dissertation a copy of the following papers in which you appear as a co-author:

"WAVELET-MULTIVARIATE RELEVANCE VECTOR MACHINE HYBRID MODEL FOR FORECASTING DAILY EVAPOTRANSPIRATION"

and

"SPATIAL DISTRIBUTION OF EVAPOTRANSPIRATION USING REMOTE SENSING AND RELEVANCE VECTOR MACHINE"

Thank you for your help,

Roula Bachour

Inga Maslova <maslova@american.edu> Monday, July 01, 2013 10:33 AM
to: Roula Bachour <bachour.roula@gmail.com>

Dear Roula,

I give my permission to reprint the following material in your dissertation:

Chapter 3 of this dissertation based on manuscript "WAVELET-MULTIVARIATE RELEVANCE VECTOR MACHINE HYBRID MODEL FOR FORECASTING DAILY EVAPOTRANSPIRATION".

and

Chapter 4 of this dissertation based on manuscript "SPATIAL DISTRIBUTION OF EVAPOTRANSPIRATION USING REMOTE SENSING AND RELEVANCE VECTOR MACHINE".

Sincerely,
Inga Maslova
Graduate Program Director
Assistant Professor
Department of Math and Stat
American University
4400 Massachusetts Ave, NW
Washington, DC 20016-8050
Office: (202) 885-3132
<https://sites.google.com/site/ingamaslova/>

From: Roula Bachour [mailto:bachour.roula@gmail.com]

Sent: Monday, July 01, 2013 10:15 AM

To: Andres Ticlavilca <andres.t@aggiemail.usu.edu>

Subject: Permission to include papers in dissertation

Dear Andres,

I am in the process of preparing my dissertation. And I'm sending this email to request your permission to include in my dissertation a copy of the following papers in which you appear as a co-author:

"WAVELET-MULTIVARIATE RELEVANCE VECTOR MACHINE HYBRID MODEL FOR FORECASTING DAILY EVAPOTRANSPIRATION"

and

"SPATIAL DISTRIBUTION OF EVAPOTRANSPIRATION USING REMOTE SENSING AND RELEVANCE VECTOR MACHINE"

Thank you for your help,

Roula Bachour

Andres Ticlavilca <andres.t@aggiemail.usu.edu> Monday, July 01, 2013 10:37 AM

to: Roula Bachour <bachour.roula@gmail.com>

Dear Roula,

You have my permission.

Andres

--

Andres M. Ticlavilca
Research Engineer
Utah Water Research Laboratory
8200 Old Main Hill
Logan, UT 84322-8200
Phone: [435-757-0851](tel:435-757-0851)
Email: andres.t@aggiemail.usu.edu
Web: <https://sites.google.com/a/aggiemail.usu.edu/ticlavilca/>

CURRICULUM VITAE

Roula Bachour

Utah State University
 Utah Water Research Laboratory,
 1600 Canyon Rd, Logan, UT, 84321, USA
 Cell phone: (435) 757-1693
 Email: bachour.roula@gmail.com

EDUCATION

Ph.D., Civil and Environmental Engineering (Irrigation Engineering), (Expected August 2013)

Utah State University, Logan, Utah, USA.

Dissertation title: Evapotranspiration Modeling and Forecasting for Efficient Management of Irrigation Command Areas.

Advisor: Dr. Wynn R. Walker

M.Sc., Irrigation, (June 2009)

American University of Beirut, Beirut, Lebanon.

Thesis title: Effects of Regular Deficit Irrigation and Root Wetting on Yield of Potato and Pepper.

Advisor: Dr. Musa Nimah

B.S., Agriculture Engineering, (July 2005)

Lebanese University, Faculty of Agricultural Sciences, Lebanon.

Thesis title: Effects of Regulated Deficit Irrigation on Quality and Radiation-Use Efficiency of Drip-Irrigated Bell Pepper.

Advisor: Dr. Fadi Karam

PROFESSIONAL SUMMARY

- Graduate Coursework and emphasis in water management, irrigation and remote sensing.
- Experience in developing decision support system models for water resources management.
- Experience in teaching undergraduate and graduate level classes.
- Over 3 years of professional experience in consulting conservation agriculture
- Extensive domestic/international experience in the field of agriculture and water management.
- Excellent interpersonal and communication skills; fluent in English, French and Arabic.

RESEARCH/WORK EXPERIENCE

Graduate Research Assistant, Utah Water Research Laboratory - 2009-Present

- Developed models to forecast evapotranspiration using wavelets techniques and statistical learning machines.
- Modeled evapotranspiration using remote sensing and developed maps for Sevier River Basin, Utah at high resolution from satellite imagery using METRIC™. ArcGIS and ERDAS Imagine were also used.

- Installed and performed maintenance work of soil moisture data collection system in Delta, UT for river management project, and in Scipio, UT, for precision agriculture project, in collaboration with the US Bureau of Reclamation (USBR)
- Developed and implemented statistical learning algorithms (Multivariate Relevance Vector Machine) for various applications in water management
- Designed a program for applied research in the Orontes River Basin, Lebanon as part of Middle East Water and Livelihood Initiative (WLI) (ICARDA, USAID)
- Assisted in organizing and training of several Lebanese and Syrian agriculture engineers and extension agents at the Agricultural Research and Education Center, Lebanon as a part of WLI activities in Lebanon
- Trained water and irrigation specialists about Irrigation Principles and Practices as part of the Iraq Agricultural Extension Revitalization project (IAER)
- Assisted in teaching and grading Water Quality Modeling course (CEE 6740), and Introduction to Statistics course (STAT 1040)

Consultant, International German Technical Cooperation (GIZ), Lebanon - 2007-2009

- Worked as a consultant to promote Conservation Agriculture (CA) to Lebanon under the "Development of rain-fed agriculture under zero-tillage in Lebanon" project
- Carried out field experiments to compare CA to traditional agriculture
- Initiated and organized seminars, workshops, and field days for engineers, service providers, students, and farmers about application of CA

Graduate Research Assistant, American University of Beirut, Lebanon - 2007-2008

- Designed and installed irrigation systems, and conducted experiments on water saving in agriculture
- Taught labs for several courses: Irrigation Principles (AGSC 228), Introduction to Soils (AGSC 215), Principles of Agronomy (AGSC 231), Introduction to Landscape Pests (LDEM 215) in the laboratories and in the field

Irrigation Engineer, Lebanese Agricultural Research Institute (LARI), Lebanon - 2005-2007

- Conducted experiments about deficit irrigation on vegetable crops
- Worked as technical site engineer for IRRIVAL FP6 Project: "Sustainable Orchard Irrigation for Improving Fruit Quality and Safety"
- Developed questionnaire for irrigation practices in Lebanon, helped farmers improve their irrigation systems

PUBLICATIONS and CONFERENCE PRESENTATIONS

Bachour R., W. R. Walker, A. Torres-Rua, M. McKee. 2013. "Assessment of Reference Evapotranspiration in the Beqa'a Valley, Lebanon". *Journal of Irrigation and Drainage Engineering* (ASCE), 10.1061/(ASCE)IR.1943-4774.0000646 (Vol. 139, No. 11, November 1, 2013)

Bachour R., I. Maslova, W. R. Walker, M. McKee, and A. Ticlavilca. 2012. "Wavelet-based Evapotranspiration Forecasts". *American Geophysical Union*, Poster presentation

Bachour R., M. Al Arab, M. Nimah, M. McKee. 2011. "Research Priorities for Water and Irrigation System Management on the Orontes Benchmark Site, Lebanon". Middle East Water and Livelihoods Initiative (WLI) Technical Report

- Farajalla N.S., Ziadeh R. and **R. Bachour**. 2011. "Impact of Climate Change on Trends of Drought Frequency and Consumptive Water Use in the Eastern Mediterranean" Symposium publication in: Water and Energy in Sustainable Food Security, Lebanon. Dec. 2-3, 2011
- Farajalla N.S., Ziadeh R. and **R. Bachour**. 2011. "Drought Frequency and Evapotranspiration Trends under a Changing Climate in the Eastern Mediterranean" Water Scarcity and Policy in the Middle East and Mediterranean Conference, Organized by Cornell University, Ithaca, NY Nov. 4-6, 2011
- Nimah M.N. and **R. Bachour** . 2010. "Modeling Deficit Irrigation Water Productivity as a Function of Crop Root Depth". ASABE Paper No. 1008942. Pittsburg, Pennsylvania, USA
- Nimah M.N, **R. Bachour** and R. Rabaa. 2010. "Comparison between Irrigation with Saline Water and Regular Deficit Irrigation on Potato Yield". USAID/MERC Potato Symposium 2010. Larnaca, Cyprus
- F. Karam, R. Masaad, **R. Bachour**, C. Rhayem and Y. Roupheal. 2009. "Water and Radiation Use Efficiencies in Drip-irrigated Pepper (*Capsicum annuum* L.): Response to Full and Deficit Irrigation Regimes". *Europ. J. Hort. Sci* 74 (2): 79-85
- Bachour R.**, K. Jouni, C. Estphan, S. Skaff, B. Hansmann and I. Bashour. 2009. "Field Experiments on Conservation Agriculture in Lebanon". 10th International Meeting on Soils with Mediterranean Type of Climate. Jun. 22-26. Beirut, Lebanon
- Farajalla N., **R. Bachour** and J. El-Jadam. 2009. "Impact of Climate Change on Water Resources in Lebanon – a Look at Crop Water Consumptive Use. European Geosciences Union. Vienna, Austria, 19-24 April 2009
- Bashour I., **R. Bachour** and K. Jouni. 2009. "Sustainable Agriculture: No-Till Farming". Expert Group Meeting on Sustainable Land Management as a Best Practice to Enhance Rural Development in the ESCWA Region. Beirut, 25-27 March 2009
- Bashour I. and **R. Bachour**. 2008. "Field Experiments: Conventional and Conservation Agriculture in Lebanon". Workshop on Investing in Sustainable Crop Intensification: The Case for Improving Soil Health, FAO, Rome, Italy 22-24 July 2008

AFFILIATIONS

Memberships

- American Geophysical Union (AGU), 2012-Present
- American Society for Civil Engineers (ASCE), 2011-Present
- American Society of Agricultural and Biological Engineers (ASABE), 2011-Present
- Soil and Water Conservation Society (SWCS), 2011-Present
- Order of Engineers and Architects, Beirut, Lebanon, 2008-Present

Leaderships

- Graduate Senator, Student Advisory Committee, Utah State University, 2010-2011
- Volunteer, Foi et Lumière (Handicap Care Community), Lebanon, 2006-2008
- Group Leader/Chief/Commissar, Cedars Scouts, Lebanon, 1996-2008
- Vice President/Treasurer, Student Representative Committee, Lebanese University, 2000-2004

SKILLS

Languages: Arabic (Native), English (Fluent), and French (Fluent)

Computer skills: ERDAS Imagine - ArcGIS - MATLAB - R Statistics - METRIC - Octave - SPSS Statistical Package - Visual Basic - AutoCad - Microsoft Office

INTERNATIONAL CONFERENCES, WORKSHOPS, SCIENTIFIC MEETINGS

American Geophysical Union Fall Meeting. San Francisco, California, 3-7 December 2012

Mapping Evapotranspiration at High Resolution using Internalized Calibration (METRIC) Workshop. Reno, Nevada, 19-23 March, 2012

Western State Remote Sensing of Evapotranspiration Workshop: Using ET to Reduce Cost and Improve Efficiency in Water Resources Management. Boise, Idaho, 12-13 October, 2011

Middle East Water and Livelihoods Initiative (WLI) Extension and Irrigation Training Workshop. Agriculture and Research and Education Center of the American University of Beirut, Lebanon, 26-30 June, 2011

Spring Runoff Conference: Water, Climate, and Sustainability of our Changing World. Logan, Utah, 29-30 March, 2011

10th International Meeting on Soils with Mediterranean Type of Climate. Beirut, 22-26 June 2009

Middle East Water and Livelihood Initiative, Rangeland Workshop. Amman, Jordan, 12-14 May 2009

Sustainable Orchard Irrigation for Improving Fruit Quality and Safety-IRRIQUAL (EU-FP6). Volos, Greece, 9-12 July 2008

Atelier de Formation en Agriculture de Conservation. El-Kef, Tunisia, 18-25 November 2007

Integrated Water Resources Management in the Arab Region. Beirut, Lebanon, 15-16 December 2006

The 4th International Water Conference in the Arab Countries: Rationalizing Water Use to Guarantee Sustainability. Beirut, Lebanon, 27-20 June 2005