

Toward a Model-Based Method for Gap Filling Latent and Sensible Heat Fluxes for a  
Semi-Arid Site

by

Andrew Lawrence Neal

---

A Thesis Submitted to the Faculty of the

DEPARTMENT OF HYDROLOGY AND WATER RESOURCES

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

WITH A MAJOR IN HYDROLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2008

## STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Andrew L. Neal

## APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

\_\_\_\_\_  
Hoshin V. Gupta  
Professor of Hydrology and Water Resources

\_\_\_\_April 23, 2008\_\_\_\_  
Date

## ACKNOWLEDGEMENTS

I would like to thank my thesis committee Paul Brooks, Jim Shuttleworth and my advisor, Hoshin Gupta, for their advice, expertise and vision while working on this project. Russ Scott at the US Department of Agriculture-Agricultural Research Service permitted use of his data and offered assistance in model development. Shirley Kurc and her students in the School of Natural Resources as well as other students in the Department of Hydrology and Water Resources provided a sounding board for this work as it coalesced. The Science and Technology Center for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA) provided financial support for this work. And for the most invaluable support, my sincerest thanks go to my family and friends, especially Rich Neal and Hayley Easton, for helping me keep focused on the important things.

## TABLE OF CONTENTS

ABSTRACT .....	5
1. INTRODUCTION .....	6
2. EDDY COVARIANCE THEORY AND THE NEED FOR GAP FILLING.....	11
3. PRIOR STUDIES .....	20
4. STUDY QUESTION, SITE AND METHODS.....	27
4.1 Study Site .....	27
4.2 Artificial Neural Network Model.....	29
4.3 Model Testing and Performance .....	32
5. RESULTS .....	36
6. DISCUSSION AND CONCLUSIONS .....	39
APPENDIX A: LAND SURFACE MODEL TESTS.....	44
A.1 Chameleon Surface Model (CHASM) .....	44
A.2 Calibration of CHASM using Shuffled Complex Evolution (SCE-UA) .....	46
A.3 CHASM Results.....	47
APPENDIX B. TABLES .....	50
APPENDIX C. FIGURES .....	53
REFERENCES .....	66

## ABSTRACT

The eddy covariance technique for measuring the exchange of mass and energy between the land surface and atmosphere yields data records with substantial gaps, reported to be as long as 30 to 40% of a time series annually (at a half-hourly time step). The application of these data sets in modeling studies as well as on varying time scales and under non-ideal conditions, requires some interpolation method to infer values for the missing data. This study will consider a neural network regression model for a flux record from a semi-arid riparian site and examine the model's responsiveness to variability in the data available for training. The neural network sensitivity to flux data used for training is evaluated. Model response worsened under reduced training data availability and was dependent on the characteristics of the data.

## 1. INTRODUCTION

The field of micrometeorology focuses on expanding the understanding of local-scale dynamic behavior of interactions between the land surface and the atmosphere. Precise measurement of land-atmosphere interactions has allowed researchers the ability to better understand the energy and mass exchange between the land surface, vegetation and the atmosphere. Among the benefits of micrometeorological techniques, the atmospheric and hydrologic sciences have benefited from improved quantification of the vertical transfer of moisture in the lower atmosphere and ecology has gained insight into the plant demands on water as well as the source/sink behavior of vegetation on atmospheric carbon dioxide. As these fields continue to progress toward a greater understanding of the complex relationships among vegetation, soil, atmospheric moisture, energy and carbon, more refined measurement and data processing techniques will be required to ensure that the data appropriately reflect the processes in question.

In the past twenty to thirty years, the eddy covariance method has become a standard technique used to measure local-scale energy and mass flux on a wide variety of landscapes (Baldocchi et al 2001). Eddy covariance provides a direct measurement of fluxes through separate measurements of the vertical wind speed and the concentration of the energy or chemical species in question. The method has enabled high-temporal-resolution (30-minute data derived from 10 Hz measurement frequency) of fluxes and has been employed in a wide array of geographic regions including agricultural lands, temperate deciduous forests, tropical rainforests and a range of arid and semi-arid landscapes (see Baldocchi et al 2001, Sun et al 2007 among others). The general

acceptance of eddy covariance as a valid method for measuring local scale energy and mass flux is evidenced by its broad use in the scientific community.

The main tools used in the eddy covariance technique are a three-dimensional anemometer to determine the vertical wind speed and a gas analyzer to measure carbon or moisture content at the same point as the anemometer. Both open- and closed-path gas analyzers produce acceptable results, though the open-path analyzers have been favored in recent studies for practical reasons (Leuning and Judd 1996). Because the measurements must be made at a high frequency (10Hz), the anemometer and gas analyzer must be highly sensitive devices. This sensitivity can lead to sensor malfunction and loss of data. The theoretical assumptions of eddy covariance require turbulent mixing of the lower atmosphere. If turbulence subsides, data may be eliminated from the record.

Due to the nature of the method and the equipment required to take measurements for eddy covariance analysis, the eddy covariance method is subject to significant gaps in the data record. One problem with these gaps is that the missing data is often serially correlated to particular events or periods of importance in the data set, e.g. extreme weather events and nighttime carbon exchange. In order to develop appropriate annual estimates of energy and mass fluxes, as well as to properly assess the diurnal variability of the fluxes, a method to approximate the missing data, and therefore fill the gap is crucial.

In some situations, gap filling can be as simple as applying average data from a similar portion of the record (Falge et al 2001). However, this simple gap filling technique inherently involves the application of an individual's conceptual model of the

flux behavior to the missing data, which can be highly subjective and requires a significant level of expertise in dealing with flux data. In the rest of this thesis, the term conceptual model will refer to an individual's understanding of the system and mental predictions made based on their knowledge of that system's function. A computational model will describe any scheme which explicitly states the relationships between components of a system. Computational models are often formulated as mathematical relationships drawn from an individual's conceptual model of a system. A statistical model is a relationship between two or more components of a system which is based purely on correlation without an explicit statement of the functional relationship between the components.

Other methods have been developed that employ computational or statistical techniques which eliminate the risks and variability associated with using a subjective conceptual model (see Moffat et al. 2007 for a complete listing of gap filling methodologies). These computational methods can provide more consistent gap-filling results because they are explicit models rather than subjective conceptual patterns. Such models can also be viewed as an explicit formulation of a relationship between two sets of variables (often between atmospheric forcing data and fluxes) rather than an implicit conceptual model. Clearly, computational models are not void of inherent disadvantages, namely the cost of development and implementation of a particular computational model for a given data set, including both time and monetary costs, and with the use of regression models, the risk of models with low physical correlation to the landscape processes.

Nevertheless, use of computational models as a means to produce complete, gap-filled data records has been established and in use for the past six years (Falge et al 2001). Data reanalysis in a model framework, as well as other, non-gap-filling studies involving model-based approaches to flux measurements, is dependent on the amount of data available and the richness, or degree of variability within the natural range, of that data.

Understanding that the eddy covariance method has become a commonly used technique for high-resolution, direct measurement of atmospheric fluxes, a need to improve the existing, limited data sets is essential to produce more comprehensive representations of fluxes. The goal of this study is to address the computational approach to gap-filling and the limitations that existing data records exert on the ability to produce reasonable, complete data records. With the understanding that a computational model is both cost-effective and methodologically robust in comparison to reliance on conceptual models, input data sets of varying length and data information content will be explored to consider the ability of a model to produce meaningful flux records from less-than-complete data.

This thesis will examine the limitations of flux gap-filling under varying constraints of input data for model calibration and training. These limitations were tested using both deterministic and regression (neural network) models, for a mesquite woodland in southern Arizona. This thesis will first address the underlying theory of eddy covariance technique, including the theoretical conditions that result in data gaps (chapter 2). The next section will review past studies of gap-filling techniques for flux

measurements (chapter 3). The fourth chapter will precisely describe the controlling question of the study, as well as describe the study site and methods used. Chapter five will present the results of the study and chapter six will discuss those results and the conclusions drawn as well as introduce future avenues for study on this topic.

## 2. EDDY COVARIANCE THEORY AND THE NEED FOR GAP FILLING

The eddy covariance (EC) technique has been widely used to quantify energy and mass fluxes in various environmental conditions (Baldocchi et al 2001, Sellers et al 1996, Fischer et al 2007). EC theory relies on high-frequency (10 Hz) measurements of atmospheric variables to produce robust statistical measures of flux. For both energy and CO<sub>2</sub> fluxes, EC can produce high temporal resolution measurements under ideal conditions. The EC method provides a consistent measurement technique that avoids the theoretical inconsistencies of other micrometeorological measurement techniques, such as the Bowen ratio method, which fails to produce measures of latent and sensible heat flux around dusk and dawn, when the ratio of the latent and sensible heat approaches infinity (Cook 2007). Other methods used to quantify fluxes are dependent on semi-empirical relationships (e.g. the Penman-Monteith equation) and so are not direct measurements of the flux process. Eddy covariance allows a direct measurement system that samples the atmospheric conditions at a high rate and yields stable results. This has led to the prominence of eddy covariance as a standard technique for flux measurement and to its widespread use at research sites worldwide.

The benefits of using the EC method are substantial in light of the measurement needs of the scientific community. As noted above the method provides a high temporal resolution of land-atmosphere interactions. Further developments by Finnigan (2004) and Foken and others (2006) have yielded a more complete theoretical understanding of the source area for eddy covariance measurements. Thus, the data collected by flux tower

measurements can be associated with specific regions in the landscape. At a broader level, the cooperative flux measurement networks (Ameriflux, FLUXNET, etc.) have developed standardized practices for eddy covariance measurements, allowing greater comparison across climatological and environmental regions.

The benefits of the EC method—improved temporal resolution of fluxes, better defined source area, methodological consistency—are substantial. However, when used for broader research purposes, complications of eddy covariance flux measurements can prove problematic. The systematic lack of measurements during nighttime and periods of intense storms suggests that, on a seasonal and annual scale, flux data sets collected using the eddy covariance method may provide an incomplete picture of the processes at work, thus leading to incorrect conclusions about the nature of energy and mass exchange. The causes of EC data loss are described below.

By looking at the theoretical underpinnings of EC, we can better understand the loss of data and where and when data must be gap filled. Assuming conservation, the fluxes of some scalar  $q$  in and out of a control volume are balanced, resulting in the following expression:

$$\frac{dq}{dt} = \frac{\partial q}{\partial t} + \left[ u \frac{\partial q}{\partial x} + v \frac{\partial q}{\partial y} + w \frac{\partial q}{\partial z} \right] = v_q \left[ \frac{\partial^2 q}{\partial x^2} + \frac{\partial^2 q}{\partial y^2} + \frac{\partial^2 q}{\partial z^2} \right] + S_q \quad (1)$$

where  $\frac{dq}{dt}$  represents the total change in the property  $q$ ,  $\frac{\partial q}{\partial t}$  represents the local change

in  $q$ ,  $u \frac{\partial q}{\partial x} + v \frac{\partial q}{\partial y} + w \frac{\partial q}{\partial z}$  represents the advection of  $q$  in the x-, y- and z-coordinates,

respectively,  $v_q \left[ \frac{\partial^2 q}{\partial x^2} + \frac{\partial^2 q}{\partial y^2} + \frac{\partial^2 q}{\partial z^2} \right]$  represents the molecular flow (diffusion) of  $q$  and

$S_q$  is the storage of  $q$  in a source or sink.

If the velocities in the  $x$ ,  $y$ , and  $z$  – direction (i.e.  $u$ ,  $v$ ,  $w$ ), and  $q$  are subject to turbulent motion, the variables above can be broken down into a mean value and a turbulent fluctuation, e.g.:

$$u(t) = \bar{u} + u' \quad (2)$$

Next, the eddy covariance is derived by applying a Reynolds' average of total flux over a unit area. Performing this Reynolds averaging, equation 1 can be rewritten as:

$$\frac{\partial \bar{q}}{\partial t} + \left[ \bar{u} \frac{\partial \bar{q}}{\partial x} + \bar{v} \frac{\partial \bar{q}}{\partial y} + \bar{w} \frac{\partial \bar{q}}{\partial z} \right] + \left[ \frac{\partial(\overline{u'q'})}{\partial x} + \frac{\partial(\overline{v'q'})}{\partial y} + \frac{\partial(\overline{w'q'})}{\partial z} \right] = v_q \left[ \frac{\partial^2 q}{\partial x^2} + \frac{\partial^2 q}{\partial y^2} + \frac{\partial^2 q}{\partial z^2} \right] + S_q \quad (3)$$

From equation 3, further simplifications can be made to reduce the number of terms, and therefore eliminate the complexity that would otherwise make characterizing this system difficult. These simplifications are based on assumptions which include the elimination of the source/sink term,  $S_q$ , no horizontal turbulent divergence, no ascent or subsidence of air locally, and no horizontal advection (all assumed in equation 3). Under these conditions, equation 3 reduces to:

$$\frac{\partial \bar{q}}{\partial t} = v_q \left[ \frac{\partial^2 q}{\partial x^2} + \frac{\partial^2 q}{\partial y^2} + \frac{\partial^2 q}{\partial z^2} \right] - \frac{\partial(\overline{w'q'})}{\partial z} \quad (4)$$

Comparing molecular diffusion to turbulent diffusion, we can expect the molecular diffusion to be negligible, further reducing equation 4 to:

$$\frac{\partial \bar{q}}{\partial t} = - \frac{\partial (\overline{w'q'})}{\partial z} \quad (5)$$

This equation is the basis for eddy covariance measurements. Using equation 5, the flux is the average of the vertical wind speed and vertical change in concentration under turbulent behavior.

In order to develop an accurate depiction of the flux behavior, high frequency measurements are necessary so that a large enough number of data points are available for averaging (Massman and Lee 2002). The frequency of measurements (usually 10Hz) demands the use of highly sensitive instruments such as three-dimensional sonic anemometers, high frequency gas analyzers, and radiometers to produce synchronous and detailed records of the mass and energy fluxes. The high sensitivity of these sensors can result in malfunctioning and failure, especially under extreme weather conditions (Richardson and Hollinger 2007, Hui et al 2004), and thus result in missing flux values in the time-series.

Due to the various assumptions made to reach the simplified basis for the eddy covariance techniques, flux records often contain missing values (Falge et al 2001, Moffat et al 2007), for example, as previously noted, equation 5 carries the assumption of no ascent or subsidence of air. This assumption breaks down at night when the atmospheric boundary layer becomes stable near the surface and vertical turbulent eddies dissipate (Goulden et al 1996, Massman and Lee 2002, Acevedo et al 2007). As these

turbulent eddies break down, the subsidence  $\overline{w} \frac{\partial \overline{q}}{\partial z}$  may become significant, while the covariance term  $\frac{\partial(\overline{w'q'})}{\partial z}$  fails as  $w'$  goes to zero. Thus, under conditions of nighttime subsidence many flux tower sites fail to produce meaningful values and the data are often removed from the record.

Similarly, the assumption of no horizontal advection can result in complications and inaccuracy in flux values obtained using the eddy covariance method (Baldocchi 2000, Finnigan 2004, Katul et al 2006). Proper site selection for eddy covariance flux measurements is critical to avoid this complication, as flat sites with uniform vegetation cover will generally follow this requirement. However, not all flux towers are located on ideal sites and the mass and energy fluxes made at the margins of various vegetation types or on non-level surfaces are increasingly becoming important areas of research (Baldocchi et al 2004, Brown-Mitic et al 2007). Unfortunately, these non-ideal sites may produce either over-estimates or under-estimates of flux values as a result of local advection toward or away from the flux tower. These flawed data are often eliminated from a flux record under quality control procedures, further reducing the total length of the flux data record.

The elimination of data points from the record raises a number of concerns about the meaning of long term flux measurement data sets. Because many of the missing data points occur during nighttime, many studies ignore flux behavior at a time when the energy and mass fluxes may be considerably active and more complicated due to three-

dimensional flow patterns (Van Gorsel et al 2007, Goulden et al 2006). At nighttime, temperatures drop, relative humidity rises and plants undergo dark photosynthetic reactions, all of which contribute to the energy and carbon fluxes, and all of which may go unrecorded when using an eddy covariance system.

Further, the physical failure of the measurement equipment may occur not only at night but also during periods of intense storms or strong winds. At these times, the pattern of energy exchange may vary significantly from periods of relatively calm atmospheric conditions, which may be of greater interest for a particular study and are also important in characterizing the full variability of energy exchange over an entire season, a year or a longer time period. If these inclement conditions result in the failure of the field equipment, then the flux data record may be missing an event which is critical to the overall characterization of the energy balance.

As discussed above, the eddy covariance method can fail to produce continuous viable data at all points in a half-hourly record over a long period (e.g. Richardson and Hollinger 2007 for a discussion of long gaps). The analysis of eddy covariance flux data sets at these longer time scales requires a more complete record for assurance that proper and valid conclusions are reached. The motivation for developing a complete record is manifold, including the ability to characterize and generalize the study site, the use of these records in comparison to other sites, resolution of the total annual energy balance, a flaw which has been identified by numerous studies (Foken 2006, Cava et al 2006, Twine et al 2000), as well as the development and refinement of flux models for the site in question.

The overall characterization of a site is dependent upon a full understanding of the variability of fluxes across a range of temporal scales. This requires a data set that includes accurate measurements across seasonal and possibly annual variations in energy and mass exchange. If large portions of the record are missing, especially at times of notable meteorological conditions, then extensive analyses may not be possible or may lead to erroneous conclusions. However, if the flux record is complete and reliable, then the validity of the conclusions can be considered dependent on the quality of the flux measurements rather than on the continuity of the flux record itself. Upon resolving difficulties with the completeness of a flux record, the patterns and trends that emerge from an analysis may be generalized to other, similar sites. The ultimate goal of developing a deeper understanding of flux processes at broader scales both temporally and spatially can be achieved when thorough records of micrometeorological variables exist (e.g. Henderson-Sellers et al 1993).

With the expanded use of eddy covariance techniques from simpler, small-scale projects to numerous sites, the scientific community has pushed for studies comparing flux behavior among different climate and vegetation types. Typified by the AMERIFLUX and EUROFLUX projects (Baldocchi et al 2001, Aubinet et al 2000, Valentini et al 2000), the comparison of fluxes across sites is dependent on a valid and accurate assessment at each individual site. Comparisons require data records at each site that are comprehensive such that they can fully depict the fluxes over long periods of time. Further, the development of complete flux records should improve methods and models used to depict the interaction between forcing variables and fluxes. Land surface

models (LSMs) simulate the resulting mass and energy fluxes through the interactions among atmospheric and vegetation characteristics and variables, and have been used as an explicit, physically-based method of calculating fluxes based on variables that are more easily measured (e.g. Sellers et al 1996). Their development was initially spurred by the need for energy exchange with the surface in general circulation models (GCMs) (see Shuttleworth 2007 for discussion). However, since that time, the models have been used independent of GCMs to simulate small-scale mass and energy exchange (e.g. Henderson-Sellers et al 1993).

By employing a complete record of fluxes, LSMs can be more effectively calibrated to a particular site so as to best reflect the relationships at work on that site. A stumbling block in model identification with respect to the use of LSMs in flux modeling is the lack of high-quality long term records of flux variables for comparison to model results (Friend et al 2007). The development of complete records according to a theoretically defensible methodology could allow for the improvement of deterministic process models by using these full-length time series for model calibration. Of course, it is important to recognize the caveat that these filled records are not entirely composed of measured values and that their use as a standard of comparison is one step further removed from the direct quantification of the fluxes.

Working from this understanding that gap-filled records can be used to provide more information to flux models, an examination of gap-filling techniques may allow users to better understand the data requirements to produce a complete flux record. This objective, while perhaps not explicitly stated, has been explored by numerous research

initiatives undertaken in the recent past (Falge et al 2001a, Falge et al 2001b, Moffat et al 2007, and citations therein ) In various attempts to produce theoretically substantiated methods for producing gap-filled flux records, many of the methods employed used either statistical methods or models, both stochastic and deterministic, to simulate and interpolate missing data from the existing record ( Falge et al 2001, Moffat et al 2007).

The purpose of this study is to use a stochastic model framework to examine the ability of model-based gap filling techniques to produce a theoretically defensible and meaningful time series of flux. This was done in light of the often limited availability and variable quality of measured flux observations. Considerations include the length of existing data record available for study as well as the richness and variability of that record. Specifically, this study will look at how the model responds to reductions in the available flux data. The expectation is that the elimination of flux data used for training will hinder the model's ability to accurately recreate the flux record, implying that the modeled data during gaps in the flux may not be appropriate as estimates of the flux during those times.

### 3. PRIOR STUDIES

The examination of data limitations in flux measurement campaigns was largely begun by Falge et al (2001a, b) to address the problem of incomplete data records in energy and carbon flux in the EUROFLUX and AMERIFLUX campaigns. Much of the work on gap-filling flux records has been done with the goal of improving estimates of carbon balance and exchange in order to better quantify the source/sink properties of the various sites studied in these global flux experiments. The authors (Baldocchi et al 1999, Aubinet et al 2000) generally note that the mission of these flux intercomparison projects is complicated by the problem of time series incompleteness, pointing out that flux records may have as many as 35% of their data points lost or eliminated due to measurement failure or quality control. At this point in time, no clear method has emerged as preferable or superior to others and the use of individual techniques is done at the discretion of the individual researcher.

The selection of an individual technique is dependent on the needs of each researcher; if the work is focused on half-hourly data then the considerations in gap-filling may be different than those for seasonal or annual measurements of fluxes due to the coarse- or fine-scale required for each analysis. At the seasonal and annual scales, mean diurnal variation have been used to interpolate missing data by assigning average values for a point in the time series taken over a defined period on the existing record. This method had been used sporadically before the Falge et al work was published (Baldocchi 1999, Foken and Wichura 1996) with some success at these longer time

scales. The Falge et al (2001) paper was the first to directly address the use of this technique for half-hourly data.

Moffat et al. (2007) found mean diurnal variation gap filling to relatively underperform compared to other methods at the half-hourly time scale, especially under scenarios where gaps are long. This poor performance under long gaps is likely due to the lack of information that exists when trying to develop mean values at gaps longer than seven days. However the authors state that the method produced good results and should not be eliminated as a method for future use. Because the mean diurnal variation method does not include information about atmospheric conditions, the method is simple to develop and operate but this may also be a causal factor in the reduced predictive ability at longer gap lengths.

In addition to the mean diurnal variation method, Falge et al. (2001a, b) tested a method termed “lookup tables”. Lookup tables are simply a set of average values for fluxes correlated to specific conditions for environmental variables. Thus, for any combination of the forcing variables, the average value of a flux would be used to replace the missing data point. This method may lead to serial over- or underestimation if there are many missing data points in a period that is known to have far different forcing conditions than otherwise exist in the dataset. As with the mean diurnal method, lookup tables are simple to develop and implement. However, they are strongly subjective, based on the user-identified forcing variables and ranges applied.

A third method used in the Falge et al (2001a) study was non-linear regression. These non-linear regressions often take the form of simple functional relationships

between the forcing and flux variables. Various researchers (Falge et al 2001, Ruppert et al 2006, etc.) have used combinations of equations for carbon and energy fluxes based on previously identified relationships for nighttime and daytime process behavior. These equations have included the Lloyd-Taylor, Arrhenius and Johnson equations for nocturnal fluxes based on air and soil temperatures, and the Michaelis-Menten relationships for daytime fluxes (see Lloyd and Taylor 1994, Michaelis and Menten 1913, Dagniele 1991 for derivations, Hollinger et al 1994, Pilegaard et al 2001, Falge et al 2001a for use in gap filling). While these relationships all focus on gap-filling for missing carbon fluxes, analogous techniques could be used for energy fluxes using, for example, the Penman equation or other non-linear relationships that relate the forcing variables to the fluxes in question. The Falge et al paper addressing energy fluxes (2001b) considers only mean diurnal variation and lookup table methods for filling flux records.

Ultimately, the discussion provided by Falge et al (2001a, b) equivocates on the comparative advantages and disadvantages of the methods of gap-filling tested. The authors note that the accuracy of gap-filling is largely dependent on the initial data quality and treatment. Since that the Falge et al publication, more extensive work has been done to compare those and other novel gap filling techniques and to move toward greater standardization of gap-filling methods to promote intercomparison of complete, filled records of fluxes among a broad array of sites (most notably by Moffat et al 2007). At the time of writing, however, no clear consensus has been reached, though many groups seem to favor lookup methods based on correlations between forcing variables

over explicit models or regression. Moffat et al (2007) suggest that the flexibility of neural networks may lead to better gap filling results in arid and semi-arid conditions where (in their case carbon flux) relationships may differ from humid locations.

Because most of the past work on gap-filling has focused on ecological studies, many of the individual linear models employed in the published literature are focused on estimating net ecosystem exchange (NEE), but general correlations can be made to similar model equations for moisture and energy transfer (e.g. Penman equation). One example of a simple model used to depict NEE is the Michaelis-Menten equation, mentioned above, which was used by Gove and Hollinger (2006) within a Kalman filter parameter estimation framework. The authors note that the more popular lookup table and moving average methods are dependent on a user-defined monthly or seasonal pattern that may skew gap-fill results. This fact coupled with the non-normality of flux measurements (Richardson and Hollinger 2005) suggests that a model calibration technique like a Kalman filter, which can be built without the assumption of a normal distribution of variables, may be better suited to properly identify a model for use in gap-filling estimation. The use of an explicit relationship between forcing variables and fluxes may produce a more appropriate value for interpolated flux values, especially when coupled with a parameter estimation method that is better suited to the true distribution of fluxes (e.g. Markov chain Monte Carlo methods).

Other research groups have developed non-linear regression techniques for gap filling, including the use of Artificial Neural Networks (Papale and Valentini 2003) and non-linear physical models to address missing data (Alavi et al 2006). Papale and

Valentini (2003) first introduced the use of artificial neural networks (ANNs) to the problem of gap filling for flux data. Though their study is used to consider carbon and not energy fluxes, it is worthwhile to consider the course of their study, since the function of an ANN can be easily adapted to fit any input-output relationship (see Model section for more detail). The Papale and Valentini (2003) study used ANNs both for temporal gap-filling as well as to generate values for net ecosystem exchange (NEE) based on spatial variables, which were subsequently compared to NEE estimates from the gap-filled records at sixteen flux tower sites in generally humid areas of Europe. Their results indicate the functional capacity of ANNs to identify input-output patterns in flux behaviors for carbon fluxes, a characteristic that may also be applicable to energy fluxes.

The use of ANNs to represent flux processes was also explored by Ooba et al (2006). Their model uses a genetic algorithm to select values for the parameters used in the ANN, including the controls on input data. Since their genetic algorithm included a component to select input data variables, Ooba et al (2006) were able to select which variables were relatively more important to the full scale modeling of carbon fluxes. Both their genetic algorithm neural network and ANN performed comparably to non-linear regression equations when modeling existing data as well as providing energy balance closure when including gap-filled data. The use of a genetic algorithm also allowed Ooba et al to determine the relative significance of their input variables based on the frequency of the selection of those variables into their genetic neural network.

While much of the work on flux data gap-filling has been focused on carbon fluxes, especially the use of non-linear regression and physically-derived models, there

have been some studies addressing gap-filling for energy fluxes. Alavi et al (2006) considered gap-filling for evapotranspiration on a humid site near Guelph, Ontario, in a similar approach to the Falge et al (2001a, b) studies. They tested the efficacy of a variety of methods, including mean diurnal variation, linear multiple regression, an averaged method based on the Priestley-Taylor equation and a linear regression based on the Penman-Monteith equation. Their linear regression took the form:

$$\lambda E = \alpha(Rn - G) + \beta D + \zeta \quad (6)$$

where,  $\lambda E$  is the latent heat,  $Rn$  is the net radiation,  $G$  is the ground heat flux,  $D$  is the vapor pressure deficit of the atmosphere, and  $\alpha$ ,  $\beta$ , and  $\zeta$  are parameters of the regression. The parameters of equation 1 were then identified using a Kalman filter to replace missing data. In their comparison, Alavi et al found that the Kalman filter-trained linear regression model outperformed the other methods in total error on existing data points. No measure was provided to indicate total energy balance. This study may have demonstrated the performance improvement suggested by Gove and Hollinger (2006) but no indication of this is made by the authors.

Moffat et al (2007), extending the work of Falge et al (2001), give an extensive discussion of the numerous methods that have been used for gap filling in past work (lookup tables, mean diurnal variation, marginal distribution sampling, non-linear regression and ANNs) and compare the results of each method in replicating datasets from the EUROFLUX measurement campaign. Their results suggest that nonlinear, lookup and marginal distribution methods perform slightly better than other methods tested, though all models produced similar results. The authors note that the tests were

conducted using datasets from humid, forested regions and that similar studies in other regions would do well to corroborate their results.

These prior studies have all examined the various techniques available for gap-filling. Here the more significant question at issue is the limitation to model performance presented by the available data. After determining a model which is appropriate for testing, that model will be iterated using training data of various length and richness to determine what compromises a sufficient flux record for gap-filling methods.

#### 4. STUDY QUESTION, SITE AND METHODS

The eddy covariance technique for flux measurement produces an accurate, high-resolution record of atmospheric fluxes; however, those measurements are subject to elimination due to an array of quality-control factors. Using computational models to interpolate missing data, we intend to examine how poor data richness and variable time series length affect model training. From this, we intend to show how model-based methods can be used to produce useful gap-filled flux records and the minimum requirements for flux data records in order to produce such results. This analysis will be performed using artificial neural networks (ANNs) trained on data sets of varying length and including different phases of flux intensity (i.e. summer and winter periods). Subsequently, the model will be tested under conditions of reduced flux data for training to identify model performance under limited data availability. This study was initially formulated to include analysis using a simple land-surface model (LSM) for comparison with the ANN results. However, due to concerns about the implementation of the LSM, those results were not included with the main analysis, but, for completeness, are included in Appendix A.

##### 4.1 Study Site

The data used in this study were collected at the Charleston Mesquite Woodland site near Sierra Vista, AZ (31° 39' 49"N, 110° 10' 40"W, 1200m elev.) by the United States Department of Agriculture-Agricultural Research Service (USDA-ARS). The site is composed mostly of velvet mesquite (*Prosopis velutina*), with ground plant

assemblages of sacaton grass (*Sporobolus wrightii*), greythorn shrub (*Zizyphus obtusifolia*) and other herbaceous plants (Scott et al. 2006). The mean canopy height is approximately 7 m and the maximum canopy height is approximately 10 m. Measurement height was 14m for sonic wind speed and gas concentrations, used to calculate the energy fluxes. These components were measured using a CSAT3 sonic anemometer (Campbell Scientific, Inc., Logan, UT) and a Li-7500 infrared gas analyzer (Li-Cor, Inc., Lincoln, NE) The meteorological forcing data used in this study (precipitation, wind speed, relative humidity and temperature) were collected at 13.5m and solar radiation was collected using a 4-component radiometer at 9m height (Scott et al 2006). Precipitation was measured with a tipping bucket rain gauge, wind speed and direction with a wind vane/anemometer (R.M. Young, Inc., Traverse City, MI), relative humidity and temperature with a HMP35D probe (Vaisala, Helsinki, Finland). Four-way radiation was detected with a CNR1 radiometer (Kipp & Zonen, Delft, The Netherlands). Data were collected using a CR5000 data logger (Campbell Scientific).

Data were aggregated to 30 minute intervals for the period from 2001 to 2003. The total length of the measurement period is 48454 30-minute periods. Data exist for 29622 intervals, representing approximately 61% of the period from the start of data collection (3/27/2001) to the close of data collection (12/31/2003). Gaps in the flux record ranged from individual half-hourly points to multiple months. The most complete record is for the calendar year 2003, which was missing 8% of its data and was thus selected for use in model training. See Figure 1 for plots of the forcing variables, Figure 2 for the fluxes and note the extended periods without data collection.

In terms of climate, the site is semi-arid with mean summer high temperatures of 34°C and mean annual high temperatures of 25°C. The mean annual precipitation is 358mm. These values are for the Tombstone, AZ weather station, located 16km from the study site, and were collected from 1971 to 2000 (Western Regional Climate Center <http://www.wrcc.dri.edu>). Precipitation is largely bimodal, with 60% of rainfall occurring between July and September, with a lesser rainy season between December and March.

Other local meteorological data associated with the Charleston Mesquite site was used to augment the existing forcing data. Since the meteorological data at the tower site had gaps similar to those in the flux record (due to problems with power and tower operations), this data from a nearby site (25km) was considered to substitute for the tower meteorological data in model runs. This record is shown in Figure 3 and the correlation statistics for the variables are shown in Table 1, indicating that most of the variables are strongly correlated to those from the flux tower. In this way a complete forcing record was assembled and applied to the Charleston Mesquite Woodland site. This alternative forcing record, which included a complete time series for meteorological variables, was used in this study as input to the models.

#### 4.2 Artificial Neural Network Model

A non-linear regression ANN technique can provide a simple relationship between inputs and outputs without explicitly defining that relationship as a mechanistic process. This results in a simple model that requires less computing time for model training and operation, when compared to explicit physical models. However, the model

itself contains no information about the physical process between the inputs and outputs, only a correlative relationship based on a series of linear and non-linear equations.

Because of the simple nature of Artificial Neural Network (ANN) structure and their ability to generate otherwise unclear relationships between input and output data, ANNs have been used in a wide variety of fields for pattern identification, including medicine, engineering as well as the physical sciences (Willis et al. 1991, Baxt 1995, Gasteiger and Zupan 1993). Neural networks have been suggested in hydrology for use in identifying spatial and temporal patterns in data (Hsu et al 1997, Hsu et al 1999), as well as for generating relationships for data extrapolation (e.g. Ooba et al 2003). In hydrology, the most often used ANN structure is a multi-layer perceptron (MLP), generally taking the form of a three-layer, feed-forward network (see Figure 4 for a diagram of such a network).

A three-layer, feed forward network consists of three levels of regression equations to build a relationship between the desired inputs and outputs. The first layer generates a set of linear combinations of the inputs and a random noise term. This layer transforms the dimensionality of the problem from  $n+1$ , where  $n$  is the number of input variables, to  $m$ , where  $m$  is the number of intermediate nodes. The second layer transforms those intermediate nodes based on a non-linear function. In this study the non-linear function is a softmax function of the form:

$$y_i = \exp \frac{(a_j)}{(\sum \exp(a_j'))} \quad (7)$$

The third layer is a linear combination based on the previous layer, i.e. a linear combination of the non-linear function values. The linear combination of the intermediate nodes superposes the nonlinear components, using that superposition to approximate the non-linear behavior of the system. This third layer transforms the dimensionality from  $m$  the number of intermediate nodes, to  $o$ , the number of outputs.

The fact that the model is known as feed-forward simply refers to the fact that the model is operated in such a way that all of the calculations occur moving from the inputs to the outputs. No recursive calculations are made and no state variable exists to store information about the previous conditions. The relationship exists strictly at that point in the time series.

In order for an ANN to produce a valid input-output relationship, the regression parameters must be fit to the data. This is done using a process called training, which iteratively takes an error measurement between the model outputs and a given measured output for comparison. In this study, the model was trained using a scaled conjugate gradient (SCG) method. The training is done by operating the model for one entire period of the input data, determining the error on the outputs, and using that error value to produce a scaling factor used to control the magnitude of the change in the parameter values. This is done to prevent the model from taking too broad a sample across the parameter space, which could preclude model convergence. Care must also be taken during model training to avoid overtraining the ANN. Since neural networks are based on statistical relationships among variables, a lengthy training period (i.e. too many iterations) may result in the model being trained on the variability within the output data,

rather than on the intended trend (Nabney 2003). This training length is largely subjective and requires the user's care to ensure that overtraining has not occurred.

The ANN model structure used in this study was developed in the Netlab toolbox (Nabney 2003). This toolbox works within the MATLAB programming environment to generate a variety of ANN structures. Since the ANN here is intended to be a proxy for the LSM (see Appendix A), a similar input and output structure was used, employing the same forcing and flux variables. Thus, the resulting MLP included six inputs (precipitation, relative humidity, air pressure, wind speed, net radiation and temperature), two outputs (latent and sensible heat flux) and 13 intermediate nodes and was trained for 300 iterations.

#### 4.3 Model Testing and Performance

The models described above were tested to ascertain their performance with varying data lengths and richness. This was done by first testing for data length with existing data richness. Model calibration/training data were selected by length, ranging from seven days (336 half-hourly data points) to 365 days (17520 points) and were then applied to the models in the following intervals: 7, 14, 30, 60, 90, 180, 365 days. These tests were done on the period starting from January 1, 2003 to December 17, 2003 which comprised the longest continuous segment of original data with minimal gaps.

After training, the models run using the entire data set as input. These trials were performed to identify a length threshold for model performance, which was measured using mean actual error, energy balance closure and regression of modeled and observed fluxes. The mean actual error was calculated as:

$$MAE = \frac{\sum |y_n^{obs} - y_n^{mod}|}{n} \quad (8)$$

where  $y_n^{obs}$  is the  $n$ th value of the observed data,  $y_n^{mod}$  is the  $n$ th data of the modeled record and  $n$  is the length of both datasets. Ideally, the mean actual error should approach zero, indicating good model correlation to the measured data, and suggesting that the continuous model record are continuing the same trend during the periods where measured values are absent. Mean actual error was used in light of the description of flux measurement errors described by Hollinger and Richardson (2005). The total energy balance closure was calculated on an annual basis as the residual of net radiation and latent and sensible heat fluxes. Because, on an annual basis, the ground heat flux can be assumed to trend to zero, total energy balance should be divided among these three components and the residual should approach zero. The linear regression analyses were done as a comparison between the modeled and measured data and were checked for the slope, intercept and coefficient of determination of the regression. Optimally, the regression analysis should produce a linear, one-to-one fit between the modeled and observed data with a coefficient of determination approaching one.

After obtaining a better understanding of the time limitations on the data, trials were conducted to examine the effects of richness and total data available on model performance. This was done by successive trials with varying amounts of data points removed from the original training data set. Random data points were selected from a training data set and removed, thereby increasing the number of gaps and reducing the richness of the data used to train the models.

When training the neural network models, it was found that the distribution of the data was non-normal. The distribution of the latent heat appears to be a gamma distribution (see Figure 4). Working from this distribution, a normalization procedure was used, similar to that of the Standardized Precipitation Index (McKee et al 1993, Mishra and Desai 2006), to convert this gamma distribution into a normal distribution. This was done by developing a cumulative probability distribution according to a gamma function (using the `gampdf` and `gamcdf` commands in MATLAB-Statistical Toolbox) using flux data that had been shifted to accommodate flux values equal to or less than zero. Once these parameters were found, the index was calculated as follows:

$$\begin{aligned}
 Z &= -\left( t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad \text{for } 0 < H(x) \leq 0.5 \\
 Z &= \left( t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad \text{for } 0.5 < H(x) < 1
 \end{aligned} \tag{9}$$

where,  $Z$  is the standardized index. The term,  $t$ , is calculated as:

$$\begin{aligned}
 t &= \sqrt{\ln \left[ \frac{1}{(H(x))^2} \right]} \quad \text{for } 0 < H(x) \leq 0.5 \\
 t &= \sqrt{\ln \left[ \frac{1}{(H(x))^2} \right]} \quad \text{for } 0.5 < H(x) < 1
 \end{aligned} \tag{10}$$

and the coefficients are:

$$\begin{aligned}
 c_0 &= 2.515517 & c_1 &= 0.802853 & c_2 &= 0.010308 \\
 d_1 &= 1.432788 & d_2 &= 0.189269 & d_3 &= 0.001308
 \end{aligned}$$

This standardized evapotranspiration index (SEI) was then used as a model output in place of latent heat. The ANN model was trained to values of SEI which were then converted to latent heat based on a third-order polynomial fit to the relationship

This new standardization index, employed here for latent heat, was used in the ANN models to improve the ability of the model to fit the data. This normalization procedure was required because the underlying structure of ANN models assumes that the inputs and outputs follow a normal distribution and that the error function applied to the model is also normal. The conversion from gamma to a normalized distribution, should improve the ability of the model to fit the flux data, as will be discussed later.

## 5. RESULTS

The ANN model runs perform well when comparing the measured and observed flux data (see Figure 6). The ANN model was constructed as an analogue to the CHASM model (i.e. taking the same inputs and outputs, see Appendix A) and was found to have a notably improved performance (see Table 2 for error measures). The ANN model run shown in Figure 6 was performed without the normalization index for latent heat but with the assumption that the distribution for both latent and sensible heat fluxes follows a Gaussian distribution. The model was trained using 300 days of available data.

It was at this point in the analysis that the need for some normalization procedure was necessary to improve the ANN model results. The Standardized Precipitation Index of McKee et al (1993) was adapted for this purpose, yielding a standardized heat index for latent heat. The distribution and fit of the latent heat data is shown in Figure 4. This fit was then used to generate a relationship between the standardized heat index and the measured latent heat (Figure 8). In subsequent ANN model runs, the latent heat data was trained to the standardized heat index and an equation fit to the relationship in Figure 8 was used to convert that index to a value of latent heat.

Using this standardized heat index, the ANN model was re-run and outputs were generated. These new model outputs, shown in Figure 8, suggest improved model function for the ANN using the standardization index to train the model in place of the (non-normal) latent heat flux data. This improvement in the latent heat flux also may have improved the model performance on sensible heat flux, since the model training

period does not require as much time to be spent improving the model fit on the latent heat data (see Table 3).

Subsequent trials were performed with the ANN using the standardization index to examine the sensitivity of the model to limitations of training data. The model was tested for data length as described in the previous chapter. These test results are shown in Figure 9. Looking at the errors for these trials (Table 3) the improvement in model fit with longer training periods is best seen in the value of  $R^2$  between measured and model fluxes, as most of the improvement at longer training times is found in values of  $R^2$  rather than in the MAE or energy balance. Scatter plots for these trials visually confirm that a longer training data set produces a model result that is closer to a one-to-one fit with the measured values (Figure 10). Because of the similar values in  $R^2$  at longer training times, confirmation of a one-to-one relationship between modeled and measured fluxes may be necessary to demonstrate good model fit.

The above trials suggest that at least one full year of data is required to ensure that the full range of behavior is captured during the model training and thus can be replicated by the model. Short duration training periods appear to result in a model that is only able to replicate flux values within the range of flux values in the training period (i.e. higher observed flux values that are outside the training period are modeled at or near the maximum value within the training period). This finding was replicated on training periods for relatively low latent heat flux values (winter, see Figure 11a) as well as for periods of high latent heat flux (summer, Figure 11b). Even for a 180-day training period, the model relationship constructed from the mostly winter period gives different results

than the model of the summer period. Figure 11 highlights the need for careful training of neural network models since training on different segments of the data record can produce widely varying patterns in the model output. Models trained on longer data sets approached better results, but were unable to model the overall trend in the measured data as the full-year training period.

After the tests for training data length, total data availability was considered. Data points were randomly removed during a 360-day training period. Tests were performed for total data availability of 70%, 80%, 90%, 95% and 100%. These percentages are taken with respect to the available data in the original record for 2003, i.e. 100% data availability is the full set as recorded with 8% of the data missing. The results for a series of tests for data availability are summarized in Table 3 and are shown in Figure 12. Figure 12a shows the model output under full data availability, while Figure 12d shows the data with up to 70% of the data removed. As could be expected, the performance of the model, based on visual inspection of the data fit, suggests that the model performance decays with less data available. In a scatter plot of these trials (Figure 13), the decay of model performance is particularly clear, with a marked shift from the one-to-one relationship desired when data is unavailable to the model. According to the error measures, however, the trials with less data show similar error as tests with the full data and may actually show improved closure of the energy budget.

## 6. DISCUSSION AND CONCLUSIONS

The ANN tests suggest that this method is viable for data gap filling, in agreement with other studies (Papale 2003). Provided that the data is properly treated before being used as input to the ANN, the resulting output can be considered a reasonable approximation for the measured fluxes. This model output can be used to develop monthly, seasonal and annual scale values of fluxes as well as being used as *a priori* estimates of flux values to apply confidence values to measured fluxes.

As is noted above, proper treatment of the data may include developing a normalization scheme for the data in question. Here, the data was found to follow a gamma distribution. A normalization scheme similar to that of McKee (1993) was used to convert this distribution to a normal distribution, so that it would fit with the assumptions of normality that underlie the formulation of ANN models. The use of this normalization scheme was demonstrated to improve the predictive ability of the ANN model with respect to latent heat.

The finding that the distribution of latent heat at this site was not normally distributed is something that has not been noted in previous work. Prior studies using neural network models give little discussion to the underlying distribution of their flux data (Papale and Valentini 2003, Moffat et al. 2007) as with other eddy covariance studies in semi-arid regions (Scott et al 2004). This finding should inform further work with flux data from similar regions, especially when used in land-atmosphere model studies, where the choice of model error estimate may alter the parameters identified by a calibration scheme and thus skew the results of that model.

The ANN model in general yielded satisfactory results as a gap-filling scheme. Error estimates indicate an appropriate degree of representativeness in the model output when compared against the measured fluxes. Visual inspection of the ANN-generated flux time series indicates adequate representation of both peak and low-flux events in both sensible and latent heat. Underestimation of some extreme events does occur and may be considered an effect of the normalization process on the signal. These extreme events, though rare, generally coincide with periods of atmospheric turbulence and high  $u$ -star values, such that gap-filled data may not be necessary to replicate those events and the original data can be considered valid.

Gap filling using ANNs appeared to be sensitive to the length of data applied for model training. The range of values for fluxes in the training period seems to be reflected in the range of values generated by the ANN. If gap filling were attempted with a record that does not include the range of possible flux response on the site, then the use of ANNs as gap filling models may not adequately replicate certain sections of the data where the fluxes are above or below that range. Similar effects have been produced when ANN models have been used to predict streamflow at progressively longer lag times—not exactly reducing the amount of input data available but calling for more output information from the same input data (Toth and Brath 2007). In attempting to predict streamflow at longer lag times, Toth and Brath (2007) found that the peak values decayed first, similar to the results here for sensible and latent heat fluxes. By training on a full range of flux conditions, the ANN will be capable of replicating flux values when similar forcing variable values are present.

The results found from this study suggest that ANNs may prove to be a reliable method for gap filling flux data at semi-arid sites, in agreement with the results found by other studies using ANNs for gap filling in humid regions (Papale and Valentini 2003, Moffat et al 2007). This conclusion indicates that there is an identifiable relationship between the forcing variables used in the model in this study and the measured fluxes.

From this work, a notion of sufficient data availability for gap filling can be obtained. At the annual level, sufficient data richness exists in 9-12 months of signal and no more than 20% of data missing. The reason for this is the need for the ANN model to “see” not only all of the possible variability in the fluxes but also in the forcing variables. Much as an incomplete lookup table would lead to poor gap filling because parts of the forcing record are absent, similarly for the ANN to identify the effects of all possible forcing conditions, the training must include a range of values both in the model inputs and outputs. To put this in terms of applications for gap filling data, the individual application may demand different levels of data. For example, if a study were only interested in summer fluxes, then a shorter record (one month or less) may be suitable though erratic model behavior may result from low total data length (see Figure 9d). For a study looking at annual patterns in fluxes, then the nine- to twelve-month training periods are advised to incorporate all possible combinations of flux and forcing values.

Overall, this finding suggests that in the model-framework of a neural network, data loss can significantly impact the ability of the model to reproduce the desired relationships between flux and forcing variables. In terms of actual data loss, nighttime data is frequently removed due to poor turbulent mixing. The results from Figure 13

show that the extent of under- and overestimation of fluxes favors under estimation under conditions of data loss even for low values of fluxes (approximately  $50\text{W}/\text{m}^2$  of latent heat in Figure 13d). This persistent underestimation may lead to gap-filled records that do not correspond well to measured values or to the process which the model is intended to replicate when data is lost.

In the future, a more rigorous treatment of the flux-forcing relationship would allow an ANN gap filling procedure to become more streamlined and promote the most accurate and appropriate system for generating gap filled data sets. Using other forms of ANNs, including other functions in the hidden layer of the model as well as ensemble ANN structures, could improve the ability of neural network models to replicate the existing flux behavior and thus to infer missing variables (see Shank 2006 for a treatment of ensemble ANNs to predict dew point temperatures). Developing a seasonality function may also improve the ANN performance by delineating between periods of similar forcing conditions under different levels of moisture availability or plant activity. Where the ANNs used in this study were employed only to predict sensible and latent heat fluxes, as an analog to the CHASM model structure, future work should also be done to predict the carbon flux from the site. The ground heat flux may also be considered a subject for gap filling analysis.

Continuing work with model-based gap filling should be performed in conjunction with the evolving standards for gap filling of flux records in the FLUXNET community. By working in conjunction with the FLUXNET developers, flux tower data sets from semi-arid sites will meet the same level of processing and treatment as those

from humid sites. In meeting these standards, these semi-arid flux records can be compared directly to those from other regions, allowing broader intercomparison of the ecohydrological interactions across plant communities as well as improved development of broad-scale land surface models for global circulation models.

## APPENDIX A: LAND SURFACE MODEL TESTS

A relatively simple LSM was selected to provide a low level of model complexity for ease in model identification procedures. With a deterministic LSM, the user is dealing with an explicit set of relationships (equations) that relate the forcing variables as inputs to the fluxes as outputs. These equations provide a more direct relationship than the statistical regression generated by an ANN, which comes with a higher degree of complexity with respect to parameterization and subsequent model identification. Since the model is presumed to express a known physical relationship between the identified forcing variables and the desired outputs, a model calibration program can be used to select optimal parameter values that yield a model structure to produce accurate calculated flux values from forcing variable inputs. The use of a physically-based model in this study was for comparison against the ANN model results for sensitivity to available data. However, as will be discussed below, there were extensive concerns about the model implementation for this study.

### A.1 Chameleon Surface Model (CHASM)

The Chameleon Surface Model (CHASM; Desborough 1999, Pitman et al 2003) was used in this study to provide the simplest possible LSM parameterization for the purpose of comparing forcing and flux variables. The CHASM method was formulated in such a way to provide varying degrees of model complexity, from a simple bucket model to including atmospheric and vegetative resistances to energy fluxes. Because this study is not concerned with carbon fluxes and NEE, the absence of these variables as outputs of

the model is not a concern for this study. This model has been used in a number of semi-arid environments including sites in Australia (Pitman and McAvaney 2004) and at desert-floor in Arizona associated with the Project for Intercomparison of Land-surface Parameterization Schemes (Pitman et al. 2003). The parameterization of CHASM is described below; each form of the model uses the same parameterization, regardless of the complexity level employed.

The surface energy balance is broken down based on land cover type (vegetation, bare ground, snow) and at the highest level of model complexity explicitly determines transpiration, soil evaporation and canopy interception. Under canopy interception, the vegetation is subsequently divided into wet and dry segments. Root zone soil moisture is treated as a bucket-style model and soil is separated into four depths for temperature, which is calculated by a finite difference method. Depending on the parameterization used, any tile can produce moisture flux by canopy evaporation, transpiration, soil evaporation and snow sublimation. These fluxes can be limited based on resistances applied based on the parameterization employed.

At the simplest level, the EB parameterization, aerodynamic resistances are determined without correcting for atmospheric stability. Evaporation occurs from root zone moisture and snow, to which the atmospheric resistance is then applied. The RS parameterization method is similar though it includes a constant surface resistance and a corrected aerodynamic resistance. Canopy interception is included in the RS-I parameterization with its own resistance to evaporation and the RSGI includes bare ground evaporation as well. The SLAM parameterization replaces the surface resistance

with a temporally-variant canopy resistance and an area-weighted tile delineation which allows for independent temperature patterns.

The CHASM model has been used extensively to track the model variability tested in the PILPS campaign (Pitman et al 2003) and has been shown to describe some of the model behaviors identified therein. One advantage to the CHASM formulation is the consistency in model structure, which allows the user to maintain a common set of parameters while altering the complexity for testing.

#### A.2 Calibration of CHASM using Shuffled Complex Evolution (SCE-UA)

Having reformulated the CHASM model from its original FORTRAN code into MATLAB, the model was calibrated using the Shuffled Complex Evolution (SCE-UA) algorithm (Duan et al 1994). This calibration technique provides a robust method for optimization. The method evolves a user-defined set, or complex, of points in the parameter space toward an optimum based on the Downhill Simplex Method (DSM) of Nelder and Mead (1965). These complexes are then randomly redistributed and evolved again using DSM. This procedure is repeated until an optimum value is obtained based on the comparison of each optima from the previous evolutions. For a more extensive description of the SCE-UA technique, including the theoretical basis for its development and implementation, see Duan et al 1994 for details.

This method has been widely used in the hydrological community for model identification and calibration (Yapo et al 2003, Vrugt et al 2003a, Vrugt et al 2003b) and has been demonstrated to provide a substantially broad sampling of the parameter space

to yield a reliable optimum parameter set. In this study, the SCE-UA method was used to optimize the CHASM model based on measured latent heat fluxes at the study site.

### A.3 CHASM Results

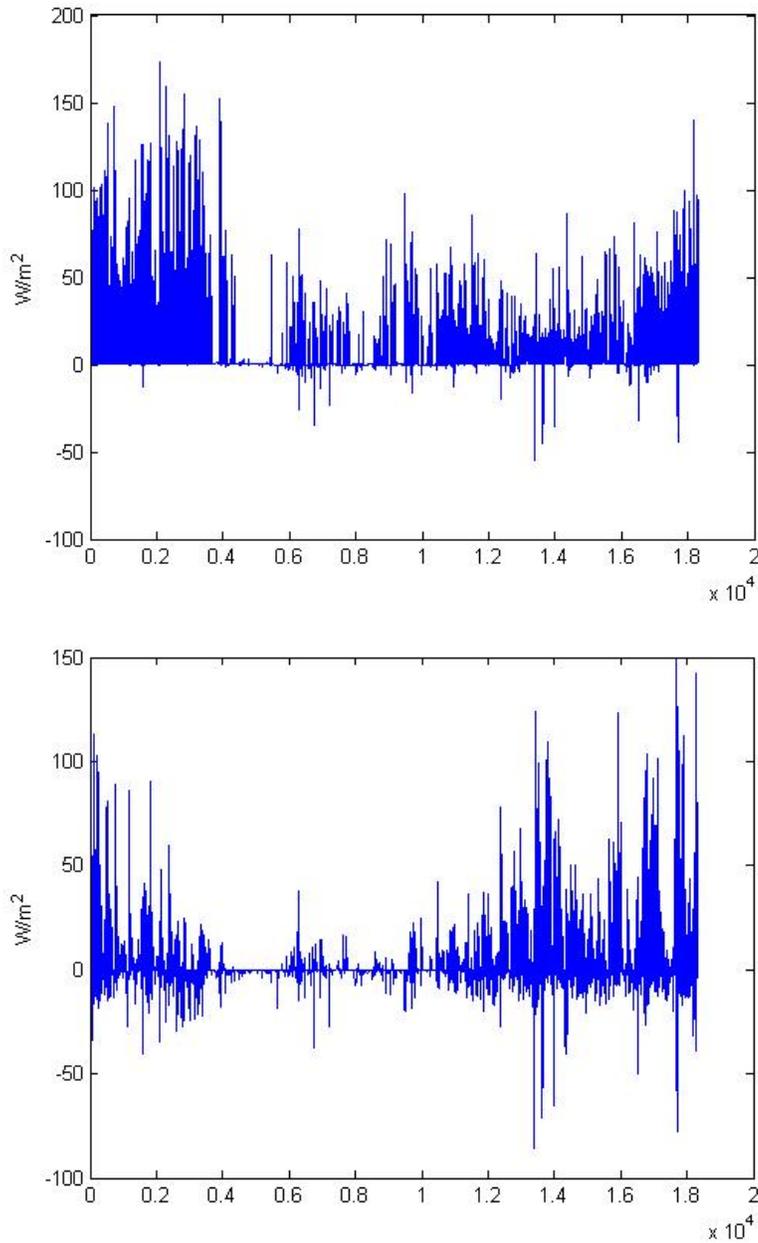
The trials performed using the CHASM model framework presented significant difficulty in this work. The model results were often not in agreement with the measured fluxes and displayed substantial variation from the expected values (Figure A1). This disagreement between modeled and observed data for fluxes was especially marked during periods of peak flux intensity. While no model can be expected to precisely represent the observed behavior, the results in Figure A1 are indicative of more severe problems with the implementation of the CHASM model for continued work with flux measurements on this site. The nature of this model-data disconnect and its implications are discussed below. As a result of this poor model behavior, the CHASM model was abandoned for further application in this study. Further work to improve the behavior of CHASM as related to this site should take place before it is used to assess the flux-forcing relationship. This also suggests that the explicit formulation of the CHASM model does not provide an appropriate replacement for the conceptual models used to assess flux data for this site.

The results shown for the CHASM trials suggest that the model is unsuited for depicting the flux processes on this study site. The modeled latent and sensible heat fluxes do not correlate well on inspection to the measured fluxes, indicating that the processes in the model may not be appropriately linking the behaviors of the forcing and flux data sets. This result indicates that some relationship between the existing soil

moisture on the site and the soil moisture depicted by the model is misrepresented in the model. While energy balance closure exists in the model, the assignment of energy appears to be incorrect even after model calibration.

The cause for the erroneous model behavior is most likely related to the soil moisture component of the model and the inability of a simple, bucket-style parameterization of soil moisture to properly represent the variable soil moisture at this site along a riparian corridor. The model has been demonstrated to perform reasonably well at semi-arid sites in previous studies (Pitman et al 2002, Hogue et al 2006), but none of those sites had been located along riparian corridors. For sites where precipitation and soil moisture may be expected to behave more uniformly, and thus soil moisture would be more closely linked to precipitation and not conditions of soil moisture up- or down-gradient from the site, the model assumptions may be more appropriate. On a site such as the one in this study, which exists along a riparian corridor, the assumptions in a bucket-style model appear to be inappropriate to depict the mechanisms by which both vegetation and evaporative demand may access soil moisture. Thus the CHASM structure does not adequately represent the process such that it would be appropriate to use for further flux data analyses.

In light of the difficulties in implementing CHASM for this site, a more in-depth examination of the model processes will be conducted to investigate the assignment of energy between latent, sensible and ground heat fluxes and ways in which the results shown here can be rectified.



*Figure A: Plots of latent (upper) and sensible (lower) heat fluxes from the CHASM model results. The erratic behavior, especially during the summer months (around 8000 half-hourly intervals, suggests poor model behavior relative to observations.*

## APPENDIX B. TABLES

	Precipitation	Temperature	Solar Rad	Wind Speed	Vapor Press	Pressure
Correlation	0.2823	0.9565	0.8933	0.7833	0.9739	0.9837

*Table 1: Correlation statistics comparing meteorological forcing variables from the flux and meteorological records at the Charleston Mesquite site.*

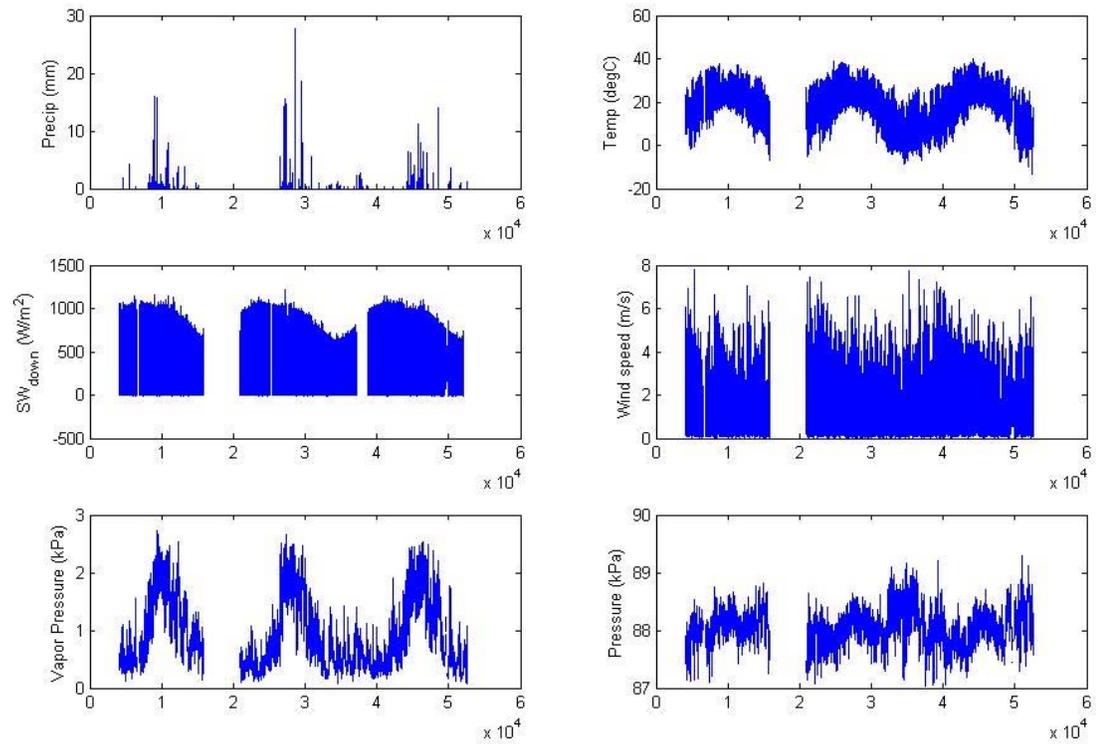
Training Length (days)	MAE		Correlation		R <sup>2</sup>		Energy Bal.
	LE	H	LE	H	LE	H	
7	133.6	122.4	0.025	0.300	-1.65	-0.22	78%
14	145.8	114.5	0.171	0.477	-2.15	-0.45	50%
60	51.85	99.64	0.685	0.576	0.60	0.82	67%
90	51.14	99.19	0.691	0.576	0.61	0.82	57%
180	49.21	90.57	0.699	0.661	0.64	0.83	67%
270	50.09	89.80	0.705	0.711	0.63	0.83	61%
360	49.75	89.98	0.694	0.700	0.63	0.83	57%

*Table 2: Summary of error estimates for training periods of different lengths.*

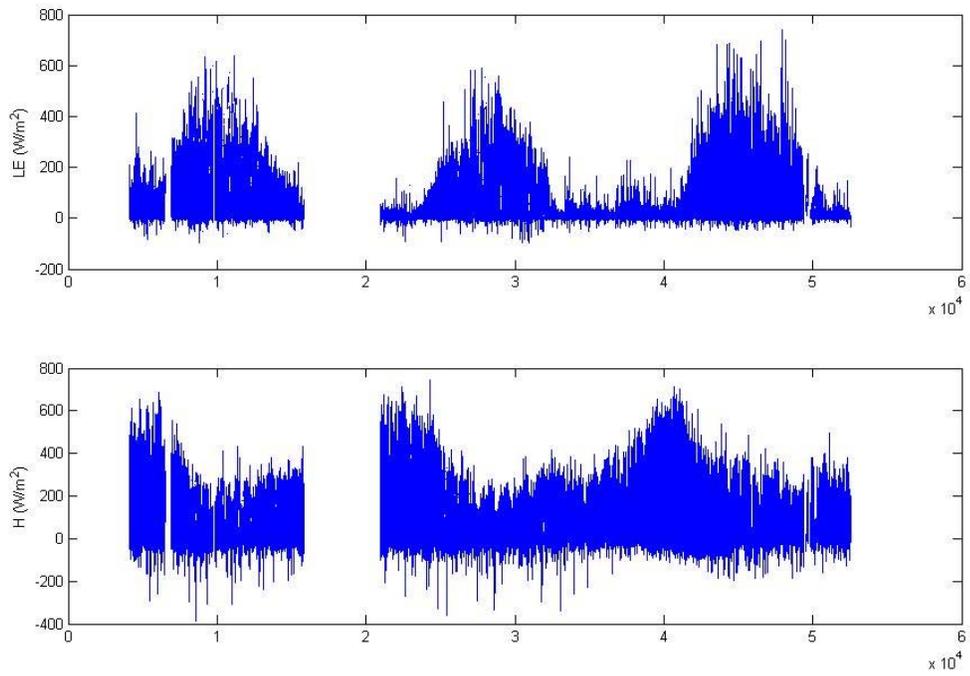
	<b>Data Coverage</b>									
	<b>100%</b>		<b>95%</b>		<b>90%</b>		<b>80%</b>		<b>70%</b>	
	<b>LE</b>	<b>H</b>	<b>LE</b>	<b>H</b>	<b>LE</b>	<b>H</b>	<b>LE</b>	<b>H</b>	<b>LE</b>	<b>H</b>
<b>RMSE</b>	49.68	88.88	49.65	88.86	50.46	87.76	53.05	86.46	57.24	86.19
<b>Correl.</b>	0.701	0.708	0.699	0.705	0.701	0.703	0.702	0.706	0.694	0.707
<b>R<sup>2</sup></b>	0.634	0.831	0.634	0.831	0.622	0.825	0.582	0.807	0.513	0.775
<b>Energy Balance</b>	57.0%		57.6%		57.8%		58.0%		60.2%	

*Table 3: Summary statistics for performance of ANN model runs under decreased data coverage. Error indices and correlation statistics decay slightly with reduced data.*

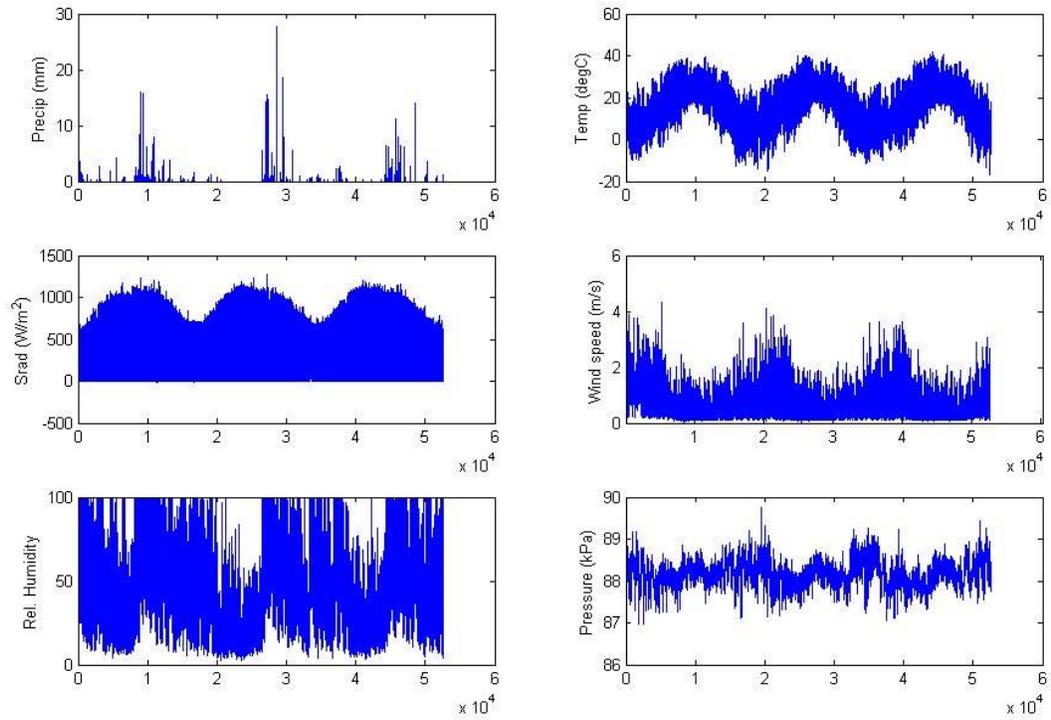
## APPENDIX C. FIGURES



*Figure 1: Meteorological variables measured from the Charleston Mesquite flux tower site. Plots include (from top left, to bottom right) precipitation, temperature, shortwave downward radiation, wind speed, vapor pressure, and ambient air pressure. The x-axis on the plots indicates the half-hourly time step (17520 time steps = 365 days).*



*Figure 2: Latent (upper) and sensible (lower) heat fluxes measured at the Charleston Mesquite flux tower site. Note the long gaps for the dormant seasons during the first two years of tower operation.*



*Figure 3.: Meteorological variables from the associated meteorological station co-located with the Charleston Mesquite flux tower. Plots include (from top left to bottom right) precipitation, temperature, net solar radiation, wind speed, relative humidity, and ambient air pressure.*

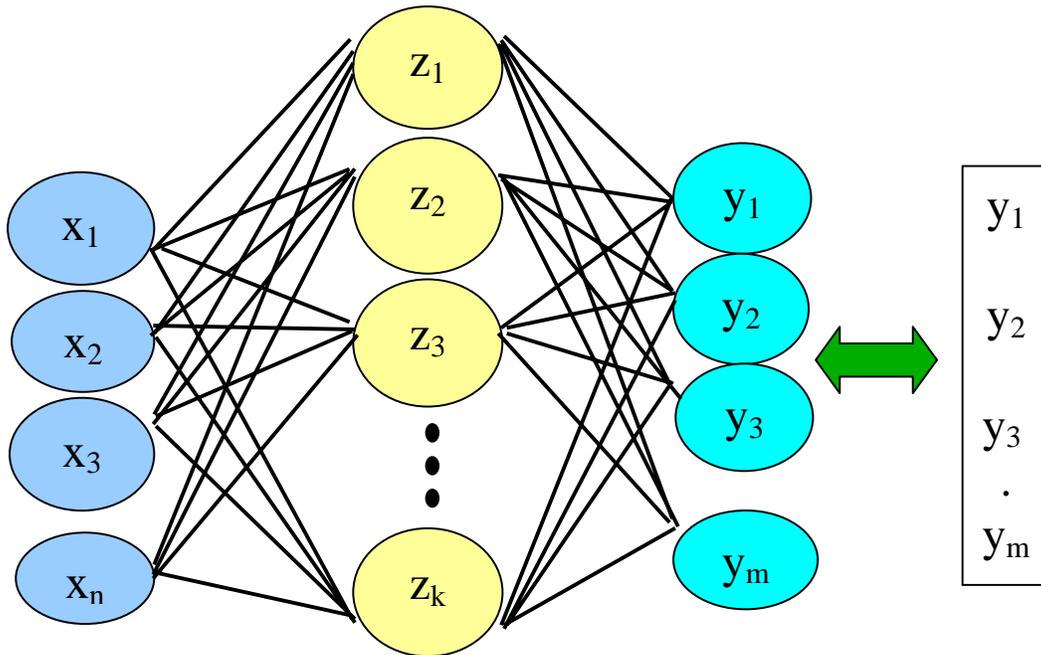
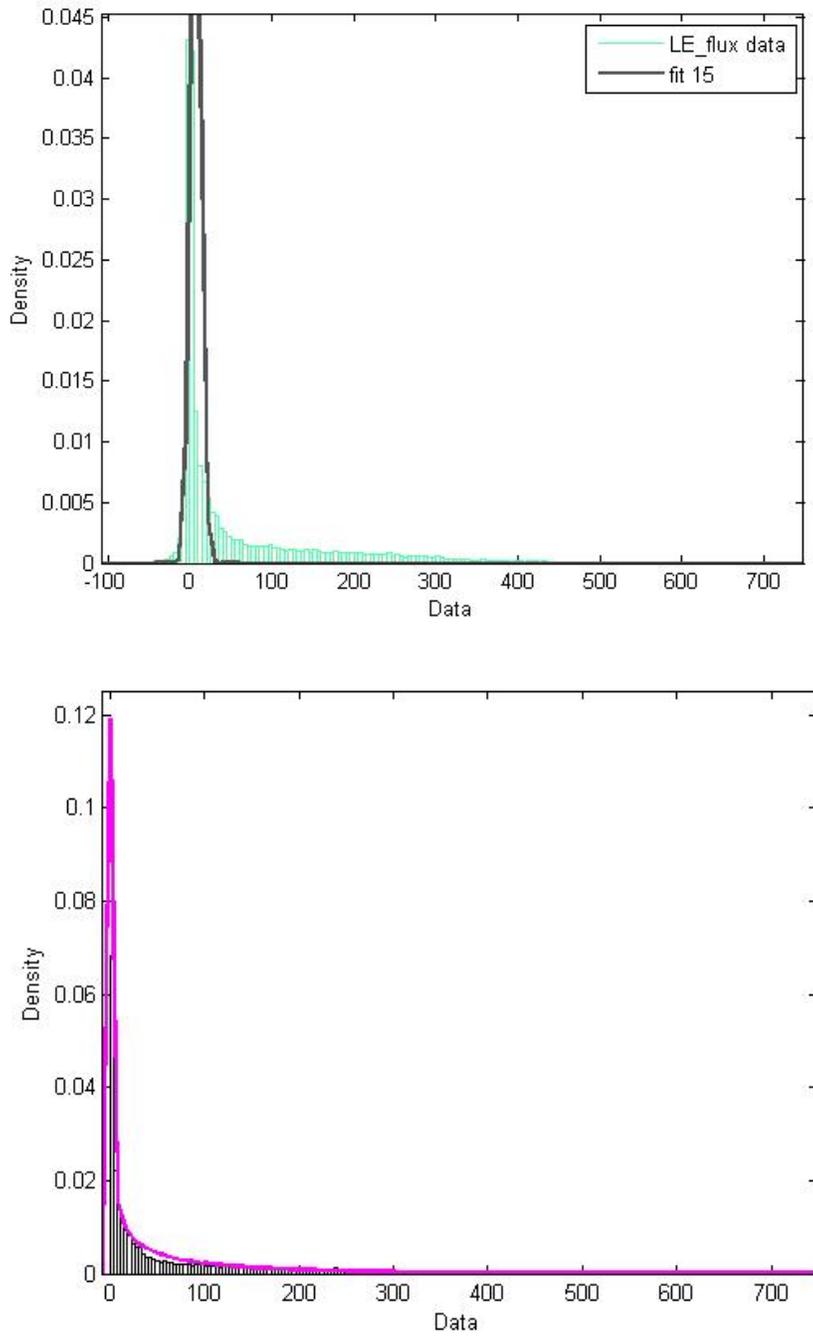
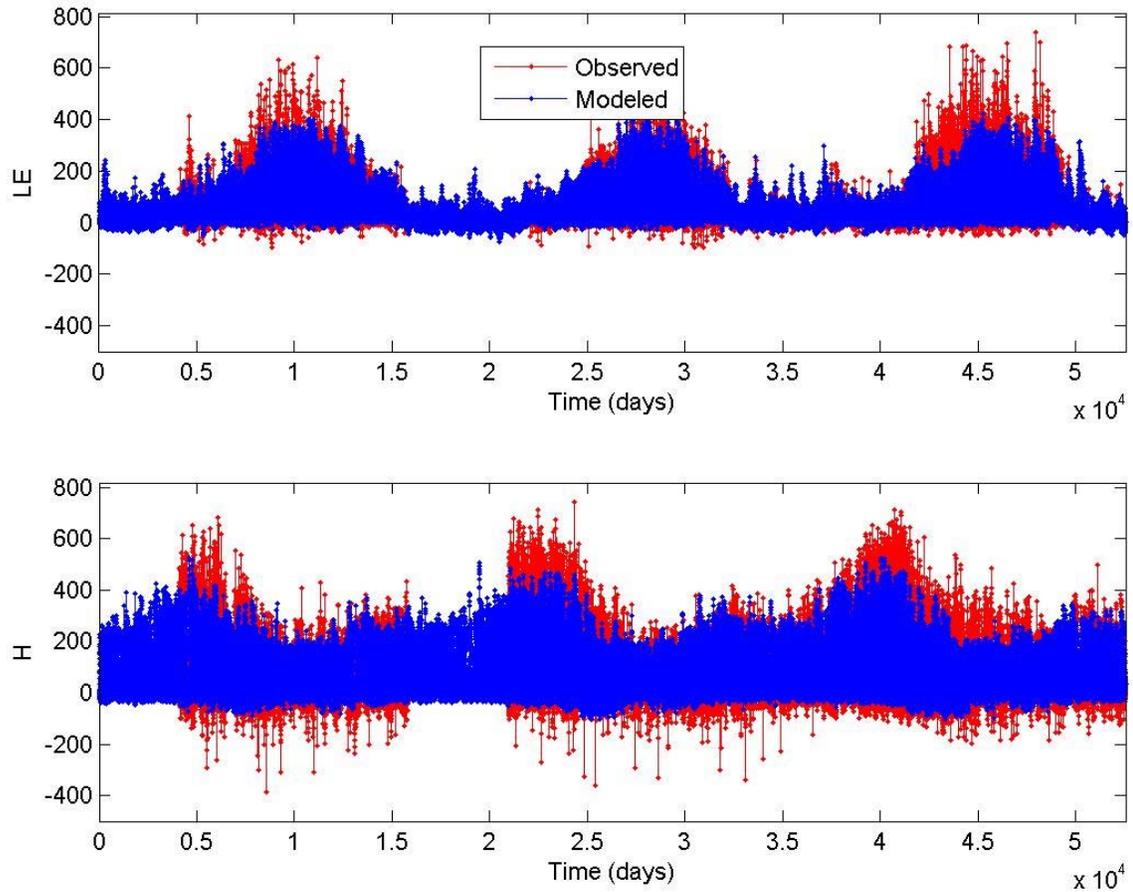


Figure 4: Schematic diagram of a three-layer neural network. The input layer is shown in blue, the middle layer in yellow and the output layer in green. The output layer is then trained against the observed output values in the white box.



*Figure 5: Frequency distributions for latent heat data. Upper plot shows data with a normal distribution fit. Lower plot shows data (shifted to remove negative values, as bars) with a gamma distribution fit (pink line).*



*Figure 6: Plots showing ANN modeled (blue) and observed (red) latent (upper) and sensible (lower) heat fluxes. This model was trained for 365 days (17520 half-hourly intervals). Peak behavior in both fluxes appears truncated.*

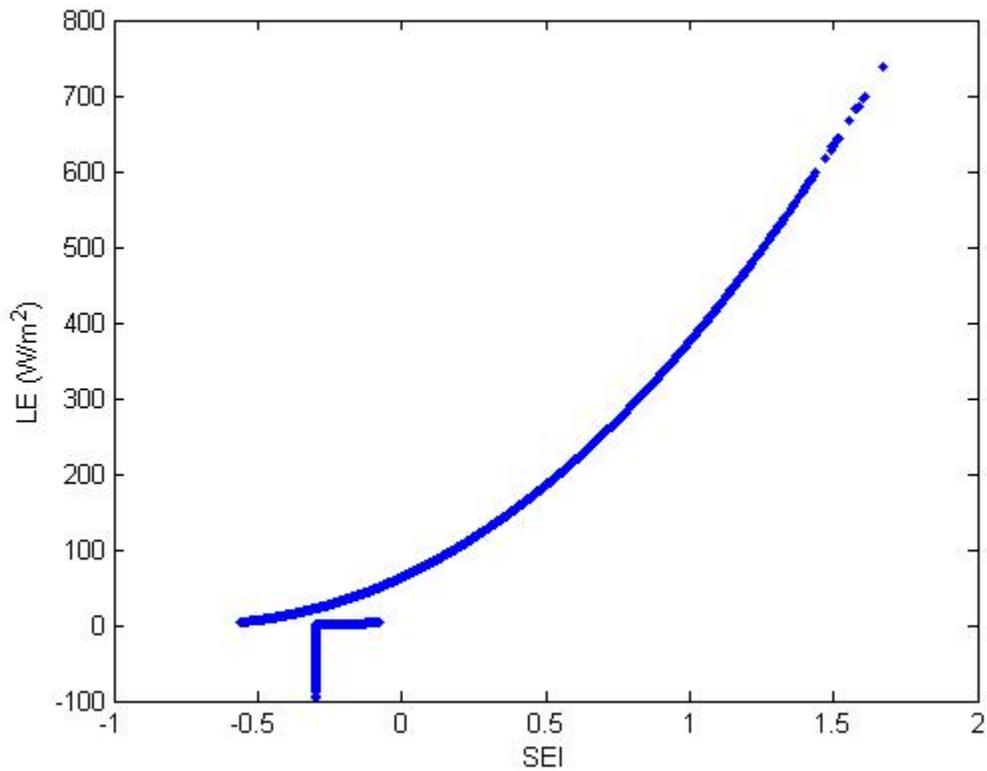
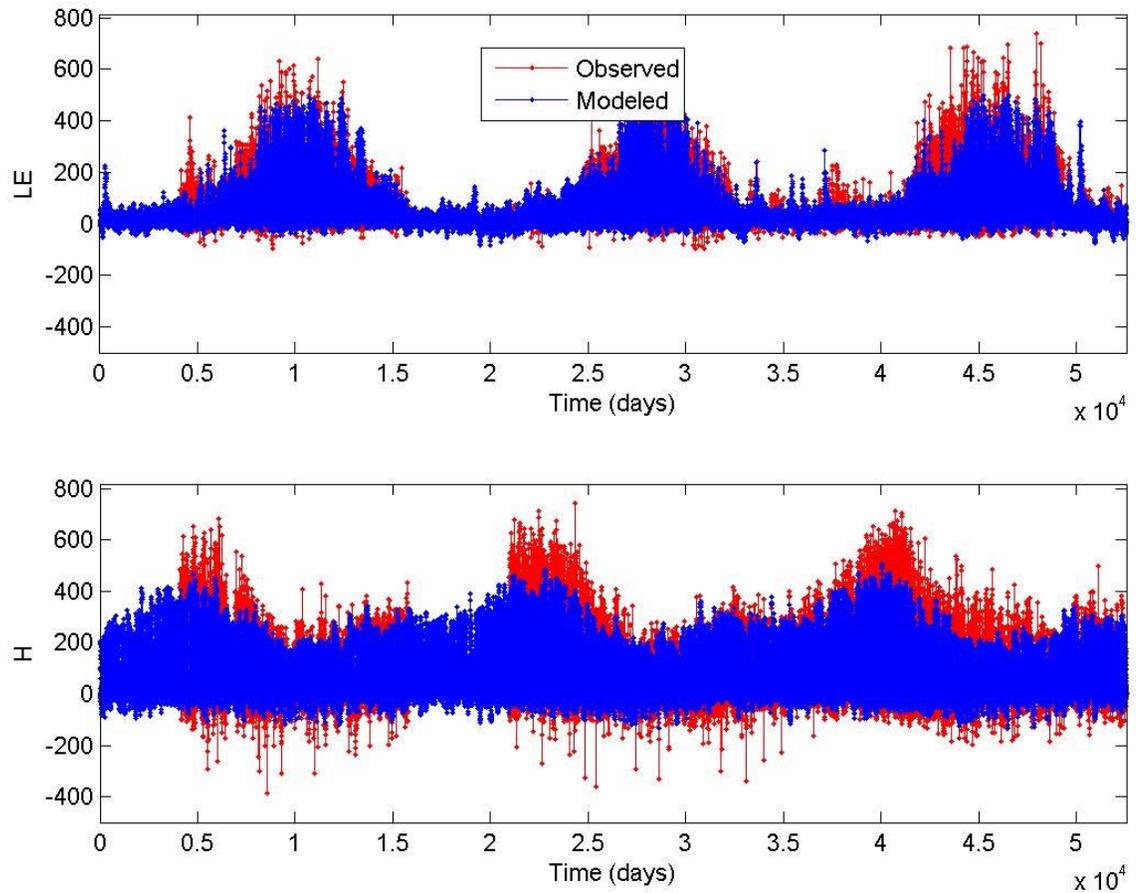


Figure 7: Plot showing values of the standardized evapotranspiration index (SEI) against the measured latent heat. This curve was fit with the second order polynomial shown to convert the model-generated SEI values into latent heat fluxes. Note that the function values for  $LE < 0$  were extrapolated according to the equation; the discontinuity results from the formulation of equations 9 and 10 for certain values of  $H(x)$ .



*Figure 8: Plots of modeled and observed latent and sensible heat fluxes (legend and plot order as in Figure 6). Comparing this plot to Figure 6, note that the peak behavior in the modeled latent heat flux better approaches the peak values of the observed flux. Similarly, the lower values more nearly approach the observations.*

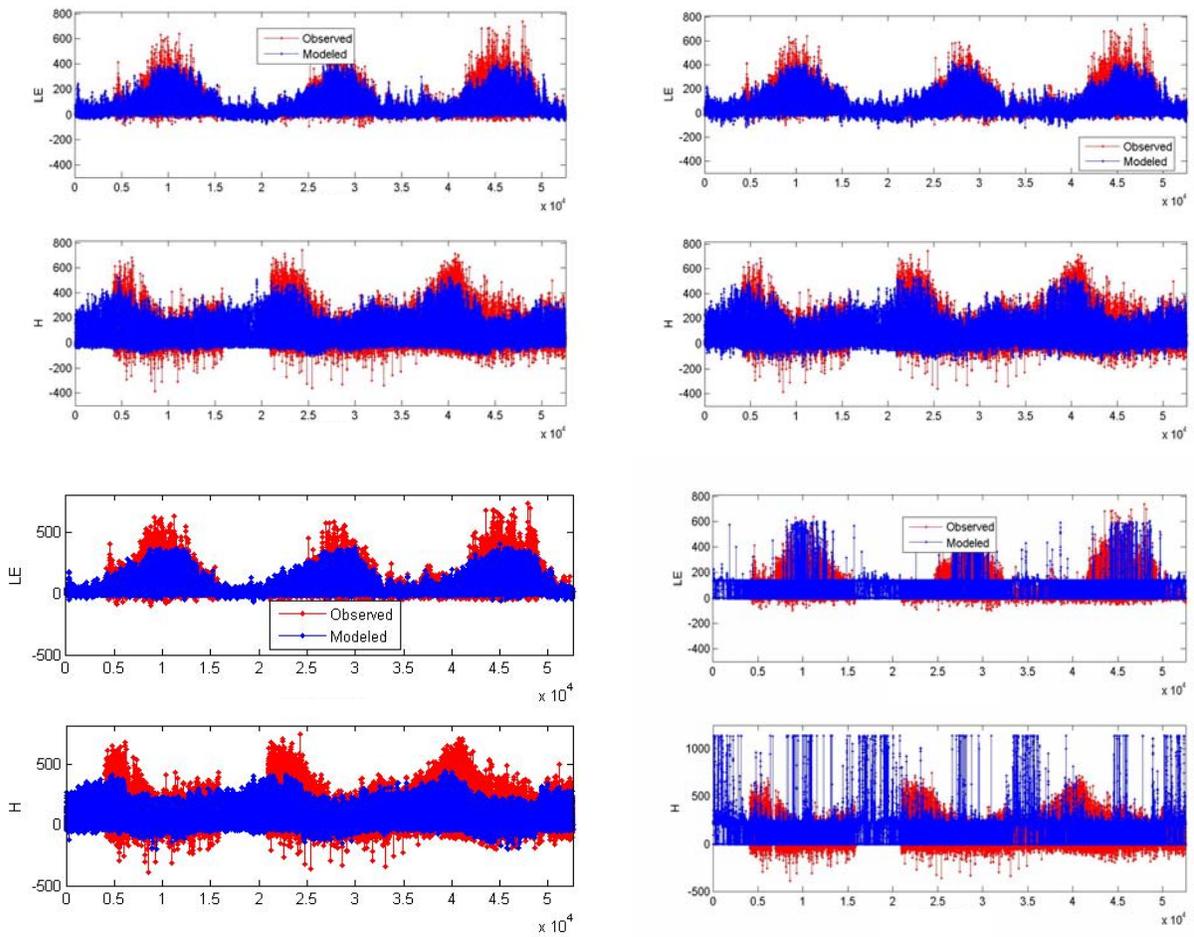
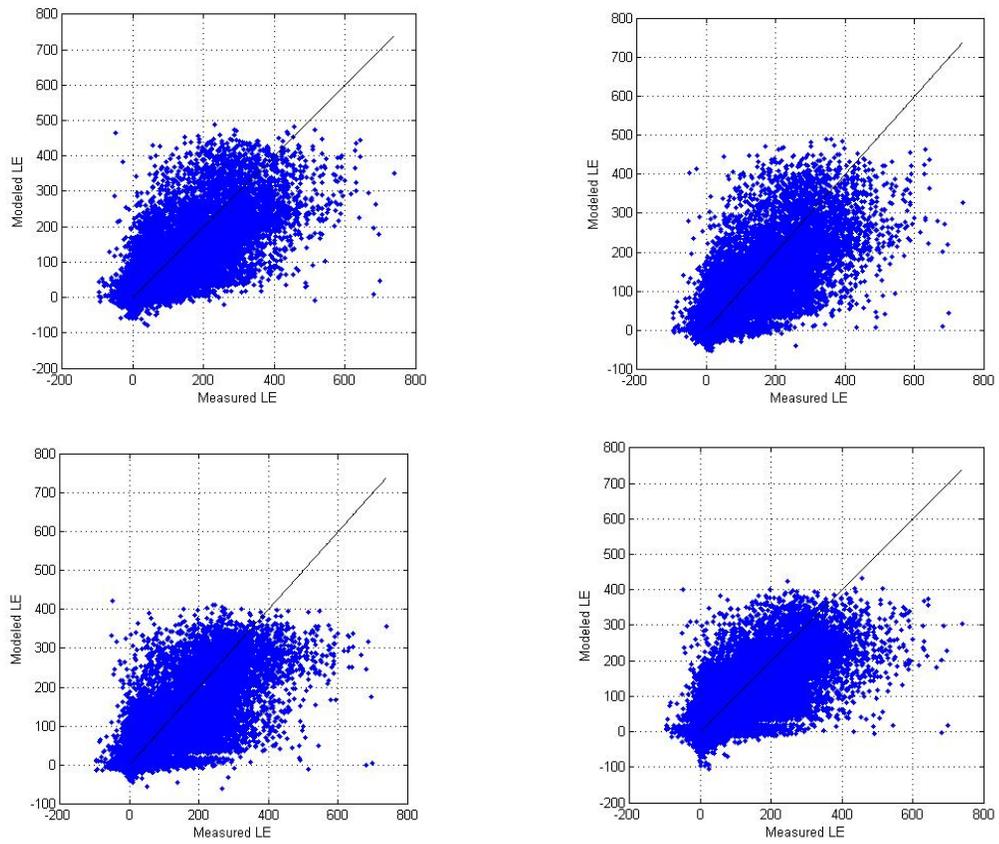


Figure 9: Plots showing the effects of shortened training data lengths on ANN-modeled latent and sensible heat fluxes. Plot legend and position in (a)-(d) as in Figure 6. Training lengths are (a) 365 days, (b) 270 days, (c) 180 days, (d) 90 days.



*Figure 10: Scatter plots of latent heat (measured vs. modeled) produced under training period lengths of (a) 360 days, (b) 270 days, (c) 180 days, (d) 90 days. Similar results were found for scatter plots of sensible heat*

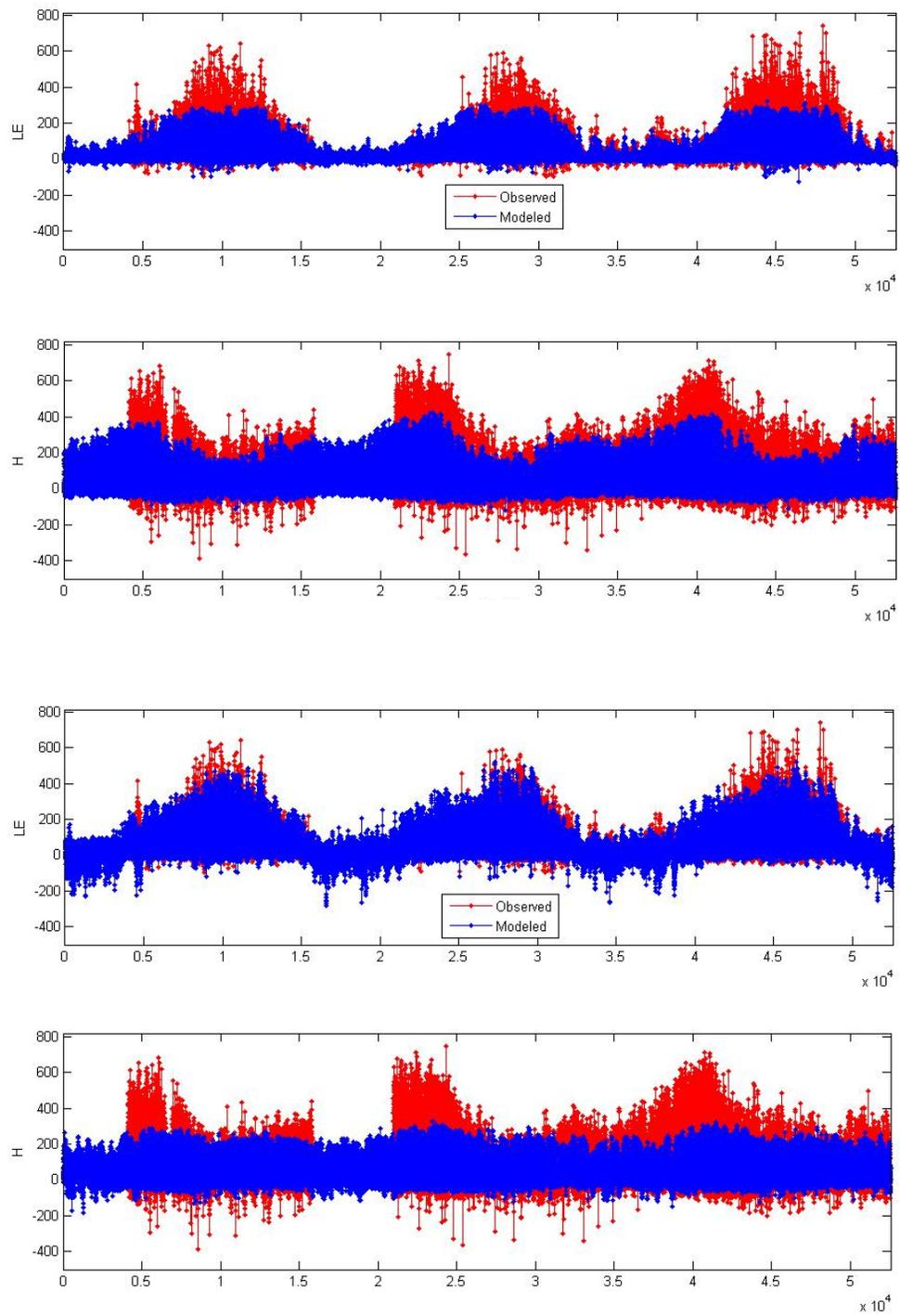


Figure 11: Comparison of model results with different patterns in training data: (a) during winter, i.e. low latent heat flux; and (b) during summer, i.e. high latent heat flux.

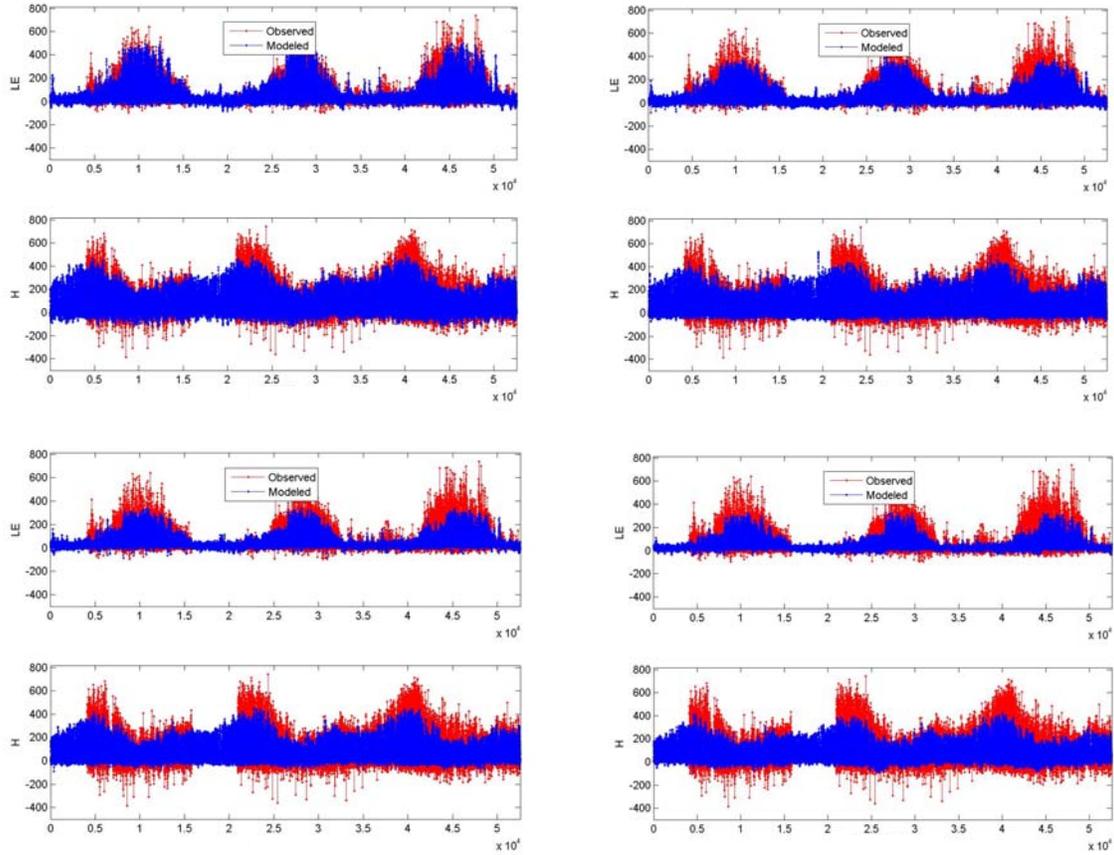
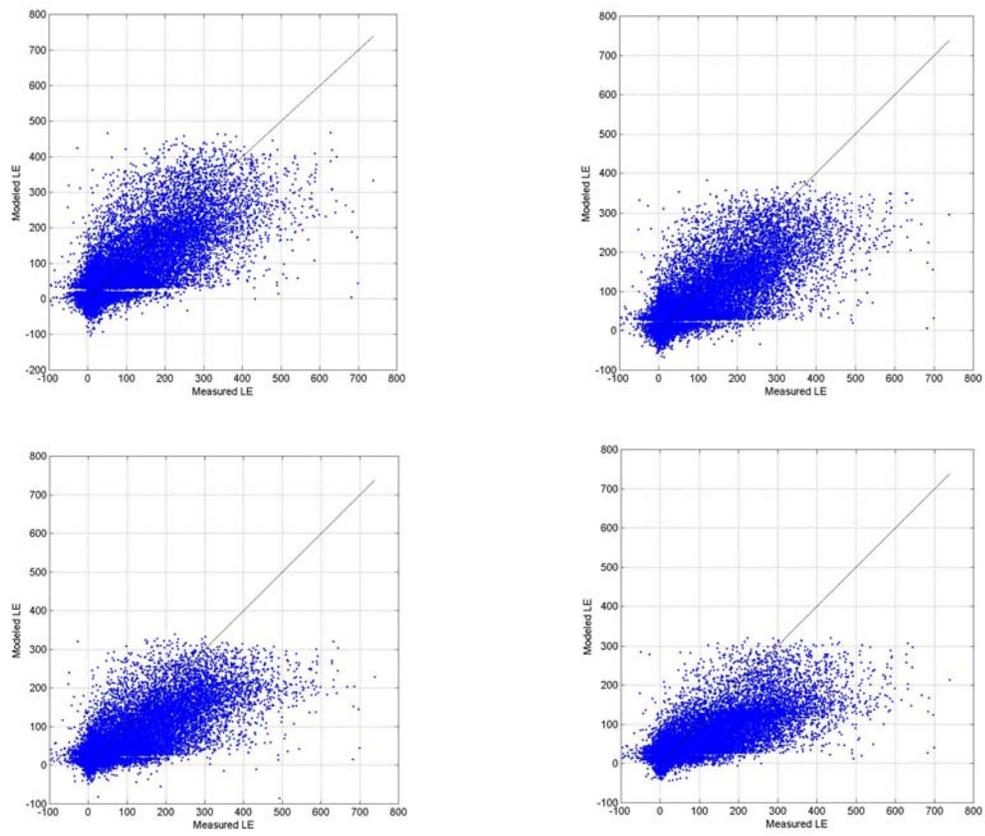


Figure 12: Plots showing the decay of ANN-modeled fluxes under reduced availability of data for training for (a) 100%, (b) 90%, (c) 80% and (d) 70% data available. Note the loss of low-frequency extreme events in the model results when training data is restricted.



*Figure 13: Scatter plots of modeled and measured latent heat under conditions of reduced training data. Arrangement (a-d) as in Figure 12.*

## REFERENCES

- Acevedo, O. C., et al. (2007), Turbulent carbon exchange in very stable conditions, *Boundary-Layer Meteorology*, 125(1), 49-61.
- Alavi, N., et al. (2006), Filling gaps in evapotranspiration measurements for water budget studies: Evaluation of a Kalman filtering approach, *Agricultural and Forest Meteorology*, 141(1), 57-66.
- Aubinet, M., et al. (2000), Estimates of the annual net carbon and water exchange of forests: The EUROFLUX methodology, in *Advances in Ecological Research, Vol 30*, edited, pp. 113-175, Academic Press Inc, San Diego.
- Baldocchi, D., et al. (2000), On measuring net ecosystem carbon exchange over tall vegetation on complex terrain, *Boundary-Layer Meteorology*, 96(1-2), 257-291.
- Baldocchi, D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bulletin of the American Meteorological Society*, 82(11), 2415-2434.
- Baldocchi, D. D., et al. (1988), MEASURING BIOSPHERE-ATMOSPHERE EXCHANGES OF BIOLOGICALLY RELATED GASES WITH MICROMETEOROLOGICAL METHODS, *Ecology*, 69(5), 1331-1340.
- Baldocchi, D. D., et al. (2004), How plant functional-type, weather, seasonal drought, and soil physical properties alter water and energy fluxes of an oak-grass savanna and an annual grassland, *Agricultural and Forest Meteorology*, 123(1-2), 13-39.
- Baxt, W. G. (1995), APPLICATION OF ARTIFICIAL NEURAL NETWORKS TO CLINICAL MEDICINE, *Lancet*, 346(8983), 1135-1138.
- Cava, D., et al. (2006), A case study of short-term closure of the surface energy balance, *Nuovo Cimento Della Societa Italiana Di Fisica B-General Physics Relativity Astronomy and Mathematical Physics and Methods*, 121(8), 845-855.
- Cook, D. R. (2007), Energy Balance Bowen Ratio Handbook, edited by A. R. Measurement, p. 23.
- Desborough, C. E. (1999), Surface energy balance complexity in GCM land surface models, *Climate Dynamics*, 15(5), 389-403.

- Duan, Q. Y., et al. (1992), EFFECTIVE AND EFFICIENT GLOBAL OPTIMIZATION FOR CONCEPTUAL RAINFALL-RUNOFF MODELS, *Water Resources Research*, 28(4), 1015-1031.
- Falge, E., et al. (2001a), Gap filling strategies for long term energy flux data sets, *Agricultural and Forest Meteorology*, 107(1), 71-77.
- Falge, E., et al. (2001b), Gap filling strategies for defensible annual sums of net ecosystem exchange, *Agricultural and Forest Meteorology*, 107(1), 43-69.
- Finnigan, J. (2004), The footprint concept in complex terrain, *Agricultural and Forest Meteorology*, 127(3-4), 117-129.
- Fischer, M. L., et al. (2007), Spatiotemporal variations in growing season exchanges of CO<sub>2</sub>, H<sub>2</sub>O, and sensible heat in agricultural fields of the Southern Great Plains, *Earth Interactions*, 11.
- Foken, T., and B. Wichura (1996), Tools for quality assessment of surface-based flux measurements, *Agricultural and Forest Meteorology*, 78(1-2), 83-105.
- Foken, T., et al. (2006), Some aspects of the energy balance closure problem, *Atmospheric Chemistry and Physics*, 6, 4395-4402.
- Friend, A. D., et al. (2007), FLUXNET and modelling the global carbon cycle, *Global Change Biology*, 13(3), 610-633.
- Gasteiger, J., and J. Zupan (1993), Neural Networks in Chemistry, in *ANGEWANDTE CHEMIE-INTERNATIONAL EDITION IN ENGLISH*, edited, pp. 503-527, VCH Publishers, Inc, Deerfield Beach, FL.
- Goulden, M. L., et al. (2006), Nocturnal cold air drainage and pooling in a tropical forest, *Journal of Geophysical Research-Atmospheres*, 111(D8), 14.
- Gove, J. H., and D. Y. Hollinger (2006), Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange, *Journal of Geophysical Research-Atmospheres*, 111(D8), 21.
- Henderson-Sellers, A., et al. (1993), THE PROJECT FOR INTERCOMPARISON OF LAND-SURFACE PARAMETERIZATION SCHEMES, *Bulletin of the American Meteorological Society*, 74(7), 1335-1349.

- Hogue, T. S., et al. (2006), Evaluating model performance and parameter behavior for varying levels of land surface model complexity, *Water Resources Research*, 42(8), 17.
- Hollinger, D. Y., and A. D. Richardson (2005), Uncertainty in eddy covariance measurements and its application to physiological models, *Tree Physiology*, 25(7), 873-885.
- Hsu, K. L., et al. (1997), Precipitation estimation from remotely sensed information using artificial neural networks, *Journal of Applied Meteorology*, 36(9), 1176-1190.
- Hsu, K. L., et al. (1999), Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation, *Water Resources Research*, 35(5), 1605-1618.
- Katul, G. G., et al. (2006), The influence of hilly terrain on canopy-atmosphere carbon dioxide exchange, *Boundary-Layer Meteorology*, 118(1), 189-216.
- Leuning, R. and M.J. Judd. (1996), "The relative merits of open- and closed-path analyzers for measurement of eddy fluxes." *Global Change Biology*, 2, 241-253.
- Massman, W. J., and X. Lee (2002), Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges, *Agricultural and Forest Meteorology*, 113(1-4), 121-144.
- McKee, T. B., et al. (1993), The relationship of drought frequency and duration to time scales, in *Proceedings of the Eight Conference on Applied Climatology Anaheim, CA, January 17-22*, edited, pp. 179-184, American Meteorological Society, Boston, MA.
- Michaelis, L., and M. L. Menten (1913), Die Kinetik der Invertinwirkung, in *Biochem. Z.*, edited, pp. 333-369.
- Mishra, A. K., and V. R. Desai (2006), Drought forecasting using feed-forward recursive neural network, *Ecological Modelling*, 198(1-2), 127-138.
- Moffat, A. M., et al. (2007), Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, *Agricultural and Forest Meteorology*, 147(3-4), 209-232.
- Nabney, I. T. (2002), *NETLAB: Algorithms for Pattern Recognition.*, 420 pp., Springer, London.

- Nelder, J. A., and R. Mead (1965), A Simplex Method for Function Minimization, in *Computer Journal*, edited, pp. 308-313.
- Ooba, M., et al. (2006), Comparisons of gap-filling methods for carbon flux dataset: A combination of a genetic algorithm and an artificial neural network, *Ecological Modelling*, 198(3-4), 473-486.
- Papale, D., and A. Valentini (2003), A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, *Global Change Biology*, 9(4), 525-535.
- Papale, D., et al. (2006), Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, *Biogeosciences*, 3(4), 571-583.
- Pilegaard, K., et al. (2001), Two years of continuous CO<sub>2</sub> eddy-flux measurements over a Danish beech forest, *Agricultural and Forest Meteorology*, 107(1), 29-41.
- Pitman, A. J., et al. (2003), The CHAmeleon Surface Model: description and use with the PILPS phase 2(e) forcing data, *Global and Planetary Change*, 38(1-2), 121-135.
- Pitman, A. J., and B. J. McAvaney (2004), Impact of varying the complexity of the land surface energy balance on the sensitivity of the Australian climate to increasing carbon dioxide, *Climate Research*, 25(3), 191-203.
- Richardson, A. D., and D. Y. Hollinger (2007), A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO<sub>2</sub> flux record, *Agricultural and Forest Meteorology*, 147(3-4), 199-208.
- Scott, R. L., et al. (2004), Interannual and seasonal variation in fluxes of water and carbon dioxide from a riparian woodland ecosystem, *Agricultural and Forest Meteorology*, 122(1-2), 65-84.
- Scott, R. L., Goodrich, D.C., Levick, L., McGuire, R., Cable, W.L., Williams, D., Gazal, R., Yopez, E.A., Elsworth, P., Huxman, T.E. (2006), San Pedro Riparian National Conservation Area (SPRNCA) water needs - final study report, edited by U. G. Survey.
- Sellers, P. J., et al. (1986), A SIMPLE BIOSPHERE MODEL (SIB) FOR USE WITHIN GENERAL-CIRCULATION MODELS, *Journal of the Atmospheric Sciences*, 43(6), 505-531.
- Shank, D. B. (2006), Dew Point Temperature Prediction Using Artificial Neural Networks, 82 pp, University of Georgia, Athens, GA.

- Shuttleworth, W. J. (2007), Putting the 'vap' into evaporation, *Hydrology and Earth System Sciences*, 11(1), 210-244.
- Sun, J. L., et al. (2007), CO<sub>2</sub> transport over complex terrain, *Agricultural and Forest Meteorology*, 145(1-2), 1-21.
- Twine, T. E., et al. (2000), Correcting eddy-covariance flux underestimates over a grassland, *Agricultural and Forest Meteorology*, 103(3), 279-300.
- Valentini, R., et al. (2000), Respiration as the main determinant of carbon balance in European forests, *Nature*, 404(6780), 861-865.
- Van Gorsel, E., et al. (2007), Nocturnal carbon efflux: reconciliation of eddy covariance and chamber measurements using an alternative to the  $u^*$ -threshold filtering technique, *Tellus Series B-Chemical and Physical Meteorology*, 59(3), 397-403.
- Vrugt, J. A., et al. (2003a), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resources Research*, 39(8).
- Vrugt, J. A., et al. (2003b), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resources Research*, 39(8).
- Williams, C. A., and J. D. Albertson (2005), Contrasting short- and long-timescale effects of vegetation dynamics on water and carbon fluxes in water-limited ecosystems, *Water Resources Research*, 41(6), 13.
- Willis, M. J., et al. (1991), ARTIFICIAL NEURAL NETWORKS IN PROCESS ENGINEERING, *Iee Proceedings-D Control Theory and Applications*, 138(3), 256-266.
- Yapo, P. O., et al. (1996), Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *Journal of Hydrology*, 181(1-4), 23-48.