QUANTIFYING WATER DEFICIT AND ITS EFFECT ON CROP YIELDS USING A SIMPLE, GENERIC DROUGHT INDEX

By

PREM WOLI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

To readers

ACKNOWLEDGMENTS

I am grateful to Agricultural and Biological Engineering Department, University of Florida, for admitting me to the PhD program and providing assistantship and other logistics; to the Embassy of the United States, Kathmandu, for letting me enter this country to pursue my studies; to my family for providing me necessary logistics and support; to my supervisory committee members for their invaluable guidance, support, and encouragement; and, ultimately, to the God, without the blessing of whom nothing would materialize.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

QUANTIFYING WATER DEFICIT AND ITS EFFECT ON CROP YIELDS USING A SIMPLE, GENERIC DROUGHT INDEX

By

Prem Woli

December 2010

Chair: James W. Jones
Cochair: Keith T. Ingram
Major: Agricultural and Biological Engineering

Simplicity, genericness, soil-plant-atmosphere-basis, and computability at high temporal resolution are basic requirements of an agricultural drought index. The index should be able to quantify drought and the effect of drought on crop yields. However, no index was found that could meet all these requirements. The objectives of this study were to develop a new index with all these features, evaluate its ability to quantify drought and estimate yield loss from drought, explore its uncertainties as a simple and generic index, and explore ways of forecasting it. The new index, Agricultural Reference Index for Drought (ARID), was evaluated by comparing it with several other drought indices and the water deficit index (WSPD) computed by DSSAT CERES-Maize, a widely-tested crop simulation model. Among the indices compared, ARID mimicked WSPD most closely with largest correlations and smallest departures. The uncertainty ranges in ARID were generally small – less than 30% of the whole range. Even in cases with large uncertainty ranges of up to 50% of the entire possible range, values of ARID concentrated around its nominal values. Sensitivity analyses showed that 'available water capacity' and 'root zone depth' are the most important parameters and explain

about 90% of the total variability in ARID, whereas ‘runoff curve number’ and ‘drainage coefficient’ have no significant influence on ARID. Various climate indices and statistical models were investigated for forecasting ARID. Values of the Nash and Sutcliffe index (NSI) indicated that the current level of forecasting, which is El Niño-Southern Oscillation (ENSO)-based, may be improved especially for the southern part of the southeast USA, such as Florida, using artificial neural network models and climate index-based indices. According to the NSI values, artificial neural network performed better then the other models for most locations in the region. Finally, the performance of ARID to estimate yield loss from drought was evaluated by comparing the ARID-estimated yields with observed yields of cotton, maize, peanut, and soybean. The ARID-based yield models provided reasonable estimations of yield loss from drought for these crops (with mean absolute error of less than 15%, and the NSI values ranging from 0.03 to 0.63).

CHAPTER 1
INTRODUCTION

Drought

Drought may be defined as a temporary condition where the amount of water due to precipitation falls short of a threshold value (Figure 1-1). From meteorological point of view, water can be precipitation and the threshold value can be normal precipitation. From hydrological perspective, water can be surface or subsurface water and the threshold value can be normal ground or underground water level. From an agricultural viewpoint, water is the plant available water in the soil and the threshold value is the atmospheric demand for evapotranspiration (ET). From socioeconomic standpoint, water is the water that is available to a society and the threshold value is the water needed by the society. Drought basically starts from a deficiency of precipitation resulting in water shortage for a certain activity. Drought is a recurring, complex, insidious, creeping, and the least understood natural hazard (AMS, 1997) that has plagued civilization throughout history (Heim, 2002). Agriculture is often the first sector to be affected by the onset of drought due to dependence on precipitation and soil moisture reserve needed to support plant growth (Narasimhan and Srinivasan, 2005).

Because drought affects different sectors with varying frequency, duration, and magnitude, operationally, it is often categorized into four types: meteorological, agricultural, hydrological, and socio-economic (Subrahmanyam, 1967; Dracup et al., 1980; Wilhite and Glantz, 1985; AMS, 1997). Meteorological drought is the departure of precipitation from normal, whereas hydrological drought is an expression of precipitation shortfall on surface or subsurface water supply (AMS, 1997). Socio-economic drought is the period of low water supply that affects society's activities (Dracup et al., 1980). An

agricultural drought occurs when the amount of water in the root zone is not sufficient to meet the need of a particular crop at a particular time (WMO, 1975; Wilhite and Glantz, 1985; Maracchi, 2000).

Drought Index

The severity of drought may be quantified using a drought index. A drought index integrates all meteorological, hydrological, and agricultural information typically into a single number and gives a comprehensible big picture on drought conditions for decision making (Hisdal and Tallaksen, 2000; Narasimhan and Srinivasan, 2005; Hayes, 2006).

Numerous drought indices are available to quantify drought, which can be categorized into three groups: meteorological, hydrological, or agricultural. Meteorological drought indices do not easily capture the effects of soil properties and crop growth characteristics that are critical for agricultural drought monitoring (Keating et al., 1996). Similarly, hydrological drought indices are not suitable indicators for agricultural drought because the hydrological drought is out of phase from agricultural drought as there is a time lag before precipitation deficiencies are detected in surface and subsurface water sources (Narasimhan and Srinivasan, 2005) and in recovery from hydrological drought. Because an agricultural drought index reflects the interactions among precipitation, other meteorological variables, and plant available water in the soil profile in relation to crop growth and development (Lourens, 1995), agricultural drought indices are used in preference to meteorological and hydrological indices for measuring crop water deficit (Pisani et al., 1998). Currently, many drought indices are available for quantifying agricultural drought. Some of the major drought indices that are being

applied to agricultural systems are presented in Table 1-1, along with some of their features that are important for agriculture.

The available drought indices vary among themselves in many respects, such as applicability (regions and conditions), basic concept (energy-balance- or water-balance-based), input requirements, purpose (monitoring crop water deficit or estimating crop yield), and target use (agricultural, hydrological, or socioeconomic). They also vary in terms of complexity, genericness (whether the parameters of these indices are crop-specific or not), and spatio-temporal resolutions. Each index has been developed for a particular purpose and has specific features. This study is concerned with the possibility for developing a new agricultural drought index that is simpler and may perform better than the many other drought indices that are available now.

New Drought Index

The list of drought indices currently available for application in agriculture (Table 1-1) shows the possibility for developing a drought index that is soil-plant-atmosphere (SPA)-based, simple, and generic and has potential for monitoring at high spatial and temporal resolutions. For the new drought index, these features are chosen for the following reasons:

SPA-basis: Plant water status is governed by atmospheric evaporative demand, soil water conditions, and plant characteristics. Water moves from the soil, through the plant, to the leaves where it evaporates and is released into the atmosphere. Plant water deficit occurs when absorption of water by roots falls short of the atmospheric demand for transpiration. To quantify plant water deficit, an agricultural drought index should take into account all three components of the SPA continuum (Lourens and Jager, 1997; Pisani et al., 1998; Narasimhan, 2004; Narasimhan and Srinivasan, 2005).

Moreover, because crop yield is often the most important factor for a producer, an agricultural drought index should be related to yield loss caused by drought. As transpiration affects carbon dioxide assimilation and ultimately crop yield, only a SPA-based drought index can estimate drought-induced yield loss.

Simplicity: Water deficit and its effect on crop yield can be estimated accurately using crop simulation models. However, crop simulation models are difficult to apply because they require a large numbers of crop-specific inputs and parameters that are difficult for users to obtain or to justify because of the spatial variability of soils and diverse crop management practices. Thus, a complex drought index derived from a crop simulation model is impractical for broad application.

Genericness: Water stress estimated by a crop model or a drought index that is crop-specific cannot generally indicate water deficit for other crops. Moreover, crop-specific indices or models do not exist for many crops. In this situation, a generic index is needed to serve as a general indicator of agricultural drought. Because a generic index provides general information to all crops without requiring their specific inputs and parameters, a generic index is both simpler and more informative than a crop-specific index (Forrester, 1970). A generic index also provides economy in both modeling and understanding (Reynolds and Acock, 1997).

Daily resolution: Because the time scale of crop responses to water deficits is closer to one day than to one week or month, an agricultural drought index should be computed daily (Acevedo et al., 1971; Steduto et al., 2009). A daily time step or less is essential to realistically account for variability related to daily changes in weather and soil conditions (Raes et al., 2009) and the dynamic nature of water deficit effects and

crop responses (FAO, 2009). A daily time step is most practical because weather data are typically collected or reported daily.

It is hypothesized that the proposed SPA-based, simple, generic, and high-resolution drought index may characterize agricultural drought better than do many other drought indices that are available now and that, being simple and generic, the index may be more convenient for users. If the new index could mimic complex and crop-specific indices that are widely-used or trusted, not only would it be more convenient for the users to apply the index but, being relatively simpler, the chance of its application as a decision support system would increase as well.

With the development of the new index, however, several questions arise. How does it compare with others that are complex, crop-specific, and not SPA-based? Does it mimic the estimations of complex indices? Does it provide a general indicator of drought for all crops? How does it compare with indices that are not based on water balance? What are the uncertainties associated with generic parameters having fixed values? For what conditions would the index be more or less applicable? Could the index be used in forecasting drought? If so, how or on what basis could drought be forecast? Could the index be used to estimate crop yields? Does it make reasonable estimations of yield losses for specific crops?

Dissertation Objectives

The purpose of this dissertation is to formulate a simple, generic, and SPA-based agricultural drought index that can be used by farmers or people involved in crop production systems as a decision support tool to monitor and or forecast agricultural drought and its effect on yields. The specific objectives are:

- To design a new agricultural drought index to be used as a generic tool for quantifying water deficit for crops, that is, agricultural drought in general.
- To evaluate the index by comparing it with several other drought indices that have been applied to agricultural systems and with the water deficit index computed by Decision Support System for Agrotechnology Transfer (DSSAT) CERES-Maize (Jones et al., 2003), a widely-used and tested crop model, to assess how the new index would mimic the work of a complex, crop-specific drought index.
- To quantify the uncertainties in the index associated with its parameters and to estimate the importance of individual parameters with respect to the uncertainties in the index.
- To explore methods and techniques for forecasting the index as it could be used to forecast agricultural droughts and compare those approaches in terms of forecasting efficiency.
- To test the ability of the index, as a generic index, to estimate yield loss from drought for peanut, cotton, soybean, and maize through exploring associations between the index and crop yields.

Dissertation Organization

This dissertation is organized around the five objectives set forth above. Chapter 2 describes the formulation of a new drought index called the Agricultural Reference Index for Drought (ARID), followed by an evaluation of the index by comparing it with other drought indices that have been widely used for agricultural purpose and with the water deficit index computed by a widely-used crop growth model, DSSAT CERES-Maize. Chapter 3 presents the uncertainties in ARID associated with the uncertainties in its parameters and the sensitivity of the index to its parameters. Chapter 4 describes various approaches and statistical techniques that could be used to forecast ARID as a predictor of agricultural droughts and then compares those approaches and techniques in terms of modeling efficiency. Chapter 5 presents a general ARID-yield relationship, ARID-based crop-specific yield models for major rainfed crops grown in the southeast USA, and the performance of ARID to estimate crop yields. Finally, Appendix A, a

byproduct of this dissertation, introduces the Lawn and Garden Moisture Index, a drought index that is being used to monitor and forecast droughts in the southeast USA, and discusses its performance across various seasons and locations in the region.

Table 1-1. Drought indices that have been reported to quantify agricultural drought

Abbr.	Drought Indicator	Author(s)	Time-step	S[a]	P[b]	A[c]	G[d]
CMI	Crop Moisture Index	Palmer (1968)	Weekly	Y[e]	N[f]	Y	Y
CSDI	Crop Specific Drought Index	Meyer et al. (1993a)	Seasonal	Y	Y	Y	N
CWSI	Crop Water Stress Index	Idso et al. (1981a)	Instant[g]	N	N	Y	N
ETDI	Evapotranspiration Deficit Index	Narasimhan and Srinivavan (2005)	Weekly	Y	Y	Y	N
KBDI	Keetch and Byram Drought Index	Keetch and Byram (1968)	Daily	N	N	Y	Y
LGMI	Lawn and Garden Moisture Index	Christy (2004)	Daily	N	N	Y	Y
PDSI	Palmer Drought Severity Index	Palmer (1965)	Monthly	Y	N	Y	Y
PZI	Palmer Z-index	Palmer (1965)	Monthly	Y	N	Y	Y
SMDI	Soil Moisture Deficit Index	Narasimhan and Srinivasan (2005)	Weekly	Y	Y	Y	N
SDI	Stress Day Index	Hiler and Clark (1971)	Daily	Y	Y	Y	N
SPI	Standardized Precipitation Index	McKee et al. (1993)	Monthly	N	N	Y	Y
WRSI	Water Requirement Satisfaction Index	Frere and Popov (1986)	Dekadal	Y	Y	Y	N
WSPD	Water stress index of the DSSAT[h]	Jones et al. (20003)	Daily	Y	Y	Y	N

[a] S = soil water balance-based; [b] P = plant physiology-based; [c] A = atmosphere-based; [d] G = generic, not requiring crop-specific parameters; [e] Y = yes; [f] N = no; [g] Instant = instantaneous, based on the vapor pressure deficit of a transpiring crop; [h] Based on Decision Support System for Agrotechnology Transfer (DSSAT) CERES-Maize model.

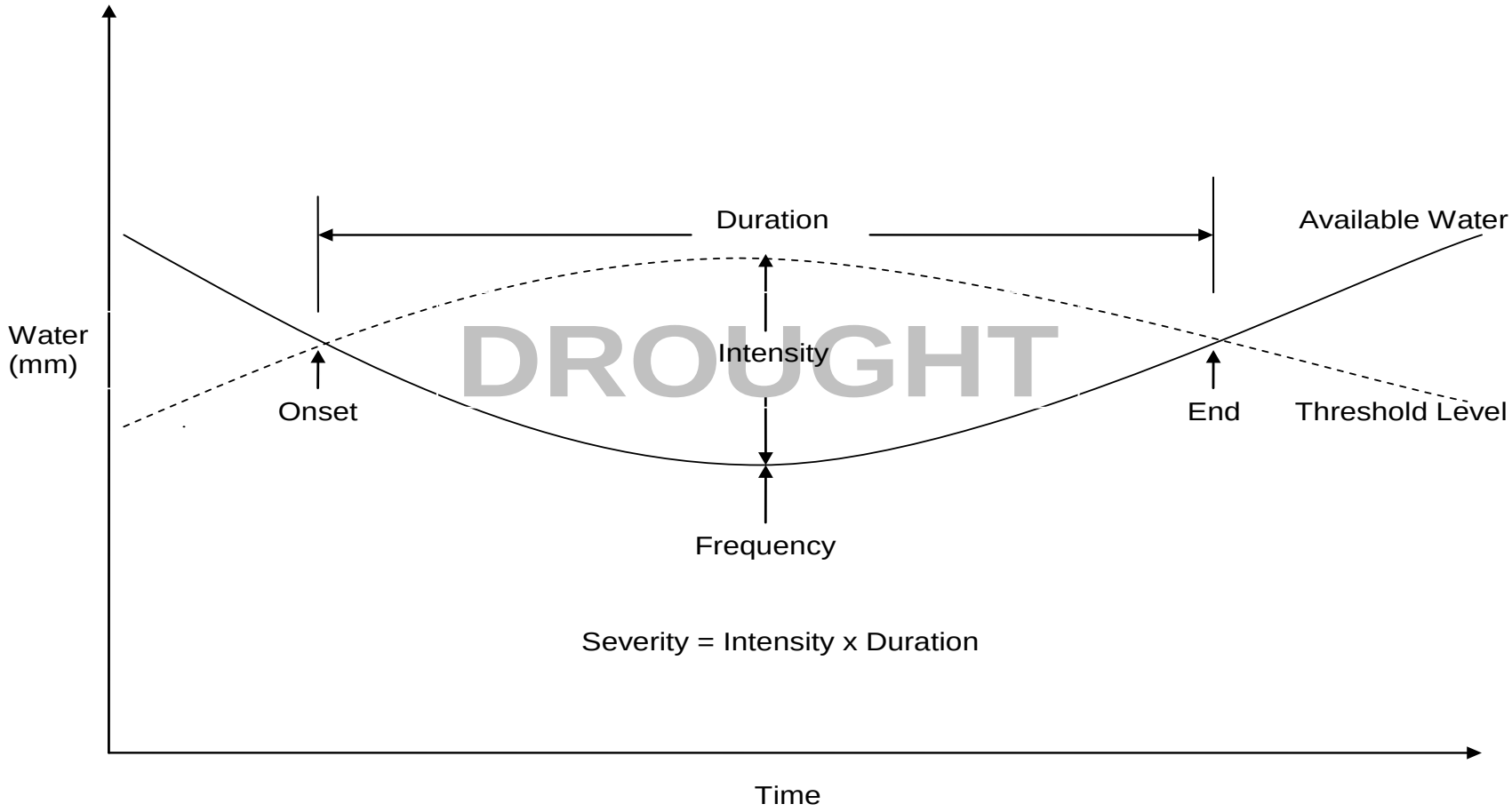

Figure 1-1. The schematic representation of drought and its components. The solid line indicates the amount of available water, whereas the dashed line shows the threshold level. Drought occurs when the threshold level is greater than the available water. The greater the deficit of available water relative to the threshold, the greater the intensity of drought.

CHAPTER 2
DEVELOPING AGRICULTURAL REFERENCE INDEX FOR DROUGHT

Introduction

Simplicity, genericness, soil-plant-atmosphere-basis, and computability at high temporal resolution are basic requirements of an agricultural drought index. The index should be able to quantify crop water deficit and the effect of the deficit on crop yields. However, no drought index was found that could meet all these requirements (Table 1-1). So, a new agricultural drought index – Agricultural Reference Index for Drought (ARID) – was developed with all these features. The purpose of ARID is to help crop producers or those involved in crop production or related fields to quantify the status of plant water deficit relative to supply and demand for water and ultimately to estimate yield loss from drought. The anticipation of drought-induced yield loss may help users minimize the losses through taking proactive measures and setting out drought mitigation strategies.

ARID is based on a reference crop, grass, that is actively growing in a well-drained, medium-textured soil and completely covers the soil surface. ARID uses a simple soil water balance for the reference grass having a 400 mm soil layer with evenly distributed roots. A single soil layer was chosen because single-layer models require few assumptions about the root zone, adequately estimate soil water status for established plants, and are relatively simple to derive and apply (Porteous et al., 1994; Eilers et al., 2007). Water from irrigation and the precipitation that remains on the soil surface after runoff enters the root-zone and exits it through transpiration, and drainage (Figure 2-1). Soil evaporation and upward flow of water from deeper layers are assumed to be insignificant. Considering complexities associated with distributed

models (Arora et al., 1987; Yokoo et al., 2001) and the dominance of vertical gradients of soil water potential as the driving force of water movement (Wheater and Sherratt, 1983), the lumped modeling approach was adopted for ARID. Although the effective rooting depth of grass depends on several factors, a 400 mm deep root-zone is reasonable (Wu, 1985; De Jong and Kabat, 1990; Haman et al., 2008; Zazueta et al., 2008).

Materials and Methods

ARID Formulation

Because transpiration is directly related to stomatal opening, which in turn, affects carbon dioxide assimilation, the amount of biomass or marketable yield produced can be estimated by the amount of water transpired during the growing season (de Wit, 1958). Plant water deficit occurs when there is insufficient water in the root zone to meet plant needs, that is, when transpiration is less than potential transpiration:

$$\text{Deficit} = TR_p - TR \tag{2-1}$$

where TRp is the potential transpiration (mm d-1), and TR is transpiration (mm d-1). The definitions of major variables and parameters of ARID are presented in Table 2-1.

An index can be used to quantify the water stress experienced by a crop on a particular day or during the growing season due to water deficit (Hiler et al., 1974). A plant water deficit index can be expressed as the ratio of water deficit to water need (Thornthwaite, 1948). Defining need as potential transpiration (TR_p), the water deficit index (WDI) can be expressed as:

$$WDI = \frac{\text{Deficit}}{\text{Need}} = \frac{TR_p - TR}{TR_p} = 1 - \frac{TR}{TR_p} \tag{2-2}$$

Based on the approach adopted to estimate TR for Equation 2-2, crop water deficit indices can be categorized into two groups: surface energy balance-based and soil water balance-based. The energy balance approach requires many variables that are not readily available or parameters that are hard to calibrate (Bouwer et al., 2008). For instance, a surface energy balance approach uses difficult-to-measure windspeed and canopy and aerodynamic resistances (Jalali-Farahani et al., 1993; Yazar et al., 1999). Moreover, this approach has limited application because of the need for sensors, which have fixed spectral bands, coarse spatial resolutions, inadequate repeat coverage, and long acquisition to delivery period (Moran, 1994; Moran et al., 1997). Similarly, dynamic or semi-dynamic water transport-based water balance models require a wide range of data that are not readily available and parameters that are difficult to calibrate (Wheater and Sherratt, 1983; Rao, 1987; Arora and Gajri, 1996; Sanchez-Cohen et al., 1997; Verdoodt et al., 2005; Raes et al., 2006). Considering the problems associated with energy balance models as well as physically-based dynamic soil water balance models, a simple budget-based soil water balance approach was adopted in ARID. The budget-based water balance approach is relatively simple and requires minimal inputs (de Jong and Bootsma, 1996) and few parameters, which are easy to estimate (Panigrahi and Panda, 2003).

ARID is generic due to its use of a reference crop that only requires site-specific weather data. Grass was chosen as the reference crop for the following reasons:

- Grass has been used previously as a reference crop to compute crop reference ET (ET_o) (Allen et al., 1998). ET_o is equal to the potential ET of the reference grass (Allen et al., 2005).
- Grass is assumed to be perennial, thus allowing ET_o to be computed throughout the year. A specific crop, on the other hand, is limited to a particular period of the

year, thus allowing ET_o to be computed only for that period. To compute ET_o for the entire year, a number of crops are needed, which makes ARID not only crop-specific but also complex.

- Grass that is well established typically is dense and has a uniform canopy that mostly covers the soil surface.

- TR and TR_p in Equation 2-2 vary with crops and their phenological stages. For a generic index, transpiration needs to be independent of crops and stages. Because grass is an indeterminate crop and has a uniform canopy throughout the growing season, this independency may be achieved by using grass.

- To estimate ET_p for a crop at a particular growth stage, ET_o needs to be used along with crop coefficients. If a crop other than a grass is used, several dynamic values of crop coefficients would be needed, which would make ARID crop-specific and increase the complexity of computations. Because the ET_p of grass is ET_o, grass does not need crop coefficients to estimate its ET_p. So using grass not only makes the index generic but also simple.

For the reference grass, which is dense and actively growing (Allen et al., 2005), transpiration may be assumed to be approximately equal to evapotranspiration (Ritchie and Burnett (1971). So, replacing TR_p by ET_o and water deficit index (WDI) by ARID in Equation 2-2 produces the following relationship:

$$ARID = 1 - \frac{TR}{ET_o} \tag{2-3}$$

When there is no transpiration, ARID equals one, the maximum water deficit. When transpiration occurs at the potential rate, ARID is zero, indicating no water deficit. Between these two extremes, water deficit decreases linearly with an increase in actual transpiration when ET_o is constant. Considering the time scale of crop responses to water deficits and daily variations in weather data (Allen et al., 1998; Steduto et al., 2009), ARID is computed daily.

Reference ET

The FAO-56 Penman-Monteith model (Allen et al., 1998) is used to estimate ET_o in ARID. This method has been recommended as the sole method of calculating ET_o (Allen et al., 1998). The Penman-Monteith equation, which estimates ET realistically from various surfaces, climatic conditions, and at different scales (Jensen et al., 1990; Fontenot, 1999; Tamm, 2002), is as follows:

$$ET_o = \frac{0.408\Delta(R_N - G) + \gamma\left(\frac{900}{T+273}\right)\mu_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34\mu_2)} \quad (2\text{-}4)$$

where ET_o is the reference ET (mm d^{-1}), Δ is the slope of vapor pressure curve (kPa $°C^{-1}$), RN is the net radiation at the grass surface (MJ m^2 d^{-1}), G is the soil heat flux density (MJ m^2 d^{-1}), γ is the psychrometric constant (kPa $°C^{-1}$), T is average daily temperature (°C), μ_2 is the wind speed at 2 m height (m s^{-1}), e_s is saturated vapor pressure (kPa), and e_a is the ambient vapor pressure (kPa).

Transpiration

The rate of transpiration is affected by ET_o and the ability of roots to extract soil water. Annual plants transpire almost all water absorbed by their roots (Jensen, 1968; Hopkins, 1999; Ingram, 2005), thus transpiration is approximately equal to root water uptake (Wu et al., 1999). To estimate transpiration, ARID uses a macroscopic modeling approach – treating the entire root system as a single unit and assuming the removal of water from the root zone as a whole. Such macroscopic models need soil and plant parameters that are readily available (Li et al., 2006) and do not require complete representation of all physical processes of water uptake, thus eliminating the need for difficult-to-obtain soil and plant parameters (Li et al., 2001). Macroscopic models are

basically of two types: water potential-based and water content-based. Because the water potential-based approach requires detailed information about root and soil properties that are difficult, time-consuming, and costly to evaluate (Taylor and Klepper, 1973; Zuo et al., 2006), ARID is based on the water content approach.

A simple root water uptake model that is not based on root length density, but based on generalizations from measured soil water, might overcome several limitations of more complex physically-based models (Dardanelli et al., 2004). Based on the exponential decay relation described by Passioura (1983), Meinke et al. (1993) developed the following simple root water uptake model:

$$RWU = \alpha\theta_a^{ad} \tag{2-5}$$

where RWU is the soil-limited root water uptake rate (mm d^{-1}), α is a generic root water uptake constant, representing the maximum fraction of available water extracted in a day, and θ_a^{ad} is the plant available water content after deep-drainage at the end of the previous day (mm mm^{-1}).

Several researchers, such as Robertson et al. (1993a), Robertson et al. (1993b), Dardanelli et al. (1997), and Dardanelli et al. (2004), have applied the water uptake model of Meinke et al. (1993). Equation 2-5 produces an exponential decline in soil water content over time. Dardanelli et al. (2004) found 0.096 to be a good approximation for α in a wide range of crops and soil types when roots are fully developed and uniformly distributed in the soil. Using the relations suggested by Passioura (1983) and Meinke et al. (1993), transpiration (mm d^{-1}) can be estimated as:

$$TR = \min\left\{\alpha\zeta\theta_a^{ad}, TR_p\right\} \tag{2-6}$$

where ζ is the root zone depth (400 mm) and the other variables as defined previously.

The use of TR_p in Equation 2-6 is because transpiration occurs at the potential rate under non water-limiting conditions, whereas Equation 2-5 is for water-limited conditions. Assuming that TR_p approximately equals the potential ET of the reference grass (ET_o), which is dense and actively growing, the actual transpiration in ARID is estimated as:

$$TR = \min\left\{\alpha\zeta\theta_a^{ad}, ET_o\right\} \quad (2\text{-}7)$$

Soil water balance

To estimate daily TR using Equation 2-7, the available soil water content in the root zone (θ_a^{ad}) has to be known. The θ_a^{ad} in a specific day is calculated by dividing the amount of available water in the root zone (W, mm) in that day by the root zone depth: $\theta_a^{ad} = W/\zeta$. ARID computes W as:

$$W_i = W_{i-1} + P_i + I_i - TR_i - D_i - R_i \quad (2\text{-}8)$$

where W_i and W_{i-1} are the amount of available water (mm) in the root zone in day 'i' and 'i-1', respectively; and P_i, I_i, TR_i, D_i, and R_i are precipitation, irrigation, transpiration, deep drainage, and surface runoff, respectively, occurred in day 'i' (mm). Whereas precipitation and irrigation are inputs to the model (Equation 2-8); the model computes transpiration, drainage, and surface runoff.

Deep-drainage

Water is subject to deep-drainage from the root zone only when the soil water content exceeds field capacity (Williams et al., 1991; Ritchie, 1998). In other words, water is drained when available water content is greater than a threshold value, called the available water capacity (θ_m). Drained water is considered to be lost from the system. ARID estimates deep-drainage as:

$$D = \begin{cases} \beta\zeta\left(\theta_a^{bd} - \theta_m\right) & \text{if } \theta_a^{bd} > \theta_m \\ 0 & \text{else} \end{cases} \quad (2\text{-}9)$$

where D is the amount of water drained (mm d^{-1}), β is the drainage coefficient – the fraction drainable water that can be drained in a day, θ_a^{bd} is the available soil water content in the previous day before drainage (mm mm^{-1}), and θ_m is the available water capacity (mm mm^{-1}).

Values of β as well as θ_m may vary with soil type. As a generic indicator of crop water status, however, ARID uses a constant value for β and θ_m each. For simplicity, ARID assumes β to be 0.55, which Ritchie (1998) found to be suitable for well drained medium textured soils. Similarly, ARID uses 0.13 mm mm^{-1} for θ_m because the water holding capacities of many soil types, except the ones having more than 65% sand, is about 0.13 mm mm^{-1} (Ratliff et al., 1983; Legowo, 1986; Ritchie et al., 1999). Although soils may vary considerably over space and their water holding capacities and rooting depths also could be very different, experience in using ARID will allow users to interpret the index values relative to their own soils. For instance, ARID, a medium textured soil-based index, would likely underestimate the true level of water deficit for a light soil. Similarly, for a heavy soil, ARID may overestimate the actual deficit. The sensitivity of ARID to differing water holding capacities is presented in Chapter 3.

Many soil water balances also take into account saturation for estimating drainage. That is, they drain the available water between field capacity and saturation point through unsaturated flow and that above saturation through saturated flow. In the case of ARID, however, the saturation component is not included in the deep-drainage model because the index is based on a well drained soil with $\beta = 0.55$. The inclusion of

saturation component actually does not make any significant effect in ARID, but increases the complexity in the index, instead.

Surface runoff

To estimate surface runoff, ARID uses the runoff curve number method developed the USDA Natural Resources Conservation Service (NRCS), formerly known as the Soil Conservation Service (SCS). The SCS runoff curve number method is computationally efficient and soil and weather data inputs are generally available (Williams et al., 1991). This method is based on the assumption that runoff occurs when the rate of rainfall is greater than that of infiltration. Runoff takes place after the initial demands of interception, infiltration, and surface storage have been satisfied. Using the USDA-SCS (1972) method, the daily runoff is estimated as:

$$R = \begin{cases} \dfrac{(P - I_a)^2}{(P - I_a) + S} & \text{if} \quad P > I_a \\ 0 & \text{else} \end{cases} \tag{2-10}$$

where R is surface runoff (mm d^{-1}); P is precipitation (mm d^{-1}); I_a is initial abstraction, including interception, retention, and infiltration (mm d^{-1}); and S is the potential maximum retention (mm d^{-1}), which is computed as $S = 25400/\eta - 254$ (where η is the runoff curve number).

Using 65 as the reference η for a meadow of well-managed, protected, and frequently mowed continuous grass (similar to the reference grass) growing in a soil that has an average antecedent moisture condition and falls in between the hydrologic groups B and C (USDA-NRCS, 1986) and calculating I_a as 0.2S (USDA-SCS, 1972), the

amount of initial abstraction is estimated to be 27.4 mm. Thus, no runoff occurs unless precipitation exceeds this amount.

ARID Evaluation

Like the DSSAT crop simulation models, ARID is SPA-based; so it can also estimate water deficit and its effect on crop yield. However, ARID is much simpler than a crop simulation model because ARID requires fewer inputs and parameters than do crop models. It was hypothesized that ARID, a simple, generic index, might adequately represent the water deficit calculated by a crop model that is complex and crop-specific.

From available crop models, the DSSAT CERES-Maize (Jones and Kiniry, 1986; Jones et al., 2003; Hoogenboom et al., 2004) was chosen because it has been widely-tested and used (Hodges et al., 1987; O'Neal et al., 2002; Asadi and Clemente, 2003; de Faria and Bowen, 2003; Lopez-Cedron, 2005). The DSSAT CERES-Maize model assumes that when root water uptake is unable to meet transpirational demand of the foliage, photosynthesis and transpiration are reduced in direct proportion to decreased water uptake. This phenomenon is expressed in terms of an index called the water deficit index for photosynthesis (WSPD). ARID was evaluated in terms of how accurately it could mimic the WSPD. The comparison of ARID with a crop simulation model can only be done during the growing season of a crop.

Also, ARID and WSPD were compared with seven other indices that have been used to quantify agricultural drought, namely Crop Moisture Index (CMI), Keetch-Byram Drought Index (KBDI), Lawn and Garden Moisture Index (LGMI), Palmer Drought Severity Index (PDSI), Palmer Z-index (PZI), Standardized Precipitation Index (SPI), and Water Requirement Satisfaction Index (WRSI). The purpose of this comparison was to find out how well each index would approximate WSPD and thus quantify crop water

deficit, given the differences in their derivations and input requirements. The eight indices were also compared among themselves in terms of the response to fluctuations in precipitation over time.

Comparisons of indices with WSPD

The ability of each of the eight drought indices to quantify crop water deficit was assessed in terms of its correlation with WSPD and the root mean squared error (RMSE). While correlation (r) was used to explore the strength and the direction of association between a drought index and WSPD, RMSE was used to assess the error associated with estimating WSPD. Values of r would show the closeness of the points to the regression line, whereas those of RMSE would depict the average proximity of the points to the line of model perfection (1:1 line). So, a larger correlation with a smaller RMSE value would imply a better performance than would any of the other combinations of r and RMSE values, such as a smaller r with a larger RMSE, a larger r with a larger RMSE, a smaller r with a larger RMSE, and so on.

Because WSPD values of DSSAT are crop-specific, maize was chosen as the study crop because it is widely grown across the southeast USA under rain-fed conditions and also is simulated well by the DSSAT CERES-Maize model (Hodges et al., 1987; O'Neal et al., 2002; Asadi and Clemente, 2003; de Faria and Bowen, 2003; Lopez-Cedron, 2005). As ARID is a grass-based index, a perennial grass could also be used as the study crop. However, grass was not used because the DSSAT grass models have not been as widely-tested and used as has been the CERES-Maize. Moreover, the objective was to explore how accurately ARID, a generic index, could estimate water deficit for a crop other than the grass.

Sixteen locations were selected in the southeast USA: Huntsville, AL; Montgomery, AL; Birmingham, AL; Mobile, AL; Rome, GA; Midville, GA; Plains, GA; Griffin, GA; Blairsville, GA; Folkston, GA; Elberton, GA; Live Oak, FL; Citra, FL; La Belle, FL; Bartow, FL; and Miami, FL (Figure 2-2). Sites were selected to give a range of geographical locations and were based on availability of historical weather data. All locations had at least 50 years of weather data, including precipitation, maximum and minimum temperatures, windspeed, and dewpoint temperature. The other criterion considered for site selection was the availability of KBDI values computed by the Florida Climate Center.

Daily historical weather data were obtained from various sources: The Florida Climate Center (ftp://coaps.fsu.edu/pub/griffin/SECC/AgroClimate/raw/ascii/) for locations in Florida, the Georgia Automated Environmental Monitoring Network (G. Hoogenboom, personal communication, 9 March 2009) for locations in Georgia, and the National Climatic Data Center (http://www.ncdc.noaa.gov/oa/climate/stationlocator.html) for locations in Alabama. Values of KBDI for all locations were obtained from the Florida Climate Center (ftp://coaps.fsu.edu/pub/griffin/SECC/Research/KBDI/).

To calculate values of WSPD, a crop management file for maize and a weather file were created for each location. Taking into account the effect of latitude on maize growing season, different planting dates were used for different locations as suggested by Bauske et al. (1996), Lee (2007), and Wright et al. (2008) (Table 2-3). For simplicity, only one planting date for each location was used. Even with multiple planting dates, the drought indices would behave similarly because the inputs (weather variables) for each of the drought indices would be the same. For each location and season, plant density

was 5 plants m^{-2} and the cultivar was McCurdy 84aa. After creating all necessary files and entering necessary inputs, DSSAT was run with the not irrigated option, thus calculating daily values of WSPD during a crop season for each year (from year 1958 through 2005) and location. Then, periodic values of WSPD were computed by averaging the daily values for the fully established (dense) period of the crop, which was assumed to comprise of the last three months of the season. This period, here in after, is referred to as the full canopy period. For comparisons, the periodic values were used because crop water requirement is generally largest and the crop is most sensitive to water deficit during the full canopy period.

Similarly, the periodic values of each drought index were computed for the full canopy period of maize by averaging the daily, weekly, or monthly values of the indices computed. The time and duration of the full canopy period for each season and location were estimated by the DSSAT CERES-Maize model while computing WSPD. Monthly values of PDSI, PZI, and SPI and weekly values of CMI were computed using PDSI and SPI software developed by the National Agriculture Decision Support System (NADSS), University of Nebraska, Lincoln (Wells, 2003; NADSS, 2008). Before computing monthly or weekly values of these indices, daily weather data were converted to mean monthly or weekly values as appropriate. For ARID, LGMI, and WRSI, daily values were computed from daily historical weather data.

Because different drought indices had different ranges, values of all drought indices were transformed so that all the indices would have the same numerical range as WSPD (0.0 to 1.0). After transformation, 0 would mean no stress and 1 would mean maximum stress in each drought index. Transformations were carried out as follows:

$$CMI_T = \begin{cases} CMI/(-4) & \text{if } CMI < 0 \\ 0 & \text{else} \end{cases} \quad (2\text{-}11)$$

$$KBDI_T = \frac{KBDI}{800} \quad (2\text{-}12)$$

$$LGMI_T = - \frac{\begin{cases} LGMI & \text{if } LGMI < 0 \\ 0 & \text{else} \end{cases}}{LET} \quad (2\text{-}13)$$

$$PDSI_T = \begin{cases} PDSI/(-5) & \text{if } PDSI < 0 \\ 0 & \text{else} \end{cases} \quad (2\text{-}14)$$

$$PZI_T = \begin{cases} PZI/(-4) & \text{if } PZI < 0 \\ 0 & \text{else} \end{cases} \quad (2\text{-}15)$$

$$SPI_T = \begin{cases} SPI/(-3) & \text{if } SPI < 0 \\ 0 & \text{else} \end{cases} \quad (2\text{-}16)$$

$$WRSI_T = 1 - \frac{WRSI}{100} \quad (2\text{-}17)$$

where the subscript T denotes the transformed index, and LET is the ET function of LGMI, whose values are dynamic. Each index was computed before transformation. To transform LGMI computed for a particular day, the index was divided by LET of that particular day. The negative denominators in many equations are based on the minimum possible values of the corresponding indices. For comparisons, the periodic values of these transformed indices were used.

Comparisons of indices against precipitation

To explore the temporal variability and the differential response of each drought index to fluctuations in precipitation, the time series of each index computed with its time-step, that is, monthly values of PDSI, PZI, and SPI, weekly values of CMI, and daily values of ARID, KBDI, LGMI, and WRSI, were compared with daily values of

precipitation. For the comparisons, one location was chosen: Huntsville/Madison County, one of the major corn growing regions in Alabama. Using historical weather data of this location over 3.5 years (1998 January through 2001 July), values of the indices were computed. The period of 1998-2001 was chosen because it included one of the worst droughts in this region. For exploring the dynamic response of a drought index to precipitation fluctuations, the dataset of 3.5 years' historical weather data of one location was considered enough as it included both short- and long-term wet and dry periods. This dataset was used only for comparing the indices against precipitation. For comparing the indices with WSPD, however, the whole dataset of 48 years (1958 through 2005) was used.

Results and Discussion

Comparisons of Indices with WSPD

ARID vs. WSPD

Of the eight drought indices compared, ARID had the largest correlation with WSPD, the water deficit factor calculated by the DSSAT CERES-Maize model, for the period when the canopy fully covered the soil for all 16 locations (Table 2-3). Similarly, RMSE values between ARID and WSPD were the smallest in all locations (Table 2-4). Although ARID differs from WSPD in that the former adopts a much simpler approach for computing soil water balance than does the latter, they are based on similar soil water balance approaches and use similar methods for estimating surface runoff, infiltration, deep-drainage, and potential ET. Both ARID and WSPD are SPA-based and incorporate a ratio of actual to potential transpiration (Equation 2-2), which may explain the good correlations between ARID and WSPD.

Although the correlations between ARID and WSPD were good, values of ARID were mostly greater than those of WSPD (Figure 2-3). ARID estimated comparatively more water deficit because the soil in ARID is shallow (40 cm) compared with soils used in the DSSAT CERES-Maize simulations and thus had less available water in the root zone. For WSPD, however, the root zone could go as deep as 215 cm depending upon the depth of the profile at each location – Birmingham: 168 cm, Bartow: 203 cm, Blairsville: 178 cm, Citra: 183 cm, Elberton: 107 cm, Folkston: 203 cm; Griffin: 107 cm, Huntsville: 152 cm, La Belle: 171 cm, Live Oak: 200 cm, Miami: 90 cm, Montgomery: 215 cm, Mobile: 173 cm, Midville: 165 cm, Plains: 183 cm, and Rome: 140 cm.

WRSI vs. WSPD and ARID

The WRSI showed relatively good correlations with WSPD for some locations (Table 2-3), which could be because, like WSPD or ARID, WRSI is based in part on the ratio of actual to potential transpiration. However, the correlation between WRSI and WSPD was not as strong as that of ARID. One of the possible reasons for the weaker association between WRSI and WSPD was the use of arbitrary values for a WRSI variable called critical soil water content, above which ET is assumed to take place at a potential rate and below which ET rate depends on soil water content (Verdin and Klaver, 2002). For WSPD, however, ET is calculated based on daily simulated values of leaf area index and root length density. Another reason for the weaker association was that WRSI considers all precipitation to be available to the plant without considering losses as run-off or drainage. WSPD, on the other hand, considers precipitation as available to the plant only after accounting for losses. This consideration resulted in relatively more available water in the soil and WRSI values mostly remained smaller than those of WSPD (Figure 2-4A) and ARID (Figure 2-4B).

LGMI vs. WSPD and ARID

Although the correlations between WSPD and LGMI were not as large as those between WSPD and ARID, LGMI showed relatively large correlations with WSPD in many locations (Table 2-3). LGMI is not explicitly based on plant transpiration and soil water storage, yet it accounts for seasonal trends in ET by using the concept of effective precipitation – rainfall that is sufficient to offset ET and maintain soil water above the wilting point. Similarly, LGMI considers soil water losses by assuming that all precipitation that occurred in the previous week is available to plants while that from 1 to 3 weeks previously declines linearly in availability.

Although LGMI correlated relatively well with WSPD, its values were larger than those of WSPD (Figure 2-4C) and ARID (Figure 2-4D) for all locations and times. The larger stress estimated by LGMI was mainly because of its ET function, whose estimations are always larger than those of other methods, such as the FAO-56 Penman-Monteith (see Appendix A for details).

KBDI vs. WSPD and ARID

Generally, the correlation of KBDI with WSPD was not as good as were those of ARID, LGMI, or WRSI (Table 2-3). The weak correlation was likely because soil and plant processes such as root water uptake and soil water balance are not considered and also because KBDI was developed for drought status representing wildfire threat to forests, not for field crops. The index is not wholly SPA-based, but it does take into account atmospheric variables such as daily maximum temperature and daily and annual mean precipitation. Values of KBDI were mostly larger than WSPD (Figure 2-4E) and ARID (Figure 2-4F). The larger values were likely due to smaller estimates of available water caused by canopy interception of some precipitation and larger

estimates of ET caused by large daily maximum temperatures and small mean annual rainfall. The KBDI assumes that 5 mm of rainfall in every event is intercepted by the tree canopy, whereas WSPD and ARID assume that intercepted rainfall is negligible. KBDI assumes evaporation rate to be an exponential function of daily maximum temperature and transpiration rate to be an inverse exponential function of mean annual rainfall. Because the maize growing season in the southeast has generally larger maximum temperatures and smaller mean annual rainfall, KBDI values were likely to be greater than those of WSPD or ARID during this season.

PDSI vs. WSPD and ARID

Compared with other drought indices, correlations of PDSI with WSPD were small because the former is based more on weather variables and less on soil and plant processes (Table 2-3). The PDSI is more useful as a meteorological than an agricultural drought index because it is based on precipitation departure from the long-term mean and soil water anomaly. The PDSI indicates only long-term droughts by calculating severe or abnormal departures, whereas WSPD and ARID reflect short-term droughts by estimating water deficit based on the demand and supply, taking into consideration several weather variables as well as soil and plant processes. Because values of PDSI are drought severity values, they remained mostly smaller than those of WSPD (Figure 2-4G) and ARID (Figure 2-4H).

CMI vs. WSPD and ARID

Similarly, CMI had a relatively small correlation with WSPD for all locations (Table 2-3, Figure 2-4I). The CMI is an index of ET anomaly, which is a measure of the cumulative severity of an agricultural drought. Both PDSI and CMI are the measures of severity of drought and belong to the same family of Palmer Indices. Whereas PDSI

quantifies the severity of a meteorological drought, CMI quantifies the severity of an agricultural drought. Because CMI is a measure of abnormally deficient ET, it shows only severe water deficit. WSPD and ARID, on the other hand, are measures of water deficit based on the demand for and supply of water for crops; so they show a wide range of water deficits, mild or severe. Because of this difference, a small correlation was found between CMI and WSPD or ARID. Because CMI shows only severe droughts, values of CMI were less frequent and always smaller than those of WSPD (Figure 2-4I) and ARID (Figure 2-4J).

PZI and SPI vs. WSPD and ARID

The PZI and SPI are meteorological drought indices based on precipitation anomalies. Because these indices are not SPA-based, they were both weakly correlated with WSPD (Table 2-3). These indices depend more on precipitation anomaly than on ET, and they generally estimate more available water and thus less water deficit relative to WSPD (Figure 2-4K and Figure 2-4L) and ARID (Figure 2-4M and Figure 2-4N).

Comparisons of Indices against Precipitation

Of the eight drought indices compared, CMI generally responded least to small rainfall amounts in many months (Figure 2-5B). However, its responses were similar to ARID in severe drought conditions (Figure 2-5C). For instance, CMI responded to the small amounts of precipitation during the summer of1998, 1999, and 2000 as did most of the other indices. CMI responded mostly only to severe water deficit because it is a measure of abnormally deficient ET and cumulates the severity of drought over time. Like CMI, values of PDSI were larger particularly when severe drought conditions existed (Figure 2-5A). The large response was likely because PDSI also is a measure of

abnormal droughts. The magnitude of water deficit shown by PDSI, PZI, and SPI each was larger than that shown by ARID for winter months but smaller for summer months (Figure 2-5A vs. Figure 2-5C). These seasonal differences were because PDSI, PZI, and SPI are mainly precipitation-based indices and not influenced by temperature, so their values fluctuate depending only on the amount of precipitation. ARID, on the other hand, is affected by temperature. In winter months, values of ARID were generally smaller than those of PDSI, PZI, and SPI because of less evaporative demand of the atmosphere. So, in the winter, even a small amount of precipitation fulfilled the water need of plants, thus leading to smaller values of ARID. Similarly, during summer, ARID was likely to have relatively larger values even with a significant amount of precipitation because of large ET losses. Precipitation-based indices did not account for these phenomena, so their values were solely governed by the amount of precipitation.

Although the overall magnitude of water deficit shown by KBDI for many occasions was about the same as that shown by ARID, the response of KBDI to precipitation fluctuations lagged by few days to a month as compared with that of ARID (Figure 6). The slow response of KBDI was probably because of its long memory. In fact, KBDI is affected significantly by past values as it is an index based on an accrual system. Generally, LGMI and WRSI responded to precipitation fluctuations in a fashion similar to ARID. However, LGMI always calculated more stress than ARID because it estimated greater ET losses. WRSI, on the other hand, mostly estimated less stress because it assumes greater amount of plant available water in the root zone (for instance, 68 mm for Huntsville) than assumed by ARID (52 mm). When the same value for water holding capacity was used, both indices estimated about the same magnitude of deficit (data

not shown). This similarity was mainly because both are SPA- and soil water balance-based and used the same method to calculate potential ET. Also, the water deficit estimated by both indices was for grass that completely covers the ground.

The PDSI showed droughts lasting more than 3 years, whereas the other indices indicated short-term droughts - monthly to seasonal (Figure 2-5). Of the daily time-step-based indices, the periods of droughts identified by KBDI and LGMI were longer by few days to a week than the period of droughts identified by ARID or WRSI (Figure 6). The identification of longer droughts by PDSI, KBDI, and LGMI was mainly due to their long memory. In fact, PDSI and KBDI are affected by their past values as these indices are based on an accrual system, and LGMI takes into account the precipitation that occurred during the previous 21 days. Droughts identified by WRSI were similar to those identified by ARID as both indices followed similar approaches and estimated water deficit for grass.

A good agricultural drought index should capture the effects of fluctuations in precipitation and temperature and reflect both short- and long-term droughts. Of the eight drought indices compared, the SPA-based indices, namely ARID and WRSI, were better than the other indices for these attributes. The weekly or monthly time-step-based indices, namely CMI, PDSI, PZI, and SPI, underestimated the magnitude of crop water deficit, especially during the winter and when drought is not so severe (Figure 2-5).

Conclusions

ARID is computationally simple, physically and physiologically sound, and generally applicable to characterize agricultural drought. The index approximated the water deficit values calculated by the DSSAT CERES-Maize model more accurately than did the seven other drought indices as the RMSE values between ARID and

WSPD were smaller than those comparing WSPD with all other drought indices by, on an average, 38%, the range being 1% to 83%. Similarly, the **r** values between ARID and WSPD were larger than those comparing WSPD with all other indices by, on an average, 41%, the range being 1% to 630%. In addition, ARID captured the effect of fluctuations in precipitation and temperature and estimated the magnitude of water deficit on crops more accurately than the other indices. Results suggest that ARID is a suitable index for characterizing drought for a crop that fully covers the soil.

Table 2-1. Definitions of variables and parameters of various functions used in ARID

Variable	Definition	Units
D	Deep-drainage	mm d^{-1}
ET	Evapotranspiration	mm d^{-1}
ET_p	Potential evapotranspiration	mm d^{-1}
ET_o	Grass reference evapotranspiration	mm d^{-1}
I	Irrigation	mm d^{-1}
P	Precipitation	mm d^{-1}
R	Surface run-off	mm d^{-1}
TR	Transpiration	mm d^{-1}
TR_p	Potential transpiration	mm d^{-1}
W	Amount of available water in the root zone	mm
α	Maximum fraction of available water extracted in a day	-
β	Fraction of free water in soil that can be drained in a day	-
θ_a	Available soil water content in a particular day	mm mm^{-1}
θ_m	Maximum plant available water capacity	mm mm^{-1}
ζ	Root zone depth	mm

Table 2-2. Planting dates of maize for various locations as used in the DSSAT CERES-Maize model

Location	Latitude (°N)	Longitude (°W)	Planting date
Miami, FL	25.77	80.19	March 01
La Belle, FL	26.75	81.44	March 10
Bartow. FL	27.90	81.84	March 20
Citra, FL	29.08	82.11	March 30
Live Oak, FL	30.28	82.98	April 10
Mobile, AL	30.68	88.04	April 15
Folkston, GA	30.73	82.02	April 15
Plains, GA	32.05	84.23	April 20
Montgomery, AL	32.30	86.30	April 20
Midville, GA	32.88	82.24	April 25
Griffin, GA	33.27	84.26	May 01
Birmingham, AL	33.57	86.80	May 05
Elberton, GA	34.15	82.87	May 10
Rome, GA	34.25	85.16	May 10
Huntsville, AL	34.65	86.62	May 15
Blairsville, GA	34.85	83.94	May 20

Table 2-3. Correlations between WSPD and the tested drought indices for several locations for the period when the maize canopy fully covered the soil

	Location															
Index	A[a]	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
ARID	0.91	0.93	0.91	0.94	0.90	0.76	0.94	0.86	0.73	0.87	0.92	0.85	0.92	0.93	0.95	0.84
CMI	0.60	0.61	0.56	0.63	0.71	0.70	0.77	0.70	0.63	0.67	0.76	0.71	0.74	0.68	0.75	0.71
KBDI	0.52	0.62	0.57	0.72	0.68	0.43	0.75	0.76	0.51	0.63	-	0.67	0.75	0.82	0.85	0.68
LGMI	0.87	0.84	0.85	0.88	0.83	0.65	0.88	0.84	0.65	0.79	0.86	0.77	0.83	0.87	0.90	0.66
PDSI	0.42	0.39	0.46	0.55	0.48	0.22	0.78	0.54	0.10	0.36	0.51	0.45	0.56	0.59	0.55	0.75
PZI	0.80	0.64	0.66	0.84	0.66	0.48	0.72	0.64	0.49	0.87	0.88	0.77	0.73	0.76	0.81	0.83
SPI	0.72	0.55	0.57	0.80	0.55	0.52	0.66	0.58	0.27	0.80	0.79	0.73	0.65	0.73	0.73	0.83
WRSI	0.90	0.84	0.77	0.87	0.80	0.50	0.84	0.68	0.54	0.70	0.89	0.64	0.85	0.82	0.91	0.71

[a] A = Miami, FL; B = La Belle, FL; C = Bartow, FL; D = Citra, FL; E = Live Oak, FL; F = Mobile, AL; G = Folkston, GA; H = Plains, GA; I = Montgomery, AL; J = Midville, GA; K = Griffin, GA; L = Birmingham, AL; M = Elberton, GA; N = Rome, GA; O = Huntsville, AL; P = Blairsville, GA.

Table 2-4. RMSE values between WSPD and the tested drought indices for several locations for the period when the maize canopy fully covered the soil

	Location															
Index	A[a]	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
ARID	0.16	0.19	0.21	0.19	0.17	0.35	0.16	0.09	0.13	0.16	0.12	0.10	0.14	0.06	0.06	0.12
CMI	0.24	0.23	0.22	0.21	0.19	0.39	0.21	0.25	0.23	0.21	0.22	0.20	0.20	0.27	0.35	0.21
KBDI	0.31	0.28	0.33	0.31	0.31	0.52	0.31	0.24	0.30	0.28	N/A	0.31	0.32	0.24	0.18	0.29
LGMI	0.29	0.24	0.25	0.32	0.20	0.47	0.30	0.20	0.28	0.24	0.30	0.32	0.34	0.28	0.20	0.30
PDSI	0.30	0.28	0.25	0.28	0.25	0.37	0.25	0.22	0.19	0.21	0.19	0.16	0.22	0.22	0.28	0.22
PZI	0.28	0.27	0.27	0.24	0.22	0.50	0.28	0.14	0.19	0.18	0.16	0.15	0.19	0.15	0.26	0.26
SPI	0.29	0.26	0.28	0.25	0.21	0.51	0.28	0.21	0.22	0.20	0.16	0.16	0.19	0.19	0.30	0.27
WRSI	0.20	0.22	0.21	0.23	0.18	0.38	0.19	0.10	0.15	0.20	0.19	0.12	0.18	0.15	0.09	0.17

[a] Locations as listed in the footnote to Table 2-3.

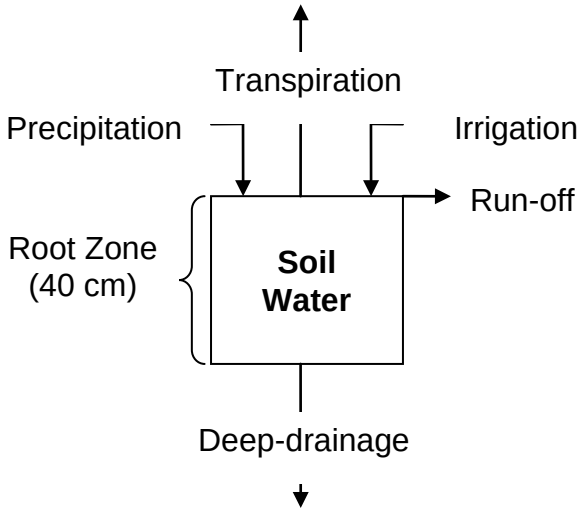

Figure 2-1. Diagram for the reference crop grass soil water balance

Figure 2-2. Sixteen locations selected for the study

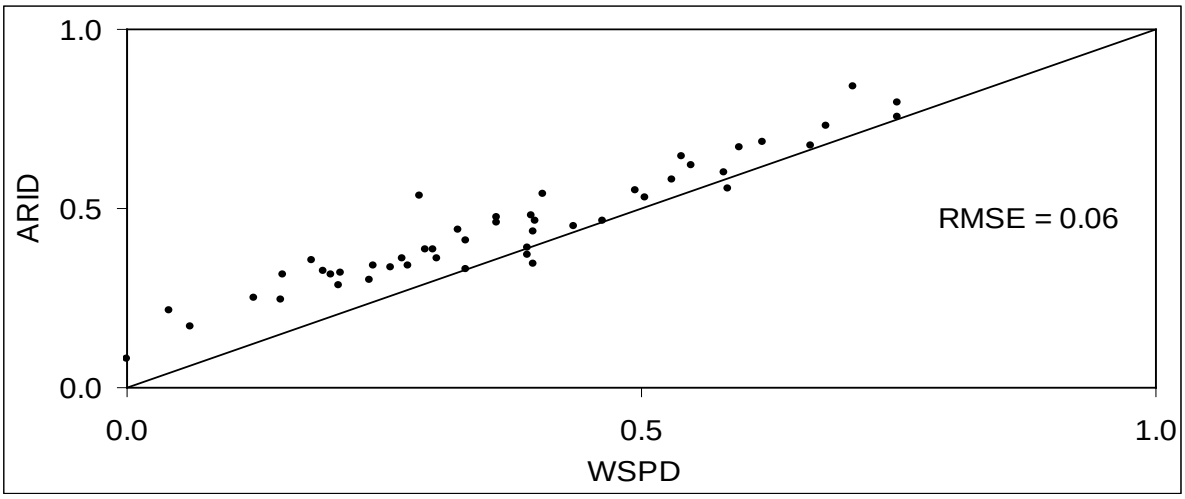

Figure 2-3. Association of ARID and WSPD for Huntsville/Madison County, a corn growing region in Alabama, over a period of 48 years (1958 through 2005). Index values are means for the period when the maize canopy fully covered the soil (3 months), and the solid line is the 1:1 line.

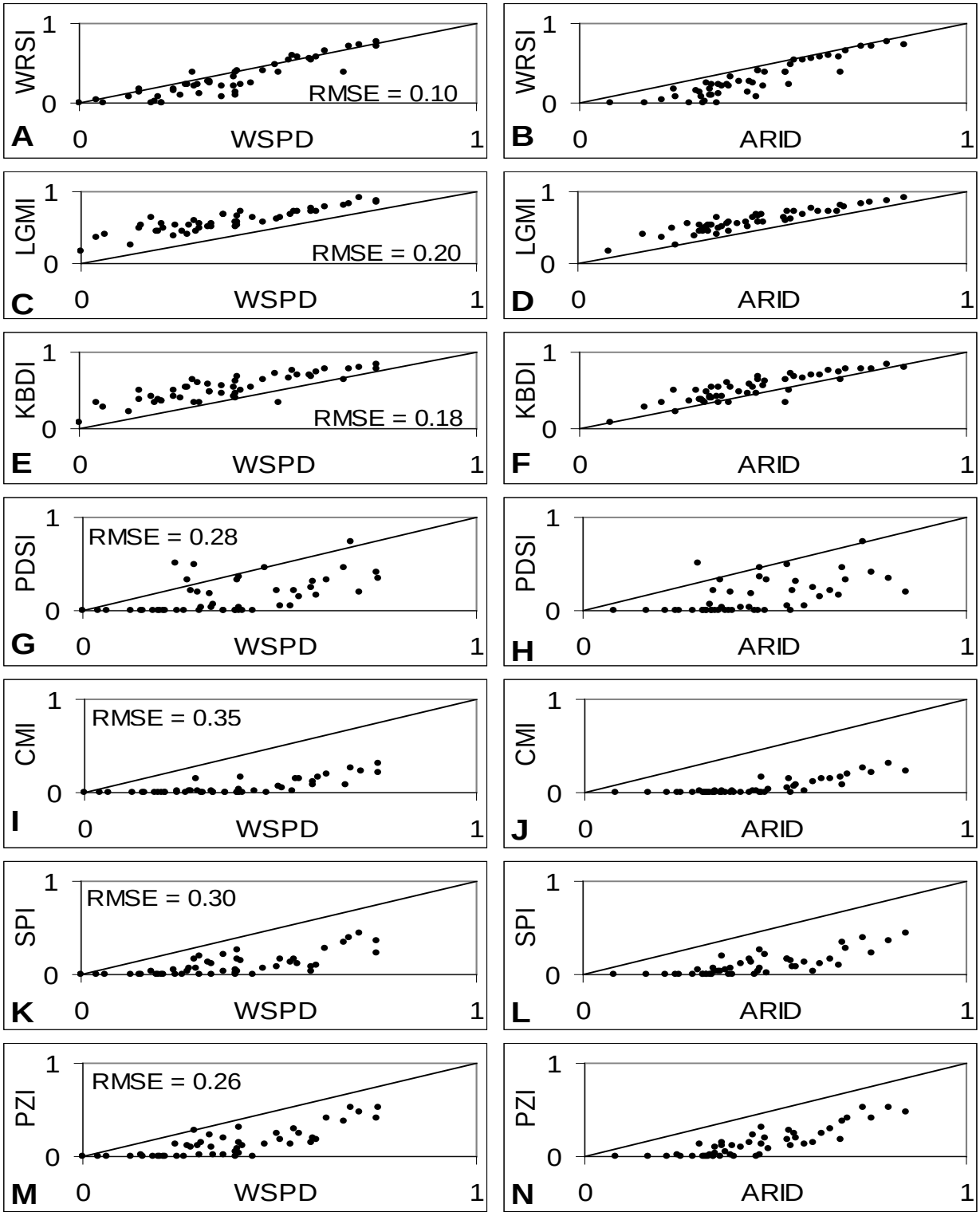

Figure 2-4. Association of WSPD with: A) WRSI, C) LGMI, E) KBDI, G) PDSI, I) CMI, K) SPI, and M) PZI for Huntsville during 1958 through 2005. Similarly, association of ARID with: B) WRSI, D) LGMI, F) KBDI, H) PDSI, J) CMI, L) SPI, and N) PZI.

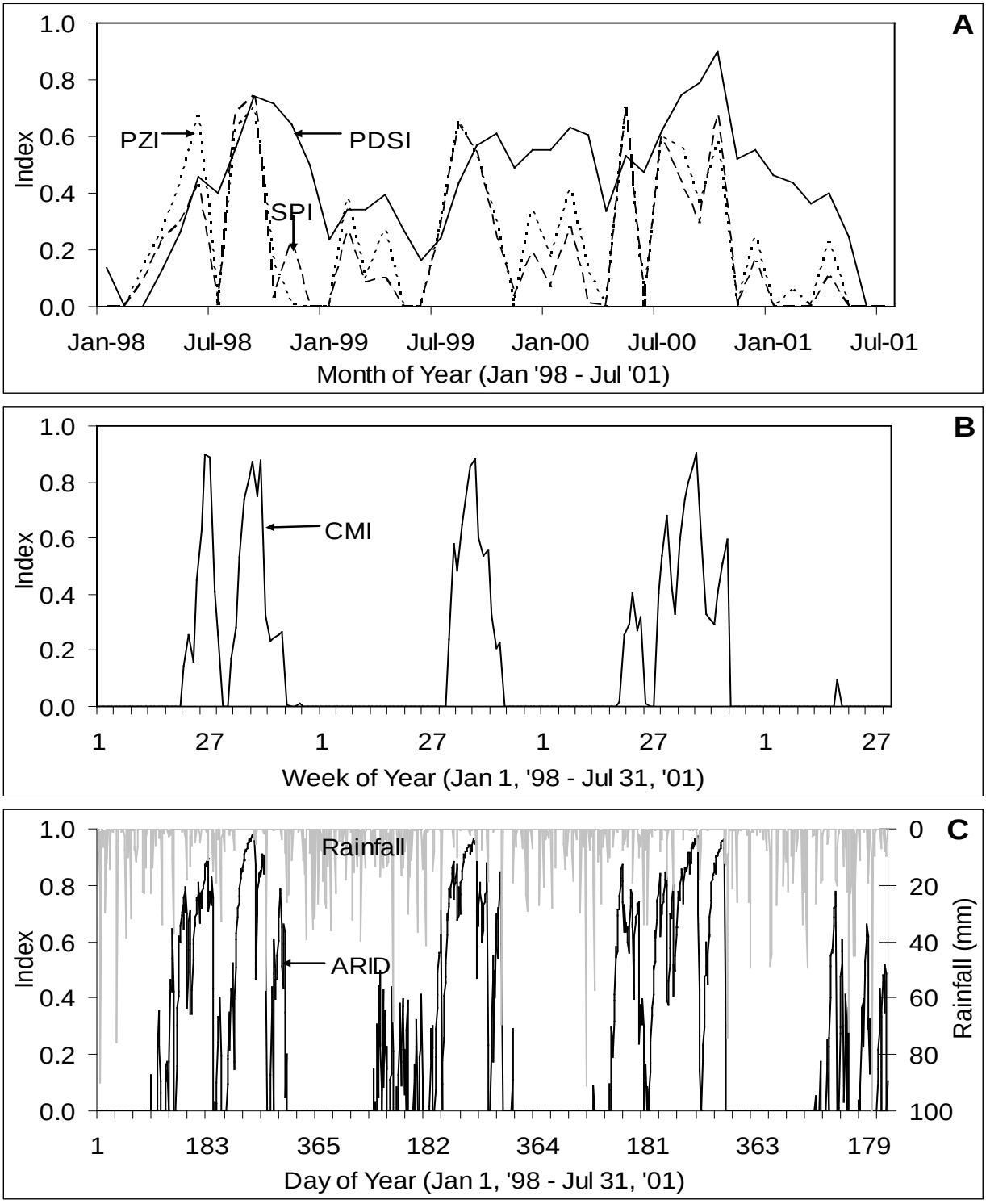

Figure 2-5. A) Monthly values of PDSI, PZI, and SPI; B) weekly values of CMI; and C) daily values of ARID and rainfall for Huntsville, AL, during Jan 1998 through Jul 2001

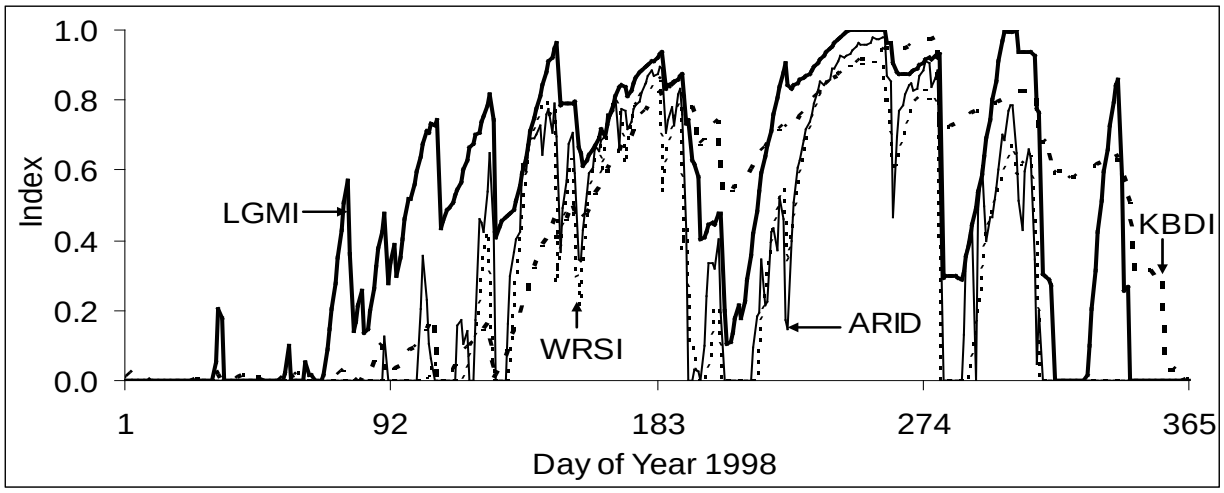

Figure 2-6. Daily values of ARID, KBDI, LGMI, and WRSI for Huntsville, AL, in year 1998

CHAPTER 3
UNCERTAINTIES IN AGRICULTURAL REFERENCE INDEX FOR DROUGHT

Introduction

Because a model is a simplification of reality, uncertainties in model outputs always exist. Structure, parameters, and input variables are three basic sources of uncertainty in a model. Choices made on the structure and functional relationships among components may be subjective. Because parameter values may be from estimations, literature reviews, or expert opinions, they may vary considerably in both space and time. Similarly, imprecise measurements or estimations may cause substantial variability in the values of input variables.

As a model, ARID may have the three sources of uncertainty mentioned above. In this study, however, only the influence of parameters was investigated for the following reasons. Inputs of ARID are weather variables, namely precipitation, solar radiation, temperature, and wind speed, whose values for a specific time and location are given. Values of weather variables vary in space, but ARID allows users to use location-specific values of these variables. So, no uncertainty in ARID is assumed to be associated with input variables. ARID has been designed for a specific purpose. Users who need a drought index that is simple, generic, SPA-based and has high temporal resolution may choose to use ARID because no other drought index is available that can meet all these specifics. Similarly, viable alternative structures for ARID are not currently available that meet the same objectives as ARID. So, with respect to meeting the specific objectives for which ARID was developed, no uncertainty in ARID is assumed to be related with its structure. However, values of the parameters of ARID may vary by time, location, or management.

ARID has five parameters: water uptake coefficient (α), drainage coefficient (β), runoff curve number (η), available water capacity (θ_m), and root zone depth (ζ). ARID uses a fixed value for each parameter, while its value may vary depending on a situation. For instance, η is influenced by soil texture, land use, land type, ground cover, slope, antecedent soil moisture conditions, and management. Similarly, values of β may vary with soil type and management. Soil texture, ground water table, and management may affect θ_m. Likewise, α and ζ may vary with topsoil depth and crop. Because soil texture, land type, water table, crop, and management may vary across locations, values of parameters may also vary accordingly, thus causing uncertainty in ARID. In this study, therefore, only parameters were considered the source of uncertainty. Based on literature review, the uncertainty ranges and nominal values of the parameters of ARID are presented in Table 3-1.

The objectives of this study were to determine the uncertainties in ARID associated with its parameters and to find out the importance of individual parameters with respect to the uncertainties in ARID. Specifically, this study was performed to seek answers to the following questions. What are the uncertainties in ARID as a simple and generic index? How sensitive is ARID to each of its parameters? How does uncertainty in ARID vary across locations or seasons? Does the sensitivity of ARID to its parameters vary by location or season? Answers of these questions could be helpful to users in understanding the behavior of ARID as a predictor or decision support tool and in applying or interpreting the index for their particular set of conditions.

Uncertainty / Sensitivity Analysis Methods

Uncertainty analysis (UA) involves deducing an uncertainty distribution for an output variable by evaluating the uncertainty or variability in the model components for a

given situation quantitatively. Sensitivity analysis (SA), on the other hand, consists of determining how sensitive the model output is with respect to the components of the model that are subject to uncertainty or variability. While UA is performed to determine the uncertainty in model outputs given the uncertainty in its parameters, SA is carried out to ascertain how a given model depends on its parameters (Saltelli et al., 1999). The basic purposes of performing uncertainty / sensitivity analysis are to identify parameters that are important in terms of influencing the output and to detect and quantify interaction effects among the parameters.

Typically, uncertainty / sensitivity analysis is carried out in four steps: 1) selecting the domain and distribution for each parameter; 2) generating a sample of points from the distribution of the parameters; 3) computing a set of model outputs after feeding the model with the sample elements; and 4) performing UA or SA analysis based on the results of model evaluations. Thus UA and SA differ only in the fourth step. For UA, output distributions are analyzed by computing means and variances; for SA, sensitivity indices are computed for each parameter by varying the parameters over their whole uncertainty ranges. Because both UA and SA follow the same procedures through the third step, a method that is used for performing SA also works for UA. That is, UA does not require a separate method to generate samples and evaluate models. So, a method is chosen based on the objective of SA.

For SA, two perspectives are used: local and global. Whereas local perspective focuses on the local (point) impact of parameters on the model, global perspective emphasizes on apportioning the output uncertainty to the uncertainty in the parameters. In local SA, the local response of a model output is investigated by varying parameters

one at a time while holding the others to their nominal values. Local SA indicates how fast the output changes locally around given values of the parameter. Although local SA can provide the role of some parameters, it is less useful than global SA when the purpose of the analysis is to explore the effect of uncertainty of several factors on model outputs (Monod et al., 2006). Local methods are not usable for a quantitative analysis; that is, they cannot find the relative importance of parameters with respect to uncertainty in the output (Ekstrom, 2005). Global SA, on the other hand, determines the importance of each parameter with respect to the uncertainty in the model output. Global SA evaluates the variability of output for the whole uncertainty domains of the parameters, thus providing a more realistic view of the model behavior when used in practice. Several methods are available for carrying out global SA that can be categorized further into screening-, factorial design-, correlation-, regression-, and variance-based methods (Saltelli et al., 2000a; Monod et al., 2006).

Screening methods isolate the set of parameters that are most important. However, they cannot quantify by how much a given parameter is more important than the others (Ekstrom, 2005). Two major screening methods are one-factor-at-a-time and Morris design (Morris, 1991). The one-factor-at-a-time method prevents interactions from being detected and quantified, whereas taking interactions into account is a key aspect of most global SA methods. Similarly, although Morris design can indicate whether or not interactions exist, it cannot identify which interactions are the most important, thus providing only qualitative information (Saltelli et al., 2000b).

Methods based on correlation measures, regression analysis, and factorial designs perform sensitivity analyses with intensive sampling. They rank factors in order

of importance on the basis of input-output relationships. However, they do not provide relative weights or quantify the percentage of output variance that is accounted for by each factor. Regression-based methods give information on the model that is used to describe the system model, not the system model itself (Saltelli et al., 1999). Also, regression techniques are not good at capturing nonlinear relationships between input and output and can be totally misleading for nonmonotonic models (Saltelli et al., 2004). The suitability of factorial designs to quantify the influence of multiple parameters is limited as they provide information on the model behavior only for specific values of the parameters (Monod et al., 2006). Examples of these types of methods are Pearson correlation coefficient, partial correlation coefficient, standardized regression coefficient, complete factorial design, and fractional factorial design. The rank-transformed versions of the correlation- and regression-based measures, namely rank correlation coefficient, partial rank correlation coefficient, and standardized rank regression coefficient, are even more qualitative than the original ones because the transformation modifies the model under analysis (Saltelli et al., 1993; Saltelli and Sobol, 1995).

Variance-based methods quantify the amount of variance that each parameter contributes to the total variance of the output. They account for nonlinear, nonmonotonic problems as they do not rely on model approximations and explore the full uncertainty ranges of the parameters (Monod et al., 2006). Variance-based methods are truly quantitative as the factors can be ranked in the order of their relative importance (Chan et al., 1997). They are model-independent, can cope with the influences of scale and shape of parameter distributions, can treat grouped factors, and include multidimensional averaging; whereas other methods such as local, regression-based,

Morris, and Monte Carlo filtering do not meet all of these properties (Saltelli et al., 2004). Some of the popular variance-based methods are importance measure (Hora and Iman, 1986; Ishigami and Hora, 1989; Iman and Hora, 1990), correlation ratio (McKay, 1995), Sobol sensitivity indices (Sobol, 1990), Jansen winding stairs (Jansen et al., 1994), and Fourier amplitude sensitivity test (FAST: Cukier et al., 1973,1975,1978; Schaibly and Shuler, 1973; Koda et al., 1979; Saltelli et al., 1999).

The importance measure and correlation ratio methods compute only main-effect contributions of parameters to the variance of output. They do not provide total-effect contributions. Moreover, they either lack robustness or are not easily interpreted (Saltelli et al., 2000b). The Sobol sensitivity index method is computationally more intensive and much more dependent on sample size (Saltelli and Bolado, 1998). It needs too many model evaluations to be practical (Ekstrom, 2005). The Jansen winding stairs method makes multiple uses of model evaluations. However, the reduction of model evaluations in the Jansen winding stairs method might affect the accuracy of the estimates (Chan et al., 2000b). Also, it needs a large number of model evaluations per parameter to obtain accurate results (Makowski et al., 2006).

The FAST, on the other hand, is one of the most elegant methods and computationally more efficient than the method of Sobol (Saltelli and Bolado, 1998; Saltelli, et al., 1999; Saltelli et al., 2000b) or Jansen winding stairs (Makowski et al., 2006). FAST is more robust than other methods as its estimates converge more rapidly even with a small sample size (Chan et al., 1997; Saltelli et al., 1999; Wang et al. (2005). FAST is fairly stable and relatively independent of sample size once the Nyquist criterion (Nyquist, 1932) is satisfied (Saltelli and Bolado, 1998). The FAST sensitivity

estimates can be evaluated independently for each factor as all the terms in a Fourier expansion are mutually orthogonal (Chan et al., 1997). FAST is good for models with no important or significant interactions among factors (Saltelli and Bolado, 1998). Moreover, FAST does not produce negative values, whereas the Jansen winding stairs and Sobol methods do so for the parameters that have a small influence on the model due to numerical errors (Makowski et al., 2006). Unlike other methods, the FAST estimates can never be greater than 1. Considering the strengths of FAST and the shortcomings of other methods as mentioned above, FAST was chosen for uncertainty / sensitivity analyses in this study.

FAST: An Overview

Sampling Approach

The key feature of FAST is that the multidimensional space of the parameters is systematically explored by a suitably defined search curve. The search curve is based on the search variable S_i (Cukier et al., 1978), which can be interpreted as the coordinate on the search curve:

$$S_i = -\pi + \frac{2i - 1}{N}\pi \tag{3-1}$$

where the subscript i stands for a scenario or solution point (i = 1, …, N), and N is the number of solution points that are selected along the search curve.

The search variable S_i forms a regularly-spaced sample on the interval (-π, π). Using values of S_i, the levels of parameter F_j are given by the following equation, which defines the search trajectory (Saltelli et al., 1999):

$$F_{i,j} = \frac{1}{2} + \frac{1}{\pi}\arcsin\left[\sin\left\{s_i\omega_j + \varphi_j\right\}\right] \tag{3-2}$$

where $F_{i,j}$ denotes a level for the i-th solution point and j-th parameter (j = α, β, η, θ_m, ζ); ω_j is the frequency associated with the parameter F_j; and φ_j is a random phase-shift parameter taking values on [0, 2π].

As S_i varies, the value of $F_{i,j}$ changes simultaneously along the search curve that systematically explores the parameter space. Equation 3-2 produces a sample for the parameter F_j on the interval [0, 1], ensuring that levels of each parameter are uniformly sampled (Monod et al., 2006).

Sensitivity Indices

A variance-based method, such as FAST, is based on the concept that variance can indicate importance for parameters. Variance-based methods assume that all information about uncertainty in output is captured by its variance. So, the basic idea of variance-based methods is to quantify the variance that each parameter contributes to the unconditional variance of the output.

The total variance of model output can be decomposed as a sum of terms of increasing dimensionality (Cukier et al., 1978). For instance, the total variance of a model that has five parameters is decomposed as:

$$V = \sum_{j=1}^{n} V_j + \sum_{j=2}^{n} \sum_{k=1}^{j-1} V_{jk} + \sum_{j=3}^{n} \sum_{k=2}^{j-1} \sum_{l=1}^{k-1} V_{jkl} + \sum_{j=4}^{n} \sum_{k=3}^{j-1} \sum_{l=2}^{k-1} \sum_{m=1}^{l-1} V_{jklm} + \sum_{j=5}^{n} \sum_{k=4}^{j-1} \sum_{l=3}^{k-1} \sum_{m=2}^{l-1} \sum_{n=1}^{m-1} V_{jklmn} \qquad (3\text{-}3)$$

where V is total variance; n is the number of parameters, which is 5; j, k, l, m, and n are parameters; V_j is the variance contributed by parameter j; and V_{jk} is the variance contributed by the interaction of parameters j and k , and so on.

A complete sensitivity analysis can be obtained by estimating all terms in the decomposition (Equation 3-3). However, depending on the number of parameters (p), terms may become too cumbersome for practical use as there will be as many as (2^p –

1) sensitivity indices. This problem is referred to as the curse of dimensionality (Bellman, 1957) because the computational cost of estimating all effects is exceptionally large. So, computing the set of first-order, also called main-effect, indices plus the set of total-effect indices is customarily preferred (Saltelli et al., 2004; Ekstrom, 2005; SimLab, 2009). The main- and total-effect indices provide a fairly good description of model sensitivities at a reasonable cost. These indices are enough to investigate the relative importance of each parameter with respect to model output variation, which is the ultimate goal in performing SA (Chan et al., 2000a).

The first term in the decomposition, V_j, corresponds to the part of total variance V arising from parameter F_j alone. Then, the ratio of V_j to V is the fraction of total variance contributed by parameter F_j. This ratio is referred to as the first-order sensitivity index of parameter F_j.

$$S_j = \frac{V_j}{V} \tag{3-4}$$

where S_j is the first-order sensitivity index.

Because S_j is related to the contribution of F_j alone, it is also called the main-effect index of F_j. The S_j is a measure of average influence of parameter F_j on the model output and can be thought of as the expected fractional reduction in variance that would be achieved if F_j were known. In many cases, the first-order indices represent an important fraction of the total indices.

If the variance of the model output arising from all-but-F_j is denoted as $V_{\sim j}$, the ratio of $V_{\sim j}$ to V may be considered as the fraction of total variance contributed by parameters other than F_j. If this ratio is subtracted from 1, the total variance, the remainder can be interpreted as the fraction of total variance contributed by parameter F_j (including the

interactions between F_j and the other parameters) and called the total-effect sensitivity index (S_j^T) because it gives a comprehensive measure of the influence of parameter F_j comprising of both main-effects and interactions.

$$S_j^T = 1 - \frac{V_{\sim j}}{V} \tag{3-5}$$

Total-effect index (S_j^T) is equal to the sum of all parameter indices involving parameter F_j and can be interpreted as the expected fraction of variance that would be left if only F_j were undetermined (Sobol, 1990; Homma and Saltelli, 1996; Saltelli et al., 1999). The difference between total-effect and main-effect sensitivities reflects the interaction between F_j and the other parameters. A difference of zero implies no interaction.

Estimating Variances

The basic principle of FAST is that if a response variable Y is sensitive to a given parameter F_j, both Y and F_j should vary simultaneously over the solution points. If parameter F_j has a strong influence on Y, oscillations of Y at frequency ω_j will be of high amplitude, which is the basis for computing a sensitivity measure.

By evaluating the spectrum for ω_j – the frequency associated with parameter F_j – and its higher harmonics, $p\omega_j$, the portion of the output variance V arising from uncertainty of parameter F_j is estimated as:

$$V_j = 2\sum_{p=1}^{M}\left(A_{p\omega_j}^2 + B_{p\omega_j}^2\right) \tag{3-6}$$

where V_j is the partial variance arising from parameter F_j, and M is the order of interference which arises when ω_j can combine to form another frequency of the set

(Cukier et al., 1978). In other words, M is the maximum number of Fourier coefficients that may be retained in calculating the partial variances without interferences between the assigned frequencies (McRae et al., 1982). $A_{p\omega_j}$ and $B_{p\omega_j}$ are Fourier coefficients of the model output at frequency ω_j and its higher harmonics $p\omega_j$. The other frequencies that are not used for computing V_j give information on the interaction effects among parameters.

Partial variance arising from all-but-F_j, denoted as $V_{\sim j}$, is estimated by assigning a frequency ω_j to parameter F_j and a different frequency $\omega_{\sim j}$ to all remaining (complementary) parameters. By evaluating the spectrum at frequency $\omega_{\sim j}$ and its higher harmonics $k\omega_{\sim j}$, partial variance in the complementary set $V_{\sim j}$ is estimated by assigning usually a large value for frequency ω_j and usually small and almost identical values for the complementary set of frequencies $\omega_{\sim j}$. Partial variance of the complementary set is computed as:

$$V_{\sim j} = 2\sum_{k=1}^{\frac{\omega_j}{2}} \left(A^2_{k\omega_{\sim j}} + B^2_{k\omega_{\sim j}}\right) \tag{3-7}$$

where $\omega_j/2$ is the frequency range of the spectral components associated with $V_{\sim j}$ which does not overlap the region of V_j. $A_{k\omega_{\sim j}}$ and $B_{k\omega_{\sim j}}$ are Fourier coefficients of model output at $\omega_{\sim j}$ – the fundamental frequency of the complementary set – and all its higher harmonics $k\omega_{\sim j}$.

Principally, a small value is assigned to $\omega_{\sim j,}$ the best possible value being 1 (Saltelli et al., 1999). M is usually set to 4 (Saltelli, et al., 1999; Monod et al., 2006). By summing all the spectrum of the Fourier series expansion, total variance (V) is computed as:

$$V = 2\sum_{k=1}^{M\omega_j} \left(A_k^2 + B_k^2\right) \tag{3-8}$$

where $M\omega_j = (N-1)/2$, the number of solution points that fall on (0, |π|); and N is the actual number of solution points that are selected along the search curve and is derived from the Nyquist criterion (Nyquist, 1932) as $N = 2M\omega_j + 1$. So, the number of points along the curve is determined after fixing ω_j. A_k and B_k are the Fourier coefficients of the model output at frequency k.

Estimating Fourier Coefficients

The Fourier coefficients A_k and B_k are numerically evaluated using the following difference expressions derived through a simple numerical quadrature technique (McRae et al., 1982):

$$A_k = \begin{cases} \frac{1}{N}\left[Y_{N''} + \sum_{q=1}^{N'} \left\{ \left(Y_{N''+q} + Y_{N''-q} \right) \cos\left(\frac{k\pi q}{N} \right) \right\} \right] & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases} \tag{3-9}$$

$$B_k = \begin{cases} \frac{1}{N}\left[\sum_{q=1}^{N'} \left\{ \left(Y_{N''+q} - Y_{N''-q} \right) \right) \sin\left(\frac{k\pi q}{N} \right) \right\} \right] & \text{if } k \text{ is odd} \\ 0 & \text{if } k \text{ is even} \end{cases} \tag{3-10}$$

In Equations 3-9 and 3-10, N is the number of solution points selected on the search curve; $N' = (N - 1)/2$, the number of solution points that fall on (0, |π|); $N'' = N' + 1$; $Y_{N''-q}$ is model output associated with negative (0, -π) values of the

search variable S_i; $Y_{N''}$ is model output associated with 0 value of S_i; and $Y_{N''+q}$ is model output associated with positive (0, π) values of S_i (Equation 3-1). Values of S_i are negative when i (i = 1, …, N) ranges from N' to 1, zero when i = N'', and positive when i ranges from $N'' + 1$ to N.

For computing V_j in Equation 3-6, which can also be written as

$$V_j = 2\sum_{k=1\omega_j}^{M\omega_j} \left(A_k^2 + B_k^2\right), \quad (3\text{-}11),$$

values of k in Equations 3-9 and 3-10 range from $1\omega_j$ to $M\omega_j$. For $V_{\sim j}$ in Equation 3-7, k ranges from 1 to $\omega_j/2$, and for V in Equation 3-8, k ranges from 1 to $M\omega_j$.

Materials and Methods

The Dynamic Nature of ARID

Before uncertainty and sensitivity analyses, a preliminary study was carried out to explore how ARID with different sets of parameter values would respond to fluctuations in weather variables, especially precipitation, over a period of one month. For this study, daily values of ARID were computed for Gainesville, Florida, for January and July 2009. Only Gainesville was chosen because one location is enough to show the dynamics in ARID. The two months were selected as they have different patterns of temperature and precipitation. For instance, January was relatively dry, whereas July was relatively wet.

To explore the effect of different sets of parameter values on the response of ARID to precipitation fluctuations, ARID was computed for each of the three values chosen for each parameter: α = 0.08, 0.10, and 0.12; β = 0.30, 0.55, and 0.80; η = 40, 65, and 90; θ_m = 0.08, 0.13, and 0.18; and ζ = 300, 400, and 500 mm. The computations were carried out using one of the three values of the given parameter and the default

(nominal) values of the other parameters. Of the three values of each parameter given above, the middle one is the nominal value. The other two values of each parameter are from their respective uncertainty ranges (Table 3-1).

After exploring the dynamic behavior of ARID, the uncertainty in the index associated with its parameters and the sensitivity of the index to its parameters were analyzed. The following sections deal exclusively about the material and methods for uncertainty and sensitivity analyses.

Study Area and Data

To assess the effect of location on the sensitivity of ARID to its parameters, five locations in the southeast USA, namely Miami, Bartow, and Live Oak, FL, and Plains and Blairsville, GA, were selected taking into consideration geographical representation (latitude) and the availability of daily historical weather data for the current climate normal period of 1971-2000 that included the weather variables dew point temperature and winds peed in addition to solar radiation, precipitation, and maximum and minimum temperatures (Figure 2-2).

Daily historical weather data for the current climate normal period for locations in Florida were obtained from ftp://coaps.fsu.edu/pub/griffin/SECC/AgroClimate/raw/ascii/, the Florida Climate Center website, and those for locations in Georgia were obtained from the Georgia Automated Environmental Monitoring Network (G. Hoogenboom, personal communication, 9 March 2009). Then, using the weather data and a given set of generated parameter values, daily values of ARID were computed for the current climate normal period of 1971-2000. To explore the effect of season on the sensitivity of ARID to its parameters, the daily values of ARID over the normal period of 30 years were separated into 4 seasons – winter (December, January, and February), spring

(March, April, and May), summer (June, July, and August), and fall (September, October, and November) – and averaged by season to compute seasonal normal values. Season was used as a temporal scale because seasonal variations in weather could be more distinct than daily or monthly differences. For the study, the weather data of the current normal period was used because the 3-decade interval is long enough to filter out many of the short-term interannual fluctuations and anomalies, but short enough to reflect long-term climatic trends (WSCO, 2003). The seasonal normal value of ARID was computed for each iteration, season, and location.

Uncertainty / Sensitivity Analysis

Parameter distributions

The range and distribution selected for each parameter based on literature review are presented in Table 3-2. Of the five parameters, β, η, and θ_m had variability in their distributions. For instance, He (2008) found that these parameters are normally distributed. Wang et al. (2005), on the other hand, concluded that η and θ_m have triangular distributions. Similarly, Aggarwal (1995) claimed θ_m to have a beta distribution. Muleta and Nicklow (2005) opined that the distribution of η is uniform or triangular. Taking into account the uncertainty in the distributions of β, η, and θ_m, therefore, three distributions were considered for each one of these parameters in this study: uniform, triangular, and normal. So an additional objective of this study was to assess how the uncertainty in ARID would vary with the type of distribution of these parameters and if the distribution type would affect the sensitivity of ARID.

For ζ, the distribution was assumed to be uniform because turfgrass can have any value for the rooting depth within the selected range depending on management, cultivar, and soil type. For α, the distribution was selected based on the results of

Dardanelli et al. (2004). Using data from various experiments, they had computed several values for α ranging from 0.06 to 0.11, 0.096 being the mode (Figure 3-1).

Parameter samples

Setting M to 4 (Saltelli et al., 1999; Ekstrom, 2005; Monod et al., 2006) and assigning 48 for ω_j, the required number of solution points along the search curve (N) was derived as $N = 2M\omega_j + 1 = 385$. Because the minimum value that can be assigned for ω_j is $2M\omega_{\sim j}$ and the maximum value that can be assigned for $\omega_{\sim j}$ for a 5-parameter model is 4 (Saltelli et al., 1999), the minimum value required for ω_j was estimated as 32. However, for better results (convergence), 48 (i.e., 1.5 times the minimum) was used instead of the minimum itself. Then, using Equation 3-2, N solution points (where N = 385) were generated on the interval (0, 1) for each parameter.

Finally, the i-th sample point (where i = 1, …, N) for the j-th parameter (where j = β, η, θ_m, or ζ) for a given distribution was generated using the inverse of the cumulative distribution function (CDF), which comprised of a solution point $F_{i,j}$ (computed using Equation 3-2) and the parameter values of the inverse CDF (Table 3-3). For α, however, samples were generated not based on the theoretical probability density functions as assumed for the other parameters, but based on the empirical distribution presented in Figure 3-1. From the empirical frequency distribution, the CDF of α was created (Figure 3-2). Then, from the inverse of the CDF, a sample of N values of α was generated following the procedure presented in Table 3-4. For instance, if an i-th solution point of α ($F_{i,\alpha}$) were less than or equal to 0.10, the i-th value for α, α_i, would be selected randomly from the range of 0.06 to 0.07 assuming a uniform distribution within the range.

Model evaluations

Once the i-th sample point for each parameter was generated, the sample set {α_i, β_i, η_i, θ_{mi}, ζ_i} was used in the model and $ARID_i$ was computed. Accordingly, ARID was computed N times (N = 385) using N sample sets. Similar computations were made for each season and location.

Uncertainty / sensitivity analyses

Uncertainty and sensitivity analyses were performed for each season and location, and their effects on the uncertainty in ARID and the sensitivity of the index to its parameters were explored. Uncertainties in ARID were determined in terms of the shape and scale of the distribution of ARID over the uncertainty ranges of its parameters and the deviation of the index from its nominal value.

Finally, implementing the FAST method, the main- and total-effect sensitivity indices (Equations 3-4 and 3-5) were computed for each parameter and the sensitivity of ARID to its parameters was analyzed based on these sensitivity indices. Effects of season and location on the sensitivity were assessed by computing the sensitivity indices for each season and location.

Results and Discussion

The Dynamic nature of ARID

Overall, ARID fluctuated inversely with precipitation. When precipitation occurred, values of ARID fell down, and during dry period, ARID went up (Figure 3-3). The index was also influenced by temperature. For instance, on January 3, ARID decreased due to less ET_o loss than the previous day; on January 20, it increased because of more ET_o loss; and on January 21, it fell down back to 0 due to less ET_o loss than the previous

day (Figure 3-3A). These changes in ARID are obvious because precipitation and ET_o are the only inputs of ARID.

Of the two months compared, January was relatively dry because it had only a few rainy days, whereas July was relatively wet as there were a number of rainy days in this month (Figure 3-3). The January and July conditions are referred to, here, as the wet and dry conditions, respectively.

Under dry conditions, soils with small θ_m, such as sandy, generally had larger values of ARID than did the ones with large θ_m, such as loam or clay (Figure 3-3A). ARID decreased with a decrease in θ_m from 0.08 to 0.13 and did not change thereafter. The phenomenon of no change in ARID above $\theta_m = 0.13$ was because the amount of available water in the soil was never greater than 52 mm, the maximum available water content at $\theta_m = 0.13$, due to dry condition. So, when θ_m was 0.13 or greater, there was no drainage loss and, thus, the amount of available water remained the same. Below $\theta_m = 0.13$, however, less and less water drained with an increase in θ_m due to increase in maximum available water in the soil. The increase in available water led to an increase in transpiration and decrease in ARID. Under wet conditions, however, ARID decreased with an increase in θ_m for all values of θ_m (Figure 3-3B). Under wet conditions, the amount of available water in the soil sometimes crossed 72 mm, the maximum available water at $\theta_m = 0.18$. Because each of the three soil types had different available water capacities – 32 mm, 52 mm, and 72 mm with θ_m equal to 0.08, 0.13, and 0.18, respectively – the amount of water lost through deep-drainage in these soils varied accordingly. Available water is subject to drainage when it is greater than that at θ_m, the threshold value. Because more water drained from the soil with smaller available water

capacity relative to the one with larger capacity, the amount of available water in the soil increased with an increase in θ_m, thus leading to increased transpiration and decreased ARID. At large precipitation, however, ARID was 0 for all soils because water was not limiting.

As normally occurs, ARID decreased with an increase in ζ from 300 mm to 400 mm (Figure 3-3C) also under dry conditions because of more and more available water in the soil. However, the index remained the same for the soils having the ζ value deeper than 400 mm because the amount of available water in the soil was never greater than 52 mm, the maximum available water content at ζ = 400 mm, due to dry condition. So, when ζ was deeper than 400 mm, there was no drainage loss and thus the amount of available water remained the same. For ζ = 300 to 400 mm, however, less and less water drained with an increase in ζ due to increase in maximum available water from 39 mm to 52 mm. So, more and more available water remained in the soil, thus leading to increased transpiration and decreased ARID. Under wet conditions, however, the amount of available water in the soil was often greater than 65 mm, the maximum available water at ζ = 500 mm. Because soils with ζ = 300 mm, 400 mm, and 500 mm had different available water capacities, namely 39 mm, 52 mm, and 65 mm, respectively, the amounts of water lost through deep-drainage in these soils were also different. Because more water drained from the soil with shallower rootzone relative to the one with deeper rootzone, the amount of available water in the soil increased with an increase in ζ, thus leading to an increase in transpiration and decrease in ARID (Figure 3-3D). At large rainfall, however, there was no difference among ζ values due to no water limitation.

ARID increased, although slightly, with an increase in α under dry conditions (Figure 3-3E). This increase was because the available water in the soil at α = 0.08 was more than $\frac{0.10}{0.08}$ times the available water at α = 0.10, and the available water at α = 0.10 was more than $\frac{0.12}{0.10}$ times the available water at α = 0.12. In other words, with an increase in α, more and more water transpired in a given day, and thus less and less water remained in the soil and became available for transpiration for the following day, leading to the larger values of ARID. Under wet conditions, however, ARID decreased with an increase in α (Figure 3-3F) because the available water at α = 0.08 was less than $\frac{0.10}{0.08}$ times the available water at α = 0.10, and the available water at α = 0.10 was less than $\frac{0.12}{0.10}$ times the available water at α =0.12. In other words, with an increase in α, more and more water transpired because of the availability of a large amount of water in the soil under wet conditions. At large rainfall, no difference in ARID was found with different values of α due to no water limitation in each case.

The number of rainy days over a given period as well as the amount of rainfall per day is smaller under dry conditions. When η was large, for instance 90, a large portion of even the scanty rainfall drained through surface runoff and only a small portion infiltered into the soil, leading to a smaller amount of available water and less transpiration. When precipitation was less than 27.4 mm but greater than 5.6 mm, run-off occurred only in the soil with η = 90. So, this soil became drier than the ones with η = 40 and 65. As this situation continued for some days, a significant difference occurred between the moisture contents of the soils with η = 90 and η = 40 or 65 (Figure 3-4A).

Under wet conditions, where the number of rainy days over a month as well as the amount of rainfall in each event is relatively larger, soils with different values of η did not vary significantly in terms of the amount of available water in the soil even though soils with larger η drained more water per event and more frequently over a month than did the ones with smaller η. So, the difference in ARID due to the difference in η was negligible although soils with greater η removed more water (Figure 3-4B). At large rainfall, for instance on January 13, the difference in ARID associated with different values of η was zero because of no water limitation.

The amount of available water in the soil throughout the month was too small for the deep-drainage to occur in each soil with β = 0.30, 0.55, and 0.80 under dry conditions (Figure 3-4C). The minimum amount of available water needed to be in the soil for the drainage to occur was 52 mm, which never occurred. So, the amount of available water in each soil remained the same, thus leading to the same magnitude of ARID. Under wet conditions, however, the amount of available water above the threshold value was subject to drainage. With an increase in β from 0.30 through 0.80, more and more water drained and thus less and less water remained in the soil, leading to an increase in ARID (Figure 3-4D). However, the differences in available water among the soils with different β values were so insignificant that values of ARID remained almost the same across β values. At large rainfall, ARID remained the same (zero) with different β values due to no water limitation.

Inputs and Parameters

ARID has basically two input variables: precipitation (P) and the potential ET of the reference grass, called the grass reference ET (ET_o). The seasonal total values of P and ET_o for each of the four seasons and five locations are presented in Figure 3-5. In

the winter, precipitation increases from south to north on the order of Miami (MI), Bartow (BT), Live Oak (LO), Plains (PL), and Blairsville (BV), whereas in the summer it increases from north to south (Figure 3-5A). While spring follows the pattern of winter, fall follows that of the summer. After spring and fall, the precipitation pattern is reversed. The ET_o increases from north to south in each season and is minimum in the winter and maximum in the summer in each location (Figure 3-5B).

Distributions of each parameter after sample generations based on the information in Table 3-2 are presented in Figure 3-6. As stated earlier, three types of distributions were considered for β, η, and θ_m – uniform, triangular, and normal. For ζ and α, however, only uniform and empirical distributions were selected, respectively. With change in the type of distribution from uniform to triangular to normal, values of β, η, and θ_m each concentrated more and more towards the center (nominal values) of the uncertainty range (Figure 3-6).

The parameters showed no significant correlations among themselves in each location and season, indicating that they are independent of each other (Table 3-5). These results ascertained the use of FAST, a variance-based technique for sensitivity analysis, where the independence of parameters is a requirement.

Uncertainty in ARID

Effect of distribution

With change in the type of distribution for parameters β, η, and θ_m from uniform to triangular to normal, the uncertainty range in ARID decreased significantly in each change for all locations and seasons (Figure 3-7). The decrease in uncertainty range was from 15 to 50%, depending on season and location. The decrease was because as the distributions moved from uniform towards normal, more and more samples of the

parameters concentrated towards the centers or nominal values (Table 3-1) of their uncertainty ranges. So, values of ARID concentrated more and more towards nominal values, thus producing smaller uncertainty ranges. The results indicated that choice for parameter distributions can have a significant impact on the uncertainty ranges of an output variable.

Effects of location and season

The variability in ARID across locations and seasons was found to be associated with the variability in its input variables – precipitation and ET_o. For instance, in winter and spring, precipitation increased northwards from Miami. Accordingly, the uncertainty ranges of ARID stretched more and more towards zero when moved from south to north (Figure 3-8). In these seasons, as the amount of ET_o decreased northwards (Figure 3-5), the ranges of ARID shifted more and more towards zero from southern to northern locations. The end result was that the uncertainty ranges shifted as well as stretched more and more towards zero and became wider and wider when moved from south to north (Figure 3-8). In summer and fall, on the other hand, the pattern of precipitation was inverse (Figure 3-5A). That is, the level of precipitation increased southwards from Blairsville. Accordingly, the uncertainty ranges of ARID extended more and more towards zero southwards. As the amount of ET_o also increased towards south, ARID values tended to move farther and farther from zero towards one. However, because of the increasing amounts of precipitation southwards, effects of ET_o were offset by those of precipitation. The consequence was that the upper values of the ranges remained about the same across locations, but the ranges extended more and more towards zero southwards. However, due to more ET_o in lower latitudes than in northern ones, the ranges of ARID in southern locations in large precipitation seasons were not as close to

zero as were the northern ones. So, the change in the uncertainty range of ARID across locations and seasons was mainly due to the change in precipitation and ET_o levels.

Overall, the uncertainty in ARID was small in comparison with its entire possible range on the interval [0, 1] (Figure 3-8). Uncertainty ranges of ARID were mostly small (less than 30% of the whole possible range of [0, 1]), except for some locations and seasons that had large amounts of precipitation (seasonal total of more than 400 mm). Even in the cases with large uncertainty ranges of up to 50% of the entire possible range of [0, 1], ARID concentrated around nominal values with small uncertainty ranges (Figure 3-8, Table 3-8). A nominal value of ARID is the one computed with its fixed set of parameters.

The uncertainty ranges presented in Figure 3-8 are based on triangular distributions of η, β, and θ_m. If beta or normal distributions were used for these parameters, as Aggarwal (1995) and He (2008) suggest, the uncertainty ranges would further decrease (Figure 3-7). Furthermore, the uncertainty ranges were relatively large especially under wet conditions, but were usually small under dry conditions. Because monitoring or forecasting moderate to severe droughts is perhaps more important than monitoring near-normal to mild droughts, estimating severe droughts with less uncertainty is more important than estimating mild droughts with more certainty.

Sensitivity of ARID

In terms of importance, Chan et al. (1997) categorizes parameters with $S_j^T > 0.8$ as very important, with $0.5 < S_j^T < 0.8$ as important, with $0.3 < S_j^T < 0.5$ as unimportant, and with $S_j^T < 0.3$ as irrelevant, where S_j^T is the total-effect sensitivity index for parameter j. According to this classification, none of the parameters of ARID

was found to be very important, and the only parameter that is important is θ_m (Table 3-6). Parameter ζ is unimportant, whereas α, β, and η are irrelevant. Makowski et al. (2006) considers a parameter to be significantly influential only when its total sensitivity index is greater than 0.1. Based on this definition, only α, θ_m, and ζ have significant influence on ARID (Table 3-6). Of the five parameters, the S_j^T values of θ_m and ζ were significantly greater than those of the other parameters in each location and season (Table 3-6), indicating that ARID is more sensitive to θ_m and ζ than to the others. Parameters β and η, on the other hand, were found to be least influential in all the cases, which was due to insignificant difference in the amount of available water in the soil with a change in these parameters. This insignificant difference was mainly because runoff and drainage do not always occur; that is, they occur only when there is precipitation. Moreover, runoff occurs only when the amount of precipitation is greater than a threshold called initial abstraction and drainage occurs only when the amount of available water is greater than a threshold called available water capacity. In most locations and seasons, α was also less influential. The sensitivity to α was relatively more in the seasons and locations that had more precipitation than the others. Based on relative importance, the parameters could be ranked as $\theta_m > \zeta > \alpha > \beta > \eta$, where the symbol > is pointed to the direction of less influence.

The main-effect sensitivity indices of θ_m, the most influential parameter, indicated that θ_m alone (without any interaction) contributes about 50 to 60% of the total variability in ARID, depending on location and season (Figure 3-9). This means that about 50 to 60% of variance in ARID would be left if only θ_m were to stay undetermined. Similarly, ζ, the second most influential parameter, was found to contribute from 25 to 37% of the

total variance. The third most influential parameter, α, contributes about 3 to 15% of the total variance in ARID, whereas β and η, the least influential parameters, each contribute less than 3% of the total variance in ARID. As the sum of all order (first and higher) indices is equal to 1, the interaction among the parameters, computed as 1 minus the sum of first-order indices, contribute about 3 to 20% of total variance depending on location and season (Figure 3-9).

Effect of distribution

With change in the type of distribution for β, η, and θ_m from uniform to triangular to normal, values of the first- as well as total-effect sensitivity indices of these parameters decreased considerably in each location and season (Figure 3-10). The decreasing influence of distribution type in the direction of normal distribution was due to concentrating more and more values of these parameters towards the centers (nominal values) of their uncertainty ranges. Because sensitivity index values are relative and the sum of indices of all orders equals to 1, the sensitivity indices of α and ζ, the parameters whose distributions were kept fixed, increased towards normal distribution to compensate for the decreased values of β, η, and θ_m, the parameters with variable distributions (Figures 3-10C and 3-10D).

Effects of location and season

For η and β, the effects of location and season on the sensitivity of ARID to these parameters were very small. Similarly, the sensitivity of ARID to α was small in most cases. However, for some locations and seasons, the sensitivity was considerable (TSI up to 0.09) (Figure 3-11A). In general, the sensitivity to α was influenced by the amounts of precipitation and ET_o (Figure 3-5). With increase in precipitation and decrease in ET_o, the sensitivity increased northwards in the winter and southwards in

the summer (Figure 3-11A and Figure 3-5). For most locations, the sensitivity was greatest in the summer and least in the fall. These results indicated that the sensitivity of ARID to α increases with an increase in the amount of water in the soil. That is, the amount of water taken up by plants through the process of transpiration (TR) fluctuates more with changes in α values in wetter conditions than in drier conditions. These indications are obvious because ARID computes TR as a linear function of available water (W) as: $TR = \alpha W$. So, when W is large, the $\frac{TR}{\alpha}$ ratio is also large, and thus the uncertainty range of TR becomes wider. Because the amounts of precipitation and ET_o in a particular season change across locations and those in a specific location change across seasons, the effects of season and location on the sensitivity of ARID to α was actually due to the variability of precipitation and ET_o across locations and seasons. If all locations and seasons had the same amounts of precipitation and ET_o, no difference would be found in the sensitivity across locations and seasons.

The sensitivity of ARID to ζ varied with W, which is influenced by the amounts of precipitation and ET_o. However, the effect of W on the sensitivity was negative. That is, the sensitivity to ζ was more in drier conditions than in the wetter conditions. For instance, Plains was drier than other locations for most of the seasons due to less precipitation and more ET_o (Figure 3-5). So, the sensitivity of ARID to ζ was more in this location (Figure 3-11B). On the other hand, the sensitivity was least in Blairsville for all seasons due to more precipitation and less ET_o (Figure 3-5A and 3-11B). Compared with other seasons, generally the sensitivity to ζ in the summer was least for all locations because of wetter conditions in this season. The smaller sensitivity in wetter conditions was because plant water uptake is relatively less soil-limited unlike under dry

conditions. In wet conditions, the amount of water even in a shallow rootzone may be sufficient to meet the evaporative demand of the atmosphere. In dry conditions, however, plant water uptake is more soil-limited. So, with an increase in the root zone depth, more and more amount of water is available for plant uptake.

As for ζ, the sensitivity of ARID to θ_m was influenced by the amount of water in the rootzone for all locations and seasons. In general, the sensitivity to θ_m was larger in drier conditions and smaller in wetter conditions. For instance, the amount of precipitation increased southwards from Blairsville to Miami during summer and fall (Figure 3-5A). Accordingly, the sensitivity to θ_m decreased from Blairsville to Miami (Figure 3-11C). Because of more precipitation and thus more amount of water in the soil, the sensitivity in the summer was less than that in other seasons for most locations (Figure 3-11C). Under wet conditions, soils with any water holding capacity may have enough water in the root zone for plant uptake or are less deficient to the evaporative demand of atmosphere. So, the rate of transpiration does not vary much among the soils that have different water holding capacities. Under dry conditions, however, plant water uptake is soil-limited. That is, plant water uptake is directly proportional to the amount of available water in the soil under dry conditions. Because the amount of available water in a soil varies with its water holding capacity, the rate of plant water uptake varies considerably among the soils that have different water holding capacities.

Interaction effects

In general, the interaction effects of the parameters were insignificant as the values of the interaction-effect indices, which were computed as $\left(S_j^T - S_j\right)$, were less than 0.1 for most of the locations and seasons (Table 3-7). The significant interactions

were found only among α, θ_m, and ζ for Plains and Blairsville in the winter, which is discussed in the following paragraphs.

The effect of interaction among parameters was less with uniform distribution than with triangular or normal distribution (data not shown). The less interaction effect with uniform distribution was likely because samples of each parameter are uniformly distributed over the parameter space. With triangular or normal distribution, however, samples are more centered towards the nominal value of each parameter. So, the process of central orientation of the samples might be responsible for more interaction effects.

For most locations and seasons, the interaction effects were insignificant $\left((S_j^T - S_j) < 0.1\right)$ (Table 3-7). Also, no significant difference in interaction was found across most locations and seasons. Considerable interaction effects were found only for Plains and Blairsville in the winter, which was probably because these locations received a considerably large amount of precipitation and lost a small amount of water through ET_o in this month. Because under wet conditions ARID is more sensitive to α, but less sensitive to θ_m and ζ, the large amount of water in the root zone could have caused more interactions among α, θ_m, and ζ. For other seasons and locations, differences in the interaction effects were insignificant as their precipitation gains as well as ET_o losses did not differ substantially.

Conclusions

Sensitivity analyses showed that β and η are relatively the least important parameters of ARID and that the influence of each of these parameters is insignificant $(S_j^T < 0.1)$. Results indicated that even if η and β are set to any value within their

uncertainty ranges, the difference in the uncertainty in ARID is negligible. Results further implied that from the view point of estimating runoff, ARID is applicable to a wide range of conditions that are represented by η ranging from 15 to 90. That is, ARID is applicable to any soil type, land use, landscape, crop, antecedent moisture condition, and management. Similarly, from a deep-drainage point of view, ARID is valid for a wide range of soil textures and management. The influence of α is also insignificant for locations or seasons that do not get large amount of precipitation although, in principle, its influence is proportional to the amount of water in the soil. Even for the significant cases, the percent of variance α contributes to the total variance in ARID is at most 15. Results indicated that except for soils that have large soil moisture content, ARID can be applied to a wide range of crops, rootzone depths, and management. Of the two most influential parameters, ζ is important, but not as influential as θ_m. The effect of ζ is significant for all locations and seasons, and its contribution to the total variance of ARID is about 30%. The most influential parameter is θ_m and contributes about 60% of the total variance in ARID. So, θ_m is the parameter that might be of most interest to the users. As a generic index, ARID provides a general indicator for crop water status for a wide range of conditions, and, as the results showed, the index works fairly well in most cases. However, if users want precise information about the water deficit for their specific soils, θ_m is the parameter whose value they may manipulate.

Results indicated that although ARID uses a fixed set of parameter values for a wide range of conditions, it has fairly small uncertainties, except for soils that have large moisture content. Even for wet conditions, the distributions of ARID peaked at nominal values with small uncertainty ranges, indicating more probability of less uncertainty.

Because θ_m is the most influential parameter, the uncertainty in ARID is mostly due to θ_m. The uncertainty is also influenced by the type of parameter distribution considered. If β, η, and θ_m are really normally distributed (He, 2008), the uncertainty in ARID decreases even further by about 10 to 30% depending on locations or seasons. Values of uncertainty ranges presented in Figure 3-9 are based on triangular distributions.

The type of distribution selected for parameters can make a significant difference in the sensitivity of and the uncertainty in ARID. By moving from uniform to normal distribution, about 15 to 50% uncertainty in ARID can be reduced. Similarly, shifting to normal from uniform distribution can alter the relative importance of each parameter and its sensitivity index values. So, caution is needed while selecting parameter distributions.

The sensitivity of and uncertainty in ARID seemed to vary by season and location. Actually, it was precipitation and ET_o, the input of ARID, that varied across locations and seasons, thus causing ARID to vary accordingly. If all locations and seasons had the same amounts of precipitation and ET_o, the uncertainty range of ARID would remain the same in all cases. So, ARID is independent of locations and seasons.

Finally, the uncertainty / sensitivity results related to ζ apply only to the grasses or crops whose effective root zone lie within the range [30 cm, 60cm]. For other ranges of ζ, the uncertainty ranges of ARID and the relative importance of each parameter could be substantially different. Similarly, the results apply only to the landscape whose water table is so deep that the upward flow of water through capillary rise does not make any significant contribution to the amount of water in the root zone considered. Generally, rainfed agricultural lands have considerably deep water table (deeper than 60 cm), so

ARID is applicable to these lands. However, if water table is shallow enough to make a significant contribution to soil water balance, drought never exist there. For such a case, no drought index is needed, actually.

Table 3-1. Uncertainty ranges and nominal values of various parameters of ARID

Parameter	Unit	Uncertainty range		Nominal value (what ARID uses)	
		Range	Source	Value	Source
Water uptake coefficient (α)	mm mm^{-1}	[0.06, 0.11]	Dardanelli et al. (2004), Dardanelli et al. (1997)	0.096	Dardanelli et al. (2004)
Drainage coefficient (β)	mm mm^{-1}	[0.25, 0.75]	Ratliff et al. (1983), Suleiman and Ritchie (2004)	0.55	Ritchie (1998), Suleiman & Ritchie (2004)
Runoff curve number (η)	number	[15, 90]	USDA-SCS (1972)	65[a]	USDA-SCS (1972)
Available water capacity (θ_m)	mm mm^{-1}	[0.05, 0.18]	Ratliff et al. (1983)	0.13	Ratliff et al. (1983), Ritchie et al. (1999)
Root zone depth (ζ)	mm	[300, 600]	USDA-NRCS (1997)	400	Haman et al. (2008), Zazueta et al. (2008)

[a] For a grass landscape that is well managed and has medium-textured soil with average antecedent moisture condition.

Table 3-2. Ranges and distributions selected for various parameters of ARID

Parameter	Distribution	Value					Reference
		Min	Max	Mode	Mean	St. dev.	
α	Empirical[a]	0.60	0.11	0.096	-	-	Dardanelli et al. (2004)
β	Uniform	0.25	0.75	-	-	-	Ritchie (1998), Suleiman and Ritchie (2004)
	Triangular	0.25	0.75	0.55	-	-	
	Normal	-	-	-	0.55	0.06	
η	Uniform	15	90	-	-	-	USDA-SCS (1972)
	Triangular	15	90	65	-	-	
	Normal	-	-	-	65	8	
θ_m	Uniform	0.05	0.18	-	-	-	Ratliff et al. (1983), Ritchie et al. (1999)
	Triangular	0.05	0.18	0.13	-	-	
	Normal	-	-	-	0.13	0.015	
ζ[b]	Uniform	300	600	-	-	-	USDA-NRCS (1997), De Jong and Kabat (1990)

[a] Based on the observations of Dardanelli et al. (2004). [b] For well-managed and frequently-clipped turf grasses, such as those grown in lawns and gardens, on which ARID is based.

Table 3-3. Inverse cumulative distribution functions used for generating samples for various parameters of ARID

Parameter	Distribution	Inverse cumulative distribution function (CDF)	Parameters of inverse CDF
β	Uniform	$\beta_i = a + (b - a)F_{i,\beta}$	a = 0.25, b = 0.75
	Triangular	$\beta_i = \left(F_{i,\beta} < (c-a)/(b-a)\right)\left(a+\sqrt{F_{i,\beta}(b-a)(c-a)}\right)+$ $\left(F_{i,\beta} \geq (c-a)/(b-a)\right)\left(b-\sqrt{(1-F_{i,\beta})(b-a)(b-c)}\right)$	a = 0.25, b = 0.75, c = 0.55
	Normal	$\beta_i = \left(-\mu\sqrt{2}\right)\text{erfcinv}\left(2F_{i,\beta}\right) + \sigma$	μ = 0.55, σ = 0.06
η	Uniform	$\eta_i = a + (b - a)F_{i,\eta}$	a = 15, b = 90
	Triangular	$\eta_i = \left(F_{i,\eta} < (c-a)/(b-a)\right)\left(a+\sqrt{F_{i,\eta}(b-a)(c-a)}\right)+$ $\left(F_{i,\eta} \geq (c-a)/(b-a)\right)\left(b-\sqrt{(1-F_{i,\eta})(b-a)(b-c)}\right)$	a = 15, b = 90, c = 65
	Normal	$\eta_i = \left(-\mu\sqrt{2}\right)\text{erfcinv}\left(2F_{i,\eta}\right) + \sigma$	μ = 65, σ = 8
θ_m	Uniform	$\theta_{m,i} = a + (b - a)F_{i,\theta_m}$	a = 0.05, b = 0.18
	Triangular	$\theta_{m,i} = \left(F_{i,\theta_m} < (c-a)/(b-a)\right)\left(a+\sqrt{F_{i,\theta_m}(b-a)(c-a)}\right)+$ $\left(F_{i,\theta_m} \geq (c-a)/(b-a)\right)\left(b-\sqrt{(1-F_{i,\theta_m})(b-a)(b-c)}\right)$	a = 0.05, b = 0.18, c = 0.13
	Normal	$\theta_{m,i} = \left(-\mu\sqrt{2}\right)\text{erfcinv}\left(2F_{i,\theta_m}\right) + \sigma$	μ = 0.13, σ = 0.015
ζ	Uniform	$\zeta_i = a + (b - a)F_{i,\zeta}$	a = 300, b = 600

Table 3-4. Procedure followed for generating samples for water uptake coefficient (α)

Solution point ($F_{i,\alpha}$)	Parameters of inverse cumulative distribution function (CDF)	Inverse CDF
$0.00 < F_{i,\alpha} \leq 0.10$	a = 0.06, b = 0.07	$\alpha_i = a + (b-a)F_{i,\alpha}$
$0.10 < F_{i,\alpha} \leq 0.20$	a = 0.07, b = 0.08	
$0.20 < F_{i,\alpha} \leq 0.28$	a = 0.08, b = 0.09	
$0.28 < F_{i,\alpha} \leq 0.73$	a = 0.09, b = 0.10	
$0.73 < F_{i,\alpha} \leq 1.00$	a = 0.10, b = 0.11	

Table 3-5. Correlations among parameters for several locations (with triangular distributions of η, β, and θ_m)

Location	Parameter	Parameter			
		η	β	α	ζ
Miami	β	0.01	1.00		
	α	0.07	0.03	1.00	
	ζ	0.01	0.00	0.00	1.00
	θ_m	0.00	0.00	0.04	0.00
Bartow	β	0.01	1.00		
	α	0.07	0.04	1.00	
	ζ	0.00	0.00	0.00	1.00
	θ_m	0.00	0.00	0.02	0.00
Live Oak	β	0.00	1.00		
	α	0.03	0.02	1.00	
	ζ	0.00	0.00	0.02	1.00
	θ_m	0.00	0.00	0.02	0.00
Plains	β	0.00	1.00		
	α	0.06	0.03	1.00	
	ζ	0.00	0.00	0.02	1.00
	θ_m	0.00	0.00	0.02	0.00
Blairsville	β	0.00	1.00		
	α	0.06	0.00	1.00	
	ζ	0.00	0.00	0.01	1.00
	θ_m	0.00	0.00	0.02	0.00

Table 3-6. Main- and total-effect sensitivity indices of various parameters for different locations and seasons (with triangular distributions of η, β, and θ_m). Values of the total-effect sensitivity index less than 0.10 indicate insignificant influence on ARID.

Season	Parameter	Location									
		Miami		Bartow		Live Oak		Plains		Blairsville	
		S_j [a]	S_j^T [b]	S_j	S_j^T	S_j	S_j^T	S_j	S_j^T	S_j	S_j^T
Winter	η	0.01	0.03	0.00	0.03	0.00	0.02	0.00	0.04	0.00	0.04
	β	0.01	0.03	0.01	0.04	0.01	0.03	0.01	0.04	0.01	0.06
	α	0.03	0.05	0.03	0.06	0.05	0.07	0.08	0.19	0.04	0.11
	ζ	0.36	0.39	0.34	0.39	0.30	0.33	0.26	0.38	0.25	0.34
	θ_m	0.55	0.61	0.56	0.62	0.55	0.61	0.51	0.65	0.54	0.69
Spring	η	0.01	0.02	0.00	0.03	0.00	0.02	0.00	0.02	0.00	0.02
	β	0.01	0.03	0.01	0.03	0.01	0.02	0.01	0.02	0.01	0.03
	α	0.09	0.11	0.06	0.08	0.05	0.06	0.07	0.10	0.08	0.10
	ζ	0.32	0.36	0.36	0.38	0.35	0.36	0.36	0.38	0.31	0.34
	θ_m	0.52	0.57	0.54	0.57	0.54	0.57	0.55	0.58	0.54	0.59
Summer	η	0.00	0.01	0.00	0.02	0.00	0.02	0.00	0.02	0.00	0.02
	β	0.02	0.04	0.02	0.05	0.01	0.03	0.01	0.03	0.01	0.03
	α	0.15	0.18	0.14	0.16	0.11	0.14	0.09	0.13	0.08	0.10
	ζ	0.31	0.34	0.31	0.34	0.31	0.33	0.33	0.36	0.32	0.34
	θ_m	0.50	0.53	0.50	0.52	0.51	0.55	0.53	0.57	0.54	0.59
Fall	η	0.00	0.01	0.00	0.02	0.00	0.01	0.00	0.02	0.00	0.02
	β	0.01	0.03	0.01	0.03	0.01	0.02	0.01	0.02	0.01	0.03
	α	0.05	0.08	0.04	0.05	0.02	0.03	0.02	0.06	0.03	0.06
	ζ	0.35	0.37	0.37	0.39	0.36	0.38	0.37	0.40	0.32	0.35
	θ_m	0.55	0.59	0.55	0.58	0.56	0.58	0.58	0.62	0.57	0.62

[a] Main-effect sensitivity index for parameter j, where j = η, β, α, ζ, or θ_m. [b] Total-effect sensitivity index for parameter j.

Table 3-7. Interaction-effect indices, computed as the difference between total-effect sensitivity index and main-effect sensitivity index, of various parameters for different locations and seasons. Values of the interaction-effect index less than 0.10 indicate insignificant interactions.

Season	Parameter	Location				
		Miami	Bartow	Live Oak	Plains	Blairsville
Winter	η	0.02	0.03	0.02	0.04	0.04
	β	0.02	0.03	0.02	0.04	0.05
	α	0.02	0.02	0.03	0.10	0.09
	ζ	0.03	0.05	0.03	0.12	0.10
	θ_m	0.05	0.06	0.07	0.14	0.15
Spring	η	0.02	0.02	0.01	0.02	0.02
	β	0.02	0.02	0.01	0.01	0.02
	α	0.02	0.02	0.01	0.03	0.03
	ζ	0.04	0.03	0.01	0.02	0.03
	θ_m	0.05	0.04	0.02	0.03	0.05
Summer	η	0.01	0.02	0.02	0.02	0.02
	β	0.02	0.03	0.02	0.02	0.02
	α	0.03	0.03	0.02	0.03	0.03
	ζ	0.03	0.03	0.02	0.03	0.03
	θ_m	0.04	0.04	0.04	0.04	0.05
Fall	η	0.01	0.02	0.01	0.02	0.02
	β	0.02	0.02	0.01	0.01	0.02
	α	0.02	0.02	0.01	0.04	0.03
	ζ	0.03	0.02	0.01	0.03	0.03
	θ_m	0.04	0.03	0.02	0.04	0.05

Table 3-8. Nominal values of ARID for different locations and seasons

Season	Location				
	Miami	Bartow	Live Oak	Plains	Blairsville
Winter	0.55	0.42	0.24	0.06	0.00
Spring	0.60	0.54	0.34	0.30	0.14
Summer	0.22	0.25	0.30	0.38	0.25
Fall	0.27	0.30	0.34	0.39	0.17

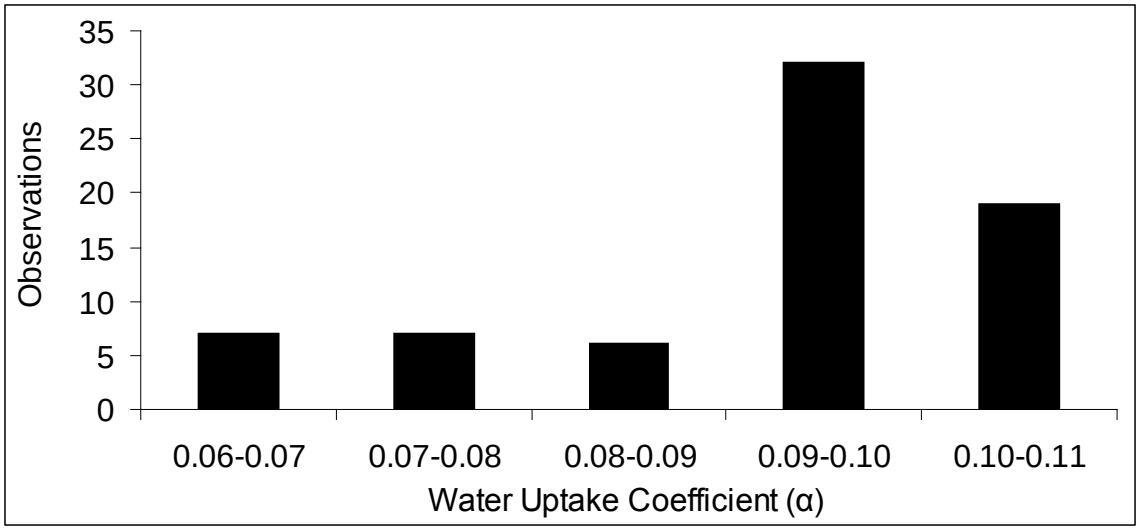

Figure 3-1. The distribution of parameter water uptake coefficient as observed by Dardanelli et al. (2004)

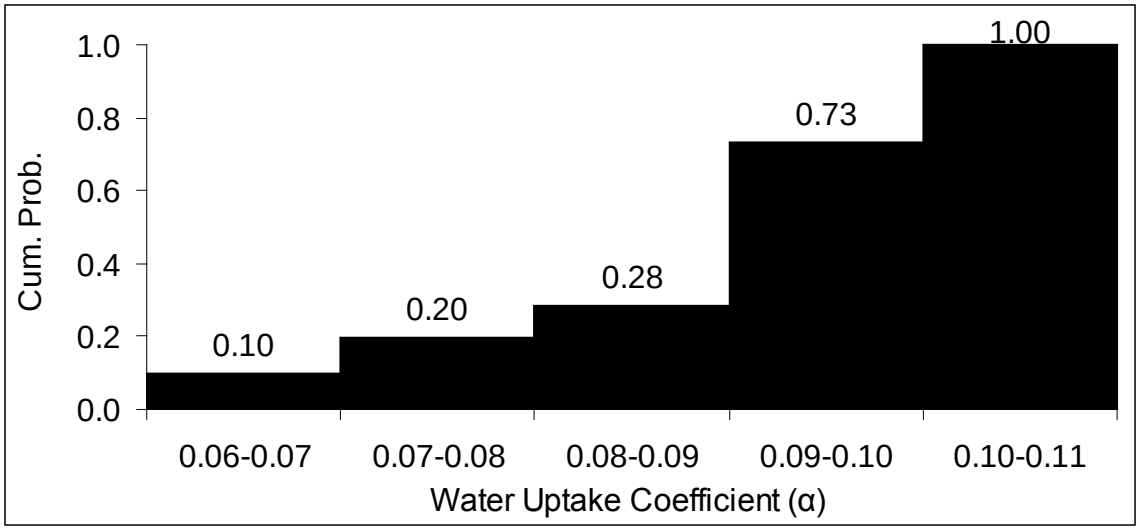

Figure 3-2. The probability histogram for the cumulative distribution of parameter water uptake coefficient

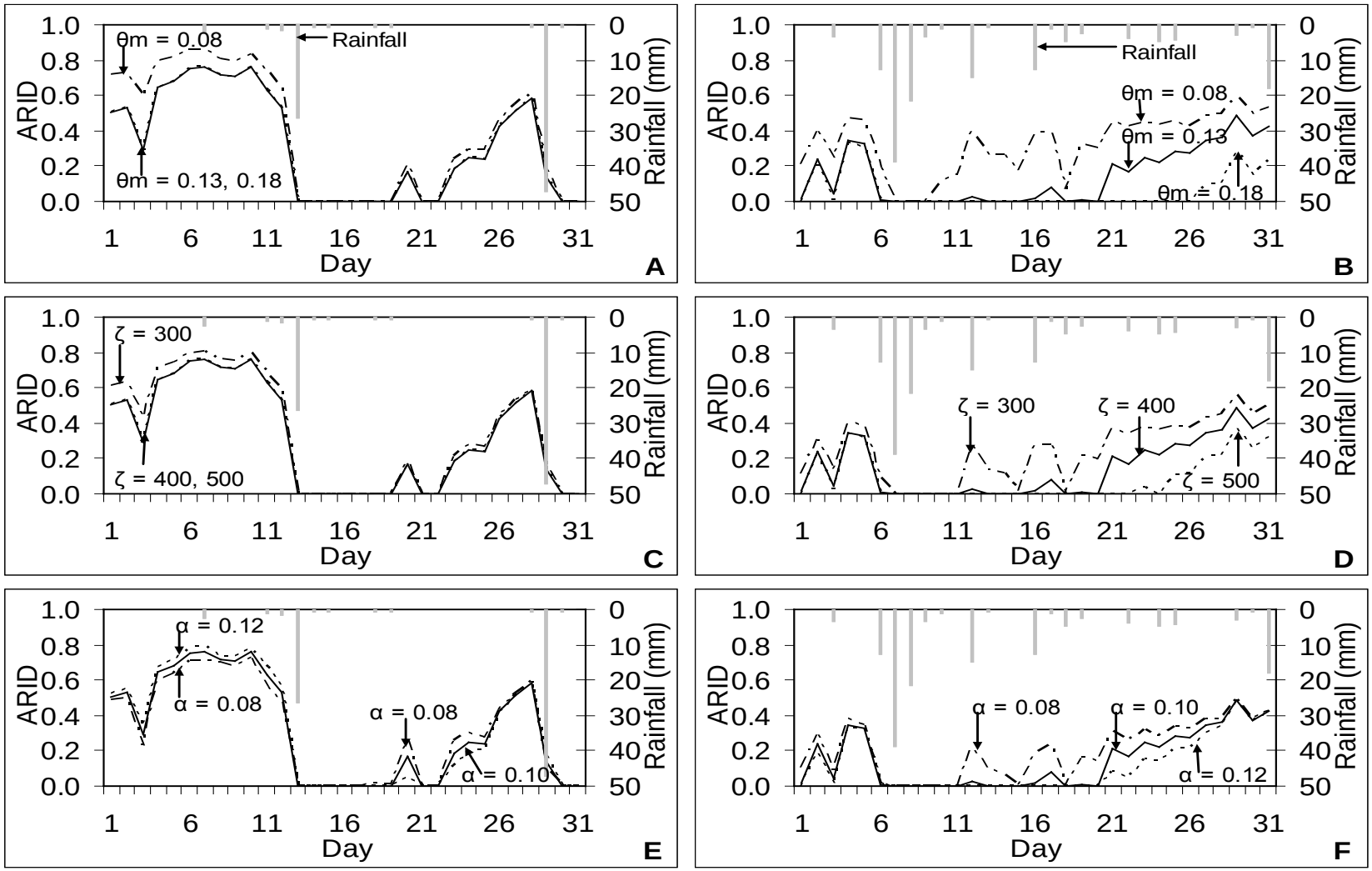

Figure 3-3. Response of ARID to precipitation in: A) January with 3 values of θ_m, B) July with 3 values of θ_m, C) January with 3 values of ζ, D) July with 3 values of ζ, E) January with 3 values of α, and F) July with 3 values of α. The ARID and precipitation values are for Gainesville, FL, in January and July 2009. θ_m = available water capacity, ζ = rootzone depth, α = water uptake coefficient.

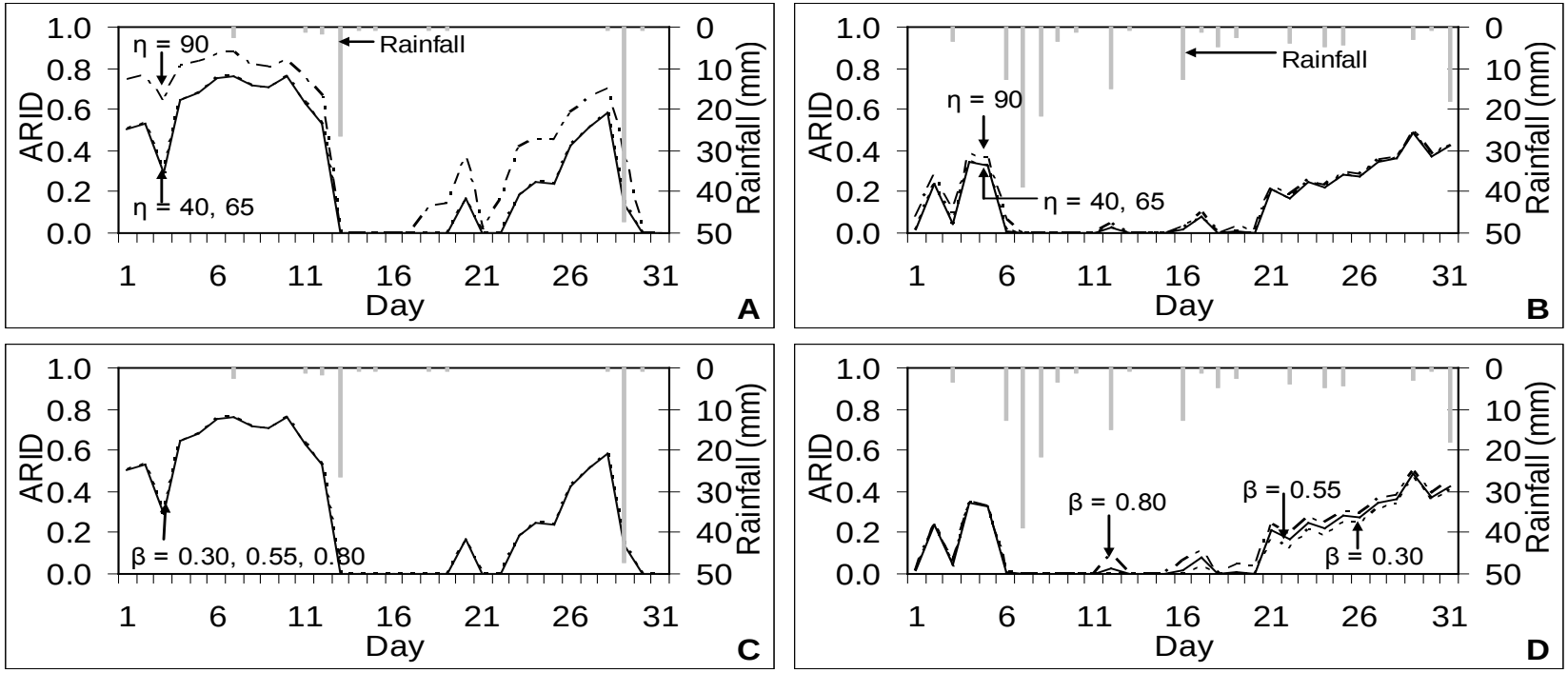

Figure 3-4. Response of ARID to precipitation in: A) January with 3 values of η, B) July with 3 values of η, C) January with 3 values of β, and D) July with 3 values of β. The ARID and precipitation values are for Gainesville, FL, in January and July 2009. η = runoff curve number, β = drainage coefficient.

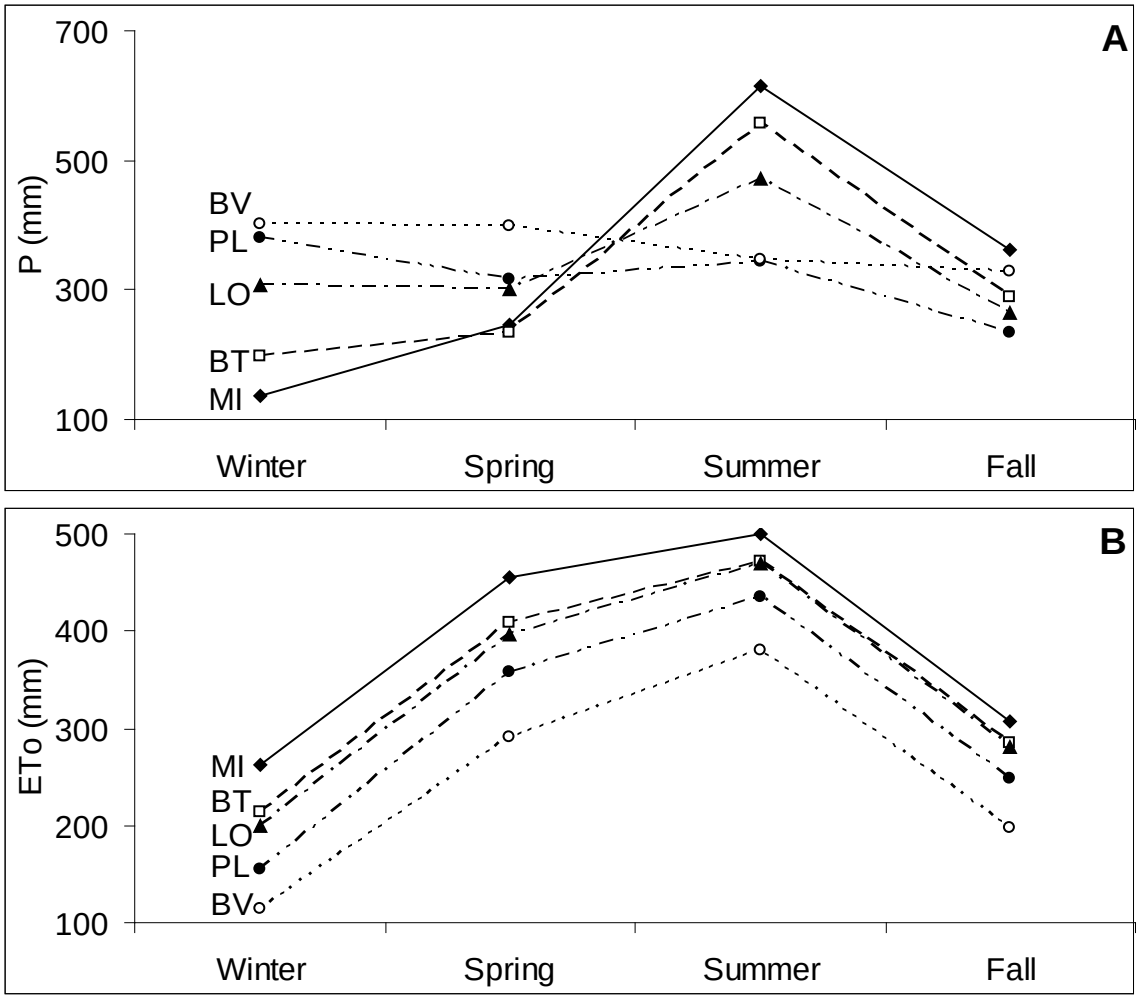

Figure 3-5. Amounts of: A) Precipitation (P) and B) ET_o for Miami (MI), Bartow (BT), Live Oak (LO), Plains (PL), and Blairsville (BV) in different seasons during the normal period 1971-2000

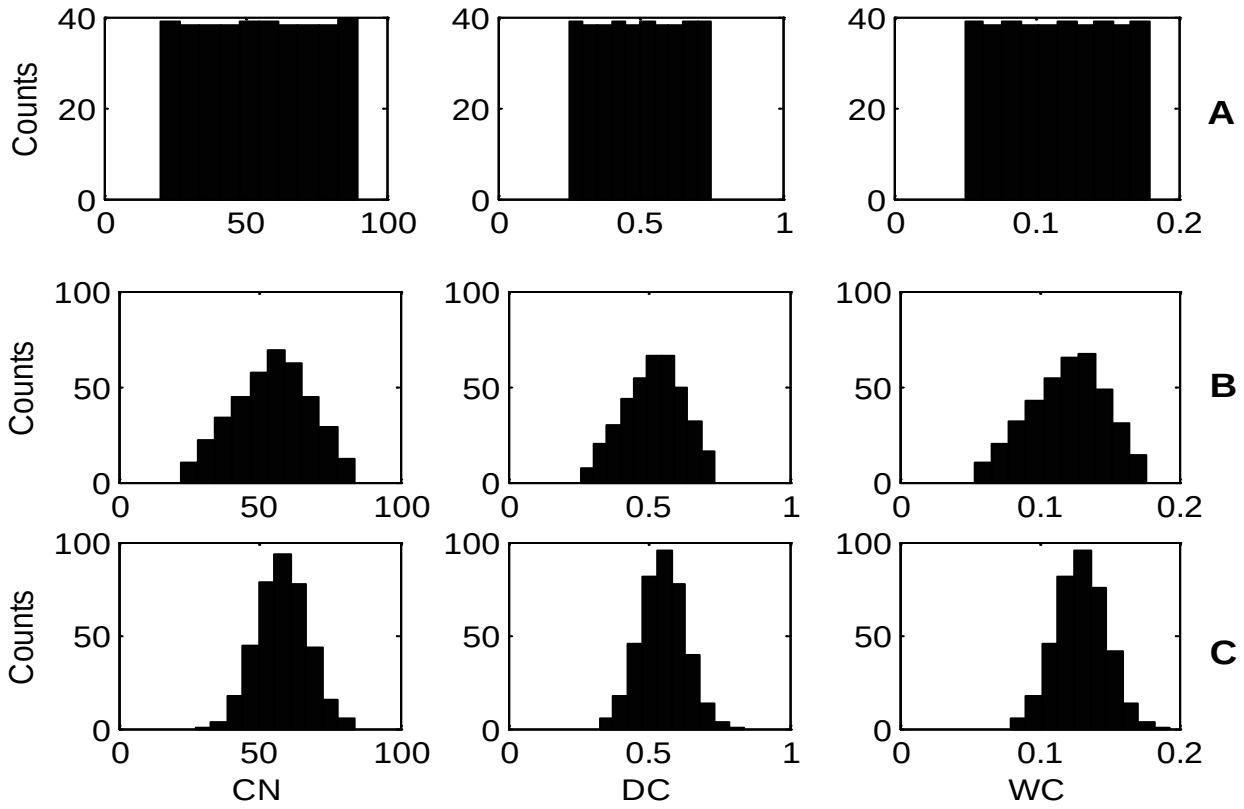

Figure 3-6. A) Uniform, B) triangular, and C) normal distributions of parameters runoff curve number (CN), drainage coefficient (DC), and available water capacity (WC)

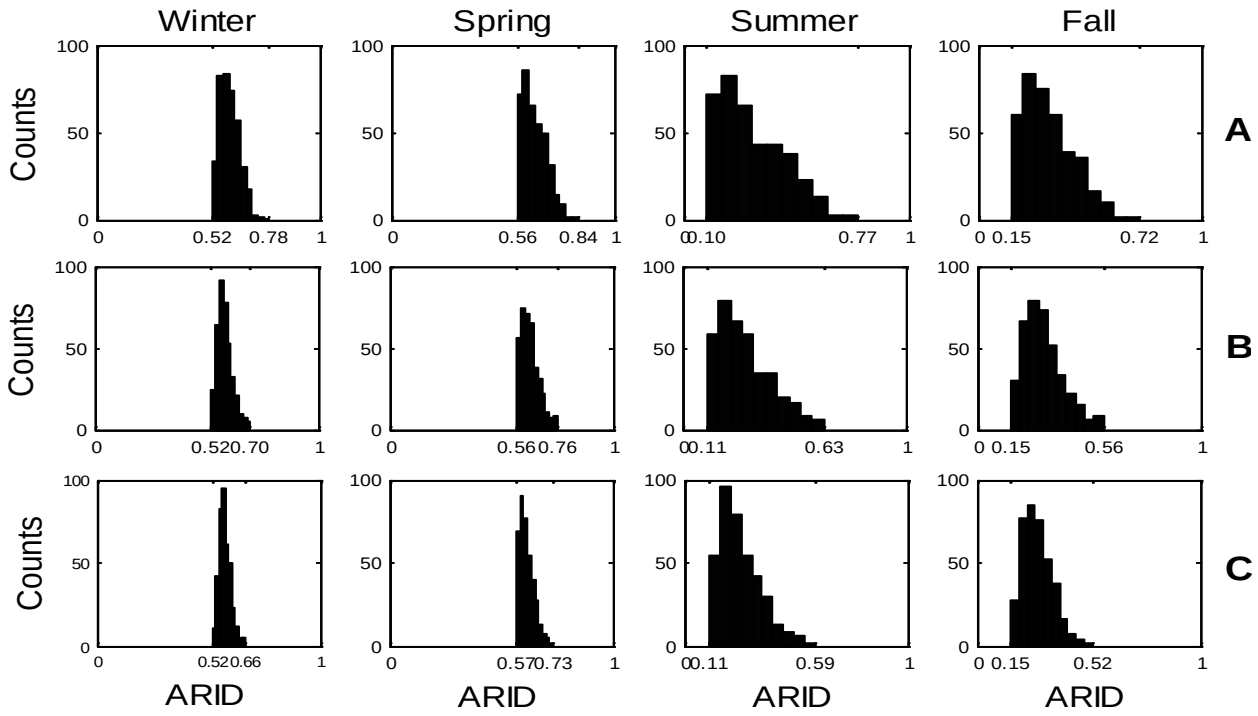

Figure 3-7. Uncertainty ranges of ARID for Miami with: A) uniform, B) triangular, and C) normal distributions of parameters drainage coefficient, curve number, and available water capacity

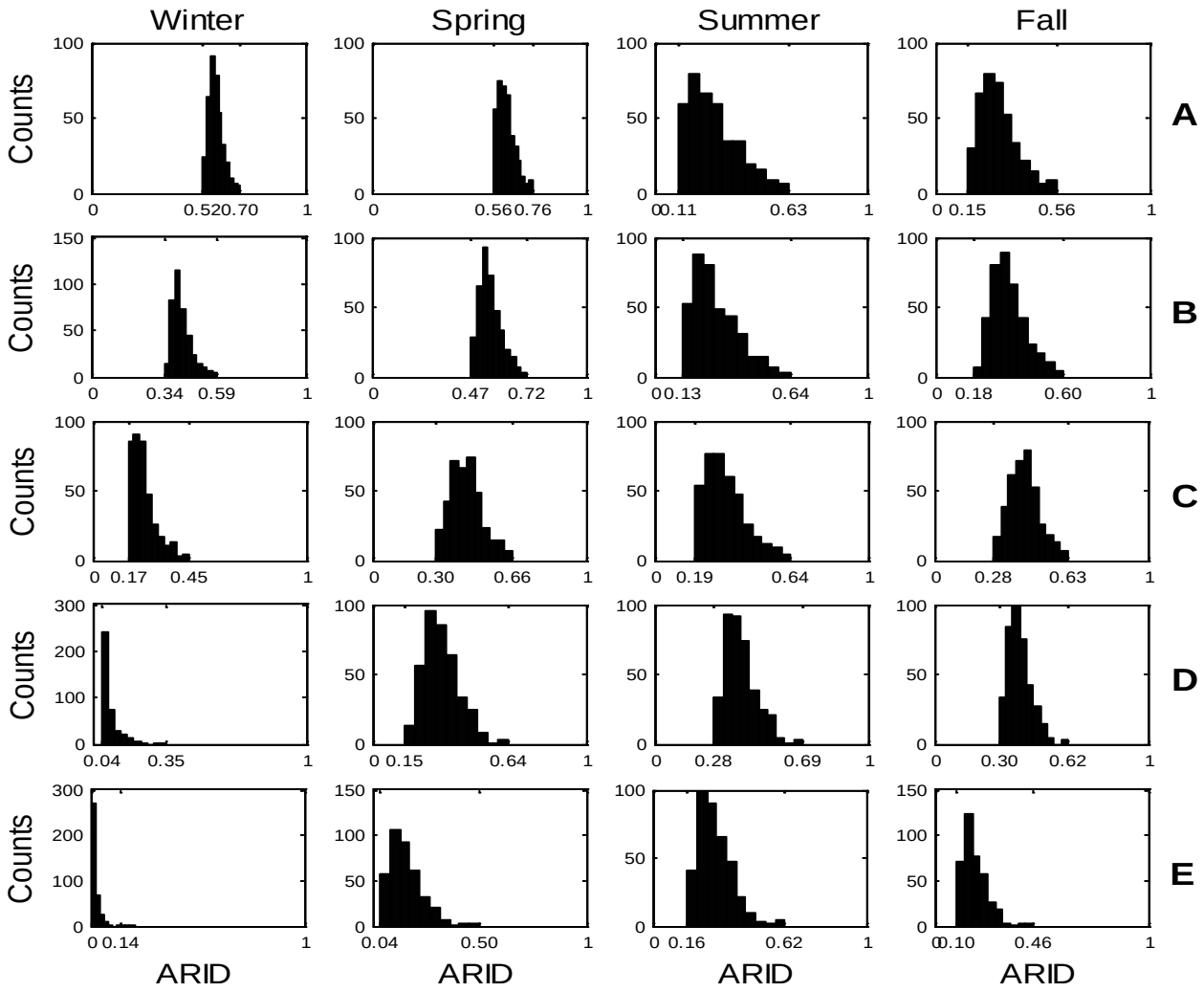

Figure 3-8. Variability in ARID in different seasons for: A) Miami, B) Bartow, C) Live Oak, D) Plains, and E) Blairsville with triangular distributions of parameters drainage coefficient, curve number, and available water capacity

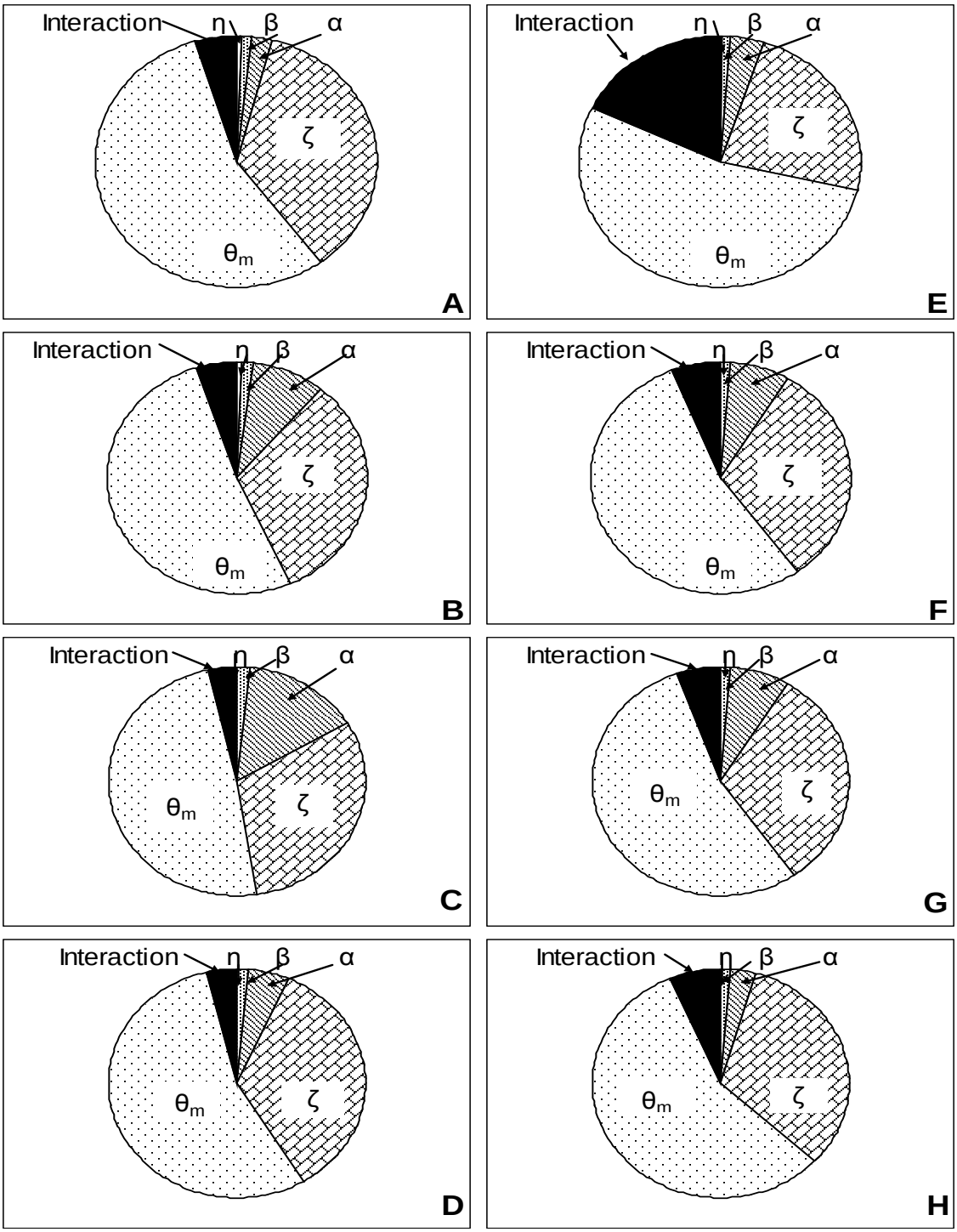

Figure 3-9. The relative proportion of the total main-effect sensitivity across parameters (η = curve number, β = drainage coefficient, ζ = rootzone depth, α = water uptake coefficient, and θ_m = available water capacity) for: A) winter, B) spring, C) summer, and D) fall for Miami (far south in the southeast USA) and for E) winter, F) spring, G) summer, and H) fall for Blairsville (far north in the southeast USA).

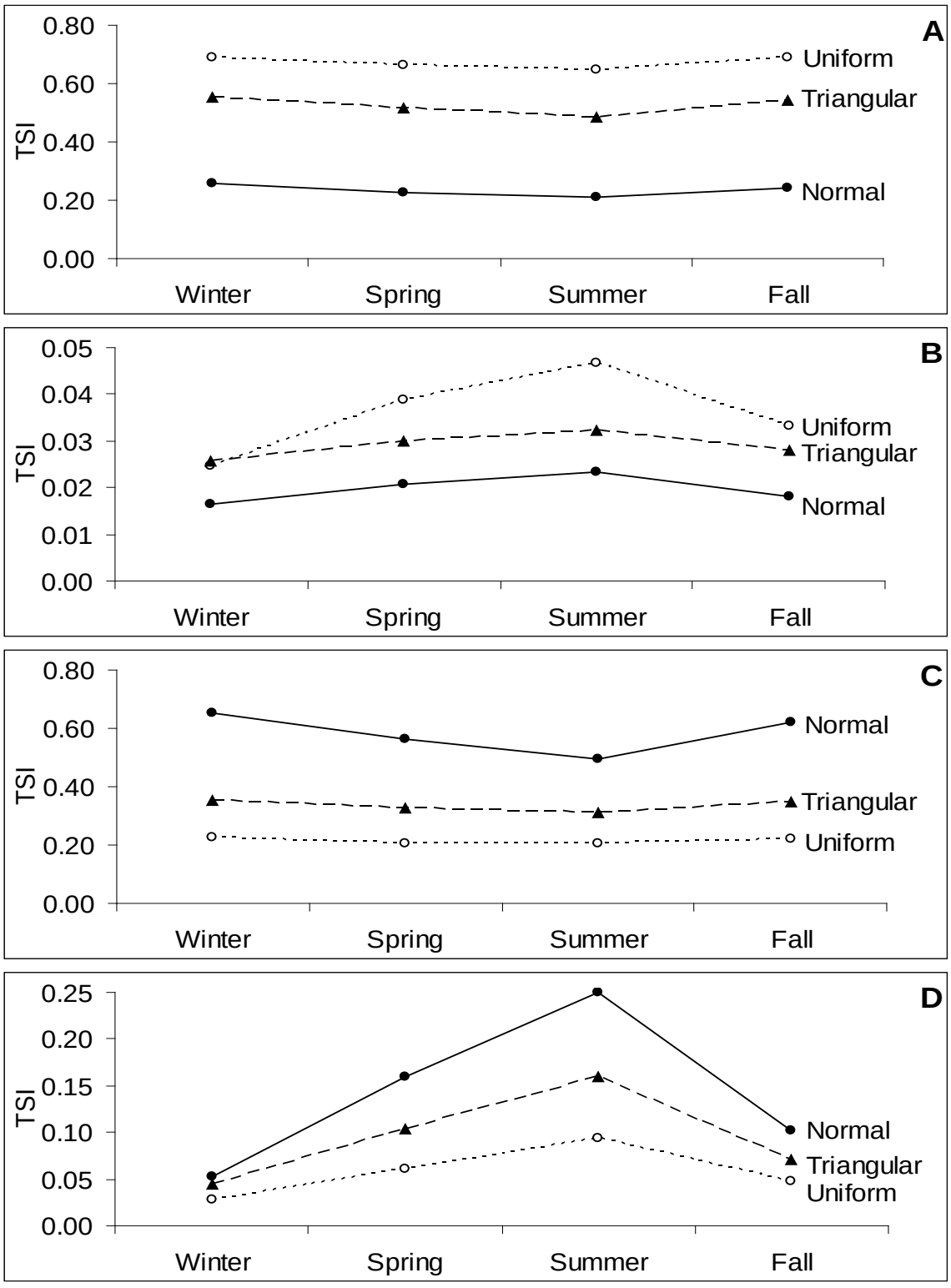

Figure 3-10. Total sensitivity index (TSI) values of: A) available water capacity, B) drainage coefficient, C) rootzone depth, and D) water uptake coefficient for different seasons and distributions for Miami

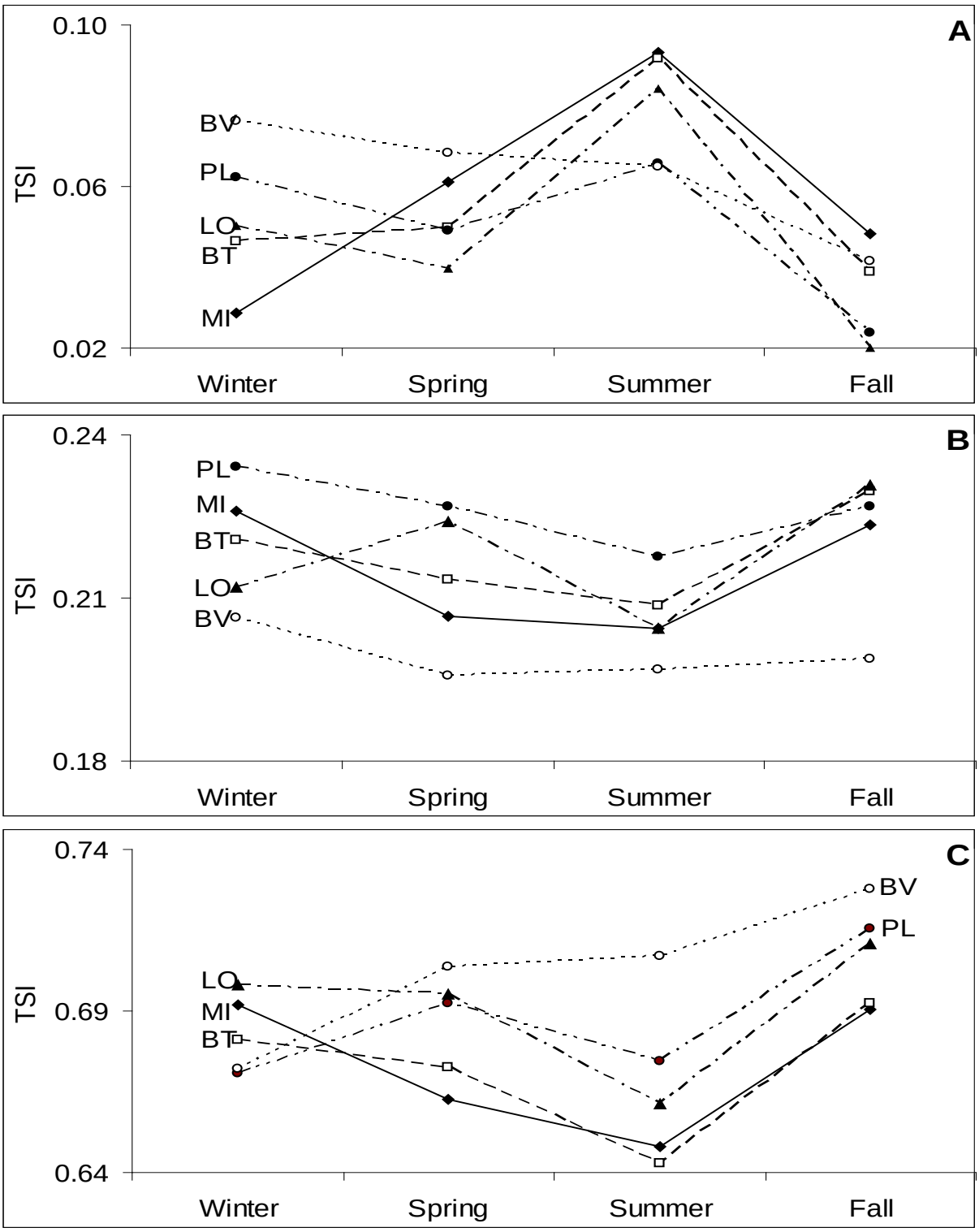

Figure 3-11. Values of the total-effect sensitivity index (TSI) of: A) water uptake coefficient, B) rootzone depth, and C) available water capacity for various locations (BT = Bartow, BV = Blairsville, LO = Live Oak, MI = Miami, and PL = Plains) and seasons with triangular distributions

CHAPTER 4
FORECASTING DROUGHT USING AGRICULTURAL REFERECE INDEX FOR DROUGHT

Introduction

Drought is a creeping and the costliest natural hazard, causing tremendous losses worldwide. In the USA, about $6 - $8 billion is lost every year through drought alone (FEMA, 1995; SDR, 2005). Although drought cannot be prevented, its losses can be minimized through taking proactive measures and setting out drought mitigation strategies if it is forecast well in advance. Because it evolves slowly, mitigation plans can be carried out more effectively for drought than for those of other natural hazards. The benefits of reliable forecasting are tremendous (Campbell, 1973).

Drought may be forecast using drought variables (Hastenrath and Greischar, 1993; Chiew et al., 1998; Cordery and McCall, 2000) or drought indicators (Nicholls, 1985; Chiew et al., 1998; Lohani et al., 1998; Mishra and Desai, 2005; Cancelliere et al., 2007; NIDIS, 2007; D'Arrigo and Wilson, 2008; Mendicino et al., 2008; Vasiliades and Loukas, 2008). A drought variable is a key element that defines drought, such as precipitation for a meteorological drought or soil moisture and ET for an agricultural drought. A drought indicator is a measure used to quantify and characterize drought conditions. For forecasting a meteorological drought, the forecasting of precipitation may suffice. However, for forecasting an agricultural drought, a shortage of water in the root zone to meet the consumptive use of plants, forecasting all associated variables is required. This is difficult, because these variables are controlled by several plant and soil processes. Therefore, for agricultural drought, the use of a drought indicator, such as a drought index, seems more appropriate than using a drought variable.

A drought index may be forecast based on various approaches. One approach is to estimate the index as a function of its own past values (Panu and Sharma, 2002; Kim and Valdes, 2003; Nayak et al., 2004; Bordi et al., 2005; Mishra and Desai, 2005; Cancelliere et al., 2007; Mendicino et al., 2008; Vasiliades and Loukas, 2008). This approach involves quantifying a relationship between the present and past values of the index by fitting the data to a time series model, such as an autoregressive moving average (ARMA) model. Another approach is to estimate the index as a function of the past values of its predictor variables. Climate or teleconnection indices may be used as predictors of drought because atmospheric circulation patterns have been shown to exert influence on the occurrence of droughts (Stahl and Demuth, 1999). The linkage between changes in the atmosphere in one place and effects on the weather in another, widely separated place is called teleconnection and is a consequence of the large-scale dynamics of the ocean and atmosphere. The indices of large-scale oceanic and atmospheric teleconnection patterns, or oscillations, are called climate or teleconnection indices. Climate indices combine many details into a generalized, overall description of the atmosphere or ocean, and this can be used to characterize the factors that affect a local or global climate system. Although climate indices are generally calculated based on measurements from a localized area, they can be statistically related with climate variables of other areas around the globe (Bond et al., 2007). Because large-scale teleconnection patterns affect weather variables such as temperature and precipitation, the key factors governing a drought index, a plausible relationship may exist between a drought index and climate indices. Some of the major, large-scale teleconnection patterns that are reported to have significant impacts on the weather conditions or

agriculture in southeastern USA are the Atlantic Multi-decadal Oscillation (AMO) (Gershunov and Barnett, 1998b; Enfield et al., 2001; Kelly, 2007), the North Atlantic Oscillation (NAO) (Rogers, 1984; Yin, 1994; Hagemeyer, 2006; Kelly, 2007), the Pacific Decadal Oscillation (PDO) (Gershunov and Barnett, 1998b; Enfield et al., 2001; Kelly, 2007), the Pacific-North American (PNA) pattern (Leathers et al., 1991; Yin, 1994; Konrad, 1998; Hagemeyer, 2006; Martinez et al., 2009), the sea surface temperature (SST) in the Niño 3.4 region (Trenberth, 1997; Gershunov and Barnett, 1998b; Schmidt et al., 2001; Hanley et al., 2003; Hagemeyer, 2006), and the SST over the region of (4°N-4°S, 150°W-90°W) (Hansen et al., 1998). The AMO index is basically an index of ongoing series of long-duration changes in the sea surface temperature of the North Atlantic Ocean over 0°N-70°N. The NAO index is the difference of sea-level pressure between two stations situated close to the centers of the Icelandic Low and Azores High. The PNA index is a linear combination of the normalized geopotential height anomalies at the 700 hPa level at four locations, namely, Hawaii (20°N, 160°W), the North Pacific Ocean (45°N, 165°W), Alberta (55°N, 115°W), and the Gulf Coast of the USA (30°N, 85°W). The PDO Index is the spatial average of the monthly sea surface temperature of the Pacific Ocean north of 20°N. The Niño 3.4 index is the departure in monthly sea surface temperatures from their long-term means over the Niño 3.4 region, i.e., the east central tropical Pacific (5°N-5°S, 170°W-120°W). The Japanese Meteorological Agency (JMA) index is a 5-month running mean of the departure in monthly sea surface temperatures from their long-term means over the region of (4°N-4°S, 150°W-90°W). The definitions of these and other abbreviations that are frequently mentioned in the following paragraphs are presented in Table 4-1.

The most commonly used approach for forecasting precipitation or drought in southeastern USA is based on the El Niño/Southern Oscillation (ENSO). The strong teleconnection between ENSO and the weather conditions in this region has enabled skillful forecasting of seasonal temperature and precipitation up to one year in advance (Steinemann, 2006; Brolley et al., 2007). However, the drought-forecasting ability of the ENSO approach varies across months and locations because the impact of ENSO differs from region to region and from season to season. That is, the ENSO-based approach might produce plausible forecasts for some months of the year and inaccurate forecasts for other months. This situation necessitates the search for other approaches. The principal objective of this study was to find out if the present level of forecasting, which is mainly ENSO-based, could be improved using the other approaches, namely, the use of ARMA models and climate indices. That is, could any of these approaches / models produce more accurate forecasts than those produced using the ENSO approach at least for the months or locations for which the predictability of ENSO is not good?

To quantify the relationship between a predictand and its predictors, various statistical models have been used, such as linear regression (LR) (Yin, 1994; Kumar and Panu, 1997; Panu and Sharma, 2002; Wood and Lettenmaier, 2006; D'Arrigo and Wilson, 2008; Vasiliades and Loukas, 2008), artificial neural network (ANN) (Werbos, 1974; Sharda and Patil, 1990; Hastenrath and Greischar, 1993; Wilpen et al., 1994; Zhang et al., 1998; Panu and Sharma, 2002; Samarasinghe, 2007; Bacanli et al., 2008; Vasiliades and Loukas, 2008), and the adaptive neuro-fuzzy inference system (ANFIS) (Nayak et al., 2004; Chang and Chang, 2006; Bacanli et al., 2008; Firat and Gungor,

2008). Several researchers have used these models, either individually or in combination, to forecast various hydrological variables. However, no agricultural droughts or their indicators were forecast using these models. The other purpose of this study was to assess the performances of the ANFIS, ANN, ARMA, and LR models relative to that of the ENSO approach in forecasting ARID. To evaluate the performance of these models, the ENSO-based forecasting approach was used as the benchmark. The assumption was that if the forecasts of any of these models were better than those of the ENSO approach for any month or location in the southeast USA, drought forecasting could be improved for that month or location in this region.

Overview of the Models

Autoregressive Moving Average Models

An ARMA model (Box and Jenkins, 1970) consists of two parts: autoregression (AR) and a moving average (MA). The AR model estimates the current value of a series, e.g., $ARID_t$, as a function of **p** past values: $ARID_{t-1}$, $ARID_{t-2}$, …, $ARID_{t-p}$, where **p** determines the number of steps into the past needed to forecast the current value. The autoregressive model of order **p**, i.e., AR(p), is of the form:

$$ARID_t = \phi_1 ARID_{t-1} + \phi_2 ARID_{t-2} + ... + \phi_p ARID_{t-p} + \varepsilon_t \quad (4\text{-}1)$$

where ϕ_1, ϕ_2, …, ϕ_p are the modal parameters and ε_t is Gaussian white noise (a random process).

The MA model of order **q**, i.e., MA(q), which assumes that the white noise ε_t can be combined linearly to form the observed data, is of the form:

$$ARID_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} \quad (4\text{-}2)$$

where there are **q** lags in the moving average; $\theta_1, \theta_2, \ldots, \theta_q$ are parameters; and ε_t is the Gaussian white noise.

Combining Equations 4-1 and 4-2 produces the autoregressive moving average (ARMA) model of order **p** and **q**, i.e., ARMA(p,q):

$$ARID_t = \phi_1 ARID_{t-1} + \phi_2 ARID_{t-2} + ... + \phi_p ARID_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} \quad (4\text{-}3)$$

Rearranging Equation 4-3 and applying a backshift operator to denote $ARID_{t-k}$ as $B^k ARID_t$ and ε_{t-k} as $B^k \varepsilon_t$ produces the following equation:

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p) ARID_t = (1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q) \varepsilon_t \quad (4\text{-}4)$$

Defining $(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$ as $\phi(B)$ and $(1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)$ as $\theta(B)$ produces the ARMA(p,q) model in concise form as:

$$\phi(B) ARID_t = \theta(B) \varepsilon_t \quad (4\text{-}5)$$

Equation 4-5 does not account for seasonality. Most natural phenomena have strong seasonal components. Often, the dependence on the past tends to occur most strongly at multiples of some underlying seasonal lag, **s**. This necessitates the introduction of ARMA polynomials that identify with the seasonal lags. Then, the pure seasonal ARMA model, i.e., $ARMA(P,Q)_s$, takes the form

$$\Phi_P(B^s) ARID_t = \Theta_Q(B^s) \varepsilon_t \quad (4\text{-}6)$$

To account for both seasonal and nonseasonal behaviors, Equations 4-5 and 4-6 can be combined to produce a multiplicative seasonal ARMA model, denoted by $ARMA(p,q) \times (P,Q)_s$:

$$\Phi_P(B^s) \phi(B) ARID_t = \Theta_Q(B^s) \theta(B) \varepsilon_t \quad (4\text{-}7)$$

A time series process can have a nonstationary trend component. Differencing such a process leads to a stationary process. The inclusion of differencing broadens an ARMA model to an integrated ARMA model, called an ARIMA model. The nonseasonal ARMA(p,q) extends to ARIMA(p,d,q) and takes the form:

$$\phi(B)\nabla^d ARID_t = \theta(B)\varepsilon_t \tag{4-8}$$

where ∇^d is the difference of order **d**, defined as $\nabla^d = (1-B)^d$.

Similarly, the pure seasonal ARIMA$(P,D,Q)_s$ has the form:

$$\Phi_P(B^s)\nabla_s^D ARID_t = \Theta_Q(B^s)\varepsilon_t \tag{4-9}$$

where ∇_s^D is the difference of order **D** with seasonal lag, **s**, defined as $\nabla_s^D = (1-B^s)^D$.

Combining Equations 4-8 and 4-9 produces the multiplicative, seasonal, autoregressive, integrated, moving average model, called the SARIMA model, denoted by ARIMA(p,d,q) x $(P,D,Q)_s$. The SARIMA model accounts for both seasonal and nonseasonal behaviors and is stationary.

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d ARID_t = \Theta_Q(B^s)\theta(B)\varepsilon_t \tag{4-10}$$

The ARMA models have several strengths. They provide systematic procedures for modeling the behavior of uncertain systems and quantifying the expected inaccuracy of the forecasts (Mishra and Desai, 2005). They can model the periodicity and nonstationarity of time series data (Mishra et al., 2007), effectively take into account serial correlations among observations, and require few parameters. As stochastic models, they can quantify the stochastic nature of drought (Modarres, 2006).

Use of the ARMA models also has some limitations. They assume that the time series are generated from linear processes. This assumption is inappropriate if the underlying mechanism is nonlinear, and drought time series are rarely linear and often

seasonal (Granger and Terasvirta, 1993; Zhang et al., 1998). They have limited ability to capture nonstationarity and nonlinearity in the data and are unsuitable for long-term forecasting (Tokar and Johnson, 1999; Kim and Valdes, 2003; Mishra et al., 2007). They are also difficult to implement and require more training in their use (Marce et al., 2004).

Linear Regression

Linear regression models the relationship between a regressand and one or more of its regressors. A LR model assumes the relationship between dependent and independent variables to be approximately linear. One of the strengths of a LR model is that it explicitly addresses causal relationships and shows optimal results when the underlying mechanism is linear. A LR model is useful in estimating and testing the parameters, testing the residuals, and carrying out diagnostics (Belsley et al., 1980).

The use of a LR model is inappropriate, however, when the relationship is nonlinear. It is limited to estimating numeric outputs and has a limited ability to capture the underlying mechanism of a complex system (Zhang et al., 1998). A LR model does not quite capture the trend of the data and is inadequate in providing good prognostication of a variable characterized by a highly nonlinear physics (Ramirez et al., 2005).

Artificial Neural Network

An artificial neural network is a system of interconnected neurons, where each neuron processes data using the concepts of learning in the brain. ANNs are adaptive learning systems that follow the observed data freely to find patterns in the data and develop nonlinear system models to make estimations. Of a variety of existing ANNs, multilayer perceptron (MLP) is the most well-known, widely-used, and influential

network for nonlinear estimation (Hornik et al., 1989; Zhang et al., 1998; Furukawa et al., 2007; Samarasinghe, 2007; Abbasi, 2009). It contains many links connecting inputs to neurons, and neurons to outputs. Figure 4-1 illustrates the structure of a simple MPL network that has one input layer, one hidden layer, and one output layer; each containing one neuron. These layers are linked by connection strengths termed weights. The 1-1-1 MLP network has one hidden layer weight, **a**, and one output layer weight, **α**. The network-input is PC1 and the network-output is $\widehat{\text{ARID}}$. The hidden neuron has a bias input of +1 with an associated weight of **b**, and the output neuron has a bias weight of **β**. The hidden-layer neuron uses a hyperbolic tangent sigmoid function as the transfer function, whereas the output neuron uses a linear function.

The network computes the final output, $\widehat{\text{ARID}}$, as follows:

Step 1: The input, PC1, is fed into the network. The most important part of processing takes place in the hidden neuron. The hidden neuron, first, calculates its weighted input, **m**. The result of this operation is to map PC1 linearly onto **m** with slope **a** and intercept **b**. As the intercept **b** can be visualized as incorporating the effects of all inputs other than PC1 that are not explicitly involved in the model, it may be called the bias weight.

$$m = aPC1 + b \tag{4-11}$$

Step 2: The hidden neuron passes its input **m** through an input transfer function, e.g., the hyperbolic tangent function, to produce the hidden-neuron output, **h**.

$$h = \frac{e^{m} - e^{-m}}{e^{m} + e^{-m}} \tag{4-12}$$

Step 3: The hidden-neuron output **h** is fed as input into the output-neuron. The output neuron calculates its weighted input, **n**, first. As in the hidden neuron, **h** is linearly mapped to **n** with a slope **α** and the intercept **β** (the bias weight).

$$n = \alpha h + \beta \tag{4-13}$$

Step 4: The output neuron now passes its input **n** through an output transfer function, such as a linear function, to produce the output-neuron output, $\hat{A}RID$, which is the network output. The linear transfer function produces the same output as its input.

$$\hat{A}RID = n \tag{4-14}$$

This concludes forward processing in the 1-1-1 network. The network output, $\hat{A}RID$, is linear with respect to **n**, but is nonlinear with respect to the original input, PC1, due to the nonlinear processing in the hidden neuron:

$$\hat{A}RID = \alpha\left(\frac{e^{aPC1+b} - e^{-(aPC1+b)}}{e^{aPC1+b} + e^{-(aPC1+b)}}\right) + \beta \tag{4-15}$$

This structure is trained to learn through supervised learning by repeated exposure to the training dataset until the network produces the correct output. During training, the transfer functions undergo transformations in shape and position through changes in the corresponding weights until the network produces the output at the desired degree of accuracy. The MLP output ($\hat{A}RID$) is compared with the target output (ARID) and the error is calculated. If the absolute error is larger than the given threshold, the error is backpropagated through the network. This process adjusts the weights, i.e., **a**, **b**, **α**, and **β**, using an appropriate learning method, such as gradient descent (backpropagation), to minimize the error in the repeated processing of the inputs by the network. The error threshold defines the accuracy of the model.

ANN is gaining tremendous momentum for forecasting because of its non-linearity, adaptability, generalization, model-independence, and arbitrary function-mapping ability (Wilpen et al., 1994; ASCE, 2000; Kim and Valdes, 2003; Sahoo et al., 2005). It is capable of acquiring a knowledge of complex processes, especially when the underlying physical relationships are not fully understood (Wilpen et al., 1994; Bacanli et al., 2008). ANN is a universal approximator due to its ability to approximate any nonlinear relationship between inputs and outputs to any degree of accuracy (Hornik et al., 1989; Smith, 1996; Zhang et al., 1998; Samarasinghe, 2007). It does not require prior knowledge of the process or a prior model structure (Zhang et al., 1998; Kim and Valdes, 2003; Nayak et al., 2004). The performance of ANN is not affected by the seasonality of time series; this is incorporated (Sharda and Patil, 1992). It performs better under less ideal conditions, such as when outliers, multicollinearity, and model misspecification are present (Zhang et al., 1998).

However, ANN acts as a black-box inference machine as there is no explicit form to explain and analyze the relationship between the inputs and outputs (Zhang et al., 1998; Marce et al., 2004). An ANN model is prone to overfitting because of the presence of a large number of parameter sets or to underfitting because its optimization algorithm may converge to a local minimum instead of the global minimum (Wilpen et al., 1994). It has no structured methods to identify the best structure (Bacanli et al., 2008). In cases where there is no underlying systematic pattern in the data, ANN does not perform any better than simpler models such as LR. The power of ANN increases only in cases where there are a large number of relevant explanatory variables with a larger sample size (Wilpen et al., 1994). If the underlying mechanism is linear and

without much irregularity, ANN performs worse than linear statistical methods (Zhang et al., 1998).

Adaptive Neuro-Fuzzy Inference System

Another artificial intelligence-based technique that has potential to make forecasts is an approach based on fuzzy logic propounded by Zadeh (1965). Fuzzy logic considers the truth of any statement as a matter of degree. In fuzzy logic, the crisp logic of yes-no is generalized. In contrast to classical binary logic, which considers the truth as yes or no, that is, 1 or 0, fuzzy logic replies to a yes-no question with a not-quite-yes-or-no answer, thus also permitting intermediate values between 0 and 1. When asked whether X is a member of set A, the answer of fuzzy logic might be yes, no, or any in-between value. Therefore, X can have partial membership in A. The degree of membership of any value of an input variable is defined by a curve, called a membership function (MF), which maps each element in the input space to a membership value between 0 and 1. The shape of a MF is defined taking into consideration simplicity, convenience, speed, and efficiency. The system of formulating the mapping from a given input to an output using fuzzy logic is called a fuzzy inference system (FIS). A FIS involves two types of MF, namely an input MF and an output MF, various logical operations, and several if-then rules. Whereas logical operations are used to combine two or more fuzzy sets, if-then rules are used to formulate the conditional statements that comprise fuzzy logic (for details, see MathWorks, 2009).

Figure 4-2 illustrates the structure of a simple FIS with PC1 as an input and ÄRID as an output. For simplicity, only one input variable is used in Figure 4-2.

The fuzzy logic approach has the ability to draw conclusions and generate responses based on vague, ambiguous, incomplete, and imprecise information (Lee et al., 2007). A FIS establishes a practical framework to include the qualitative aspects of human experience and reasoning processes in modeling (Jang, 1993; Borri et al., 1998). It is a promising tool for modeling complex, non-linear systems (Marce et al., 2004; Lee et al., 2007; Bacanli et al., 2008; Altun et al., 2009). FIS is a model-free approach that is easy to implement (Marce et al., 2004).

However, fuzzy logic does not have a systematic procedure to define MF parameters and fuzzy rules (Nayak, et al., 2004; Bacanli et al., 2008). The MFs and their parameters are determined arbitrarily from looking at the data or by trial and error (MathWorks, 2009).

To eliminate the requirement for manual optimization of MFs and their parameters in a conventional FIS as mentioned in previous paragraphs, Jang (1993) developed the Adaptive Neuro-Fuzzy Inference System (ANFIS) by merging the fuzzy logic approach with an artificial neural network that has automatic learning ability. The combination of fuzzy logic and a neural network enables the system to learn and improve, based on the data. Like a neural network, an ANFIS learns information about the data and automatically selects MFs and tunes their parameters. As in an MLP network, an ANFIS has multiple layers with signals passing from inputs to outputs via the neurons of several intermediate layers through connections. Once an output has been estimated, it is compared with the target value and the error is calculated. The error is then backpropagated through the network to adjust the parameters of an input MF using the backpropagation (gradient descent) learning method. The parameters of an output MF,

however, are updated using the least squares estimate (LSE) during the forward pass of the network. Figure 4-3 illustrates the structure of an ANFIS with PC1 as an input and $\widehat{A}RID$ as an output.

Using PC1 as an input variable, the ANFIS computes the final output, $\widehat{A}RID$, in the following order:

Fuzzification: The crisp values of the input variable are transformed into fuzzy membership values using a generalized bell input MF. This is done in the node of layer 1 (Figure 4-3).

$$f_i = \frac{1}{1+\left|\frac{PC1-c_i}{a_i}\right|^{2b_i}} \tag{4-16}$$

The subscript i in Equation 4-16 stands for the i-th rule; f is the input MF; PC1 is the input; and **a**, **b**, and **c** are the antecedent parameters, whose values are updated in every backward pass of the ANFIS using the backpropagation algorithm (which uses the gradient descent approach of error minimization).

Integration: In this step, usually the firing strength, also called the weight, is computed for rule i by integrating the input MFs of various input variables using the fuzzy intersection/conjunction, fuzzy union/disjunction, or fuzzy complement operators (for details, see MathWorks, 2009). This operation is done inside the nodes of the second layer. However, for the single input variable used in this example, no integration of the various MFs in each rule is needed. Therefore, the same values of MF are used also for the weight:

$$w_i = f_i \tag{4-17}$$

where w_i is the firing strength of rule i.

Normalization: For each node in the third layer, the normalized firing strength is calculated as the ratio of the i-th rule's weight to the sum of the weights of all n rules:

$$w_i^* = \frac{w_i}{\sum_{i=1}^{n} w_i} \tag{4-18}$$

where w^*_i is the normalized firing strength of rule i and n is the number of rules.

Defuzzification: The fuzzy input membership values are transformed back to crisp output values using a linear output MF.

$$g_i = \alpha_i PC1 + \beta_i \tag{4-19}$$

The term **g** in equation 4-19 is the output MF; PC1 is the input; and **α** and **β** are the consequent parameters whose values are updated during every forward pass of the ANFIS algorithm using the LSE. Then, using the output MF and the normalized firing strength, the contribution of each rule towards the model output is computed in each node of the fourth layer:

$$O_i = w_i^* g_i \tag{4-20}$$

where O_i is the output contributed by the i-th rule towards the overall ANFIS output.

Summation: Finally, the overall output of the model is computed as the summation of outputs contributed by all rules. This operation is performed in the fifth layer, which has a single node.

$$\widehat{A}RID = \sum_{i=1}^{n} O_i \tag{4-21}$$

$\hat{A}RID$ is the overall output of the ANFIS. This step concludes the forward processing in the ANFIS network. The network is trained to learn by repeated exposure to the training dataset until the correct output is produced.

An ANFIS removes the basic problem of conventional FIS, i.e., defining MF parameters and fuzzy if-then rules (Lee et al., 1997). An ANFIS also has an advantage over ANN in that the latter needs to perform a trial and error procedure to develop the best network structure, whereas the former does not (Bacanli et al., 2008). Therefore, an ANFIS combines the benefits of both FIS and ANN in a single framework (Nayak, et al., 2004). An ANFIS has the potential to capture the essential components of the underlying dynamics of complex systems with unclear input-output relationships (Nayak et al., 2004).

However, ANFIS is not free from limitations. The performance of ANFIS depends on the completeness of the database because sufficient samples are needed to have significant results (Costa Branco and Dente, 2001). The scarcity of data may lead to information holes and ultimately nonsense values (Marce et al., 2004). Moreover, if the training and checking datasets are not sufficiently distinct, ANFIS renders trivial results (Sahoo et al., 2005).

Overview of the ENSO

In the tropical Pacific Ocean, normally, the persistent easterly trade winds blow westward from the eastern Pacific, a region of higher pressure, toward Indonesia, a region of lower pressure. The trade winds create upwelling that brings cold water to the surface in the eastern Pacific. As this water moves westward, it is heated by sunlight and the atmosphere. Consequently, surface water along the equator is usually cool in

the eastern Pacific and warm in the western Pacific. In addition, the dragging of surface water by the trade winds raises the sea level and produces a thick layer of warm water in the western Pacific.

Every two to seven years, air pressure rises over the western Pacific and falls over the eastern Pacific (Yin, 1994). This change in pressure weakens the easterly trade winds, which are then replaced by westerly winds which strengthen the countercurrent. Surface water warms over the tropical Pacific and heads eastward towards South America, causing a remarkable increase in the equatorial sea surface temperatures over the eastern Pacific. Because the warming in the eastern Pacific peaks around the Christmas season, this phenomenon is called El Niño, referring to the Christ child. Toward the end of the warming period, atmospheric pressure over the eastern Pacific reverses and begins to rise; whereas, over the western Pacific, it falls (Maunder, 1992). This seesaw pattern of reversing surface air pressure at opposite ends of the Pacific Ocean is called the Southern Oscillation (Philander, 1990; Glantz et al., 1991). Because the pressure reversals and ocean warming are more or less simultaneous, scientists term this phenomenon as the El Niño-Southern Oscillation.

Following an El Niño event, the trade winds usually return to normal. However, if the trades are exceptionally strong, unusually cold surface water moves over the eastern Pacific, and the warm water is confined mainly to the western Pacific. This cold-water episode, which is the opposite of El Niño conditions, is given the feminine name La Niña in contrast to the masculine El Niño. Because El Niño and La Niña are two faces of the same large-scale ocean-atmosphere interaction, the ENSO phenomenon comprises of three phases: El Niño, La Niña, and neutral conditions.

The impact of ENSO varies across regions and seasons. Generally, the impacts are stronger in tropical regions than in the mid-latitudes (NDMC, 2006). In southeastern USA, fall, winter, and spring are generally wetter than usual during El Niño events (Ropelewski and Halpert, 1986; Green et al., 1997; Smith et al., 1998; Brolley et al., 2007; Kelly, 2007). La Niña events, in contrast, typically involve wetter summers and falls, and drier winters and springs (Gershunov and Barnett, 1986a; Green et al., 1997; Smith et al., 1998). These ENSO phenomena indicate that drought, a drier condition than normal, might be forecast for this region for any month of the year if an ENSO episode could be forecast for that month. In fact, the use of sophisticated predictive numerical models and the availability of expanded datasets collected through an array of moored buoys in the equatorial Pacific Ocean has allowed for skillful forecasts of ENSO episodes for this region up to a year in advance (Steinemann, 2006; Brolley et al., 2007; NDMC, 2009).

An ENSO episode is generally characterized using an indicator. Some indicators, such as the Japanese Meteorological Association Index (JMAI), are based on sea surface temperature. Others, such as the Southern Oscillation Index (SOI), are based on atmospheric pressure. In comparison to other indicators, the JMAI is more widely-used because it is more sensitive to both El Niño and La Niña conditions (Trenberth, 1997; Hanley et al., 2003; P. G. F. Gerard-Marchant, personal communication). The Southeast Climate Consortium (SECC) also uses JMAI for its own ENSO outlook (Trenberth, 1997; Hanley et al., 2003). The JMAI is a 5-month running mean of spatially averaged sea surface temperature anomalies over the tropical pacific region (4°S-4°N, 150°W-90°W). In its standard application, an El Niño (or La Niña) episode is defined

when the JMAI is greater than or equal to 0.5°C (or less than or equal to -0.5°C) for six consecutive months, including October through December (Hanley et al., 2003). The episode then lasts from October through the following September. The episode for all other values of JMAI is termed Neutral.

Materials and Methods

Study Area and Data

Five locations in southeastern USA, namely Miami, Bartow, and Live Oak , FL, and Plains and Blairsville, GA, were selected taking into consideration the geographical representation (latitude) and the availability of daily historical weather data for more than 50 years that also included the weather variables: dewpoint temperature and windspeed (Figure 2-2).

For locations in Florida, daily historical weather data spanning 56 years (1951 through 2006) were obtained from the website of the Florida Climate Center (ftp://coaps.fsu.edu/pub/griffin/SECC/AgroClimate/raw/ascii/). For Georgia, the weather data for the same years were obtained from the Georgia Automated Environmental Monitoring Network (G. Hoogenboom, personal communication, 9 March 2009). Similarly, values of various climate indices for the same years were collected from their respective websites (Table 4-2).

Basic Computations

Using the historical weather data, daily values of ARID were computed and later converted to monthly average values. The computations were carried out monthly, because the monthly time scale is more appropriate for monitoring drought from a farm management viewpoint and the sensitivity of crop growth stages to water deficit than

are daily or weekly time steps (Panu and Sharma, 2002). Moreover, most of the climate indices used in the study recorded only monthly values.

Of the five methods used to forecast ARID, namely ANFIS, ANN, LR, ARMA, and ENSO, only the first three methods used climate indices as regressors. Out of the six climate indices selected, only four were used in a set of predictor variables: AMO or NAO; JMA or Niño 3.4; PDO; and PNA. From NAO and AMO, generally the former was chosen. But in the cases where ARID was not significantly correlated with NAO, but was correlated with AMO, the latter was used. Similarly, from Niño 3.4 and JMA, the former was chosen. But JMA was used when Niño 3.4 was not significantly correlated. Such selections were carried out because AMO and NAO are similar, as are JMA and Niño 3.4 (J. O'Brien, personal communication, 7 May 2009). Because the predictor variables might be correlated to each other, whereas the use of regression techniques requires them to be independent (Cordery and McCall, 2000), the principal component analysis (PCA) was carried out on the four predictor variables, and the first principal component (PC1) was used as the predictor of ARID. Before applying PCA, each input variable was normalized individually by dividing the difference of each of its observations from its mean by its standard deviation (Weigend et al., 1992) to avoid computational problems, meet algorithm requirements, and facilitate network learning (Zhang et al., 1998).

To assess the predictability of ARID using different lead-times, ARID was forecast with 1-month to 3-month lead-times using the past values of PC1. Therefore, to forecast ARID for month t, denoted as $ARID_t$, three series of PC1 were created: $PC1_{t-1}$, $PC1_{t-2}$, and $PC1_{t-3}$. Each of the three series was used individually as the predictor of $ARID_t$. Assuming that the degree of linkage between large-scale teleconnection patterns and

the weather in this region varies across months, the dataset of each predictor and predictand, namely $ARID_t$, $PC1_{t-1}$, $PC1_{t-2}$, and $PC1_{t-3}$, were separated by months, thus creating 12 subsets, one for each month. These monthly datasets were then used in ANFIS, ANN, and LR models to forecast ARID with 1-, 2-, and 3-month lead-times. Finally, the performance of each of these models was evaluated using the technique as explained in the Specific Computations section below.

Specific Computations

Autoregressive moving average models

For each location, 12 time series were created, one for each month, from the original time series of ARID that were constructed in the Basic Computations section above. For instance, the time series for January comprised the ARID values from January 1951 through December 2005. Similarly, the February time series started from February 1951 through January 2006 and so on. These monthly time series were created assuming that meteorological processes vary across months.

To compute the root mean squared error of prediction (RMSEP) for each monthly model, a range of model-estimated values were needed. Probably the best approach is to use the leave-one-out technique of cross-validation. However, this technique would discontinue the time series. Therefore, from each monthly time series created above, 16 sub-series were created by truncating the original time series by 1 to 15 years. For instance, the first time series of January ended in December 2005, the second in December 2004, and the sixteenth in December 1990. Similarly, 16 time series were created for each month, and 16 ARMA models were developed for each month, accordingly. The number 16 was chosen, that is, truncating was stopped after the 16th

truncation, considering 40 years as the minimum length of the time series required to reflect any long-term periodicity in the data.

To observe the relationships between present and past values of ARID, autocorrelation functions (ACF) and partial autocorrelation functions (PACF) were calculated for each of the 16 time series for each month in each location. In each series, the values of ACF and PACF exhibited periodicities corresponding to the positive correlation between values separated by 12 months, indicating a need for seasonal differencing (Figure 4-4). Thus, each time series was seasonally differenced as:

$$\nabla_s^D \mathrm{ARID}_t = (1-B^s)^D \mathrm{ARID}_t = (1-B^{12})\mathrm{ARID}_t = \mathrm{ARID}_t - \mathrm{ARID}_{t-12} \quad (4\text{-}22)$$

where $\nabla_s^D \mathrm{ARID}_t$ is the ARID series seasonally differenced by order **D**, which is 1.

To study the behavior of the seasonally differenced ARID series, $\nabla_{12}^1 \mathrm{ARID}_t$, values for the ACFs and PACFs of the differenced series were calculated and analyzed in terms of finding any correlation among the ARID values (Figure 4-5). Because the values of ACF and PACF of the first-differenced series did not exhibit any periodicities, no further differencing was needed. Therefore, the value of **D** was set to 1. However, the characteristics of the ACF and PACF of the seasonally differenced series showed a strong peak at a lag, **h**, of 12 months in the autocorrelation function. Smaller peaks appeared at 24 and 36 **h** in the ACF, combined with peaks at 12, 24, 36, and 48 **h** in the PACF (Figure 4-5). The cutting off of ACF after lag **h** = 12 months, that is, a seasonal lag of 1 year, and the tailing off of the PACF of the seasonally differenced series indicated a seasonal moving average (MA) of order **Q** = 1. Therefore, the orders of **P**, **D**, and **Q** for the seasonal part were 0, 1, and 1, respectively. The same orders were found for each series, month, and location.

For the non-seasonal component, values of ACF and PACF of the seasonally differenced series within seasonal lags, **h** = 1, 2, …11, did not show any periodicities. Therefore, no non-seasonal differencing was needed (**d** = 0). Because the non-seasonal behavior of ACF and PACF was the same in each series for each month and location, the order of **d** was also the same for all cases. For the order of **p** and **q**, inspection of the ACF and the PACF within seasonal lags suggested three possibilities in each series of each month and location (Figure 4-5):

- The tailing off of ACF and the cutting off of PACF after lag **h** = 1 indicated a non-seasonal AR of order **p** = 1.
- The cutting off of ACF after lag **h** = 2 and the tailing off of PACF indicated a non-seasonal MA of order **q** = 2.
- The tailing off of both ACF and PACF suggested, at first, **p** = 1 and **q** = 1.

These observations suggested three plausible SARIMA models for each month and location: ARIMA(1, 0, 0) x $(0, 1, 1)_{12}$, ARIMA(0, 0, 2) x $(0, 1, 1)_{12}$, and ARIMA(1, 0, 1) x $(0, 1, 1)_{12}$. From the three SARMIA models, ARIMA(1, 0, 0) x $(0, 1, 1)_{12}$, the most appropriate model, was selected for each month and location based on the lowest values of the Akaike's Information Criterion (AIC: Akaike, 1969) of the fitted models. The preferred SARIMA models and their fitted versions for various locations in the region are presented in Table 4-3. The diagnostic analysis of these models led to the conclusion that the models were adequate, that is, fit well to the data.

Using the 16 SARIMA models, one for each time series, for each month, ARID was forecast with 1-, 2-, and 3-month lead times. Finally, to evaluate the performance of the SARIMA models, values of the cross-validated RMSEP and ME were computed for each month and each lead-time forecasting using these 16 forecasts and the

corresponding 16 original values of ARID. The same procedure was followed for each location.

Linear regression models

To generate the forecast of ARID **m** months ahead with a validation dataset using Equation 4-23, the training dataset comprised of the current values of ARID, denoted as $ARID_t$, and the past values of PC1, denoted as $PC1_{t-m}$, were used in the regression to estimate the coefficients α and β for the validation dataset as:

$$ARID_t = \alpha PC1_{t-m} + \beta \quad (4\text{-}23)$$

where the subscripts t and m stand for the month and lead-time, respectively. The values of m ranged from 1 to 3 months.

The performance of a LR model was evaluated applying the leave-one-out technique of cross-validation, for which each monthly dataset was divided into two parts: one for training and the other for validation. That is, of the total 56 input-output combinations, 55 combinations were used as the training set for parameter estimation and one combination as the validation set for output estimation. Leaving one combination out and adding one combination in, this procedure was repeated 56 times such that each of the 56 input-output combinations was used as the validation set. Finally, the values of RMSEP and ME were computed using the 56 model-estimated and 56 original values of ARID. The same procedure was followed for each month, location, and lead-time forecast.

Artificial neural network models

In this study, a 3-layer MLP (Figure 4-1) was used as it can provide a general framework for representing the nonlinear functional mapping of a set of input and output

variables (Kim and Valdes, 2003). In the MLP architecture, a single hidden layer was used because one hidden layer is sufficient to approximate any complex nonlinear function to any desired accuracy (Cybenko, 1989; Hornik et al., 1989; Zhang et al., 1998; Maier and Dandy, 2000). The number of hidden nodes in the hidden layer was set to one because the networks whose hidden nodes are equal to the input nodes provide better forecasting results (De Groot and Wurtz, 1991; Chakraborty et al., 1992; Sharda and Patil, 1992; Tang and Fishwick, 1993; Zhang et al., 1998). For the hidden layer transfer function, a hyperbolic tangent function was used as it learns more quickly (Maier and Dandy, 1998), produces better forecasts (Klimasauskas, 1991) and is widely-used (Samarasinghe, 2007). For the output layer, a linear transfer function was used because it is appropriate for a forecasting problem that involves continuous target values (Rumelhart et al., 1995; Zhang et al., 1998; Sahoo, et al., 2005).

For training the network, backpropagation was used because it is the most popular learning algorithm (Zhang et al., 1998; ASCE, 2000a; Mishra and Desai, 2006; Altun et al., 2009). Backpropagation uses the gradient descent approach, which is an efficient method to find the bottom of the error surface more quickly during training (Samarasinghe, 2007). Also, it provides effective solutions as it allows a comprehensive search in the weight space. The learning rate, which indicates how far in the direction of the steepest descent the weights must be shifted per epoch, was set to a small value of 0.01, assuming that the input-output relationships were complex (Tang et al., 1991; Hagan et al., 1996; Principe et al., 1999; Mishra and Desai, 2006; Mishra et al., 2007; MathWorks, 2009). To improve the gradient descent method by allowing larger learning rates to result in faster convergence while minimizing the tendency to oscillate, an

additional parameter, the momentum, was used. The momentum indicates the relative importance of the past weight change on the new weight increment and forces the next weight change in the same direction as the previous one. In this study, the momentum was set to 0.9, assuming that the input-output data were complex (Tang et al., 1991; Mishra and Desai, 2006; Mishra et al., 2007; MathWorks, 2009). For training, the batch learning approach was used because it is the most preferred method and forces the search to move in the direction of the true gradient at each weight update (Maier and Dandy, 2000). It also avoids oscillations or instability (Samarasinghe, 2007).

Before feeding the data into the artificial neural network, each monthly dataset of 56 input-output combinations was divided into three subsets: 44 combinations for training, 11 combinations for testing, and 1 combination for validation (Granger, 1993). Leaving one combination out and adding one combination in, the validation set and the window of each training and testing dataset were moved a further 55 times to produce 56 estimates. The testing set was used to avoid overfitting the data through early stopping, whereas the validation set was used to evaluate the model. The performance of an ANN model was evaluated applying the leave-one-out technique of cross-validation as mentioned above. Finally, the values of RMSEP and ME were computed using the 56 model-estimated and 56 original values of ARID. The same procedure was followed for each month, location, and lead-time forecast.

Adaptive neuro-fuzzy inference system models

Of the three types of fuzzy inference system that are available for implementation, namely, Mamdani (Mamdani and Assilian, 1975), Sugeno (Takagi and Sugeno, 1985), and Tsukamoto (Tsukamoto, 1979), the Sugeno-type was used because it is computationally efficient, works well with linear and adaptive techniques, has a

guaranteed continuity of the output surface, and is well suited for modeling non-linear systems (Lee et al., 2007; MathWorks, 2009). Moreover, the Sugeno-type is specifically used in ANFIS to fit parameters (Jang, 1993). Of the various orders of the Sugeno-type, the first order was used because higher orders often add an unwanted level of complexity (Lee et al., 2007). For the input MF, the generalized bell function was used because it is smooth and nonzero at all points (Marja and Esa, 1997; Lee et al., 2007; MathWorks, 2009). For the output MF, a linear function was used. Of the several methods used for partitioning the input space into a number of fuzzy regions, grid partitioning was used because it generates a single-output fuzzy inference system and usually gives more precise modeling (Altun et al., 2009). For training ANFIS, the combination of LSE and backpropagation methods, called the hybrid optimization method, was used (Figure 4-3) because it speeds up the learning process substantially (Jang, 1993; Lee et al., 1997).

As for ANN, each monthly dataset was divided into three subsets – a training set (44/56), a testing set (11/56), and a validation set (1/56) – before feeding the data into the system. Similarly, 56 model estimations were made for each month in each location and each lead-time forecast. Accordingly, the values of RMSEP and ME were computed using the 56 model-estimated and 56 original values of ARID to evaluate the performance of an ANFIS model.

El Niño-Southern Oscillation approach

For the ENSO-based forecasting of ARID, the historical monthly time series of ARID was separated into three ENSO categories as characterized by the modified JMA index (D. Zierden, personal communication, 7 May 2009). In this study, a modified version of the JMA index was used because it gives more significant results than the

original index (P. G. F. Gerard-Marchant, personal communication, 7 May 2009). In the modified version, the El Niño (or La Niña) episode stops as soon as the temperature conditions are no longer met (D. Zierden, personal communication, 7 May 2009). In each ENSO category, the ARID values were further separated into 12 months and then averaged individually. The monthly average value of ARID in a particular month and ENSO phase was considered as the forecast for that month during that ENSO phase. For instance, the average value of ARID in the month of January during an El Niño phase was considered as the forecast value of ARID for this month during this ENSO phase.

Comparison

The predicting abilities of ANFIS, ANN, ARMA, ENSO, and LR models were evaluated in terms of the modeling efficiency (Nash and Sutcliffe, 1970). The ME, which assesses the predictive power of a model, is defined as:

$$ME = 1 - \left[\frac{\sum_{i=1}^{n} \left(ARID - \hat{A}RID \right)^2}{\sum_{i=1}^{n} \left(ARID - \bar{A}RID \right)^2} \right] \quad (4\text{-}24)$$

where ARID, $\hat{A}RID$, and $\bar{A}RID$ are the observed, estimated, and mean observed values of ARID, respectively, and n is the total number of observations.

An ME denotes the average distance between the observed and estimated values relative to the average distance between the observed and the mean observed values. A greater ME value indicates that the estimated values are closer to the observed values relative to the mean observed values; that is, the average relative distance is shorter. Values of ME can range from -∞ to 1. Whereas a negative value indicates that

the observed mean is a better predictor than the model, a positive value signifies that using model estimations is better than using the observed mean. An ME of 0 indicates that the model estimations are as accurate as the mean of the observed data, whereas an ME of 1 corresponds to a perfect match of the modeled values to the observed data. In fact, the closer the ME is to 1, the more accurate is the model. To evaluate the performance of the various models used in this study, ME was chosen because it is very commonly used in describing the predictive accuracy of models that simulate hydro-meteorological systems (Moriasi et al., 2007) and also because it is recommended by several researchers, including Sevat and Dezetter (1991), ASCE (1993), Legates and McCabe (1999), and Moriasi et al. (2007), for validating hydrologic and hydroclimatic models.

To assess the predictability of ARID for different lead-times, the forecasts of ARID with one-, two-, and three-month lead times were compared using RMSEP. These comparisons were carried out for each model, month, and location.

Probabilistic Forecasting of ARID

Basically, forecasting approaches are of two types: deterministic and probabilistic. A deterministic approach specifies a point estimate of the variable that is forecast, whereas a probabilistic approach specifies a probability distribution function of forecast variable. A deterministic forecast may create the illusion of certainty in a user's mind, which can easily lead the user to suboptimal action, whereas a probabilistic forecast is scientifically more honest as it allows the forecaster to admit the uncertainty and to express the degree of certitude, enables risk-based warning with explicitly-stated detection probabilities, apprises the user of the uncertainty and provide information necessary for making rational decisions, and offers potential for additional economic

benefits of forecasts (Krzysztofowicz, 2001). As long as the forecast is not given a 0% or 100% chance of occurring, a probabilistic forecast can never be wrong (Mason, 2002). Although deterministic forecasts are more useful for users, they are more difficult for forecasters to produce. Therefore, most forecasts are probabilistic.

Probabilistic forecasting needs the probability distribution function of the variable that is forecast. The distribution of prediction error, computed as observed minus predicted ARID, can be used to make the probabilistic forecasts of ARID. Using values of a forecast (F), a prediction error threshold (ε), $P(ARID < F+\varepsilon)$, and $P(ARID < F-\varepsilon)$, $P(F-\varepsilon < ARID < F+\varepsilon)$ can be computed as $P(ARID < F+\varepsilon)$ minus $P(ARID < F-\varepsilon)$. The term P(…) means the probability of (…).

Based on the values of ME as computed in the preceding section, three best performing models were selected for the probabilistic forecasting purpose. From the distribution of the 56 (or 16 in the case of SARIMA) values of prediction error, a cumulative probability distribution function was created, from which $P(-0.1 < \varepsilon < 0.1)$ was computed by subtracting $P(\varepsilon < -0.1)$ from $P(\varepsilon < 0.1)$. Such computations were made for each location, month, and model, thus producing a table of the probability of prediction error values which can be used to make the probabilistic forecasting of ARID. The prediction error threshold 0.1 was chosen to indicate a small error. In place of 0.1, any value could be used for indicating a small error. If, for instance, 0.05 were used, the width of the forecast would be narrower. If 0.15 were used, the forecast would have a wider range. The idea of using the error threshold was not to show the forecasts in absolute terms, but to demonstrate a way for making the probabilistic forecasts of ARID using the climate indices and forecasting models discussed in the Introduction section.

Results and Discussion

Comparing Models with ENSO

The modeling efficiency of each of the four models, namely ANFIS, ANN, LR, and SARIMA, varied depending on the month and location (Figure 4-6). The variation in efficiencies of the models gave an indication that for a specific month at a specific location, one specific model may be more appropriate than the others. In general, the efficiencies of the climate index (CI)-based models (namely, ANFIS, ANN, and LR) and the ENSO approach were large during the winter and small during the summer in most of the locations. Furthermore, these models indicated an acceptable level of performance during the winter as their ME values were generally positive. During the summer, they indicated an unacceptable level of performance as their ME values were mostly negative. This seasonal difference was possibly due to the stronger signals of ENOS and large-scale teleconnections during winter than in the summer. The time series-based model, SARIMA, however, did not show this pattern. Its efficiency largely depended on the periodicity of the time series data. For instance, SARIMA would be highly efficient for a month that received a large amount of precipitation in each year of the time series.

Generally, the efficiency of each model, including the ENSO approach, was larger in southern locations, such as Miami, Bartow, and Live Oak than in the northern ones, such as Plains and Blairsville (Figure 4-6). The modeling efficiencies decreased in the direction of south to north. Miami, south Florida, had the largest efficiencies, whereas Blairsville, north Georgia, showed the lowest efficiencies. This differential efficiency probably arose because of stronger ENSO and teleconnection signals and larger precipitation periodicities in the south than in the north.

In terms of modeling efficiency, ANN generally ranked number 1 (the largest efficiency) for southern locations such as Miami, Bartow, and Live Oak (Table 4-4). For northern locations, SARIMA ranked number 1 for Plains and ENSO ranked number 1 for Blairsville. Of the five approaches compared, the performances of LR and ANFIS were generally poor for all locations. The poor efficiency of LR was likely the result of its being a linear model, such that it could not represent the nonlinearity that possibly exists between ARID and climate indices. For ANFIS, even though it is supposed to account for nonlinearity in a system, it performed poorly. This was most probably due to limited data. The total number of parameters associated with an ANFIS model was 15, whereas the number of samples in the training data set was only 44. Because the training data points were fewer than four times the number of model parameters (a threshold value (Marce et al., 2004)), the small training dataset probably caused overfitting so that the generality was lost. Because of limited data, ANFIS probably could not generate proper fuzzy rules and derive all the complex information of the system properly. The use of ANN, on the other hand, does not necessarily require a large sample size (Zhang et al., 1998). The ANN models perform well even with a sample size of less than 50 (Kang, 1991). Being a universal approximator due to its ability to approximate any nonlinear relationship between inputs and outputs to any degree of accuracy (Hornik et al., 1989; Smith, 1996; Zhang et al., 1998; Samarasinghe, 2007), ANN performed better than the other CI-based models, namely ANFIS and LR. The better performance of ANN for southern locations than for northern locations indicated that the relationships between climate indices and ARID are stronger in the southern locations. Probably because of weak teleconnections between the

precipitation pattern in northern locations and the climate indices, ANN performed poorly for Plains and Blairsville. Generally, both ENSO- and teleconnection-based methods performed poorly for the northern part of the region. The performance of teleconnection-based methods was even poorer than that of the ENSO approach, indicating that the signals of large-scale teleconnection are even weaker than those of the ENSO in northern locations. Also in southern locations, the ENSO approach performed better than the other models for some months: February and April for Miami; January, April, and October for Bartow; and July through October for Live Oak. These results indicated that the ENSO signal is stronger in these months and locations than are the teleconnections represented by the climate indices. For the other months, signals represented by the climate indices seemed stronger than that of the ENSO. The efficiency of SARIMA models also varied depending upon month and location. Generally, the performance of SARIMA was smaller than those of ANN and ENSO for most of the months in southern locations. The poor performance probably resulted because a SARIMA model assumes that time series are generated from linear processes, whereas the drought time series are often nonlinear and seasonal (Granger and Terasvirta, 1993; Zhang et al., 1998). Generally, SARIMA performs better when periodicity exists in the data, as it effectively takes into account the serial correlation among observations (Mishra et al., 2007). Probably because of the lack of seasonality in the data in these months, the SARIMA models performed poorly. In the northern region, however, especially at Plains, SARIMA performed better than the other models. Even in southern locations, the performance was better than those of the other approaches for some months, namely, January, August, September, and December for

Miami; June and November for Bartow; and April and June for Live Oak. The better performance in these months was probably due to the existence of seasonality and nonstationarity in the time series data. For instance, December and January of Miami are dry months as they usually receive less precipitation, whereas August and September are wet months, the period of hurricanes. Therefore, the SARIMA models performed better when periodicity existed in the time series.

The results indicated that ARID may be forecast more accurately for several months in most locations using one or more models than using the current ENSO approach (Table 4-5). For instance, whereas the forecasting of ARID could be improved for January in Miami using the time series-based SARIMA models, the index could be better forecast for March and May through December using the CI-based ANN models. Similarly, forecasting for December in Miami could be improved with any of the following models: ANFIS, ANN, LR, or SARIMA. Forecasting of ARID could not be improved with the CI-based models for the other months, probably because of the weaker signals of the large-scale teleconnections than that of the ENSO. Similarly, the SARIMA models could not generate more accurate forecasts than that of ENSO due to the lack of strong periodicity in the time series data in these months.

Comparing Models in terms of Lead-Time Forecasting

The error of prediction (RMSEP) associated with 1-, 2-, and 3-month ahead forecasting for each month and location varied depending on the models used. For the CI-based models, namely ANFIS, ANN, and LR, the occurrence of $RMSEP_{t+1} < RMSEP_{t+2} < RMSEP_{t+3}$ was less than 25% (Figure 4-7). The terms $RMSEP_{t+1}$, $RMSEP_{t+2}$, and $RMSEP_{t+3}$ denote the RMSEP associated with 1-, 2-, and 3-month ahead forecasting, respectively. The percentage of occurrence shows the number of

times $RMSEP_{t+1} < RMSEP_{t+2} < RMSEP_{t+3}$ occurred out of the total 60 cases (5 locations x 12 months). For SARIMA, the occurrence of $RMSEP_{t+1} < RMSEP_{t+2} < RMSEP_{t+3}$ was 50% (i.e., 30 out of 60).

Although $RMSEP_{t+1} < RMSEP_{t+2} < RMSEP_{t+3}$ did not occur for 100% of the cases with each model, the results gave an indication that short-term forecasting (for one month in advance, for instance) can generally make more accurate estimations than can long-term forecasting (for two or three months in advance, for instance). The SARIMA models produced less accurate forecasts than the CI-based models for the months that were more distant because of weaker auto-correlations in the time series data with increasing lag times. The occurrence of $RMSEP_{t+1} < RMSEP_{t+2} < RMSEP_{t+3}$ for fewer cases with the CI-based models than with the SARIMA models gave an impression that, although ARID is slightly more correlated with climate indices with shorter lags, the correlations between ARID and the climate indices with longer lags are also possible.

Probabilistic Forecasts of ARID

The distribution of error in terms of box and whisker plots for each of the 12 months, 5 locations, and the three methods, namely ANN, ENSO, and SARIMA, whose performances were better than those of the others as shown by the ME values presented in the Comparing Models with the ENSO section above, are presented in Figure 4-8. In line with the modeling efficiency results, the performance of each of the methods in terms of small errors varied depending on months and locations. For instance, the interquartile range of errors for Miami in January was smallest with the ENSO approach, whereas that for Miami in March was smallest with ANN. Similarly, ANN produced the smallest errors for Miami in June and July, whereas SARIMA

produced the smallest errors in August and September. Why different methods behaved differently in different months or locations has already been discussed in the Comparing Models with the ENSO section above.

The probability of prediction error falling inside -0.1 and 0.1, denoted as $P(-0.1 < \varepsilon < 0.1)$, computed for each location, month, and method is presented in Table 4-6. Using this table, the probabilistic forecasting of ARID could be carried out for a given each location, month, and method. For instance, $P(-0.1 < \varepsilon < 0.1)$ for Miami in January as estimated by the ENSO approach was 0.39 (Table 4-6). If the forecast (F) made this approach for Miami in January were 0.26, $P(F-0.1 < ARID < F+0.1)$, that is, $P(0.16 < ARID < 0.36)$ for this month and location would be 0.39. That is, there was 39% chance that the forecast of ARID would fall inside 0.26±0.1. Similarly, if the forecast made by an ANN model for Bartow in July were 0.27, $P(F-0.1 < ARID < F+0.1)$, that is, $P(0.17 < ARID < 0.37)$ for this month and location would be 0.82 (Table 4-6).

The large probability values in Table 4-6 may be due to extremely wet conditions, extremely dry conditions, or good teleconnections between the large-scale oceno-atmospheric signals (climate indices) and ARID in this region. Under extremely wet conditions, almost all values of ARID are 0, such as those in January, February, and December in Plains and Blairsville (Figure 4-9d and 4-9e). Under extremely dry conditions, values of ARID concentrate at 1. So, the width of prediction error is almost zero, and $P(-0.1 < \varepsilon < 0.1)$ is high in both severely wet and dry conditions. When the distribution of ARID is not close to 0 or 1, such as those in July, August, and September in Miami (Figure 4-9a); July and August in Bartow (Figure 4-9b); and February in Live Oak (Figure 4-9c), the large value of $P(-0.1 < \varepsilon < 0.1)$ indicates a strong teleconnection

between climate indices and ARID and the good performance of the forecasting model such as ANN.

Figure 4-9 also illustrates 75% probability that the mean value of ARID as forecast by an ANN model would fall inside a certain range. For January in Miami, for instance, there was 75% chance that the forecast of ARID would fall inside 0.30 and 0.63 (Figure 4-9). The 75% probability was computed as the difference between the 87.5th percentile and the 12.5th percentile (Figure 4-9).

Conclusions

The performance of the ANFIS, ANN, ENSO, LR, and SARIMA models varied depending upon month and location. Generally, the CI-based models and the ENSO approach performed better during the winter than the summer due to stronger signals of the ENSO and large-scale teleconnections in the winter than in the summer. The efficiency of SARIMA models depended largely on the periodicity of precipitation in the time series data. In general, the efficiency of each model was larger in southern locations than in northern locations due to stronger ENSO and teleconnection signals and more distinct precipitation periodicities in the south than in the north. Whereas ANN generally outperformed the other models for the southern region, the ENSO approach performed better than the other models for the north. In general, LR and ANFIS performed poorly as the former could not represent the nonlinearity between ARID and the climate indices, and the latter lost generality mainly because of limited data. The relatively better performance of ANN was due to its ability to approximate the nonlinear relationship between the climate indices and ARID. The CI-based models performed worse than the ENSO approach for northern locations because of the weaker signals of the teleconnections relative to those of the ENSO in this region.

The results gave an indication that the current level of forecasting, which is essentially ENSO-based, may be improved for several months in most locations in the southeast USA using one or more of the CI-based and SARIMA models. The results further indicated that the climate indices used in this study have good potential for use in forecasting drought for both winter and summer months, especially in the southern part of the region.

Table 4-1. Definitions of various abbreviations that have been mentioned frequently in Chapter 4

Abbreviation	Definition
ACF	Auto-correlation Function
AMO	Atlantic Multi-decadal Oscillation
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
AR	Autoregression, Autoregressive
ARID	Agricultural Reference Index for Drought
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ENSO	El Niño-Southern Oscillation
FIS	Fuzzy Inference System
JMA	Japanese Meteorological Agency
LR	Linear Regression
LSE	Least Squares Estimate
MA	Moving Average
ME	Modeling Efficiency
MF	Membership Function
MLP	Multi-Layer Perceptron
NAO	North Atlantic Oscillation
PACF	Partial Auto-correlation Function
PC1	First Principal Component
PDO	Pacific Decadal Oscillation
PNA	Pacific North American pattern
RMSE	Root Mean Squared Error
SARIMA	Seasonal Autoregressive Integrated Moving Average

Table 4-2. Various URL sources from where the data of the various climate indices were obtained

Climate index	Website	Source
AMO	http://www.esrl.noaa.gov/psd/data/correlation/amon.us.data	Earth System Research Laboratory, NOAA
JMA	ftp://www.coaps.fsu.edu/pub/JMA_SST_Index/jmasst1949-today.filter-5	COAPS, Florida State University
NAO	ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/tele_index.nh	Climate Prediction Center, NOAA
Niño 3.4	http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices	Climate Prediction Center, NOAA
PDO	http://jisao.washington.edu/pdo/PDO.latest	JISAO, University of Washington
PNA	ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/tele_index.nh	Climate Prediction Center, NOAA

Table 4-3. The preferred SARIMA models and their fitted versions for various locations in the southeast USA

Location	SARIMA Model	Equivalent Fitted Model
Miami	ARIMA $(1, 0, 0) \times (0, 1, 1)_{12}$	$ARID_t = 0.31ARID_{t-1} + ARID_{t-12} - 0.31ARID_{t-13} + \varepsilon_t - 1.00\,\varepsilon_{t-12}$
Bartow	ARIMA $(1, 0, 0) \times (0, 1, 1)_{12}$	$ARID_t = 0.30ARID_{t-1} + ARID_{t-12} - 0.30ARID_{t-13} + \varepsilon_t - 1.00\,\varepsilon_{t-12}$
Live Oak	ARIMA $(1, 0, 0) \times (0, 1, 1)_{12}$	$ARID_t = 0.20ARID_{t-1} + ARID_{t-12} - 0.20ARID_{t-13} + \varepsilon_t - 0.99\,\varepsilon_{t-12}$
Plains	ARIMA $(1, 0, 0) \times (0, 1, 1)_{12}$	$ARID_t = 0.19ARID_{t-1} + ARID_{t-12} - 0.19ARID_{t-13} + \varepsilon_t - 1.00\,\varepsilon_{t-12}$
Blairsville	ARIMA $(1, 0, 0) \times (0, 1, 1)_{12}$	$ARID_t = 0.18ARID_{t-1} + ARID_{t-12} - 0.18ARID_{t-13} + \varepsilon_t - 0.95\,\varepsilon_{t-12}$

Table 4-4. The ranks of various models for different months and locations based on modeling efficiency. Whereas rank 1 indicates the largest efficiency, rank 5 indicates the smallest efficiency.

Month	Miami					Bartow					Live Oak					Plains					Blairsville				
	E[a]	A	N	L	S	E	A	N	L	S	E	A	N	L	S	E	A	N	L	S	E	A	N	L	S
Jan	2	4	3	5	1	1	4	3	2	5	2	4	1	3	5	2	3	4	5	1	1	3	5	4	2
Feb	1	4	3	5	2	2	4	1	3	5	5	2	1	3	4	2	5	3	4	1	1	5	4	2	3
Mar	4	2	3	1	5	2	5	1	3	4	4	2	3	1	5	1	3	2	4	5	1	2	3	4	5
Apr	1	3	2	4	5	1	3	4	5	2	3	5	2	4	1	2	5	3	4	1	1	5	2	4	3
May	4	1	2	3	5	2	3	1	4	5	2	4	1	3	5	1	5	3	4	2	1	5	3	4	2
Jun	3	5	1	2	4	3	5	2	4	1	2	5	3	4	1	1	5	3	4	2	1	5	3	4	2
Jul	2	4	1	3	5	2	4	1	3	5	1	3	4	5	2	2	5	3	4	1	1	5	2	4	3
Aug	2	5	3	4	1	2	4	1	3	5	1	5	3	4	2	1	4	2	5	3	1	5	3	4	2
Sep	4	2	3	5	1	2	4	1	3	5	1	5	3	4	2	2	4	3	5	1	1	5	3	4	2
Oct	2	5	1	4	3	1	4	2	3	5	1	4	2	3	5	4	2	5	3	1	1	3	2	4	5
Nov	2	3	1	4	5	3	4	2	5	1	4	3	1	2	5	4	5	2	3	1	2	4	3	5	1
Dec	5	3	2	4	1	3	4	1	5	2	5	4	1	3	2	3	4	1	2	5	1	3	4	5	2
Overall	2	4	1	5	3	2	5	1	3	4	2	5	1	4	3	2	5	3	4	1	1	5	3	4	2

[a] E = ENSO, A = ANFIS, N = ANN, L = LR, S = SARIMA

Table 4-5. Months and locations for which ARID was forecast better using the given one or more models than using the ENSO approach.

Month	Miami	Bartow	Live Oak	Plains	Blairsville
Jan	SARIMA	-	ANN	SARIMA	-
Feb	-	ANN	ANN	SARIMA	-
Mar	ANFIS / ANN / LR	ANN	ANFIS / ANN / LR	-	-
Apr	-	-	ANN / SARIMA	SARIMA	-
May	ANFIS / ANN / LR	ANN	ANN	-	-
Jun	ANN	ANN / SARIMA	SARIMA	-	-
Jul	ANN	ANN	-	SARIMA	-
Aug	ANN / SARIMA	ANN	-	-	-
Sep	ANN / SARIMA	ANN	-	SARIMA	-
Oct	ANN	-	-	ANFIS / LR/ SARIMA	-
Nov	ANN	ANN / SARIMA	ANFIS / ANN / LR	ANN / LR / SARIMA	SARIMA
Dec	ANFIS / ANN / LR / SARIMA	ANN / SARIMA	ANFIS / ANN / LR / SARIMA	ANN / LR	-

Table 4-6. Values of the probability of prediction error falling inside -0.1 and 0.1 as estimated by ANN, ENSO, and SARIMA models for various months and locations in the southeast USA

	Miami			Bartow			Live Oak			Plains			Blairsville		
Month	E[a]	A	S	E	A	S	E	A	S	E	A	S	E	A	S
Jan	0.39	0.29	0.16	0.23	0.14	0.36	0.44	0.48	0.56	0.94	0.91	0.96	0.99	0.98	0.96
Feb	0.32	0.34	0.36	0.30	0.30	0.16	0.45	0.64	0.44	0.89	0.88	0.92	0.99	0.98	0.96
Mar	0.32	0.38	0.36	0.34	0.27	0.20	0.39	0.43	0.24	0.53	0.55	0.52	0.94	0.91	0.84
Apr	0.49	0.48	0.60	0.33	0.36	0.28	0.31	0.34	0.32	0.37	0.29	0.40	0.56	0.54	0.52
May	0.45	0.46	0.28	0.35	0.39	0.40	0.35	0.38	0.44	0.37	0.36	0.32	0.49	0.45	0.40
Jun	0.52	0.50	0.32	0.42	0.50	0.36	0.46	0.41	0.40	0.39	0.38	0.28	0.46	0.41	0.40
Jul	0.67	0.75	0.60	0.72	0.82	0.60	0.53	0.54	0.64	0.51	0.38	0.52	0.38	0.34	0.28
Aug	0.59	0.63	0.68	0.67	0.64	0.56	0.49	0.52	0.52	0.42	0.43	0.40	0.44	0.45	0.36
Sep	0.80	0.77	0.84	0.46	0.46	0.32	0.46	0.36	0.40	0.34	0.36	0.20	0.35	0.27	0.32
Oct	0.38	0.36	0.28	0.34	0.32	0.12	0.33	0.29	0.32	0.30	0.30	0.32	0.18	0.38	0.28
Nov	0.22	0.23	0.16	0.24	0.30	0.24	0.24	0.27	0.20	0.28	0.38	0.32	0.51	0.68	0.76
Dec	0.32	0.32	0.28	0.23	0.23	0.32	0.20	0.20	0.04	0.73	0.73	0.84	0.97	0.98	0.96

[a] E = ENSO, A = ANN, S = SARIMA

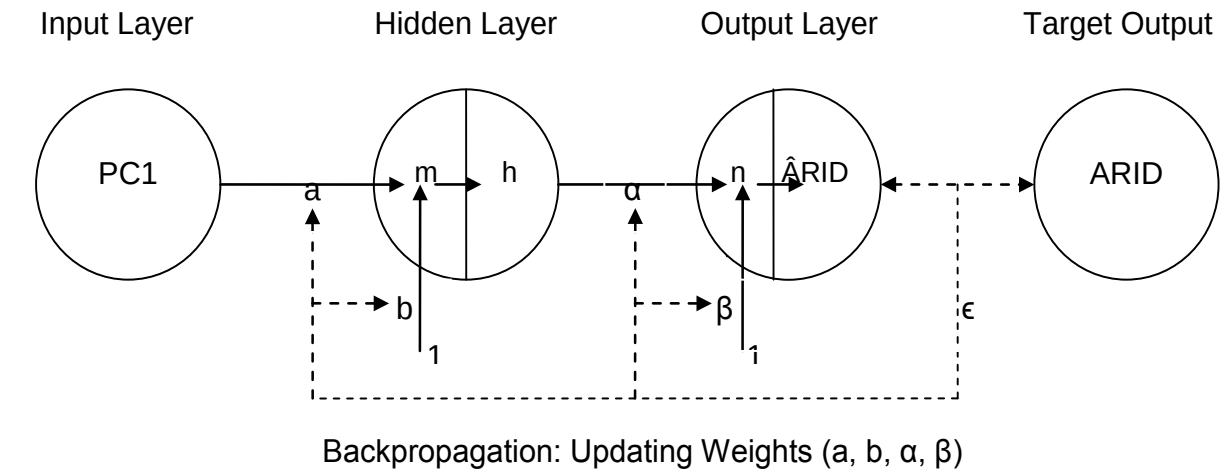

Figure 4-1. The architecture of a Multi-layer Perceptron (MLP) with one input layer, one hidden layer, and one output layer; each with one neuron

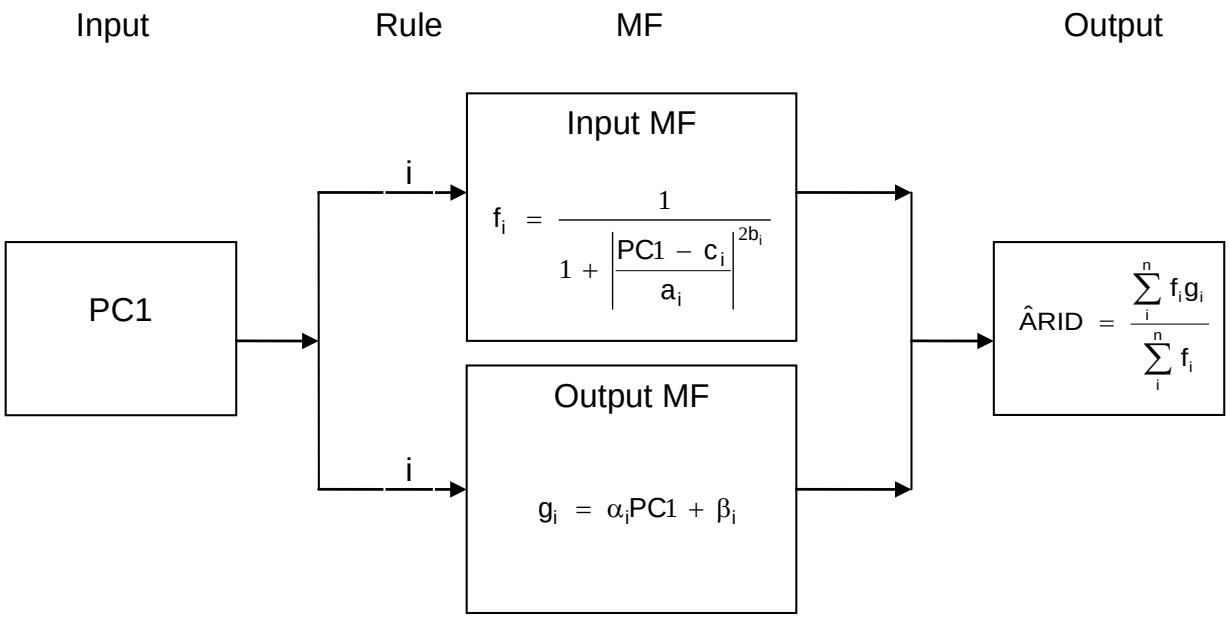

Figure 4-2. A first-order Sugeno-type Fuzzy Inference System with one input variable, PC1

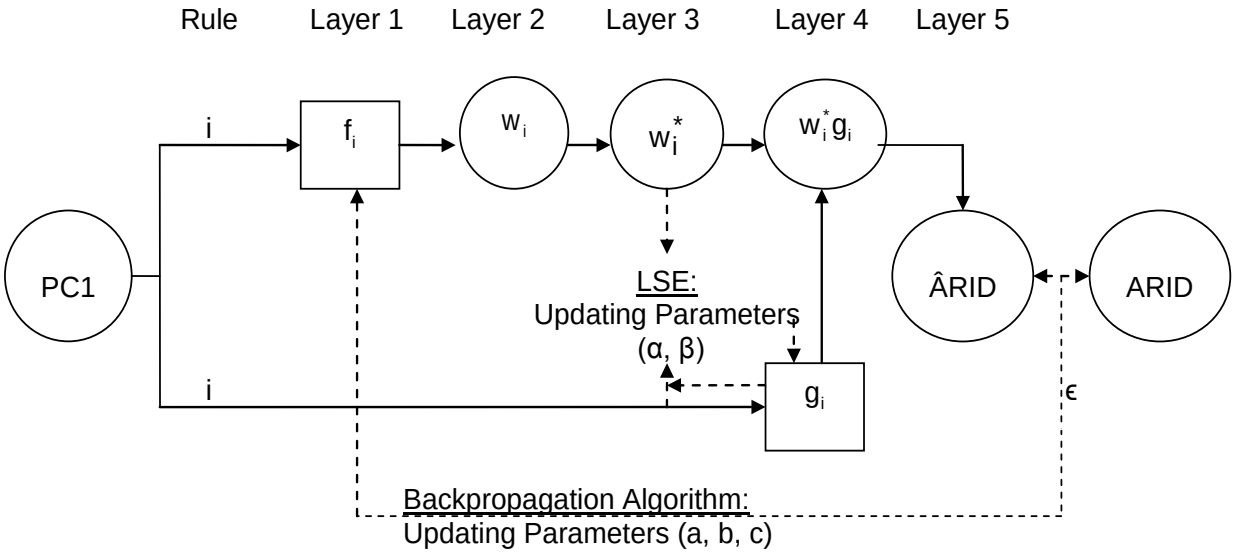

Figure 4-3. The architecture of an Adaptive Neuro-fuzzy Inference System (ANFIS) with one input variable, PC1

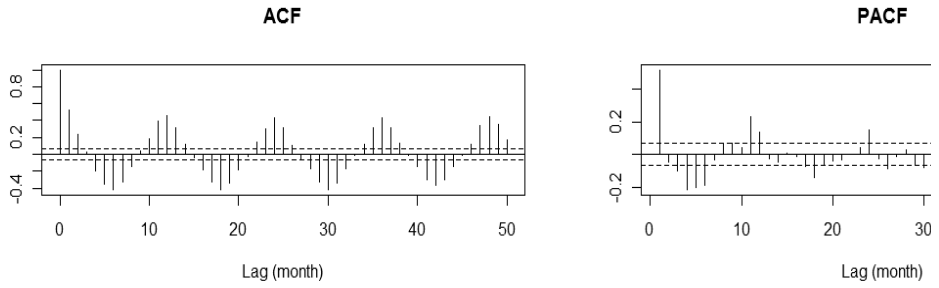

Figure 4-4. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) of ARID for Miami in January

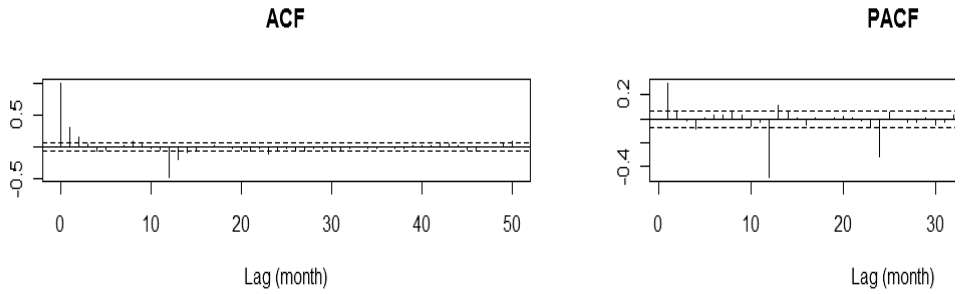

Figure 4-5. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the differenced ARID for Miami in January

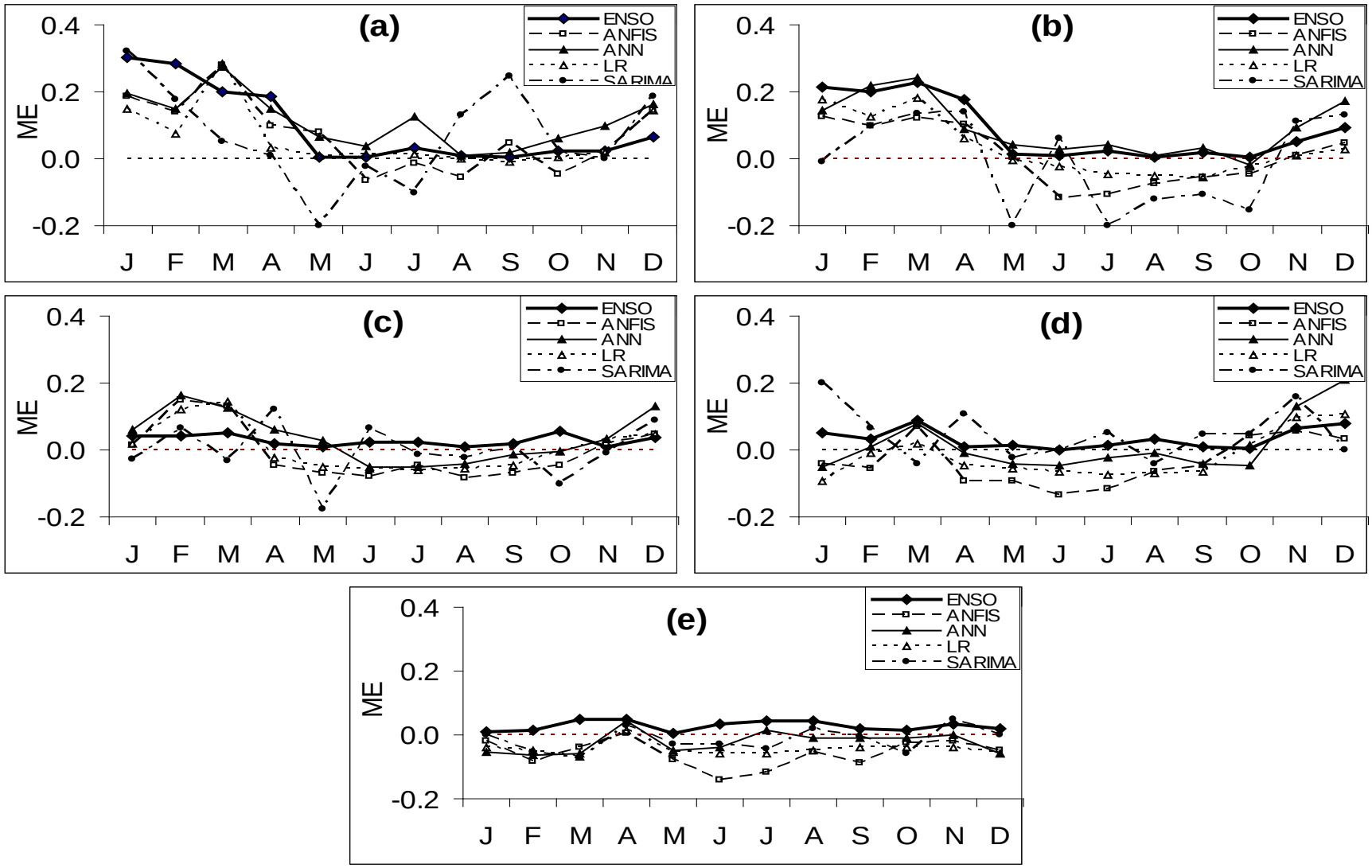

Figure 4-6. Values of modeling efficiency (ME) of various models associated with 1-month ahead forecasting across months for: a) Miami, b) Bartow, c) Live Oak, d) Plains, and e) Blairsville. Note: J = January, …, D = December.

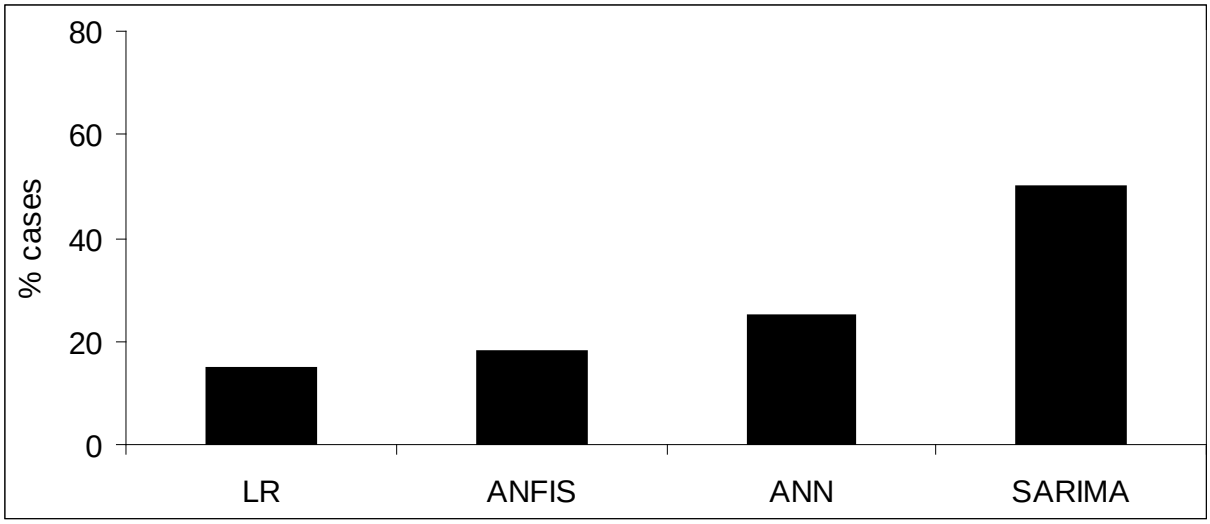

Figure 4-7. The percentage of occurrence of $RMSEP_{t+1} < RMSEP_{t+2} < RMSEP_{t+3}$ for different forecasting models, where $RMSEP_{t+1}$, $RMSEP_{t+2}$, and $RMSEP_{t+3}$ denote the RMSEP values associated with 1-, 2-, and 3-month ahead forecasting, respectively. There were 60 cases in total (5 locations x 12 months), so 50% means 30 cases.

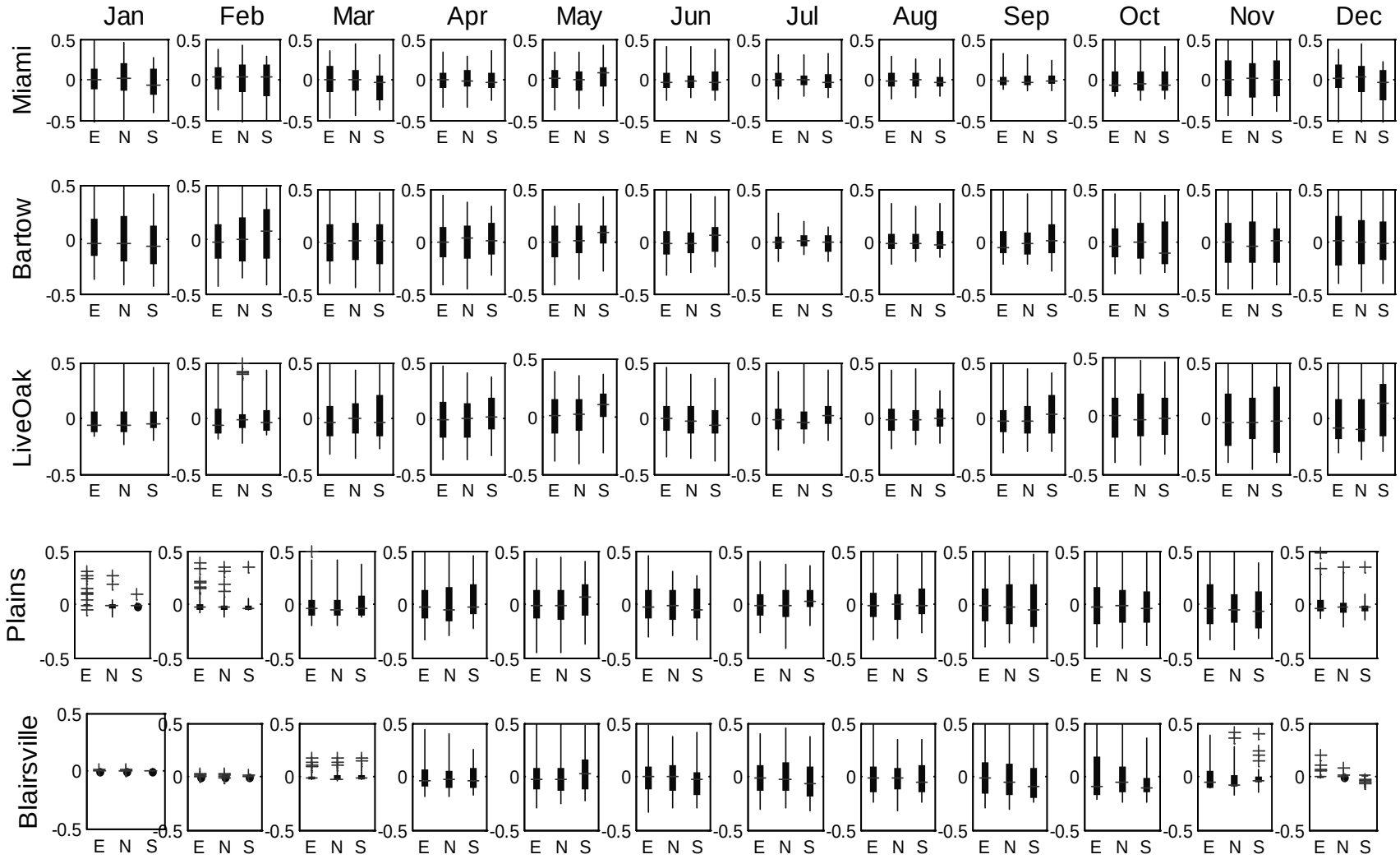

Figure 4-8. Box and whisker plots of errors, computed as the difference between observed and estimated ARID by ANN, ENSO, and SARIMA models, for various months and locations. E = ENSO, N = ANN, and S = SARIMA.

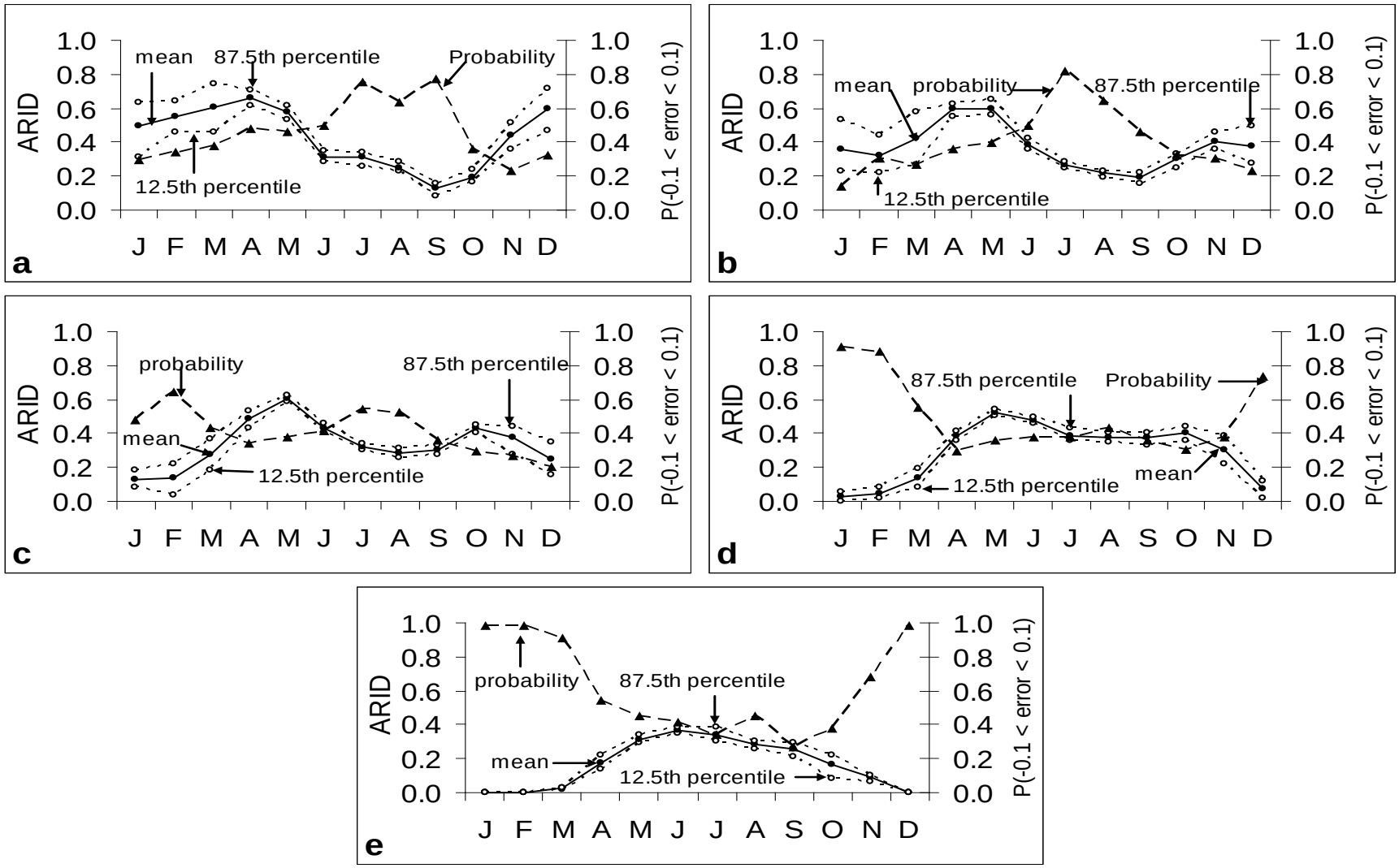

Figure 4-9. Distributions of ARID with mean and the 12.5th and 87.5th percentile values as estimated by ANN models across months for: a) Miami, b) Bartow, c) Live Oak, d) Plains, and e) Blairsville. Also shown are values of P(-0.1 < prediction error < 0.1). Note: J = January, F = February, …, D = December.

CHAPTER 5
ESTIMATING CROP YIELDS WITH AGRICULTURAL REFERENCE INDEX FOR DROUGHT

Introduction

One of the intended, and probably the most important, uses of an agricultural drought index is to estimate yield loss from drought (Sakamoto, 1978; Kumar and Panu, 1997; Quiring and Papakryiakou, 2003). The purpose of this study was to test the ability of ARID to estimate yield loss from drought for some of the major crops grown under rainfed conditions, such as cotton, maize, peanut, and soybean, through exploring associations between ARID and crop yields. Specifically, this study was carried out to seek answers to the following questions. How to estimate crop yields or the yield losses caused by drought using ARID? Can ARID, a generic index, make accurate estimations of yields or yield losses from drought for specific crops?

Materials and Methods

Deriving ARID-Yield Relationship

Briggs and Shantz (1913a) found a close relationship between crop transpiration and crop dry matter production, which they quantified using the transpiration ratio, defined as the ratio of transpiration to dry matter yield. This work was probably the most widely recognized investigation of water use by crop plants at the time (Shantz and Piemeisel, 1927; Jensen, 1968). The relationship used to model crop yield vs. ARID is based on this relationship, which is:

$$TRR = \frac{TR}{Y_{DM}} \tag{5-1}$$

where TRR is the transpiration ratio (mm kg^{-1}), TR the transpiration (mm), and Y_{DM} the crop drymatter produced (kg dry matter or simply kg dm).

The Briggs and Shantz equation can be rearranged with TRR^{-1} replaced by α to result in the following relation between dry matter yield and transpiration:

$$Y_{DM} = \kappa TR \quad (5\text{-}2a)$$

where κ is a proportionality constant (kg dm mm^{-1}), a parameter that depends on crop species and growing conditions.

Because κ is the reciprocal of transpiration ratio, it can be regarded as water use efficiency (Viets, 1962), which is the amount of dry matter produced per unit of water transpired. The dry matter water use efficiency, κ, depends on climatic factors that are associated with evaporative demand and is controlled to a large extent by atmospheric relative humidity (Arkley, 1963).

The ratio of economic yield, such as grain yield, to total dry matter yield is called harvest index. Multiplying both sides of Equation 5-2a by harvest index produces:

$$\chi Y_{DM} = \chi \kappa TR \quad (5\text{-}2b)$$

where χ is the harvest index (kg economic yield per kg dm yield).

Defining the product of harvest index and total dry matter yield as economic yield (or, simply, yield) and defining the product of χ and κ as ψ, yield can be estimated as:

$$Y = \psi TR \quad (5\text{-}3a)$$

where Y is economic yield (kg); and ψ is the water use efficiency for economic yield (kg Y mm^{-1}), also known as net water use efficiency (Jensen et al., 1990).

Equation 5-3a also shows the relationship between potential yield – the yield when soil moisture does not limit growth – and potential transpiration, the transpiration that occurs when soil water is not limiting:

$$P = \psi TR_p \quad (5\text{-}3b)$$

where P is potential yield (kg), and TR_p is potential transpiration (mm).

Because ψ is assumed to be constant for a given crop (Hanks, 1974), the following general relationship between relative yield and relative transpiration can be obtained for a given crop and year by dividing Equation 5-3a by Equation 5-3b.

$$\frac{Y}{P} = \frac{TR}{TR_p} \tag{5-4}$$

where Y/P is the relative yield of a crop, and TR/TR_p is the relative transpiration occurred during the crop season. Equation 5-4 does not contain ψ as it cancels out.

Equation 5-4 is applicable only if the quantities on both sides of the equality sign belong to the same set of conditions such as of weather, soil, genotype, and management. For instance, the relative yield of a crop growing in a fertile soil is directly proportional to the relative transpiration of the crop growing in the same soil, other variables other than water remaining constant.

Because plant physiological processes are different during different crop phenological stages, a differential yield response to water deficiency occurs during each stage of the growing season (Jensen, 1968; Hiler and Clark, 1971; Hiler et al., 1974; Doorenbos and Kassam, 1979). A yield model that takes into account a series of stages within the growing season can reflect the effect of water deficit on yield better than the one that considers the whole season as a single stage (Jensen, 1968; Doorenbos and Kassam, 1979). Because crop yield depends on many factors besides drought, relative yield may be more reliably estimated than absolute yield based on a series of stage-specific relative transpiration ratios during the season. Generally, the effect of water deficit on crop yield at different growth stages can be expressed using either geometric or arithmetic principles (Tsakiris, 1982). The geometric method combines effects of

deficiencies occurring during different stages in a multiplicative form; the arithmetic method combines effects in an additive form. Recognizing that crop yield may not be linearly related to total water use when plants are stressed, Jensen (1968) chose the geometric method to reflect the possible drastic reduction of crop yield if a severe water deficit occurs during a particularly sensitive crop stage. He applied this approach for determinate, flowering crops, such as grain crops, which have several distinct growth stages with differing drought sensitivities. Applying the geometric approach, Equation 5-4 can be extended to the following relation to compute the relative yield for determinate crops:

$$\frac{Y}{P} = \prod_{m=1}^{M} \left(\frac{TR}{TR_p} \right)_m^{\lambda_m} \tag{5-5}$$

In Equation 5-5, the symbol Π indicates a product, 'm' is a crop stage, ' M' is the number of stages considered in the crop growing season, and λ_m is the relative sensitivity of the crop to water deficit during the m-th stage of growth. Using Equation 5-5, yield of a determinate crop may be estimated if potential yield and ratios of actual to potential transpiration and sensitivity coefficients for the different crop growth stages are known.

As explained in Chapter 1, ARID is calculated as a function of TR to ET_o ratio. In ARID, the ET_o is assumed to be approximately equal to TR_p for the reference grass which is dense and actively growing (Ritchie and Burnett, 1971).

$$\text{ARID} = 1 - \frac{TR}{ET_O} = 1 - \frac{TR}{TR_p} \tag{5-6}$$

Combining and rearranging Equations 5-5 and 5-6 results in the following general equation for estimating yield for a determinate crop based on the stage-specific average values of ARID:

$$Y = P\prod_{m=1}^{M}(1 - ARID)_m^{\lambda_m} \quad (5\text{-}7)$$

Equation 5-7 may be used to estimate yields from ARID if values of the parameters, namely potential yield (P) and stage-specific sensitivity coefficients (λ_m), are available. This equation, after linear transformation, may also be used for estimating P and λ_m using regression methods.

Sometimes, however, P may be estimated from observed values of non-water limited yield rather than having to estimate it as a parameter. In such cases, P can be used to compute relative yield (R) as a ratio of Y to P. Relative yield is defined as the yield that is produced under water-stress condition relative to the one that could be produced under non-water-stress condition, the other inputs remaining the same. For instance, the relative yield of 0.7 means if a crop yields 1.0 Mg ha^{-1} under irrigated conditions, it would yield 0.7 Mg ha^{-1} under rainfed conditions. The fraction of yield loss due to water deficit during the crop season is then computed by (1-R). To estimate relative yield (R), Equation 5-7 can be modified to:

$$R = \prod_{m=1}^{M}(1 - ARID)_m^{\lambda_m} \quad (5\text{-}8)$$

Data

For exploring associations between ARID and crop yields, four major crops grown in the southeast USA were studied – maize, cotton, soybean, and peanut. Locations in this region were selected based on the availability of both rainfed and irrigated yield

data from previous studies. While rainfed yields were needed for testing ARID, irrigated yields were used, when available, to estimate potential yields for each crop and season, and these values were used to compute relative yields. The rainfed crop yield data used in this study were obtained from the University of Georgia's Statewide Variety Testing (SWVT) program and from the database of the DSSAT crop growth model (Hoogenboom et al., 2004). For cotton, maize, and peanut, the SWVT website was the sole data source (www.swvt.uga.edu/index.html). The maize data were obtained for two locations in Georgia, whereas the cotton and peanut yields were from three locations each (Table 5-1). None of the studies had both irrigated and rainfed soybean yields for a given year in any location in the region. However, the locations with rainfed yields had irrigated yields but in different years. The soybean yield data for seven locations in Georgia were obtained from SWVT. The data for 1987 through 1996 were obtained through Mavromatis et al. (2001), whereas those for 1997 through 2008 were obtained from the SWVT website. The soybean data for one more location, Gainesville, FL, were obtained from the DSSAT database (Zur et al., 1982; Wilkerson et al., 1983; Hoogenboom et al., 2004; K. J. Boote, personal communication, 14 January 2009).

Although the yield data were available, cultivars were not consistent across locations and years. That is, for each crop, there was no common cultivar grown in all locations and years, and management was also different across locations and years. The SWVT website, the principal data source, consisted of a sheet of data for each crop, location, and year, each containing a range of yield records obtained from several trials conducted for a number of cultivars available in a year (http://www.swvt.uga.edu/2009/cn09/AP101-Coastal.pdf). To minimize the variability

and maximize consistency in the data and to make sure that the data reflect standard management, except water, the varieties that were within 95% of the highest yield each year were used as the rainfed yield data. Like rainfed data, the irrigated crops also had a range of yields for several cultivars and management. The highest two irrigated yields were averaged to approximate the potential yield for a given crop, year, and location in order to reduce observation errors. The observed data for cotton, maize, and peanut each had combinations of rainfed and irrigated yields each year and location.

Along with the rainfed yields, the planting and harvesting dates of each crop in each location and year were obtained and, from them, the duration of each crop growing season was computed. In general, each crop season had a duration of about five months across locations and years. To account for the differential responses of crops to water deficit in different periods of time during the growing seasons, each crop season was divided into five periods, each with 30 days. The first 30-day period was counted from the day of planting. The five months were used to approximate five growing stages, namely juvenile, vegetative development, flowering, grain-filling, and ripening.

To compute ARID values, daily historical weather data for the corresponding locations and years were collected from two different sources: the Florida Climate Center (ftp://coaps.fsu.edu/pub/griffin/SECC/AgroClimate/raw/ascii/) for locations in Florida and the Georgia Automated Environmental Monitoring Network (G. Hoogenboom, personal communication, 9 March 2009) for locations in Georgia.

Computations

Using the historical daily weather data, values of $ARID_{j,k,t}$ were computed for every location j, year k, and day t and later converted to $(1\text{-}ARID_{j,k,t})$ values. Definitions of

these and the other variables are presented in Table 5-2. Then, starting from the day of planting, daily values of ($1\text{-}ARID_{j,k,t}$) for each location and crop growing season were averaged to create five successive 30-day (monthly) mean values, one for each designated crop stage m ($1\text{-}ARID_{j,k,m}$). These values and the corresponding observed yields for crop i ($Y_{i,j,k}$) were regressed using Equation 5-9, 5-10, 5-11, or 5-12, depending upon the case described below, to estimate the potential yield ($P_{i,j,k}$) and stage-specific sensitivity coefficient parameters ($\lambda_{i,j,m}$).

Equations 5-7 and 5-8 are based on Equation 5-4, which expresses the relationship between relative yield and relative transpiration. Two approaches were used to estimate potential yield (P) for the regression analyses:

- Using Equation 5-7. In this case, potential yield ($P_{i,j}$) was assumed to be a parameter and estimated using the regression analysis.
- Using Equation 5-8. In this case, potential yield ($P_{i,j}$) was not assumed to be a parameter. Using rainfed yields as actual yields ($Y_{i,j,k}$) and irrigated yields as potential yields ($P_{i,j,k}$), relative yields ($R_{i,j,k}$) were computed as the ratio of actual to potential yields. Values of $R_{i,j,k}$ were then used in the regression analysis.

Considering physiological processes, values of sensitivity coefficients of a crop may be the same across locations. However, because of the absence of solid evidence to support this assumption, two more analyses were carried out:

- Assuming that stage-specific sensitivities are local. That is, they vary across locations. In this case, $\lambda_{i,j,m}$ were estimated.
- Assuming that stage-specific sensitivities are global. That is, they do not vary across locations. In this case, $\lambda_{i,m}$ were estimated.

The four cases mentioned above were analyzed as explained in the following paragraphs. Because basic plant physiological processes are expected to be independent of year, the stage-specific sensitivities to water deficit were assumed to be the same across years in all cases. They were assumed to depend only on the crop,

location, and phenological stage in case one and three and only on crop and phenological stage in cases two and four.

Case one

In this case, $P_{i,j}$ was assumed to be a parameter, and sensitivity coefficients ($\lambda_{i,j,m}$) were assumed to vary from location to location (Table 5-3). The parameters $P_{i,j}$ and $\lambda_{i,j,m}$ were estimated by fitting the data using Equation 5-9, the linearized form of Equation 5-7. In the equation, $P_{i,j}$ was assumed to be the same across years for each location.

$$\ln\left(Y_{i,j,k}\right) = \ln\left(P_{i,j}\right) + \sum_{m=1}^{5}\left[\lambda_{i,j,m}\ln\left(1-\mathrm{ARID}_{j,k,m}\right)\right] \tag{5-9}$$

Case two

In this case, potential yield ($P_{i,j}$) was assumed to be a parameter that varies by crop and location but remains the same across years (Table 5-4). Sensitivity coefficients ($\lambda_{i,m}$) were assumed to be the same for all locations in the region. The parameters, $P_{i,j}$ and $\lambda_{i,m}$, were estimated using Equation 5-10, a variant of Equation 5-9.

$$\left[\ln\left(Y_{i,j,k}\right) - \ln\left(P_{i,j}\right)\right] = \sum_{m=1}^{5}\left[\lambda_{i,m}\ln\left(1-\mathrm{ARID}_{j,k,m}\right)\right] \tag{5-10}$$

For regression, Equation 5-10 was used instead of Equation 5-9 because the latter with $\lambda_{i,m}$, which was assumed to be the same for all locations, would provide not $P_{i,j}$, but P_i, a single value for potential yield for the entire region, which is unrealistic. Because potential yield of a crop is defined by climatic factors that vary from location to location, P is expected to vary by location ($P_{i,j}$). So, using Equation 5-10, the location-specific potential yields ($P_{i,j}$) were estimated through an iterative non-linear regression procedure described in the following paragraph.

Using values computed from irrigated yields for initial estimates, the optimum values for $P_{i,j}$ were obtained through iterative regression. The highest irrigated yield of crop i observed in location j over a period for which the data were available was used as an initial estimate for $P_{i,j}$, denoted as $(P_{i,j})_1$, where the subscript 1 stands for the first (initial) estimate (Table 5-5). The initial estimate for crop i in location j – the first set – was $(P_{i,j})_1$, for j = 1 to J, the total number of locations for crop i. Using the multiplier 1±0.5, two other values were created from $(P_{i,j})_1$, namely $(1-0.5)x(P_{i,j})_1$ and $(1+0.5)x(P_{i,j})_1$. Using the three values of each of J locations – $[0.5(P_{i,j})_1, (P_{i,j})_1, 1.5(P_{i,j})_1]$, – 3^J combinations of $(P_{i,j})_1$ were created, where each combination had J elements, one for each location. Then, each combination was used in Equation 5-10 to estimate the parameters and calculate the root mean squared error (RMSE). From the 3^J combinations, the one with the least RMSE was selected as the second set of estimates with elements $(P_{i,j})_2$,. For this case, the multiplier 1±0.5 was reduced by 0.001to 1±0.499. Using the new three values of $P_{i,j}$ for each of J locations – 3^J combinations of $(P_{i,j})_2$ were created, applied in Equation 5-10, and RMSE values calculated. From these 3^J combinations, the one with the least RMSE was selected as the third set of estimates, and this process was continued 500 times. The RMSE values reached the global minimum and the estimates became stable by the 500th iteration. The final set of estimates – $(P_{i,1})_{500}$, $(P_{i,2})_{500}$, … , $(P_{i,j})_{500}$ – was selected as the optimum set of estimates for $P_{i,j}$ for J locations, and the corresponding coefficients, $\lambda_{i,m}$, were the estimates for the stage-specific sensitivity coefficients.

Case three

In this case, P was not a parameter, and sensitivity coefficients were assumed to be different across locations (Table 5-6). Because Equation 5-8 does not contain P, only

sensitivity coefficients ($\lambda_{i,j,m}$) were estimated using Equation 5-11, the linearized form of Equation 5-8. The relative yield ($R_{i,j,k}$) was calculated as the ratio of rainfed to irrigated yield ($Y_{i,j,k}/P_{i,j,k}$), considering the irrigated yield as the potential yield. From the observed values of rainfed yield ($Y_{i,j,k}$) and irrigated yield ($P_{i,j,k}$), values of $R_{i,j,k}$ were computed for crop i in location j and year k.

$$\ln\left(R_{i,j,k}\right) = \sum_{m=1}^{5}\left[\lambda_{i,j,m}\ln\left(1-\mathrm{ARID}_{j,k,m}\right)\right] \tag{5-11}$$

Case four

In this case, P was not a parameter, and sensitivity coefficients were assumed to be the same across locations (Table 5-7). The sensitivity coefficients, $\lambda_{i,m}$, were estimated using Equation 3-12. As in case three, $R_{i,j,k}$ was computed as the ratio of rainfed yield to irrigated yield ($Y_{i,j,k}/P_{i,j,k}$), both of which were observed values.

$$\ln\left(R_{i,j,k}\right) = \sum_{m=1}^{5}\left[\lambda_{i,m}\ln\left(1-\mathrm{ARID}_{j,k,m}\right)\right] \tag{5-12}$$

Once the values for the parameters of Equations 5-7 and 5-8 were estimated for each of the four cases, four ARID-based empirical yield models, one for each case, were developed for each crop. Then, using stage-specific monthly (30-day) values of $\mathrm{ARID}_{i,j,m}$ in the empirical yield models, absolute ($Y_{i,j,k}$) and relative ($R_{i,j,k}$) crop yields for the same locations and years belonging to the observed yields were estimated for each crop and case. Finally, to evaluate the performance of the yield models, estimated relative yields ($R_{i,j,k}$) were compared with computed relative yields ($Y_{i,j,k}$ / $P_{i,j,k}$) using the root mean squared error (RMSE), the mean absolute error (MAE), the Willmott Index (d-index: Willmott, 1981), and the modeling efficiency (ME: Nash and Sutcliffe, 1970) as measures of goodness-of-fit.

Results and Discussion

Figure 5-1 shows trends in ARID values that were used to estimate parameters for various crop yield models. For cotton, the mean growing season extended from the last week of May through the first week of October, and the values of ARID were greatest in the first month and decreased afterwards in each location. For maize, the ARID values were almost same in each month in Tifton, whereas those in Calhoun increased from the first month, April, through the last month, August. For peanut, whose growing season generally fell in between May and mid October, values of ARID in each location decreased from May to September and increased afterwards. For soybean, the planting date varied from late May to late June and the harvesting date from late October to late November depending upon the latitude of the locations. The ARID values for soybean in Tifton and Blairsville were nearly same across months, whereas those in Midville, Griffin, and Calhoun decreased in every month from planting to harvesting. However, the values in Gainesville, Quincy, and Plains fluctuated with no apparent pattern. In general, values of ARID for each crop and month in northern locations were greater than those in southern locations. This variation was mainly because of the difference in precipitation and temperature across locations.

Case One

The estimated values for $P_{i,j}$ for each crop and location were considerably different from the observed values, indicating that the approach used in case one was not appropriate for estimating potential yields (Table 5-8). In some cases, the estimates were unrealistic: 117174 kg ha^{-1} for soybean in Tifton, 11854 kg ha^{-1} for soybean in Plains, 904 kg ha^{-1} for cotton in Plains, and 172 kg ha^{-1} for cotton in Midville. These

unrealistic values of $P_{i,j}$ indicated that Equation 5-7 does not express the relationship between ARID and crop yields.

The $\lambda_{i,j,m}$ values for cotton in most of the months in Midville and Plains were negative (Table 5-9) for case one, in which location-specific coefficient estimates were made. Similarly, negative values of $\lambda_{i,j,m}$ were found for each of the four crops in most locations during their early development, pre-flowering, or full development stages. A negative value of the sensitivity coefficient during any stage indicates that water deficit has a positive effect on yield, which may occur to a small extent in some crops (de Bruyn, 1964; Singh, 1975). But it is unlikely that this effect would be large. For cotton, the fourth month is the most sensitive period as peak flowering and boll development occur during this period (Salter and Goode, 1967; Grimes et al., 1970; Hiler and Clark, 1971; Hiler et al., 1974; Doorenbos and Kassam, 1979). However, case one results showed that the second month – during vegetative development – and the fifth month – during crop maturation – are the most sensitive stages for Tifton and Plains, respectively. For peanut, the second month has the largest susceptibility to drought when peak flowering and early pegging stages occur (Salter and Goode, 1967; Hiler et al., 1974; Doorenbos and Kassam, 1979; Boote and Ketring, 1990). In contrast, the fourth month – when the crop is maturing – is the most sensitive stage for peanut grown in Midville and Tifton. For soybean, the largest sensitivity to drought occurs during the fourth month during the pod filling stage (Salter and Goode, 1967; Doss et al., 1974; Brown et al., 1985; Reicosky and Heatherly, 1990). Parameters estimated using the case one method were inconsistent relative to the most sensitive stage. The most sensitive month for Calhoun, Griffin, and Tifton was shown to be the first month – when

the crop is in the juvenile stage – and that for Plains to be the third month. Moreover, the $\lambda_{i,j,m}$ values for soybean ranged from -3.54 to 7.02 and were inconsistent across locations. Thus, these unrealistic $\lambda_{i,j,m}$ values, based on previous findings, estimated in case one did not support the hypothesis that the stage-specific sensitivity of a crop to water deficit varies across locations.

Case Two

As in case one, the potential yields estimated for various crops and locations, $P_{i,j}$, in case two were significantly different from the corresponding observed potential yields (Table 5-10). The estimated regional $\lambda_{i,m}$ values were negative in the first, third, and fifth months for cotton and in the first and fifth months for peanut, indicating that water scarcity during each of these months has a positive effect on yield (Table 5-11). Because water is essential during each of these months for good crop growth and development, it is likely that these negative values were caused by the statistical methods used to estimate potential yield. Maize was shown to be insensitive to water deficit except in the third month when flowering and pollination occur. Although the third month is the most critical, the second and fourth months are also susceptible to water deficit (Musick and Dusek, 1980). For soybean, the same level of sensitivity was shown for each of the first four months, contrary to the findings of Salter and Goode (1967), Doss et al. (1974), Brown et al. (1985), and other researchers that the pod filling stage, which occurs in the fourth month, is the most sensitive followed by flowering and vegetative stages, which fall in the third and second month, respectively. The unrealistic values of $P_{i,j}$ and $\lambda_{i,m}$ were probably because Equation 5-7 does not express the relationship between crop yields and ARID as P can vary across years. Accordingly, $\lambda_{i,m}$ estimated based on unlikely relationship were also unrealistic.

Results of cases one and two suggested that reasonable values for the parameter $P_{i,j}$ cannot be estimated either through linear regression using Equation 5-9 or through iterative regression using Equation 5-10. The results further indicated that reasonable absolute yields ($Y_{i,j,k}$) cannot be estimated using the ARID-based absolute yield model – Equation 5-7. In other words, P cannot be a parameter because P in Equation 5-4, from where Equation 5-7 was derived, is, in fact, not constant across years. Instead, every P, non-water-limited yield, is associated with the corresponding Y, rainfed yield, both of which belong to a given production environment except water. Since P in (Y/P) is not from an environment that is different from Y, it cannot take values from a different environment. In other words, potential yields cannot be the same across years due to variation in weather variables such as temperature and solar radiation.

Case Three

Cases three and four were based on relative yields (R) and were used to test the validity of Equation 5-8 only. Because this equation does not have P as a parameter, only sensitivity coefficients were analyzed in these cases.

The location-specific sensitivity coefficients ($\lambda_{i,j,m}$) estimated in case three for Equation 5-8 are presented in Table 5-12. For each of the four crops, the $\lambda_{i,j,m}$ values were negative for at least one month in each location. For instance, the $\lambda_{i,j,m}$ values for cotton were negative in the third month in Plains and Tifton and in the fourth month in Plains when the crop was in the preflowering and boll formations stage, respectively. For soybean, the negative values were estimated for several locations in the fourth month, the most sensitive period (Reicosky and Heatherly, 1990). For cotton, the second, third, and fifth months were shown to be the most sensitive for Tifton, Midville, and Plains, respectively, in contrast to previous findings that the most sensitive month is

the fourth month (Salter and Goode, 1967; Grimes et al., 1970; Hiler and Clark, 1971; Hiler et al., 1974; Doorenbos and Kassam, 1979). Moreover, the location-specific differences were not supported by any differences in precipitation, temperature, or soil type across locations. For maize, the fourth month was depicted as the most susceptible for Calhoun, whereas other researchers have demonstrated that the third month, which approximates the flowering stage, is the most susceptible (Salter and Goode, 1967; Musick and Dusek, 1980). For peanut, the most sensitive month was shown to be the first month for Tifton, the third month for Midville, and the fourth month for Plains, which was inconsistent with the results of other researchers that the second month, which approximates the pegging stage, is the most sensitive to water deficit (Salter and Goode, 1967; Hiler et al., 1974; Doorenbos and Kassam, 1979; Boote and Ketring, 1990). For soybean, case three results showed months other than the fourth to be the most sensitive in contrast to the findings of other researchers that the fourth month is the most sensitive (Salter and Goode, 1967; Doss et al., 1974; Brown et al., 1985). The unreasonable values of $\lambda_{i,j,m}$ indicated that stage-specific sensitivity coefficients are not likely to exist at a location level.

Case Four

Values for the stage-specific sensitivity coefficients at a regional level ($\lambda_{i,m}$) estimated in case four are presented in Table 5-13. In this case, crop stage during the second month – from 31 through 60 days after planting (DAP) – was found to be the most sensitive to drought for peanut. For maize, the third month – from 61 through 90 DAP – was the most sensitive. Both cotton and soybean were found to be most sensitive to water deficit during the fourth month – from 91 through 120 DAP. These

results were consistent with those of other researchers as described in the following paragraphs.

Peak flowering and boll development stages are the most susceptible stages to water deficit for cotton, and the stress occurring during this period causes severe yield loss (Salter and Goode, 1967). Grimes et al. (1970), Hiler and Clark (1971), Hiler et al. (1974), and Doorenbos and Kassam (1979) also found that peak flowering is the most sensitive stage. Because peak flowering and boll development stages fall in the fourth month of the growing season (Grimes and El-Zik, 1990), results of my study were in good agreement with those of the previous studies. Some researchers also suggested that a mild water deficit stimulates flower production (Cothren, 1999). Water stress during the preflowering period (square to first bloom), which falls in the second month of the season in the southeast USA (Grimes and El-Zik, 1990), increases the number of flowers and bolls per plant in some cultivars causing increased yield (de Bruyn, 1964; Singh, 1975). This phenomenon was reflected by the negative value of the sensitivity coefficient (-0.10) estimated by the present study for the second month.

For maize, Salter and Goode (1967) found flowering and early grain formation stages to be the most critical period for water deficits, followed by the yield formation stage. Similarly, Hiler and Clark (1971) and Hiler et al. (1974) found reproductive stage – silking and tasseling to soft dough – the most sensitive of all stages, followed by the maturity stage – after soft dough and ear development. Later, Doorenbos and Kassam (1979) found that flowering is the most sensitive to water deficit period. Because water deficit at silking, tasseling, and pollination are most detrimental to yield, the sensitivity of crop growth stages to water deficit occurs in the following order: flowering and

pollination, grain filling, and vegetative stress (Musick and Dusek, 1980). Because the vegetative stage falls in the second month; the silking, tasseling, and pollination stages in the third month after planting (Rhoads and Bennett, 1990; Nielson et al., 2002); and grain filling in the fourth month, this study showed the same phenomena as found by previous researchers.

For peanut, Salter and Goode (1967) found that the period between flowering and beginning of seed, which falls between 35 and 65 DAP, is the most critical growth period for water deficit. Similarly, Hiler et al. (1974) found the vegetative to peak flowering and early pegging stages, which fall from 37 through 57 DAP, to be the most susceptible stages to drought. Doorenbos and Kassam (1979) also found that the most critical stage of peanut crop to water deficit is flowering, which starts at about 35 DAP and peaks around 63 DAP (Boote and Ketring, 1990). The flowering and pegging stage is the most sensitive stage; water deficit at this stage causes greater yield loss than that at other growth stages (Boote and Ketring, 1990). Generally, the greatest effect of water deficit on economic yields has been reported to have been during flowering or anthesis because of the reduction in grain numbers (Howell, 1990). These findings indicate that the sensitivity coefficients estimated based on case four are consistent with those of previous studies.

The most sensitive stages of soybean to water deficit are peak flowering and pod set (Salter and Goode, 1967; Hiler and Clark, 1971; Hiler et al., 1974; Doorenbos and Kassam, 1979). Brown et al. (1985) found the crop to be most sensitive to water deficit around 100 DAP, with both weight and number of seed reduced. Doss et al. (1974) concluded that the pod filling stage, which falls around 90 to 120 DAP (Reicosky and

Heatherly, 1990), is the most critical period in terms of obtaining maximum yield. Our study showed similar phenomenon as peak flowering and podfill stages fall during the fourth month of the season.

Because the information given by the regional sensitivity coefficients estimated in case four was in line with the findings of other researchers, the hypothesis that all locations in a region have the same stage-specific sensitivities of a crop to water deficit was not rejected. The reasonable values of $\lambda_{i,m}$ led to the conclusion that Equation 5-8 expresses the relationship between ARID and relative crop yields. The likelihood of regional sensitivity coefficients also supported the results in cases one and two that Equation 5-7 does not express the relationship between ARID and absolute crop yields. The results further indicated that ARID can be used as a useful tool to detect the stage-specific sensitivity of crops to water deficit as well as the yield loss from drought.

Relative Yield and ARID Performance

The potential yield values for case four were estimated using observed irrigated data instead of estimating them using regression methods. And, there was only one set of sensitivity coefficients across locations for each crop. The relative yield model, Equation 5-8, of each of the four crops depicted the stage-specific sensitivities of the corresponding crop to water deficit fairly well. The yield models produced reasonable estimates of relative yields ($R_{i,j,k}$) of different crops (Figure 5-2).

To evaluate the performance of the ARID-based relative yield models, various measures of goodness-of-fit, such as RMSE, MAE, d-index, and ME were used. Values of RMSE, which denotes the average distance of a data point from the fitted line measured along a vertical line, were 0.18, 0.10, 0.11, and 0.17 kg ha^{-1} Y/ kg ha^{-1} P for the cotton, maize, peanut, and soybean models, respectively (Table 5-14). Values of

MAE, a measure of the average magnitude of the error that is not influenced by extreme outliers, also suggested that the yield models performed reasonably well. The difference of RMSE and MAE revealed that peanut had the smallest variance in individual errors, whereas soybean had the largest. The value of percent error of the model, computed as the ratio of RMSE to mean relative yield, was less than 27% for all crops. Values of the Willmott Index, a measure of the degree to which the observed values are approached by the model estimated values, indicated that the maize model agreed most closely with the observed data (d-index = 0.87), whereas the cotton model agreed the least (d-index = 0.55). Values of ME, a measure to assess the predictive power of models, also indicated that the agreement between the model-estimated and observed values was best for the maize model (ME = 0.63) and worst for the cotton model (ME = 0.03). Although modeling efficiency varied across the yield models, all yield models, except for cotton, performed well. Positive value of ME indicated that even the model estimated corn yields were more accurate than taking the mean of the observed data. Figure 5-2 illustrates the performance of the relative yield models. Of the four crop yield models, the performance of maize model was the best probably because maize is relatively more sensitive to water deficit than are the other crops compared. The performance of cotton model was relatively poor because cotton is the most drought-tolerant crop among the four crops compared. These results indicated that ARID-based yield models perform better especially for the crops that are sensitive to water deficit. In general, each of the four yield models did fairly well at estimating relative yields. The ARID-based yield models estimated relative yields of the crops with mean absolute errors of 0.08-0.15 kg ha^{-1} Y/ kg ha^{-1} P.

Conclusions

Despite large variations in observed yield data due to location differences in weather, soil, management, cultivars, and dates of planting and harvesting (Table 5-1, Figure 5-1), the ARID-based relative yield models quantified the overall effects of drought on yields of cotton, maize, peanut, and soybean. The results indicated that ARID has potential to estimate relative yields or the yield loss from drought for several crops. The ARID-based relative yield models estimated yields better than did the absolute yield models. The absolute yield models could not express the relationship between ARID and crop yields because the regression analysis was not able to estimate potential yield in addition to the sensitivity parameters in the models. Also, the poor performance of absolute yield models was probably because potential yields vary across years due to variation in temperature and solar radiation.

The stage-specific sensitivities of each of the four crops to water deficit were not found to vary by location. These results were likely because the sensitivity of a plant to water deficit, a physiological phenomenon, is expected to be influenced more by internal characteristics of the plant, which remains the same everywhere, than by external environmental factors, which vary from location to location.

Table 5-1. Number of rainfed crop growing seasons for cotton, maize, peanut, and soybean grown in several locations in the southeast USA

	Crop (i)							
	Cotton		Maize		Peanut		Soybean	
Location(j)	Seasons	DOP[a]	Seasons	DOP	Seasons	DOP	Seasons	DOP
Gainesville, FL							4 (78-85)	Jun 28 [14]
Quincy, FL							5 (98-00)	Jun 14 [15]
Tifton, GA	25 (99-08) [b]	Apr 27 [3] [c]	47 (97-08)	Mar 24 [14]	15 (97-08)	Apr 22 [15]	20 (87-92)	May 23 [5]
Plains, GA	17 (99-08)	May 02 [2]			27 (97-08)	May 07 [8]	10 (94-98)	May 14 [3[
Midville, GA	20 (99-08)	May 11 [12]			13 (97-08)	May 17 [10]	24 (87-00)	May 22 [8]
Griffin, GA							28 (87-00)	May 15 [4]
Calhoun, GA			17 (98-08)	Apr 16 [4]			24 (90-00)	May 20 [3]
Blairsville, GA							15 (87-95)	May 26 [3]
Total	62		64		55		130	

[a] Date of planting, mean values. [b] Figures in parentheses are years during which the yield data were available [(99-08) means 'from 1999 through 2008']. [c] Figures in square brackets are standard deviation of planting dates.

Table 5-2. Definitions of the subscripted variables used in Chapter 5

Variable	Definition
$ARID_{j,k,m}$	Values of ARID for location j, year k, and month m (independent of crop).
$ARID_{i,k,t}$	Values of ARID for location j, year k, and day t (independent of crop).
$ARID_m$	Values of ARID for month m (independent of crop).
i	Subscript referring to crops (1=cotton, 2=maize, 3=peanut, 4=soybean).
j	Subscript referring to location (1= Gainesville, 2=Quincy, 3=Tifton, 4=Plains, 5= Midville, 6=Griffin, 7=Calhoun, 8=Blairsville).
J	Total locations used in the study for a particular crop (2=maize, 3=cotton and peanut, 8=soybean).
k	Subscript referring to year, which vary depending on crop and location (see Table 5-1).
m	Subscript referring to crop stage of growth; m=1, 2,…, 5.
M	Total number of phenological stages considered in the crop growing season, which is 5.
n	Subscript referring to an iteration of potential yield ($P_{i,j}$) in Case two; n=1, 2,…, 5000.
$P_{i,i}$	Potential yield for crop i and location j (assumed not to vary across years).
$P_{i,j,k}$	Potential yield for crop i, location j, and year k.
$(P_{i,i})_n$	An estimate for $P_{i,i}$ for the nth iteration in Case two.
$R_{i,i,k}$	Relative yield for crop i, location j, and year k.
$Y_{i,j,k}$	Absolute yield for crop i, location j, and year k.
$\lambda_{i,i,m}$	Sensitivity coefficient for crop i, location j, and month m (assumed not to vary across years).
$\lambda_{i,m}$	Sensitivity coefficient for crop i and month m (assumed not to vary across locations or years).
λ_m	Sensitivity coefficient for month m.

Table 5-3. Summary of the technique used to estimate location-specific $P_{i,j}$ and $\lambda_{i,j,m}$ for yield models in Case one

Particular	Description
Purpose	To estimate parameters for Equation 5-7 for computing yields ($Y_{i,j,k}$) using location-specific sensitivity coefficients
Hypothesis	Potential yield ($P_{i,j}$) and sensitivity coefficients ($\lambda_{i,j,m}$) are parameters that vary across locations.
Fitting method	Multiple linear regression using Equation 5-9.
Regression analysis	Parameters are fit for each crop i and location j, individually.
Data used in regression	$Y_{i,i,k}$ and $ARID_{i,k,m}$.
Parameters estimated	$P_{i,j}$: crop- and location-specific but assumed to be the same for all years. $\lambda_{i,j,m}$: crop-, location-, and stage-specific but assumed to be the same for all years.

Table 5-4. Summary of the technique used to estimate location-specific $P_{i,j}$ and regional $\lambda_{i,m}$ for yield models in Case two

Particular	Description
Purpose	To estimate parameters for Equation 5-7 for computing yields ($Y_{i,j,k}$) using regional level sensitivity coefficients.
Hypothesis	Potential yield ($P_{i,j}$) is a parameter, and sensitivity coefficients ($\lambda_{i,m}$) do not vary across locations.
Fitting method	Iterative regression using Equation 5-10, varying $P_{i,i}$ until best fit was obtained.
Regression analysis	Parameters are fit for each crop i individually, using the data of all locations in the region.
Data used in regression	$Y_{i,j,k}$, and $ARID_{j,k,m}$. Values used for $P_{i,j}$ in Equation 5-10 were the initial estimates using irrigated yields for the iterative procedure.
Parameters estimated	$P_{i,j}$: crop- and location-specific but assumed to be the same for all years. $\lambda_{i,m}$: crop- and stage-specific but assumed to be the same for all locations and years.

Table 5-5. The highest irrigated crop yields (kg ha^{-1}) observed in several locations in the southeast USA over the periods for which data were available. These were used as initial estimates for $P_{i,j}$ and assumed to be the same across years for crop i and location j.

	Crop (i)			
Location (j)	Cotton	Maize	Peanut	Soybean
Gainesville, FL	-	-	-	4350 (533)
Quincy, FL	-	-	-	4320 (681)
Tifton, GA	2339 (343)[a]	14746 (1597)	6850 (783)	5003 (816)
Plains, GA	2303 (226)	-	6941 (679)	4465 (717)
Midville, GA	2539 (409)	-	6999 (1018)	4513 (722)
Griffin, GA	-	-	-	4055 (602)
Calhoun, GA	-	12841 (1436)	-	4270 (528)
Blairsville, GA	-	-	-	4371 (271)
No. of locations (n)	3	2	3	8

[a] Figures in parentheses are standard deviation values (kg ha^{-1}).

Table 5-6. Summary of the technique used to estimate location-specific sensitivity coefficients ($\lambda_{i,j,m}$) for yield models in Case three

Particular	Description
Purpose	To estimate parameters for Equation 5-8 to compute relative yields ($R_{i,j,k}$) with location-specific sensitivity coefficients
Hypothesis	P is not a parameter, and sensitivity coefficients ($\lambda_{i,j,m}$) vary across locations.
Fitting method	Multiple linear regression using Equation 5-11.
Regression analysis	Parameters were fit for each crop i and location j, individually.
Data used in regression	$R_{i,j,k}$ and $ARID_{i,k,m}$.
Parameters estimated	$\lambda_{i,j,m}$: crop-, location-, and stage-specific but assumed to be the same for all years.

Table 5-7. Summary of the technique used to estimate regional level sensitivity coefficients ($\lambda_{i,m}$) for yield models in Case four

Particular	Description
Purpose	To estimate parameters for Equation 5-8 to compute relative yields ($R_{i,j,k}$) with regional level sensitivity coefficients
Hypothesis	P is not a parameter, and sensitivity coefficients ($\lambda_{i,m}$) do not vary across locations.
Fitting method	Multiple linear regression using Equation 5-12.
Regression analysis	Parameters were fit for each crop i individually, using the data of all locations in the region.
Data used in regression	$R_{i,j,k}$ and $ARID_{j,k,m}$.
Parameters estimated	$\lambda_{i,m}$: crop- and stage-specific but assumed to be the same for all locations and years.

Table 5-8. Observed and estimated potential yields ($P_{i,j}$: kg ha^{-1}) in Case one for various crops and several locations in southeast USA. In Case one, it was assumed that potential yield is a parameter and sensitivity coefficients are local.

	Crop (i)							
	Cotton		Maize		Peanut		Soybean	
Location (j)	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated
Gainesville, FL	-	-	-	-	-	-	4350	-
Quincy, FL	-	-	-	-	-	-	4320	-
Tifton, GA	2339	1514	14746	9171	6850	5875	5003	117174
Plains, GA	2303	904	-	-	6941	3602	4465	11854
Midville, GA	2539	172	-	-	6999	7400	4513	6284
Griffin, GA	-	-	-	-	-	-	4055	4094
Calhoun, GA	-	-	12841	15353	-	-	4270	5956
Blairsville, GA	-	-	-	-	-	-	4371	2508

Table 5-9. Values of location level sensitivity coefficients ($\lambda_{i,j,m}$) for various crops and their stages as estimated in Case one, where it was assumed that potential yield is a parameter and sensitivity coefficients are local.

	Crop (i)													
	Cotton			Maize		Peanut			Soybean					
Stage/month (m)	M[a]	P	T	C	T	M	P	T	B	C	G	M	P	T
First (m=1)	-0.56	0.13	-0.11	-0.22	0.00	0.10	-0.27	-0.10	-0.41	0.70	0.36	0.23	-3.54	2.21
Second (m=2)	-0.22	-0.05	0.82	0.28	-0.02	0.59	0.09	0.33	0.01	-0.08	-0.08	0.23	0.76	1.59
Third (m=3)	-1.83	-1.01	-1.28	0.63	0.27	0.44	0.00	-0.06	-0.37	0.48	0.09	0.56	7.02	0.10
Fourth (m=4)	1.39	-0.09	0.38	0.53	0.01	0.67	0.21	0.18	0.29	0.13	0.17	0.61	-0.67	2.07
Fifth (m=5)	-0.72	0.46	0.01	0.10	-0.18	-0.70	-0.04	-0.08	0.01	0.00	-0.34	0.14	5.33	-0.03

[a] Location (j): M = Midville, P = Plains, T = Tifton, C = Calhoun, G = Griffin, and B = Blairsville

Table 5-10. Observed and estimated potential yields ($P_{i,j}$: kg ha^{-1}) in Case two for various crops in several locations in southeast USA. In Case two, it was assumed that P is a parameter and sensitivity coefficients are global.

	Crop (i)							
	Cotton		Maize		Peanut		Soybean	
Location (j)	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated
Gainesville, FL	-	-	-	-	-	-	4350	4371
Quincy, FL	-	-	-	-	-	-	4320	5726
Tifton, GA	2339	1257	14746	12500	6850	5429	5003	5003
Plains, GA	2303	973	-	-	6941	4429	4465	3661
Midville, GA	2539	825	-	-	6999	4060	4513	4739
Griffin, GA	-	-	-	-	-	-	4055	6447
Calhoun, GA	-	-	12841	10417	-	-	4270	5551
Blairsville, GA	-	-	-	-	-	-	4371	4371

Table 5-11. Values of regional level sensitivity coefficients ($\lambda_{i,m}$) for various crops and stages as estimated in Case two, where it was assumed that potential yield is a parameter and sensitivity coefficients are global.

	Crop (i)			
Stage/month (m)	Cotton	Maize	Peanut	Soybean
First (m=1)	-0.02	0.00	-0.21	0.27
Second (m=2)	0.02	0.00	0.22	0.18
Third (m=3)	-0.43	0.39	0.08	0.29
Fourth (m=4)	0.46	0.05	0.08	0.26
Fifth (m=5)	-0.29	0.05	-0.01	0.01

Table 5-12. Values of location level sensitivity coefficients ($\lambda_{i,j,m}$) for various crops and stages as estimated in Case three, where it was assumed that potential yield is not a parameter and sensitivity coefficients are local.

	Crop (i)													
	Cotton			Maize		Peanut			Soybean					
Stage/month (m)	M[a]	P	T	C	T	M	P	T	B	C	G	M	P	T
First (m=1)	0.18	0.23	-0.03	-0.22	-0.01	-0.31	-0.14	0.20	0.12	0.58	0.35	0.04	10.17	-1.44
Second (m=2)	0.00	0.01	0.40	0.02	0.13	0.09	0.23	0.08	-0.53	-0.25	-0.08	0.16	-0.04	0.50
Third (m=3)	0.51	-0.04	-0.58	0.41	0.26	0.62	0.17	-0.05	0.42	0.46	0.09	0.43	14.96	2.24
Fourth (m=4)	0.21	-0.03	0.27	0.74	0.14	0.21	0.31	0.17	0.78	-0.19	0.16	0.65	-3.22	-0.02
Fifth (m=5)	-0.12	0.46	0.20	-0.09	-0.11	0.26	0.09	-0.14	0.16	0.11	-0.33	0.10	16.97	0.32

[a] Location (j): M = Midville, P = Plains, T = Tifton, C = Calhoun, G = Griffin, and B = Blairsville

Table 5-13. Values of regional level sensitivity coefficients ($\lambda_{i,m}$) for various crops and their stages as estimated in Case four, where it was assumed that potential yield is not a parameter and sensitivity coefficients are global.

	Crop (i)			
Stage/month (m)	Cotton	Maize	Peanut	Soybean
First (m=1)	0.12	0.05	0.04	0.16
Second (m=2)	-0.10	0.09	0.27	0.15
Third (m=3)	0.17	0.36	0.13	0.21
Fourth (m=4)	0.34	0.11	0.08	0.30
Fifth (m=5)	0.04	0.03	0.07	0.01

Table 5-14. Values of the goodness-of-fit measures that were used to evaluate the performance of ARID-based relative yield models for various crops in Case four, where it was assumed that potential yield is not a parameter and sensitivity coefficients are global

	Crop			
Measure	Cotton	Maize	Peanut	Soybean
RMSE(kg ha^{-1} Y/ kg ha^{-1} P)	0.18	0.10	0.11	0.17
MAE (kg ha^{-1} Y/ kg ha^{-1} P)	0.15	0.08	0.09	0.14
RMSE minus MAE	0.029	0.026	0.021	0.034
Willmott d-index (-)	0.55	0.87	0.78	0.72
Modeling efficiency (-)	0.03	0.63	0.42	0.32

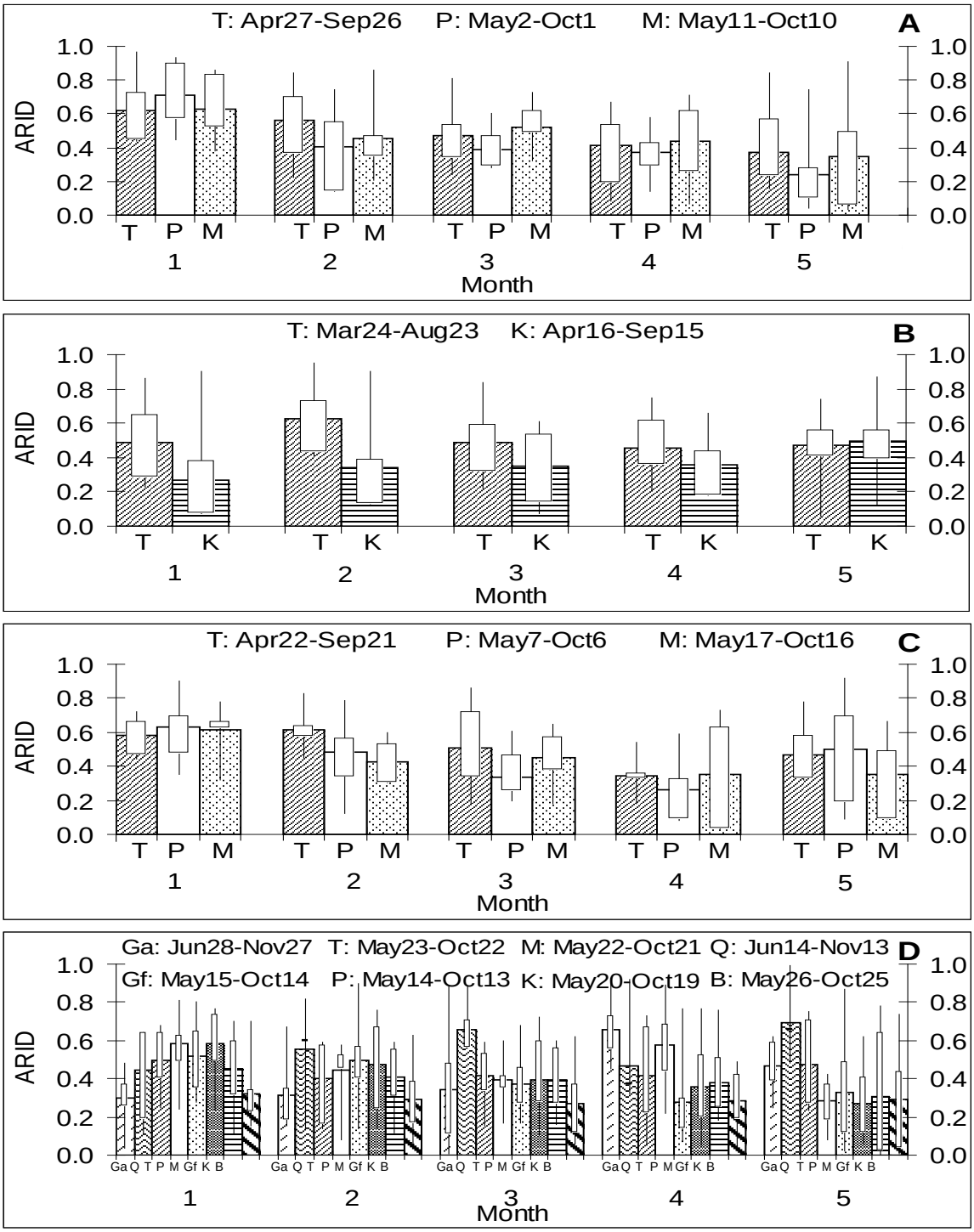

Figure 5-1. Average values of ARID across locations and months for: A) cotton, B) maize, C) peanut, and D) soybean. These are the means of all seasons (see Table 5-1). While the bars in each panel denote the mean values, the boxes and whiskers reflect the variability. Abbreviations: Ga = Gainesville, Q = Quincy, T = Tifton, P = Plains, M = Midville, Gf = Griffin, K = Calhoun, B = Blairsville.

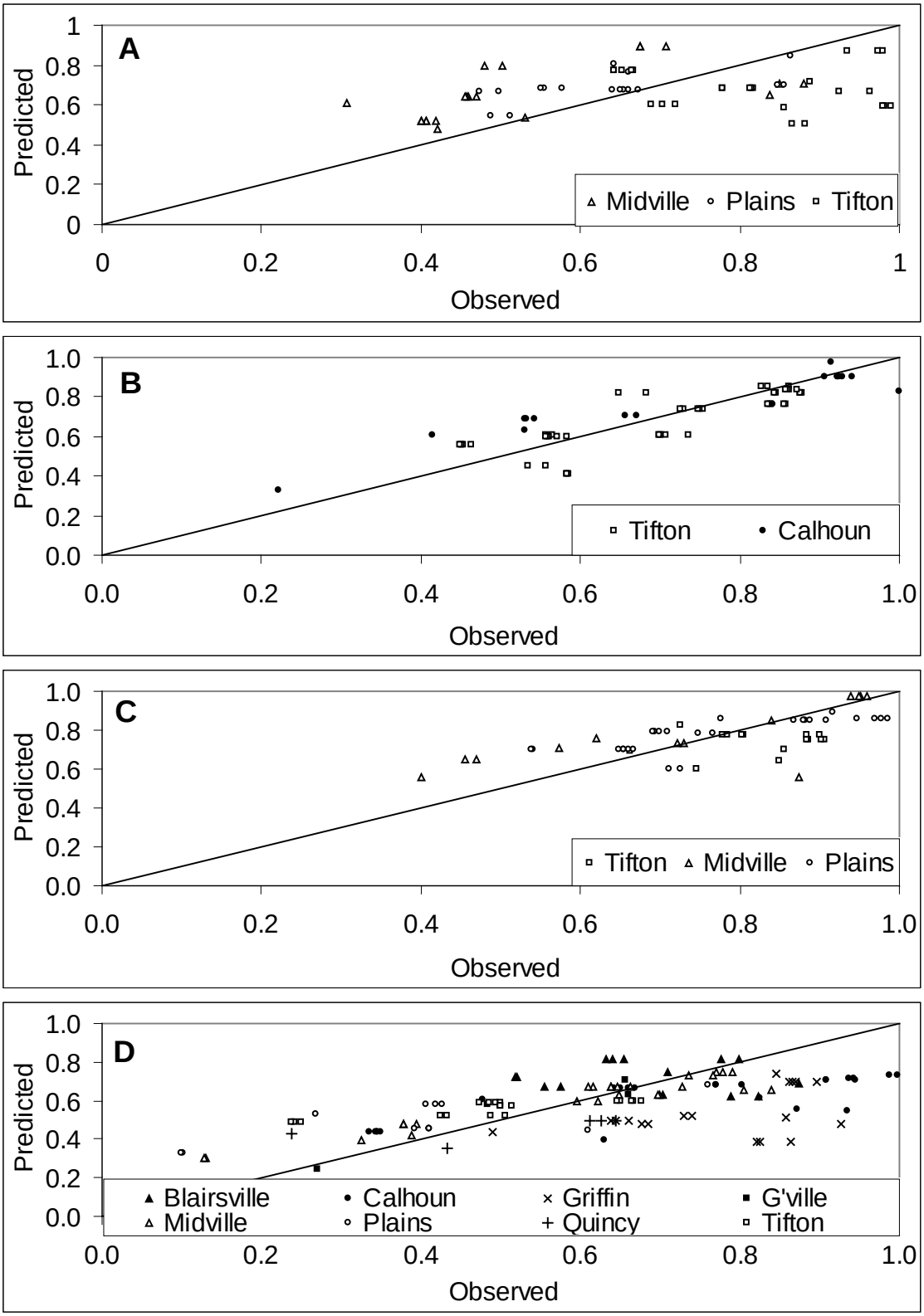

Figure 5-2. Model-estimated relative yields versus observed relative yields: A) cotton, B) maize, C) peanut, and D) soybean. The solid lines are the 1:1 lines.

CHAPTER 6
SUMMARY AND CONCLUSIONS

This study formulated a relatively simple, generic, and SPA-based drought index that can be used by farmers, lawn and garden managers, or people involved in crop production or similar systems as a decision support tool to monitor and forecast drought and estimate drought effects on yields. Specifically, this study: A) designed a generic drought index, called Agricultural Reference Index for Drought (ARID), for quantifying water deficit and its influence on crop yields; B) evaluated ARID by comparing it with the water deficit index computed by DSSAT CERES-Maize, a widely-tested and used crop model, and with eight other drought indices that have been applied to agricultural systems; C) quantified the uncertainties in ARID associated with its parameters and the importance of individual parameters with respect to the uncertainties of the index; D) explored methods for forecasting ARID as a predictor of agricultural droughts and compared their performances; and E) developed ARID-based yield models for several crops and assessed the ability of the index to estimate yield loss from drought. This chapter summarizes the conclusions of each of these phases.

ARID is expressed as the ratio of plant water deficit to plant water need. The latter, also called potential transpiration (TR_p), is estimated using the FAO-56 Penman-Monteith model. The former is computed as the difference between TR_p and actual transpiration (TR). In turn, TR is estimated based on simple soil water balance and a simple water uptake model, which gives ARID its simplicity. The genericness of ARID, on the other hand, comes from the use of grass, a reference crop that only requires site-specific weather data. Considering the time scale of crop responses to water deficits and daily variations in weather data, ARID is computed daily.

Results showed that ARID can mimic a complex, crop-specific crop simulation model. The index accurately approximated WSPD, the water deficit index calculated by the DSSAT CERES-Maize model. Values of ARID were better correlated with WSPD than were values of the other drought indices tested. Root mean squared errors between ARID and WSPD were smaller than those comparing WSPD with all other drought indices. Moreover, by capturing the effect of fluctuations in precipitation, temperature, solar radiation, and windspeed, ARID estimated the absolute magnitude of water deficit effects on crops more accurately than did the other indices. ARID is computationally simple, physically and physiologically sound, and generally applicable to characterize agricultural drought.), except for some locations and seasons that had. Even in the cases with large uncertainty ranges,

In general, uncertainty ranges in ARID were small, that is, less than 30% of the entire possible range of [0, 1]. Even when the index had relatively large uncertainty ranges of up to 50% of the total range at large amounts of precipitation (seasonal total of more than 400 mm), distributions of ARID peaked at its nominal values. That is, the majority of the values of ARID concentrated around its nominal values with small uncertainty ranges, thus indicating less uncertainty. Uncertainty was considerably influenced by the parameter distribution. With a change from uniform to normal distribution, uncertainty in ARID decreased as much as 50 percent depending on location and season. Sensitivity analysis showed that available soil water capacity was the most influential parameter and explained about 60% of the total variance in ARID. The second most important parameter was rootzone depth, which contributed about 30% to the total variability in ARID. Of the five parameters, runoff curve number and

drainage coefficient were the least influential due to insignificant difference in the amount of available water in the soil with a change in these parameters. The influence of uptake coefficient was also unimportant in most cases. This parameter never contributed more than 15% to the total variance in ARID. Being the most influential, available water capacity is the parameter that might be of most interest to users. If they want precise information about water deficit for their specific soils, they can adjust the value of this parameter.

Results related to the predictability of ARID showed that the index can be forecast using ENSO and teleconnection signals as inputs and artificial neural network and SARIMA models as forecasting methods. However, performances of these models varied depending on months and locations. Generally, climate index-based models and ENSO-based approaches performed better in the winter than in the summer due to stronger signals of ENSO and large-scale teleconnections in the winter. The efficiency of time series-based models depended largely on the periodicity of precipitation. Modeling efficiencies of ENSO-based and climate index-based models were larger in southern locations, such as Miami and Bartow, FL, than in the northern ones, such as Plains and Blairsville, GA, due to stronger signals of ENSO and teleconnections in the southern part of the region. The SARIMA models also performed better in southern locations due to more distinct periodicities of precipitation in the south. Of methods compared, artificial neural network performed best for most locations and seasons due to its ability to approximate the nonlinear relationship between inputs and outputs, whereas linear regression performed most poorly as it could not represent the nonlinearity between ARID and climate indices. Generally, both ENSO- and

teleconnection-based methods performed poorly for the northern part of the region. The performance of teleconnection-based methods was even poorer than that of the ENSO approach, indicating that the signals of large-scale teleconnection are weaker than those of ENSO in northern locations. Using one or more climate index- and time series-based models, the current level of forecasting, which is essentially ENSO-based, can be improved for several months, including summer, especially for the southern locations in the southeast USA. The climate indices used in this study have good potential to be used in forecasting drought particularly for the southern part of the region not only in the winter but also in the summer.

ARID-based crop yield models provided reasonable estimations of yield loss from drought for cotton, maize, peanut, and soybean even though ARID is not a crop-specific drought index. Compared with absolute yield models, which estimate drought-impaired yields, relative yield models, which estimate drought-impaired yields relative to non-water-limited yields, made better estimations. Stage-specific sensitivities of each of the four crops to water deficit did not vary by location, indicating that the sensitivity of a plant to water deficit in each specific stage is influenced more by genotype than environmental factors.

The findings of this study are summarized as follows:

- ARID, a simple, generic index, can mimic the water deficit estimations of the DSSAT CERES-Maize, a complex, crop-specific simulation model, which is widely-tested.
- ARID is able to estimate yield loss from drought for specific crops.
- ARID can quantify plant water deficit better than can other available drought indices.
- ARID provides a general indicator of water deficit for crops and is generally applicable to a wide range of conditions. However, if users want precise

information for their specific conditions, they can change values of available soil water capacity, the most influential parameter.

- Choice for parameter distributions can significantly affect the uncertainties in ARID.
- Stage-specific sensitivities of a crop to water deficit do not vary by location.
- Relative yield models give more accurate estimations of yield loss from drought than do absolute yield models.
- Using SARIMA and artificial neural network models and climate indices, namely AMO, NAO, JMA, Niño 3.4, PDO, and PNA, forecasting of drought can be improved especially for the southern part of the region both in winter and summer months.

Finally, ARID, an outcome of the study, may be used as a decision support tool by crop producers, lawn and garden managers, or people involved in plant production systems:

- To monitor plant water deficit by computing daily values of ARID,
- To forecast drought by computing monthly lead-time values of ARID using artificial neural network and past values of climate indices and ARID, and
- To estimate yield loss from drought for crops using the ARID-based yield models.

APPENDIX A
ASSESSING THE LAWN AND GARDEN MOISTURE INDEX AS A TOOL FOR MONITORING DROUGHT

Introduction

The Lawn and Garden Moisture Index (LGMI) is a drought index that is being used in the southeastern U.S. to monitor (http://nsstc.uah.edu/aosc/lawn_garden_se.html) and forecast (http://agroclimate.org/forecasts/LGMI/) droughts. The LGMI is calculated daily from Doppler radar estimates of precipitation with a 4 km × 4 km resolution.

The LGMI is a weighted estimate of the difference between available soil water and evaporative demand of shallow-rooted plants. Positive values indicate adequate precipitation or water surplus, negative values indicate water deficit.

$$LGMI_i = SM_i - ET_i \tag{A-1}$$

where $LGMI_i$ is the index for the i-th day; SM is the available soil moisture; and ET is the ET of shallow-rooted plants.

The SM is calculated as the weighted sum of precipitation over the previous three weeks. For the previous week, all precipitation is considered to be effective; for 2 and weeks previous, effective precipitation is reduced from actual precipitation by a linear scale (Figure A-1):

$$SM_i = \sum_{b=1}^{7} P_{i-b} + \frac{1}{14} \sum_{b=8}^{21} P_{i-b}(21.5 - b) \tag{A-2}$$

where b is the previous day, whose values range from 1 to 21; ‘i’ is the current day; and P_i is the amount of precipitation on the i-th day.

For instance, if precipitation in each day were 1 mm and uniform throughout the previous 21-day period, the total precipitation amount for the period would be 21 mm,

and the SM for day i would be 14 mm, comprised of 7 mm for the 7 days before present and 7 mm for 8 to 21 days before present.

While SM changes daily in response to precipitation amount, ET does not change with precipitation or environmental conditions. To keep LGMI simple and allow it to be computed from daily rainfall data, LGMI assumes the same ET function for every location, year, crop, and soil type (Figure A-2):

$$ET_i = \begin{cases} 31.75 - 19.05\cos[(\pi/120)(D-1)] & \text{for } 1 \le D \le 120 \\ 50.80 & \text{for } 121 \le D \le 250 \\ 31.75 - 19.05\cos[(\pi/115)(D-250)] & \text{for } 251 \le D \le 365 \end{cases} \quad \text{(A-3)}$$

where ET_i is the ET of plants on the i-th day (mm), and D is the day of year.

The curve in Figure A-2 shows the amount of moisture considered to be adequate for a specific day to sustain shallow-rooted plants. In winter, about 12.7 mm of rainfall per week is assumed sufficient to meet plant needs; whereas during the hottest time of the year, the requirement is assumed to be 50.80 mm per week.

Generally, a good agricultural drought index should be both easily computed and relevant for agricultural management. To make rational decisions based on an index, it is important to understand its predictability (Chen et al., 1997). The simplicity, high temporal (daily) and spatial (4 km × 4 km) resolutions, as well as the potential to forecast LGMI, make it attractive as a tool for monitoring agricultural drought. Although the LGMI was being used for monitoring drought in Alabama, Florida, Georgia, and South Carolina, its response to latitude, longitude, season, and ENSO phase were not quantified yet. The objectives of this study were to find out the relationships of LGMI with latitude and longitude and the responses of the index to changes in seasons and ENSO phases.

Materials and Methods

Locations and Precipitation Data

Seven counties were selected in the southeast U.S.: Cullman and Dallas, AL; Floyd, Jefferson, and Sumter, GA; and Columbia and Lee, FL (Figure A-3). The locations were chosen to represent the broad geographical region, at least 100 years of historical weather data, minimum missing precipitation data, and significant production area of corn, cotton, and peanut.

Selected locations had historical precipitation data for 112 years from 1893 to 2004. The dataset of each location was split into three categories based on ENSO phase: El Niño, Neutral, and La Niña, as defined by Japan Meteorological Agency (http://www.coaps.fsu.edu/products/jma_index.php). Of the 112 years, 25 were El Niño, 60 were Neutral, and 27 were La Niña. For any location, no more than 5% of the total days were missing precipitation data. These data gaps were filled in with the corresponding precipitation values from computer-generated bootstrapped data (Preston Leftwich, personal communication).

LGMI Computation

Using Equation A-2 and the gap-filled historical precipitation data, daily SM values were calculated for each year of each location. The SM for each location was first calculated for 22 January 1893 from the precipitation records of 1 to 21 January 1893 and then daily thereafter for the remainder of the dataset. Then, daily values of LGMI were computed by subtracting the daily values of ET calculated with Equation A-3 from the corresponding daily values of SM. The LGMI computation was made before dividing the dataset into ENSO phases.

After calculating daily LGMI for all the years of each ENSO phase and location, the index dataset of each location was split by month. For each month, there were 60 neutral, 25 El Niño, and 27 La Niña values. Then, daily LGMI values for each month were averaged and used for the analysis.

Analysis

Regression analyses were conducted to study the relationship of LGMI with latitude and longitude. Spatial variation in the index was determined using an analysis of variance (ANOVA) test. The hypothesis was whether or not mean LGMI values of the seven locations were different.

To identify seasonal effects on the relationships between the index and latitude or longitude, 3-month seasonal means were used in regression analyses: winter (January, February, and March); spring (April, May, and June); summer (July, August, and September); and fall (October, November, and December). For analyzing temporal variation in the index, the ANOVA was used to test the hypothesis that mean LGMI values of the four seasons differed significantly.

To explore the ENSO-phase based variation in LGMI, testing was carried out at two levels: phase-wise and month-wise. Thus, a test contained three phases: El Niño, La Niña, and neutral. As the test was conducted for each month-location combination, there were a total of 84 tests (12 months times 7 locations).

The driving variable for the index is precipitation, which is the only input. To understand how the index would behave with changes in precipitation, LGMI values were compared with corresponding precipitation. Comparisons were made for each location, season and ENSO phase.

When LGMI is negative, water deficit is indicated for shallow-rooted plants. Based on monthly average values of the index, the probabilities of water deficit as calculated by the index were computed for two locations, one in southern Florida and the other in northern Alabama; and for two months, one in summer and the other in winter.

Results and Discussion

Spatial Variation

In the winter, LGMI correlated positively with latitude; in the summer, the correlation was negative (Figure A-4). During the spring and fall seasons, no significant correlations were found. Similarly, LGMI correlated positively with longitude in the fall and winter, but there was no significant correlation in the winter or summer (Figure A-5). The increase in LGMI with latitude in the winter was due to an increase in precipitation from south to north in the winter. Conversely, precipitation appeared to follow the opposite trend in the summer (Figure A-6A). These results were in agreement with monthly mean historical precipitation values (Figure A-6B). Increase in LGMI with longitude in the fall and winter was due to an increase in precipitation from east (Jefferson) to west (Dallas) in winter months.

Temporal Variation

In general, LGMI values were smallest in the spring and greatest in the winter. This variation was because the precipitation was less than ET in the spring but greater than ET the winter. The ANOVA test further showed that mean LGMI values among the four seasons differed significantly ($P < 0.0001$). This difference was because of the direct influence of precipitation as seasonal mean precipitation values were also significantly different from each other. The positive correlation between LGMI and latitude during winter and the negative correlation during summer were due to the

southern locations receiving less precipitation in winter time but more in summer time. Similarly, the positive correlation between the index and longitude in winter was because of the western locations receiving comparatively more precipitation than the eastern locations.

ENSO Phase-Based Variation

The overall phase-wise comparison showed that LGMI differed significantly among ENSO phases ($P < 0.0001$). Among monthly mean values, LGMI was usually largest in El Niño years and smallest in La Niña years (Figure A-7), which reflects the fact that ENSO phases vary significantly in the amount and distribution of precipitation (Winsberg, 2003). Except for a few summer months, precipitation was generally greatest in El Niño phase and least in La Niña, which caused the index to vary accordingly. These results agree with the well established knowledge that the southern part of the region gets more precipitation in El Niño years and less in La Niña years than in the normal years, especially in winter months (Ahrens, 2003).

Month-wise comparisons showed significant LGMI differences among ENSO phases, especially in fall and winter months of November through March, and comparatively more so in southern locations (Table A-1). There were no significant differences among ENSO phases from May through October (Figure A-8). A month-wise comparison of precipitation showed the same pattern, with more significant differences in winter months and in southern locations. Significant differences in LGMI across ENSO phases reflect precipitation anomalies during El Niño or La Niña episodes (Figure A-9). During El Niño, precipitation was generally above normal – the long-term average of 112 years; whereas during La Niña, it was below normal. This result shows that ENSO has a strong effect during the winter in this region. Similarly, the significant

differences in the southern locations were due to an ENSO amplification effect. Although the amplification is greatest in the winter, ENSO amplifies weather events also in the spring, summer, and fall (Winsberg, 2003). Because of the amplification effect, southern locations received more rainfall in El Niño years and less in La Niña years than in neutral years. Precipitation differences across phases in summer were greater in the southern locations, which led to greater ENSO phase effects in southern locations.

Water Deficit Probability

For Columbia County, Florida, LGMI showed larger probabilities of water deficit (LGMI < 0.0) in November relative to July in all the ENSO phases (Figures A-10A and A-10B). For instance, the probability of deficit during a La Niña November was 1.0, whereas in July it was 0.7. For Cullman County in northern Alabama, however, the probability was smaller in November but larger in July in each phase (Figures A-10C and A-10D). For example, the probability of deficit during an El Niño July was 0.9, whereas in November it was 0.5. Probability of water deficit in the southern location was greater than in the northern location in winter for each ENSO phase (Figures A-10A and A-10C). On the other hand, the water deficit probability in July was less in the southern U.S. than in northern locations (Figures A-10B and A-10D). For instance, in November, the probability of deficit in La Niña was 1.0 for Columbia, while for Cullman it was 0.9. But in July, the probability in the same phase was 0.7 for Columbia, whereas for Cullman it was 1.0. The larger probability of water deficit in the southern locations in winter and smaller probability in summer were because of less precipitation in winter and more so in summer (Figure A-6B). Similarly, smaller water deficit probabilities in the northern locations during winter and larger during summer were because they received more precipitation during winter and less so during summer.

Conclusions

A positive correlation between LGMI and latitude was found in the winter; whereas in the summer, the correlation was negative. Similarly, a positive correlation between the index and longitude existed in the winter and the correlation was negative in the summer. The variations in LGMI with latitude and longitude over seasons were because of the corresponding differences in precipitation. Overall, the LGMI was significantly different across seasons and ENSO phases. At a monthly level, however, the variations across ENSO phases were significant mainly in winter months and in southern locations.

These results suggest that forecasts of LGMI based on ENSO may be possible for some locations and seasons in the southeast U.S. The historical weather data of a specific ENSO phase can be used to forecast the index for a corresponding ENSO year as LGMI was significantly different across ENSO phases, especially during winter and in southern locations. Particularly, gardeners, lawn owners, and golf course managers may benefit from these forecasts.

The LGMI, however, has some shortcomings. The ET function is vague and not dynamic; that is, it does not account for crop-specific differences in transpiration and the location-specific variations in weather. The SM function is also vague; that is, it does not take into consideration the specific soil processes and properties. Despite these limitations, its application might be continued after some improvement for quantifying water stress for lawns and gardens because it is simpler than many other drought indices and is also calculated daily.

For quantifying water deficit for agricultural crops, however, developing a new index is recommended than fixing LGMI because it does not account for many aspects

that are important from agricultural point of view. For instance, LGMI cannot be used for estimating crop yield loss from drought, probably the most important use of an agricultural drought index, because this index is not SPA-based, or physiology-based in particular. Because incorporating these aspects into LGMI completely changes its structure, it is better to develop a new agricultural drought index than to improve LGMI.

Table A-1. Monthly average LGMI values (mm) for seven locations in the southeast U.S. during different ENSO phases. Locations area arranged from southernmost (Lee) to northernmost (Floyd). There were no significant differences among ENSO phases for May through October (Figure A-7), so those data are not presented here.

Month	ENSO	Lee			Columbia			Sumter			Dallas			Jefferson			Cullman			Floyd		
Jan	El Niño	2.03	a*		8.64	a		16.51	a		13.21	a		9.65	a		14.99	a		12.19	a	
	Neutral	-4.57		b	5.08	a		12.19	a		14.22	a		7.37	a		18.80	a	b	17.53	a	
	La Niña	-7.62		b	2.54	a		10.92	a		13.21	a		7.11	a		24.64		b	17.02	a	
Feb	El Niño	-7.37	a		5.59	a		7.11	a		5.84	a		3.05	a		2.03	a		4.32	a	
	Neutral	-11.94	a		-3.56		b	5.08	a		6.35	a		1.02	a		8.38	a	b	9.14	a	
	La Niña	-17.02		b	-8.89		b	0.76	a		0.51	a		0.00	a		11.68		b	6.86	a	
Mar	El Niño	-17.02	a		-6.86	a		-3.81	a		-1.02	a		-4.06	a		-3.56	a		-1.27	a	
	Neutral	-25.65		b	-13.97	a		-10.92	a	b	-6.60	a		-14.99		b	-5.84	a		-6.10	a	
	La Niña	-31.75		c	-21.08		b	-13.97		b	-5.33	a		-16.26		b	-4.06	a		-3.81	a	
Apr	El Niño	-31.50	a		-24.38	a		-23.62	a		-18.80	a		-26.16	a		-19.05	a		-15.49	a	
	Neutral	-36.07	a	b	-28.19	a		-23.11	a		-18.80	a		-26.92	a		-16.00	a		-18.03	a	
	La Niña	-38.10		b	-32.77	a		-26.92	a		-21.59	a		-31.24	a		-18.54	a		-18.80	a	
Nov	El Niño	-12.19	a		-9.65	a		-4.57	a		-5.59	a		-7.87	a		4.57	a		0.25	a	
	Neutral	-17.02	a		-12.95	a		-12.19		b	-7.62	a		-14.48		b	-2.29	a	b	-6.60	a	
	La Niña	-15.24	a		-19.05		b	-17.53		b	-14.99		b	-17.53		b	-8.64		b	-9.91		b
Dec	El Niño	-3.05	a		7.62	a		12.45	a		13.72	a		6.86	a		15.75	a		14.99	a	
	Neutral	-7.62		b	-0.76		b	7.62	a	b	12.95	a		2.79	a	b	16.00	a		12.19	a	
	La Niña	-9.40		b	0.00		b	2.54		b	7.62	a		0.00		b	14.99	a		8.13	a	

* Means followed by the same letter within a month and location are not significantly different at $P < 0.05$ by Tukey-Kramer LSD test.

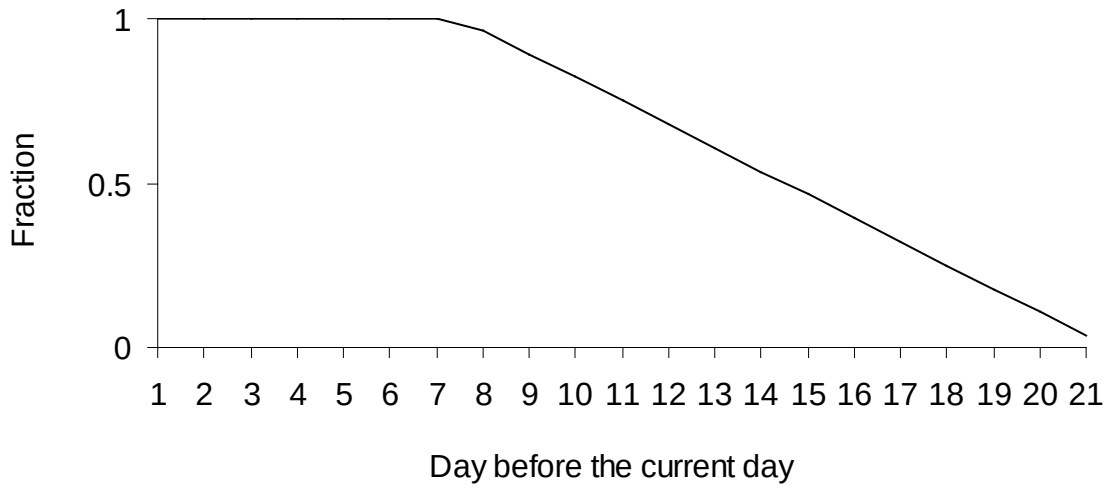

Figure A-1. Fraction of daily precipitation contributed to total available soil moisture (SM). While previous seven days contribute their whole amount, days from 8 to 21 contribute according to a linear scale.

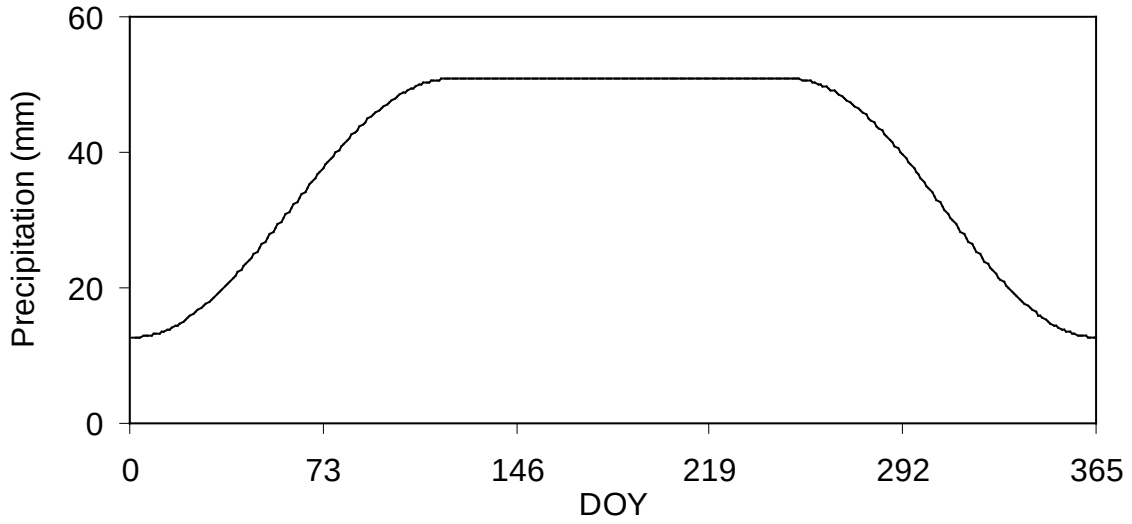

Figure A-2. The evapotranspiration (ET) function showing the amount of weekly effective precipitation considered to be enough for healthy lawns and gardens each day of the year.

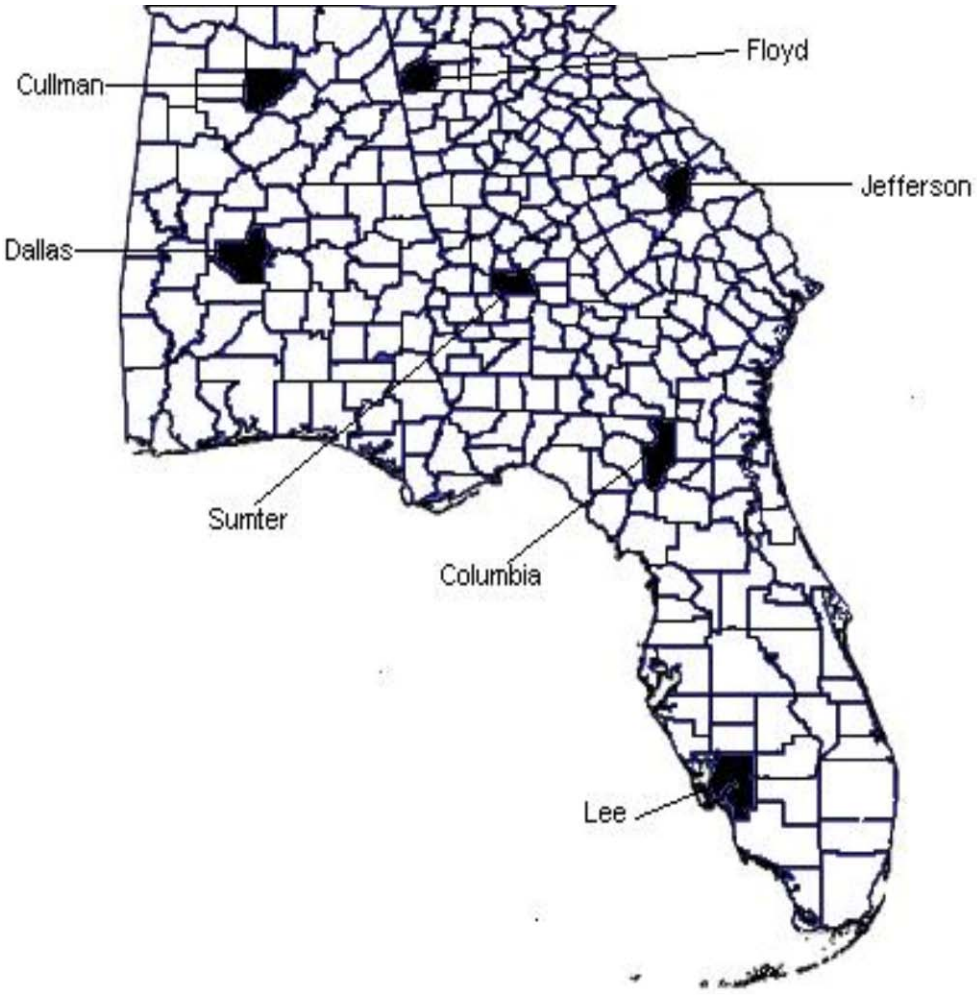

Figure A-3. Map of the SECC region showing the seven counties whose historical precipitation data were used for the study: Cullman and Dallas, AL; Columbia and Lee, FL; and Floyd, Jefferson, and Sumter, GA.

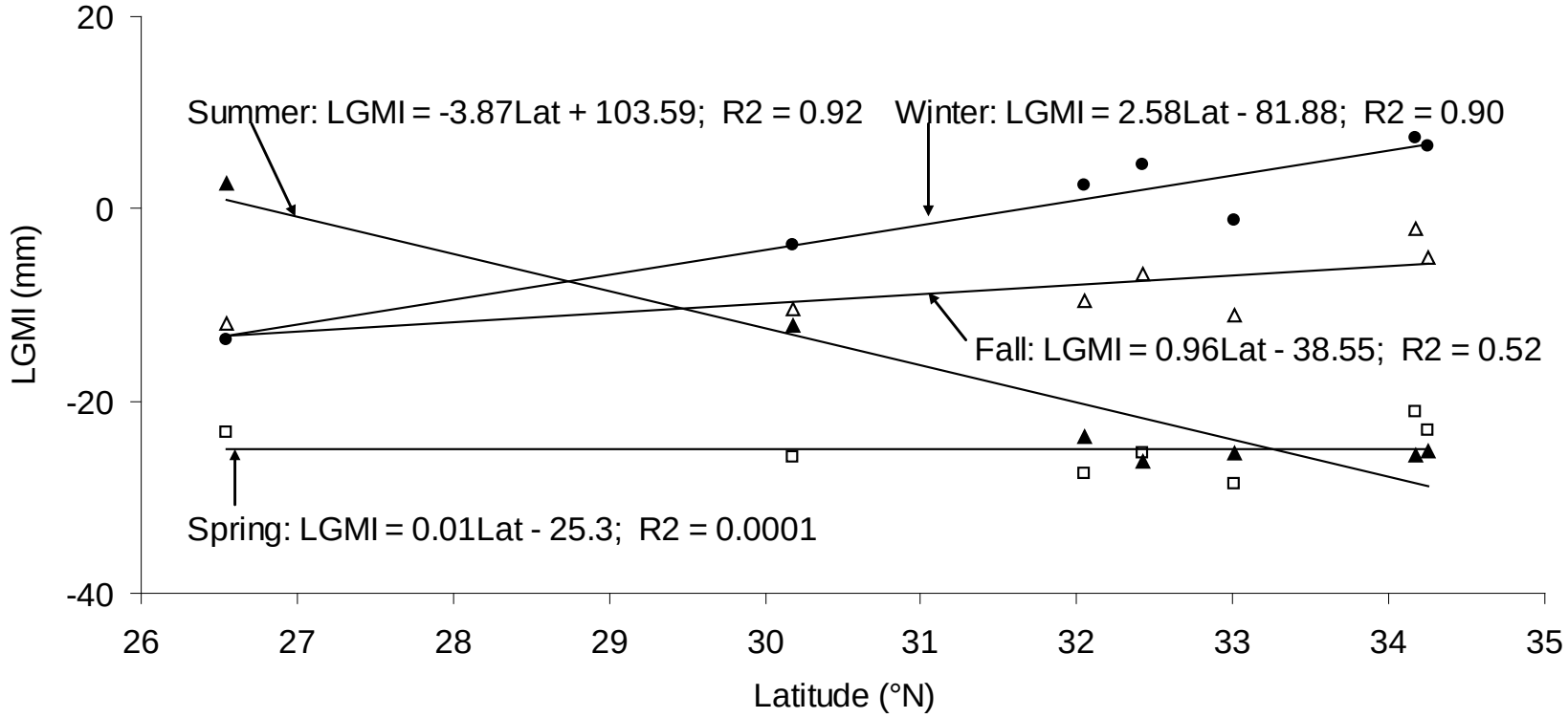

Figure A-4. Relationship of LGMI with latitude (Lat) in different seasons

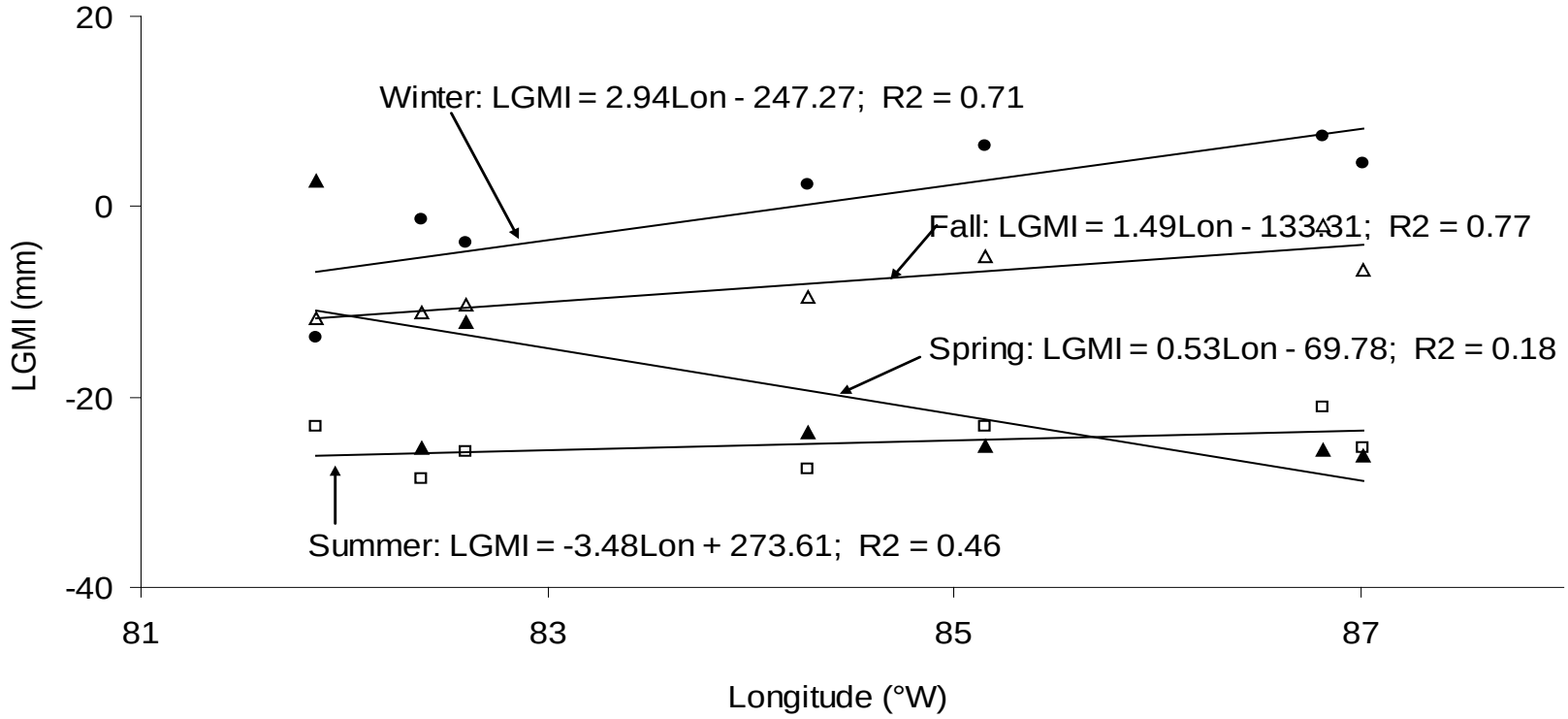

Figure A-5. Relationship of LGMI with longitude (Lon) in different seasons

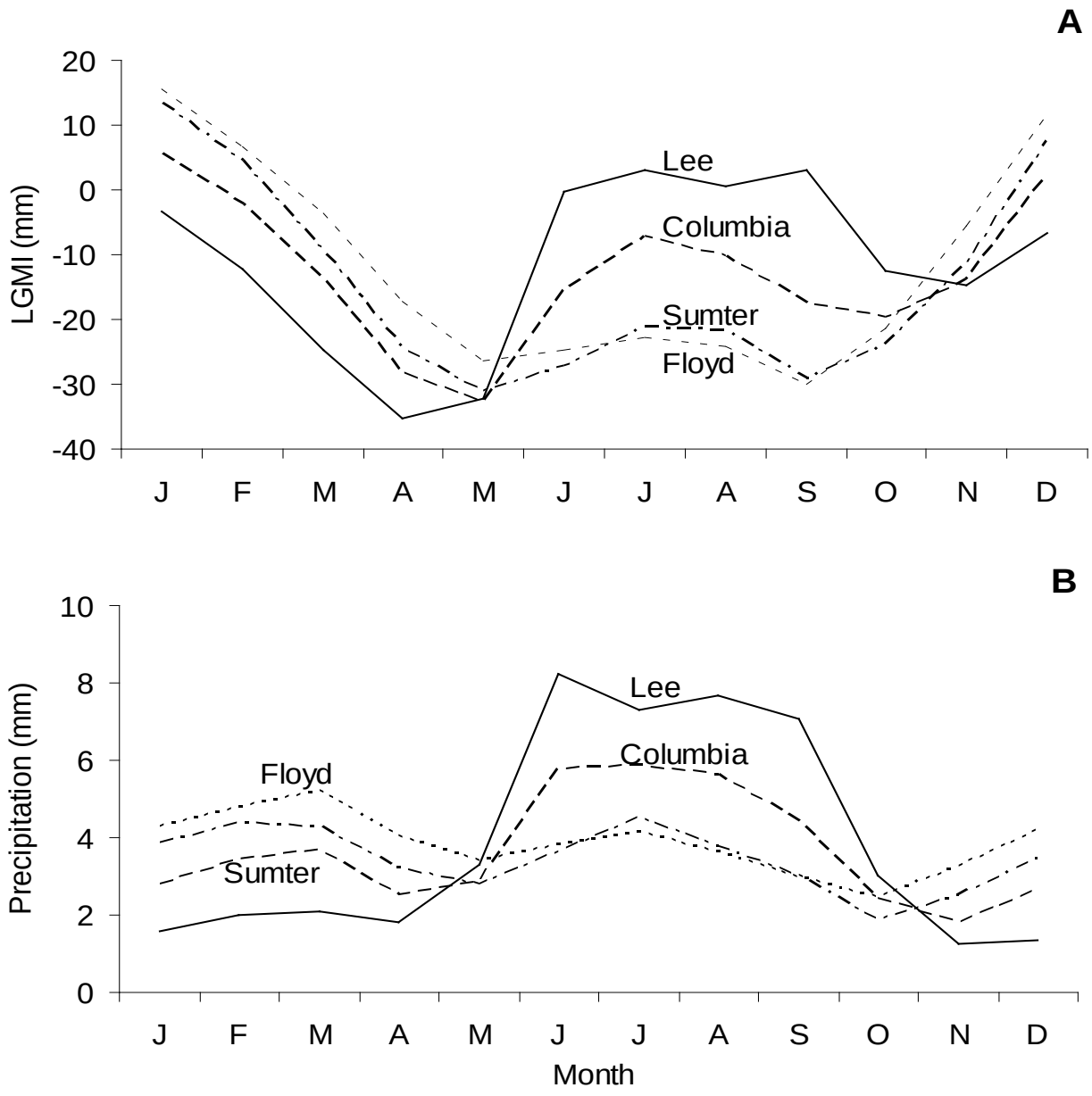

Figure A-6. A) Monthly mean LGMI and B) monthly mean precipitation for different locations over a year. Latitude of locations increases from Lee, Columbia, Sumter, to Floyd.

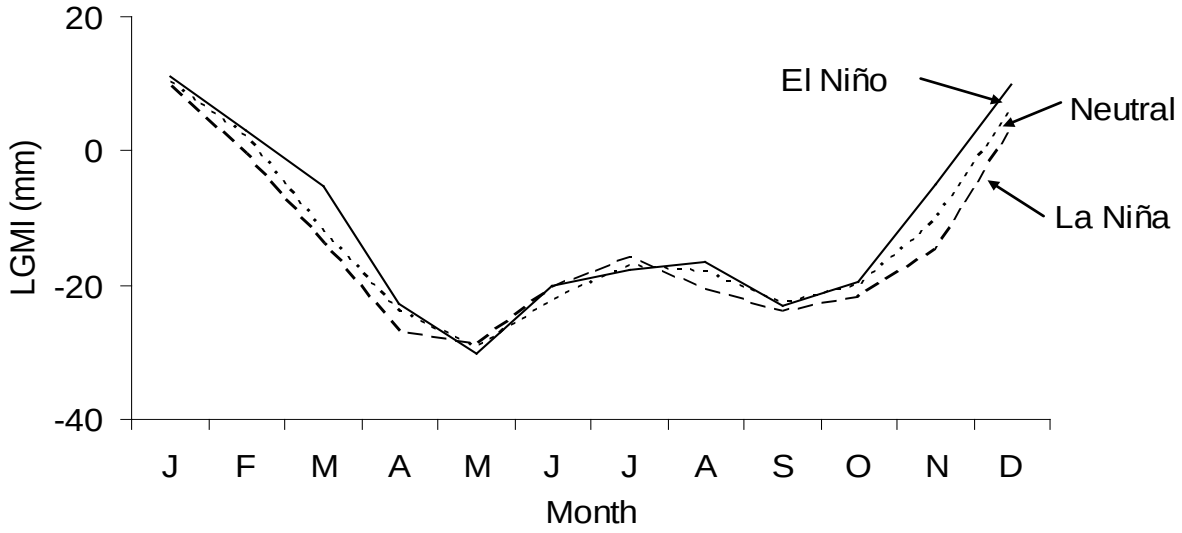

Figure A-7. Monthly LGMI values for different ENSO phases

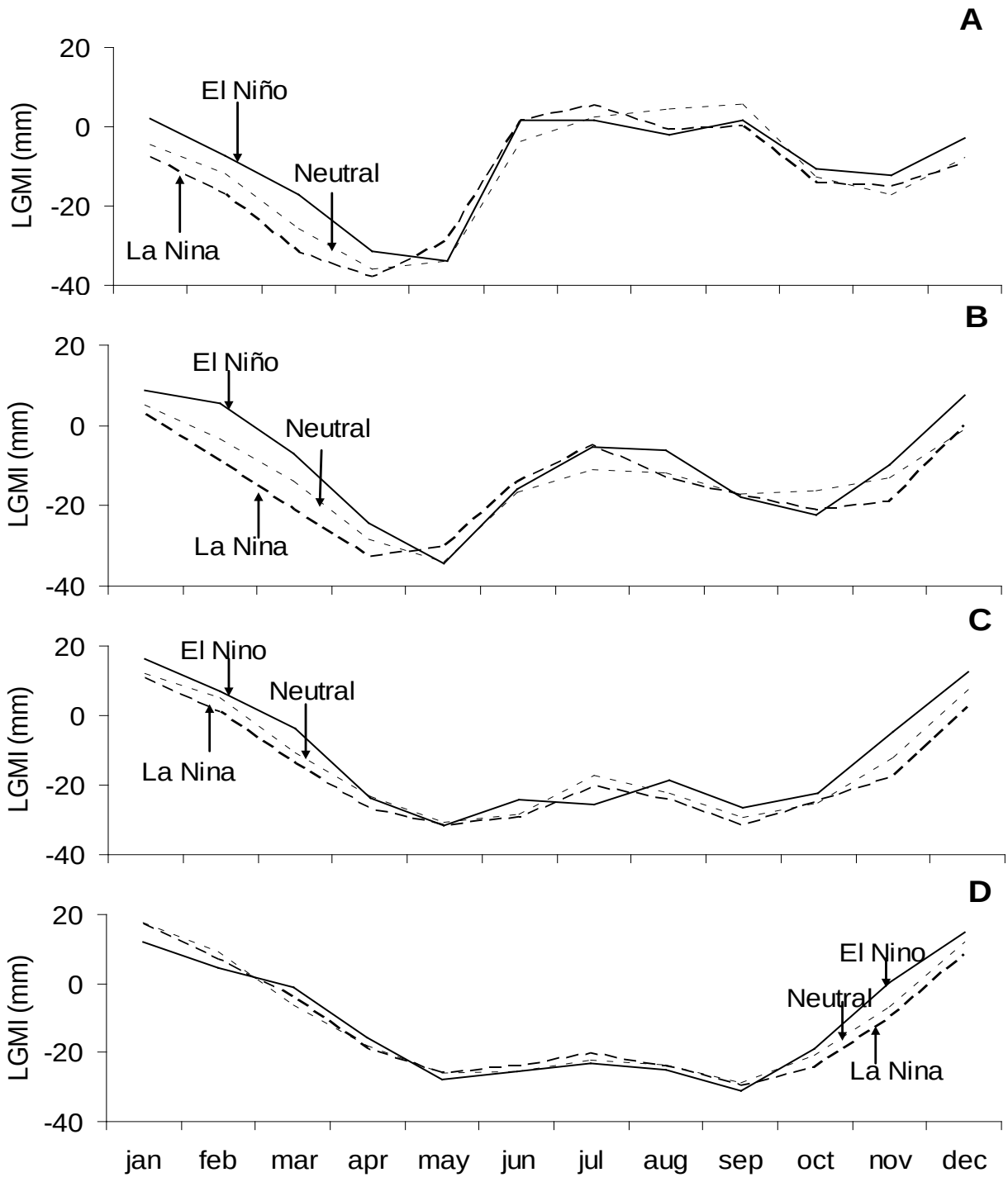

Figure A-8. Mean monthly LGMI values for different ENSO phases in four locations: A) Lee, B) Sumter, C) Columbia, and D) Floyd

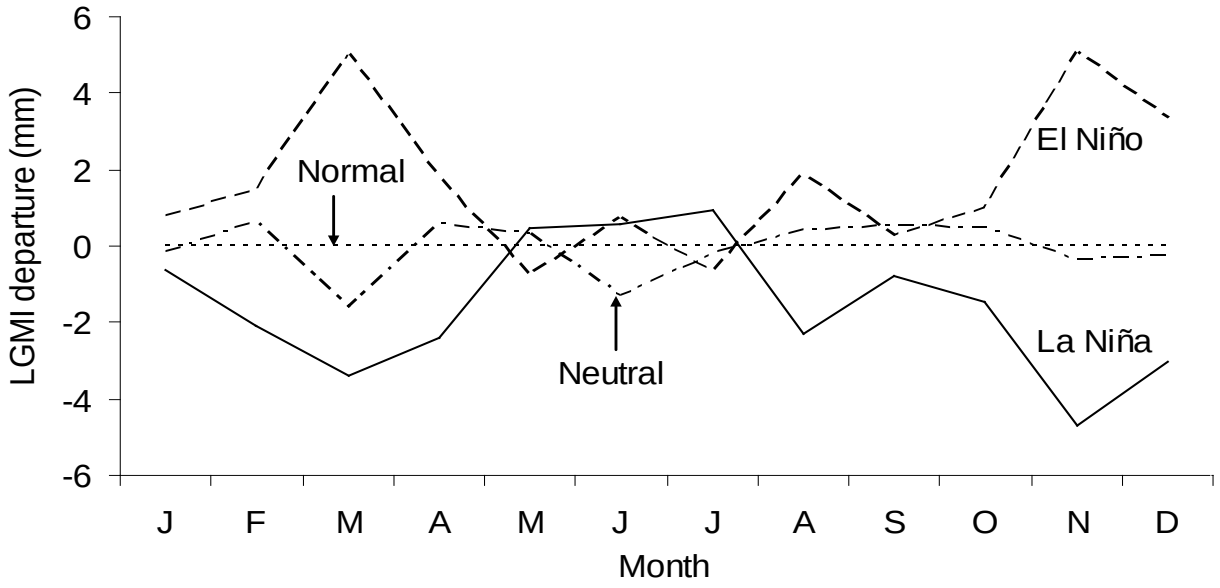

Figure A-9. Departure of monthly precipitation in ENSO phases from normal - the mean monthly precipitation of 112 years

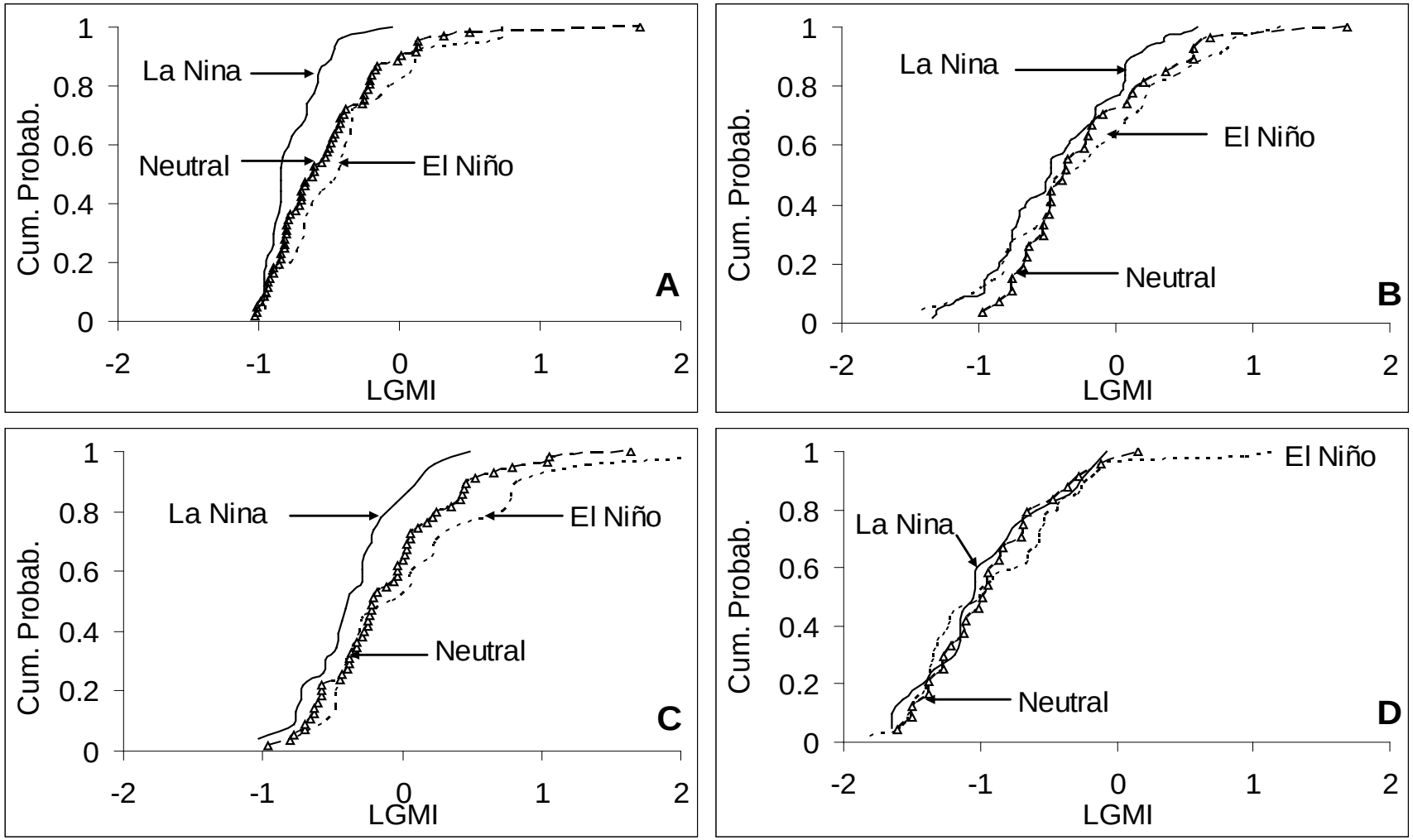

Figure A-10. Probability of water deficit (LGMI < 0) in: A) Columbia in November, B) Columbia in July, C) Cullman in November, and D) Cullman in July. Cumulative probability values are based on monthly average LGMI.

APPENDIX B
FULLY-FUNCTIONAL MATLAB PROGRAM FOR COMPUTING ARID

% Variables Defined

% arid the ARID index
% ARID arid transposed
% CN runoff curve number
% CSR clear-sky radiation
% CWBD change in water before drainage
% DC drainage coefficient
% DR drainage
% ea ambient vapor pressure
% es saturation vapor pressure
% ETo reference ET
% extra extra-terrestrial radiation
% FC water content at field capacity
% IA initial abstraction
% IRDES Inverse relative distance Earth-Sun
% IW initial water content
% lat latitude in radiation
% LWR long-wave radiation
% MUF water uptake coefficient
% NRAD net radiation
% PSC psychrometric constant
% RO runoff
% RRAD relative radiation
% S maximum abstraction
% SD solar declination
% slope slope of vapor pressure curve
% SSA sun-set hour angle
% SWR short-wave radiation
% TR transpiration
% WAD water after drainage
% WAT water after transpiration
% W_AT water after transpiration (temporary)
% WBD water before drainage
% WP water content at wilting point
% ws2 wind speed at 2 m
% Z rootzone depth
%{

Inputs

This program uses a dataset that has 7 columns – column 1: solar radiation (MJ m^2 d^{-1}), column 2: maximum temperature (°C), column 3: minimum temperature (°C), column 4: dewpoint temperature (°C), column 5: precipitation (mm), column 6: windspeed at 10 m height (m s^{-1}), and column 7: day-of-year (number). If your dataset has different columns, change the column index of a variable accordingly.
%}

```
load data.dat;                  % Load data from a file named data.dat
rain = data(:,5);               % the 5th column of the data is precipitation
latitude = 28.102;              % latitude of the weather station (degree)
elevation = 53.0;               % elevation of the weather station (m)
%
% Parameters                    % the parameter values are the default values
DC = 0.55;
Z = 400;
CN = 65;
MUF = 0.096;
WP = 0.06;
FC = WP+0.13;
IW = FC;
S = 25400/CN-254;
IA = 0.2*S;
lat = latitude*pi/180;
PSC = 0.665*10^-3*101.3*((293-0.0065*elevation)/293)^5.26;
%
%ARID Computation (computes a vector of ARID using the dataset)
for i=1:length(data)
        % First, compute ETo (Allen et al., 1998).
        if mod(data(i,7),4)==0
                days = 366;
        else
                days = 365;
        end
        ws2 = data(i,6)*4.87/log(67.8*10-5.42);
        es = ((0.6108*exp(17.27*data(i,2)/(data(i,2)+237.3)))+...
                (0.6108*exp(17.27*data(i,3)/(data(i,3)+237.3))))/2;
        slope = (0.6108*exp(17.27*((data(i,2)+data(i,3))/2)/(((data(i,2)+...
                data(i,3))/2)+237.3))*4098)/((data(i,2)+data(i,3))/2+237.3)^2;
        SWR = (1-0.23)*data(i,1);
        IRDES = 1+0.033*cos(2*pi*data(i,7)/days);
        SD = 0.409*sin(2*pi*data(i,7)/days-1.39);
        SSA = acos(-tan(lat)*tan(SD));
        extra = 24*60*0.082/pi*IRDES*(SSA*sin(lat)*sin(SD)+...
                cos(lat)*cos(SD)*sin(SSA));
        CSR = (0.75+2*10^-5*elevation)*extra;
        RRAD = data(i,1)/CSR;
        ea = 0.6108*exp(17.27*data(i,4)/(data(i,4)+237.3));
        LWR = 4.903*10^-9*((data(i,2)+273.16)^4+(data(i,3)+273.16)^4)/2*...
                (0.34-0.14*sqrt(ea))*(1.35*RRAD-0.35);
        NRAD = SWR-LWR;
        ETo = (0.408*slope*NRAD+PSC*(900/((data(i,2)+data(i,3))/2+273))*ws2*...
                (es-ea))/(slope+PSC*(1+0.34*ws2));
        %
```

```
        % Then, compute ARID.
        if rain(i)>IA
                RO = (rain(i)-0.2*S)^2/(rain(i)+0.8*S);
        else
                RO = 0;
        end
        CWBD = rain(i)- RO;
        if i==1
                W_AT = Z*IW;
        else
                W_AT = WAT(i-1);
        end
        WBD(i) = CWBD + W_AT;
        if WBD(i)/Z > FC
                DR = Z*DC*(WBD(i)/Z - FC);
        else
                DR = 0;
        end
        WAD = WBD(i) - DR;
        TR = min(MUF*Z*(WAD/Z-WP),ETo);
        WAT(i) = WAD - TR;
        arid(i) = 1-TR/ETo;
end
ARID = arid';
```

LIST OF REFERENCES

Abbasi, B. 2009. A neural network applied to estimate process capability of non-normal processes. *Expert Systems with Applications* 36: 3093-3100.

Acevedo, E., T. C. Hsiao, and D. W. Henderson. 1971. Immediate and subsequent growth responses of maize leaves to changes in water status. *Plant Physiology* 48: 631–636.

Aggarwal, P. K. 1995. Uncertainties in crop, soil, and weather inputs used in growth models: implications for simulated outputs and their applications. *Agricultural Systems* 48: 361-384.

Ahrens, D. C. 2003. *Meteorology Today an Introduction to Weather, Climate, and the Environment*. 7th ed. Pacific Grove, California: Brooks/Cole-Thomson Learning.

Akaike, H. 1969. Fitting autoregressive model for prediction. *Ann. Inst. Stat. Math*. 21: 243-247.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration – guidelines for computing crop water requirements. Irrigation and Drainage Paper No. 56. Rome, Italy: Food and Agriculture Organization.

Allen, R. G., I. A. Walter, R. Elliott, and T. Howell. 2005. The ASCE standardized reference evapotranspiration equation. Reston, Virginia: American Society of Civil Engineers.

Altun, S., A. B. Goktepe, A. M. Ansal, and C. Akguner. 2009. Simulation of torsional shear test results with neuron-fuzzy control system. *Soil Dynamics and Earthquake Engineering* 29: 253-260.

AMS. 1997. Meteorological drought – policy statement. *BAMS* 78(5): 847-849.

Arkley, R. J. 1963. Relationships between plant growth and transpiration. *Hilgardia* 34(13): 559-584.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large area hydrologic modeling and assessment part I: model development. *Journal of American Water Resources Association* 34(1): 73-89.

Arora, V. K., S. S. Prihar, and P. R. Gajri. 1987. Synthesis of simplified water use simulation model for predicting wheat yields. *Water Resources Research* 23(5): 903-910.

Arora, V. K., and P. R. Gajri. 1996. Performance of simplified water balance models under maize in a semi-arid subtropical environment. *Agricultural Water Management* 31: 51-64.

Asadi, M. E., and R. S. Clemente. 2003. Evaluation of CERES-Maize of DSSAT model to simulate nitrate leaching, yield, and soil moisture content under tropical conditions. *Food, Agriculture & Environment* 1(3&4): 270-276.

ASCE. 1993. Criteria for evaluation of watershed models. *Journal of Irrigation and Drainage Eng.* 119(3): 429-442.

ASCE. 2000. Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering* 5(2): 124-137.

Bacanli, U. G., M. Firat, and F. Dikbas. 2008. Adaptive neuron-fuzzy inference system for drought forecasting. *Stochastic Environmental Research and Risk Assessment* 23: 1143-1154.

Bauske, E. M., P. L. Mask, K. Harker, C. D. Monks, and J. Kemble. 1996. Delayed planting may help corn growth. *Highlights of Agricultural Research* 43(2).

Bellman, R. E. 1957. *Dynamic Programming*. Princeton, N.J.: Princeton University Press.

Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression diagnostics: Identifying influential data and sources of collinearity*. New York, N.Y.: Wiley-Interscience.

Bond, P., R. Dada, and G. Erion. 2007. *Climate change, carbon trading and civil society: Negative returns on South African investments*. Amsterdam, The Netherlands: Rozenberg Publishers.

Boote, K. J., and D. L. Ketring. 1990. Peanut. In *Irrigation of agricultural crops*, 675-717. B. A. Stewart and D. R. Nielsen, eds. Madison, Wisconsin: ASA-CSSA-SSSA.

Bordi, I., K. Fraedrich, M. Petitta, and A. Sutera. 2005. Methods for predicting drought occurrences. In *Proc. of the 6th international conference of the European water resources association*. Menton, France: European Water Resources Association.

Borri, D., G. Concilio, and E. Conte. 1998. A fuzzy approach for modeling knowledge in environmental systems evaluation. *Computers, Environment, and Urban Systems* 22: 299-313.

Bouwer, L. M., T. W. Biggs, and J. C. J. H. Aerts. 2008. Estimates of spatial variation in evaporation using satellite-derived surface temperature and a water balance model. *Hydrological Processes* 22: 670-682.

Box, G. E. P., and G. M. Jenkins. 1970. *Time series analysis, forecasting, and control*. Oakland, California: Holden-Day, Inc.

Briggs, L. J., and H. L. Shantz. 1913a. The water requirement of plants. I. Investigations in the Great Plains in 1910 and 1911. USDA Bureau of Plant Industry Bulletin No. 284. Washington, D.C.: U.S. Department of Agriculture.

Brolley, J. M., J. J. O'Brien, J. Schoof, and D. Zierden. 2007. Experimental drought threat forecast for Florida. *Agricultural and Forest Meteorology* 145: 84-96.

Brown, E. A., C. E. Caviness, and D. A. Brown. 1985. Response of selected soybean cultivars to soil moisture deficit. *Agronomy Journal* 77: 274-278.

Budyko, M. I. 1958. *The Heat Balance of the Earth's Surface*. Washington, D.C.: U.S. Department of Commerce.

Campbell, C. O. 1973. The future role of agriculture in the Australian economy. In *The environmental, economic, and social significance of drought*. J. V. Lovett, eds. Sidney, Australia: Angus and Robertson.

Cancelliere, A., G. D. Mauro, B. Bonaccorso, and G. Rossi. 2007. Drought forecasting using the Standardized Precipitation Index. *Water Resources Management* 21: 801-819.

Chakraborty, K., K. Mehrotra, C, K. Mohan, and S. Ranka. 1992. Forecasting the behavior of multivariate time series using neural networks. *Neural Networks* 5: 961-970.

Chan, K., A. Saltelli, and S. Tarantola. 1997. Sensitivity analysis of model output: variance-based methods make the difference. Proceedings of the 1997 Winter Simulation Conference, 261-268.

Chan, K., A. Saltelli, and S. Tarantola. 2000a. Winding stairs: a sampling tool to compute sensitivity indices. *Statistics and Sampling* 10: 187-196.

Chan, K., S. Tarantola, A. Saltelli, and I. M. Sobol. 2000b. Variance-based methods. In *Sensitivity analysis, Wiley Series in Probability and Statistics*, 167-197. A. Saltelli, K. Chan, and E. M. Ecott, eds. West Sussex, England: John Wiley & Sons Ltd.

Chang, F. J., and Y. T. Chang. 2006. Adaptive neuron-fuzzy inference system for prediction of water level in reservoir. *Advances in Water Resources* 29: 1-10.

Chen, Y. Q., D. S. Battisti, T. N. Palmer, J. Barsugli, and E. S. Sarachick. 1997. A study of the predictability of tropical Pacific SST in a coupled atmosphere-ocean model using singular vector analysis: the role of the annual cycle and the ENSO cycle. *Monthly Weather Review* 125: 831-845.

Chiew, F. H. S., T. C. Piechota, J. A. Dracup, and T. A. McMahon. 1998. El Nino/Southern Oscillation and Australian rainfall, stream flow, and drought: links and potential for forecasting. *Journal of Hydrology* 204: 138-149.

Christy, J. 2004. The lawn-and-garden moisture index. Huntsville, Alabama: Alabama Office of the State Climatologist, National Space Science and Technology Center, University of Alabama. Available at: nsstc.uah.edu/aosc/lawn_garden2.htm. Accessed 27 January 2009.

Cole, J. S., and O. R. Mathews. 1923. Use of water by spring wheat on the Great Plains. USDA Bureau of Plant Industry Bulletin No. 1004. Washington, D.C.: U.S. Department of Agriculture.

Cordery, I., and M. McCall. 2000. A model for forecasting drought from teleconnections. *Water Resources Research* 36(3): 763-768.

Costa Branco, P. J., and J. A. Dente. 2001. Fuzzy systems modeling in practice. *Fuzzy Sets Systems* 121: 73-93.

Cothren, J. T. 1999. Physiology of the cotton plant. In *Cotton: Origin, History, Technology, and Production*, 207-268. W. C. Smith and J. T. Cothren, eds. New York, N.Y.: John Wiley & Sons, Inc.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petshek, and J. H. Schaibly. 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics* 59(8): 3873-3878.

Cukier, R. I., J. H. Schaibly, and K. E. Shuler. 1975. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations. *The Journal of Chemical Physics* 63(3): 1140-1149.

Cukier, R. I., H. B. Levine, and K. E. Shuler. 1978. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics* 26: 1-42.

Cybenko, G. 1989. Approximation by superposition of a sigmoidal function. *Mathematical Control Signals Systems* 2: 303-314.

Dardanelli, J. L., O. A. Bachmeier, R. Sereno, and R. Gill. 1997. Rooting depth and soil water extraction patterns of different crops in a silty loam Haplustoll. *Field Crops Research* 54: 29-38.

Dardanelli, J. L., J. T. Ritchie, M. Calmon, J. M. Andriani, and D. J. Collino. 2004. An empirical model for root water uptake. *Field Crops Research* 87: 59-71.

D'Arrigo, R., and R. Wilson. 2008. Short communication El Nino and Indian Ocean influences on Indonesian drought: implications for forecasting rainfall and crop productivity. *International Journal of Climatology* 28: 611-616.

De Bruyn, L. P. 1964. Influence of soil moisture deficit on growth and production of cotton. *Abs. Emp. Cotton Grow. Rev.* 43: 69.

De Faria, R. T., and W. T. Bowen. 2003. Evaluation of DSSAT soil-water balance module under cropped and bare soil conditions. *Brazilian Archives of Biology and Technology* 46(4): 489-498.

De Groot, C., and D. Wurtz. 1991. Analysis of univariate time series with connectionist nets: a case study of two classical examples. *Neurocomputing* 3: 177-192.

De Jong, R., and A. Bootsma. 1996. Review of recent developments in soil water simulation models. *Canadian Journal of Soil Science* 76: 263-273.

De Jong, R., and P. Kabat. 1990. Modeling water balance and grass production. *Soil Science Society of America Journal* 54: 1725-1732.

de Wit, C. T. 1958. Transpiration and crop yields. Versl Landbouwk Onderz No. 64.6. Wageningen, the Netherlands: Institute for Biological and Chemical Research on Field Crops and Herbage.

De Wit, C. T. 1967. Photosynthesis: Its relationship to overpopulation. In *Harvesting the sun*, 315-320. A. San Pietro et al., ed. New York, N.Y.: Academic Press.

Dolling, K., P. S. Chu, and F. Fujioka. 2005. A climatological study of the Keetch/Byram drought index and fire activity in the Hawaiian Islands. *Agricultural and Forest Meteorology* 133: 17-27.

Doorenbos, J., and A. H. Kassam. 1979. Yield response to water. Irrigation and Drainage Paper No. 33. Rome, Italy: Food and Agriculture Organization.

Doss, B. D., R. W. Pearson, and H. T. Rogers. 1974. Effect of soil water stress at various growth stages on soybean yield. *Agronomy Journal* 66: 297-299.

Dracup, J. A., K. S. Lee, and E. D. Paulson Jr. 1980. On the definition of Droughts. *Water Resources Research* 16(2): 297-302.

Eilers, V. H. M., R. C. Carter, and K. R. Rushton. 2007. A single layer soil water balance model for estimating deep drainage (potential recharge): An application to cropped land in semi-arid North-east Nigeria. *Geoderma* 140: 119-131.

Ekstrom, P-A. 2005. Eikos – a simulation toolbox for sensitivity analysis. MS thesis. Uppsala, Sweden: Uppsala University, Faculty of Science and Technology.

Enfield, D. B., A. M. Mestas-Nunez, and P. J. Trimble. 2001. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters* 28(10): 2077-2080.

FAO. 2009. AquaCrop. Rome, Italy: Food and Agriculture Organization. Available at: www.fao.org/nr/water/infores_databases_aquacrop.html. Accessed 24 December 2009.

FEMA. 1995. National mitigation strategy: partnerships for building safer communities. Washington, D.C.: Federal Emergency Management Agency.

Firat, M., and M. Gungor. 2008. Hydrological time-series modeling using adaptive neuron-fuzzy inference system. *Hydrological processes* 22(13): 2122-2132.

Fontenot, R. L. 1999. An evaluation of reference evapotranspiration models in Louisiana. MS thesis. Baton Rouge, Louisiana: Louisiana State University, A & M College.

Forrester, J. W. 1970. Systems analysis as a tool for urban planning. *IEEE Trans. Syst. Sci. Cybernetics* 6: 259-265.

Frere, M., and G. Popov. 1986. Early agrometeorological crop yield assessment. Plant Production and Protection Paper No. 73. Rome, Italy: Food and Agricultural Organization.

Furukawa, T., K. Tokunaga, S. Yasui, H. Tamukoh, K. Ishii, M. Ishikawa, K. Horio, and K. Natsume. 2007. Modular network self-organizing map: Can it be an artificial cortex? *International Congress Series* 1301: 43-47.

Gershunov, A., and T. P. Barnett. 1998a. ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *Journal of Climate* 11: 1575-1586.

Gershunov, A., and T. P. Barnett. 1998b. Interdecadal modulation of ENSO teleconnections. *BAMS* 79: 2715-2725.

Gibbs, W. J., and J. V. Maher. 1967. Rainfall Deciles as Drought Indicators. Bulletin No. 48. Melbourne, Australia: Australian Bureau of Meteorology.

Glantz, M., R. Katz, and N. Nicholls. 1991. *Teleconnections Linking Worldwide Climate Anomalies*. Cambridge, United Kingdom: Cambridge University Press.

Gommes, R., and F. Petrassi. 1994. Rainfall variability and drought in sub-Saharan Africa since 1960. Agrometeorology Series Working Paper No. 9. Rome, Italy: Food and Agriculture Organization.

Granger, C. W. J. 1993. Strategies for modeling nonlinear time series relationships. *The Economic Record* 69(206): 233-238.

Granger, C. W. J., and T. Terasvirta. 1993. *Modeling nonlinear economic relationships*. Oxford, United Kingdom: Oxford University Press.

Green, P. M., D. M. Leager, C. J. Miranda, J. J. O'Brien. 1997. The North American climate patterns associated with the El Niño-Southern Oscillation. COAPS Report Series 97-1. Tallahassee, Florida: Center for Ocean-Atmosphere Prediction Studies, Florida State University. Available at: http://www.coaps.fsu.edu/lib/booklet/. Assessed on: October 16, 2009.

Grimes, D. W., and K. M. El-Zik. 1990. Cotton. In *Irrigation of agricultural crops*, 741-773. B. A. Stewart and D. R. Nielsen, eds. Madison, Wisconsin: ASA-CSSA-SSSA.

Grimes, D. W., R. J. Miller, and L. Dickens. 1970. Water stress during flowering of cotton. *California Agriculture* 24(3): 4-6.

Hagan, M. T., H. P. Demuth, and M. Beale. 1996. *Neural network design*. Boston, Massachusetts: PWS Publishing.

Hagemeyer, B. C. 2006. ENSO, PNA and NAO Scenarios for extreme storminess, rainfall and temperature variability during the Florida dry season. In *Preprints of the 18th Conference on Climate Variability and Change*, CD-ROM P2.4. Atlanta, Georgia: American Meteorological Society. Available at: http://www.srh.noaa.gov/mlb/research.html. Accessed 24 September 2009.

Haman, D. Z., G. A. Clark, and A. G. Smajstrla. 2008. Irrigation of lawns and gardens. Gainesville, Florida: Electronic Data Information Source – UF/IFAS Extension. Available at: edis.ifas.ufl.edu/WI003. Accessed 27 January 2009.

Hanks, R. J. 1974. Model for predicting plant yield as influenced by water use. *Agronomy Journal* 66(5): 660-665.

Hanley, D. E., M. A. Bourassa, J. J. O'Brien, S. R. Smith, and E. R. Spade. 2003. A quantitative evaluation of ENSO indices. *Journal of Climate* 16: 1249-1258.

Hansen, J. W., A. W. Hodges, and J. W. Jones. 1998. ENSO influences on agriculture in the southeastern United States. *Journal of Climate* 11: 404-411.

Hastenrath, S., and L. Greischar. 1993. Further work on the prediction of Northeast Brazil rainfall anomalies. *Journal of Climate* 6: 743-758.

Hayes, M. J. 2006. What is drought? Drought Indices. Lincoln, Nebraska: University of Nebraska, National Drought Mitigation Center. Available at: www.drought.unl.edu/whatis/indices.htm. Accessed 27 January 2009.

Hayes, M. J., M. D. Svoboda, D. A. Wilhite, and O. V. Vanyarkho. 1999. Monitoring the 1996 drought using the standard precipitation index. *BAMS* 80: 429-440.

Schmidt, N., E. K. Lipp, J. B. Rose, and M. E. Luther. 2001. ENSO Influences on Seasonal Rainfall and River Discharge in Florida. *Journal of Climate* 14: 615-628.

He, J. 2008. Best management practice development with the CERES-Maize model for sweet corn production in north Florida. PhD diss. Gainesville, Florida: University of Florida, Department of Agricultural and Biological Engineering.

Heim, R. R. 2002. A review of twentieth-century drought indices used in the United States. BAMS 83(8): 1149-1166.

Hiler, E. A., and R. N. Clark. 1971. Stress day index to characterize effects of water stress on crop yields. *Trans. ASAE* 14(4): 757-761.

Hiler, E. A., T. A. Howell, R. B. Lewis, and R. P. Boos. 1974. Irrigation timing by the stress day index method. *Trans. ASAE* 65: 393-398.

Hisdal, H., and L. M. Tallaksen. 2000. Drought event definition. Assessment of the regional impact of droughts in Europe, Technical Report No. 6. Oslo, Norway: University of Oslo, Department of Geophysics.

Hodges, T., D. Botner, C. Sakamoto, and H. J. Hays. 1987. Using the CERES-Maize model to estimate production for the U.S. Corn Belt. *Agricultural and Forest Meteorology* 40: 293-303.

Homma, T., and A. Saltelli. 1996. Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety* 52(1): 1-17.

Hoogenboom, G., J. W. Jones, P. W. Wilkens, C. H. Porter, W. D. Batchelor, L. A. Hunt, K. J. Boote, U. Singh, O. Uryasev, W. T. Bowen, A. J. Gijsman, A. S. Du Toit, J. W. White, and G. Y. Tsuji. 2004. *Decision Support System for Agrotechnology Transfer*. ver. 4.0. Honolulu, Hawaii: University of Hawaii.

Hopkins, W. G. 1999. Water relations of the whole plant. In *Introduction to Plant Physiology*, 7-59. W. G. Hopkins, ed. New York, N.Y.: John Wiley and Sons, Inc.

Hora, S. C., and R. L. Iman. 1986. A comparison of maximum/bounding and Bayesian/Monte Carlo for fault tree uncertainty analysis. Technical Report No. SAND85-2839. Albuquerque, New Mexico: Sandia National Laboratories.

Hornik, K., M. Stichcombe, and H. White. 1989. Multi-layer feed-forward networks are universal approximators. *Neural Networks* 2: 359-366.

Howell, T. A. 1990. Relationships between crop production and transpiration, evapotranspiration, and irrigation. In *Irrigation of agricultural crops*, 391-434. B. A. Stewart and D. R. Nielsen, eds. Madison, Wisconsin: ASA-CSSA-SSSA.

Idso, S. B., R. D. Jackson, P. J. Pinter Jr., R. J. Reginato, and J. L. Hatfield. 1981a. Normalizing the stress-degree day parameter for environmental variability. *Agricultural Meteorology* 24: 45-55.

Iman, R. L., J. C. Helton, and J. E. Campbell. 1981. An approach to sensitivity analysis of computer models, Part 1. Introduction, input variable selection, and preliminary variable assessment. *Journal of Quality Technology* 13(3): 174-183.

Iman, R. L., and S. C. Hora. 1990. A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Analysis* 10: 401-406.

Ingram, K. T. 2005. Drought-related characteristics of important cereal crops. In *Monitoring and Predicting Agricultural Drought – A Global Study*, 11-27. V. K. Boken, A. P. Cracknell, and R. L. Heathcote, eds. New York, N.Y.: Oxford University Press, Inc.

Ishigami, T., and T. Homma. 1989. An importance quantification technique in uncertainty analysis for computer models. Report No. JAERI-M 89-111. Tokai-Mura, Japan: Japan Atomic Energy Research Institute.

Jalali-Farahani, H. R., D. C. Slack, D. M. Kopec, and A. D. Matthias. 1993. Crop water stress index models for Bermudagrass turf: a comparison. *Agronomy Journal* 85: 1210-1217.

Jang, J. R. 1993. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* 23: 665-685.

Jansen, M. J. W., W. A. H. Rossing, and R. A. Daamen. 1994. Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In *Predictability and non-linear modeling in natural sciences and economics*, 334-343. J. Gasman, and G. van Straten, eds. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Jensen, M. E. 1968. Chapter 1: Water consumption by agricultural plants. In *Water Deficits and Plant Growth*, Vol. II, 1–22. T. T. Kozlowski, ed. New York, N.Y.: Academic Press Inc.

Jensen, M. E., R. D. Burman, and R. G. Allen. 1990. Evapotranspiration and Irrigation water requirements. ASCE Manuals and Reports on Engineering Practice No. 70. New York, N.Y.: American Society of Civil Engineers.

Johnson, M. B., and G. Forthum. 2001. Spatial mapping of KBDI for the southeast United States. In *Proc. 4th Symposium on Fire and Forest Meteorology*, 64-65. Reno, Nevada: American Meteorological Society.

Jones, C. A., and J. R. Kiniry. 1986. CERES-Maize – A simulation model of maize growth and development. College Station, Texas: Texas Agricultural and Mechanical University Press.

Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. 2003. The DSSAT cropping system model. *European Journal of Agronomy* 18: 235-265.

Kang, S. 1991. An investigation of the use of feedforward neural networks for forecasting. PhD diss. Kent, Ohio: Kent State University.

Keating, B. A., H. Meinke, and J. P. Dimes. 1996. Prospects for using a cropping system simulator to assess exceptional drought. Consultancy report. Canberra, Australia: Bureau of Resource Sciences, Department of Primary Industries and Energy.

Keetch, J. J., and G. M. Byram. 1968. A drought index for forest fire control. USDA Forest Service Research Paper SE – 38. Asheville, North Carolina: U.S. Department of Agriculture - Forest Service.

Kelly, S. 2007. Statistical associations between large scale climate oscillations and mesoscale surface meteorological variability in the Apalachicola-Chattahoochee-Flint river basin. MS thesis. Tallahassee, Florida: Florida State University, Department of Meteorology.

Keyantash, J., and J. A. Dracup. 2002. The quantification of drought: an evaluation of drought indices. *BAMS* 1167-1180.

Kim, T., and J. B. Valdes. 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *Journal of Hydrologic Engineering* 8(6): 319-328.

Kipkorir, E. C., and D. Raes. 2002: Transformation of yield response factor into Jensen's sensitivity index. *Irrigation and Drainage Systems* 16: 47-52.

Klimasauskas, C. C. 1991. Applying neural networks Part 3: Training a neural network. *PCAI* 5(3): 20-24.

Koda, M. G. J. McRae, and J. H. Seinfeld. 1979. Automatic sensitivity analysis of kinetic mechanisms. *International Journal of Chemical Kinetics* 11: 427-444.

Kogan, F. N. 1990: Remote sensing of weather impacts on vegetation in non-homogenous areas. *International Journal of Remote Sensing* 11: 1405-1419.

Konrad, C. E. 1998. Intramonthly indices of the Pacific/North American teleconnection pattern and temperature regimes over the United States. *Theoretical and Applied Climatology* 60: 11-19.

Krzysztofowicz, R. 2001. The case for probabilistic forecasting in hydrology. *Journal of Hydrology* 249: 2-9.

Kumar, V., and U. Panu. 1997. Predictive assessment of severity of agricultural droughts based on agro-climatic factors. *Journal of the American Water Resources Association* 33(6): 1255-1264.

Leathers, D. J., B. Yarnal, and M. A. Palecki. 1991. The Pacific/North American pattern and United States climate. *Journal of Climate* 4(5): 517-528.

Lee, D. 2007. A guide to corn production in Georgia 2007. Athens, Georgia: University of Georgia, College of Agriculture and Environmental Sciences, Cooperative Extension Service. Available at: www.caes.uga.edu/commodities/fieldcrops/gagrains/pdf/2007cornproduction.pdf. Accessed 27 January 2009.

Lee, S. H., R. J. Howlett, C. Crua, and S. D. Walters. 2007. Fuzzy logic and neuron-fuzzy modeling of diesel spray penetration: a comparative study. *Journal of Intelligent and Fuzzy Systems: Applications in Engineering and Technology* 18(1): 43-56.

Legates, D. R., and G. J. McCabe. 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research* 35(1): 233-241.

Legowo, E. 1986. Estimation of water extractability and hydraulic conductivity in tropical mollisols, ultisols, and andisols. PhD diss. Honolulu, Hawaii: University of Hawaii, Department of Soil Science.

Li, K. Y., R. D. Jong, and J. B. Boisvert. 2001. An exponential root-water-uptake model with water stress compensation. *Journal of Hydrology* 252: 89-204.

Li, K.Y., R. D. Jong, M. T. Coe, and N. Ramankutty. 2006. Root water uptake based upon a new water stress reduction and an asymptotic root distribution function. *Earth Interactions* 10(14).

Lohani, V. K., G. V. Loganathan, and S. Mostaghimi. 1998. Long-term analysis and short-term forecasting of dry spells by Palmer Severity Drought Index. *Nord Hydrology* 29(1): 21-40.

Lopez-Cedron, F. X., K. J. Boote, B. Ruiz-Nogueira, and F. Sau. 2005. Testing CERES-Maize versions to estimate maize production in a cool environment. *European J. Agronomy* 23: 89-102.

Lourens, U. W. 1995. A system for drought monitoring and severity assessment, PhD diss. Bloemfontein, South Africa: University of Orange Free State.

Lourens, U. W., and J. M. de Jager. 1997. A computerized crop-specific drought monitoring system: design concepts and initial testing. *Agricultural Systems* 53: 303-315.

Maier, H. R., and G. C. Dandy. 1998. The effect of internal parameters and geometry on the performance of back propagation neural networks: an empirical study. *Environmental Modeling & Software* 13(2): 193-209.

Maier, H. R., and G. C. Dandy. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environmental Modeling & Software* 15: 101-124.

Makowski, D. C. Naud, M. H. Jeuffroy, A. Barbottin, and H. Monod. 2006. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model prediction. *Reliability Engineering and System Safety* 91: 1142-1147.

Mamdani, E. H., and S. Assilian. 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies* 7(1): 1-13.

Maracchi, G. 2000. Agricultural drought – a practical approach to definition, assessment and mitigation strategies. In *Drought and Drought Mitigation in Europe*, 64-75. J. V. Vogt and F. Somma, eds. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Marce, R., M. Comerma, J. C. Garcia, and J. Armengo. 2004. A neuron-fuzzy modeling tool to estimate fluvial nutrient loads in watersheds under time-varying human impact. *Limnology and Oceanography Methods* 2: 342-355.

Marja, M., and M. Esa. 1997. Neuro-fuzzy modelling of spectroscopic data. Part A – Modelling of dye solutions. *Journal of the Society of Dyers and Colourists* 113(1): 13-17.

Martinez, C. J., G. A. Baigorria, and J. W. Jones. 2009. Use of climate indices to predict corn yields in southeast USA. *International Journal of Climatology* 29: 1680-1691.

Mason, S. J. 2002. *Analyzing and communicating forecast quality and probability* (Asian Climate Training Manual, Module 3.3). Pathumthani, Thailand: Extreme Climate Events Program, Asian Disaster Preparedness Center. Available at: http://www.adpc.net/ece/ACT_man/ACT-3.3-ForecastQuality.pdf. Accessed 11 July 2010.

Mathews, O. R., and L. A. Brown. 1938. Winter wheat and sorghum in the southern Great Plains under limited rainfall. USDA Circular No. 477. Washington, D.C.: U.S. Department of Agriculture.

MathWorks. 2009. *Fuzzy Logic Toolbox™ 2 User's Guide*. Natick, Massachusetts: The MathWorks, Inc. Available at: http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/fuzzy.pdf. Accessed 6 October 2009.

Maunder, W. J. 1992. *Dictionary of Global Climate Change*. New York, N.Y.: Chapman and Hall.

Mavromatis, M., K. J. Boote, J. W. Jones, A. Irmak, D. Shinde, and G. Hoogenboom. 2001. Developing genetic coefficients for crop simulation models with data from crop performance trials. *Crop Science* 41: 40-51.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2): 239-245.

McKay, M. D. 1995. Evaluating prediction uncertainty. Technical Report NUREG/CR-6311 LA-12915-MS. Washington, D.C: U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research.

McKee, T. B., N. J. Doesken, and J. Kleist. 1993. The relationship of drought frequency and duration to time scales. In *Preprints of the 8th Conference on Applied Climatology*. Anaheim, California: American Meteorological Society.

McRae, G. J., J. W. Tilden, and J. H. Seinfeld. 1982. Global sensitivity analysis – a computational implementation of the Fourier amplitude sensitivity test (FAST). *Computers and Chemical Engineering* 6(1): 15-25.

Meinke, H., G. L. Hammer, and P. Want. 1993. Potential soil water extraction by sunflower on a range of soils. *Field Crops Research* 32: 59-81.

Mendicino, G., A. Senatore, and P. Versace. 2008. A Groundwater Resource Index (GRI) for drought monitoring and forecasting in a Mediterranean climate. *Journal of Hydrology* 357: 282-302.

Meyer, S. J., K. G. Hubbard, and D. A. Wilhite. 1993a. A crop specific drought index for corn: I. Model development and validation. *Agronomy Journal* 85: 396-399.

Mika, J., S. Horvath, L. Makra, and Z. Dunkel. 2005. The Palmer Drought Severity Index (PDSI) as an indicator of soil moisture. *Physics and Chemistry of the Earth* 30: 223-230.

Mishra, A. K., and V. R. Desai. 2005. Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment* 19: 326-339.

Mishra, A. K., and V. R. Desai. 2006. Drought forecasting using feed-forward recursive neural network. *Ecological Modeling* 198: 127-138.

Mishra, A. K., V. R. Desai, and V. P. Singh. 2007. Drought forecasting using a hybrid stochastic and neural network model. *Journal of hydrological Engineering* 12(6): 626-638.

Modarres, R. 2006. Streamflow drought time series forecasting. Stochastic *Environmental Research and Risk Assessment* 21: 223-233.

Monod, H., C. Naud, and D. Makowski. 2006. Uncertainty and sensitivity analysis for crop models. In *Working with dynamic crop models – evaluation, analysis, parameterization, and applications*, 55-96. D. Wallach, D. Makowski, and J. W. Jones, eds. Amsterdam, The Netherlands: Elsevier.

Moran, M. S. 1994. Irrigation management in Arizona using satellites and airplanes. *Irrigation Science* 15: 35-44.

Moran, M. S., Y. Inoue, and E. M. Barnes. 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing Environment* 61: 319-346.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50(3): 885-900.

Morris, M. D. 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33: 161-174.

Muleta, M. K., and J. W. Nicklow. 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology* 306: 127-145.

Musick, J. T., and D. A. Dusek. 1980. Irrigated corn yield response to water. *Trans. ASAE* 23: 92-98.

NADSS. 2008: Weekly SPI Users Manual. Lincoln, Nebraska: University of Nebraska, National Agricultural Decision Support System. Available at: greenleaf.unl.edu/downloads/SPI_Manual.pdf. Accessed 27 January 2009.

Narasimhan, B. 2004. Development of indices for agricultural drought monitoring using a spatially distributed hydrologic model. PhD diss. Kingsville, Texas: Texas A & M University, Department of Biological and Agricultural Engineering.

Narasimhan, B., and R. Srinivasan. 2005. Development and evaluation of soil moisture deficit index and evapotranspiration deficit index for agricultural drought monitoring. *Agricultural and Forest Meteorology* 133: 69-88.

Nash, J. E., and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology* 10(3): 282-290.

Nayak, P. C., K. P. Sudheer, D. M. Rangan, and K. S. Ramasastri. 2004. A neuron-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology* 291: 52-66.

NDMC. 2006. What is Drought? Understanding ENSO and Forecasting Drought. Lincoln, Nebraska: University of Nebraska, National Drought Mitigation Center. Available at: http://www.drought.unl.edu/whatis/elnino.htm. Assessed 24 October 2009.

Nicholls, N. 1985. Towards the prediction of major Australian droughts. *Australian Meteorological Magazine* 33: 161-166.

NIDIS. 2007. The national integrated drought information system implementation plan: A pathway for national resilience. Boulder, Colorado: National Integrated Drought Information System.

Nielsen, R. L., P. R. Thomison, G. A. Brown, A. L. Halter, J. Wells, and K. L. Wuethrich. 2002. Corn: Delayed planting effects on flowering and grain maturation of dent corn. *Agronomy Journal* 94: 549-558.

Nyquist, H. 1932. Regeneration theory. *Bell System Technical Journal* 11: 126-147.

O'Neal, M. R., J. R. Frankenberger, and D. R. Ess. 2002. Use of CERES-Maize to study effect of spatial precipitation variability on yield. *Agricultural Systems* 73: 205-225.

Palmer, W. C. 1965. Meteorological drought. Research Paper No. 45. Washington, D.C.: U.S. Department of Commerce.

Palmer, W. C. 1968. Keeping track of moisture conditions, nationwide: The new crop moisture index. *Weatherwise* 21: 156-161.

Panigrahi, B., and S. N. Panda. 2003. Field test of a soil water balance simulation model. *Agricultural Water Management* 58: 223-240.

Panu, U. S., and T. C. Sharma. 2002. Challenges in drought research: Some perspectives and future directions. *Hydrological Sciences* 47(S): S19-S30.

Passioura, J. B. 1983. Roots and drought resistance. *Agricultural Water Management* 7: 265-280.

Philander, S. G. 1990. *El Niño, La Nina, and the Southern Oscillation*. New York, N.Y.: Academic Press.

Pisani, L. G., H. J. Fouche, and J. C. Venter. 1998. Assessing range drought in South Africa. *Agricultural Systems* 57: 367-380.

Porteous, A. S., R. E. Basher, and M. J. Salinger. 1994. Calibration and performance of the single-layer soil water balance model for pasture sites. *New Zealand Journal of Agricultural Research* 37: 107-118.

Principe, J. C., N. R. Euliano, and W. C. Lefebvre. 1999. *Neural and adaptive systems: fundamentals through simulations*. New York, N.Y.: John Wiley and Sons Inc.

Quiring S. M., and T. N. Papakryiakou. 2003. An evaluation of agricultural drought indices for the Canadian prairies. *Agricultural and Forest Meteorology* 118: 49-62.

Raes, D., S. Geerts, E. Kipkorir, J. Wellens, and A. Sahli. 2006. Simulation of yield decline as a result of water stress with a robust soil water balance model. *Agricultural Water Management* 81: 335-357.

Raes, D., P. Steduto, T. C. Hsiao, and E. Fereres. 2009. Chapter 1: AquaCrop - The FAO crop model to simulate yield response to water. In *AquaCrop Reference Manual*. Rome, Italy: Food and Agriculture Organization, Land and Water Division.

Ramirez, M. C. V., H. F. C. Velho, and N. J. Ferreira. 2005. Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. *Journal of Hydrology* 301: 146-162.

Rao, N. H. 1987. Field test of a simple soil-water balance model for irrigated areas. *Journal of Hydrology* 91: 179-186.

Rasmussen, E. M., R. E. Dickinson, J. E. Kutzbach, and M. K. Cleaveland. 1993. Climatology. In *Handbook of Hydrology*, 2.1-2.44. D. R. Maidment, ed. New York, N.Y.: McGraw-Hill, Inc.

Ratliff, L. F., J. T. Ritchie, and D. K. Cassel. 1983. Field-measured limits of soil water availability as related to laboratory-measured properties. *Soil Science Society of America Journal* 47: 770-775.

Redmond, K. 1991. Climate monitoring and indices. In *Proc. seminar and workshop on drought management and planning*. D. A. Wilhite, D. A. Wood, and P. A. Kay, eds. Lincoln, Nebraska: University of Nebraska.

Reicosky, D. C., and L. G. Heatherly. 1990. Soybean. In *Irrigation of agricultural crops*, 639-674. B. A. Stewart and D. R. Nielsen, eds. Madison, Wisconsin: ASA-CSSA-SSSA.

Reynolds, J. F., and B. Acock. 1997. Modularity and genericness in plant and ecosystem models. *Ecological Modelling* 94: 7-16.

Rhoads, F. M., and J. M. Bennett. 1990. Corn. In *Irrigation of agricultural crops*, 569-596. B. A. Stewart and D. R. Nielsen, eds. Madison, Wisconsin: ASA-CSSA-SSSA.

Ritchie, J. T., and E. Burnett. 1971. Dryland evaporative flux in a subhumid climate: 2. Plant influences. *Agronomy Journal* 63: 56-62.

Ritchie, J. T. 1998. Soil water balance and plant water stress. In *Systems Approaches for Sustainable Agricultural Development vol. 7 – Understanding options for agricultural production*, 41-54. Y. T. Gordon, G. Hoogenboom, and P. K. Thornton, eds. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ritchie, J. T., A. Gerakis, and A. Suleiman, 1999. Simple model to estimate field-measured soil water limits. *Trans. ASAE* 42(6): 1609-1614.

Robertson, M. J., S. Fukai, M. M. Ludlow, and G. L. Hammer. 1993a. Water extraction by grain sorghum in a sub-humid environment. I. Analysis of the water extraction pattern. *Field Crops Research* 33: 81-97.

Robertson, M. J., S. Fukai, M. M. Ludlow, and G. L. Hammer. 1993b. Water extraction by grain sorghum in a sub-humid environment. II. Extraction in relation to root growth. *Field Crops Research* 33: 99-112.

Rogers, J. C. 1984. The association between the North Atlantic Oscillation and the Southern Oscillation in the Northern Hemisphere. *Monthly Weather Review* 112(10): 1999-2015.

Ropelewski, C. F., and M. S. Halpert. 1986. North American precipitation and temperature patterns associated with the El Niño/South Oscillation (ENSO). *Monthly Weather Review* 115: 2352-2362.

Rumelhart, D. E., R. Durbin, R. Golden, and Y. Chauvin. 1995. Backpropagation: the basic theory. In *Backpropagation: Theory, architectures, and applications*, 1-34. Y. Chauvin and D. E. Rumelhart, eds. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Sahoo, G. B., C. Ray, and H. F. Wade. 2005. Pesticide prediction in ground water in North Carolina domestic wells using artificial neural networks. *Ecological Modeling* 183: 29-46.

Sakamoto, C.M. 1978. The Z-index as a variable for crop yield estimation. *Agricultural Meteorology* 19: 305-313.

Saltelli, A., T. H. Andres, and T. Homma. 1993. Sensitivity analysis of model output: An investigation of new techniques. *Computational Statistics and Data Analysis* 15: 211-238.

Saltelli, A., and I. M. Sobol. 1995. About the use of rank transformation in sensitivity analysis of model output. *Reliability Engineering and System Safety* 50: 225-239.

Saltelli, A., and R. Bolado. 1998. An alternative way to compute Fourier amplitude sensitivity test (FAST). *Computational Statistics & Data Analysis* 26: 445-460.

Saltelli, A., S. Tarantola, and K. P. S. Chan. 1999. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41(1): 39-56.

Saltelli, A., K. Chan, and M. Scott. 2000a. *Sensitivity analysis*. New York, N.Y.: John Wiley and Sons Ltd.

Saltelli, A., S. Tarantola, and F. Campolongo. 2000b. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15(4): 277-395.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto. 2004. *Sensitivity analysis in practice – a guide to assessing scientific models*. West Sussex, England: John Wiley & Sons Ltd.

Salter, P. J., and J. E. Goode. 1967. *Crop responses to water at different stages of growth*. Buckinghamshire, England: Commonwealth Agricultural Bureau.

Samarasinghe, S. 2007. *Neural networks for applied sciences and engineering*. Boca Raton, Florida: Auerbach Publications.

Sanchez-Cohen, I., V. L. Lopes, D. C. Slack, and M. M. Fogel. 1997. Water balance model for small scale water harvesting systems. *Journal of Irrigation and Drainage Engineering* 123(2): 123-128.

Saxton, K. E., W. J. Rawls, J. S. Romberger, and R. I. Papendick. 1986. Estimating generalized soil-water characteristics from texture. *Soil Science Society of America Journal 5*0: 1031-1036.

Schaibly, J. H., and K. E. Shuler. 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II Applications. *The Journal of Chemical Physics* 59(8): 3879-3888.

SDR. 2005. Grand challenges for disaster reduction. A report of the Subcommittee on Disaster Reduction. Washington, D.C.: National Science and Technological Council, Committee on Environment and Natural Resources, Subcommittee on Disaster Reduction (SDR).

Sevat, E., and A. Dezetter. 1991. Selection of calibration objective functions in the context of rainfall-runoff modeling in a Sudanese savannah area. *Hydrological Sciences Journal* 36(4): 307-330.

Shafer, B. A., and L. E. Dezman. 1982. Development of a surface water supply index to assess the severity of drought conditions in snow pack runoff areas. In *Proc. 50th Annual Western Snow Conference*, 164–175. Fort Collins, Colorado: Colorado State University.

Shantz, H. L., and L. N. Piemeisel. 1927. The water requirement of plants at Akron, Colorado. *Journal of Agricultural Research* 34: 1093-1190.

Sharda, R., and R. B. Patil. 1990. Neural networks as forecasting experts: An empirical test. In *Proceedings of the International Joint Conference on Neural Networks, vol. II*, 491-494. Washington, D.C.

Sharda, R., and R. B. Patil. 1992. Connectionist approach to time series prediction: an empirical approach. *Journal of Intelligent Manufacturing* 3: 317-323.

SimLab, 2009. *Documentation. SimLab ver. 3.2.5*. Ispra, Italy: Joint Research Centre of the European Commission, Econometrics and Applied Statistics Unit. Available at: http://simlab.jrc.ec.europa.eu/docs/html/index.html. Assessed 11 March 2010.

Singh, S. P. 1975. Studies on the effects of soil moisture stress on the yield of cotton. Indian *Journal of Plant Physiology* 18: 49-55.

Smith, M. 1996. *Neural networks for statistical modeling*. London, England: International Thompson Computers Press.

Smith, S. R., P. M. Green, A. P. Leonardi, and J. J. O'Brien. 1998. Role of multiple-level troposphereic circulations in forcing ENSO winter precipitation anomalies. *Monthly Weather Review* 126(12): 3102-3116.

Sobol, I. M. 1990. Sensitivity estimates for nonlinear mathematical models. *Matematicheskoe Modelirovanie* 2: 112-118 (in Russian). [English translation: Sobol, I. M. 1993. Sensitivity analysis for nonlinear mathematical models. Mathematical Modelling and Computational Experiment 1(4): 407-414].

Stahl, K., and S. Demuth. 1999. Methods for regional classification of streamflow drought series: cluster analysis. ARIDE Technical Report 1. Freiburg, Germany: University of Freiburg, Institute of Hydrology.

Steduto, P., T. C. Hsiao, D. Raes, and E. Fereres. 2009. AquaCrop - The FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agronomy Journal* 101(3): 426-437.

Steinemann, A. C. 2006. Using climate forecasts for drought management. *Journal of Applied Meteorology and Climatology* 45: 1353-1361.

Stewart, B. A. 2005. Adequate tools and technologies to support soil moisture management. In *Drought resistant soils: optimization of soil moisture for sustainable plant production (FAO Land and Water Bulletin 11)*, 49-58. A. Bot and J. Benites, eds. Rome, Italy: Food and Agriculture Organization.

Subrahmanyam, V. P. 1967. Incidence and spread of continental drought. WMO Report No. 2. Geneva, Switzerland: World Meteorological Organization.

Suleiman, A. A., and J. T. Ritchie. 2004. Modifications to the DSSAT vertical drainage model for more accurate soil water dynamics estimation. *Soil Science* 169(11): 745-757.

Szep, I. S., J. Mika, and Z. Dunkel. 2005. Palmer drought severity index as soil moisture indicator: physical interpretation, statistical behavior and relation to global climate. *Physics and Chemistry of the Earth* 30: 231-243.

Takagi, T., and M. Sugeno. 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans Systems, Man, and Cybernetics* 15: 116-132.

Tamm, T. 2002. Effects of meteorological conditions and water management on hydrological processes in agricultural fields – parameterization and modeling of Estonian case studies. PhD diss. Helsinki, Finland: Helsinki University of Technology.

Tang, Z., C. Almeida, and P. A. Fishwick. 1991. Time series forecasting using neural networks vs. Box-Jenkins methodology. *Simulation* 57(5): 303-310.

Tang, Z., and P. A. Fishwick. 1993. Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing* 5(4): 374-385.

Taylor, H. M., and B. Klepper. 1973. Rooting density and water extraction patterns for corn (Zea mays L.). *Agronomy Journal* 65: 965-968.

Thornthwaite, C. W. 1948. An approach toward a rational classification of climate. *Geographical Review* 38: 55-94.

Tokar, A. S., and P. A. Johnson. 1999. Rainfall runoff modeling using artificial neural network. *Journal of Hydrologic Engineering* 4(3): 232-239.

Trenberth, K. E. 1997. The definition of El Niño. *BAMS* 78(12): 2771-2777.

Tsakiris, G. P. 1982. A method for applying crop sensitivity factors in irrigation scheduling. *Agricultural Water Management* 5: 335-343.

Tsukamoto, Y. 1979. An approach to reasoning method. In *Advances in fuzzy set theory and applications*, 137-149. M. Gupta, R. K. Ragade, and R. R. Yager, eds. New York, N.Y.: North-Holland Publishing Company.

Tucker, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of the Environment* 8: 127-150.

USDA-NRCS. 1986. Urban hydrology for small watersheds. Technical Release 55. Washington, D.C.: U.S. Department of Agriculture, Natural Resources Conservation Service, Conservation Engineering Division.

USDA-NRCS. 1997. Chapter 3: Crops. In *210 - National Engineering Handbook (NEH) Part 652 - Irrigation Guide*. Washington, D.C.: U.S. Department of Agriculture, Natural Resources Conservation Service.

USDA-SCS. 1972. Estimation of direct runoff from storm rainfall. In *National Engineering Handbook, Section 4: Hydrology*. Washington, D.C.: U.S. Department of Agriculture, Soil Conservation Service.

Van Rooy, M. P. 1965. A rainfall anomaly index independent of time and space. *Notos* 14: 43–48.

Vasiliades, L., and A. Loukas. 2008. Meteorological drought forecasting models. *Geophysical Research Abstracts* 10: EGU2008-A-08458.

Verdin, J., and R. Klaver. 2002. Grid-cell based crop water accounting for the famine early warning system. *Hydrological Processes* 16: 1617-1630.

Verdoodt, A., E. van Ranst, L. Ye, and H. Verplancke. 2005. A daily multilayered water balance model to predict water and oxygen availability in tropical cropping systems. *Soil Use and Management* 21: 312-321.

Viets, F. G. 1962. Fertilizers and the efficient use of water. *Advances in Agronomy* 14: 223-264.

Wang, C., D. B. Enfield, S. K. Lee, and C. W. Landsea. 2006. Influences of the Atlantic Warm Pool on Western Hemisphere Summer Rainfall and Atlantic Hurricanes. *Journal of climate* 19: 3011-3028.

Wang, X., X. He, J. R. Williums, R. C. Izaurralde, and J. D. Atwood. 2005. Sensitivity and uncertainty analysis of crop yields and soil organic carbon simulated with EPIC. *Trans. ASAE* 48(3): 1041-1054.

Webner, L., and L. Nkemdirim. 1988. Palmer's drought indices revisited. *Geografiska Annaler* 80: 153-172.

Weghorst, K. M. 1996. *The reclamation drought index: Guidelines and practical applications*. Denver, Colorado: Bureau of Reclamation.

Weigend, A. S., B. A. Huberman, and D. E. Rumelhart. 1992. Predicting sun-spots and exchange rates with connectionist networks. In *Nonlinear Modeling and Forecasting*, 395-432. M. Casdagli and S. Eubank, eds. Redwood City, California: Addison-Wesley.

Wells, N. 2003. *PDSI User's Manual ver. 2.0.* Lincoln, Nebraska: University of Nebraska, National Agricultural Decision Support System. Available at: greenleaf.unl.edu/downloads/PDSI_Manual.pdf. Accessed 27 January 2009.

Werbos, P. J. 1974. Beyond regression: New tools for prediction and analysis in the behavioral sciences. PhD diss. Cambridge, Massachusetts: Harvard University.

Wheater, H. S., and D. J. Sherratt. 1983. Field validation of soil moisture models for water balance estimation. In *Proc. New Approaches in Water Balance Computations Workshop*, 61-74. Hamburg, Germany: International Association of Hydrological Sciences (IAHS).

Wilhite, D. A., and M. H. Glantz. 1985. Understanding the drought phenomenon: the role of definitions. *Water International* 10(3): 111-120.

Wilkerson, G. G., J. W. Jones, K. J. Boote, K. T. Ingram, and J. W. Mishoe. 1983. Modeling soybean growth for crop management. *Trans. ASAE* 26: 63-73.

Williams, J., P. J. Ross, and K. L. Bristow. 1991. Perspicacity, precision and pragmatism in modeling crop water supply. In *Climatic Risk in Crop Production: Models and Management for the Semiarid Tropics and Subtropics*, 73-96. R. C. Muchow and J. A. Bellamy, eds. London, England: CAB International.

Willmott, C. J. 1981. On the validation of models. *Phys. Geog.* 2: 184-194.

Wilpen, L. G., D. Nagin, and J. Szczypula. 1994. Comparative study of artificial neural network and statistical models for predicting student grade point averages. *International Journal of forecasting* 10: 17-34.

Winsberg, M. D. 2003. *Florida Weather*. 2nd ed. Gainesville, Florida: University Press of Florida.

WMO. 1975. Drought and agriculture. WMO Technical Note 138. Geneva, Switzerland: World Meteorological Organization.

Wood, A. W., and D. P. Lettenmaier, 2006. A test bed for new seasonal hydrologic forecasting approaches in the western United States. *BAMS* 1699-1712.

Wright, D., J. Marois, J. Rich, and R. Sprenkel. 2008. Field corn production guide. Gainesville, Florida: University of Florida, Florida Cooperative Extension Service, IFAS. Available at: edis.ifas.ufl.edu/pdffiles/AG/AG20200.pdf. Accessed 27 January 2009.

WSCO, 2003. Climatic normals. Madison, Wisconsin: Wisconsin State Climatology Office (WSCO). Available at: http://www.aos.wisc.edu/~sco/normals.html. Accessed 23 March 2010.

Wu, L. 1985. Matching irrigation to turfgrass root depth. *California turfgrass culture* 35: 1-4.

Wu, J., R. Zhang, and S. Gui. 1999. Modeling soil water movement with water uptake by roots. *Plant and Soil* 215: 7-17.

Yazar, A., T. A. Howell, D. A. Dusek, and K. S. Copeland. 1999. Evaluation of crop water stress index for LEPA irrigated corn. *Irrigation Science* 18: 171-180.

Yin, Z. Y. 1994. Moisture condition in the south-eastern USA and teleconnection patterns. *International Journal of Climatology* 14(9): 947-967.

Yokoo, Y., S. Kazama, M. Sawamoto, and H. Nishimura. 2001. Regionalization of lumped water balance model parameters based on multiple regression. *Journal of Hydrology* 246: 209-222.

Zadeh, L. A. 1965. Fuzzy sets. *Infection Control* 8(3): 338-353.

Zazueta, F., S, A. Brockway, L. Landrum, and B. McCarty. 2008. Turf Irrigation for the home. Gainesville, Florida: Electronic Data Information Source (EDIS) – UF/IFAS Extension. Available at: edis.ifas.ufl.edu/BODY_AE144#IMAGE%20AE:AE144F2. Accessed 27 January 2009.

Zhang, G., B. E. Patuwo, and M. Y. Hu. 1998. Forecasting with artificial neural networks: The state of art. *International Journal of Forecasting* 14: 35-62.

Zuo, Q., J. Shi, Y. Li, and R. Zhang. 2006. Root length density and water uptake distributions of winter wheat under sub-irrigation. *Plant and Soil* 285: 45-55.

Zur, B., J. W. Jones, K. J. Boote, and L. C. Hammond. 1982. Total resistance to water flow in field soybeans: II. Limiting soil moisture. *Agronomy Journal* 74: 99-105.

BIOGRAPHICAL SKETCH

Prem Woli was born in a rural village of Nepal into a family whose ancestors, including his parents, were completely illiterate and virtually ignorant of the world. Because of poverty and other hindrances, he could not continue his education in time. He received his bachelor's, master's, and doctorate degrees only after about 10, 20, and 30 years of his high school graduation in 1980, respectively. In 2010, he acquired his PhD degree from University of Florida, an institution that is one of the best in a country that is perhaps the most advanced in the world and the capital of the gator nation, becoming a proud citizen of which is a dream of numerous people in the world. Can there be a greater achievement than this for him?