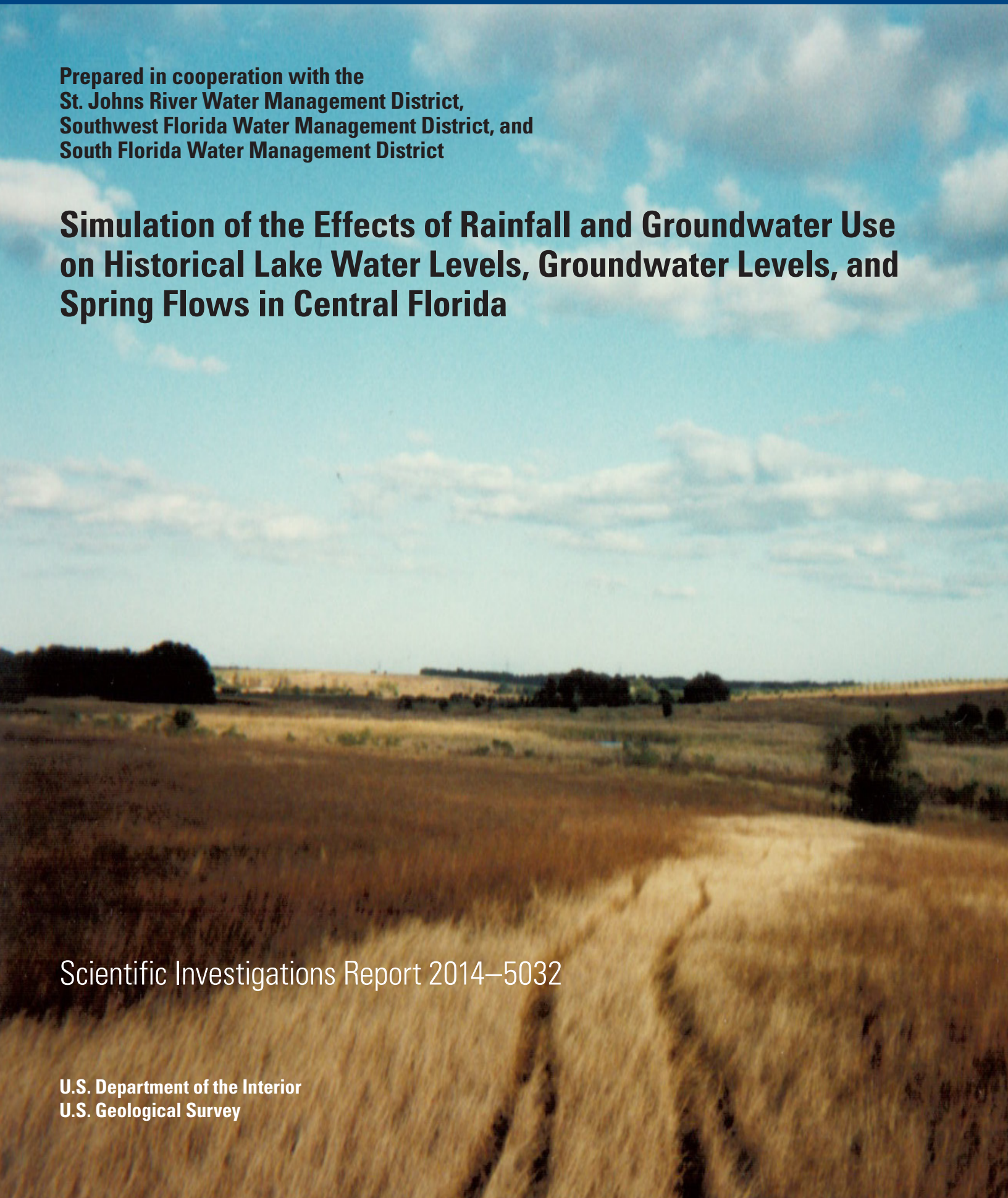


Prepared in cooperation with the
St. Johns River Water Management District,
Southwest Florida Water Management District, and
South Florida Water Management District

Simulation of the Effects of Rainfall and Groundwater Use on Historical Lake Water Levels, Groundwater Levels, and Spring Flows in Central Florida

Scientific Investigations Report 2014–5032

U.S. Department of the Interior
U.S. Geological Survey

Cover: Photo showing grassland area (formerly citrus grove) in the Lake Wales Ridge physiographic region, south Lake County, Florida (photograph by A.M. O'Reilly).

Simulation of the Effects of Rainfall and Groundwater Use on Historical Lake Water Levels, Groundwater Levels, and Spring Flows in Central Florida

By Andrew M. O'Reilly, Edwin A. Roehl, Jr., Paul A. Conrads, Ruby C. Daamen, and Matthew D. Petkewich

Prepared in cooperation with the
St. Johns River Water Management District,
Southwest Florida Water Management District, and
South Florida Water Management District

Scientific Investigations Report 2014–5032

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2014

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit *http://www.usgs.gov* or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit *http://www.usgs.gov/pubprod*

To order this and other USGS information products, visit *http://store.usgs.gov*

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:
O'Reilly, A.M., Roehl, E.A., Jr., Conrads, P.A., Daamen, R.C., and Petkewich, M.D., 2014, Simulation of the effects of rainfall and groundwater use on historical lake water levels, groundwater levels, and spring flows in central Florida: U.S. Geological Survey Scientific Investigations Report 2014–5032, 153 p., *http://dx.doi.org/10.3133/sir20145032*.

ISSN 2328-0328 (online)

Acknowledgments

The provision of water-use data by the following individuals is greatly appreciated: Jill Hood and Mike Kelley (Southwest Florida Water Management District) for providing reported or estimated public-supply and phosphate industry groundwater withdrawals; Douglas Munch and Jay McLeod (St. Johns River Water Management District) for providing estimated groundwater withdrawals for citrus, non-citrus agricultural, and golf course irrigation areas; Douglas Munch, Bruce Florence, and George Robinson (St. Johns River Water Management District), Lindsey Killian, Jennifer Kasper, and Katherine Graf (University of Florida), and Richard L. Marella (U.S. Geological Survey) for providing reported or estimated public-supply groundwater withdrawals. The authors thank Paul G. Biscardi (formerly at the U.S. Geological Survey; currently (2014) at the University of Central Florida) for acquisition and compilation of historical data and development of the hydrologic database on which this study was based.

Contents

Abstract 1
Introduction 2
  Purpose and Scope 4
  Description of Study Area 5
    Climate 5
    Hydrogeologic Setting 5
  Previous Studies 6
  Approach 6
Historical Data 7
  Hydrologic Data 7
    Historical Hydrologic Database 7
    Rainfall, Air Temperature, and Potential Evapotranspiration 9
    Lake Water Level 9
    Groundwater Level 9
    Spring Flow 11
  Groundwater Use 11
    Public Supply 11
    Agricultural 11
    Phosphate Mining 14
    Drainage Wells 14
  Land Use and Land Cover 16
Characterization of Historical Data 19
  Hydrologic Data 19
    Range and Linear Trend 19
    Time-Series Cluster Analysis 21
  Groundwater Use 30

Methods for Simulation of Historical Lake Water Levels, Groundwater Levels, and Spring Flows 45
Limitations of Datasets....45
System Dynamics and Analysis 45
Signal Filtering and Decomposition 45
Correlation Analysis 48
State-Space Reconstruction 51
Input-Output Mapping and Problem Representation 54
Artificial Neural Network Models 57
Statistical Measures of Prediction Accuracy....59
Development of Artificial Neural Network Models in Central Florida 60
Model Construction and Training....60
Rainfall Models....61
Groundwater-Use Models....62
Model Prediction Accuracy and Example Model Simulations 63
Lake Water-Level Models....63
Crooked Lake near Babson Park....63
Prevatt Lake 66
Groundwater-Level Models 66
OR-47 Floridan Well 66
Taft Surficial Well....69
Spring-Flow Models 69
Rock Springs....70
Sanlando Springs....72
Development of the Decision Support System 74
System Architecture....74
Model Simulation Control and Graphics 75
Comparison of Rainfall and Groundwater-Use Effects....76
Sensitivity Analysis....76
Potential Application to Vulnerability Assesment 82
Limitations of Artificial Neural Network Models 82
Comparison of Artificial Neural Network Models With a Physics-Based Model....84
Summary and Conclusions....87
References Cited....91
Appendixes 1–8.
1. Description of hydrologic database 100
2. Description of Groundwater-Use Data Viewer 106
3. Summary of artificial neural network models 107
4. Input variables used in artificial neural network models 108
5. Summary statistics for lake water-level, groundwater-level, and spring-flow models ...114
6. User's manual for the Central Florida Artificial Neural Network Decision Support System (CFANN DSS) 122
7. Description of Model Data Viewer 140
8. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data 142

Figures

1. Map showing location of study area and the Central Florida Coordination Area boundaries 3
2. Map showing locations of sites for which historical hydrologic data were acquired for this study 8
3. Map showing lakes classified as either flowthrough or closed basin for which historical water-level data were acquired 10
4. Map showing locations of water-treatment plants in the Central Florida Coordination Area and vicinity for which public water-supply groundwater withdrawal data from 1978 through 2008 were available 12
5. Map showing locations of citrus, non-citrus agricultural, and golf course irrigation areas in the Central Florida Coordination Area and vicinity where groundwater withdrawal estimates were available 13
6. Map showing locations of phosphate industry wells for which reported groundwater withdrawal rates from 1978 through 2010 were available, phosphate mining areas and phosphate chemical plants where groundwater withdrawal estimates from 1935 through 1977 were available, and drainage wells in the Central Florida Coordination Area and vicinity 15
7. Map showing generalized land use and land cover in the study area for *A*, 1977, and *B*, 2006 17
8. Map showing range of daily water level for lakes in the study area with at least 10 measurements and a period of record of at least 10 years 20
9. Graph showing exceedance probability of range in daily lake water level by lake type (closed basin or flowthrough) and physiography (ridge or non-ridge physiographic region) for lakes in the study area with at least 10 measurements and a period of record of at least 10 years 21
10. Map showing range of daily groundwater level for wells in the study area with at least 10 measurements and a period of record of at least 10 years 22
11. Map showing range of daily flow and mean flow for springs in the study area with at least 10 measurements and a period of record of at least 10 years 23
12. Map showing linear trend in daily average air temperature for sites in the study area with a period of record of at least 45 years 24
13. Map showing linear trend in daily rainfall for sites in the study area with a period of record of at least 95 years (1901 or 1914 through 2008) 25
14. Graph showing exceedance probability of linear trend in daily lake water level by lake type (closed basin or flowthrough) and physiography (ridge or non-ridge physiographic region) for lakes in the study area with at least 10 measurements and a period of record of at least 10 years 26
15. Map showing linear trend in daily water level for lakes in the study area with at least 10 measurements and a period of record of at least 10 years 27
16. Map showing linear trend in daily groundwater level for wells in the study area with at least 10 measurements and a period of record of at least 10 years 28
17. Map showing linear trend in daily flow and mean flow for springs in the study area with at least 10 measurements and a period of record of at least 10 years 29
18. Graph showing root-mean-square error for k-means cluster analysis of water-level data for lakes and wells and flow data for springs in the study area 30
19. Map showing locations of lakes in the study area identified by group number based on the 31-year cluster analysis of lake water-level data from 1978 through 2008 31

20. Graph showing distribution of lakes in the study area identified by lake type (closed basin or flowthrough) and physiography (ridge or non-ridge physiographic region) based on the 31-year cluster analysis of lake water-level data from 1978 through 2008 32
21. Map showing locations of wells in the study area identified by group number and aquifer based on the 31-year cluster analysis of groundwater-level data from 1978 through 2008 33
22. Map showing locations of springs in the study area identified by group number based on the 41-year cluster analysis of spring-flow data from 1969 through 2009 and locations of wells by group number and aquifer based on the 41-year cluster analysis of groundwater-level data from 1968 through 2008....34
23. Map showing direction of flow for filling missing rainfall data using "IF-THEN-ELSE" rules 35
24. Map showing locations of National Oceanic and Atmospheric Administration rain gages in the study area identified by group number based on the 67-year cluster analysis of rainfall data from 1942 through 2008....36
25. Graphs showing moving window averages of rainfall data for the individual National Oceanic and Atmospheric Administration rain gages in the study area composing the four groups derived from the 67-year cluster analysis....37
26. Map showing locations of National Oceanic and Atmospheric Administration air temperature measurement sites in the study area identified by group number based on the 67-year cluster analysis of computed Hargreaves potential evapotranspiration data from 1942 through 2008 38
27. Graphs showing moving window averages of computed Hargreaves potential evapotranspiration data for the individual National Oceanic and Atmospheric Administration air temperature measurement sites in the study area composing the three groups derived from the 67-year cluster analysis 39
28. Map showing grid used for aggregating groundwater-use data in the Groundwater-Use Data Viewer showing the identification label for each cell 40
29. Graphs showing total groundwater use in the Central Florida Coordination Area and vicinity by category from 1950 through 2008 based on the Groundwater-Use Data Viewer for *A*, monthly data, and *B*, annual average data....41
30. Diagram showing screen capture of the Groundwater-Use Data Viewer showing groundwater use by category and sum total in million gallons per day for a dry period (May 2000) 43
31. Diagram showing screen capture of the Groundwater-Use Data Viewer showing groundwater use by category and sum total in million gallons per day for a wet period (August 2004)....44
32. Map showing sites in central Florida for which artificial neural network models were developed 46
33. Graphs showing average daily and low-pass filtered rainfall for the standard rainfall time series for *A*, 1945 through 2008, and *B*, 1997 through 2001....47
34. Graphs showing average monthly and low-pass filtered groundwater use in the Central Florida Coordination Area and vicinity for the sum total of all water-use categories for *A*, 1957 through 2008, and *B*, 1997 through 2001....49
35. Graph showing pearson product-moment correlation coefficient (R) between rainfall (for each rain gage and rainfall-signal pair) and hydrologic response at each site....50

36. Graphs showing comparison of the Pearson product-moment correlation coefficients (R) between rainfall (for each rain gage and rainfall-signal pair) and hydrologic response and the distance between the site and each rain gage for *A*, Lake Clinch (site number 50), *B*, ROMP 60 Floridan well (site number 472), and *C*, Rock Springs (site number 309)....51
37. Graph showing pearson product-moment correlation coefficient (R) between the R shown in figure 35 (between rainfall and hydrologic response) and the distances between the site and each rain gage for each site....52
38. Graph showing pearson product-moment correlation coefficient (R) between groundwater-use signals and hydrologic response at each site....52
39. Graph showing pearson product-moment correlation coefficient (R) between groundwater-use signals and rainfall signals at each National Oceanic and Atmospheric Administration (NOAA) rain gage....53
40. Diagram showing two-stage model architecture used to simulate water level at lakes or wells, or flow at springs....55
41. Pie charts showing water budgets for the combined surficial and Floridan aquifer systems simulated by the Lake County/Ocala National Forest east-central Florida, and east-central Florida transient models for *A*, inflows, and *B*, outflows....56
42. Diagram showing multilayer perceptron artificial neural network architecture....59
43. Graphs showing model-fit statistics for the final water-level or flow models for each site for the 1965 to 2008 simulation period: *A*, coefficient of determination, *B*, root-mean-square error, and *C*, and percent model error....64
44. Graphs showing results of lake water-level simulation for Crooked Lake near Babson Park (site number 49) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines....65
45. Graphs showing results of lake water-level simulation for Prevatt Lake (site number 5) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines....67
46. Graphs showing results of groundwater-level simulation for the OR-47 Floridan well (site number 648) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines....68
47. Graphs showing results of groundwater-level simulation for the Taft Surficial well (site number 712) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines....70
48. Graphs showing results of spring-flow simulation for Rock Springs (site number 309) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines....71
49. Graphs showing results of spring-flow simulation for Sanlando Springs (site number 312) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines....73
50. Diagram showing architectural elements of the Central Florida Artificial Neural Network Decision Support System....74
51. Diagram showing simulator controls used to run a simulation of the Central Florida Artificial Neural Network Decision Support System and the associated graphical display....75

52. Graphs showing sensitivity of lake water-level, groundwater-level, and spring-flow models as indicated by the maximum change in water level or flow simulated during 1965 to 2008 caused by changes in *A*, rainfall, and *B*, groundwater use.....77
53. Graphs showing sensitivity of lake water-level, groundwater-level, and spring-flow models as indicated by the minimum change in water level or flow simulated during 1965 to 2008 caused by changes in *A*, rainfall, and *B*, groundwater use.....78
54. Graphs showing simulated water levels at the ROMP 60 Floridan well (site 472) for historical conditions and hypothetical rainfall and groundwater-use scenarios showing *A*, water-level altitude for historical and scenario conditions, and *B*, the change in water level between historical and scenario conditions.....80
55. Graphs showing simulated water levels at the Cocoa P Floridan well (site 622) for historical conditions and hypothetical rainfall and groundwater-use scenarios showing *A*, water-level altitude for historical and scenario conditions, and *B*, the change in water level between historical and scenario conditions.....81
56. Graphs showing model-fit statistics for the final water-level or flow models for the 51 sites incorporated in the Central Florida Artificial Neural Network Decision Support System for the 1995 to 2006 period: *A*, coefficient of determination, *B*, Nash Sutcliffe coefficient of efficiency, *C*, root-mean-square error, and *D*, percent model error 85
57. Graphs showing model-fit statistics for 48 sites in the east-central Florida transient (ECFT) model (Sepúlveda and others, 2012) for the 1995 to 2006 period: *A*, coefficient of determination, *B*, Nash Sutcliffe coefficient of efficiency, *C*, root-mean-square error, and *D*, percent model error 86

Tables

1. Linear trend statistics for air temperature, rainfall, lake water levels, groundwater levels, and spring flow for study area sites in central Florida.....21
2. Site information for lakes, wells, and springs for which artificial neural network models were developed. 96
3. Standard deviation of rainfall and groundwater-use signals in central Florida. 57

Conversion Factors, Abbreviations, and Datums

Inch/Pound to SI

Multiply	By	To obtain
	Length	
inch (in)	2.54	centimeter (cm)
inch (in)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
square foot (ft^2)	0.09290	square meter (m^2)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
gallon (gal)	0.003785	cubic meter (m^3)
million gallons (Mgal)	3,785	cubic meter (m^3)
cubic inch (in^3)	16.39	cubic centimeter (cm^3)
cubic foot (ft^3)	0.02832	cubic meter (m^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
gallon per day (gal/d)	0.003785	cubic meter per day (m^3/d)
million gallons per day (Mgal/d)	0.04381	cubic meter per second (m^3/s)
inch per year (in/yr)	25.4	millimeter per year (mm/yr)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

°F=(1.8×°C)+32

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

°C=(°F-32)/1.8

Vertical coordinate information is referenced to the National Geodetic Vertical Datum of 1929 (NGVD 29).

Horizontal coordinate information is referenced to the North American Datum of 1927 (NAD 27).

Altitude, as used in this report, refers to distance above the vertical datum.

Abbreviations

ANN	artificial neural network
CFANN	central Florida artificial neural network
CFCA	Central Florida Coordination Area
CFWI	Central Florida Water Initiative
DSS	decision support system
ECF	east-central Florida
ECFT	east-central Florida transient
ET	evapotranspiration
FDEP	Florida Department of Environmental Protection
GIS	geographic information system
GUDV	Groundwater-Use Data Viewer
GUI	graphical user interface
LCONF	Lake County/Ocala National Forest
ME	mean error
MOR	monthly operating report
MDV	Model Data Viewer
MWA	moving window average
MWD	moving window difference
NASS	National Agricultural Statistics Service
NOAA	National Atmospheric and Oceanic Administration
NSCE	Nash-Sutcliffe coefficient of efficiency

PET	potential evapotranspiration
PME	percent model error
R	Pearson product-moment correlation coefficient
ROMP	Regional Observation Monitoring Program
RMSE	Root-mean-square error
SFWMD	South Florida Water Management District
SJRWMD	St. Johns River Water Management District
SWFWMD	Southwest Florida Water Management District
USGS	U.S. Geological Survey
WUP	water use permit

Simulation of the Effects of Rainfall and Groundwater Use on Historical Lake Water Levels, Groundwater Levels, and Spring Flows in Central Florida

By Andrew M. O'Reilly,[1] Edwin A. Roehl, Jr.,[2] Paul A. Conrads,[1] Ruby C. Daamen,[2] and Matthew D. Petkewich[1]

Abstract

The urbanization of central Florida has progressed substantially in recent decades, and the total population in Lake, Orange, Osceola, Polk, and Seminole Counties more than quadrupled from 1960 to 2010. The Floridan aquifer system is the primary source of water for potable, industrial, and agricultural purposes in central Florida. Despite increases in groundwater withdrawals to meet the demand of population growth, recharge derived by infiltration of rainfall in the well-drained karst terrain of central Florida is the largest component of the long-term water balance of the Floridan aquifer system. To complement existing physics-based groundwater flow models, artificial neural networks and other data-mining techniques were used to simulate historical lake water level, groundwater level, and spring flow at sites throughout the area.

Historical data were examined using descriptive statistics, cluster analysis, and other exploratory analysis techniques to assess their suitability for more intensive data-mining analysis. Linear trend analyses of meteorological data collected by the National Oceanic and Atmospheric Administration at 21 sites indicate 67 percent of sites exhibited upward trends in air temperature over at least a 45-year period of record, whereas 76 percent exhibited downward trends in rainfall over at least a 95-year period of record. Likewise, linear trend analyses of hydrologic response data, which have varied periods of record ranging in length from 10 to 79 years, indicate that water levels in lakes (307 sites) were about evenly split between upward and downward trends, whereas water levels in 69 percent of wells (out of 455 sites) and flows in 68 percent of springs (out of 19 sites) exhibited downward trends. Total groundwater use in the study area increased from about 250 million gallons per day (Mgal/d) in 1958 to about 590 Mgal/d in 1980 and remained relatively stable from 1981 to 2008, with a minimum of 559 Mgal/d in 1994 and a maximum of 773 Mgal/d in 2000. The change in groundwater-use trend in the early 1980s and the following period of relatively slight trend is attributable to the concomitant effects of increasing public-supply withdrawals and decreasing use of water by the phosphate industry and agriculture.

On the basis of available historical data and exploratory analyses, empirical lake water-level, groundwater-level, and spring-flow models were developed for 22 lakes, 23 wells, and 6 springs. Input time series consisting of various frequencies and frequency-band components of daily rainfall (1942 to 2008) and monthly total groundwater use (1957 to 2008) resulted in hybrid signal-decomposition artificial neural network models. The final models explained much of the variability in observed hydrologic data, with 43 of the 51 sites having coefficients of determination exceeding 0.6, and the models matched the magnitude of the observed data reasonably well, such that models for 32 of the 51 sites had root-mean-square errors less than 10 percent of the measured range of the data. The Central Florida Artificial Neural Network Decision Support System was developed to integrate historical databases and the 102 site-specific artificial neural network models, model controls, and model output into a spreadsheet application with a graphical user interface that allows the user to simulate scenarios of interest.

Overall, the data-mining analyses indicate that the Floridan aquifer system in central Florida is a highly conductive, dynamic, open system that is strongly influenced by external forcing. The most important external forcing appears to be rainfall, which explains much of the multiyear cyclic variability and long-term downward trends observed in lake water levels, groundwater levels, and spring flows. For most sites, groundwater use explains less of the observed variability in water levels and flows than rainfall. Relative groundwater-use impacts are greater during droughts, however, and long-term trends in water levels and flows were identified that are consistent with historical groundwater-use patterns. The sensitivity of the hydrologic system to rainfall is expected, owing to the well-drained karst terrain and relatively thin confinement of the Floridan aquifer system in much of central Florida. These

[1]U.S. Geological Survey.

[2]Advanced Data Mining, LLC, Greenville, South Carolina.

characteristics facilitate the relatively rapid transmission of infiltrating water from rainfall to the water table and contribute to downward leakage of water to the Floridan aquifer system. The areally distributed nature of rainfall, as opposed to the site-specific nature of groundwater use, and the generally high transmissivity and low storativity properties of the semiconfined Floridan aquifer system contribute to the prevalence of water-level and flow patterns that mimic rainfall patterns. In general, the data-mining analyses demonstrate that the hydrologic system in central Florida is affected by groundwater use differently during wet periods, when little or no system storage is available (high water levels), compared to dry periods, when there is excess system storage (low water levels). Thus, by driving the overall behavior of the system, rainfall indirectly influences the degree to which groundwater use will effect persistent trends in water levels and flows, with groundwater-use impacts more prevalent during periods of low water levels and spring flows caused by low rainfall and less prevalent during periods of high water levels and spring flows caused by high rainfall. Differences in the magnitudes of rainfall and groundwater use during wet and dry periods also are important determinants of hydrologic response.

An important implication of the data-mining analyses is that rainfall variability at subannual to multidecadal timescales must be considered in combination with groundwater use to provide robust system-response predictions that enhance sustainable resource management in an open karst aquifer system. The data-driven approach was limited, however, by the confounding effects of correlation between rainfall and groundwater use, the quality and completeness of the historical databases, and the spatial variations in groundwater use. The data-mining analyses indicate that available historical data when used alone do not contain sufficient information to definitively quantify the related individual effects of rainfall and groundwater use on hydrologic response. The knowledge gained from data-driven modeling and the results from physics-based modeling, when compared and used in combination, can yield a more comprehensive assessment and a more robust understanding of the hydrologic system than either of the approaches used separately.

Introduction

The urbanization of central Florida has progressed substantially in recent decades, with total population in Lake, Orange, Osceola, Polk, and Seminole Counties more than quadrupling from 590,000 in 1960 to 2,740,000 in 2010 (Florida Office of Economic and Demographic Research, 2012). This development has led to an increasing demand for water, which has been met primarily by groundwater withdrawals from the Floridan aquifer system. During the period 1965–2005, groundwater withdrawals increased 53 percent, and in 2005, groundwater supplied 762 million gallons per day (Mgal/d) in Lake, Orange, Osceola, Polk, and Seminole Counties, or 95 percent of total water withdrawals in central Florida (Marella, 1995, 2009). The population of central Florida is projected to grow 57 percent from 2011 to 2040 (Smith and Rayer, 2012), and the demand for water is expected to increase as a result. Despite important anthropogenic influences, recharge to groundwater by infiltration of rainfall in the well-drained karst terrain of central Florida is the largest component of the water budget for the Floridan aquifer system during years with normal rainfall (Sepúlveda and others, 2012). The variability in historical seasonal to multidecadal rainfall patterns in Florida can be attributed to both local/global and natural/anthropogenic factors (Obeysekera and others, 2011), and climate variability is predicted to become more extreme (Berry and others, 2011). Consequently, variations in both rainfall and groundwater use may influence surface and subsurface water-level and flow conditions, potentially affecting the ability of the regional water resources to meet both human and environmental needs.

In some areas of central Florida, declining water levels and increasing salinity of groundwater have led to concerns by local and State water managers. The area of concern has been designated as the Central Florida Coordination Area (CFCA) and includes the southern part of Lake County and all of Orange, Osceola, Polk, and Seminole Counties (fig. 1). An analysis of data collected for various periods from 1941 to 2009 at 115 sites in the CFCA indicates that water levels at 31 of 62 wells, flow at 4 of 6 springs, and water levels at 13 of 47 lakes have exhibited statistically significant (80 percent confidence level) downward trends over their respective periods of record (Murch and Tara, 2010). The data indicate downward trends in water levels primarily in wells penetrating the Upper Floridan aquifer, which is the primary source of water for potable, industrial, and agricultural uses in the CFCA; these downward trends have been attributed to a combination of increases in groundwater withdrawals and long-term below-average rainfall (Spechler and Halford, 2001). Increases in groundwater levels have been observed in some areas, particularly in Polk County, and were attributed to decreases in groundwater withdrawals due to changes in phosphate mining practices (Spechler and Kroening, 2007; Murch and Tara, 2010). Multiyear cyclic variations in rainfall may contribute to observed periods of generally increasing or decreasing lake levels (German and Adamski, 2005). Upward trends in the concentrations of chloride in groundwater have occurred in the vicinity of well fields in eastern Seminole and Orange Counties, suggesting the upward movement of saline water, possibly the consequence of groundwater withdrawals (Spechler and Halford, 2001; Adamski and German, 2004). It is important to note, however, that observed trends in hydrologic data in the CFCA may be attributable to factors other than rainfall or groundwater withdrawals. For example, land application of reclaimed water has been shown to cause increases in water-table altitudes as well as water levels in the Upper Floridan aquifer (O'Reilly, 1998; Adamski and German, 2004). Similarly, other land-use changes, such as increases in irrigated land areas and construction of drainage ditches to

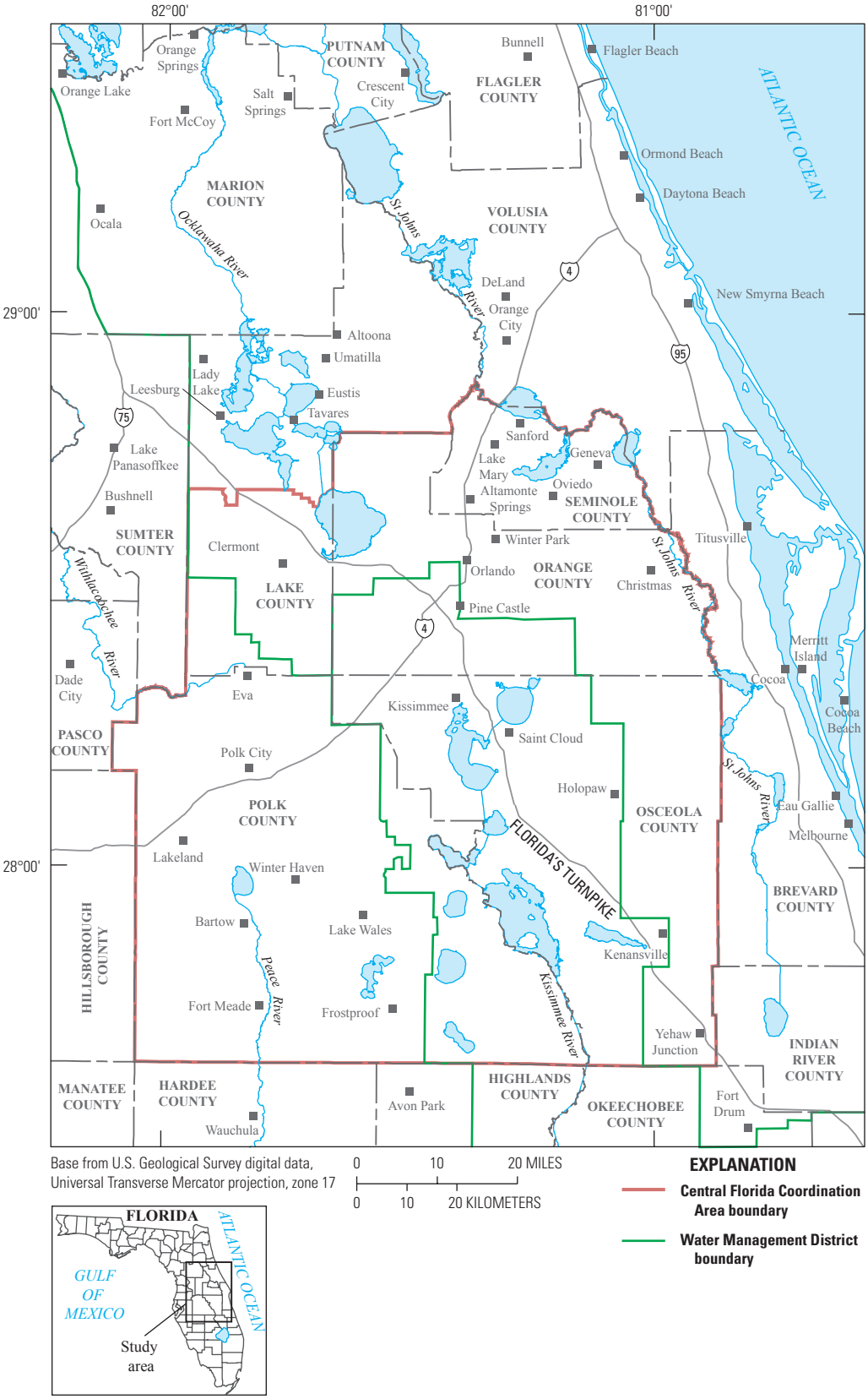


Figure 1. Location of study area and the Central Florida Coordination Area boundaries.

lower the water table and make land more suitable for development, have been suggested as possible causes of increased groundwater seepage to streams as inferred from observed long-term increases in 7-day low flow in some streams in urbanizing areas in Orange County (German and Adamski, 2005). Complex feedback mechanisms among hydroclimatic processes can affect the response of the hydrologic system. For example, in an area where the water table is near land surface and groundwater is subject to loss by evapotranspiration (ET), which occurs throughout much of the CFCA, small changes in water-table depth can result in changes in the water available for groundwater recharge (Knowles and others, 2002). Additionally, land-cover changes, such as those caused by urban or agricultural development, can affect spatial ET patterns. Atmospheric model simulations performed by Pielke and others (1999) showed that land-cover changes from 1900 to 1993 in south Florida potentially influenced local rainfall patterns, resulting in average summer rainfall as much as 11 percent less than what would have occurred in the absence of any land-cover changes. The cumulative effects of changes in the hydroclimatic system in the CFCA caused by natural or anthropogenic factors may result in changes in the long-term balance between the recharge and discharge of groundwater. Over time, such changes could cause changes in groundwater levels, spring flows, and the position of the saline-water/freshwater interface.

The primary agencies that manage water resources in the region, the St. Johns River Water Management District (SJRWMD), the South Florida Water Management District (SFWMD), and the Southwest Florida Water Management District (SWFWMD), are concerned that the groundwater resource is managed to maintain sufficient water supplies to meet the needs of future population growth and the environment. To address these concerns, the Central Florida Water Initiative (CFWI) was undertaken. The CFWI is a collaborative process—among the SJRWMD, SFWMD, SWFWMD, Florida Department of Environmental Protection, Florida Department of Agriculture and Consumer Services, central Florida water utilities, and other stakeholders—to assess whether the Floridan aquifer system is reaching its sustainable limits of use in the CFCA and to develop a regional water supply plan (*http://cfwiwater.com/*).

In 2005, the U.S. Geological Survey (USGS) began a 3-year study in cooperation with the SJRWMD to compile and analyze historical lake water-level, groundwater-level, spring-flow, rainfall, and groundwater-use data in Lake, Orange, and Seminole Counties to assess utility of artificial neural networks (ANNs) and other data-mining techniques for identifying the effects of rainfall and groundwater use. On the basis of the effectiveness of the data-mining techniques demonstrated in the initial results, the original study was expanded in 2008 to include an additional 4 years of investigation to address water-resource management concerns in the entire CFCA. The expanded USGS study was conducted in cooperation with SJRWMD, SFWMD, and SWFWMD to compile and assess historical lake water-level, groundwater-level, spring-flow, meteorological, and groundwater-use data in the entire CFCA with the purpose of quantifying, to the extent possible given the available data, the effects of weather variation and groundwater use. In 2011, the USGS became a participant in the CFWI, collaborating with other stakeholders in the study and assessment of historical hydrologic data in the CFCA. In addition to providing supporting science for local water-resource managers and stakeholders, this study addresses Federal science interests by assessing the status of and change in freshwater resources—part of the USGS science strategy direction "Water Census of the United States" (U.S. Geological Survey, 2008).

Purpose and Scope

This report presents the results of an investigation in which the response of lake water levels, groundwater levels, and spring flows to changing rainfall and groundwater-use conditions over a multidecadal period was analyzed. The report documents the development of the Central Florida Artificial Neural Network Decision Support System (CFANN DSS) and provides examples of the application of the CFANN DSS to simulate hydrologic response to historical rainfall and groundwater use from 1965 through 2008. Additionally, the report provides examples of how the CFANN DSS may be used to simulate scenarios of interest based on user-specified rainfall and groundwater use. The compilation of hydrologic response data was limited to Lake, Orange, Osceola, Polk, and Seminole Counties, which includes all of the CFCA (fig. 1). Streamflow data were not analyzed in this study.

An important part of the USGS mission is to provide scientific information for the effective water-resources management of the Nation. To assess the quantity and quality of the Nation's surface water and groundwater, the USGS collects hydrologic and water-quality data from rivers, lakes, and estuaries by using standardized methods and maintains the data from these stations in a national database. Additionally, many local and State agencies collect and store such data in their respective institutional databases. These databases commonly are underutilized and underinterpreted for addressing contemporary hydrologic issues. The techniques described in this report demonstrate how valuable information can be extracted from existing disparate databases to assist local, State, and Federal agencies in understanding and managing complex hydrologic systems. The application of data-mining techniques, including ANN models, to the CFCA demonstrates how empirical models of complex hydrologic systems can be developed, disparate databases and models can be integrated, and study results can easily be disseminated to meet the needs of a broad range of end users.

Description of Study Area

The study area encompasses Lake, Orange, Osceola, Polk, and Seminole Counties in central Florida (fig. 1). The Orlando metropolitan area, approximately centered on the Interstate 4 corridor from northern Osceola to western Seminole Counties, encompasses many smaller towns and cities and is the major urbanized area in central Florida. The remainder of the study area is predominantly rural with a few, mostly widely scattered, but steadily growing towns such as Clermont, Eustis, Lady Lake, Lakeland, Leesburg, Oviedo, and Winter Haven. Land-surface altitudes range from less than 5 feet (ft) along the St. Johns River (from Orange County downstream) to more than 300 ft in Polk and Lake Counties.

Climate

The climate of the study area is classified as subtropical and is characterized by warm, typically wet summers and mild, dry winters. Maximum daily temperatures typically exceed 90 degrees Fahrenheit (°F) during the summer, but may fall below freezing for several days in the winter. The mean annual air temperatures for 1981–2010 are 72.7 °F at the city of Sanford in the northern part of the CFCA and 73.6 °F at the city of Bartow in the southern part of the CFCA (fig. 1) (National Oceanic and Atmospheric Administration, 2011). In January, mean monthly air temperatures are 59.7 °F at Sanford and 61.8 °F at Bartow, and in July are 83.3 °F at Sanford and 82.8 °F at Bartow. Mean annual rainfall at Sanford and Bartow is 53.04 and 52.08 inches (in.), respectively, for 1981–2010, with 55 and 59 percent, occurring during June through September (National Oceanic and Atmospheric Administration, 2011). Frequent thunderstorms during the summer as well as occasional tropical storms and hurricanes during the summer and early fall can bring heavy rainfall to the area. Winter rainfall generally is associated with large, frontal-type, cold air masses from the north.

Hydrogeologic Setting

The study area generally is underlain by unconsolidated sand and clay sediments that form the surficial aquifer system; less permeable clay and carbonate rocks that form the intermediate confining unit; more permeable carbonate rocks than those of the intermediate confining unit that form the intermediate aquifer system; a thick sequence of limestone and dolomitic limestone of variable permeability that forms the Floridan aquifer system; and low permeability dolomite and evaporite beds that form the sub-Floridan confining unit, which functions as the base of the fresh groundwater flow system (Miller, 1986). Generally the intermediate confining unit serves as a confining unit for the Floridan aquifer system, but where multiple permeable zones are present, the sequence of hydrogeologic units is referred to as the intermediate aquifer system. The intermediate aquifer system in the CFCA is located in the southern two-thirds of Polk County (Sepúlveda and others, 2012, p. 28). For the purpose of this report, the intermediate confining unit and intermediate aquifer system are discussed collectively. The Floridan aquifer system—the principal source of groundwater in the study area—is divided into two aquifers of relatively high permeability, referred to as the Upper Floridan aquifer and the Lower Floridan aquifer (Miller, 1986). These aquifers are separated by a less permeable unit called the middle confining unit I in east-central Florida. In west-central Florida, these two aquifers are separated by the middle confining unit II, which is composed of gypsiferous dolomite and dolomitic limestone of considerably lower permeability than the middle confining unit I in east-central Florida (Miller, 1986; O'Reilly and others, 2002). The reader is referred to Sepúlveda and others (2012) for a more detailed description of the hydrogeologic framework in the CFCA.

The aquifer system in the study area is recharged primarily by rainfall. Rainfall that exceeds ET, and does not become surface runoff, can recharge the aquifer system after infiltrating at land surface and percolating through the unsaturated zone. Sources of water to the aquifer system, in addition to net recharge from rainfall, are artificial recharge (for example, irrigation or rapid infiltration basins) and subsurface inflow from outside the study area. Inflow to the aquifer system in the study area is eventually discharged by springs, leakage to some surface-water bodies, well withdrawals, and subsurface outflow.

In a geologic setting where limestone is at or near land surface, net recharge interacts with the carbonate rocks, resulting in karst terrain. Karst is characterized by the absence of a well-defined surface drainage system and is drained internally, that is, rainfall not lost to ET infiltrates and recharges the aquifer. Internal drainage results in higher net recharge rates, which are conducive to the dissolution of limestone and the formation of such features as voids and conduits in the limestone and closed depressions at land surface. Numerous karst features, including sinkholes and springs, are present in the study area. Sinkholes in all stages of development are common throughout much of the study area and range from small depressions a few feet in diameter to large lakes. Sinkholes can be dry or water-filled, and many are areas of high recharge to the underlying aquifers. Numerous springs, located in the northern half of the study area, discharge water from the Upper Floridan aquifer into rivers and streams that eventually flow into the Atlantic Ocean.

The well-drained karst terrain in the western and central parts of the study area combined with the coastward topographic relief is sufficient to support regional flow in the Floridan aquifer system. In central Florida, the highest water levels in the Floridan aquifer system occur in the Upper Floridan aquifer in north Polk County; the lowest water levels occur in the Upper Floridan aquifer near springs, the St. Johns River, and the Atlantic Ocean. Water recharges the Upper Floridan aquifer at the highest rates in sandy ridge areas of west Orange, south Lake, and east-central Polk Counties as well as in more isolated, local areas where sandy soils with a

deep water table are common (Sepúlveda and others, 2012). In these recharge areas, water generally moves laterally in the Upper Floridan aquifer with a smaller amount moving downward into the Lower Floridan aquifer. In both the Upper and Lower Floridan aquifers, water generally flows to the northeast and east. Water exits the Lower Floridan aquifer by upward leakage to the Upper Floridan aquifer subsequently exiting the groundwater flow system by discharge at springs, upward leakage to the St. Johns River and other areas of lower land-surface altitude where it may contribute to streamflow or be removed by evaporation or withdrawal by plants (transpiration), and lateral flow to the Atlantic Ocean. Water also exits the Upper and Lower Floridan aquifers by withdrawals from water-supply wells. The reader is referred to Sepúlveda and others (2012) for a more detailed description of the conceptual groundwater flow system and water budget in the CFCA.

Previous Studies

Studies on the status and trends of the groundwater system over long-term (multidecadal) periods are limited for central Florida. Murch and Tara (2010) analyzed hydrologic and rainfall data from 1941 to 2009 for 120 sites in the CFCA to determine if long-term statistically significant trends are present in groundwater levels, lake water levels, spring discharge, and rainfall measurements using exploratory data analysis, trend analysis, and cluster analysis techniques. Other earlier general water-resource studies include some qualitative or quantitative analyses of long-term behavior of the hydrologic system in parts of central Florida, including Osceola County (Schiner, 1993), Seminole County (Spechler and Halford, 2001), Lake County (Knowles and others, 2002), Orange County (Adamski and German, 2004; German and Adamski, 2005), and Polk County (Spechler and Kroening, 2007). Basso (2003) examined the effects of groundwater withdrawals on groundwater levels and groundwater discharge to the upper Peace River Basin in south-central Polk County using estimated and measured data from the 1940s to the early 2000s. Osburn and others (2006) compiled period-of-record discharge measurements and used locally weighted scatterplot smoothing to indicate possible patterns in discharge over time for springs in the SJRWMD; an updated, online version of this publication is available at *http://floridaswater.com/springs/* (accessed April 2013). Intera (2007a, 2007b) developed regression models to predict daily discharge, starting as early as 1949 and extending as late as 2005, at several springs in central Florida.

Approach

In central Florida, a need exists to identify and quantify the salient factors influencing the behavior of the groundwater system, and several physics-based models have been developed for this purpose (Knowles and others, 2002; McGurk and Presley, 2002; Sepúlveda, 2002; Environmental Simulations Inc., 2007; Sepúlveda and others, 2012). However, complex interactions between the surface and subsurface environments in a karst terrain are difficult to simulate with physics-based models. Alternatively, substantial historical hydrologic data, which are well suited for empirical modeling, are available for central Florida in the databases of local, State, and Federal agencies.

The emerging field of data mining involves extracting valuable knowledge from large databases (Weiss and Indurkhya, 1998). Data mining comprises several technologies that include signal decomposition, advanced statistics, multidimensional visualization, machine learning/artificial intelligence, and chaos theory. Data mining can provide insight into complex problems and has been applied to an array of problems related to the interactions between natural and human-made systems. These interactions are becoming increasingly important as growing populations and development place heavier burdens on the environment. The application of data-mining technologies to surface-water and groundwater systems has been particularly successful. Models of hydraulic and water-quality parameters based in part on ANNs, a form of machine learning, have been accepted by State and Federal agencies for regulatory applications for streams in western Oregon (Risley and others, 2002), and the Beaufort (Conrads and others, 2003), Lower Savannah (Conrads and others, 2006), and Pee Dee (Conrads and Roehl, 2007) estuaries in South Carolina and Georgia. Other studies have focused on groundwater, surface-water, and water-quality applications of ANNs, such as the simulation of groundwater levels in the surficial aquifer system and Upper Floridan aquifer (Coppola, Szidarovszky, and others, 2003) and the karstic Edward's aquifer in Texas (Trichakis and others, 2011), groundwater management using multi-objective optimization (Coppola, Poulton, and others, 2003), Everglades hydrology (Conrads and Roehl, 2006) and water quality (Conrads and Roehl, 2010), stream temperatures in Wisconsin for fisheries management (Stewart and others, 2006), Upper Floridan aquifer water levels in the Suwannee River valley (Roehl, Risley, and others, 2006), Lake Okeechobee inflow volumes (Trimble and others, 2006), fate and transport of volatile organic compounds in groundwater at the Savannah River nuclear site in South Carolina (Conrads and others, 2007), dam-regulated wetland hydrology in South Carolina (Conrads and others, 2008), prediction of Upper Floridan aquifer spring flows (Sepúlveda, 2009), and specific conductance and chloride concentration in a tidal tributary of the Savannah River (Conrads and others, 2011). These studies have demonstrated that ANN models, combined with other data-mining techniques, can provide an effective approach for simulating complex hydrologic systems.

The variability of lake water levels, groundwater levels, and spring flows in the CFCA is a result of many factors, including rainfall and groundwater-use conditions. As part of the CFWI, a variety of empirical and physics-based modeling approaches have been conducted. The empirical modeling approach described in this report used correlation functions that were synthesized directly from data to predict lake water

levels, groundwater levels, and spring flows at selected sites in the CFCA in order to simulate response to rainfall and groundwater-use conditions. Hydrologic datasets at daily to semiannual measurement frequencies were available. Empirical lake water-level, groundwater-level, and spring-flow models were developed directly from these data by using ANN models and other data-mining techniques, such as signal (time series) processing methods including clustering, filtering, and signal decomposition. Because the intent of this study was to ascertain the salient factors influencing the behavior of the groundwater system as inferred from available historical data, restraint was exercised in prescribing assumptions based on previous models, whether empirical or physical, in the development of the ANN models. Given the empirical nature of data-driven correlation-based modeling, however, data were prepared in such a way to limit the potential confounding effects of cross correlations among explanatory variables, and the ANN models were developed to be consistent with hydrologic principles and the basic components of a water budget.

The application of data-mining techniques to develop empirical models to simulate lake water levels, groundwater levels, and spring flows and analyze response to rainfall and groundwater-use conditions was undertaken in four phases:

1. obtaining and evaluating the suitability of historical lake water-level, groundwater-level, spring-flow, meteorological, and groundwater-use data for developing the models;
2. developing ANN models to simulate the hydrologic response at selected sites with long-term record (at least 25 years);
3. developing the CFANN DSS to integrate historical databases and the site-specific ANN models, model controls, and model output into a spreadsheet application with a graphical user interface (GUI) and simulating four hypothetical scenarios to assess ANN model sensitivity; and
4. comparing results of the simulations made with the ANN models with those from the physics-based east-central Florida transient (ECFT) groundwater flow model developed by Sepúlveda and others (2012), using a spreadsheet application with a GUI that allows the user to select sites of interest for comparison.

Historical Data

Many Federal, State, and local agencies have collected data in the CFCA. For this study, data collected by the USGS, National Oceanic and Atmospheric Administration (NOAA), SJRWMD, SFWMD, SWFWMD, Orange County, and Seminole County were used for analysis and to develop ANN models. These data consist of hydrologic data, consisting of forcing (meteorological) and response (water level and flow), and groundwater-use data. Some meteorological data extend back to 1900 and some hydrologic response data extend back to the 1930s, whereas little groundwater-use data are available prior to 1957.

Hydrologic Data

Hydrologic data were compiled for sites in the CFCA (including all of Lake County) with a 10-year or greater period of record up through 2008, which includes data collected by the USGS, NOAA, SJRWMD, SFWMD, SWFWMD, Orange County, and Seminole County. One exception to this criterion is lake water levels measured by Seminole and Orange Counties. Owing to the lower frequency of collection of these data (typically monthly), only those lakes with a 20-year or greater period of record and six or more measurements per year were included in the analysis.

Historical Hydrologic Database

A Microsoft Access® database was created to store all historical hydrologic data. The database contains information for 963 sites comprising about 4.5 million records for 307 lakes, 470 wells, 22 springs, 143 rain gages, and 21 air temperature measurement sites (fig. 2). Additionally, at the 21 air temperature measurement sites, computed potential evapotranspiration (PET) data were included in the database. It is important to note that the hydrologic database developed for this study was constructed from data received from the collecting agencies during 2008 and 2009. Data could be revised by the collecting agency if changes in data processing techniques are made in the future. For this reason, it is noted here that spring-flow data were acquired from the SJRWMD (*http://floridaswater.com/toolsGISdata/*) on September 22, 2011, for the six springs for which ANN models were developed. The hydrologic database was updated with these data, but not for the other 16 springs. For future use of this database, the user is encouraged to verify that all data are consistent with those that are most recently published by the collecting agency.

Data in the hydrologic database were examined for apparent errors, and identified issues were resolved. Data for duplicate sites (sites monitored by more than one agency) were merged, where consistency in data quality among the collecting agencies could be verified, in order to increase data density and extend the period of record for individual sites. Duplicate sites generally had non-overlapping periods of record; however, when data were available for the same date at duplicate sites, the data value for the site with the longer period of record was retained. As a result, a single time series of data was created for each site. Simple statistics computed for the entire period of record for each site were examined for anomalies: period-of-record length in years, average number of measurements per year, minimum and maximum values, range, mean, linear slope, and Pearson product-moment correlation coefficient (R). Errors in site coordinates were identified and resolved by verifying the coordinates against known site

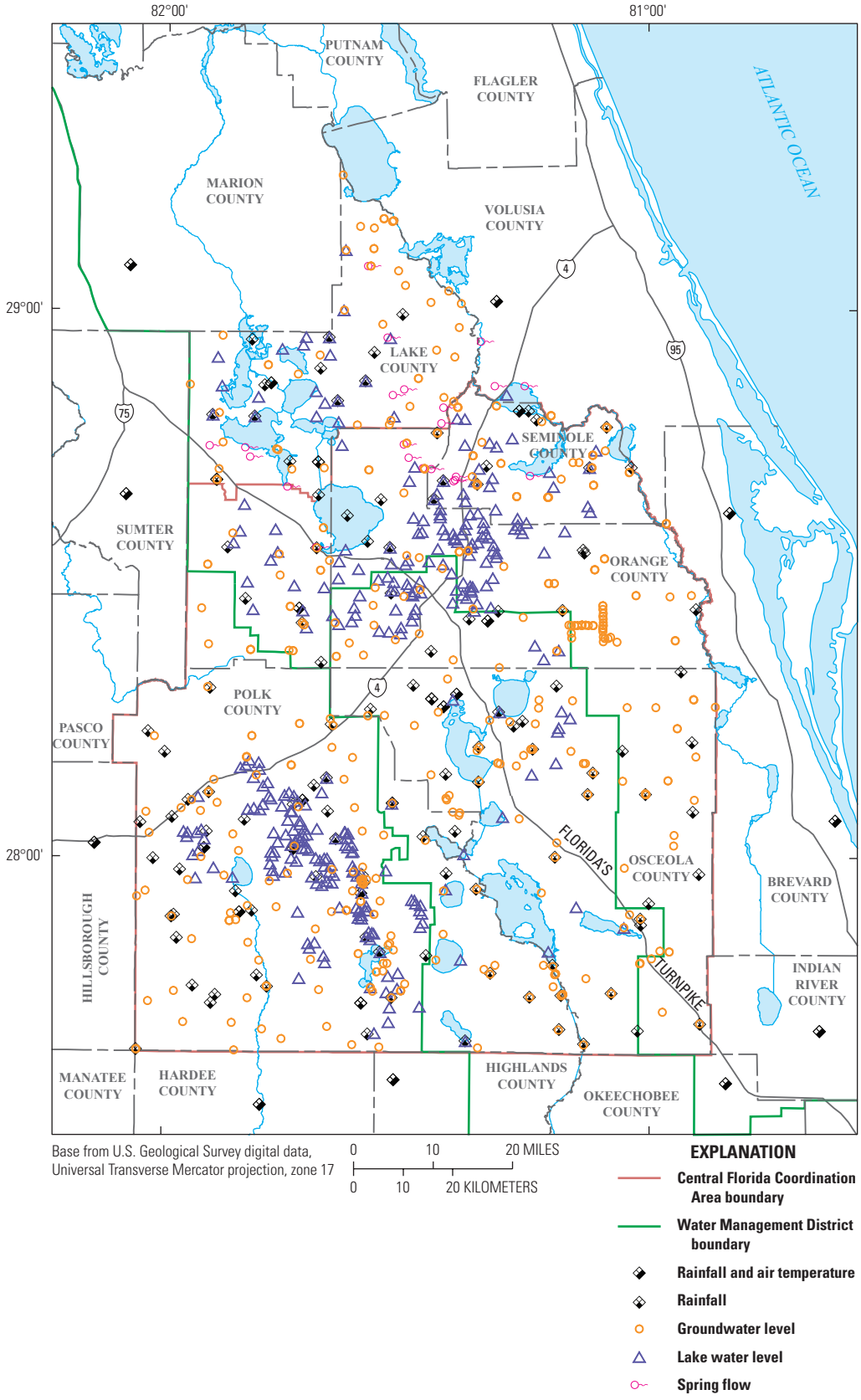


Figure 2. Locations of sites for which historical hydrologic data were acquired for this study.

locations. A conservative approach was taken in addressing anomalies attributable to apparent errors in the original data, in that erroneous data were deleted if a definitive correction could not be identified. Likewise, anomalous data that were not clearly erroneous were retained. Additional details on the organization of the hydrologic database in Access as well as information on how to obtain the database are provided in appendix 1, and a summary of the data contained in the database is provided in the following sections.

Rainfall, Air Temperature, and Potential Evapotranspiration

Climate is an important external forcing factor that influences the behavior of the groundwater system in the CFCA. Meteorological data acquired and assessed for use in this study were rainfall, air temperature, and PET. The most complete set of long-term daily rainfall data were available from NOAA. Twenty-three NOAA weather stations exist in or near the CFCA. Both rainfall and air temperature data are collected at 21 of the stations, whereas only rainfall is measured at the other two stations (fig. 2). Rainfall data at six of the NOAA stations (station identification numbers 2229, 3137, 4502, 6628, 7982, and 8942; see hydrologic database described in appendix 1) were provided by SJRWMD (David Clapp, written commun., 2009), and data at the other 17 stations were provided by SWFWMD (Jill Hood, written commun., 2009). For NOAA rainfall data, periods of record start in the period 1901–46 and end in the period 2006–08, and range in length from 61 to 108 years. For accumulated data values in the original NOAA records, which represent rainfall that was accumulated over 2 days to a month and reported as a single value, daily values were estimated by SJRWMD and SWFWMD by disaggregating using nearby rain gages. For station records provided by SJRWMD, periods of missing data were filled with estimates published by NOAA or computed by SJRWMD on the basis of data from nearby rain gages. For station records provided by SWFWMD, periods of missing data were filled with the estimates computed by Aly (2008). Rainfall data are also available for 120 additional sites operated by the SJRWMD, SFWMD, or SWFWMD (fig. 2). It is important to note that some of the SJRWMD, SFWMD, or SWFWMD rainfall sites report accumulated data that represent several days or even a month or more, which were not disaggregated into daily values. As a result, statistics computed for such rainfall sites would be skewed due to the treatment of accumulated data as daily data. All accumulated rainfall data are identified in the hydrologic database by a quality code assigned by the collecting agency (appendix 1).

Air temperature data were obtained directly from NOAA for the 21 weather stations (fig. 2). Daily measurements of minimum, maximum, and mean air temperature were available. Periods of record start in the period 1900–59 and end in the period 1961–2009, and range in length from 47 to 109 years, with 252 to 365 measurements per year.

Direct measurements or estimates of actual ET are sparse in the study area, and data that are available do not extend prior to 1993 and generally have periods of record less than 10 years (O'Reilly, 2007). As alternatives, PET data were acquired from the USGS Statewide ET dataset (*http://fl.water.usgs.gov/et/*) (accessed July 2010) and were computed using the temperature-based Hargreaves equation (Hargreaves and Samani, 1985). Because the USGS Statewide ET dataset is a spatial product providing data at 2-kilometer (km) resolution across the entire land area of Florida (Mecikalski and others, 2011), for the purposes of this study, data were extracted for the 2-km pixel overlying each of the 21 NOAA air temperature measurement sites. The USGS Statewide ET data are available from June 1, 1995, and data through 2009 were included in the hydrologic database. PET data estimated using the Hargreaves equation were computed using the entire period of record for the 21 NOAA air temperature measurement sites. Given its simplicity, Hargreaves-derived PET approximates actual ET reasonably well after multiplying by a crop factor (O'Reilly, 2007).

Lake Water Level

Lake water-level measurements were available for 307 lakes in the study area (fig. 2). Periods of record start in the period 1933–99 and end in the period 1968–2009, and range in length from 10 to 76 years, with an average of 5 to 365 measurements per year. The water-level range for individual lakes varies from 2.38 to 32.96 ft for their respective periods of record. Lakes were classified as either flowthrough (containing at least one stream inflow or outflow) or closed basin (containing no stream inflow or outflow) on the basis of existing maps and interpretation of aerial imagery. Slightly more than half of the lakes (164) were classified as flowthrough, and the remaining 143 were classified as closed basin (fig. 3).

Groundwater Level

Water-level measurements were available for 470 wells in the study area (fig. 2). Periods of record start in the period 1930–2000 and end in the period 1981–2009, and range in length from 8 to 79 years, with an average of less than 1 to 365 measurements per year. The water-level range for individual wells varies from 2.57 to 62.87 ft for their respective periods of record. Well-construction information typically was incomplete, but total depths were available for all but 27 wells. Of the remaining 443 wells, total depths ranged from 4 to 2,090 ft. Wells were classified as being completed in the surficial aquifer system, intermediate confining unit/intermediate aquifer system, Upper Floridan aquifer, Lower Floridan aquifer, or Floridan aquifer system on the basis of records of the collecting agency and well depth. Thus, of the 470 wells, 139 are in the surficial aquifer system, 40 are in the intermediate confining unit/intermediate aquifer system, 47 are in the Upper Floridan aquifer, 13 are in the Lower Floridan aquifer,

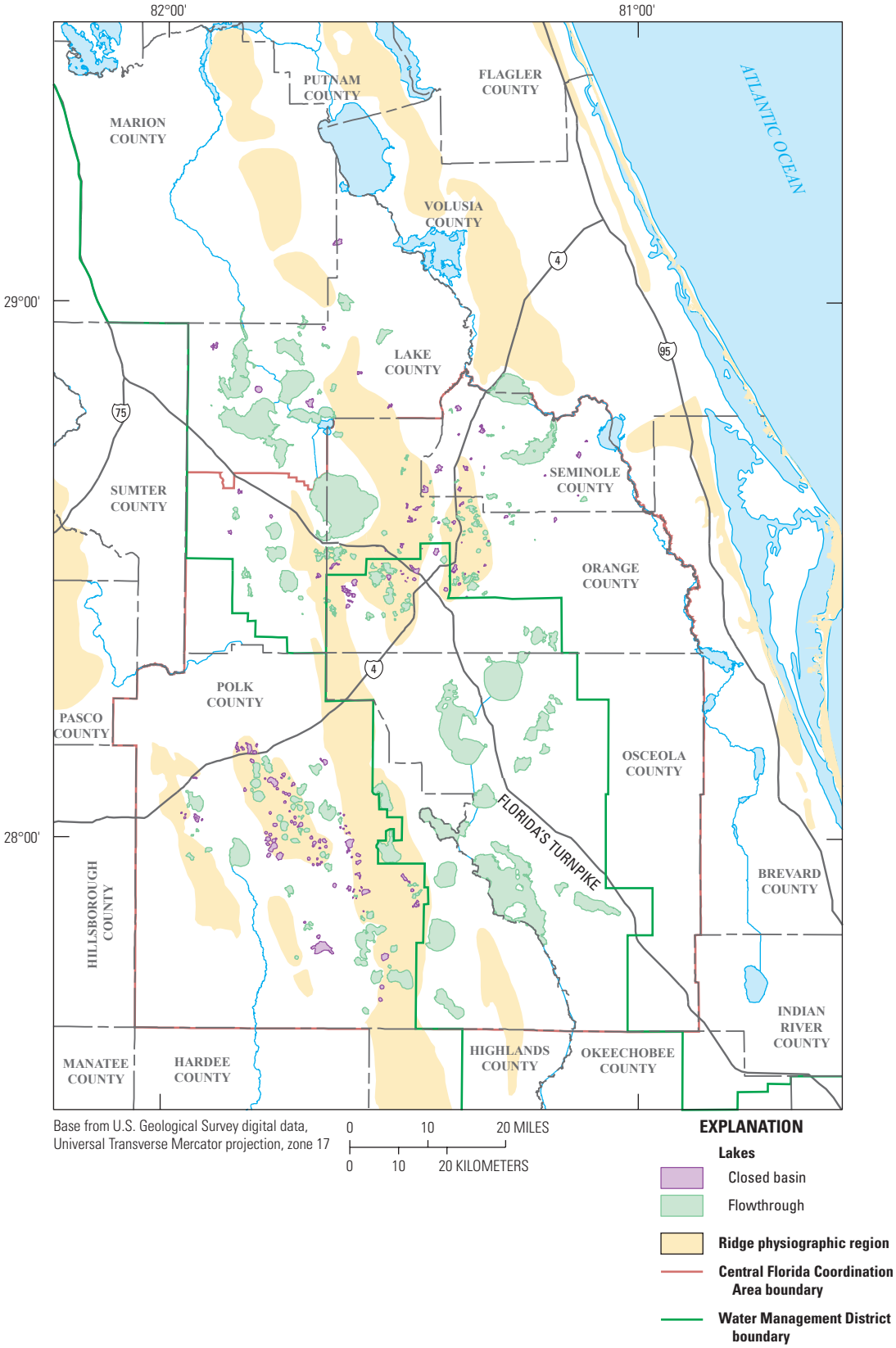


Figure 3. Lakes classified as either flowthrough or closed basin for which historical water-level data were acquired.

and 231 are in the Floridan aquifer system (known to be in the Floridan aquifer system, but undetermined whether in the Upper Floridan aquifer, Lower Floridan aquifer, or both).

Spring Flow

Spring-flow measurements were available for 22 springs in the study area (fig. 2). Periods of record start in the period 1931–91 and end in the period 1995–2009, and range in length from 18 to 79 years, with an average of less than 1 to 40 measurements per year. Mean (arithmetic average) spring flow for individual springs ranges from 0.8 to 149 cubic feet per second (ft^3/s), whereas the spring-flow range for individual springs varies from 0.6 to 169 ft^3/s for their respective periods of record. The temporal data density for springs is extremely variable. For an individual spring, it is not unusual to have only one flow measurement per decade in the early part of the period of record and daily flow data for more recent periods.

Groundwater Use

Groundwater-use data were compiled for the categories public supply, citrus irrigation, non-citrus agricultural irrigation, golf course irrigation, phosphate mining, and drainage well recharge. Monthly data were compiled for all categories, although for some categories only annual estimates were available and were disaggregated into monthly values. For sites for which data were available or could be estimated, data were compiled as far back as 1950, although for most categories data are sparse or poorly estimated for years prior to 1957. Groundwater use was not categorized by aquifer for this study, although Sepúlveda and others (2012) report that more than 95 percent of the total groundwater withdrawn from 1995 through 2006 in east-central Florida was from the Floridan aquifer system. Surface-water-use data were not analyzed in this study.

The various types of groundwater withdrawal rates or direct water inflow (drainage well recharge) data were provided by the SJRWMD, SFWMD, SWFWMD, and the Florida Department of Environmental Protection (FDEP) and supplemented with data available from the USGS. These data represent the best available data at the time of this study. For future analyses, the user is strongly encouraged to verify that all data are consistent with those that are most recently reported by the collecting agency. For some categories, the reported and estimated water withdrawals and inflows represent amounts for all of the CFCA and small portions of surrounding areas. Further details are provided in the following sections on the reported data as well as the methodologies used to make estimates for categories without reported data or for periods of missing data.

Public Supply

The public water supply category includes all permit holders that pump from wells with a 4-inch or larger diameter at a rate equal to or greater than 100,000 gallons per day. This category represents all water distributed from the water-treatment facilities for public water supply on a monthly basis. Reported data generally were available for the years 1978 to 2008 at the water-treatment plants shown in figure 4, but such data were sparse prior to 1978 and were estimated for the years 1950 through 1977.

Data on groundwater withdrawals for public water supply for SJRWMD and SFWMD were gathered from FDEP Monthly Operating Reports (MORs) for 1978 through 2008, and periods of missing record were filled by linear interpolation prorated by average monthly flow as a percentage of annual total flow for 1978–2008 (Katherine Graf, University of Florida, written commun., 2011). Groundwater withdrawals for public water supply in the SWFWMD area were compiled using reported data from the SWFWMD Water Use Permit (WUP) records for most water-treatment plants or from estimates for a few plants (Jill Hood, Southwest Florida Water Management District, written commun., 2010). These public water-supply groundwater withdrawal data for the CFCA and vicinity are provided by Munch (2014), including documentation of the methods used to estimate missing data. As described later in the report, all groundwater-use data were spatially aggregated into 20 × 20-km grid cells for input into the ANN models; therefore, only total monthly water amounts delivered from a water-treatment plant were compiled and no attempt was made to distribute this total flow to active wells that supply the plant. After aggregating the data into 20 × 20-km grid cells, public water-supply groundwater withdrawal data from 1950 through 1977 were estimated by extrapolation and comparison with the limited amount of reported data during this period, as described later in the report.

Agricultural

Estimates of groundwater withdrawals for citrus irrigation, non-citrus agricultural irrigation, and golf course irrigation developed by McLeod and Munch (2012) for the CFCA, including all of Lake County, were used in this study. Even though golf course irrigation is not traditionally categorized as an agricultural water use, McLeod and Munch (2012) applied the same methodology to estimate groundwater withdrawals for golf course irrigation as that used to estimate withdrawals for citrus and non-citrus agricultural irrigation. Monthly groundwater withdrawal data were available for citrus and non-citrus agricultural irrigation for the period 1957–2010 and for golf course irrigation for the period 1977–2010 in the areas shown in figure 5. The methodology applied by McLeod and Munch (2012) involved the following procedures:
(1) aggregating annual total water use by county using data compiled by the SJRWMD, SFWMD, SWFWMD, and

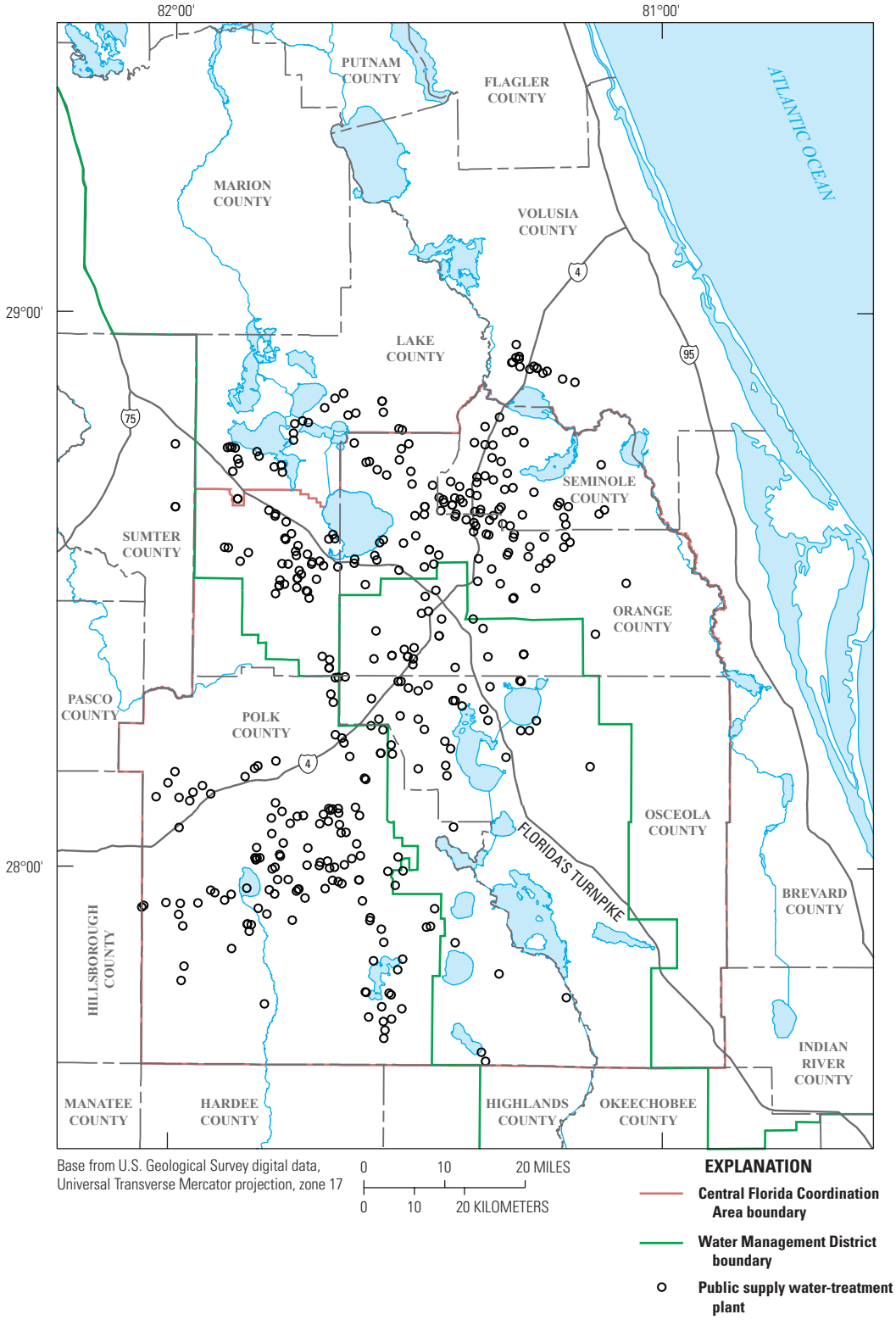


Figure 4. Locations of water-treatment plants in the Central Florida Coordination Area and vicinity for which public water-supply groundwater withdrawal data from 1978 through 2008 were available.

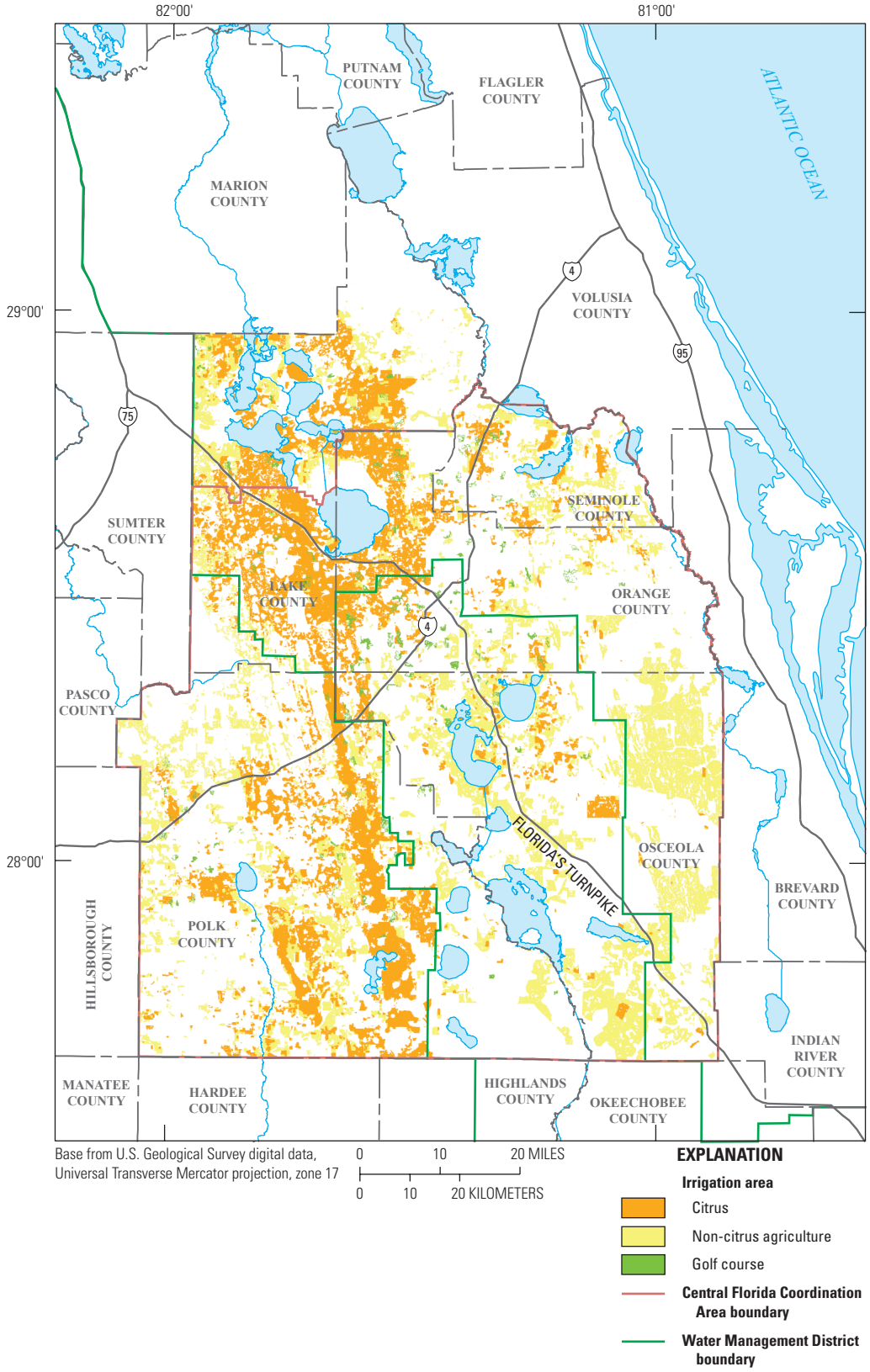


Figure 5. Locations of citrus, non-citrus agricultural, and golf course irrigation areas in the Central Florida Coordination Area and vicinity where groundwater withdrawal estimates were available. Areas denote where irrigation occurred sometime during the analysis period for each cover type (1957 to 2010 for citrus and non-citrus agriculture and 1977 to 2010 for golf course) as estimated by McLeod and Munch (2012); some areas did not receive irrigation during all years owing to changes in agricultural practices and land development.

USGS; (2) interpolating and extrapolating the total water use into groundwater use based on the best available sources (generally the National Agricultural Statistics Service [NASS] for citrus irrigation and the SJRWMD, SFWMD, SWFWMD, and USGS for non-citrus agricultural and golf course irrigation); (3) disaggregating the total use spatially by proportionally distributing the use according to estimated acreage polygons, which were based on spatial land-use and land-cover data available from the Florida Geographic Data Library (*http://www.fgdl.org*), NASS, SJRWMD, SFWMD, SWFWMD, and USGS; and (4) disaggregating the data temporally to yield monthly groundwater use based on monthly irrigation schedules for recommended crop irrigation requirements. Given the paucity of spatial data on historical irrigated acreage, McLeod and Munch (2012, p. v) state that "…only the groundwater use estimates generated from these models are considered to be accurate (on a semi-regional scale); not the polygon acreages." This limitation of the data is another reason why all groundwater-use data were aggregated into 20 × 20-km grid cells, as described later in the report.

Phosphate Mining

Phosphate mining in the CFCA has been confined to the southwestern part of Polk County (fig. 6). Water used by the phosphate industry consists of that required for the mining (ore extraction) and phosphate chemical production processes. Groundwater withdrawals for both categories of water use by the industry were compiled from reported data for 1978–2010 on a monthly basis (Jill Hood, Southwest Florida Water Management District, written commun., 2011) and well locations from the SWFWMD WUP records (fig. 6). No WUP data were available for years prior to 1978, and groundwater withdrawals for mining and chemical production were estimated separately on the basis of annual estimates of total water use by the phosphate industry, using the stepwise methodology described below (Jill Hood, Southwest Florida Water Management District, written commun., 2011):

1. The 5-year average ratio of groundwater to surface-water use of 96 percent was calculated for the phosphate industry using reported 1978–82 data from SWFWMD WUP records. This period was chosen on the basis of the assumption that as the phosphate industry started using alternative sources, the 5-year average ratio would change over time.

2. The average monthly variability of groundwater use for mining and chemical production WUPs were calculated separately. The use of water for mining generally showed more seasonal variations.

3. The 5-year average ratio of groundwater use for mining (69 percent) and chemical production (31 percent) to total groundwater use by the phosphate industry was calculated using reported 1978–82 data from SWFWMD WUP records.

4. The estimated annual water use for the phosphate industry for 1935–74 reported by Robertson and others (1978, p. 12) was used. The groundwater ratio (96 percent) established in step one was applied to the 1935–74 data, followed by linear interpolation between estimated 1974 and reported WUP 1978 groundwater-use data.

5. Ratios were initially applied to the total annual groundwater use for mining and chemical production from step three, followed by ratios representing monthly variability for mining and chemical production from step two.

6. For spatial distribution of chemical production groundwater use, the WUPs active in 1978 were assumed to have been active since 1935. The contributing ratio for each WUP was calculated and used to prorate the total chemical production groundwater use. The resulting estimated withdrawal was placed at the centroid of the wells associated with each WUP (fig. 6).

7. For estimating the spatial distribution of mining groundwater use, spatial data from FDEP showing historical evolution of mining areas were used. These data were available for the periods up to 1930, 1940, 1950, 1960, 1965, 1970, 1975, and 1980. The contributing ratio was calculated for each parcel with the same mining year. For years with no data, the ratio was assumed to be the same as that for the following year. The contributing ratio for each mining parcel was used to prorate the total mining groundwater use, and the resulting estimated withdrawal was associated with each respective parcel (fig. 6).

Drainage Wells

From the early 1900s to the 1960s, drainage wells were drilled in central Florida primarily for the purposes of disposing of stormwater runoff and controlling lake levels (Kimrey and Fayard, 1984; Bradner, 1996). In the CFCA, direct inflow to the Upper Floridan aquifer takes place through 210 active drainage wells, which are concentrated in the Orlando metropolitan area (fig. 6). The status and locations of these wells were verified by Hartman and Associates, Inc. (2003). These wells, which are cased to the top of the Upper Floridan aquifer and then drilled open-hole into the Upper Floridan aquifer, either receive street runoff from storm drains (street drainage wells) or control lake water levels (lake-level control wells). The estimated monthly recharge to the Upper Floridan aquifer was computed for each drainage well using a water-budget analysis and empirical runoff coefficients developed by Sepúlveda (2002, p. 34–35). The surface drainage areas of these wells, which were delineated by Sepúlveda and others (2012) on the basis of a digital elevation model, were used with monthly rainfall intensity (from the NOAA Orlando gage, station identification number 6628) and runoff coefficients (0.478 for street drainage wells and 0.225 for lake-level control wells) to compute the monthly volume of water

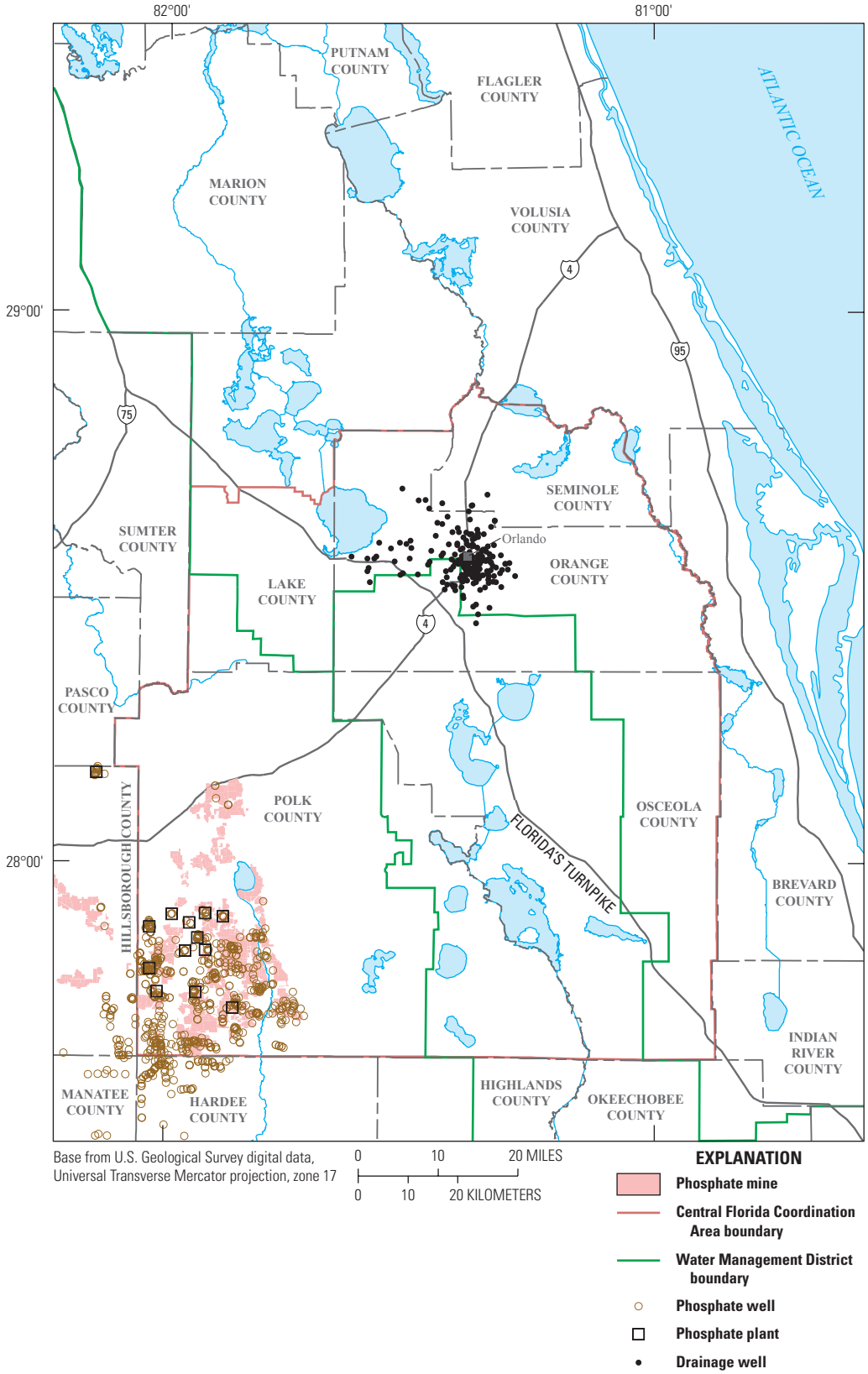


Figure 6. Locations of phosphate industry wells for which reported groundwater withdrawal rates from 1978 through 2010 were available, phosphate mining areas and phosphate chemical plants where groundwater withdrawal estimates from 1935 through 1977 were available, and drainage wells in the Central Florida Coordination Area and vicinity.

contributing recharge to the Upper Floridan aquifer from 1950 through 2008. Therefore, the estimated recharge rates to the Upper Floridan aquifer at drainage wells are strongly correlated to the amount of total rainfall for any period. Sepúlveda and others (2012) developed more refined drainage-well recharge rates using the Green-Ampt infiltration equations (Chow and others, 1988), but these data are only available for the period 1995–2006 and were not used in this study because of the short period of record. Due to the lack of historical data, for the purposes of the estimates used in this study, the number of drainage wells was not changed over time from the 210 wells inventoried by Hartman and Associates, Inc. (2003).

Land Use and Land Cover

Landscape change resulting from urbanization may be an important determinant of hydrologic change (Renken and others, 2005). Because of paucity of historical data on land cover and the coarse temporal resolution of the data that are available, landscape change was not included as a quantitative factor in this study. In order to illustrate the potential for landscape-change-induced impacts on central Florida hydrology, however, a brief comparison of historical land-use and land-cover data was made.

Comparison of land-use and land-cover data for the years 1977 and 2006 indicates substantial changes in the study area, particularly regarding the extent of developed areas (fig. 7). It is important to note that such changes may be due in part to differences in land-use and land-cover data sources as well as to differences in the interpretation and classification techniques. The 1977 data were compiled from high-altitude aerial photographs and the minimum digitized polygon size generally was 4 hectares (U.S. Geological Survey, 1990), whereas the 2006 data were derived from Landsat images and released in a raster format at 30-meter resolution (equivalent to 0.09 hectare) (Fry and others, 2011). For this reason, only the following more generalized land-use and land-cover categories were directly compared: developed areas (low-, medium-, and high-intensity urban areas and associated open areas, such as lawns, parks, and aesthetic landscaping), cropland/pasture, open water, and other areas (barren land, forest, scrub/grassland, and wetland). In 1977, the study area (6,019 square miles (mi^2) covering all of Lake, Orange, Osceola, Polk, and Seminole Counties) consisted of 7 percent developed land, 35 percent cropland/pasture, 9 percent open water, and 49 percent other areas. In 2006, the study area consisted of 18 percent developed land, 21 percent cropland/pasture, 9 percent open water, and 52 percent other areas. These results indicate the extent of increased urbanization for this 29-year period during which the area of developed land increased by about 680 mi^2, representing a change of about 160 percent, and the area of cropland/pasture decreased by about 810 mi^2, representing a change of about –40 percent (fig. 7). During a similar 30-year period (1980–2010), population in the study area increased from about 1.12 million to 2.74 million, or about 140 percent (Florida Office of Economic and Demographic Research, 2012). Increases in the area of developed land are particularly important because such changes could cause an increase in runoff as a result of reduced infiltration through impervious surfaces, soil compaction during development activities, or installation of stormwater drainage systems. In contrast, depending in part on the design of the stormwater management system, such as the use of stormwater infiltration basins rather than detention basins or ditch/canal systems that eventually direct overland runoff to surface-water bodies, urbanization could result in an average increase in infiltration on an areal basis.

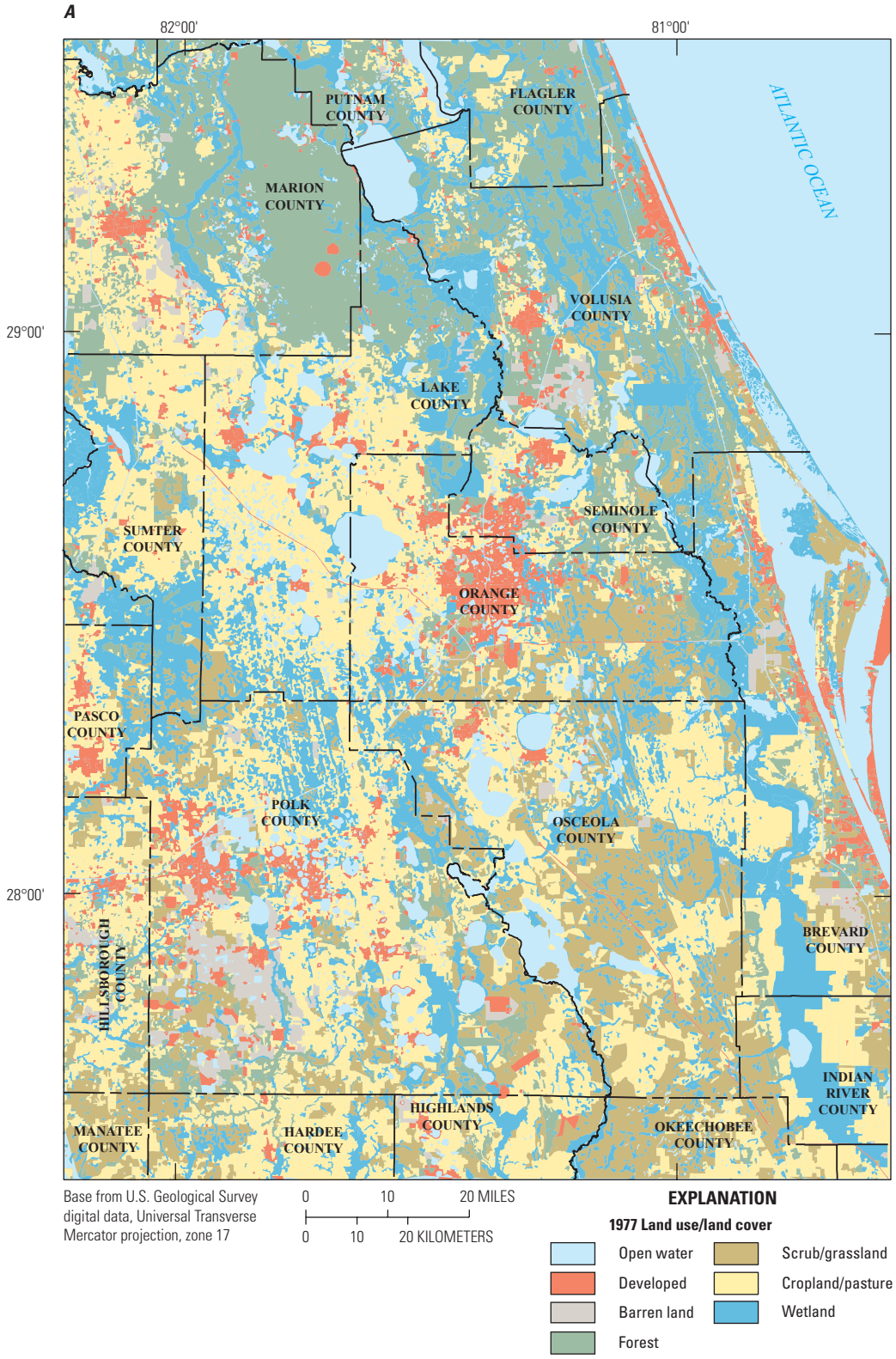


Figure 7. Generalized land use and land cover in the study area for *A*, 1977 (U.S. Geological Survey, 1990), and *B*, 2006 (Fry and others, 2011).

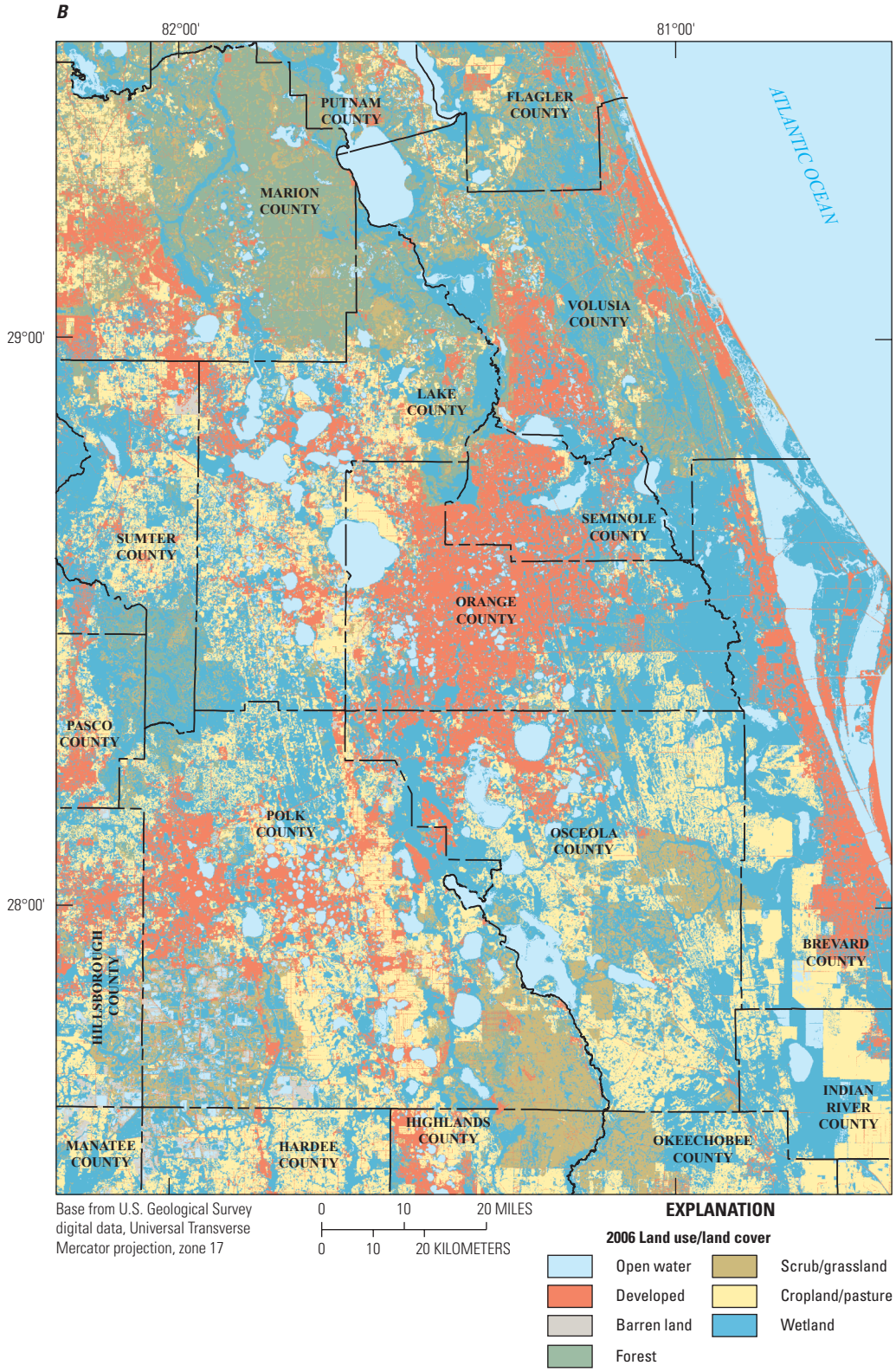


Figure 7. Generalized land use and land cover in the study area for *A*, 1977 (U.S. Geological Survey, 1990), and *B*, 2006 (Fry and others, 2011).—Continued

Characterization of Historical Data

Historical data were examined using descriptive statistics, cluster analysis, and other exploratory analysis techniques to assess their suitability for more intensive data-mining analysis. Based on these analyses and input from SJRWMD and SWFWMD on sites of priority water management concern, a subset of lakes, wells, and springs (51 sites in all) was selected for the development of ANN models.

Hydrologic Data

Hydrologic data were available for a wide range of time periods and temporal resolutions. These data were examined by computing range and linear trend for the entire period of available data for each site with at least 10 years of data consisting of at least 10 measurements. Cluster analyses were performed for sites with longer periods of record.

Range and Linear Trend

In order to better understand long-term variability in the historical hydrologic data, ranges and linear trends of the time series were computed. The range of a time series is a measure of the maximum fluctuation in an observed parameter over a specified period and is computed as the difference between maximum and minimum values. For each site, ranges were computed for the period of record for the purposes of quality assuring the dataset, wherein an anomalous value could be identified from an unusually large range. For rainfall sites, unusually large maximum values can be present in data accumulated over long periods, such as in some of the rainfall datasets collected by SJRWMD, SFWMD, or SWFWMD. Thus, for the purposes of further analysis in this study, only NOAA daily rainfall data were examined. The linear trend of a time series is the slope of a linear regression line computed for the period of record data. For this simple exploratory analysis, no attempt was made to identify cyclical, nonmonotonic, or statistically significant trends.

Ranges of water-level data (in lakes or wells) varied widely across the study area. Variation in lake water levels ranged from 2.38 to 32.96 ft with a mean of 8.43 ft and median of 7.22 ft. Some spatial trends are apparent. Of all 307 lakes, 59 percent are in ridge physiographic regions comprising 32 percent closed-basin and 27 percent flowthrough lakes (figs. 8, 9). In contrast, 26 percent of the 307 lakes have a historical range in water level greater than 10 ft, of which 19 percent are in ridge physiographic regions comprising 12 percent closed-basin lakes and 7 percent flowthrough lakes (fig. 9). These results indicate that closed-basin lakes in ridge areas consistently have greater variation in water levels than do flowthrough lakes in ridge areas (fig. 9), perhaps as a result of the lack of surface-water inflows or outflows. Additionally, both closed-basin and flowthrough lakes in ridge areas generally have greater variation in water levels than lakes in non-ridge areas, suggesting that lakes in ridge areas are more susceptible to large changes in water level as a result of greater hydraulic connection with the Floridan aquifer system. Ranges of groundwater levels exhibit larger variations, from 2.57 to 62.87 ft with a mean of 13.23 ft and median of 10.41 ft. Spatial trends in groundwater levels are due to spatial variations in groundwater withdrawals, recharge, and hydrogeologic conditions (fig. 10). The largest ranges of groundwater levels occurred in southwest Polk County, likely attributable in part to historical changes in water-use practices by the phosphate industry (Basso, 2003; Spechler and Kroening, 2007).

Variation in spring flows ranged from 0.9 to 169 ft^3/s with a mean of 32.8 ft^3/s and median of 13.8 ft^3/s. The range of flow for an individual spring generally is greater for large magnitude springs relative to that for small springs (fig. 11). However, spring-flow ranges as a percentage of mean flow are not strongly correlated with mean flows (R of –0.15).

Linear trend statistics of time series for each data type computed for the respective periods of record indicate a mixture of upward and downward trends (table 1). Meteorological data indicate 67 percent of sites exhibited upward trends in air temperature, whereas 76 percent exhibited downward trends in rainfall. Hydrologic response data indicate about an even split between upward and downward trends in lake water levels, whereas water levels in 69 percent of wells and flows in 68 percent of the springs exhibited downward trends. Spatial patterns are apparent in some data. Upward trends in air temperature occur throughout the study area, but downward trends occur only at some sites in the central portion of the study area (fig. 12). Downward rainfall trends occur throughout the study area, and upward trends occur only at five sites scattered across the study area (fig. 13). For lake water levels, closed-basin lakes in ridge areas consistently have stronger linear trends, whether upward or downward, than flowthrough lakes in ridge areas, perhaps a result of lack of surface-water inflows or outflows (figs. 14, 15). Additionally, both closed-basin and flowthrough lakes in ridge areas generally have stronger upward or downward trends in water levels than lakes in non-ridge areas, suggesting that lakes in ridge areas are more susceptible to long-term changes in water level as a result of greater hydraulic connection with the Floridan aquifer system (figs. 14, 15). Trends in groundwater levels generally are more pronounced than those in lake water levels, with downward trends throughout the study area (fig. 16). Upward trends in groundwater levels, although fewer in number, also occur throughout the study area. The strongest upward trends were computed for wells in southwest Polk County (fig. 16), likely attributable in part to historical changes in water-use practices by the phosphate industry (Basso, 2003; Spechler and Kroening, 2007). The linear trend in flow for an individual spring, whether upward or downward, generally is greater for large magnitude springs than for small springs (fig. 17). However, the magnitudes of spring flow trends (absolute values of the slopes of the linear trend lines for each spring) as a percentage of mean flow are weakly inversely correlated with mean flows (R = –0.33), indicating smaller springs tend to have trends in

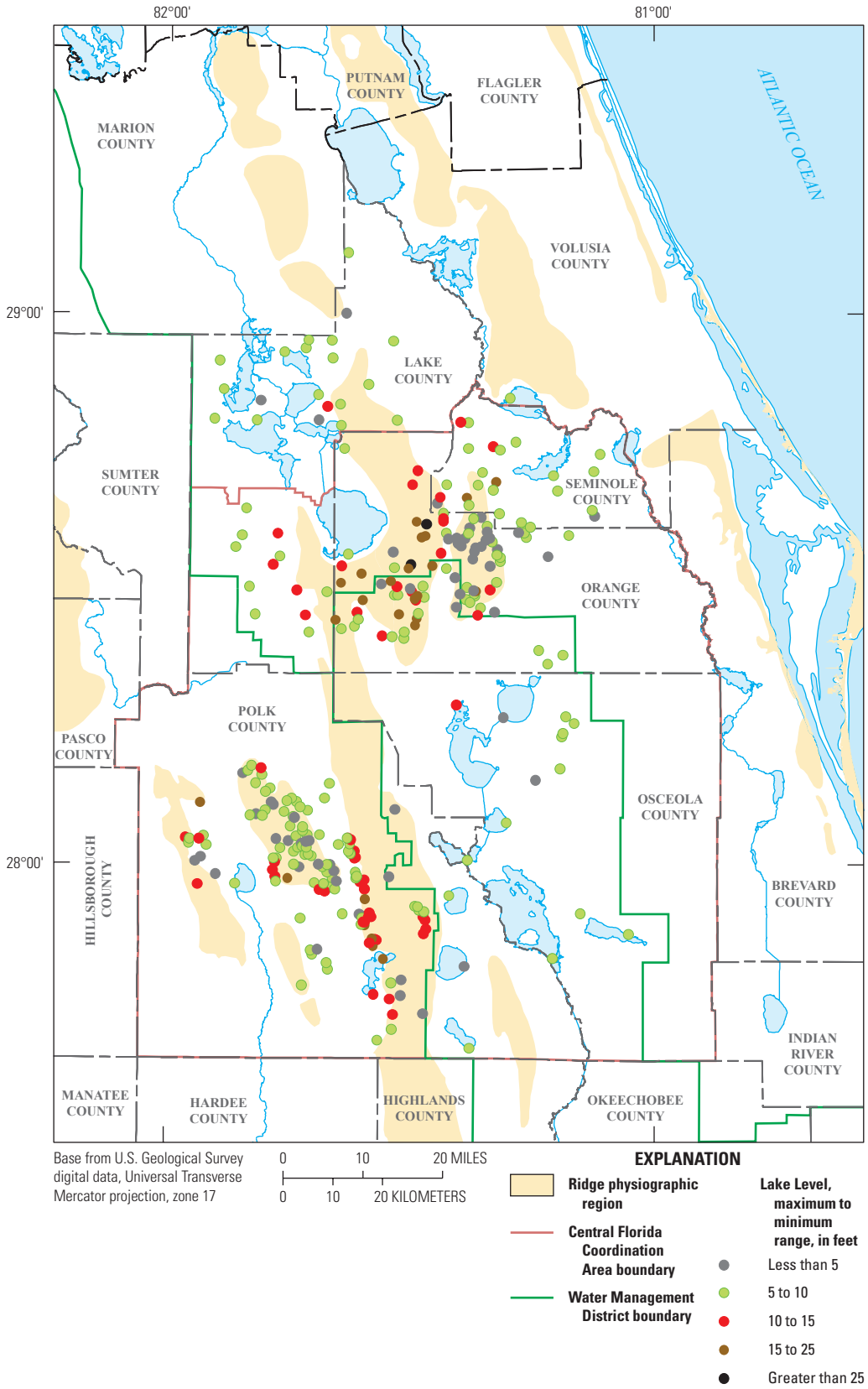


Figure 8. Range of daily water level for lakes in the study area with at least 10 measurements and a period of record of at least 10 years.

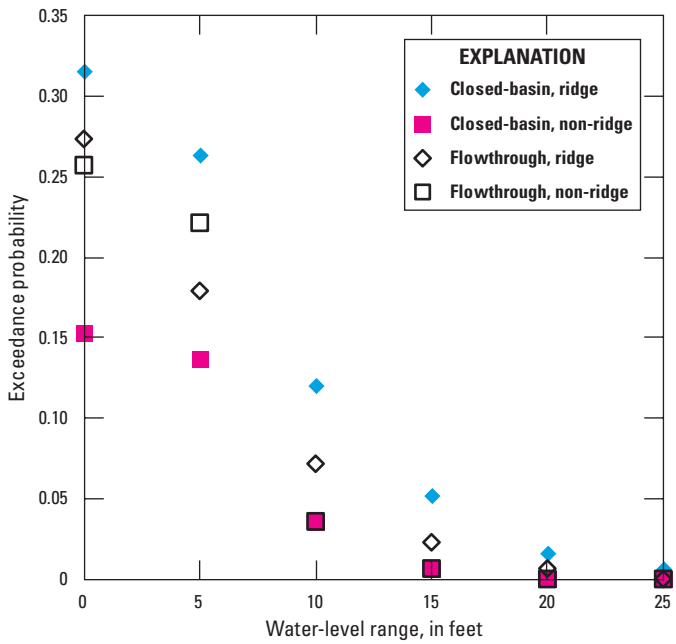


Figure 9. Exceedance probability of range in daily lake water level by lake type (closed basin or flowthrough) and physiography (ridge or non-ridge physiographic region) for lakes in the study area with at least 10 measurements and a period of record of at least 10 years.

flow, whether upward or downward, that are proportionately greater on the basis of flow magnitude than those of larger springs. The analyst must exercise caution in comparing trends among these sites because of large differences in periods of record, measurement frequency, and hydrologic setting. Furthermore, the intent of the linear trend analysis described here is for exploratory purposes only and should not be construed as a statistical analysis. No attempt was made to assess normality of the data or imply statistical inference. Murch and Tara (2010) presented a thorough statistical analysis of trends in hydrologic data for 115 sites in the CFCA. Nevertheless, the data ranges and linear trends described above indicate that substantial variability exists in the hydrologic system in the study area.

Time-Series Cluster Analysis

A time-series cluster analysis was performed for each type of hydrologic response data (lake water level, groundwater level, and spring flow) to identify dynamic similarities and differences among sites of the same type. The criteria for selection of sites for cluster analysis was that the site had at least 30 years of data and that no more than 25 percent of the record was missing during the period 1976–2008 (data collection at some sites began well before 1976). For wells, daily values of water level were required, but for springs and lakes, monthly or more frequent data values were permitted. A total of 176 sites—consisting of 121 lakes, 46 wells, and 9 springs—were used for cluster analysis.

A multistep process was followed to prepare and cluster each type of data. First, each time series was processed by applying a smoothing function, consisting of a trailing running average hereafter called a moving window average (MWA). The window size was selected on the basis of the original data frequency: a 15-day MWA with 15-day resampling for wells and a 30-day MWA with 30-day resampling for lakes and springs. Next, the data were centered by computing the difference from a "standard" time series (arithmetic average by time step of all the time series for that data type), leaving the residuals that accentuate differences between signals. Finally, a correlation matrix (values of R) of the residuals was constructed as a measure of the dynamic similarity among sites, and the k-means clustering technique was applied to the matrix. The k-means clustering technique provided in the Data Miner Software Kit of Weiss and Indurkhya (1998) was implemented, which is based on the algorithm of Hartigan and

Table 1. Linear trend statistics for air temperature, rainfall, lake water levels, groundwater levels, and spring flow for study area sites in central Florida.

[Temperature sites have periods of record of at least 45 years; rainfall sites have periods of record of at least 95 years; other sites have periods of record of at least 10 years; °F/yr, degrees Fahrenheit per year; (in/d)/yr, inches per day per year; in/yr, inches per year; (ft^3/s)/yr, cubic feet per second per year]

	Number of sites	Percentage of sites		Linear trend			
		Increasing	Decreasing	Minimum	Maximum	Mean	Median
Temperature, °F/yr	21	66.7	33.3	–0.014	0.052	0.007	0.003
Rainfall, (in/d)/yr	21	23.8	76.2	-1.36×10^{-4}	4.72×10^{-4}	-6.56×10^{-6}	-4.02×10^{-5}
Lake water level, in/yr	307	48.5	51.5	–2.85	13.26	0.305	–0.0111
Groundwater level, in/yr	455	31.4	68.6	–25.83	22.97	–0.431	–0.381
Spring flow, (ft^3/s)/yr	19	31.6	68.4	–0.599	0.238	–0.075	–0.036

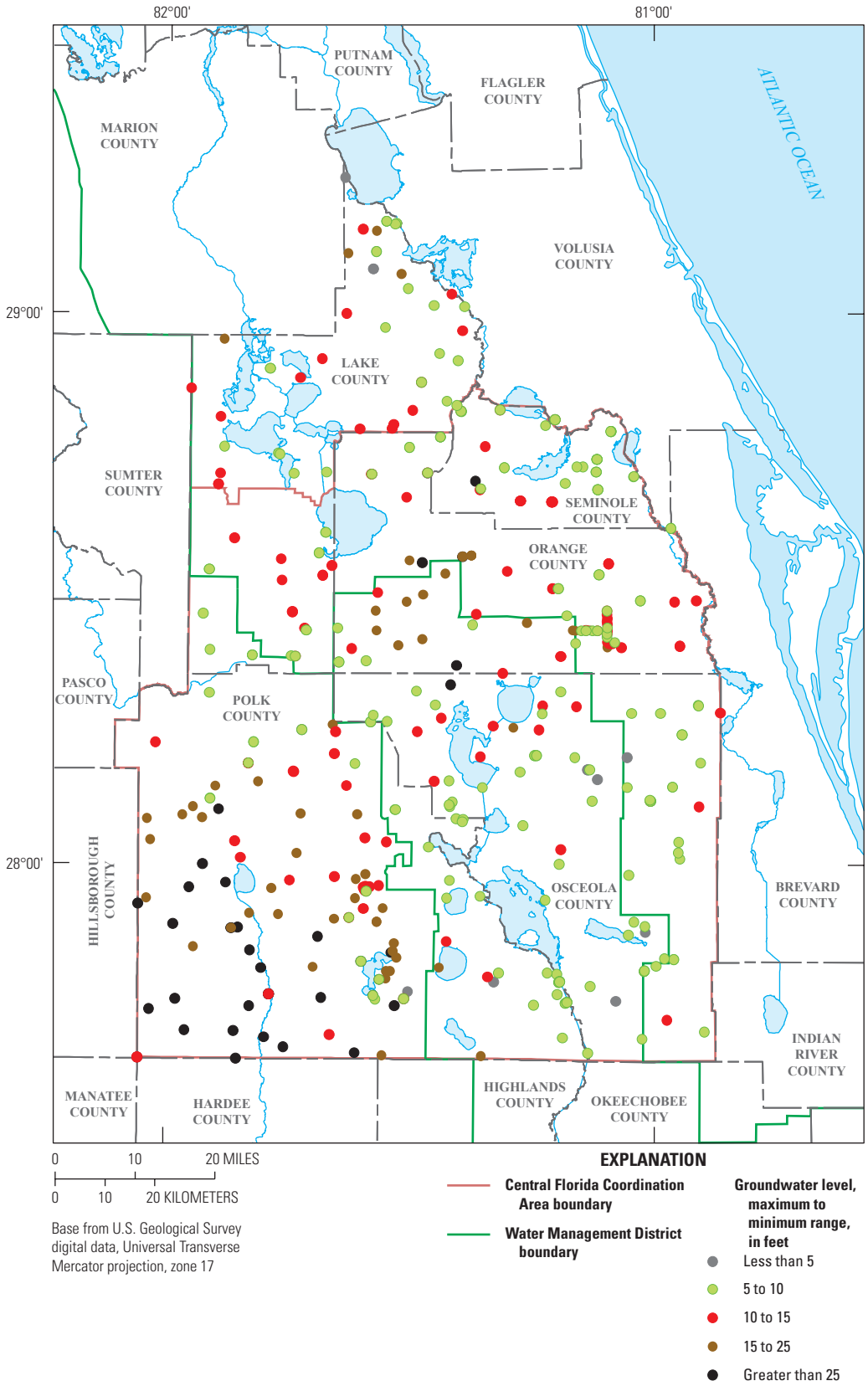


Figure 10. Range of daily groundwater level for wells in the study area with at least 10 measurements and a period of record of at least 10 years.

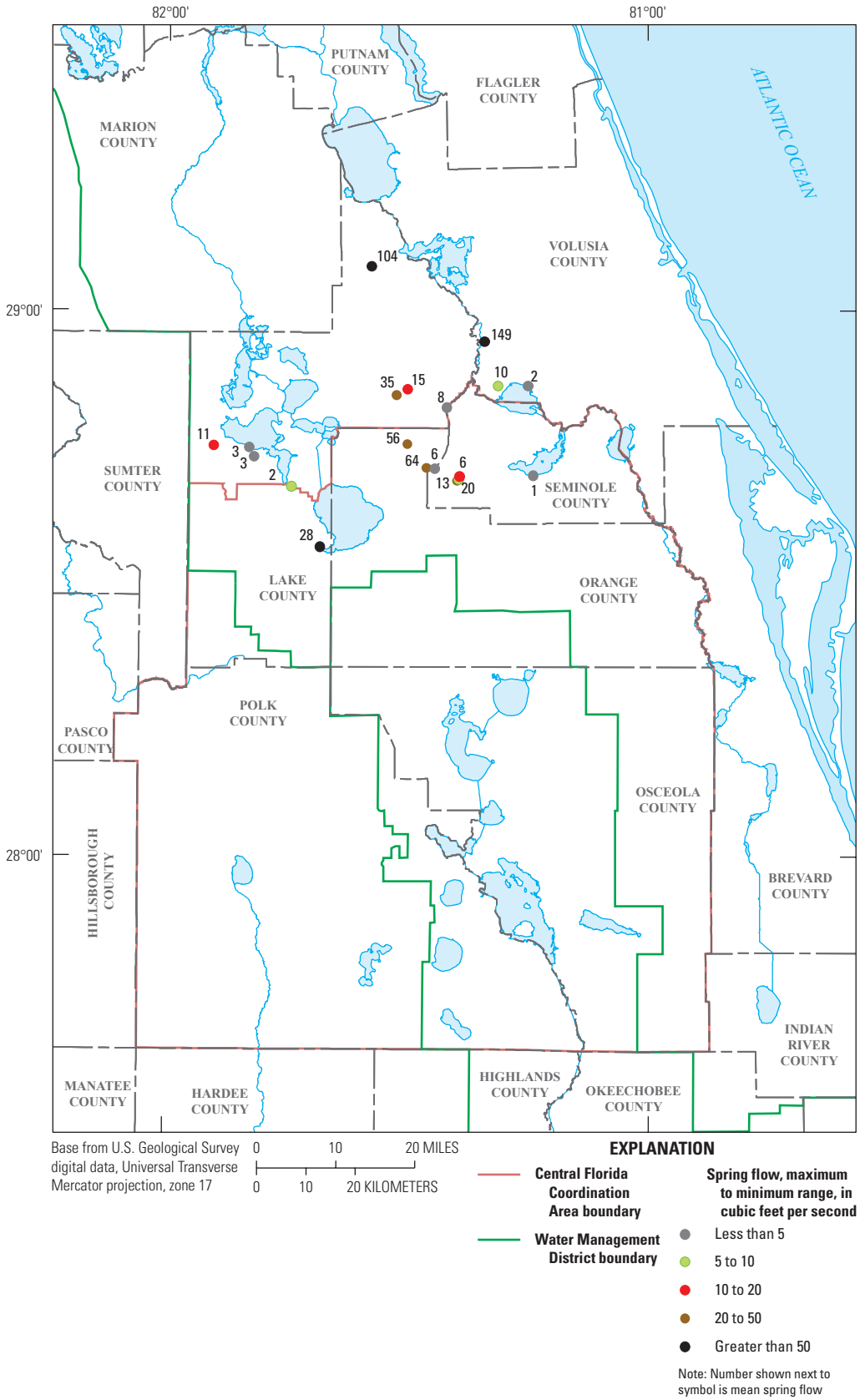


Figure 11. Range of daily flow and mean flow for springs in the study area with at least 10 measurements and a period of record of at least 10 years.

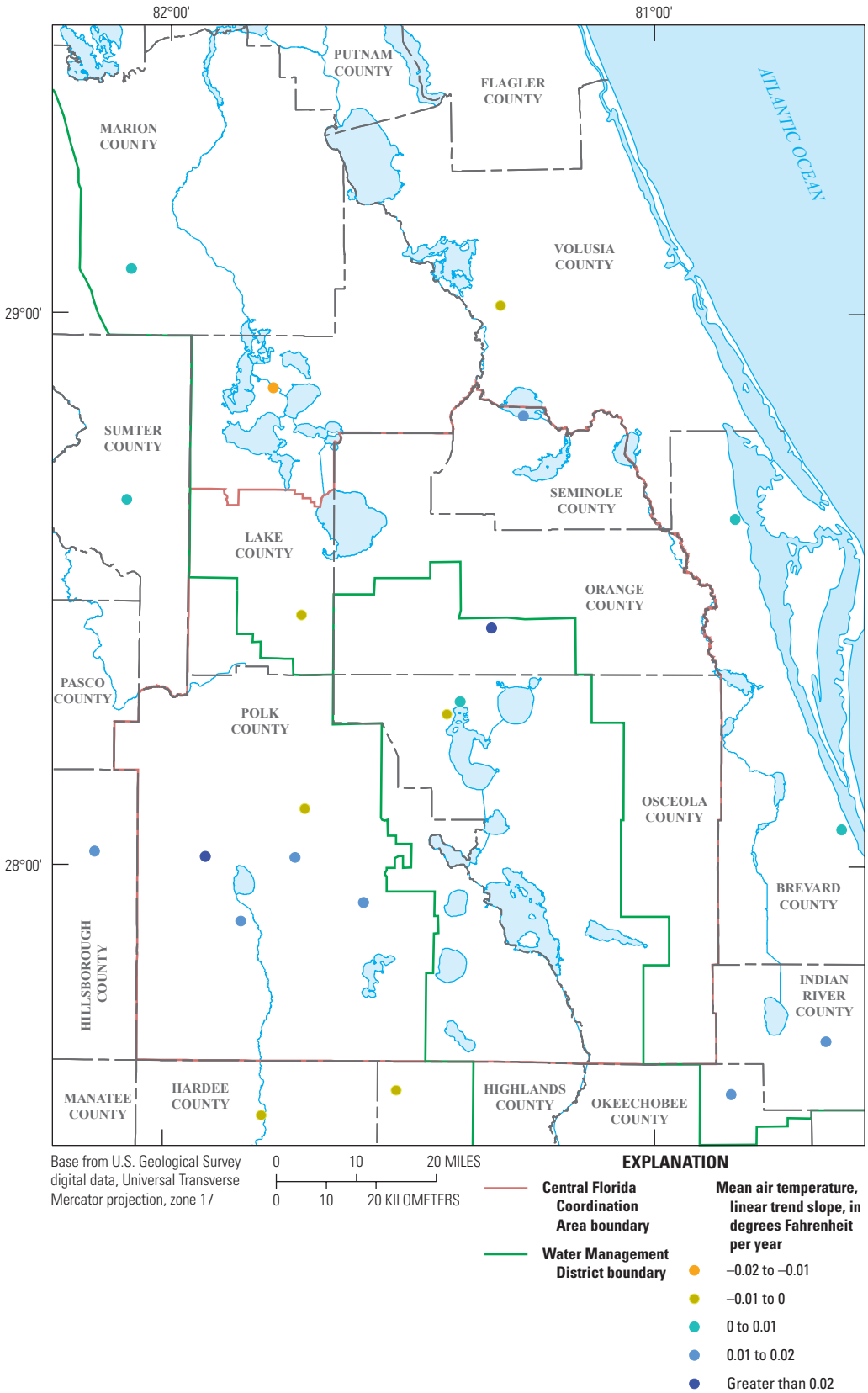


Figure 12. Linear trend in daily average air temperature for sites in the study area with a period of record of at least 45 years.

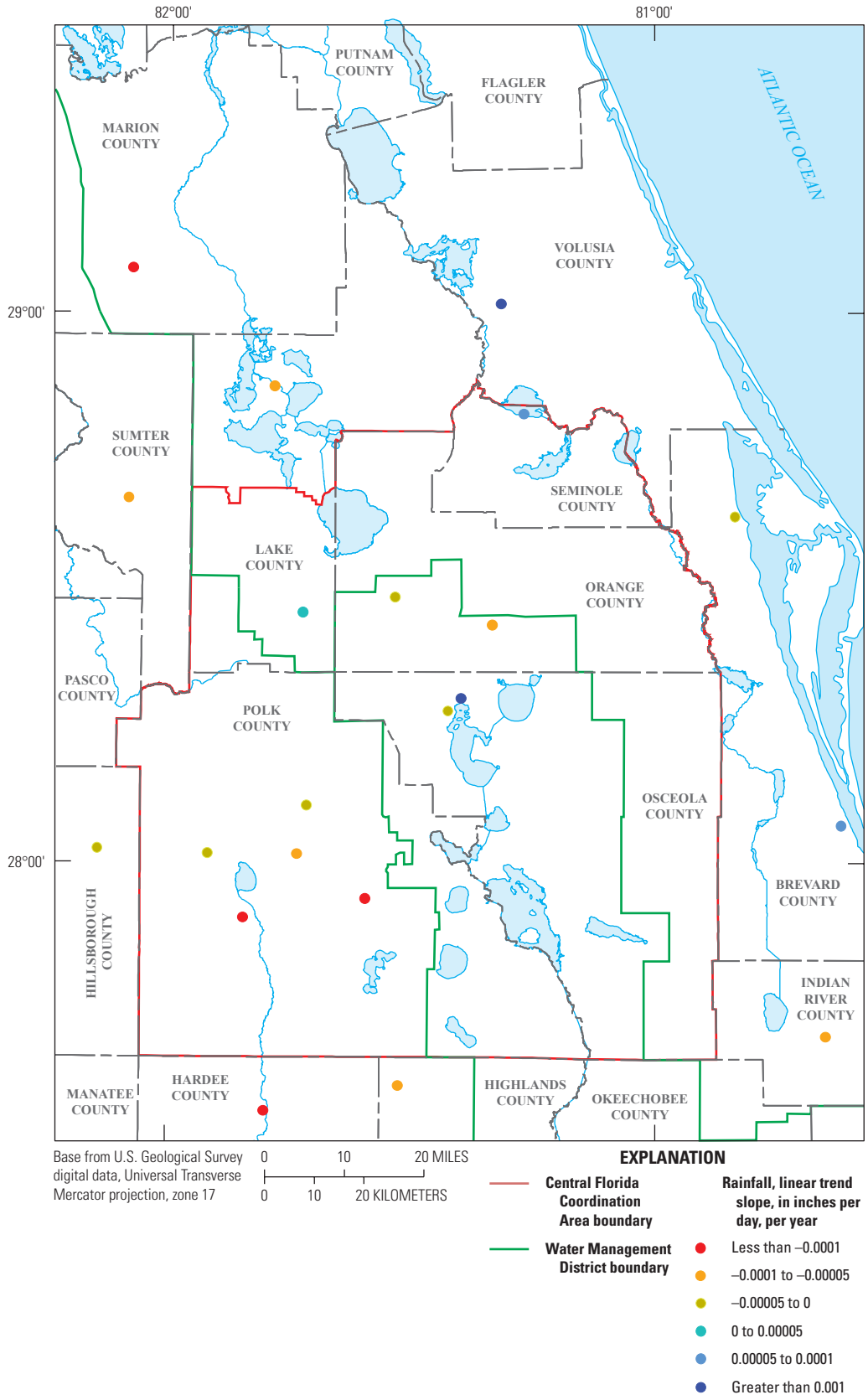


Figure 13. Linear trend in daily rainfall for sites in the study area with a period of record of at least 95 years (1901 or 1914 through 2008).

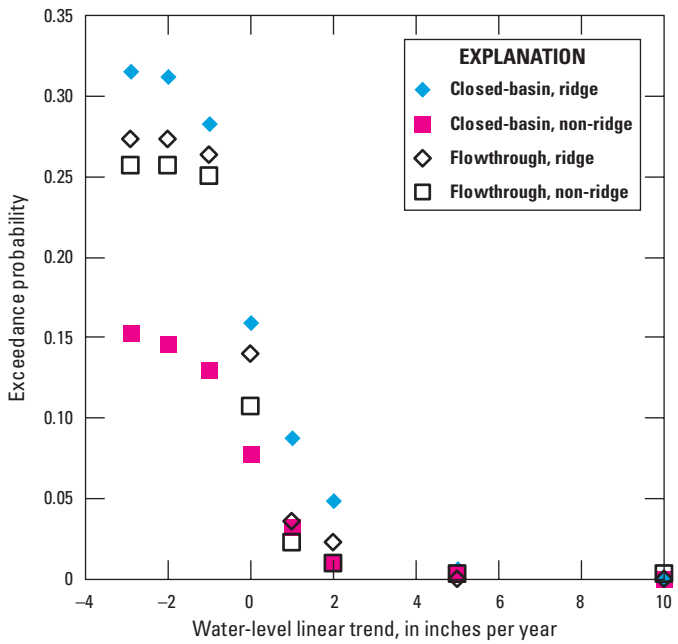


Figure 14. Exceedance probability of linear trend in daily lake water level by lake type (closed basin or flowthrough) and physiography (ridge or non-ridge physiographic region) for lakes in the study area with at least 10 measurements and a period of record of at least 10 years.

Wong (1979). For k number of groups, the k-means algorithm optimizes which members of the overall group should be in groups 1 through k. The optimal partitioning of groups is determined by using the root-mean-square error (RMSE) as a measure of the difference in distance between each member and the mean of the group such that movement of any point from one group to another will not decrease the RMSE for either group. The k-means clustering technique was repeated for a range of group sizes (k values), and a mean RMSE for all groups was computed for each k value and plotted (fig. 18). In some cases the optimal number of groups can be selected at the inflection point between a sharp vertical decline in mean RMSE and a horizontal plateau, for example, for the spring results showing a distinct break at three groups (fig. 18). In other cases a more gradual reduction in RMSE with increasing number of groups occurs, for example, for the lake results showing breaks at two and six groups (fig. 18).

Cluster analyses were performed for several different time periods: 31 years (1978–2008) for lakes, 31 years (1978–2008) and 41 years (1968–2008) for wells, and 41 years (1969–2009) for springs. For lakes, the optimal number of groups selected was six (fig. 18). There is no apparent geographic association among these lake groups, as each group has members generally located throughout the study area (fig. 19). However, groups 2 and 3 have proportionately more closed-basin lakes (10 of 22 for group 2 and 9 of 20 for group 3) than all groups combined (32 of 121), whereas groups 4 and 5 have proportionately more flowthrough lakes (20 of 21 for group 4 and 19 of 22 for group 5) than all groups combined (89 of 121) (fig. 20). For wells, the optimal number of groups selected was three for both the 31- and 41-year cluster analyses (fig. 18). For the 31-year cluster analysis, wells generally are associated geographically (fig. 21). Group 1 wells are in southwest Polk County and comprise all Floridan aquifer system wells and two intermediate aquifer system wells in this area. Group 2 wells are generally in south Lake and Orange Counties, east Polk County, and all of Osceola County, and include all but one of the surficial aquifer system wells used in the analysis. Group 3 wells are generally in the central and northern parts of the study area. Wells from the 31-year cluster analysis with long periods of record were used in the 41-year cluster analysis and generally were associated into similar groups (fig. 22). For springs, the optimal number of groups selected was three (fig. 18). Spring groups are loosely characterized by geographic location and magnitude of flow. Group 1 springs comprise Alexander and Blue Springs, both of which are first magnitude springs. Group 2 springs comprise Miami, Palm, Sanlando, and Starbuck Springs, which are small second magnitude or third magnitude springs where all but Miami Springs are in close proximity to each other (fig. 22). Group 3 springs comprise Bugg, Rock, and Wekiwa Springs, where both Rock are Wekiwa are second magnitude springs.

Time-series cluster analyses also were performed on long-term meteorological data reported by NOAA. The period 1942–2008 was selected for analysis because it had the least amount of missing data. However, large amounts of data were missing for 5 of the 23 NOAA rainfall sites and 1 of the 21 NOAA air temperature measurement sites; the data for these sites were not used in the analyses. Estimates of missing rainfall data were included in the hydrologic database (appendix 1), but most of these estimates were developed by Aly (2008) for long-term rainfall analyses, such as those used in the calculation of long-term water budgets, and not for event-based predictions; thus, a daily prediction from this method may have a large error (Aly, 2008, p. 97–98). Therefore, the relatively small amounts of missing data in the remaining 18 rainfall time series and 20 PET time series were filled using "IF-THEN-ELSE" rules in which a missing value at a gage is filled using values from one or more neighboring stations. Missing rainfall data in particular are difficult to fill because of their "spiky," spatially discontinuous nature; therefore, these filling rules were selected to maintain consistent temporal variability in the data. Figure 23 shows how data were propagated across the network, with the arrows indicating the direction of flow. The order of filling emphasized the use of gages with the most abundant data to fill the records for gages with the next most abundant data and so on. In some instances filled data at one gage were used to fill the record for another gage. Data for three additional rain gages (sites 510206, 540106, and 1880101 monitored by the SJRWMD and located in Indian River County; fig. 23) were used to fill the record for NOAA site 2936. The rainfall filling rules also were used for

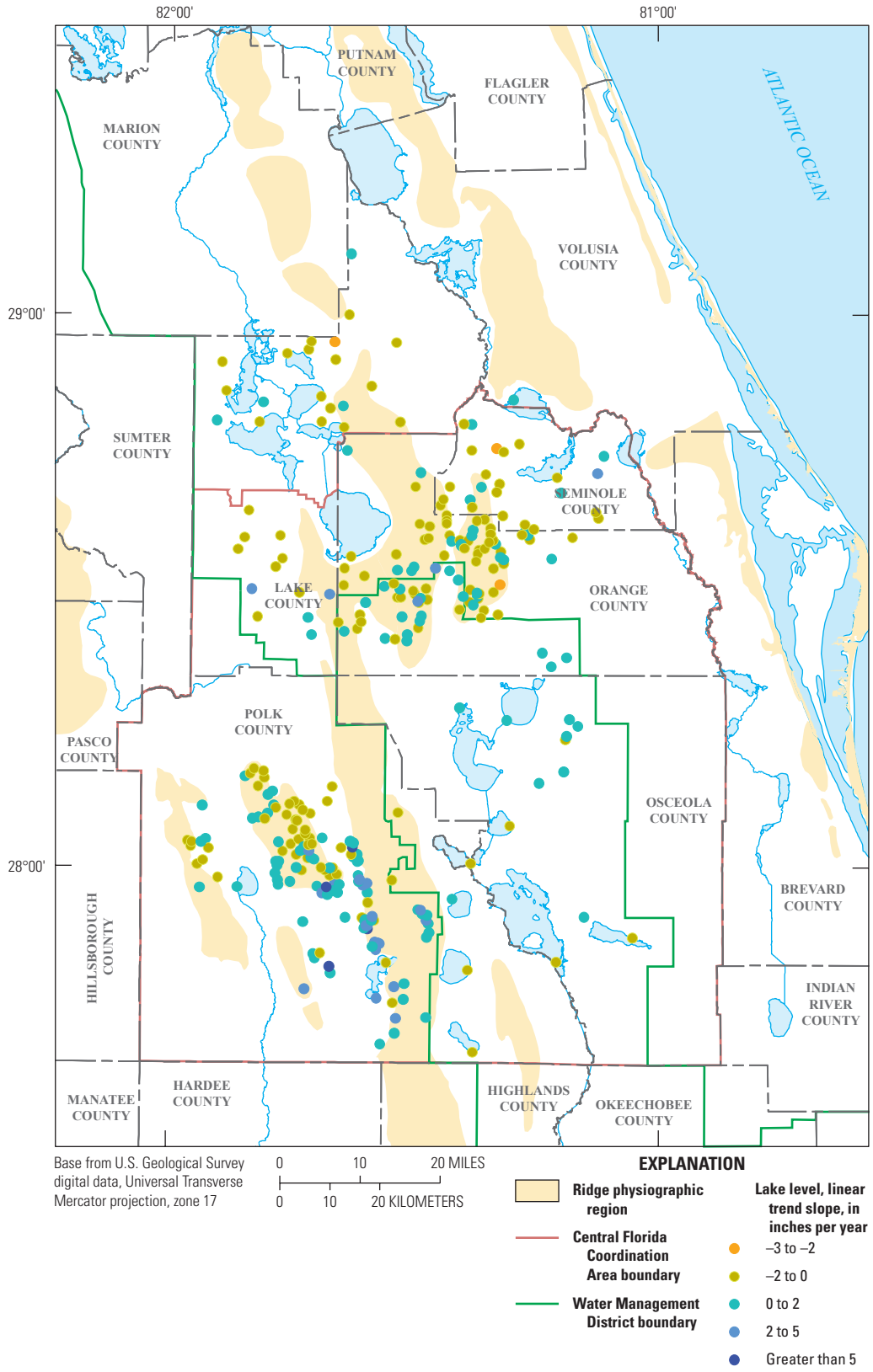


Figure 15. Linear trend in daily water level for lakes in the study area with at least 10 measurements and a period of record of at least 10 years.

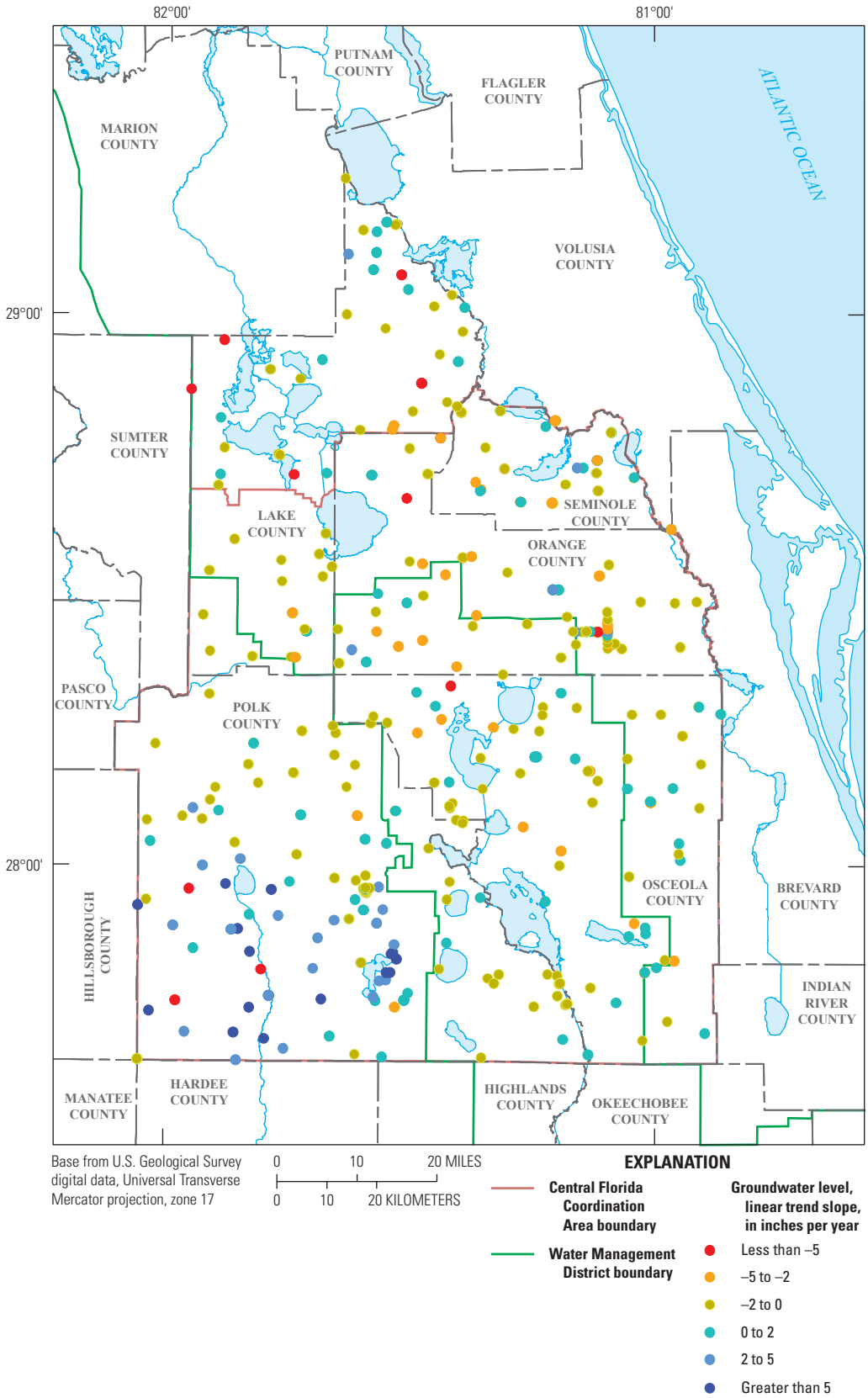


Figure 16. Linear trend in daily groundwater level for wells in the study area with at least 10 measurements and a period of record of at least 10 years.

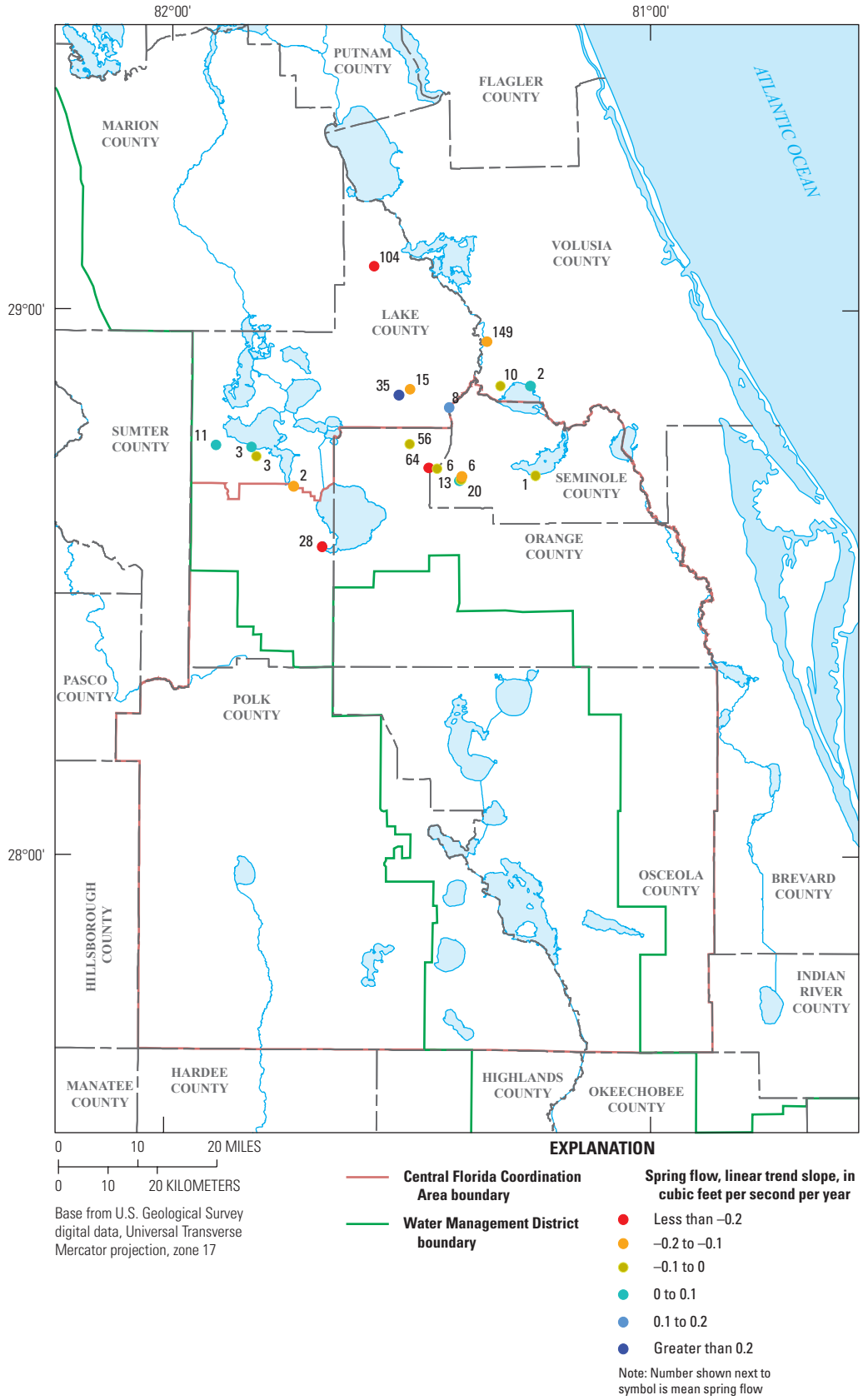


Figure 17. Linear trend in daily flow and mean flow for springs in the study area with at least 10 measurements and a period of record of at least 10 years.

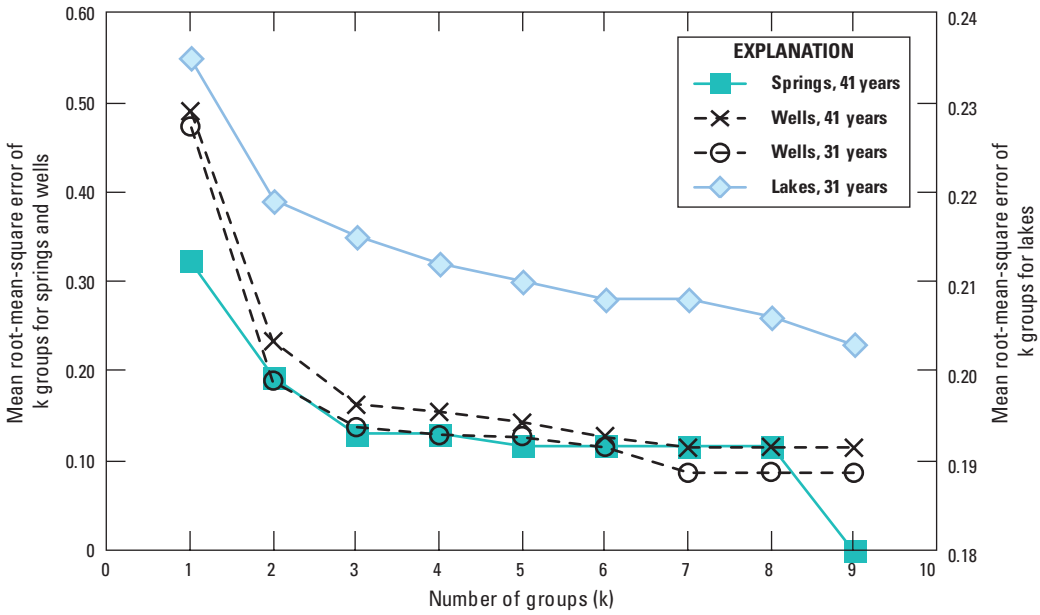


Figure 18. Root-mean-square error for k-means cluster analysis of water-level data for lakes and wells and flow data for springs in the study area.

filling missing PET data rather than developing a new set of rules unique to the PET data. This approach is expected to be sufficient given the more spatially and temporally consistent nature of PET data. To assess the longer-term variability in rainfall and PET data, cluster analyses were performed using 1,440-day (approximately 4-year) MWAs computed from the filled datasets.

Cluster analysis of rainfall data from 18 NOAA gages indicates 4 groups of sites with relatively similar dynamic behavior exhibiting north-south and east-west geographic associations (fig. 24). Comparisons with the long-term linear trend analysis (fig. 13) show the rain gages in the western half of the study area belonging to cluster groups 3 and 4 had downward trends, with the exception of NOAA gage 1641, which had a slightly upward trend. In contrast, rain gages in the eastern half of the study area belonging to cluster groups 1 and 2 had a mix of upward and downward long-term trends (fig. 13). Such spatial patterns in long-term trends may potentially affect the hydrologic system, especially the groundwater flow system because most of the areas of high groundwater recharge within the study area occur in the western half (Sepúlveda and others, 2012), and long-term downward trends are apparent in groundwater levels and spring flows (figs. 16, 17; table 1). Closer examination of 4-year MWAs of the daily time series, however, indicates that substantial intragroup differences exist (fig. 25), because rainfall data commonly are highly variable in space and time. Therefore, for developing empirical models such as ANNs, the use of data from all 18 rain gages rather than generalizing into 4 groups should improve model predictive capability.

Cluster analysis of Hargreaves PET data computed from air temperature at 20 NOAA gages indicates 3 groups of sites with relatively similar dynamic behavior and a generally east-west geographic association (fig. 26). The PET data generally are more smoothly varying in time than the rainfall data, although some intragroup differences indicate spatial variation (figs. 25, 27). These results are consistent with those of Sumner (2006) and O'Reilly (2007), who noted that the temporal variability of ET in central Florida is considerably less than that of rainfall for a wide range of time scales (from daily to annual). Such differences in the stochastic characteristics of rainfall and ET lead to rainfall explaining more than 90 percent of the temporal variability in available water (rainfall minus ET) (O'Reilly, 2007, p. 15). Because available water drives the hydrologic system and its variability is controlled by rainfall, PET time series were not further analyzed in this study and were excluded as explanatory variables for ANN modeling.

Groundwater Use

To view general spatial trends in groundwater use over time, an application was developed that aggregates the monthly data using a grid cell size of approximately 20 × 20 km (fig. 28). This application—called the Groundwater-Use Data Viewer (GUDV)—is described in further detail in appendix 2. The relatively large grid cell size was selected because the effects of groundwater use can propagate over fairly long distances in a karst aquifer, such as the Floridan aquifer system. Murray (2010) analyzed the effects

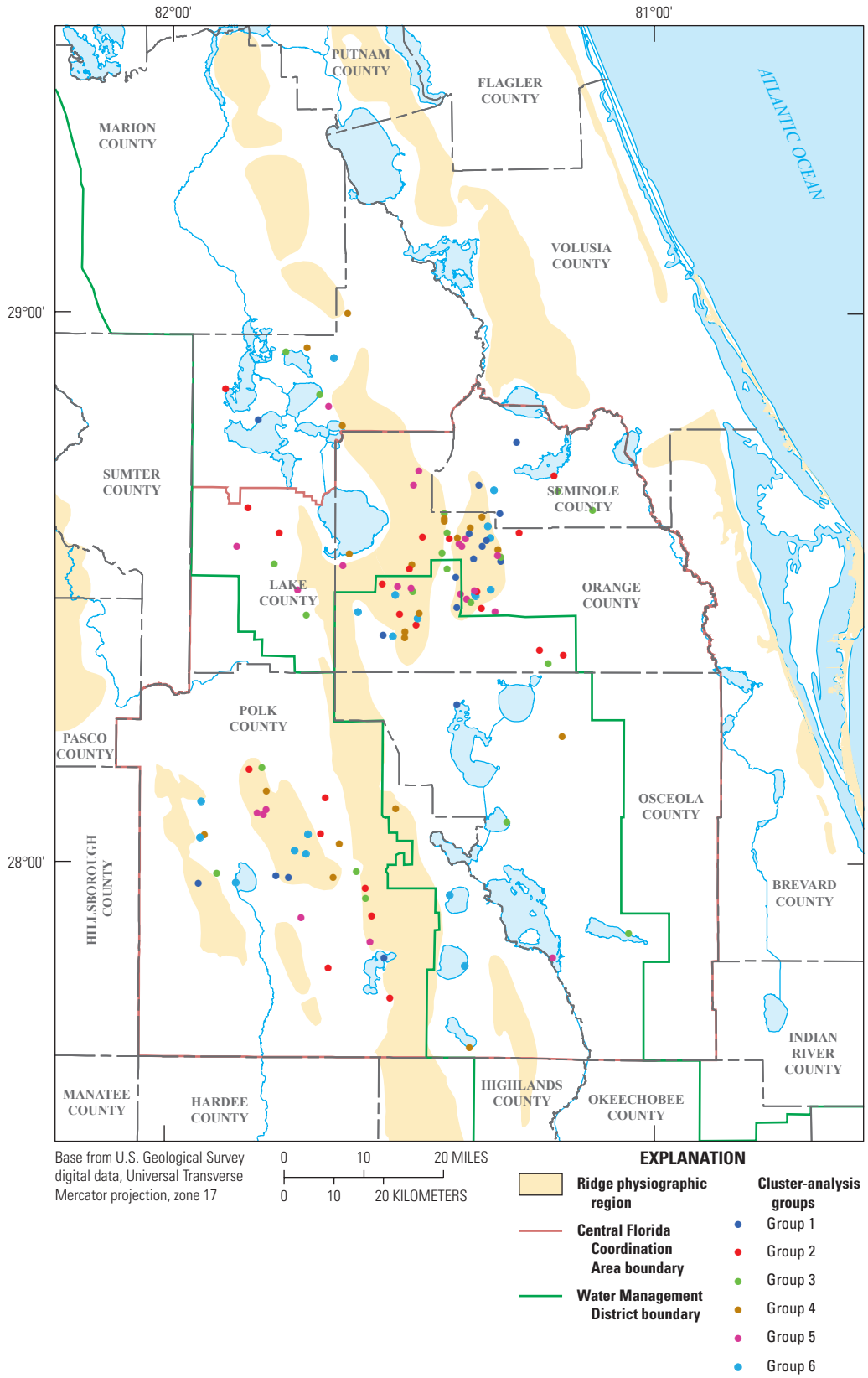


Figure 19. Locations of lakes in the study area identified by group number based on the 31-year cluster analysis of lake water-level data from 1978 through 2008.

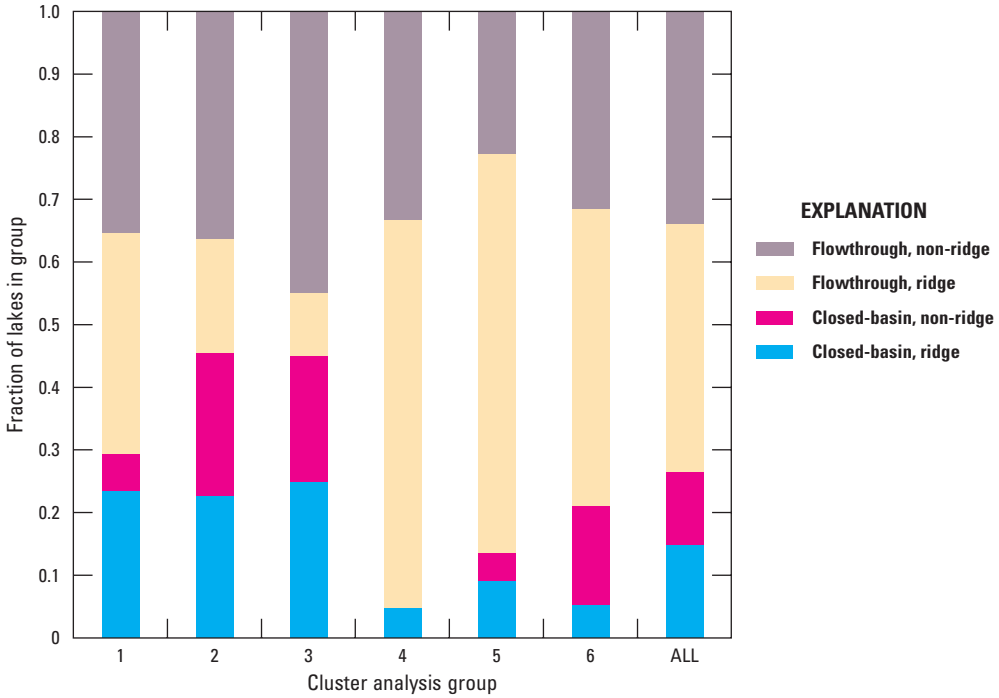


Figure 20. Distribution of lakes in the study area identified by lake type (closed basin or flowthrough) and physiography (ridge or non-ridge physiographic region) based on the 31-year cluster analysis of lake water-level data from 1978 through 2008.

of groundwater withdrawals and water levels at two monitoring well sites in central Florida and reported no changes in correlation between groundwater levels and withdrawals at a range of radial distances from 2.5 to 10 miles, suggesting that that the GUDV cell width of 20 km (12.4 miles) is not too large. The grid structure of the GUDV also was used to estimate missing data for public-supply groundwater withdrawals. Data reported on water distributed by public-supply water-treatment plants in the study area (fig. 4) are sparse prior to 1978; therefore, data from 1950 through 1977 were estimated by hindcasting measured data. Hindcasting consisted of linear extrapolation of log-transformed 1978–2008 data with average monthly variations from 1978 to 2008 added back in. Hindcasting was done per GUDV grid cell (fig. 28), not for individual users, and only for cells where 1978 total public-supply groundwater withdrawals exceeded 0.1 Mgal/d. Also, five cells for which public-supply data were missing in the early part of the 1978–2008 period were filled in the same way. Hindcasted data compare better with the limited amount of reported data in the period 1965–77, when percentage differences generally were less than 10 percent. Results compare more poorly with data for the period 1950–64, when percentage differences generally exceeded 10 percent. The paucity of measured data affects the accuracy of such comparisons, however, and the hindcasted data have a similar upward trend as the public-supply data estimated by Munch (2014) for the same period.

Spatial and temporal trends in groundwater-use data were examined using the GUDV. Long-term patterns in total groundwater use exist in both monthly and annual data for all water-use categories (fig. 29). Seasonal variations in groundwater use are apparent in the monthly data, especially in the citrus and non-citrus agricultural irrigation categories (fig. 29*A*). Examination of annual data clearly indicates that all categories of groundwater withdrawal were increasing during the 1950s and throughout most of the 1960s, with the largest quantities from the phosphate mining category increasing from 44 Mgal/d in 1951 to 252 Mgal/d in 1966 (fig. 29*B*). Groundwater withdrawals by the phosphate mining industry began to decline in 1967 and generally have continued to decline, reaching a minimum of 44 Mgal/d in 2006. Groundwater withdrawals for citrus and non-citrus agricultural irrigation have increased from 63 and 50 Mgal/d, respectively, in 1957 to a maximum of 225 Mgal/d for citrus in 1981 and 139 Mgal/d for non-citrus agricultural in 1984. Since the early 1980s, however, citrus and non-citrus agricultural groundwater withdrawals generally have decreased, reaching a minimum of 126 Mgal/d for citrus irrigation in 2008 and 29 Mgal/d for non-citrus agricultural irrigation in 2002. The amount of water used for golf course irrigation is small relative to that in other categories but has shown a slight upward trend since the earliest available records, from 3 Mgal/d in 1977 to 18 Mgal/d in 2008 and a maximum of 34 Mgal/d in 2000. As expected, drainage well recharge generally follows rainfall patterns with

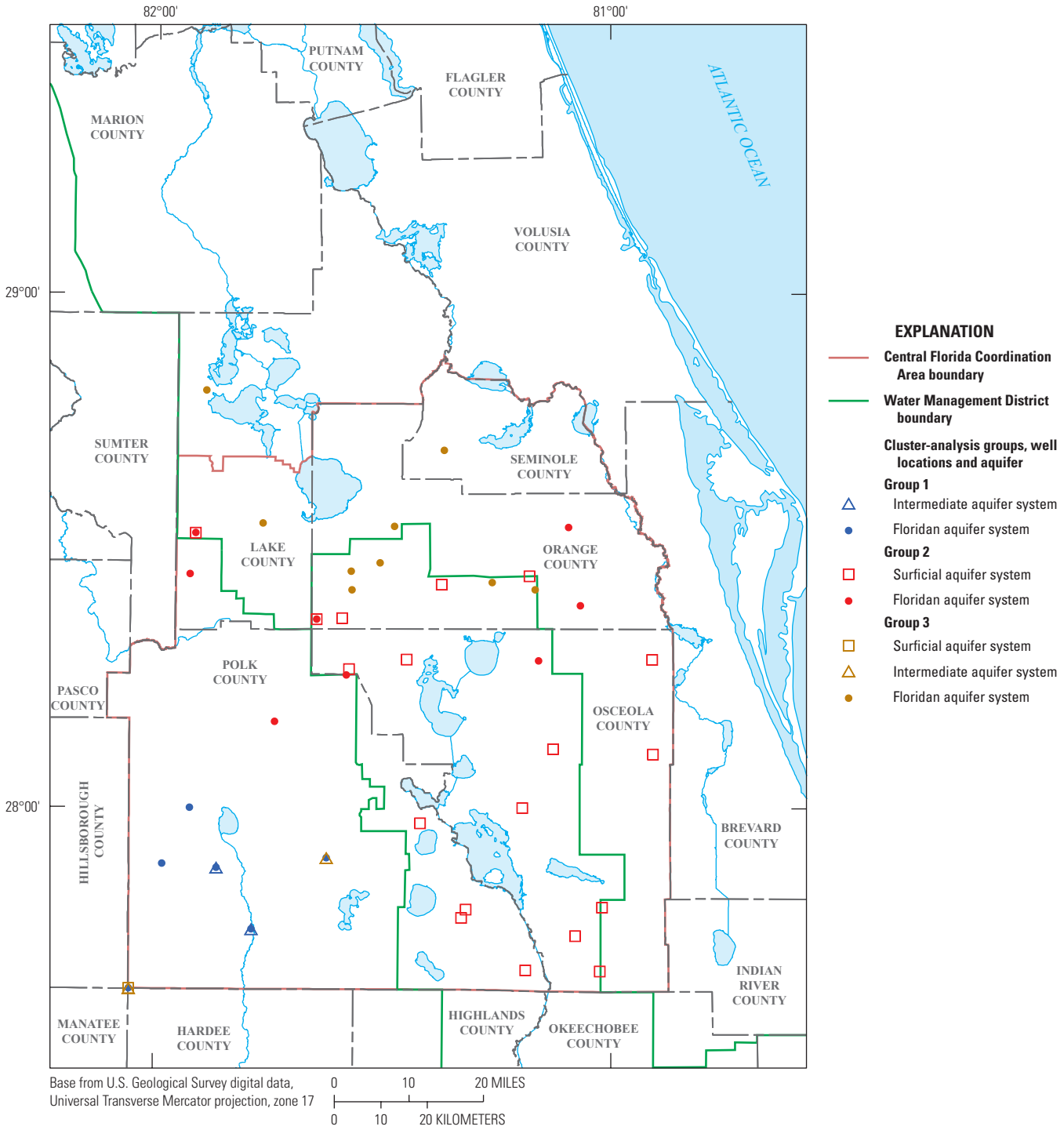


Figure 21. Locations of wells in the study area identified by group number and aquifer based on the 31-year cluster analysis of groundwater-level data from 1978 through 2008.

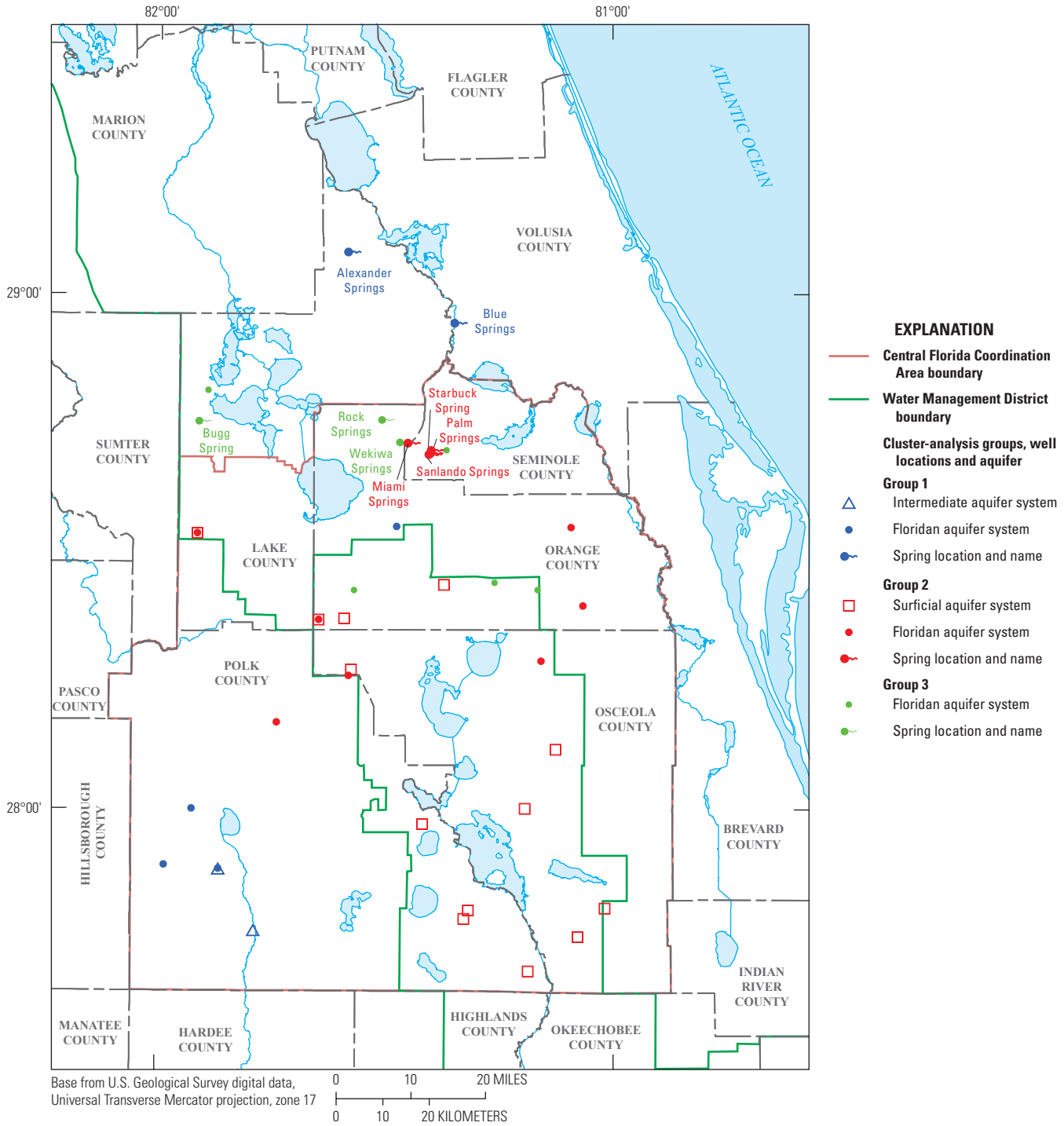


Figure 22. Locations of springs in the study area identified by group number based on the 41-year cluster analysis of spring-flow data from 1969 through 2009 and locations of wells by group number and aquifer based on the 41-year cluster analysis of groundwater-level data from 1968 through 2008.

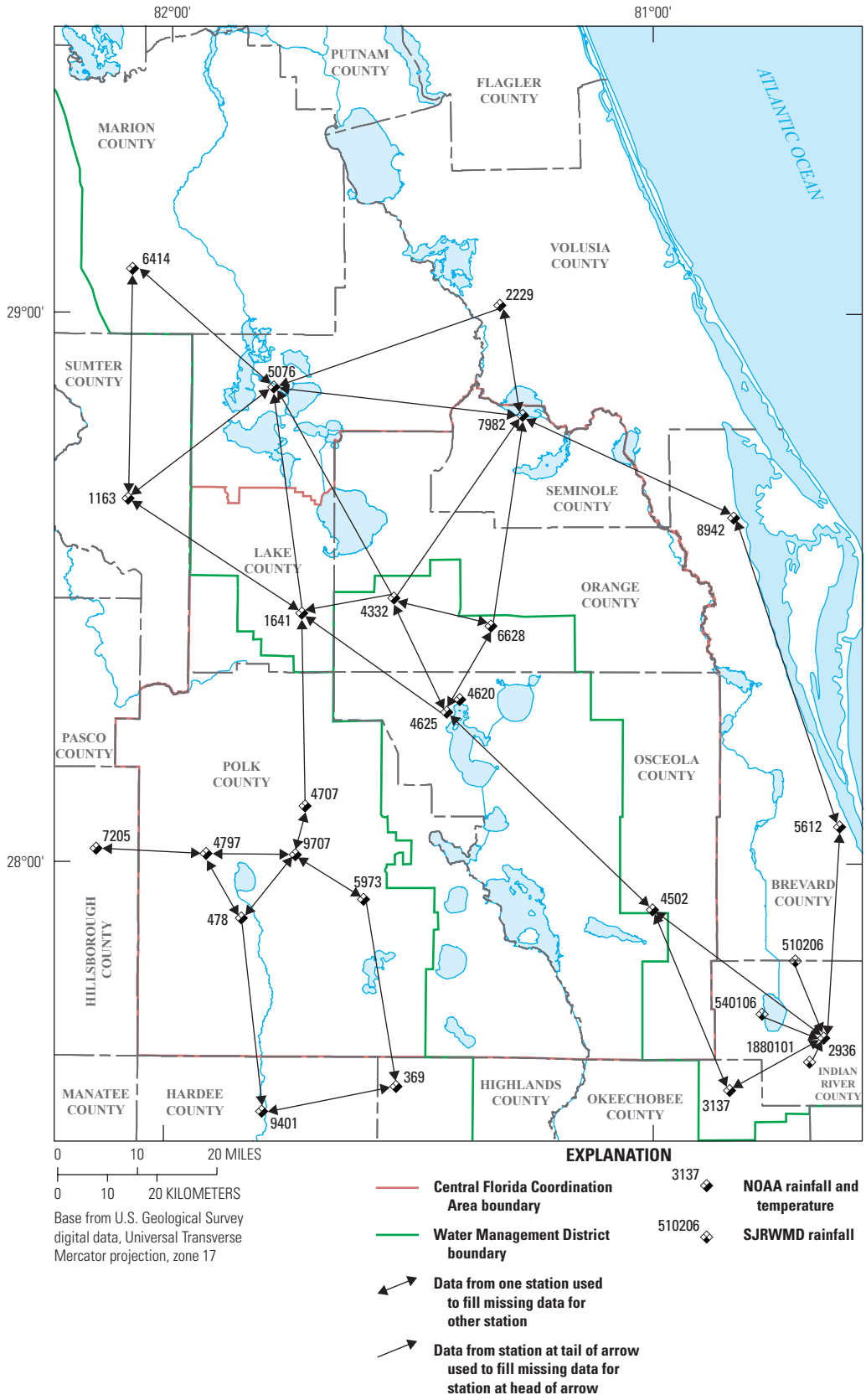


Figure 23. Direction of flow for filling missing rainfall data using "IF-THEN-ELSE" rules.

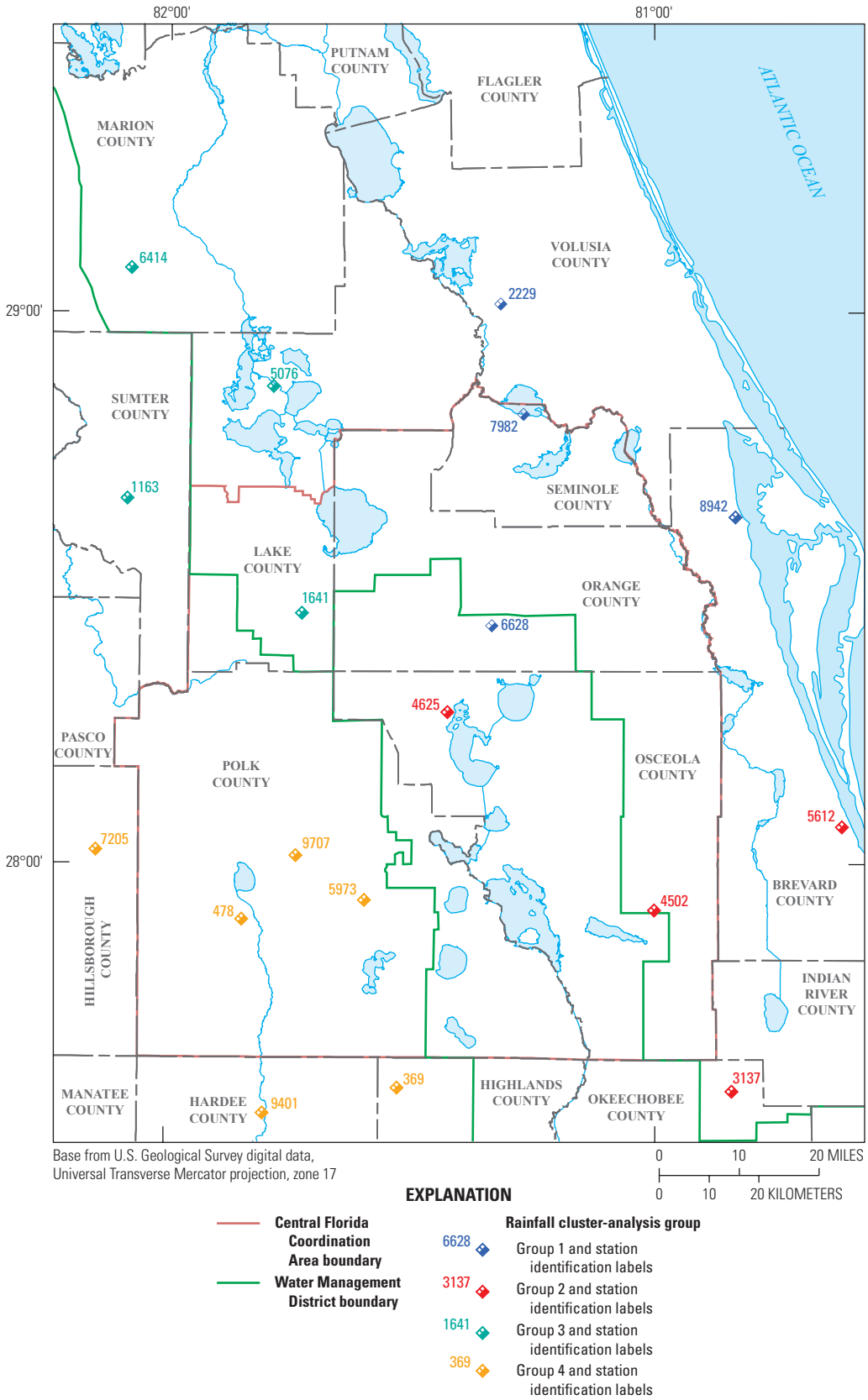


Figure 24. Locations of National Oceanic and Atmospheric Administration (NOAA) rain gages in the study area identified by group number based on the 67-year cluster analysis of rainfall data from 1942 through 2008. Information and data for each site are included in the hydrologic database (described in appendix 1) and can be referenced by the station identification label.

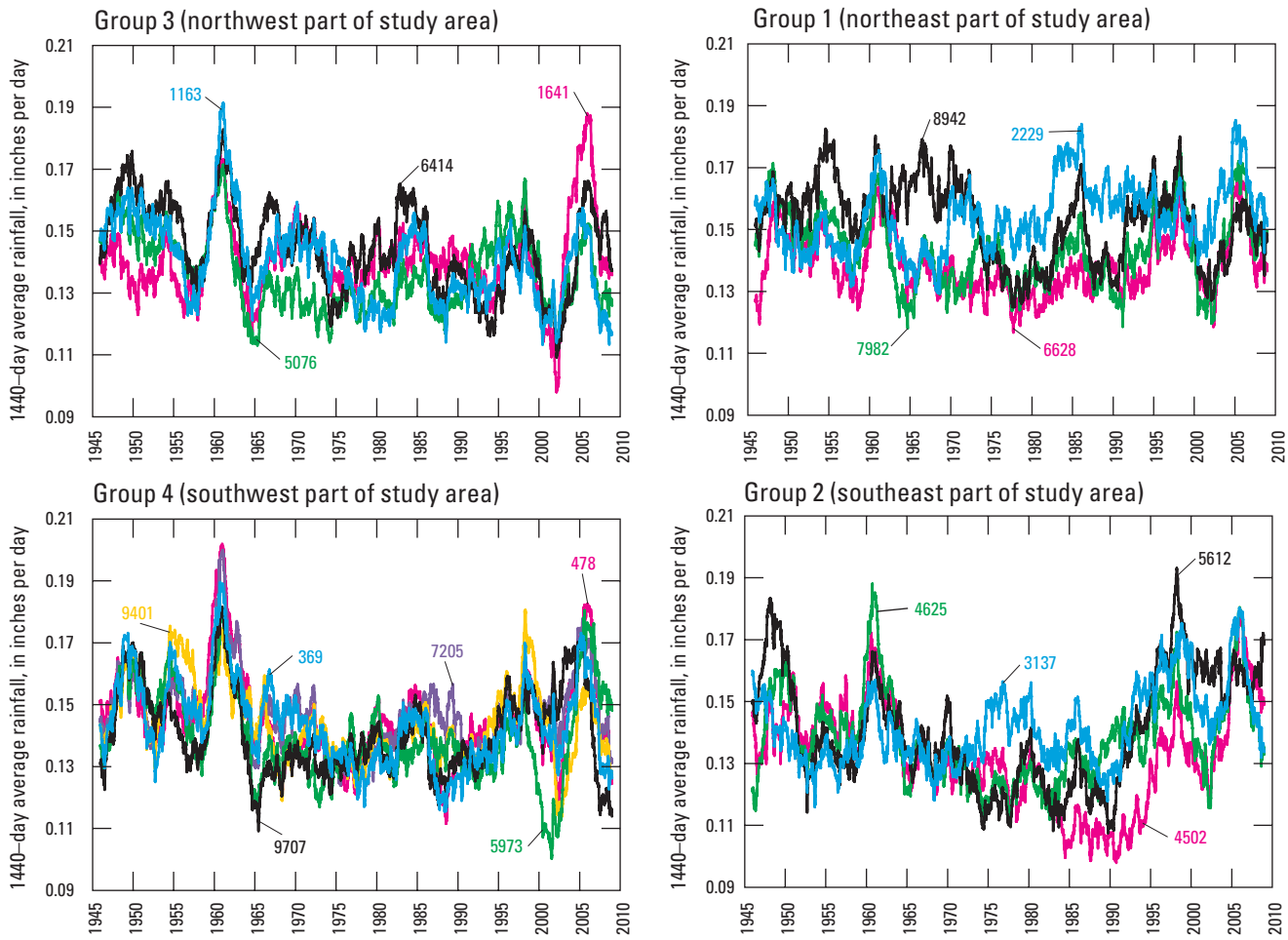


Figure 25. Moving window averages (1,440 days ≈ 4 years) of rainfall data for the individual National Oceanic and Atmospheric Administration (NOAA) rain gages in the study area composing the four groups derived from the 67-year cluster analysis. Site locations are shown in figure 24.

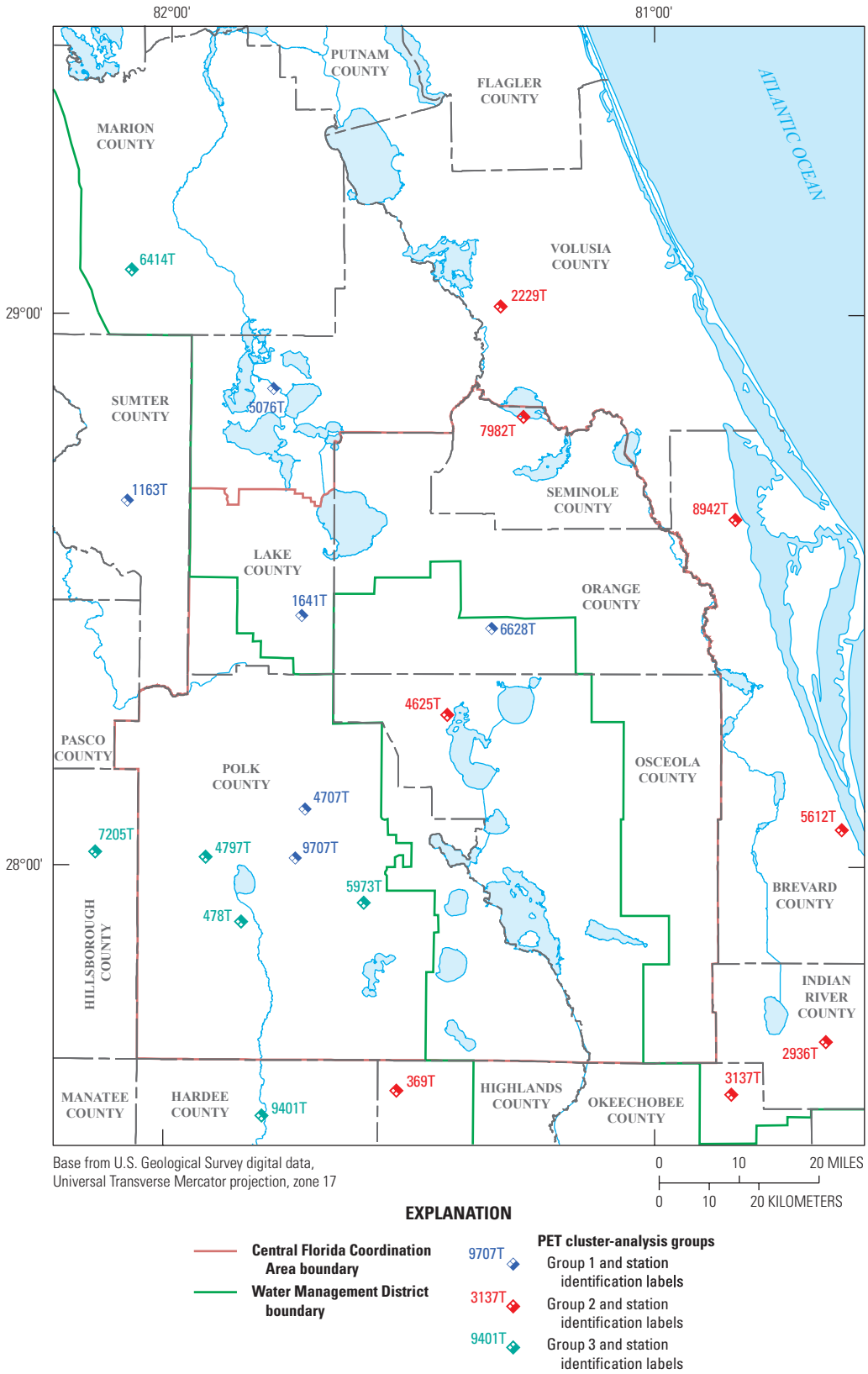


Figure 26. Locations of National Oceanic and Atmospheric Administration (NOAA) air temperature measurement sites in the study area identified by group number based on the 67-year cluster analysis of computed Hargreaves potential evapotranspiration (PET) data from 1942 through 2008. Information and data for each site is included in the hydrologic database (described in appendix 1) and can be referenced by the station identification label.

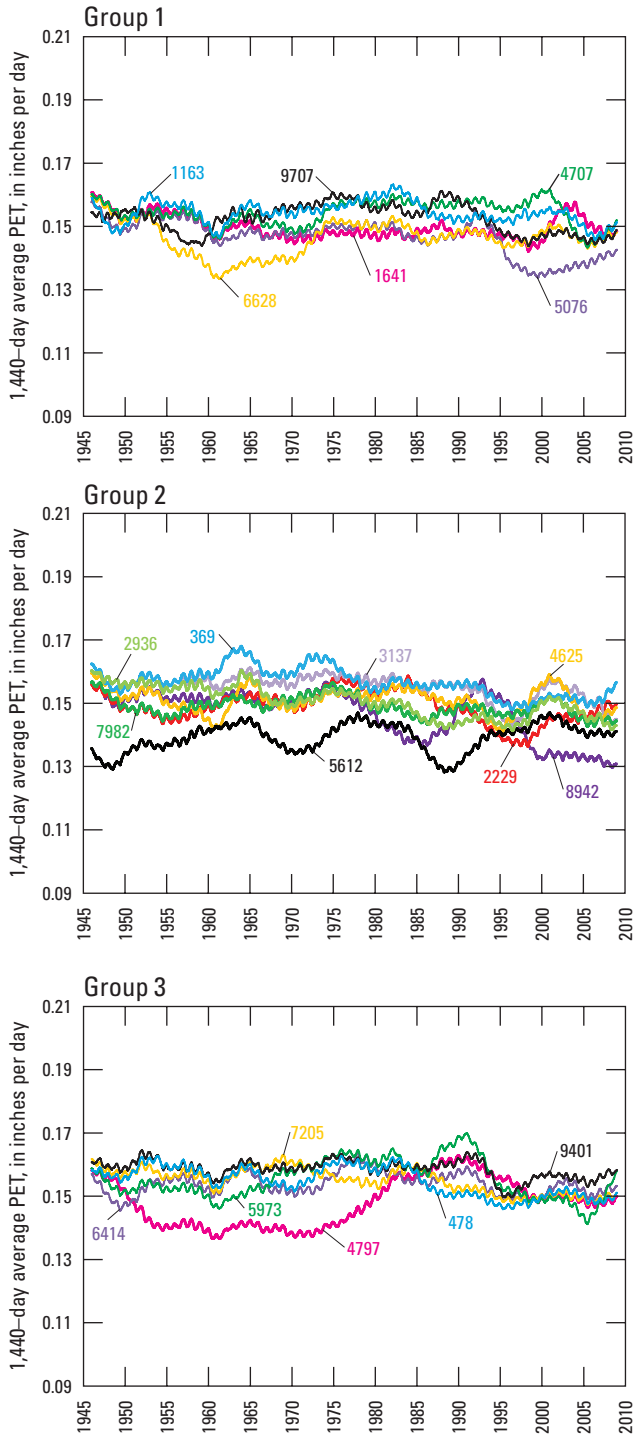


Figure 27. Moving window averages (1,440 days ≈ 4 years) of computed Hargreaves potential evapotranspiration (PET) data for the individual National Oceanic and Atmospheric Administration (NOAA) air temperature measurement sites in the study area composing the three groups derived from the 67-year cluster analysis. Site locations are shown in figure 26.

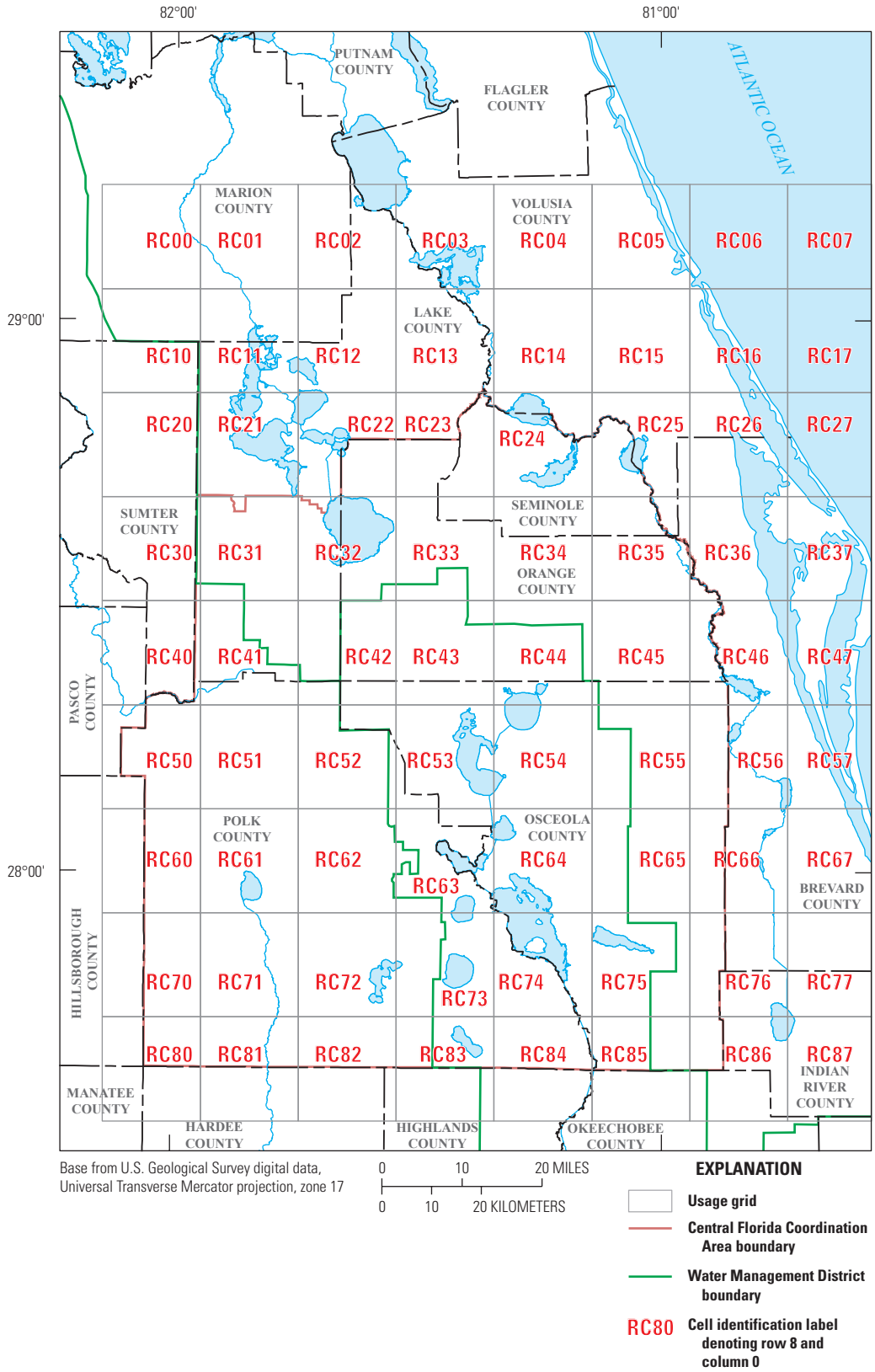


Figure 28. Grid used for aggregating groundwater-use data in the Groundwater-Use Data Viewer showing the identification label for each cell.

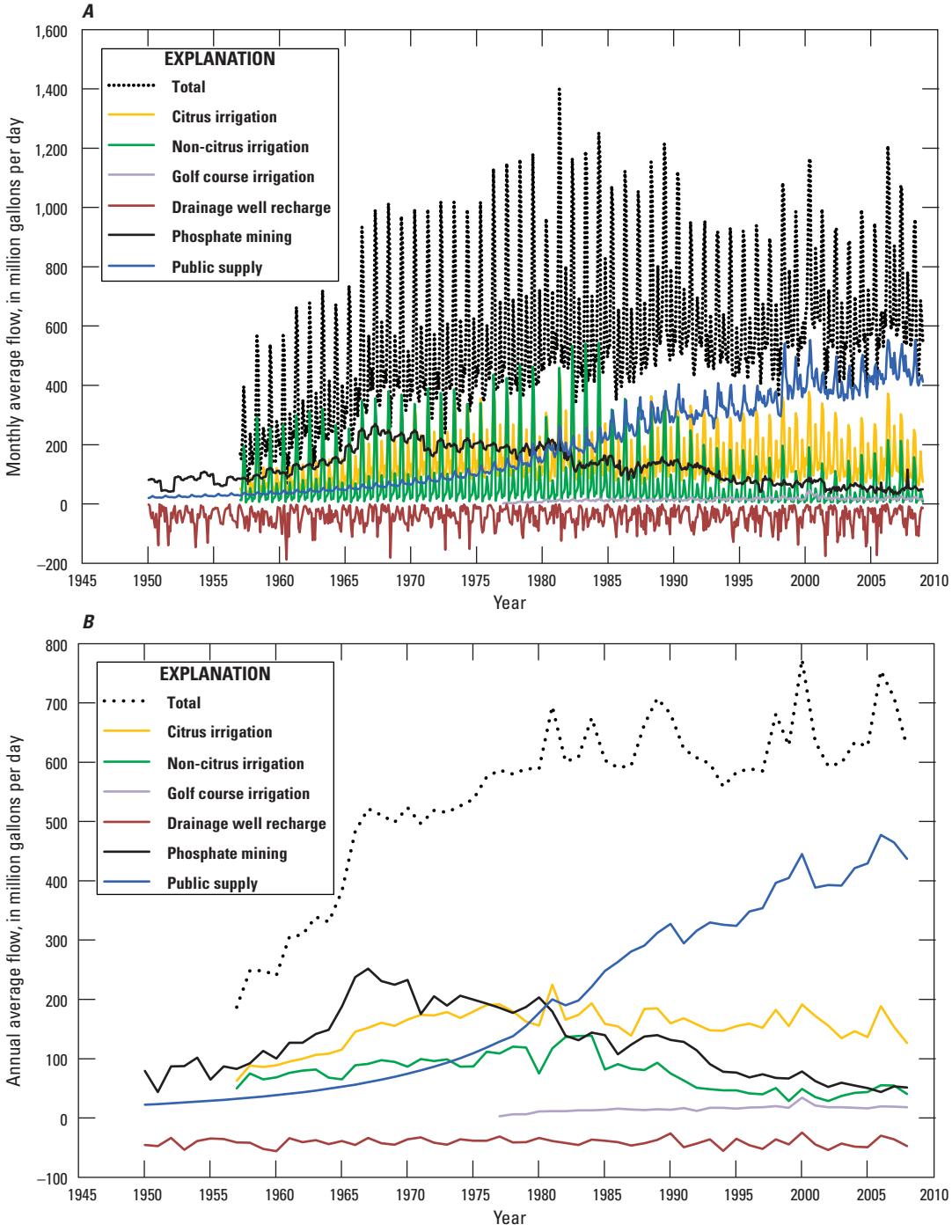
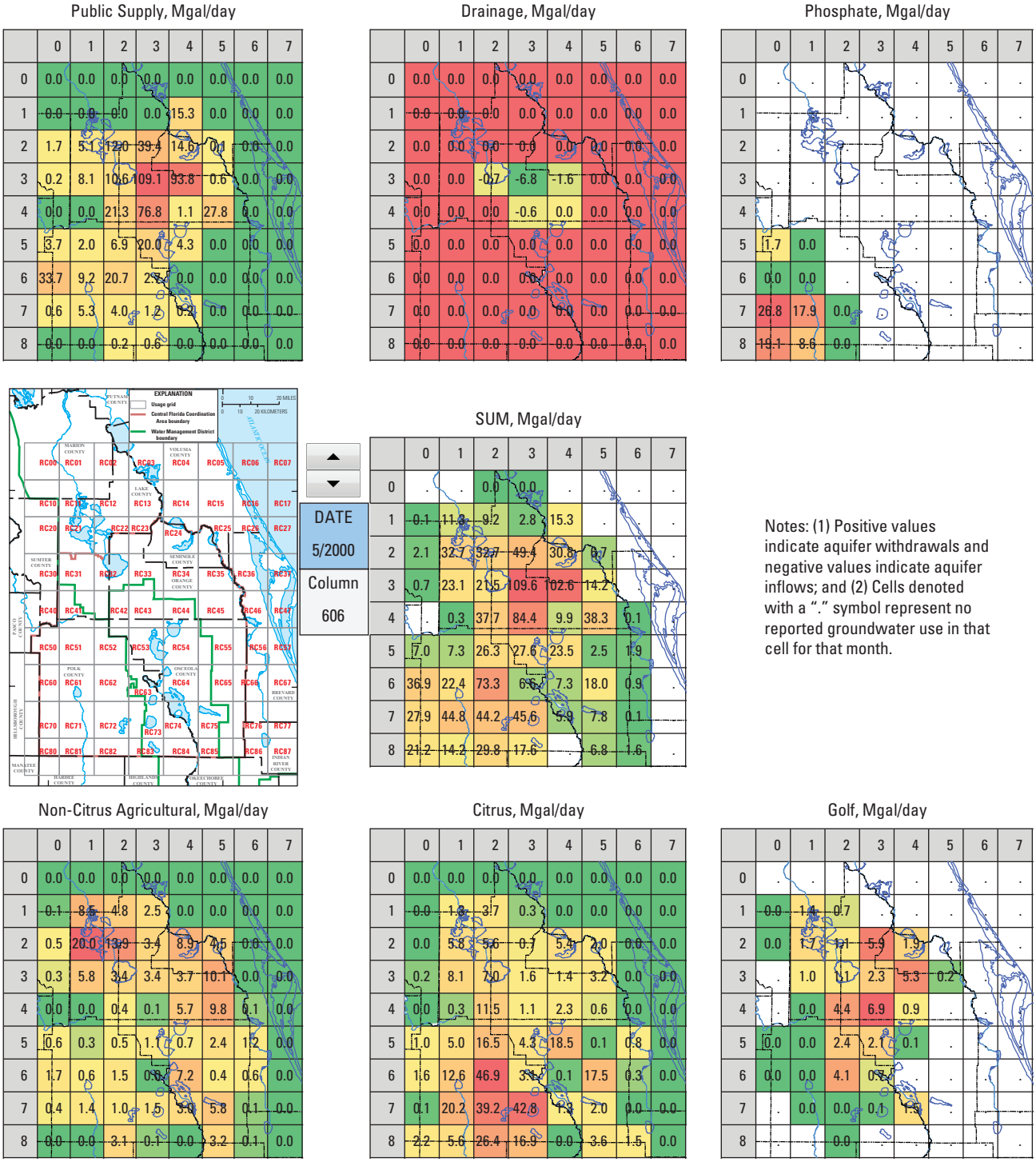


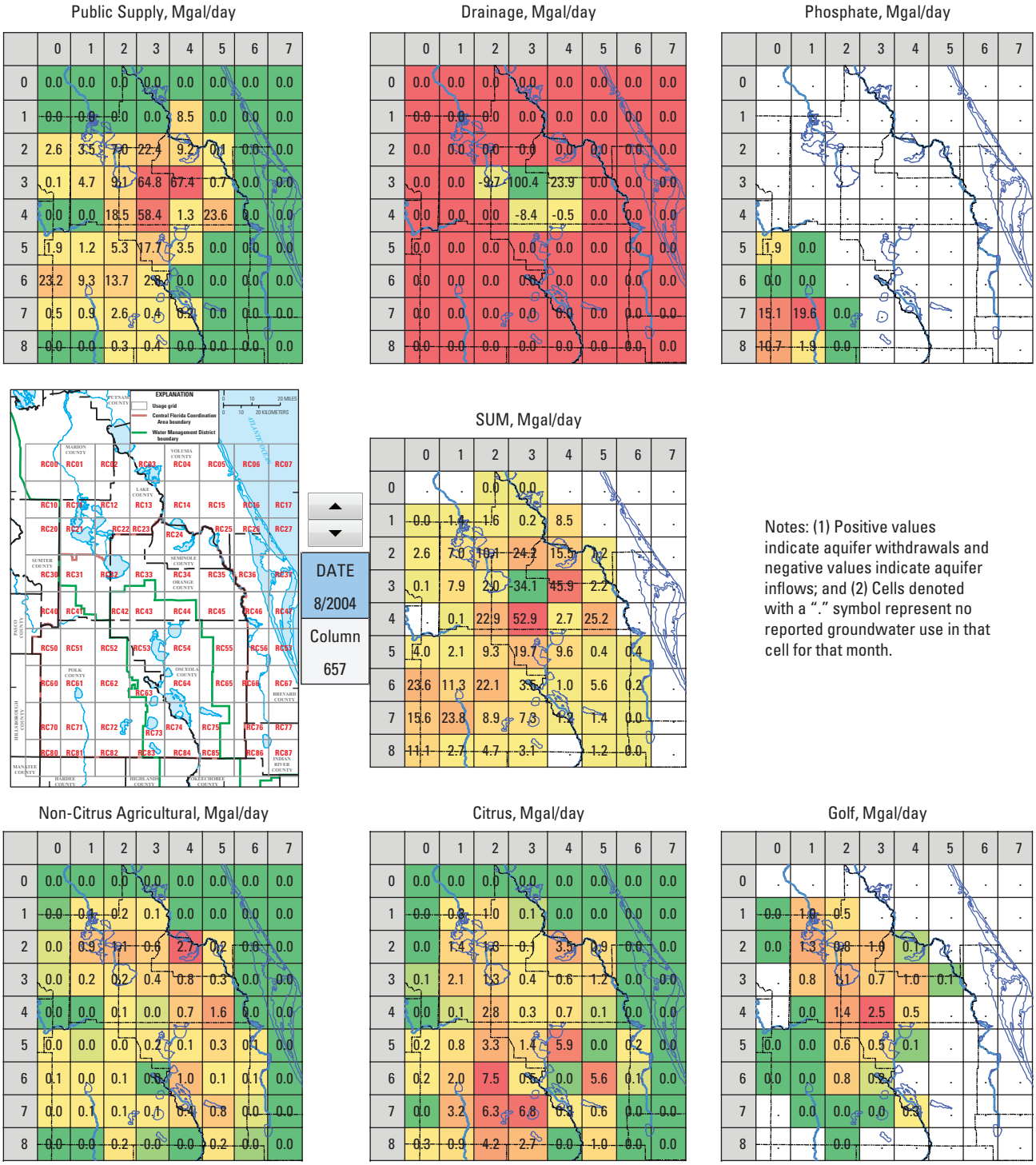
Figure 29. Total groundwater use in the Central Florida Coordination Area and vicinity by category from 1950 through 2008 based on the Groundwater-Use Data Viewer for *A*, monthly data, and *B*, annual average data. Drainage well recharge depicted as negative values because these flows represent inputs to, rather than withdrawals from, the aquifer system.

a minimum of 25 Mgal/d in 2000, maximum of 56 Mgal/d in 1960, and mean of 41 Mgal/d for 1950–2008. In contrast to these other categories, public-supply groundwater withdrawals have increased substantially from 22 Mgal/d in 1950 to 437 Mgal/d in 2008, although short periods of decrease occurred in 1981–82, 1990–91, 2000–2001, and 2006–08. The net result is that total groundwater use in the study area (excluding 1950–57 due to missing data) increased from about 250 Mgal/d in 1958 to about 590 Mgal/d in 1980 and has remained relatively stable in the period 1981–2008 with a minimum of 559 Mgal/d in 1994, maximum of 773 Mgal/d in 2000, and mean of 636 Mgal/d. The change in trend in the early 1980s and the following period of relatively slight trend is attributable to the concomitant effects of increasing public-supply withdrawals and decreasing phosphate mining and citrus and non-citrus agricultural irrigation (fig. 29*B*). A comparison of aggregated data for a dry period (May 2000) and a wet period (August 2004) using the GUDV shows the effect of rainfall on all water-use categories, with generally high groundwater withdrawals and low drainage well recharge in May 2000 (fig. 30) and low groundwater withdrawals and high drainage well recharge in August 2004 (fig. 31). May 2000 was a dry period with only 0.53 in. of rainfall, whereas August 2004 was a wet period with 13.07 in. of rainfall, based on average data for the 18 rain gages shown in figure 24. During wet periods, the effect of drainage well recharge in offsetting public-supply withdrawals in the Orlando area is substantial, resulting in a net inflow to the aquifer in one area (fig. 31). Such insight provided by the GUDV was used to guide the development of ANN models described in the following section.



Notes: (1) Positive values indicate aquifer withdrawals and negative values indicate aquifer inflows; and (2) Cells denoted with a "." symbol represent no reported groundwater use in that cell for that month.

Figure 30. Screen capture of the Groundwater-Use Data Viewer showing groundwater use by category and sum total in million gallons per day (Mgal/day) for a dry period (May 2000).



Notes: (1) Positive values indicate aquifer withdrawals and negative values indicate aquifer inflows; and (2) Cells denoted with a "." symbol represent no reported groundwater use in that cell for that month.

Figure 31. Screen capture of the Groundwater-Use Data Viewer showing groundwater use by category and sum total in million gallons per day (Mgal/day) for a wet period (August 2004).

Methods for Simulation of Historical Lake Water Levels, Groundwater Levels, and Spring Flows

Water levels and flows in a hydrologic system typically are simulated using dynamic deterministic models that incorporate the mathematical descriptions of the physics of surface and subsurface flow. Although physics-based models have been the state of the practice for resource management evaluations of anthropogenic effects on hydrologic systems, developments in the field of advanced statistics, machine learning, and data mining offer opportunities to develop empirical ANN models that often more closely replicate observed conditions. ANN models also have the capability to reliably predict system response if the full range of input variables was considered during the model development and training process.

Limitations of Datasets

The reliability of all mathematical models, whether empirical or physics based, is dependent on the quality of the data and range of measured conditions used for training or calibrating the model. The available period of record for forcing variables (such as rainfall and groundwater use) and hydrologic response variables can limit the range of lake water levels, groundwater levels, and spring flows that a model can accurately simulate. A long period of continuous record and large range of historical conditions are critical for developing accurate empirical models; therefore, only the sites used in cluster analyses, which satisfied these conditions, were considered for potential ANN modeling.

To assist in the selection of potential sites for more detailed analysis, SJRWMD and SWFWMD provided a list of "priority sites" that are of important management or regulatory concern, such as a well with a long period of water-level record or a lake for which a regulatory minimum water level must be established. Priority sites with longer periods of record, as well as other sites included in the cluster analysis with long-term data records, were initially selected. For the groups of sites exhibiting similar behavior as indicated by cluster analysis, representative priority sites were selected for ANN modeling. Owing to the sparseness of spring-flow data, only priority springs that also have a relatively long and complete data record were selected for ANN modeling. A final dataset of 51 sites, consisting of 22 lakes, 23 wells, and 6 springs, was selected for intensive analysis and development of ANN models [fig. 32; table 2 (table 2 in back of report)].

System Dynamics and Analysis

The behavior, or dynamics, of a natural system results from the cause-and-effect relations between multiple physical forces. For example, temporal variations in water-table altitude at a fixed location are subject to weather conditions, such as rainfall and solar radiation, and anthropogenic influences, such as artificial recharge and groundwater withdrawals, over a range of time scales from daily to multiyear. For the application of ANN models to sites in the CFCA, data-mining techniques were applied to maximize the information content in raw data while diminishing the influence of poor or missing measurements. Methods included digital filtering; the use of time derivatives, time delays, and moving averages; and comparison of differences between stations.

Signals, or time series, manifest three types of behavior: periodic, noise, and chaotic (Conrads and others, 2006). Periodic behavior is precisely predictable. Examples of periodic behavior are the diurnal sunlight and temperature patterns caused by the rising and setting sun or fluctuations in tidal water levels attributed to orbital mechanics. Noise is unpredictable. Noise refers to unexplained variations in data values and commonly is attributed to measurement error. Chaotic behavior is neither totally periodic nor noise and always has a physical cause. Variations in weather conditions, such as fluctuations in rainfall or air temperature, are examples of chaotic behavior. Chaotic behavior is to an extent predictable, especially over short time frames and prediction horizons. Such a short prediction horizon points to a classic characteristic of chaos that small permutations in the initial state of a deterministic system can lead to large changes in a later state (Motter and Campbell, 2013, p. 27).

Signal Filtering and Decomposition

Signal filtering involves removal of certain frequencies or range of frequencies from an input signal, preserving all other frequencies in the output signal. A MWA is a form of low-pass filter, which preserves the low-frequency variations of the input signal and removes the high-frequency variations, where larger window sizes lead to greater smoothing of the input signal. All MWAs in this study were trailing moving averages such that, for example, a 30-day MWA was computed as the arithmetic average of data for the current and preceding 29 days. Daily rainfall data at the 18 NOAA rain gages (fig. 24) were used to compute MWAs for a "standard" rainfall signal (arithmetic average by time step of all gages) using the following window sizes: 30; 90; 270; 630; 1,440; and 2,250 days. Using these same window sizes, MWAs of normalized rainfall (gage rainfall minus standard rainfall) also were computed for each NOAA rain gage. Multiyear trends in rainfall (for example, a 630-day MWA, fig. 33*A*) and seasonal variations in rainfall (for example, a 90-day MWA, fig. 33*B*) are clearly evident in the study area as manifest in the standard rainfall signal. Likewise, monthly groundwater-use data (fig. 29*A*) were used to compute MWAs

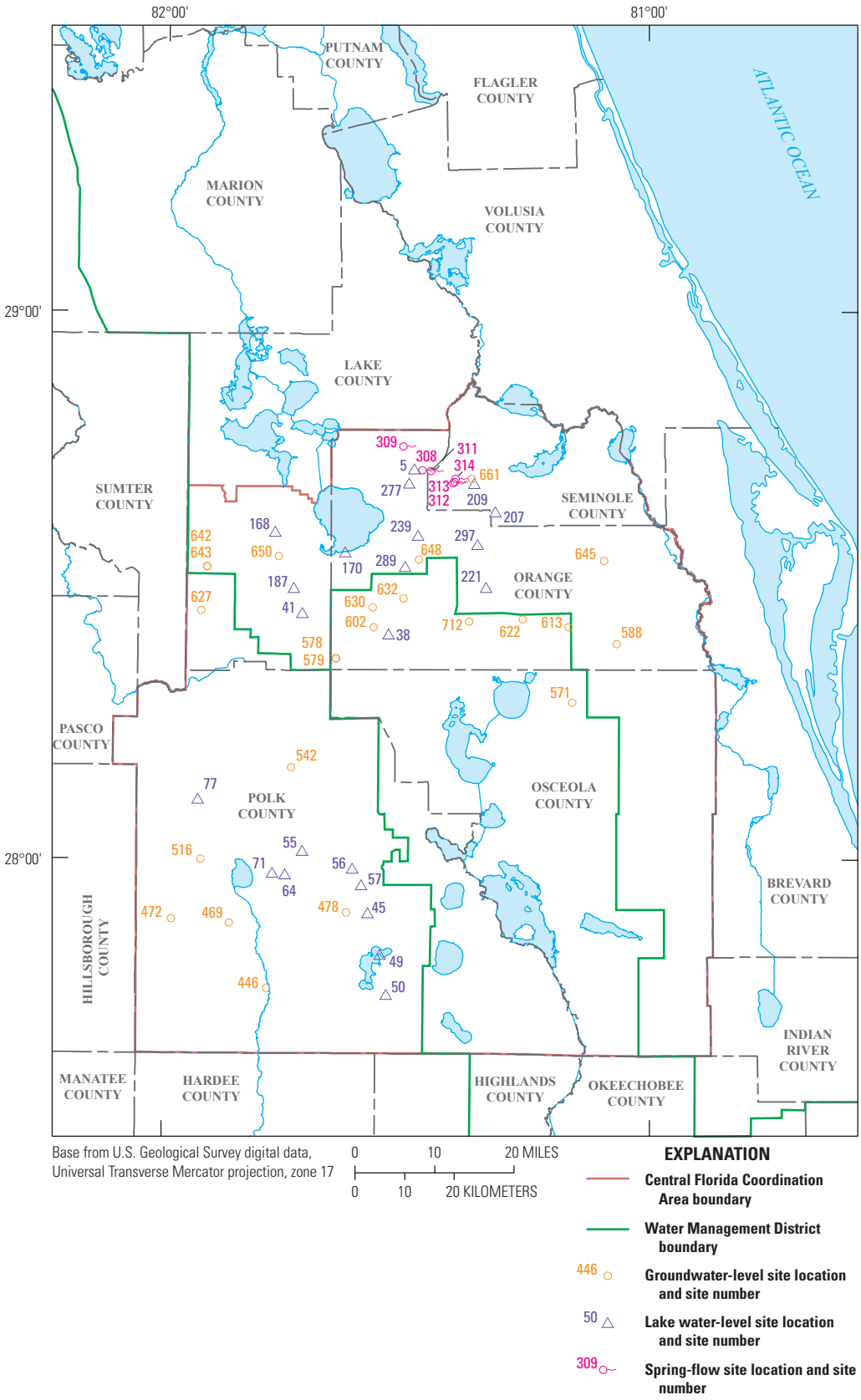


Figure 32. Sites in central Florida for which artificial neural network models were developed. Information for each site is referenced in table 2 by the site number.

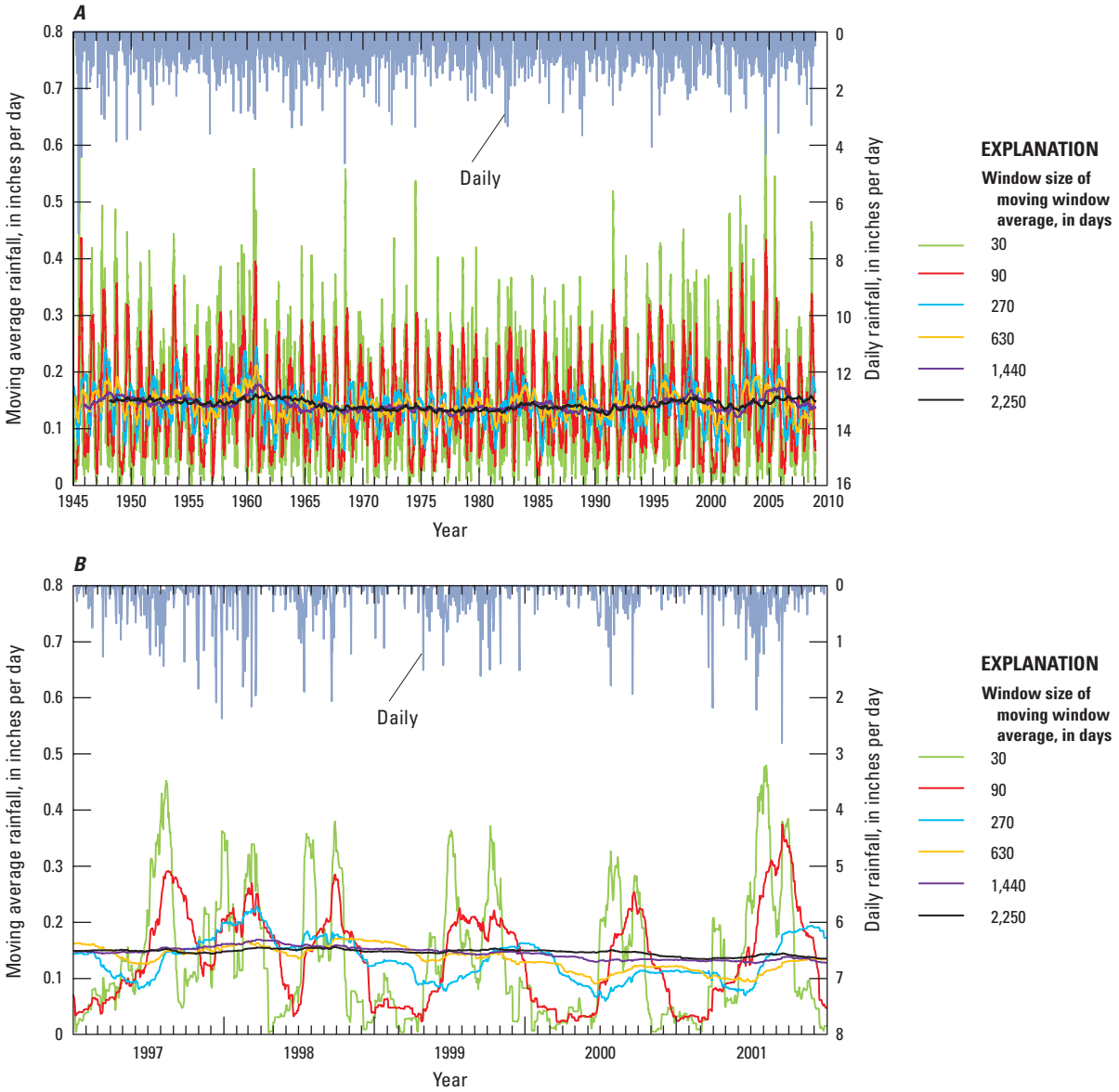


Figure 33. Average daily and low-pass filtered rainfall for the standard rainfall time series for *A*, 1945 through 2008, and *B*, 1997 through 2001. The standard rainfall time series represents the arithmetic average of daily rainfall at the 18 National Oceanic and Atmospheric Administration (NOAA) rain gages shown in figure 24.

for total groundwater-use flow rates for the study area using the following window sizes: 3, 6, 12, 24, 48, and 96 months. Similar to rainfall, multiyear trends in groundwater use (for example, 24-month MWA, fig. 34*A*) and seasonal variations in groundwater use (for example, 3-month MWA, fig. 34*B*) are clearly evident in the study area. Unlike rainfall, however, a notable long-term upward trend in groundwater use existed prior to the mid-1980s, with a steeper upward trend during the 1960s and a milder upward trend during the 1970s and early 1980s (48-month MWA, fig. 34*A*). From about the mid-1980s through 2008, long-term trends in groundwater use were slight, and during this period multiyear variability on the scale of 5 to 10 years was more common (fig. 34*A*).

Signal decomposition involves splitting a signal into subsignals, called "components," which are independently attributable to different physical forces (Conrads and others, 2006). A moving window difference (MWD) is a form of band-pass filter, which preserves variations of the input signal within a certain range of frequencies and removes variations of frequencies outside this range. Therefore, MWDs can be used to represent various frequency-band components of the forces that modulate water-level or spring-flow behavioral dynamics. In this study, all MWDs were computed from successive pairs of MWAs, for example, the 2,250-day MWA was subtracted from the 1,440-day MWA, the 1,440-day MWA was subtracted from the 630-day MWA, and so on. This process transforms the raw input signal, such as rainfall or groundwater use, into a waveform that may be more strongly correlated (than the raw input signal) with a water-level or spring-flow signal and more representative of the input-output behavior of the physical process. The behavioral dynamics manifest in the rainfall and groundwater use as represented by MWAs and MWDs were used to assess correlations with system response (lake water levels, groundwater levels, and spring flows) and ultimately for incorporation in ANN models.

Correlation Analysis

The relations between the forcing variables, response variables, and their various low-pass filtered signals are ascertained through correlation analyses to provide greater understanding of system dynamics. For example, spring flow is dependent, in part, on rainfall and groundwater use, and correlation analysis provides a measure of the relative contribution of each variable. Sensitivity analysis quantifies the relations between a dependent variable of interest and causal variables. Computing sensitivities requires defining the relation between variables through modeling.

Using statistical and (or) ANN software, correlations can be computed between factors that most influence system characteristics of interest (for example, groundwater level) and candidate combinations of controlled and uncontrolled variables (for example, rainfall and groundwater use). Correlation methods based on statistics as well as ANNs are applied. Promising results identified by the analysis are validated by comparing them to known patterns of behavior. For the current study, correlations among rainfall, groundwater use, and hydrologic response (lake water level, groundwater level, and spring flow) were computed. Daily time series extending from 1942 (or the first year of hydrologic response record) through 2008 were used for identifying correlations between rainfall and hydrologic response. Because groundwater-use data were available only at monthly resolution and over a shorter period of record, monthly averages of rainfall and hydrologic response were computed and correlations were identified for time series extending from 1957 through 2008.

The correlations between 3 rainfall signals (270-, 630-, and 1,440-day MWAs) at the 18 NOAA rain gages (fig. 24) were compared to the hydrologic response at 51 sites (fig. 32; table 2). These three rainfall signals were selected because they generally showed the highest correlations with hydrologic response based on preliminary analyses using a greater number and wider range of MWAs. Correlation coefficients were computed for all rain gage and rainfall-signal pairs, yielding a total of 54 values of R for each site (3 MWAs for each of the 18 rain gages). Maximum correlations ranged from 0.35 to 0.90, and average correlations (for all 54 R values) ranged from –0.03 to 0.61, whereas for most sites negative values were found for minimum correlations ranging from –0.45 to 0.27 (fig. 35). Negative correlations likely are spurious and have no direct causative significance; however, the relatively high correlations suggest that rainfall is an important control on hydrologic response.

Relatively high correlations exist for many rain gages and rainfall signals. For example, groundwater level at site number 469 (ROMP 59 Avon Park well, table 2; fig. 32) is most highly correlated (R values range from 0.68 to 0.76) with rainfall at three relatively nearby rain gages (NOAA gages 5973, 7205, and 9401; see figure 24 for rain gage locations) and rainfall at two distant rain gages (NOAA gages 4502 and 6628) at relatively short time scales (270- and 630-day MWAs). Groundwater level at site number 602 (Bay Lake Floridan well, table 2; fig. 32) is most highly correlated (R values range from 0.38 to 0.56) with one nearby rain gage (NOAA gage 1641) and two distant rain gages (NOAA gages 1163 and 8942) at various time scales (270-, 630-, and 1,440-day MWAs).

Correlations between rainfall and hydrologic response also were examined as a function of distance between the sites and the rain gages. Fairly high correlations between rainfall and hydrologic response were found for both nearby and distant rain gages as illustrated at three lake, well, and spring sites (fig. 36). With some exceptions, correlations tend to decrease with increasing distance between the site and the rain gage, as indicated by the negative trends shown in figure 36 and by the predominantly negative R values shown in figure 37. Correlations that have an upward trend with distance (positive R value in figure 37) may be a consequence of other factors such as groundwater use. If groundwater use is a significant factor affecting water levels and flows, it is possible that low positive correlations or even negative correlations with nearby rainfall may actually represent the confounding effects of inverse correlation between rainfall and

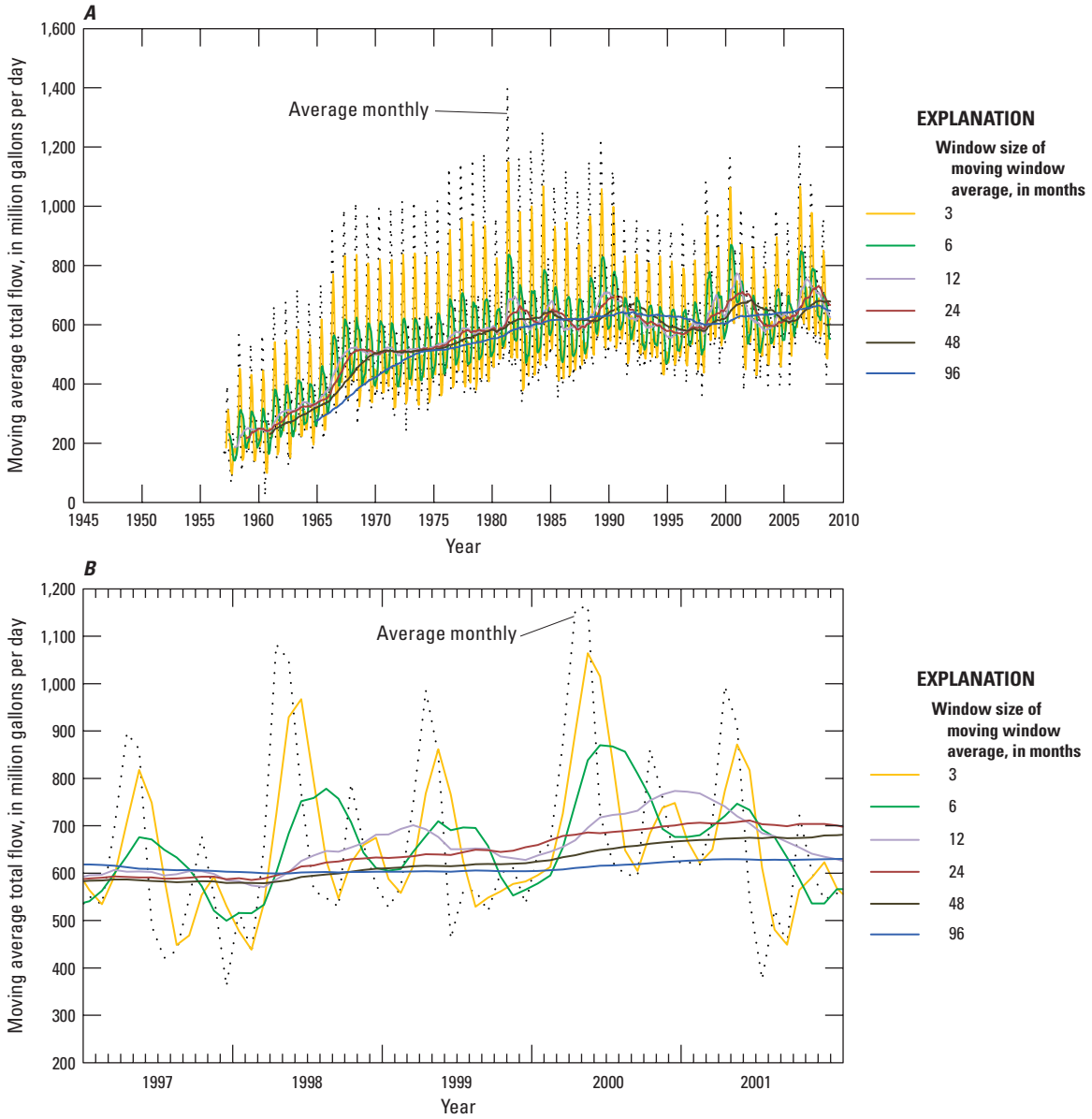


Figure 34. Average monthly and low-pass filtered groundwater use in the Central Florida Coordination Area and vicinity for the sum total of all water-use categories for *A*, 1957 through 2008, and *B*, 1997 through 2001. Data for individual water-use categories shown in figure 29*A*.

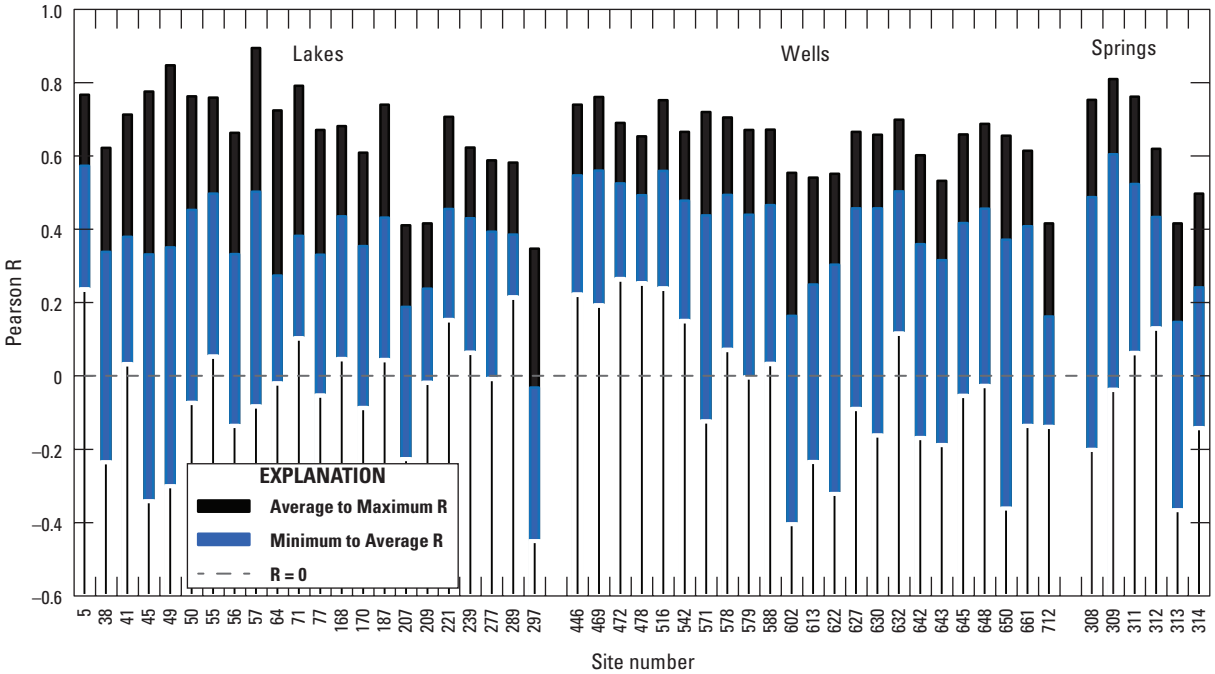


Figure 35. Pearson product-moment correlation coefficient (R) between rainfall (for each rain gage and rainfall-signal pair) and hydrologic response at each site. Ranges are depicted indicating those values falling between the minimum and average R for the site and those values falling between the average and maximum R for the site. Rainfall signals are 270-, 630-, and 1,440-day moving window averages of rainfall at each of the 18 National Oceanic and Atmospheric Administration (NOAA) rain gages (locations shown in figure 24). Site information and locations are shown in table 2 and figure 32.

nearby groundwater use. Such spurious correlations may be more common at nearby rain gages than at distant rain gages where groundwater-use effects may tend to be attenuated by distance. In general, however, correlations with distance are weak, where 78 percent of sites have R values less than 0.3 in absolute value (fig. 37). A possible explanation for the generally weak correlations with distance is that rainfall at the location of the hydrologic response may be correlated with rainfall at a distant location. Rainfall signals (30-, 90-, 270-, 630-, 1,440-, and 2,250-day MWAs) from the 18 NOAA gages are correlated between sites, where R values range from –0.31 to 0.91. Only 25 percent of the rainfall signals among all 18 NOAA gages have R values greater than 0.5 and 6 percent have R values exceeding 0.75, suggesting that correlation of rainfall between gages is not particularly strong (consistent with the cluster analysis that indicated rainfall data are highly variable in space and time). All these results taken together suggest an “interconnectedness” of the hydrologic system in which rainfall forcing may possibly be transmitted over long distances (up to about 100 miles based on the extent of the current study area). Such behavior is not inconsistent with that of a semiconfined or confined karst aquifer, such as the Floridan aquifer system in central Florida, where pressure transients move relatively quickly through the system owing to the high transmissivity and low storativity properties of the hydrogeologic units.

Groundwater use is a potentially important forcing that may influence behavior of the hydrologic system. Correlations were examined between groundwater-use signals (monthly values and 3-, 6-, 12-, 24-, 48-, and 96-month MWAs for the sum total of all groundwater-use types; fig. 34*A*) and the hydrologic response at 51 sites (fig. 32; table 2). The correlation was computed for each groundwater-use signal, yielding a total of seven R values for each site (monthly signal plus six MWAs). Correlations generally are negative because water levels and flows will decrease in response to an increase in groundwater withdrawals. Maximum correlations ranged from –0.57 to 0.49 and average correlations (for all seven R values) ranged from –0.73 to 0.30, whereas minimum correlations were negative for nearly all sites and ranged from –0.82 to 0.12 (fig. 38). Positive correlations likely are spurious and have no direct causative significance. The relatively strong negative correlations suggest that groundwater use is an important control on hydrologic response, although correlations with groundwater use generally are lower (in absolute value) than with rainfall. However, cross correlation with rainfall may affect the relative explanatory strength of groundwater use or rainfall for modeling hydrologic response.

Correlated variation between groundwater use and rainfall for a range of time scales was examined to ascertain the degree to which there may be confounding interactions between these two fundamental forcing variables and

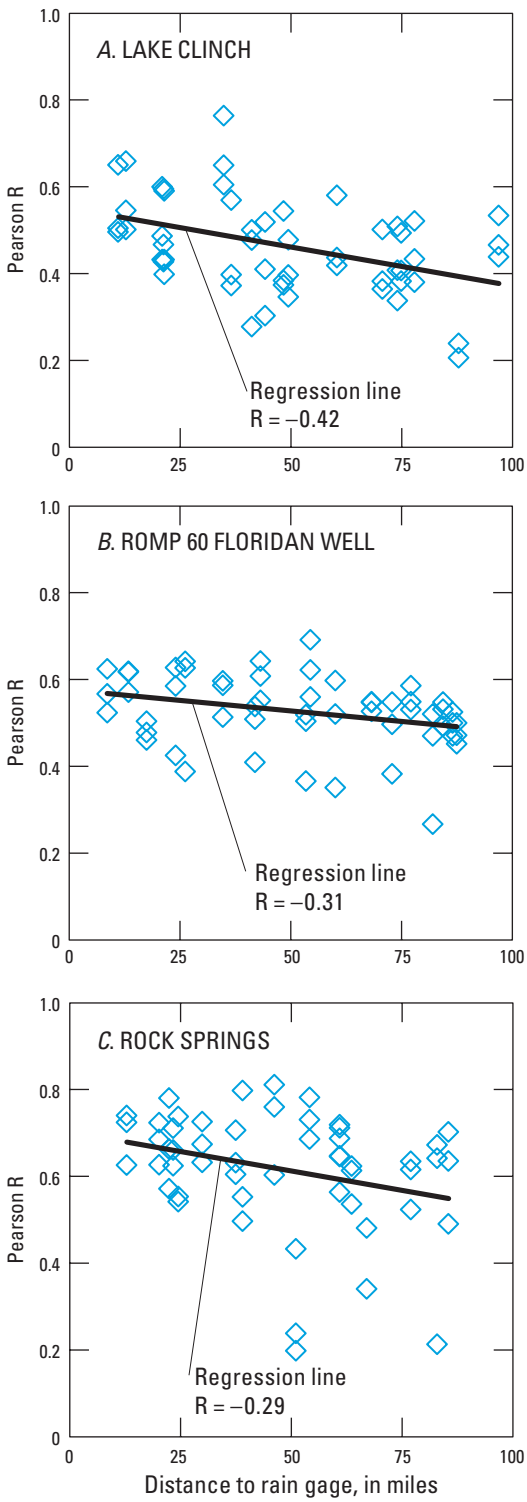


Figure 36. Comparison of the Pearson product-moment correlation coefficients (R) between rainfall (for each rain gage and rainfall-signal pair) and hydrologic response and the distance between the site and each rain gage for *A*, Lake Clinch (site number 50), *B*, ROMP 60 Floridan well (site number 472), and *C*, Rock Springs (site number 309). Site information and locations are shown in table 2 and figure 32.

hydrologic response. Correlations were examined between rainfall signals (30-, 90-, 270-, 630-, 1,440-, and 2,250-day MWAs) at the 18 NOAA rain gages (fig. 24) and groundwater-use signals (monthly values and 3-, 6-, 12-, 24-, 48-, and 96-month MWAs for the sum total of all groundwater-use types; fig. 34*A*). The correlation was computed for each rain gage and rainfall-signal pair using monthly averages computed from the MWAs of daily rainfall data (fig. 33*A*), yielding a total of 42 values of R for each rain gage (6 rainfall MWAs for each of the 7 groundwater-use MWAs). Correlations between rainfall and groundwater use tended to be negative, likely attributable to rainfall control of certain types of water use (for example, irrigation) such that groundwater withdrawals increase in response to a decrease in rainfall (fig. 39). For example, the strongest negative correlation of –0.76 occurs for NOAA gage 1163 (2,250-day MWA of rainfall) for a groundwater-use 24-month MWA. At all rain gages, however, positive correlations were found for at least some rainfall signals. For example, the strongest positive correlation of 0.71 occurs for NOAA gage 2229 (2,250-day MWA of rainfall) for a groundwater-use 96-month MWA. Positive correlations tend to be more prevalent for rain gages in the eastern part of the study area (cluster analysis groups 1 and 2, figs. 24, 39), where several gages had upward trends in long-term rainfall (fig. 13). Conversely, negative correlations tend to be more prevalent for rain gages in the western part of the study area (cluster analysis groups 3 and 4, figs. 24, 39), where several gages showed downward trends in long-term rainfall (fig. 13). Correlations tend to be stronger (both negative and positive) with increasing groundwater-use MWA size (see correlations composited for all rain gages shown for each MWA size in figure 39), reflecting long-term trends in both rainfall and groundwater use. It is important to note, though, that there may not be a causative relation at this time scale. Other factors, such as trends in population, likely influence long-term trends in groundwater use more than do weather variations.

State-Space Reconstruction

Achieving accurate predictions of the behavior of dynamic systems can be notoriously difficult owing to chaotic behavior (Motter and Campbell, 2013). Difficulties range from microscale problems of computing point velocities in turbulent pipe flow to macroscale problems of computing mass and energy flows in regional hydrologic systems. Chaos theory provides a conceptual framework called “state-space reconstruction” for representing dynamic relations in systems with inherent chaotic behaviors (Kugiumtzis and others, 1994; Conrads and others, 2006). Data collected at a point in time can be organized as a vector of measurements; for example, element one of the vector might be the water level, element two the rainfall, and so on. Engineers assert that a process evolves from one state to another, through time, and that a vector of measurements, also referred to as a “state vector,” represents the process state at the moment the measurements were taken. A sequence of state vectors represents a “state

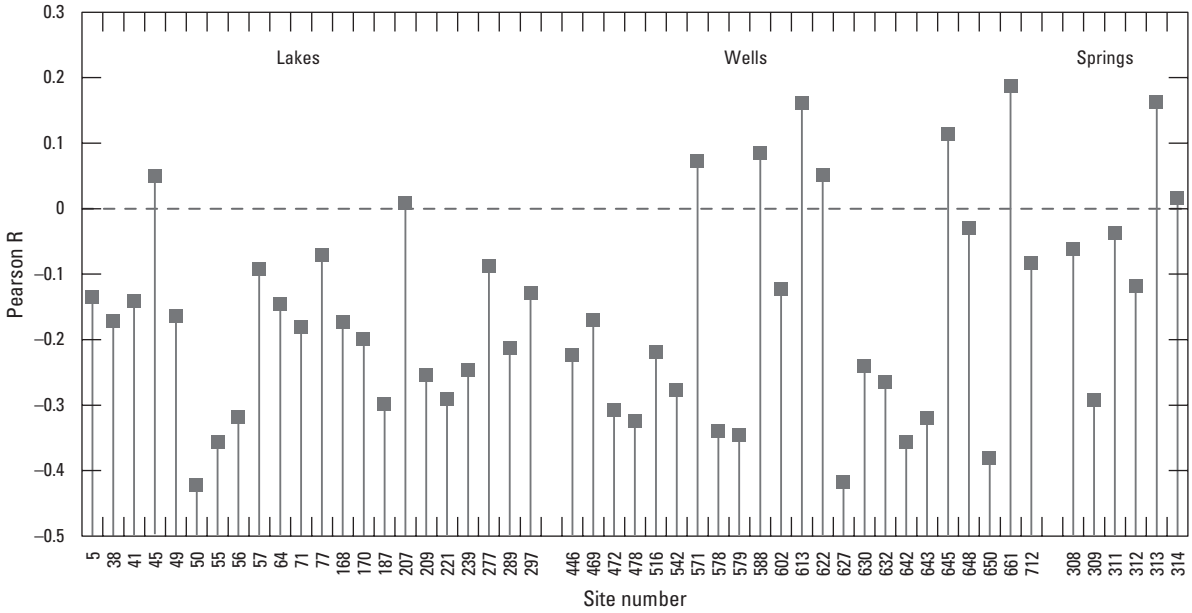


Figure 37. Pearson product-moment correlation coefficient (R) between the R shown in figure 35 (between rainfall and hydrologic response) and the distances between the site and each rain gage for each site. Site information and locations are shown in table 2 and figure 32.

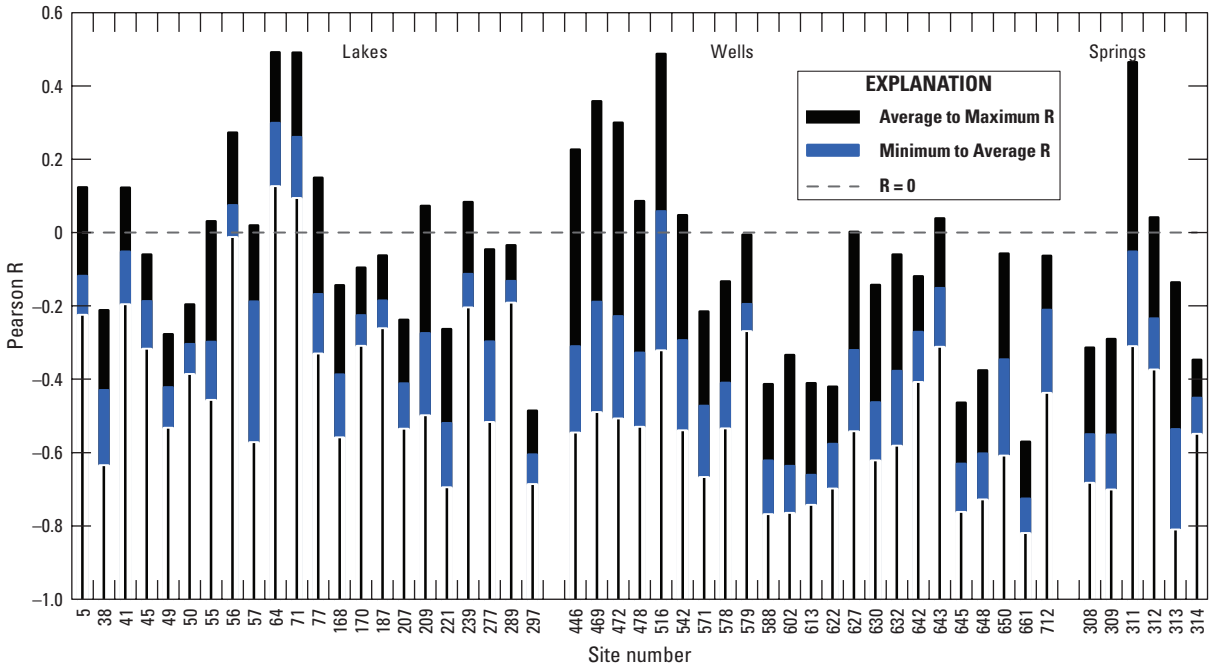


Figure 38. Pearson product-moment correlation coefficient (R) between groundwater-use signals and hydrologic response at each site. Ranges are shown for those values falling between the minimum and average R for the site and those values falling between the average and maximum R for the site. Groundwater-use signals are the monthly data and 3-, 6-, 12-, 24-, 48-, and 96-month moving window averages of the sum total groundwater use (fig. 34*A*). Site information and locations are shown in table 2 and figure 32.

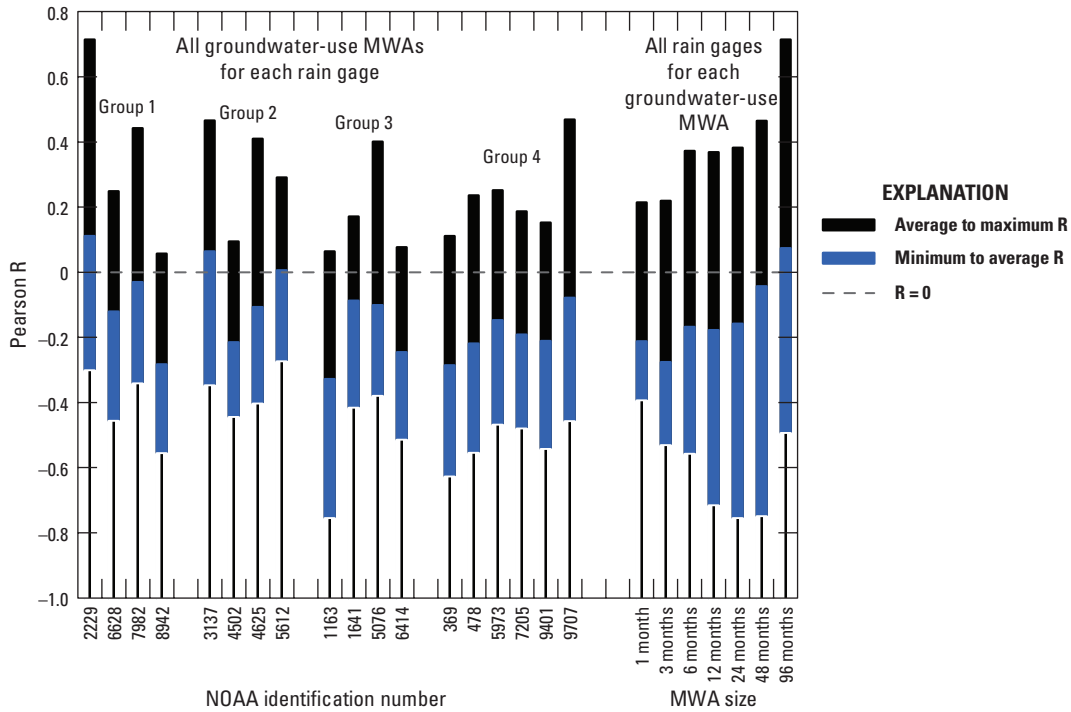


Figure 39. Pearson product-moment correlation coefficient (R) between groundwater-use signals and rainfall signals at each National Oceanic and Atmospheric Administration (NOAA) rain gage. Ranges are shown for those values falling between the minimum and average R for the site and those values falling between the average and maximum R for the site. Groundwater-use signals are the monthly data and 3-, 6-, 12-, 24-, 48-, and 96-month moving window averages (MWAs) of the sum total groundwater use (fig. 34*A*), and rainfall signals are monthly averages of 30-, 90-, 270-, 630-, 1,440-, and 2,250-day MWAs of daily rainfall (fig. 33*A*).

history." Mathematicians define the state vector as a point in a "state space" with a number of dimensions equal to the number of elements in the vector. For example, eight vector elements equate to eight dimensions. Engineers, mathematicians, and scientists use this conceptual framework to develop empirical models of the behavior of dynamic systems. Empirical modeling is the fitting of a multidimensional surface to the points arrayed in state space.

Chaos theory proposes that a process can be optimally represented (reconstructed) by a collection of state vectors, $Y(t)$, using an optimal number of measurements equal to "local dimension" d_L that are spaced in time by integer multiples of an optimal time delay, τ_d (Abarbanel, 1996, p. 4–12, 39). The sets of k parameters $\{d_{L_i}\}$ and $\{\tau_{d_i}\}$ for i = 1,2,..., k are called "dynamical invariants," which characterize the behavior of a process in state space similar to how the amplitude, frequency, and phase angle can be used to define a periodic time series. Thus, for a multivariate process of k independent variables:

$$Y(t) = \left\{ \begin{array}{l} \left[x_1(t), x_1(t-\tau_{d_1}), \ldots, x_1(t-(d_{L_1}-1)\tau_{d_1}) \right], \ldots, \\ \left[x_k(t), x_k(t-\tau_{d_k}), \ldots, x_k(t-(d_{L_k}-1)\tau_{d_k}) \right] \end{array} \right\}, \quad (1)$$

where each $x_i(t,\tau_{d_i})$ for i is 1,2,..., k represents a different dimension in state space and, therefore, a different element in a state vector. Values of $\{d_{L_i}\}$ and $\{\tau_{d_i}\}$ are estimated analytically or experimentally from the data. The mathematical formulations for models are derived from these state vectors. To predict a dependent variable of interest $y(t)$ from prior measurements (also known as forecasting) of k independent variables (Roehl and Conrads, 2006):

$$y(t) = F \left\{ \begin{array}{l} \left[\begin{array}{l} x_1(t-\tau_{p_1}), x_1(t-\tau_{p_1}-\tau_{d_1}), \\ \ldots, x_1(t-\tau_{p_1}-(d_{M_1}-1)\tau_{d_1}) \end{array} \right], \ldots, \\ \left[\begin{array}{l} x_k(t-\tau_{p_k}), x_k(t-\tau_{p_k}-\tau_{d_k}), \\ \ldots, x_k(t-\tau_{p_k}-(d_{M_k}-1)\tau_{d_k}) \end{array} \right] \end{array} \right\}, \quad (2)$$

where F is an empirical function such as an ANN, each $x_i(t,\tau_{p_i}, \tau_{d_i})$ for i = 1,2,..., k is a different input to F, and $\{\tau_{p_i}\}$ for i = 1,2,..., k is another time delay. For each variable, $\{\tau_{p_i}\}$ is specified according to one of the following constraints: time delay at which an input variable becomes uncorrelated to all other inputs, but can still provide useful information

about $y(t)$; time delay of the most recent available measurement of x_i; or time delay at which an input variable is most highly correlated to $y(t)$. In equation 2, the state-space local dimension $\{d_{L_i}\}$ of equation 1 is replaced with a model input variable dimension $\{d_{M_i}\}$, which is determined experimentally. It is noted that $\{d_{M_i}\} \leq \{d_{L_i}\}$ for $i = 1,2,..., k$, and $\{d_{M_i}\}$ tends to decrease with increasing k. The final model for the predicted variable $y(t)$ can be a superposition of disparate behaviors, $y_j(t)$, originating from n different forcing functions (Roehl and Conrads, 2006):

$$y(t) = \sum_{j=1}^{n} y_j(t) = \sum_{j=1}^{n} F_j \quad . \tag{3}$$

In different terms, a goal of state-space reconstruction is to determine the optimal temporal spacing (time delays) and number of successive measurements (dimension) of an input signal required to satisfactorily predict an output signal, which amounts to solution of an inverse problem (Lillekjendlie and others, 1994). State-space reconstruction is focused on univariate systems, that is, the system is described by a single time-series variable and the prediction being made is the system's state at a near-future time. This "delays and dimension" conceptual framework is applied in this study to a complex multivariate system where multicollinearity among input signals (rainfall and groundwater use) is common. Rather than explicit identification of time delays and dimensions, however, signals and frequency-band components that best predict the observed hydrologic response are selected from candidate MWAs and MWDs of rainfall and groundwater use using a method to minimize correlated inputs. Ultimately, the synthesis of one or more empirical functions (formalized by equations 2 and 3 in the context of state-space reconstruction) by judicious construction and training of an ANN is the key task in developing a model useful for simulating system behavior and is discussed in further detail in the following sections.

Input-Output Mapping and Problem Representation

The development of ANN models to predict the water level or flow at the 51 sites (table 2) was undertaken in two phases, resulting in a two-stage model architecture (fig. 40). The two-stage approach was motivated in part by the different temporal resolutions of the forcing time series, consisting of daily rainfall and monthly groundwater use. The first phase was to train a daily resolution ANN for each site to simulate the observed hydrograph using daily rainfall inputs, herein called the "rainfall model." The second phase was to train a monthly resolution ANN, herein called the "groundwater-use model," for each site to simulate the residual error time series from the rainfall model using monthly groundwater-use inputs. At a few sites, the predicted water level or flow from the rainfall model also was used as an input. The final simulated water level or flow is the sum of the predictions of both ANNs for the site (fig. 40), thus following the superposition approach of equation 3 where each ANN represents the empirical function F of equation 2. The residual error, also called "the residuals," is a model's prediction error and represents variability in an output that is not accounted for by variability in the inputs. Residuals are simply calculated by subtracting the predicted values from the measured data. By predicting variability in the residuals, a groundwater-use model predicts hydrograph variability that is unaccounted for by rainfall, but can be accounted for by groundwater use.

The decision to place the rainfall model first in the two-stage model sequence, rather than model groundwater use first or both rainfall and groundwater use simultaneously, was based on the following constraints and rationale:

- rainfall data records generally are more complete (fewer missing data), and the data are of higher quality (more measured and fewer estimated values) and finer temporal resolution than the groundwater-use data;
- a paucity of previously compiled historical monthly water-use data in the CFCA required a substantial effort by SJRWMD and SWFWMD to compile reported data and estimate missing values, most of which is reported by McLeod and Munch (2012) and Munch (2014), resulting in the groundwater-use data not being available until late in the study;
- groundwater return flows—groundwater withdrawals that subsequently are returned to the aquifer system, such as by land application of treated wastewater or deep percolation of irrigation—are poorly known, and few data are available to enable estimation of historical temporal and spatial patterns of these return flows;
- water-budget analyses derived from physics-based model simulations for central Florida (Knowles and others, 2002; McGurk and Presley, 2002; and Sepúlveda and others, 2012) indicate groundwater recharge at the water table (from rainfall and artificial recharge) is the largest inflow to the aquifer system (79 to 97 percent of total inflow for the combined surficial and Floridan aquifer systems), and groundwater withdrawals are a small percentage (6 to 12 percent) of flow in the aquifer system (combined surficial and Floridan aquifer systems) during years with normal rainfall; and
- the variability of rainfall is much greater (more than an order of magnitude) than the variability of groundwater use over a wide range of time scales from monthly to multiyear.

For empirical correlation-based models such as ANNs, the use of independent knowledge is essential to constrain model results to established or hypothesized physics. The correlation analyses indicated that hydrologic response was correlated with both rainfall (fig. 35) and groundwater use (fig. 38),

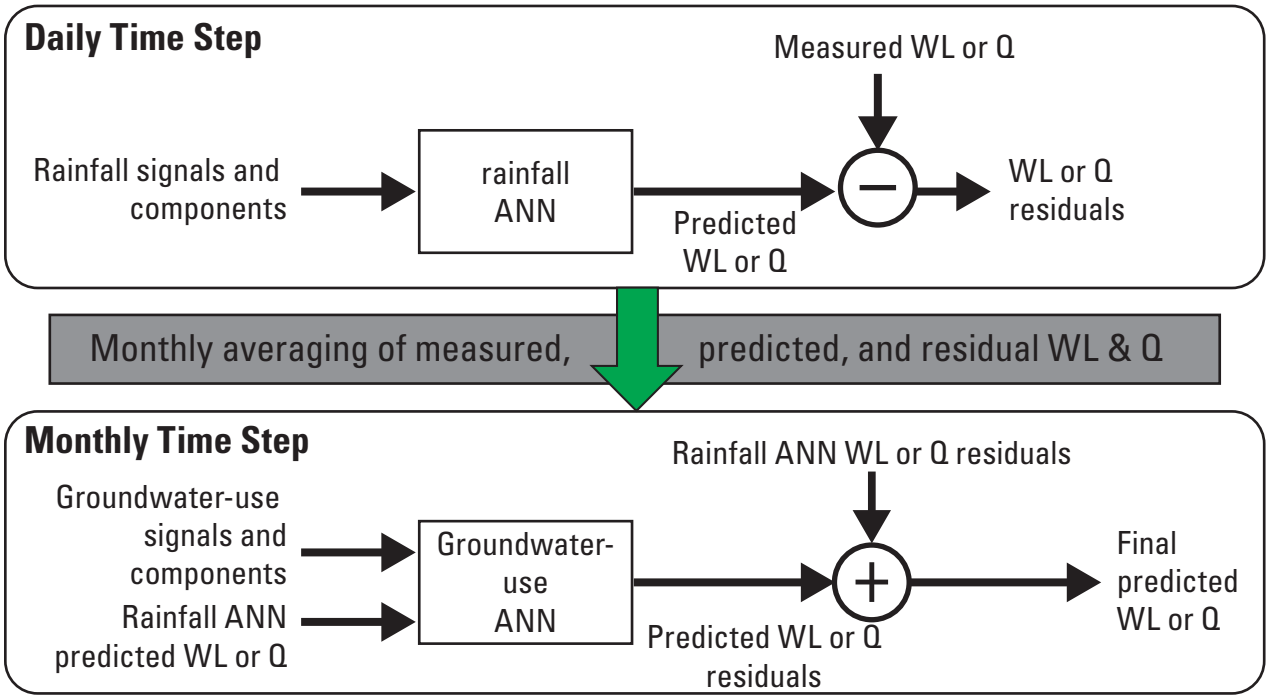


Figure 40. TDiagram wo-stage model architecture used to simulate water level at lakes or wells, or flow at springs. ANN, artificial neural network; Q, spring flow; WL, lake water level or groundwater level.

and that rainfall and groundwater use were correlated with each other (fig. 39). Additionally, other factors, such as population growth and landscape change, can confound interpretation of the actual causative relations. In this study, previous independent water-budget analyses provided a physics-based constraint on ANN development and influenced the selection of the two-stage model architecture. Several physics-based models based on the USGS three-dimensional, finite-difference groundwater flow model code MODFLOW (Harbaugh and McDonald, 1996; Harbaugh and others, 2000; Harbaugh, 2005) have been developed for central Florida that cover part or nearly all of the CFCA. The models were developed to simulate either steady-state conditions, where model inputs and outputs are constant over time, or transient conditions, where model inputs and outputs vary over a selected time period and changes in water stored in the aquifer are simulated. The Lake County/Ocala National Forest (LCONF) model developed by Knowles and others (2002) covers all of Lake County and the Ocala National Forest and parts of adjoining counties, and simulates steady-state, groundwater flow for average 1998 conditions (December 1997 to December 1998). The east-central Florida (ECF) model developed by McGurk and Presley (2002) covers all of Orange and Seminole Counties, most of Brevard and Lake Counties, and parts of adjoining counties, and simulates steady-state, groundwater flow for average 1995 conditions (January 1995 to December 1995). The east-central Florida transient (ECFT) model developed by Sepúlveda and others (2012) covers all of the CFCA (except the westernmost edge in Polk County) and parts of adjoining counties, and simulates monthly groundwater flow, lake and stream water levels, and streamflow from January 1995 through December 2006. All these models actively simulate the surficial and Floridan aquifer systems; therefore, water budgets of the entire aquifer system could be compared for their respective simulation periods (fig. 41). It is important to note that the ECFT model simulates changes in groundwater storage that constitute substantial components of aquifer inflow (release of water stored in the aquifer) and outflow (increase in volume of water stored in the aquifer). These storage changes can influence the temporal dynamics of the aquifer system even though the net change in storage over long periods typically is small; for example, the difference in storage between inflow and outflow in the ECFT model from 1995 through 2006 was 0.08 inch per year (in/yr) (6.32 in/yr storage inflow minus 6.24 in/yr storage outflow, fig. 41). In contrast, the steady-state LCONF and ECF models do not simulate storage changes, and only average conditions during the respective simulation periods are represented by the model based on the assumption that net storage effects are small. Despite these differences in conceptual model, simulation period, and geographic area, all three models indicate that groundwater recharge is the largest inflow (79 to 97 percent) to the aquifer system (fig. 41*A*). Groundwater recharge is nearly exclusively derived from infiltration of rainfall because artificial recharge (land based application of reclaimed water) constitutes only 0.1 to 0.4 in/yr. Groundwater withdrawals are a relatively small component of aquifer

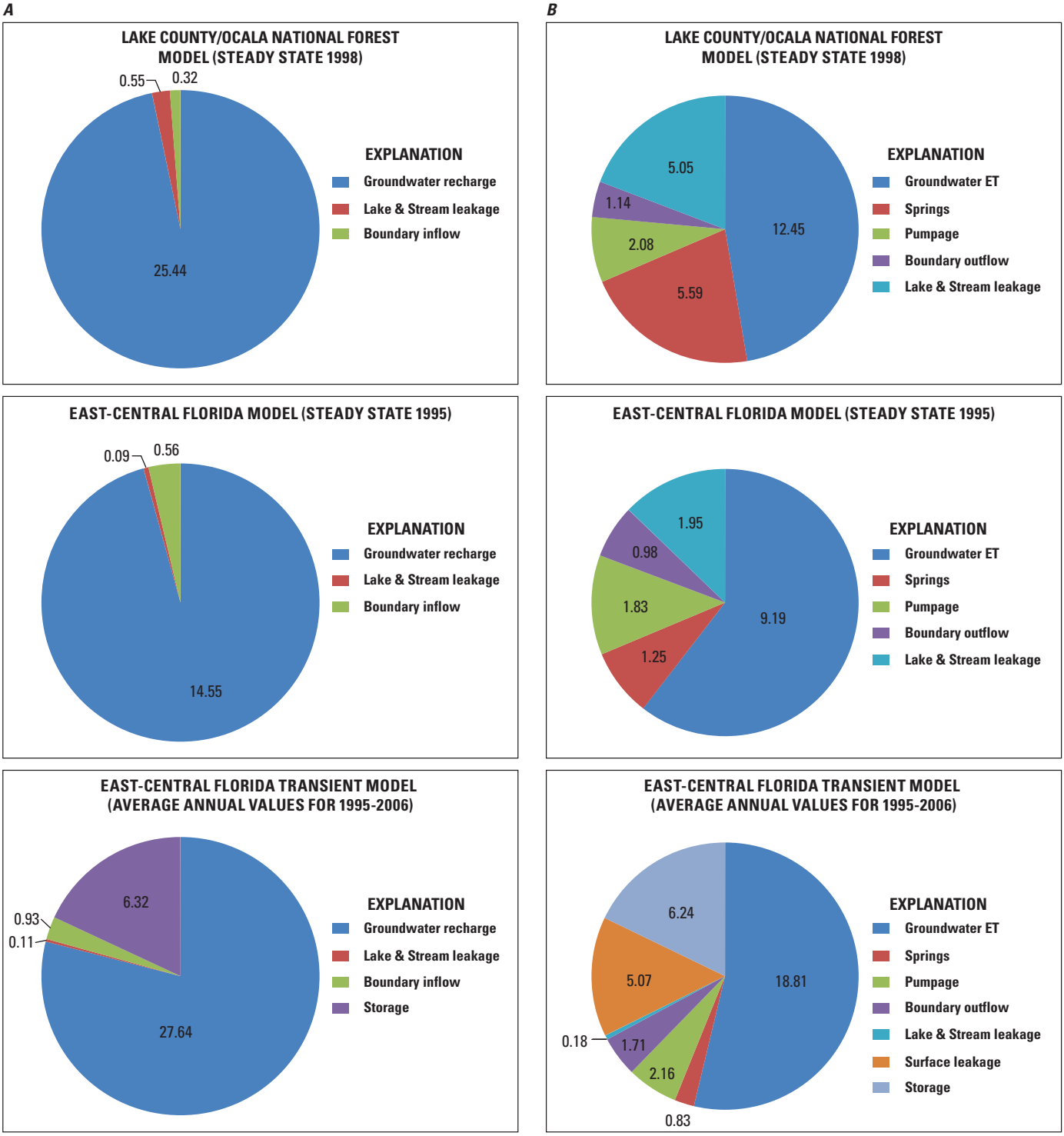


Figure 41. Water budgets for the combined surficial and Floridan aquifer systems simulated by the Lake County/Ocala National Forest (LCONF; Knowles and others, 2002), east-central Florida (ECF; McGurk and Presley, 2002), and east-central Florida transient (ECFT; Sepúlveda and others, 2012) models for *A*, inflows, and *B*, outflows. Flows expressed in equivalent inches per year based on surface area of surficial aquifer system simulated by the model (approximately 4,800 square miles (mi^2) for the LCONF model, 6,480 mi^2 for the ECF model, and 9,004 mi^2 for the ECF model). Surface leakage simulated by the ECFT model conceptually represents groundwater seepage at model cells where the water table rises above land surface and drains to lakes, streams, wetlands, and closed basins. Sepúlveda and others (2012, p. 124 and 128) report that 24 percent of surface leakage is to cells in closed basins or cells draining to streams or lakes outside the model area. Thus total leakage to cells draining to lakes, streams, and wetlands within the model area would be the sum of lake and stream leakage and 76 percent of surface leakage, or 4.05 inches per year, which is comparable to lake and stream leakage rates simulated by the LCONF and ECF models.

outflow (6 to 12 percent), and ET is the largest loss of water (47 to 60 percent of aquifer outflow) (fig. 41*B*). Because rainfall-derived recharge is the largest component of the aquifer water budget, the two-stage model architecture was developed using rainfall first followed by groundwater use (fig. 40). This approach provided models more consistent with generally well accepted physics-based models and mitigated potential spurious results of a pure correlation-based approach.

Hydrologic response is driven by variability in system forcings in addition to the relative magnitudes of the forcings. Therefore, a comparison of the temporal variability of rainfall and groundwater use was made over a range of time scales. In order to make an even comparison, monthly rainfall was computed as an arithmetic average of daily values (based on standard rainfall, fig. 33*A*), and monthly groundwater use (fig. 34*A*) was converted to a flux by dividing the volumetric rate by the total area where withdrawals historically have occurred, which corresponds to 53 grid cells of the GUDV (appendix 2) or 8,415 mi^2. The standard deviation was computed for each monthly signal and for 3-, 6-, 12-, 24-, 48-, and 96-month MWAs. For the period of contemporaneous monthly data extending from 1957 through 2008, the standard deviation of rainfall was 37.62 in/yr compared to 0.59 in/yr for groundwater use (table 3). Likewise, over longer time scales, the variability of rainfall is much greater than that of groundwater use based on the standard deviations of low-pass filtered signals (table 3). Even if groundwater use were distributed over an area only one-tenth the size of the historical coverage (making the standard deviations for groundwater use in table 3 ten times larger), only for the 96-month MWA would the variability of groundwater use exceed that of rainfall. Therefore, not only is rainfall-derived groundwater recharge the largest component of the water budget for the groundwater system in central Florida for years with normal rainfall, but also the temporal variability of rainfall far exceeds that of groundwater use over a wide range of time scales from monthly to multi-year. These factors provide further basis for modeling rainfall first followed by groundwater use for the two-stage model architecture (fig. 40).

Other potential causes of variability in hydrologic response include meteorological forcing and groundwater use outside the study area as well as changes in land use and land cover. It was beyond the scope of the current study to explore many of the other potential explanatory variables, but experiments were conducted that evaluated the use of air temperature and estimated PET in simulating the observed hydrographs. The estimated PET was computed using the temperature-based Hargreaves equation and, therefore, is not independent of air temperature. Additionally, air temperature is highly correlated to MWAs of rainfall in the study area. The experiments found that model output sensitivities to inputs representing air temperature and estimated PET were low compared to those representing rainfall. Thus, including air temperature and estimated PET as explanatory variables did not substantially improve prediction accuracy, and these variables were not included in the final ANN models.

Table 3. Standard deviation of rainfall and groundwater-use signals in central Florida.

[MWA, moving window average; --, not available; rainfall signals based on standard rainfall shown in figure 33*A*; groundwater-use signals based on sum total of all water use categories shown in figure 34*A*; groundwater use converted to a flux by dividing by the total area where withdrawals historically have occurred, which corresponds to 53 grid cells of the Groundwater-Use Data Viewer (appendix 2) or 8,415 square miles]

Data range	Data type	Standard deviation, in inches per year	
		Rainfall	Groundwater use
1942–2008	monthly	37.98	--
1957–2008	monthly	37.62	0.59
1965–2008	monthly	37.57	0.54
1965–2008	3-month MWA	26.75	0.40
1965–2008	6-month MWA	18.11	0.26
1965–2008	12-month MWA	8.10	0.19
1965–2008	24-month MWA	5.64	0.19
1965–2008	48-month MWA	3.52	0.20
1965–2008	96-month MWA	2.13	0.24

Artificial Neural Network Models

ANN models are empirical models that are developed directly from data. The most common empirical modeling approach is a linear model fit using least squares (Hastie and others, 2001), which relates variables using straight lines, planes, or hyper-planes, whether the actual relations are linear or not. Sahoo and Jha (2013) conducted a comparison of linear and ANN models and noted the superior performance of the ANN technique over multiple linear regression for the simulation of transient groundwater levels. The principal advantages of empirical models, such as ANN models, over physics-based models are that they can be developed faster and typically are more accurate provided that the modeled systems are well characterized by data (Conrads and others, 2006). Empirical models, however, are prone to problems when poorly applied. Overfitting and multicollinearity caused by correlated input variables can lead to invalid “mappings,” or relations, between input and output variables (Roehl and others, 2003).

An ANN model is a flexible mathematical structure capable of describing complex nonlinear relations between input and output datasets (Conrads and others, 2006). The structure of ANN models is loosely based on the biological nervous

system with interconnections of neurons and synapses (Hinton, 1992). Although numerous types of ANN models exist, the most commonly used type of ANN model is the multilayer perceptron, which is used in this study and described in detail by Jensen (1994). Multilayer perceptron ANNs can synthesize functions to fit multidimensional, nonlinear data. Devine and others (2003) and Roehl and Conrads (1999) describe the application of multilayer perceptron ANNs to simulate and control combined manmade and natural systems, including disinfection byproduct formation, industrial air emissions monitoring, and surface-water systems affected by point- and nonpoint-source pollution.

Multilayer perceptron ANNs are constructed from layers of interconnected processing elements called neurons that execute a simple "transfer function" (fig. 42). All input-layer neurons are connected to all hidden-layer neurons, and all hidden-layer neurons are connected to all output-layer neurons. Multiple hidden layers are possible, but a single hidden layer as shown in figure 42 is sufficient to address most problems (Bishop, 1995, p. 132).

Typically, as well as for this study, linear transfer functions are used to scale input values from the input layer to the hidden layer, yielding values that generally fall within the range that corresponds to the most linear part of the s-shaped sigmoid transfer functions (hyperbolic tangent) used to scale values from the hidden layer to the output layer (fig. 42). Each connection has a "weight," w_i, associated with it, which scales the output received by a neuron from a neuron in an antecedent layer. The output of a neuron is a simple combination of the values it receives through its input connections and the associated weights, as well as the neuron's transfer function.

An ANN is "trained" by iteratively adjusting its weights to minimize the error by which it maps inputs to outputs for a dataset composed of input/output vector pairs. Prediction accuracy, during and after training, can be measured by several metrics, including coefficient of determination (R^2) and RMSE. An algorithm that is commonly used to train multilayer perceptron ANNs is the error back-propagation training algorithm (Rumelhart and others, 1986; Bishop, 1995, p. 140–141). This algorithm optimally minimizes the error in the objective function by adjusting the weights into and out of the hidden layer of the model (fig. 42).

Experimentation with various ANN model architectural and training parameters is a typical part of the modeling process. For correlation analysis or predictive modeling applications, a number of potential ANN models are trained and evaluated for their statistical accuracy and their representation of dynamics of the system. Interactions between combinations of variables are considered. The models here were calibrated using a form of K-fold cross-validation, which involves partitioning the data into K approximately equal-size subsets and training the model recursively on K – 1 subsets starting with subset 1 up to K (Hastie and others, 2001, p. 214). A value of 2 was used for K (2-fold cross-validation), resulting in a training dataset that is evaluated with a testing dataset. Rather than using equal-size training and testing datasets, however, the data were partitioned based on the characteristics of data. For models with a large amount of data that adequately represents the range of historical behaviors, a small percentage of the data (10–30 percent) was selected for the training dataset, with the balance assigned to a testing dataset. For models with limited data, a larger percentage of the data (greater than 75 percent) was used in the training dataset. In some cases, the data were too sparse to allow use of a training dataset. There are many strategies for selecting the individual data points composing the training and testing datasets, but a common method is random selection of a specified percentage of the total population of measurements (Hastie and others, 2001). Serial correlation, which is common in hydrologic time-series data, influences the degree of independence between the training and testing datasets and can result in overly optimistic error statistics for the testing dataset potentially causing poor generalization (or overfitting) of the model. Using disjoint time periods for the training and testing datasets can mitigate this effect, but such an approach also limits the range of historical behaviors to which the model is trained. For the ANN models developed here, the goal was representing to the fullest extent possible the effects of the limited number of permutations of extended high and low forcing events (rainfall and groundwater use); thus, random selection throughout the historical period of record was used to populate the training and testing datasets. Other measures taken to guard against overfitting are described in a later section on ANN model construction and training.

To mitigate extrapolation and data sparseness issues, the ANN models were conservatively trained using a method called "stop training" or "early stopping" (Bishop, 1995, p. 343), thus allowing the ANN to both fit the data and extrapolate in a minimally nonlinear and, therefore, predictable manner. Stop training simply means stopping the training process before the ANN has fit the data to the maximum extent possible. Training is commonly stopped when the testing dataset, rather than the training dataset, R^2 or RMSE values cease to improve with additional training. Adjusting model architectural and training parameters allows the modeler to control the geometric complexity of the surface that the ANN fits to the data. In general, a high-quality predictive model can be obtained and a satisfactory model can be exported for implementation in an end-user application when the following conditions are satisfied (Conrads and others, 2006):

- the data are well distributed throughout the state space of variables describing the physical system of interest;
- the input variables selected by the modeler share mutual information about the output variables; and
- the functional form "prescribed" or "synthesized" by the model to "map" (correlate) input variables to output variables is a good one, that is, the model closely matches observed data but is sufficiently generalized to not overfit the data as guided by the training method and the modeler's knowledge of the real-world system

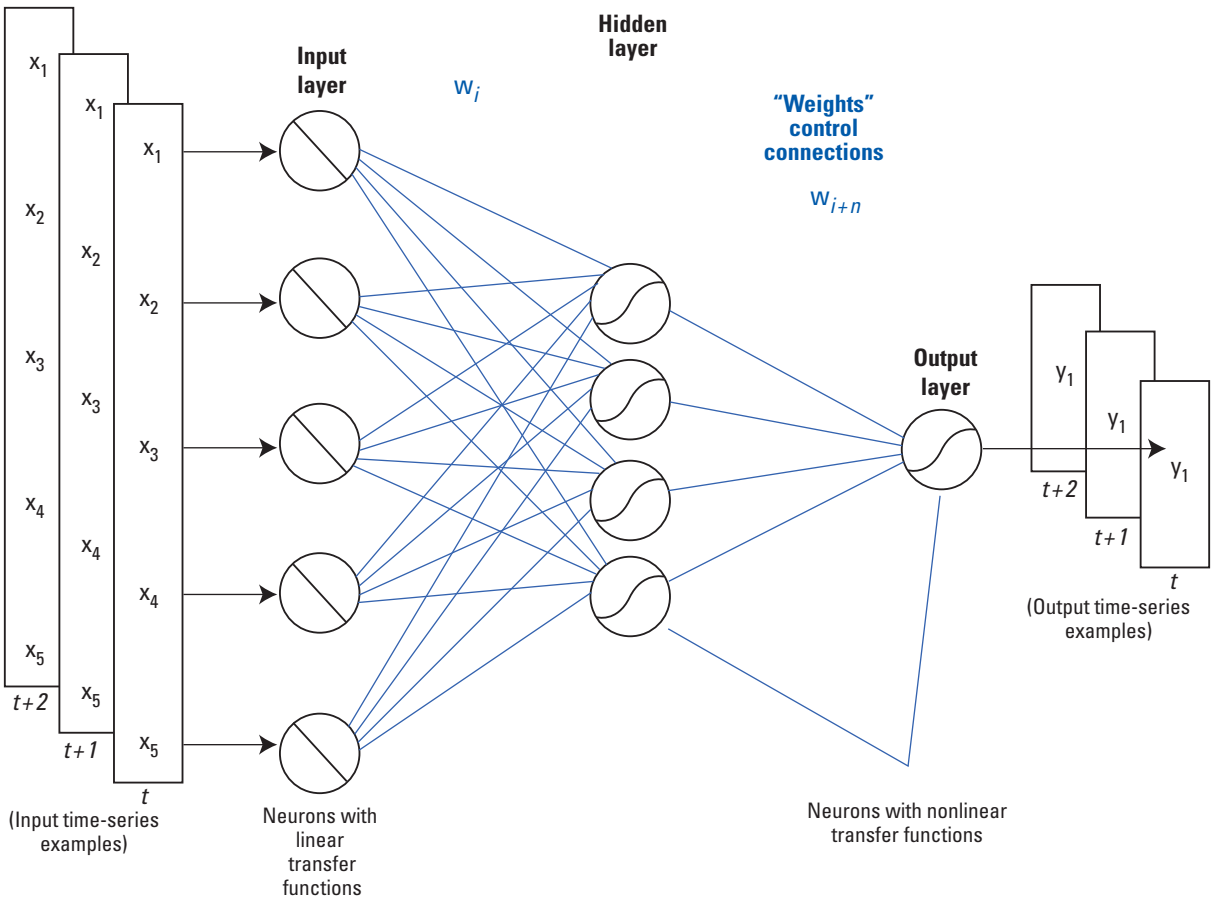


Figure 42. Multilayer perceptron artificial neural network architecture (modified from Conrads and Roehl, 2007).

and relevant physical principles. Machine-learning techniques, like ANN models, synthesize a best fit to the data. Techniques such as ordinary least squares and physics-based models prescribe the functional form of the model's fit to the calibration data.

Subdividing a complex modeling problem into subproblems and then addressing each is an effective means of achieving the optimal results. A collection of submodels whose calculations are coordinated by a computer program constitutes a "super-model" (Daamen and others, 2006). For the study described here, ANN models (submodels) were first developed to predict water level or flow at a particular gaging station. These submodels were then incorporated into the CFANN DSS super-model that integrated the model controls, model database, and model outputs. The ANN models described in this report were developed using the iQuest™ data-mining software[3] (Version 2.03C DM Rev31). The ANN models were deployed in the DSS using the Visual Basic run-time library of the iQuest R/T™ software.

[3] The iQuest™ software is exclusively distributed by Advanced Data Mining, LLC, 3620 Pelham Road, PMB 351, Greenville, SC 29615-5044 Phone: 864–616–9876, email: *ruby.daamen@advdmi.com*, *http://www.advdmi.com*.

Statistical Measures of Prediction Accuracy

Statistical measures of prediction accuracy were computed for the final water-level and flow estimates (output from the two-stage model architecture) and for the individual ANN models. The statistics for the final predicted values capture the capability of the two-stage modeling approach to accurately estimate the water levels or flows at the site. The statistics for the individual models (rainfall and groundwater-use ANNs) document these intermediate models. Because two models are used, the statistics for the individual models may not be an indication of the quality of the final water-level or flow estimates.

The R^2, the mean error (ME), RMSE, and percent model error (PME) were computed for the training and testing datasets for both the rainfall and groundwater-use ANN models for each site. In addition to these statistics, the Nash-Sutcliffe coefficient of efficiency was computed for the final water-level and flow estimates. Model performance often is evaluated in terms of R^2, which commonly is interpreted as a measure of the "goodness of the fit" of a model. A second interpretation is one of answering the question, "How much information does one variable or group of variables provide about the behavior

of another variable?" In the first context, an R^2 of 0.6 might be disappointing, whereas in the latter, it is merely an accounting of how much information is shared by the variables being used as defined by the selected model. As a correlation statistic, R^2 is insensitive to additive and proportional differences between predicted and measured values, and thus should not be used as the sole measure of the goodness of fit of hydrologic models (Legates and McCabe, 1999).

The ME and RMSE statistics provide a measure of model prediction accuracy. The ME is a measure of the bias of model predictions—whether the model over- or underpredicts the measured data. The ME is presented as the adjustment needed to make the simulated values equal to the measured values. Therefore, a negative ME indicates an overprediction by the model and a positive ME indicates an underprediction by the ANN model. Mean errors near zero may be misleading because negative and positive discrepancies in the simulations can cancel each other. As an absolute-error statistic, RMSE addresses the limitations of ME by computing the magnitude, rather than the direction (sign) of the discrepancies. The units of the ME and RMSE statistics are the same as those of the simulated variable of the model.

The accuracy of the models, as indicated by RMSE, should be evaluated with respect to the range of the output variable. A model may have a low RMSE, but if the range of the output variable is small, the model may be accurate only for a small range of conditions and the model error may be a relatively large percentage of the model response. Likewise, a model may have a large RMSE, but if the range of the output variable is large, the model error may be a relatively small percentage of the total model response. The PME was computed by dividing the RMSE by the range of the measured data.

Nash and Sutcliffe (1970) developed an efficiency index to evaluate the goodness of fit of hydrologic models, commonly referred to as the Nash-Sutcliffe coefficient of efficiency (NSCE). The NSCE can range from minus infinity to 1. A value of 1 corresponds to a perfect match of predicted values to the observed data. A value of 0 indicates that the model predictions are as accurate as the mean of the observed data used to develop the model, whereas a value less than 0 indicates that the observed mean is a better predictor than the model. In summary, models with an NSCE from 0 to 1 provide a better estimate than the mean of the observed data, and the higher the value, the better the estimates. McCuen and others (2006) noted that only subjective evaluations of the NSCE were possible and that the index is influenced by sample size, model bias, timing errors, and outliers. Legates and McCabe (1999) recommend using a combination of correlation-based and absolute-error statistics as well as graphical analysis to evaluate model goodness of fit.

Development of Artificial Neural Network Models in Central Florida

The following sections describe how the ANN models were developed for predicting hydrologic response at each of the 51 lake, well, and spring sites. Development of models for all types of sites—lake water level, groundwater level, and spring flow—followed a similar approach, which is described in the section below on model construction and training. Following this discussion is a section on prediction accuracy of the final, trained models, including sections on the models for each type of site. Finally, a general description of the performance of all the models for each type of site is provided followed by a discussion of two example models.

Model Construction and Training

The ANN models were developed iteratively starting with selecting a candidate pool of input variables, training the ANN, and then using prediction accuracy statistics, such as R^2 and input-output sensitivities, to cull the least important input variables. Effectively, a hybrid modeling approach was applied that combines signal decomposition with ANN models, whereby rainfall and groundwater-use data are decomposed to capture variability in time series of these key forcing variables over a variety of frequency ranges and ANN models are used to reconstruct the original time series of response variable (water level or flow) using these decomposed subsignals. This approach is conceptually similar to wavelet-neural network hybrid models, which combine wavelet transforms and ANN models, such as those developed to forecast monthly groundwater levels and daily river flow (Wang and Ding, 2003), the Palmer drought severity index (Kim and Valdés, 2003), daily intermittent streamflow (Kişi, 2009), and monthly river flow (Wei and others, 2013). Additionally, Özger and others (2012) developed a wavelet fuzzy logic model to forecast the Palmer modified drought index that combined wavelet transforms with a fuzzy logic model. In the current study, in place of the wavelet analysis, signal decomposition was conducted using MWDs computed from MWAs of prescribed window sizes. It is left to the ANN to learn which explanatory variables and associated frequency ranges are the best predictors of behaviors that are manifest in the hydrologic signal. As previously described, two-stage model architecture was used that modeled rainfall and groundwater-use effects separately (fig. 40). Summaries of the rainfall and groundwater-use ANN models for each of the 51 sites, including model construction details and prediction accuracy statistics, are listed in appendix 3, and descriptions of the explanatory variables used in the final ANNs are provided in appendix 4.

The quality and completeness of both the forcing and response data are critical for any empirical modeling approach, such as that applied in the current study. In terms of accuracy, completeness, and representativeness of physical process

behavior, the rainfall data are better than the hydrologic response data, which in turn are much better than the groundwater-use data. Some of the hydrologic response data yield hydrographs that exhibit late starts, long periods of missing data, sporadic sampling, and values that appear to be anomalous from inspection. In general, the groundwater-level hydrographs are better than the lake water-level and spring-flow hydrographs. More consistent data frequency led to the ability to use random sampling to partition all groundwater-level and most lake water-level hydrographs into training and testing datasets. For some lakes and one spring (sites 5, 207, 209, 221, 239, 277, 289, 297, and 313; table 2; appendix 3), sparse data necessitated using all data for the training dataset. The spring-flow hydrographs are the most problematic. The springs were measured sporadically for most of the study period, but more frequent measurements were made in the last decade. The disparate data frequency was particularly pronounced for three springs (sites 311, 312, and 314; table 2; appendix 3), and data collected during the last decade were subsampled so as not to be disproportionately represented when combined with the earlier sparse sampling period to partition the hydrographs into training and testing datasets. Spring flow "flat-lining" (consecutive days of identical flows), which were apparent anomalies in some datasets, are possibly artifacts of procedures used to estimate daily data from direct measurements, and data during these periods were removed from the dataset. Some of the lake and well hydrographs have occasional short duration (60 days or shorter) gaps when data are missing. For sites where there was little submonthly cyclic variability in the hydrographs, these data gaps were filled using simple linear interpolation prior to training the ANN models so as not to bias the training dataset toward some periods over others. At these sites, most of the variability is due to seasonal (3-month) and longer term trends. In addition, the final predictions generated by the two-stage modeling approach use a 1-month time step; therefore, interpolating over a period of 60 days or less means estimating across only one time step. Notes are included in the model summaries (appendix 3) for the sites where 60-day or shorter gaps were filled in this way.

The majority of the groundwater-use data were estimates and, by a wide margin, lacked the accuracy, representativeness, and temporal and spatial resolution of the meteorological and hydrologic response data. Data that describe the extent to which groundwater withdrawals are returned to the hydrologic system are also largely unavailable. Given these limitations, a single total groundwater-use signal was created by aggregating all of the categories of groundwater use (figs. 29, 34). Potential inaccuracies introduced by this approach are likely reduced by the generally highly permeable nature of the study area's karst terrain (Kuniansky and others (2012) mapped large values of transmissivity for the Upper Floridan aquifer in much of the study area), which would tend to dissipate spatial variability in responses to groundwater use, and the 1-month simulation time step, which is longer than the short-term fluctuations in groundwater use. Using the two-stage model architecture (fig. 40), separate ANN models were developed for rainfall and groundwater use and are described in detail in the following sections.

Rainfall Models

The rainfall models were developed using a 1-day time step (fig. 40) enabled by daily rainfall (available from 1942 through 2008) and hydrologic response data (available for the periods of record shown in table 2). To varying extents, the rainfall data from the 18 NOAA sites are correlated between sites (R values range from –0.31 to 0.91; 25 percent have R values greater than 0.5). The general approach to developing accurate models of dynamic processes involves filtering each candidate input time series into signals representing a range of frequencies and decomposing the filtered signals into multiple frequency-band components. Then the supervised machine learning of the ANN model is allowed to determine which combination of input signals and components are the best predictors while culling poor predicting inputs.

Given the large number of candidate input signals and components, it was necessary to employ an input representation and model development approach that mitigated problems associated with so many potentially correlated inputs. The approach includes the following steps:

1. A "standard" rainfall signal was created for each hydrograph. The standard signal was the same for all of the wells and springs and was computed as the arithmetic average by time step of the values for the 18 NOAA sites for each day (fig. 33). As a starting point, the standard for each lake was a weighted average of the rainfall within the lake drainage basin based on Thiessen polygons fitted to the NOAA sites. For some lakes, the standard was modified to use fewer or other nearby NOAA sites to improve prediction accuracy. The formula used to compute the rainfall standard for each lake is provided in the model summary table for the respective lake (appendix 3).

2. A "normalized" rainfall signal was created for each of the 18 NOAA rainfall gages by subtracting the standard rainfall signal from the individual gage rainfall.

3. The standard and normalized rainfall signals were processed as follows:

    - Low-pass filters represented by MWAs of 30; 90; 270; 630; 1,440; and 2,250 days were applied to the standard rainfall signal (fig. 33) as well as each normalized rainfall signal. The selection of the 270-, 630-, and 1,440-day MWAs followed an analysis that determined the window sizes at which correlations peaked between different rainfall MWAs and hydrologic response at a variety of sites; this analysis used the super tau tool described by Conrads and Roehl (2010, p. 19–20). The other three MWAs provide for computation

of additional higher and lower frequency-band components.

- The standard rainfall signal's frequency-band components were computed using MWDs of successive MWA pairs. It should be noted that one MWA plus the MWDs completely represent any of the defined low-pass signals without loss of information.
- Subtracting the standard rainfall signals from the individual NOAA rainfall signals decorrelates the resulting MWAs from the standard signal. The resulting normalized signals, however, are not consequentially decorrelated from each other, because, for example, two highly correlated pre-normalized signals will still be correlated after normalization. The effects of such cross-correlations are addressed in item 4 below. Combining the 6 MWAs for the standard rainfall signal and each of the 18 normalized rainfall signals with the 5 MWDs of standard rainfall results in a total of 119 candidate input variables for an ANN model.

4. An ANN model was iteratively developed for each site to simulate hydrologic response. During the ANN development and training process, input variables were culled according to the procedure described below, resulting in a final set of variables for each site (appendixes 3, 4). This process is purposefully conservative with regard to overtraining, resulting in relatively simple ANNs with limited numbers of parameters (weights; for example, see figure 42) relative to the size of the training datasets. Furthermore, Amari and others (1996) note that a large testing dataset for cross-validation stop training is not necessary when the training dataset is greater than 30 times larger than the number of ANN parameters—a criterion that was met by each of the final ANN models—and use of a large testing dataset at the expense of a small training dataset may result in undertraining of the ANN. The goal of the ANN development and training process was to maximize R^2 under the following constraints:

   - Retain input variables representing standard rainfall signals and frequency-band components.
   - Disallow input variables with an R value exceeding 0.4 in order to minimize confounding effects attributable to cross-correlations.
   - Allow only positive input-output sensitivities, such that increasing or decreasing any individual rainfall input variable correspondingly increases or decreases the predicted value, in order to remove input variables that may yield a predicted value at odds with known physics (in particular, principles of mass conservation). Imposing constraints on the basis of plausible physics is important, because relying solely on cross-validation training techniques does not assure a satisfactory ANN that reasonably approximates the behavior of the physical system (Kingston and others, 2005).
   - Avoid overfitting. The hydrologic response data were partitioned, where the data quantity allowed, into training and testing datasets by random selection. The prediction accuracy statistics for the training and testing data should be similar to verify model accuracy and ensure that the training data were not overfitted. Training datasets as small as 10 percent of the available data were used where possible, although 31 percent or less was used for 39 sites and greater than 80 percent was used for the remaining 12 sites. Several of the hydrographs were not partitioned because of sparse data and concern that using a testing dataset would exclude important data for training. The size of the training and testing datasets for each ANN model are listed in appendix 3.
   - Limit the ANN models to one hidden-layer neuron (instead of four neurons shown in the hidden layer of the generic multilayer perceptron ANN depicted in figure 42), thereby allowing the enforcement of only positive input-output sensitivities and minimizing overfitting. This constraint prevents fitting multivariate data with non-monotonic surfaces (surfaces with saddle points), which would cause the signs of sensitivities to change as inputs are ranged. The constraint is appropriate because hydrograph values should always increase with rainfall increases (and decrease with groundwater-use increases), and the constraint mitigates the problem of unavailable testing data for some sites.

Groundwater-Use Models

The groundwater-use models were developed using a 1-month time step (fig. 40), which was the minimum resolution of the groundwater-use data (available from 1957 through 2008). The rainfall data and the rainfall model predictions and residuals had be converted to the same time step, which was done by applying 1-month MWAs and subsampling the last day of each month. The amount of the available data for training and testing the ANN models was reduced substantially as a result of the monthly time step; therefore, testing datasets were small or eliminated completely for the groundwater-use models to maximize the amount of data for model training (appendix 3).

The groundwater-use models simulate the residual error signal (residuals) for each of the rainfall models and were developed similarly to the rainfall models. The window sizes

of the MWAs applied to total groundwater use were 3, 6, 12, 24, 48, and 96 months (fig. 34), and MWDs were computed for successive MWA pairs by subtracting each MWA from the MWA with the next smaller window size. A total of 13 potential inputs for an ANN model were available, considering the original monthly data, 6 MWAs, and 6 MWDs. A single MWA was selected on the basis of the highest negative correlation to the residuals, thus enforcing a known physical constraint that an increase in groundwater use causes a decrease in water level or flow (that is, negative input-output sensitivity). An ANN model with only one hidden-layer neuron was used to allow enforcement of only negative input-output sensitivity. Other potential inputs to the model included the MWDs and the monthly average of the predicted signal (water level or flow) from the respective rainfall model. ANN training was an iterative process wherein only those inputs with a negative input-output sensitivity were included. The rainfall-model predicted water level or flow was used as an input for a groundwater-use model only when it resulted in an increase in R^2 of at least 0.01 for the final model (the combined rainfall and groundwater-use models) or when it was needed to maintain a negative input-output sensitivity. In addition, collinear inputs with an R value exceeding 0.5 were excluded. After training, the largest groundwater-use frequency-band components retained were MWDs using 96-month MWAs for some sites. Hence, even though groundwater-use estimates were available from 1957 onward and even earlier rainfall data were available (from 1942 onward), the simulation period for the final two-stage model architecture began in 1965 in order to accommodate the shorter period of record for groundwater use and an 8-year background period required by the 96-month MWA.

Model Prediction Accuracy and Example Model Simulations

Statistical measures of prediction accuracy were computed for the final water-level and flow estimates based on the monthly simulated values from January 1965 through December 2008. The statistics for the final predicted values demonstrate the generally good to excellent capability of the two-stage architecture that uses the hybrid signal-decomposition ANN models to accurately estimate the water levels or flows at each site. The models explained much of the variability in observed data, with models for 43 of the 51 sites having R^2 values exceeding 0.6, as well as matching well the magnitude of the observed data, with models for 32 of the 51 sites having PME values less than 10 percent (fig. 43). Generally, groundwater-level models (R^2 averaging 0.774 and PME averaging 8.90 percent) performed slightly better than lake water-level models (R^2 averaging 0.718 and PME averaging 9.96 percent). Spring-flow models had the poorest model fit (R^2 averaging 0.546 and PME averaging 13.14 percent). Summary statistics of both measured and simulated data and model fit are provided in appendix 5 for each site. The lake water-level, groundwater-level, and spring-flow models are described in more detail in the following sections. For each type of site, a general description of the performance of all the models is provided, followed by a discussion of two example models—one with good model-fit statistics and one with a relatively poor match to the observed data—in order to provide a sense of the range in model performance. By discussing model construction (input variables) and model performance in detail for six sites, example interpretation techniques are provided that can be applied by the reader to the models for any of the other 45 sites of interest using data in appendixes 3, 4, and 5 and the CFANN DSS.

Lake Water-Level Models

The lake water-level models explained from 43 to 89 percent of the variability in the observed data and matched observed water levels quite well (RMSE ranged from 0.39 to 2.25 ft). Values of PME ranged from 7.44 to 13.3 percent, and values of ME generally were near zero (median of 0.009 ft for 22 lakes) and ranged from –0.26 to 0.14 ft. Because the absolute errors were generally small, values of the NSCE were nearly identical to those of R^2, ranging from 0.431 to 0.894.

Crooked Lake near Babson Park

Of all the lake water-level models, the model for Crooked Lake near Babson Park (southeast Polk County; site 49, table 2; fig. 32) had the most complete data record (no missing data from 1965 to 2008) and the second highest set (ranked in terms of accuracy) of model-fit statistics (R^2 of 0.887, ME of 0.02 ft, RMSE of 1.28 ft, PME of 7.44 percent, and NSCE of 0.881; appendix 5). The model simulates well multidecadal trends (for example, the generally downward trend from 1965 to 1991 and the generally upward trend from 1991 to 2008; fig. 44*A*) and shorter term multiyear trends (for example, the downward trend from 1979 to 1981; fig. 44*A*). The simulation of intraannual variability tends to be less accurate; seasonal dry- to wet-period changes were sometimes overpredicted (for example, the September 1976 to May 1977 simulated water-level drop was 4.08 ft and the measured drop was 2.24 ft) and sometimes underpredicted (for example, the April to October 1992 simulated water-level rise was 2.69 ft and the measured rise was 3.76 ft). Lake water level was measured at approximately weekly intervals, resulting in a large amount of data for training the rainfall model. Therefore, 30 percent of the data were randomly selected and used for the training dataset, and the remaining 70 percent were used for the testing dataset. Model-fit statistics for the daily rainfall model for the training and testing datasets were virtually the same, with R^2 of 0.863 and 0.864 and PME of 8.23 and 8.13 percent, respectively (appendix 3).

The final model for Crooked Lake near Babson Park is relatively simple, with four inputs for the rainfall model and two for the groundwater-use model. The rainfall model uses the following inputs: 1,440-day MWA and a 630-day minus 1,440-day MWD of standard rainfall, a 2,250-day MWA of normalized rainfall at NOAA gage 4502 in south Osceola

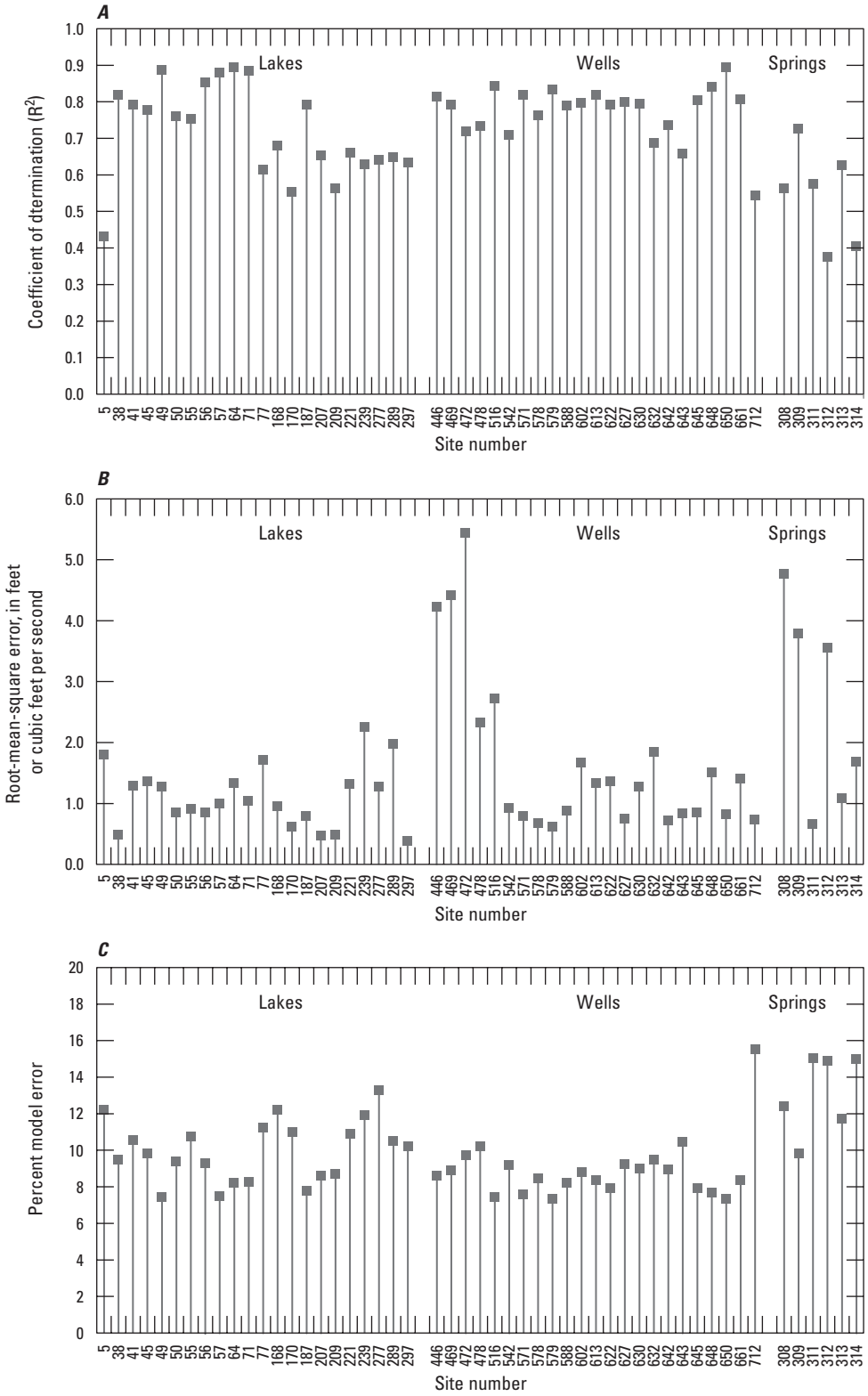


Figure 43. Model-fit statistics for the final water-level or flow models for each site for the 1965 to 2008 simulation period: *A*, coefficient of determination (R^2), *B*, root-mean-square error (RMSE), and *C*, and percent model error (RMSE divided by the range of data). Site information and locations are shown in table 2 and figure 32.

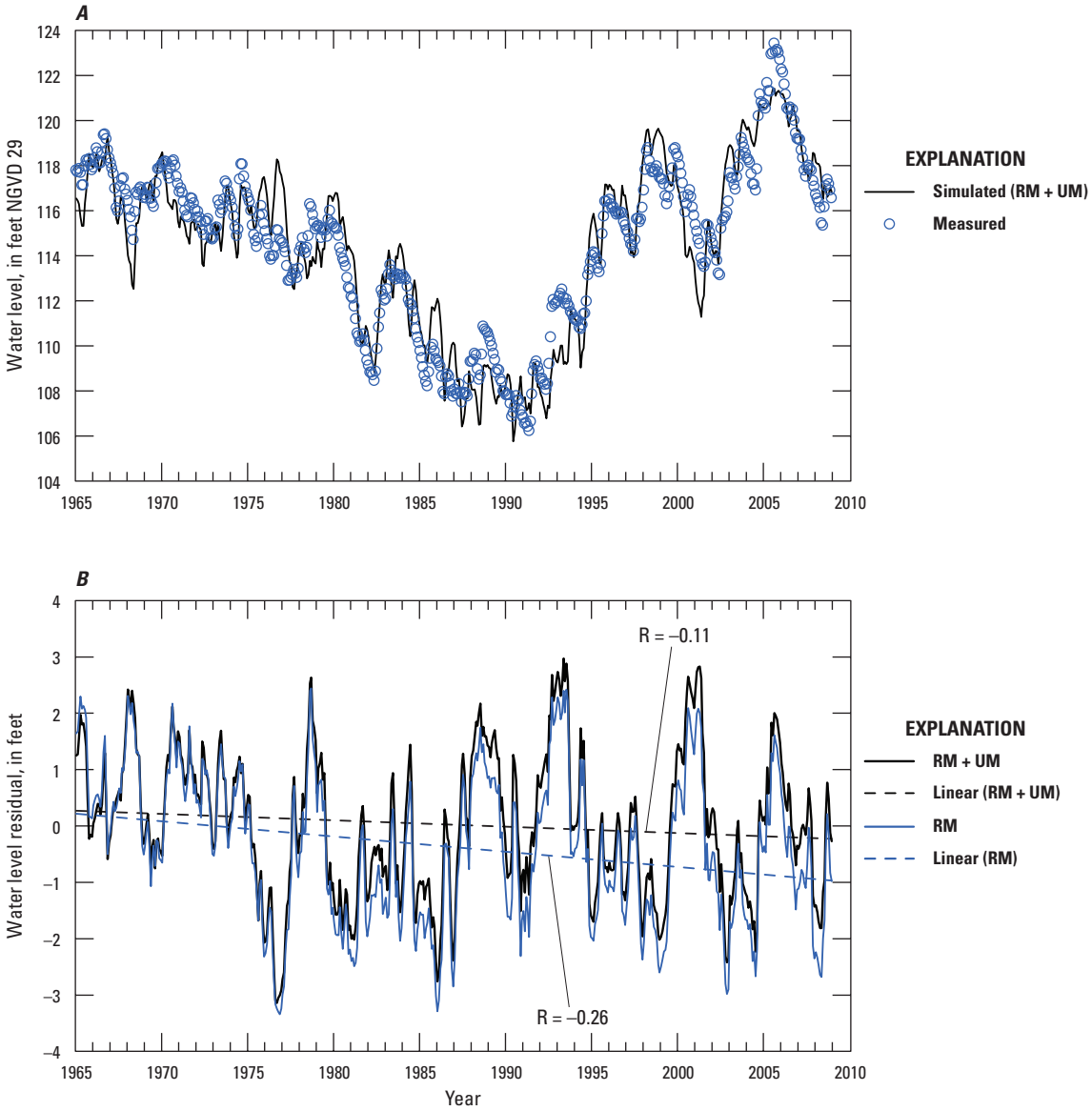


Figure 44. Results of lake water-level simulation for Crooked Lake near Babson Park (site number 49; table 2; fig. 32) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines. R is the Pearson product-moment correlation coefficient for the linear regression of residuals and time.

County (fig. 24), and a 630-day MWA of normalized rainfall at NOAA gage 9401 in north Hardee County (fig. 24) (appendix 3). The standard rainfall for Crooked Lake was computed as the weighted average of values from the NOAA rain gages 5973 (96.2 percent) and 369 (3.8 percent) within the lake's drainage basin (fig. 24). The sensitivity of the rainfall model to lower frequency rainfall signals is consistent with the model accuracy over longer time scales noted above and indicative of the nature of the lake water-level response where lower-frequency variability is more evident in the observed water-level data than subannual frequencies at this site. A bias exists in the rainfall model residuals, leading to a downward trend (linear correlation) in residual error (R = –0.26; fig. 44*B*), which is qualitatively consistent with possible decreases in lake water level attributable to increases in groundwater use.

Because of the bias in rainfall model residuals, the groundwater-use model was included in the two-stage model architecture (fig. 40) in order to better explain the observed hydrologic response. Inputs for the groundwater-use model for this site are a 24-month MWA and a 3-month minus 6-month MWD of total groundwater use (appendix 3). Addition of the groundwater-use model explained about 18 percent of the variability in the rainfall model residuals (R^2 = 0.177; appendix 3)

and eliminated much of this bias, such that the long-term trend in residuals for the final model (rainfall plus groundwater use) is only slightly downward (R = –0.11; fig. 44*B*). The groundwater-use model tended to mitigate negative residuals from the rainfall model, indicating overpredicted water levels from the rainfall model were reduced by the groundwater-use model (fig. 44*B*). The overall average absolute residual error decreased 0.11 ft from the rainfall model to the final model, indicating that inclusion of the groundwater-use model improved final model accuracy.

Prevatt Lake

Of all the lake water-level models, the model for Prevatt Lake (northwest Orange County; site 5, table 2; fig. 32) had one of the most incomplete data records (42.8 percent missing data from 1965 to 2008) and the lowest set (ranked in terms of accuracy) of model-fit statistics (R^2 of 0.432, ME of 0.03 ft, RMSE of 1.80 ft, PME of 12.20 percent, and NSCE of 0.431; appendix 5). Little long-term trend is apparent in the measured data, and the model focuses more on attempting to replicate shorter-term variability with limited success (fig. 45*A*). Intraannual variability attributable to seasonal dry- and wet-period cycles is reflected in the simulated water levels, although the simulated water levels often did not coincide in magnitude with measured data (fig. 45*A*). Multiyear trends are simulated fairly well, such as during drought and recovery periods in the early 1980s and early 2000s (fig. 45*A*). Lake water level was measured sporadically prior to 1982. After 1982, lake water level generally was measured at monthly intervals with a short period (April 2004 to December 2008) of daily data. Nevertheless, the more complete record after 1982 was interspersed with numerous periods of missing data ranging from 1 to 14 months in duration. Given these sporadic data, 100 percent of the data were used for the training dataset.

The final model for Prevatt Lake is relatively simple, with five inputs for the rainfall model and two for the groundwater-use model. The rainfall model uses the following inputs: 630-day MWA, a 90-day minus 270-day MWD, and 270-day minus 630-day MWD of standard rainfall; a 630-day MWA of normalized rainfall at NOAA gage 3137 in northeast Okeechobee County (fig. 24); and a 630-day MWA of normalized rainfall at NOAA gage 6414 in south Marion County (fig. 24) (appendix 3). The standard rainfall for this lake was set equal to that of NOAA rain gage 7982 in north Seminole County (fig. 24). The sensitivity of the rainfall model to higher frequency rainfall signals and frequency-band components is indicative of the nature of the lake water-level response at this site where high-frequency variability is more evident in the observed water-level data than low frequency multidecadal variability. A bias exists in the rainfall model residuals, leading to a slight upward trend (linear correlation) in residual error (R = 0.10; fig. 45*B*), which is qualitatively consistent with possible increases in lake water level attributable to decreases in groundwater use.

The groundwater-use model was used to better explain the observed hydrologic response by addressing the bias in rainfall model residuals. Inputs for the groundwater-use model for this site are a 12-month MWA and a 48-month minus 96-month MWD of total groundwater use (appendix 3). Addition of the groundwater-use model explained about 5 percent of the variability in the rainfall model residuals (R^2 = 0.0463; appendix 3) and eliminated some of this bias such that the long-term trend in residuals for the final model (rainfall plus groundwater use) is slightly less upward (R = 0.088; fig. 45*B*). The long-term trend in total groundwater use is slight since the early 1980s (fig. 34*A*), which coincides with the period of greatest data availability for model training at this site. The low frequency 48- and 96-month groundwater-use MWAs show smooth upward and downward trends but little long-term trend over this period (fig. 34*A*); these smooth variations in groundwater use may correlate, by way of the 48-month minus 96-month MWD, with measured lake water-level variations. The groundwater-use model tended to mitigate negative residuals from the rainfall model, indicating that overpredicted water levels from the rainfall model were reduced by the groundwater-use model (fig. 45*B*). The overall average absolute residual error decreased 0.03 ft from the rainfall model to the final model, indicating that inclusion of the groundwater-use model slightly improved final model accuracy. The large amount of missing data and irregular measurement frequency are important factors that contributed to the relatively poor fit of the water-level model for Prevatt Lake by limiting the representation of complete cyclical patterns in the model training dataset.

Groundwater-Level Models

The groundwater-level models explained from 54 to 89 percent of the variability in the observed data and matched observed water levels well (RMSE ranged from 0.62 to 5.44 ft). Values of PME ranged from 7.33 to 15.5 percent, and values of ME generally were near zero (median of –0.02 ft for 23 wells) and ranged from –1.17 to 0.08 ft. Because the absolute errors were generally small, values of the NSCE were nearly identical to those of R^2, ranging from 0.540 to 0.894.

OR-47 Floridan Well

Of all the groundwater-level models, the model for the OR-47 Floridan well (west-central Orange County; site 648, table 2; fig. 32) had a nearly complete data record (6.3 percent missing data from 1965 to 2008) and the third highest (ranked in terms of accuracy) set of model-fit statistics (R^2 of 0.840, ME of 0.02 ft, RMSE of 1.50 ft, PME of 7.69 percent, and NSCE of 0.836; appendix 5). The model simulates well multidecadal trends (for example, the generally downward trend from 1965 to 2008; fig. 46*A*) and shorter term multiyear trends (for example, the drought and recovery periods in the early 1980s and early 2000s; fig. 46*A*). Intraannual variability also is simulated well, but the model sometimes underpredicts (for example, August 1994 to March 1995; fig. 46*A*) or overpredicts (for example, March 1981 to January 1982; fig. 46*A*) the

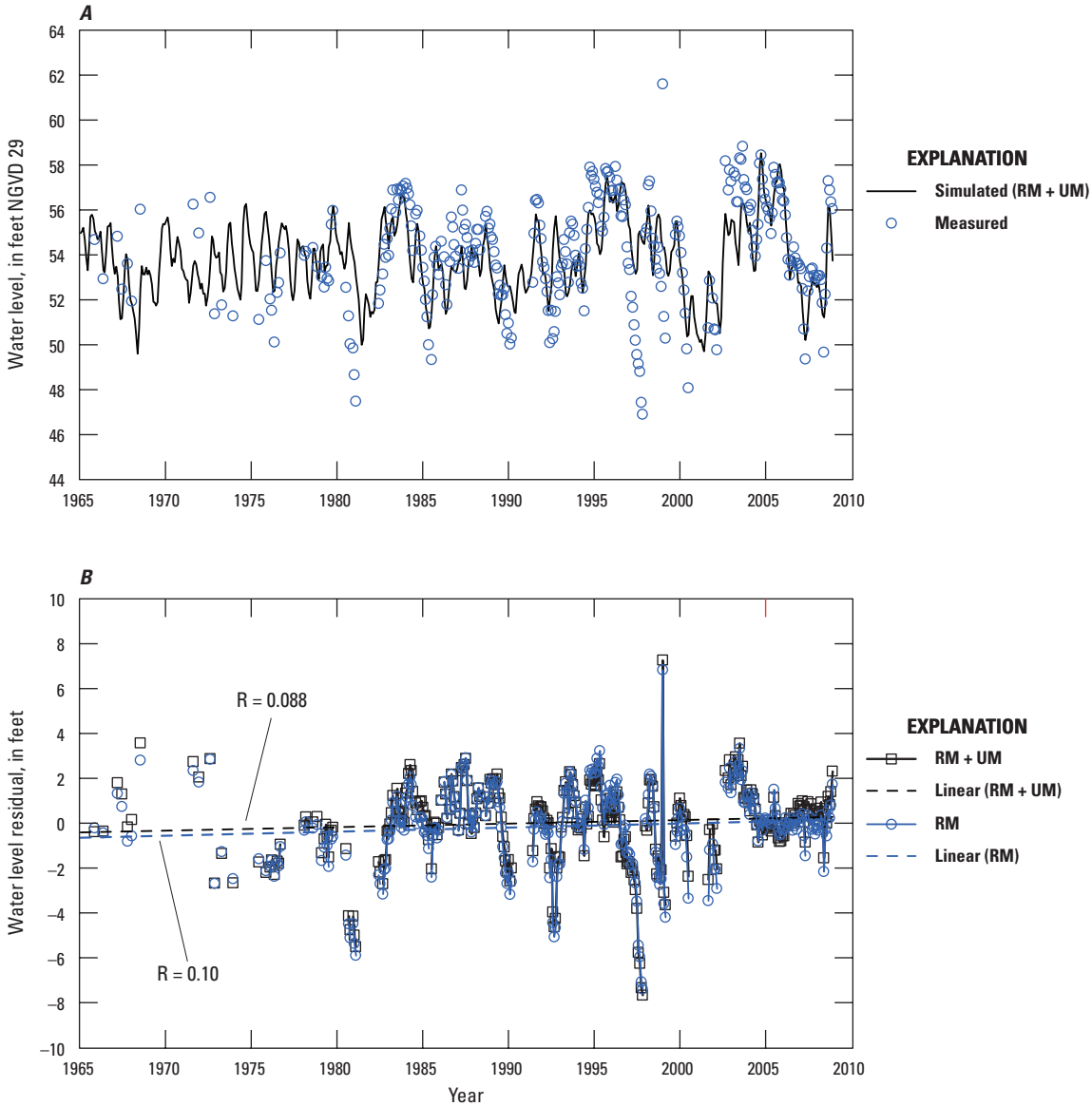


Figure 45. Results of lake water-level simulation for Prevatt Lake (site number 5; table 2; fig. 32) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines. For the residuals, lines connecting points indicate periods of consecutive monthly data. R is the Pearson product-moment correlation coefficient for the linear regression of residuals and time.

seasonal dry- and wet-period water-level extremes. Groundwater level was measured at daily intervals, and, given this large amount of data for the rainfall model, 30 percent of the data were randomly selected and used for the training dataset and the remaining 70 percent were used for the testing dataset. Model-fit statistics for the daily rainfall model were virtually the same, with R^2 of 0.905 and 0.906 and PME of 4.80 and 4.81 percent for the training and testing datasets, respectively (appendix 3).

The model for the OR-47 Floridan well is relatively complex, with 18 inputs for the rainfall model and 3 for the groundwater-use model. The rainfall model uses the following 5 inputs based on the standard rainfall: 630-day MWA, 30-day minus 90-day MWD, 90-day minus 270-day MWD, 270-day minus 630-day MWD, and 1,440-day minus 2,250-day MWD. In addition, the rainfall model uses 13 inputs with MWAs (ranging from 270 to 2,250 days in length) of normalized rainfall at NOAA gages 369, 478, 1163, 4502, 4625, 5076, 5612, 5973, 6414, 6628, 7205, 7982, and 9401 (see appendix 3 for further details). The standard rainfall for this well (and all other wells) was computed as the arithmetic average of the values for the 18 NOAA rain gages (fig. 33). The sensitivity of the rainfall model to a wide variety of rainfall frequency-band components is indicative of the nature of the

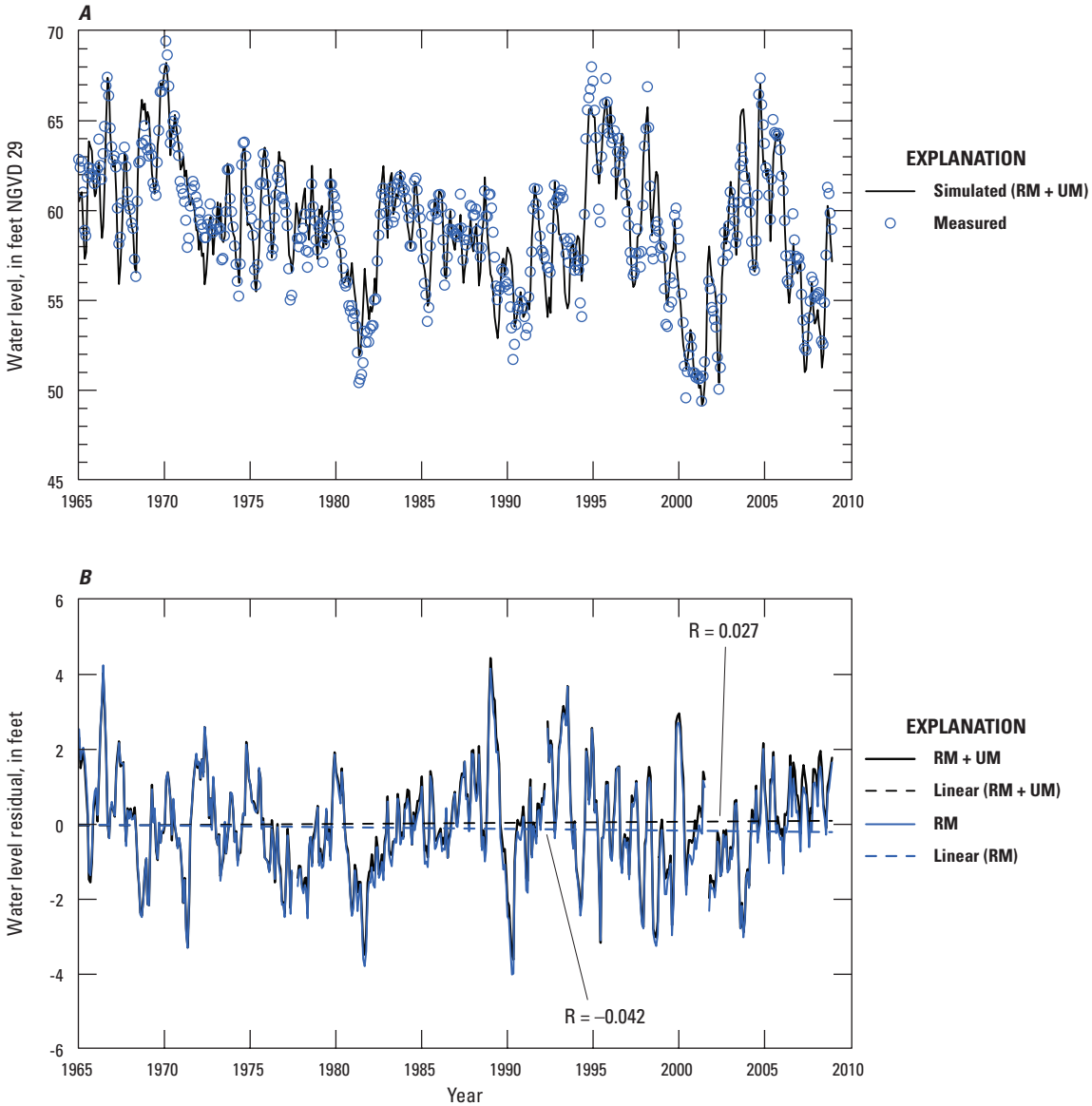


Figure 46. Results of groundwater-level simulation for the OR-47 Floridan well (site number 648; table 2; fig. 32) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines. R is the Pearson product-moment correlation coefficient for the linear regression of residuals and time.

groundwater-level response in which variability over a range of time scales, from multidecadal to subannual, is evident at this site. A slight bias exists in the rainfall model residuals, leading to a downward trend (linear correlation) in residual error (R = –0.042; fig. 46*B*), which is qualitatively consistent with possible decreases in groundwater level attributable to increases in groundwater use.

The groundwater-use model was used to better explain the observed hydrologic response by addressing the bias in rainfall model residuals. Inputs for the groundwater-use model for this site are a 12-month MWA, 1-month minus 3-month MWD, and 24-month minus 48-month MWD of total groundwater use (appendix 3). Addition of the groundwater-use model eliminated this negative bias such that the long-term trend in residuals for the final model (rainfall plus groundwater use) is slightly upward (R = 0.027; fig. 46*B*). Such a small change may not be significant, however, and suggests that long-term (multidecadal) trends in groundwater level at this site attributable to groundwater use are small or cannot be definitively inferred from the available data. The sensitivity of the groundwater-use model to a higher frequency signal and frequency-band component of total groundwater use (12-month MWA and 1-month minus 3-month MWD) is indicative of the lesser importance of low-frequency variations

in groundwater use for explaining measured groundwater-level fluctuations. The groundwater-use model explained about 3 percent of the variability in the rainfall model residuals (R^2 = 0.0279; appendix 3) and tended to mitigate negative residuals from the rainfall model, indicating that overpredicted water levels from the rainfall model were reduced by the groundwater-use model (fig. 44*B*). The overall average absolute residual error decreased 0.02 ft from the rainfall model to the final model, indicating that inclusion of the groundwater-use model slightly improved final model accuracy.

Taft Surficial Well

Of all the groundwater-level models, the model for the Taft surficial well (south-central Orange County; site 712, table 2; fig. 32) had about an average amount of missing data (21.6 percent missing data from 1965 to 2008) and the lowest (ranked in terms of accuracy) set of model-fit statistics (R^2 of 0.543, ME of –0.03 ft, RMSE of 0.73 ft, PME of 15.53 percent, and NSCE of 0.540; appendix 5). Variability in groundwater level at this site predominantly occurred at higher frequencies, from multiyear to subannual time scales, and the model focused on attempting to replicate this short-term variability but with limited success (fig. 47*A*). Intraannual variability attributable to seasonal dry- and wet-period cycles is reflected in the simulated water levels, but often does not coincide in magnitude with measured data (fig. 47*A*). The water-level record for this well has extended periods of missing data at the beginning (January 1965 to May 1969) and end of the simulation period (August 2004 to December 2008), but groundwater level was measured at daily intervals during the intervening 35-year period. Given this large amount of data for the rainfall model, 30 percent of the data were randomly selected and used for the training dataset and the remaining 70 percent were used for the testing dataset. Model-fit statistics for the daily rainfall model were nearly equal, with R^2 of 0.573 and 0.557 and PME of 13.44 and 13.64 percent for the training and testing datasets, respectively (appendix 3).

The model for the Taft surficial well is relatively complex, having 13 inputs for the rainfall model and 3 for the groundwater-use model. The rainfall model uses 2 inputs based on the standard rainfall consisting of a 90-day MWA and a 30-day minus 90-day MWD. In addition, the rainfall model uses 11 inputs with MWAs (30 or 90 days in length) of normalized rainfall at NOAA gages 1163, 1641, 4502, 4625, 5076, 5973, 6414, 6628, 7205, 7982, and 8942 (see appendix 3 for further details). The standard rainfall was computed as the arithmetic average of the values for the 18 NOAA rain gages (fig. 33). The sensitivity of the rainfall model to only high frequency rainfall signals and a high frequency frequency-band component is indicative of the nature of the groundwater-level response in which variability over short time scales, 30 and 90 days, dominates at this site. A bias exists in the rainfall model residuals, leading to a downward trend (linear correlation) in residual error (R = –0.28; fig. 47*B*), which is qualitatively consistent with possible decreases in groundwater level attributable to increases in groundwater use.

The groundwater-use model was used to better explain the observed hydrologic response by addressing the bias in rainfall model residuals. Inputs for the groundwater-use model for this site are a 6-month MWA, 3-month minus 6-month MWD, and 24-month minus 48-month MWD of total groundwater use (appendix 3). Addition of the groundwater-use model eliminated some of the negative bias such that the long-term downward trend in residuals for the final model (rainfall plus groundwater use) (R = –0.17; fig. 47*B*) was less than that for the rainfall model residuals. The sensitivity of the groundwater-use model to higher frequency (6-month MWA) signals as well as both higher and lower frequency frequency-band components (3-month minus 6-month MWD and 24-month minus 48-month MWD) of total groundwater use is indicative of the importance of short- and long-term variations in groundwater use for explaining measured groundwater-level fluctuations at this site. The groundwater-use model explained about 15 percent of the variability in the rainfall model residuals (R^2 = 0.153; appendix 3) and tended to mitigate negative residuals from the rainfall model, indicating that overpredicted water levels from the rainfall model were reduced by the groundwater-use model (fig. 47*B*). The overall average absolute residual error decreased 0.04 ft from the rainfall model to the final model, indicating that inclusion of the groundwater-use model slightly improved final model accuracy. Data collected at the Taft surficial well represent the altitude of the water table that is near land surface (well depth of 7 ft; site 712, table 2); therefore, other physical processes not represented in the two-stage model architecture but that may affect the dynamics of a shallow water table, such as specific yield variability caused by gas bubble entrapment (Fayer and Hillel, 1986; Faybishenko, 1995) and ponding of surface runoff, likely contribute to the relatively poor fit of this model.

Spring-Flow Models

The spring-flow models explain from 38 to 72 percent of the variability in the observed data and matched observed flows fairly well (RMSE ranged from 0.66 to 4.76 ft^3/s). Values of PME ranged from 9.85 to 15.0 percent, and values of ME generally were near zero (median of 0.01 ft^3/s for six springs) and ranged from –0.17 to 0.07 ft^3/s. Because the absolute errors were generally small, values of the NSCE were close to those of R^2, ranging from 0.345 to 0.721. The flow data for all six springs were sparse (missing data range from 43.8 to 67.8 percent for 1965 to 2008; appendix 5). The springs were sporadically measured, and quality of the flow data was poor compared to the water-level data for the lakes and wells. As a result, spring-flow model training focused on lower frequency variability in an attempt to match multidecadal and multiyear trends, resulting in a generally poorer match to seasonal variability.

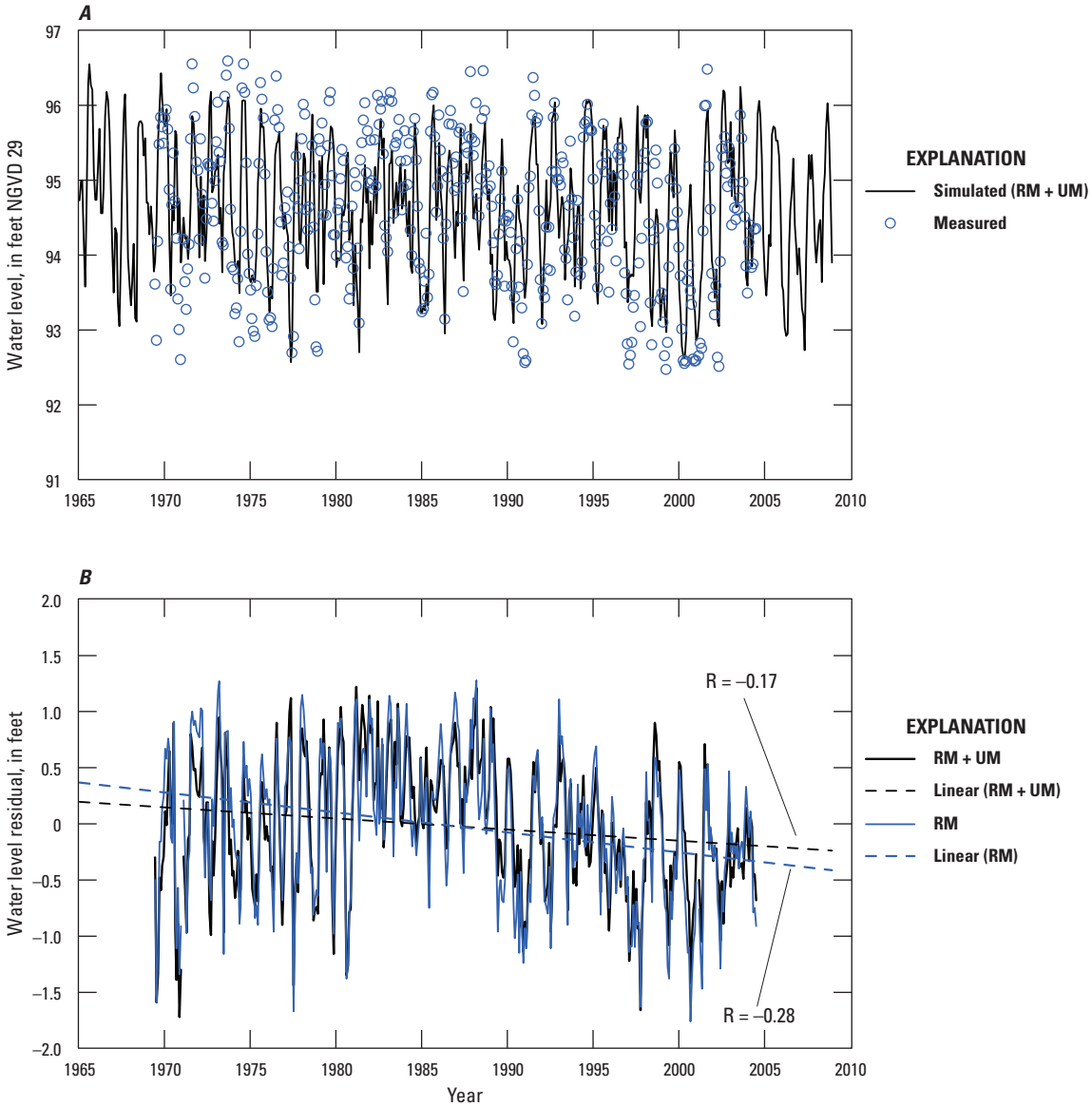


Figure 47. Results of groundwater-level simulation for the Taft Surficial well (site number 712; table 2; fig. 32) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines. R is the Pearson product-moment correlation coefficient for the linear regression of residuals and time.

Rock Springs

Of all the spring-flow models, the model for Rock Springs (site 309, table 2; fig. 32) had the most complete data record (43.8 percent missing data from 1965 to 2008) and the highest (ranked in terms of accuracy) set of model-fit statistics (R^2 of 0.725, ME of –0.08 ft^3/s, RMSE of 3.79 ft^3/s, PME of 9.85 percent, and NSCE of 0.724; appendix 5). The model simulates well multidecadal trends (for example, the generally downward trend from 1965 to 1990; fig. 48*A*) and shorter term multiyear trends (for example, the drought and recovery periods in the early 1980s and the late 1990s to early 2000s; fig. 48*A*). Intraannual variability attributable to seasonal dry- and wet-period cycles is reflected in the simulated flows, although simulated flows often did not coincide in magnitude with measured data (fig. 48*A*). Spring-flow data were available at irregular intervals ranging from daily to once per year or even less frequent. The daily spring-flow data, however, are not field measurements and were computed from model-derived (rating curve) estimates made by the collecting agency. Given the relatively large amount of data available for the rainfall model (considering the spring-flow models were developed to replicate multiyear and longer trends), 29 percent of the data were randomly selected and used for the training

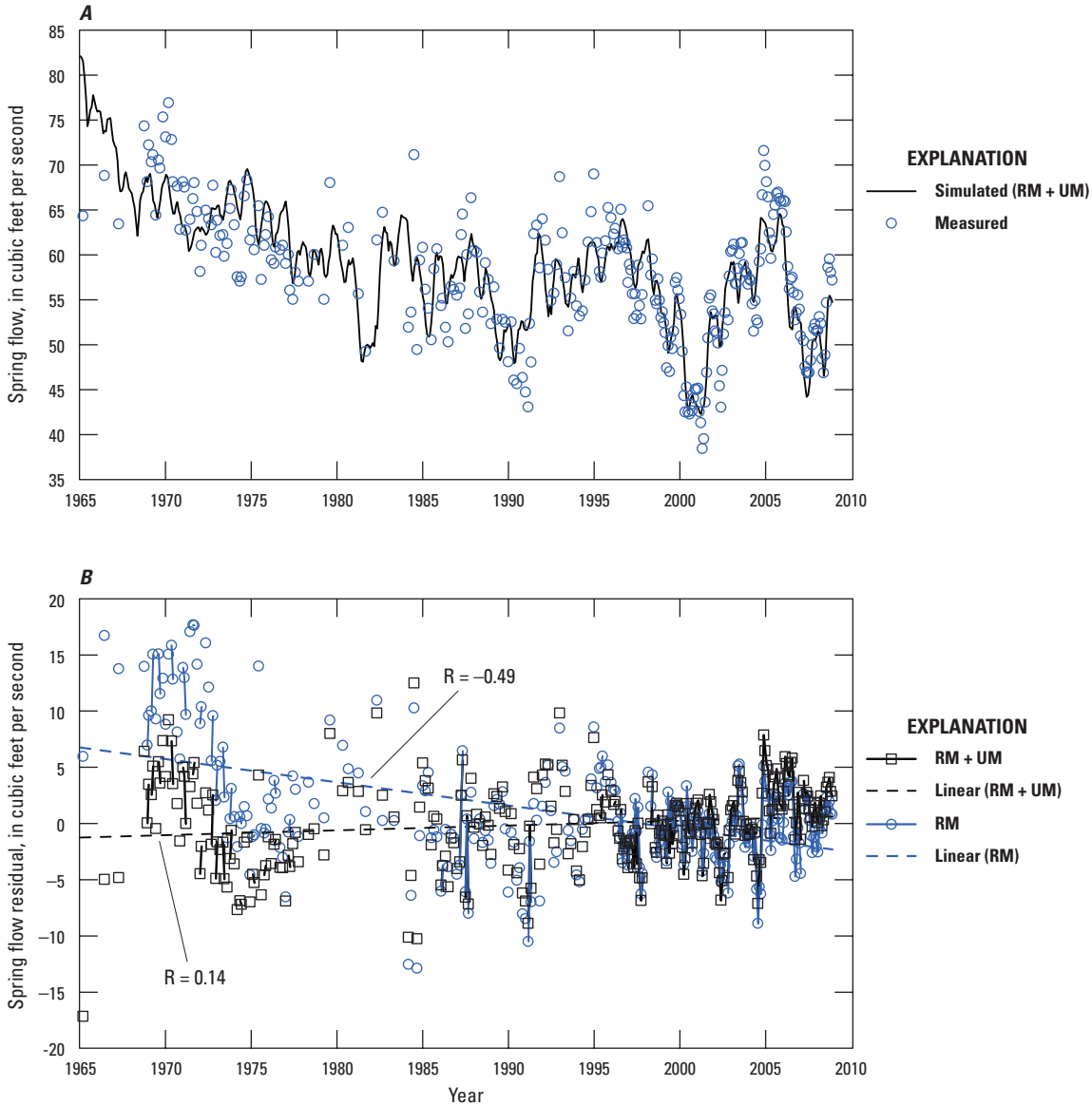


Figure 48. Results of spring-flow simulation for Rock Springs (site number 309; table 2; fig. 32) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines. For the residuals, lines connecting points indicate periods of consecutive monthly data. R is the Pearson product-moment correlation coefficient for the linear regression of residuals and time.

dataset and the remaining 71 percent were used for the testing dataset. Model-fit statistics for the daily rainfall model were close in value, with R^2 of 0.796 and 0.777 and PME of 8.04 and 8.44 percent for the training and testing datasets, respectively (appendix 3).

The final model for Rock Springs is relatively complex, having 12 inputs for the rainfall model and 4 for the groundwater-use model. The rainfall model uses 2 inputs based on the standard rainfall consisting of a 630-day MWA and a 270-day minus 630-day MWD. In addition, the rainfall model uses 10 inputs with MWAs (270, 630, or 1440 days in length) of normalized rainfall at NOAA gages 1163, 2229, 3137, 5973, 6414, 6628, 7205, 7982, 8942, and 9401 (see appendix 3 for further details). The standard rainfall for this spring (and all other springs) was computed as the arithmetic average of the values for the 18 NOAA rain gages (fig. 33). A bias exists in the rainfall model residuals, leading to a downward trend (linear correlation) in residual error (R = –0.49; fig. 48*B*), which is qualitatively consistent with possible decreases in groundwater level attributable to increases in groundwater use.

The groundwater-use model was used to better explain the observed hydrologic response by addressing the bias in rainfall model residuals. Inputs for the groundwater-use model for this site are a 48-month MWA, 12-month minus 24-month

MWD, and 24-month minus 48-month MWD of total groundwater use; the predicted spring flow from the rainfall model was also used (appendix 3). Addition of the groundwater-use model eliminated this negative bias such that the long-term trend in residuals for the final model (rainfall plus groundwater use) is slightly upward (R = 0.14; fig. 48*B*). This relatively large improvement suggests that long-term (multidecadal) trends in spring flow at this site attributable to groundwater use are important. The sensitivity of the groundwater-use model to a lower frequency signal (48-month MWA) and lower frequency frequency-band components (12-month minus 24-month MWD and 24-month minus 48-month MWD) of total groundwater use is indicative of the importance of low-frequency variations in groundwater use for explaining measured spring-flow fluctuations. The inclusion of predicted spring flow from the rainfall model as an input parameter resulted in an increase in R^2 from 0.329 to 0.477 for the groundwater-use model, although an increase of only 0.03 was realized in R^2 for the final model (combined rainfall and groundwater-use models). The groundwater-use model was most sensitive to the 48-month MWA of total groundwater use and less sensitive to predicted spring flow. These two input parameters were inversely correlated such that periods of lower predicted spring flow tended to occur during periods of greater groundwater use. Therefore, predicted spring flow may be acting as a surrogate for some error in the current input parameters, such as poorly estimated groundwater use, or some unaccounted for factor in the current groundwater-use model, such as spatial variations in groundwater withdrawals masked by using total groundwater use. Nevertheless, the groundwater-use model explained about 48 percent of the variability in the rainfall model residuals (R^2 = 0.477; appendix 3) and tended to mitigate negative residuals from the rainfall model, indicating that overpredicted water levels from the rainfall model were reduced by the groundwater-use model (fig. 48*B*). The overall average absolute residual error decreased 0.72 ft^3/s from the rainfall model to the final model, indicating that inclusion of the groundwater-use model improved final model accuracy.

Sanlando Springs

Of all the spring-flow models, the model for Sanlando Springs (site 312, table 2; fig. 32) had an above average amount of missing data (64.0 percent missing data from 1965 to 2008) and the lowest (ranked in terms of accuracy) set of model-fit statistics (R^2 of 0.376, ME of 0.05 ft^3/s, RMSE of 3.56 ft^3/s, PME of 14.88 percent, and NSCE of 0.375; appendix 5). Little long-term trend is apparent in the measured data, and the model attempts to replicate shorter-term variability, although with limited success (fig. 49*A*). Multiyear trends are roughly simulated, such as during drought and recovery periods in the early 1980s and late 1990s to early 2000s (fig. 49*A*). Intraannual variability attributable to seasonal dry- and wet-period cycles is reflected in the simulated flows, although simulated flows often did not coincide in magnitude with measured data (fig. 49*A*). Spring-flow data were available at irregular intervals ranging from daily to once per year; the first measurement made after the beginning of the model simulation period (1965) was in April 1972. The daily spring-flow data, however, are not field measurements and were computed from model-derived (rating curve) estimates made by the collecting agency. The data available for the rainfall model were sparse until a 6-year period at the end of the record (2003–08), when daily data exhibited high frequency variability having a range less than that of the previous 24 years of sparse data. The training dataset was hand assembled and incorporated all of the earlier sparse data and a subsample from the later period of daily data, thereby removing bias in the training dataset (caused by oversampling the daily-data period) to better represent the long-term record. The testing dataset incorporates data only from the daily-data period and poorly represents the full historical record. The overall split between training and testing datasets is 90 and 10 percent, respectively. Model-fit statistics for the daily rainfall model were substantially different, with R^2 of 0.521 and 0.190 and PME of 10.26 and 13.68 percent for the training and testing datasets, respectively (appendix 3), which is a direct result of the peculiarities of the datasets.

The final model for Sanlando Springs is relatively complex, with 10 inputs for the rainfall model and 5 for the groundwater-use model. The rainfall model uses one input based on the standard rainfall consisting of a 270-day MWA. In addition, the rainfall model uses nine inputs with MWAs (90, 270, 630, or 1,440 days in length) for normalized rainfall at NOAA rain gages 1641, 2229, 3137, 5076, 5612, 5973, and 6628 (see appendix 3 for further details). The standard rainfall was computed as the arithmetic average of the values for the 18 NOAA rain gages (fig. 33). A slight bias exists in the rainfall model residuals, leading to a downward trend (linear correlation) in residual error (R = –0.021; fig. 49*B*), which is qualitatively consistent with possible decreases in groundwater level attributable to increases in groundwater use.

The groundwater-use model was used to better explain the observed hydrologic response by addressing the bias in rainfall model residuals. Inputs for the groundwater-use model for this site are a 12-month MWA, 1-month minus 3-month MWD, 3-month minus 6-month MWD, and 24-month minus 48-month MWD of total groundwater use; the predicted spring flow from the rainfall model also was used (appendix 3). Addition of the groundwater-use model eliminated this negative bias such that the long-term trend in residuals for the final model (rainfall plus groundwater use) is slightly upward (R = 0.043; fig. 49*B*). Such a small change may not be significant, however, and suggests that long-term (multidecadal) trends in spring flow at this site attributable to groundwater use are small or cannot be definitively inferred from the available data. The groundwater-use model explained only about 4 percent of the variability in the rainfall model residuals (R^2 = 0.0417; appendix 3). The inclusion of predicted spring flow from the rainfall model as an input parameter caused an insignificant change in R^2, but was necessary in order to obtain a model with negative input-output sensitivity (an increase or

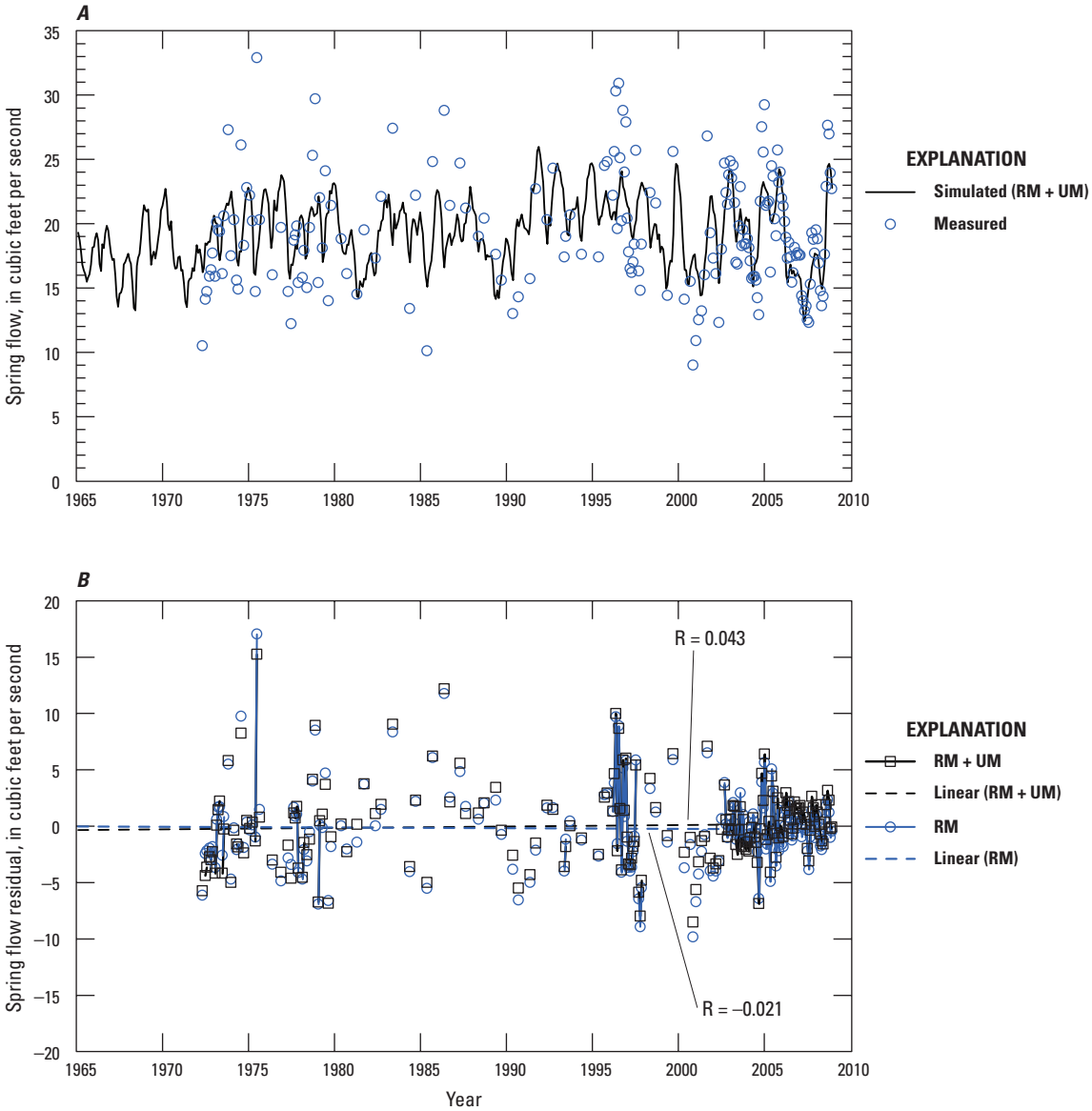


Figure 49. Results of spring-flow simulation for Sanlando Springs (site number 312; table 2; fig. 32) showing *A*, measured and simulated water levels, and *B*, residuals for the rainfall model (RM) and the final combined rainfall and groundwater-use model (RM + UM) with associated linear regression lines. For the residuals, lines connecting points indicate periods of consecutive monthly data. R is the Pearson product-moment correlation coefficient for the linear regression of residuals and time.

decrease in groundwater-use causes a decrease or increase, respectively, in spring flow). These poor correlations and difficulty maintaining correct input-output sensitivity suggest either that there is little impact of groundwater use on spring flow or that such a relation cannot be discerned with the available data. Nevertheless, the groundwater-use model tended to mitigate negative residuals from the rainfall model, indicating that overpredicted water levels from the rainfall model were reduced by the groundwater-use model (fig. 49*B*). The overall average absolute residual error decreased 0.004 ft^3/s from the rainfall model to the final model, indicating that inclusion of the groundwater-use model slightly improved final model accuracy. The large amount of missing data, irregular measurement frequency, and differences in data variability between the sparse older data and the later period of daily data are important factors that contributed to the relatively poor fit of the spring-flow model for Sanlando Springs.

Development of the Decision Support System

Dutta and others (1997, p. 912) define decision support systems, or DSSs, as "...systems helping decision-makers to solve various semi-structured and unstructured problems involving multiple attributes, objectives, and goals. Historically, the majority of DSSs have been either computer implementations of mathematical models or extensions of database systems and traditional management information systems." DSS technology can help provide the knowledge and tools required for informed decision making (Roehl, Conrads, and Daamen, 2006). Even though the collective interests and computer skills within the community of water-resource managers, scientists, and other stakeholders are wide ranging and unequal, equal access to the broad scope of current scientific knowledge is needed to make the best possible decisions.

The CFANN DSS integrates historical databases and the site-specific ANN models, model controls, and model output into a spreadsheet application with a GUI that allows the user to simulate water-resource scenarios of interest. Scenarios can consist of adjustments to rainfall and groundwater-use data, whereby predictions of lake water levels, groundwater levels, and spring flows are provided by the CFANN DSS. Development of the DSS required the following: (1) merging all the data into a single comprehensive database; (2) developing rainfall and groundwater-use ANN models; and (3) developing a Microsoft Excel® application that integrates the new database, ANN models, model inputs and outputs, and graphical display routines into a single package that is easy to use and disseminate. The user's manual for the installation and operation of the CFANN DSS is provided in appendix 6.

System Architecture

The basic architectural elements, or functional components, of the CFANN DSS are shown in figure 50. The DSS reads and writes files for the various run-time options that can be selected by the user through the system's GUI. A historical database, containing 59 years (1950 to 2008) of monthly average rainfall, hydrologic response data (lake water level, groundwater level, and spring flow), and groundwater-use data, is read into the simulator along with the ANN models at the start of a simulation. By using GUI controls, the user can evaluate scenarios for alternative rainfall and groundwater use. The outputs generated by the simulator are written to files for post processing in Microsoft Excel® or other analysis software packages. During simulations of lake water level, groundwater level, or spring flow response by the ANN models, the DSS provides a tabular display of simulated values for the current computational time step for all sites and streaming graphics for one selected site.

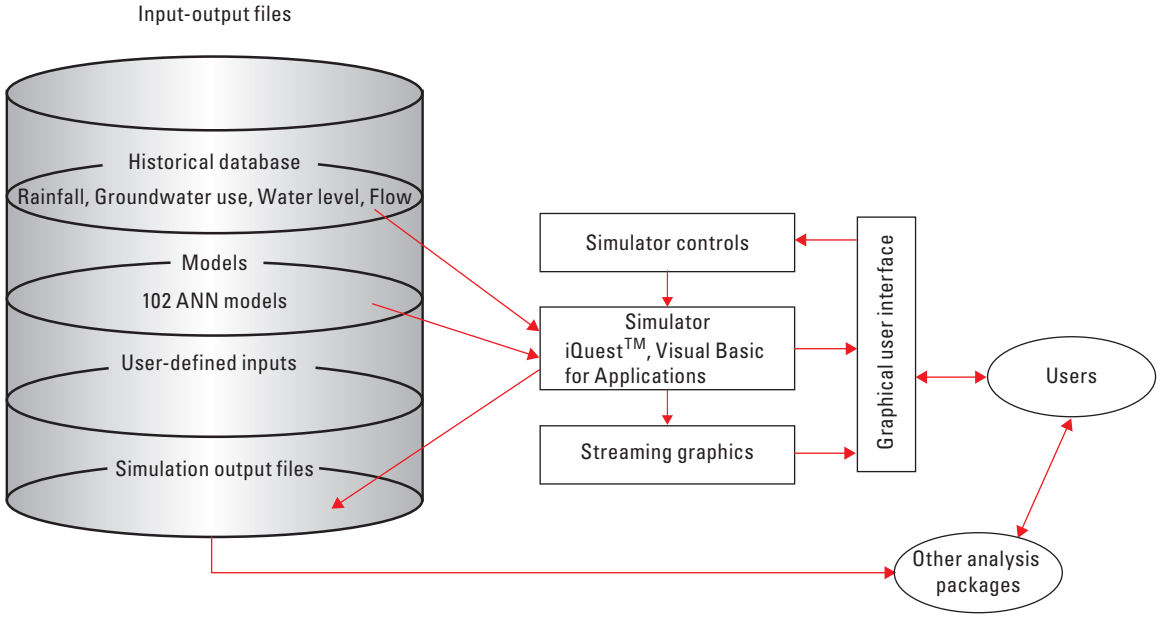


Figure 50. Architectural elements of the Central Florida Artificial Neural Network Decision Support System.

Model Simulation Control and Graphics

The simulator in the CFANN DSS integrates the historical database with the 102 ANN models. The date/time controls on the user control panel (fig. 51) are used to adjust the start and end dates and graphical and tabular output for a simulation. The simulator allows the user to run "what-if" simulations by varying the rainfall and groundwater-use inputs from the historical values. The user has two simulation options for changing rainfall inputs and three options for changing groundwater-use inputs:

- as a percentage of historical rainfall or groundwater-use values,
- as a constant offset from historical rainfall or groundwater values, and
- as a user-defined time series of groundwater use.

Explanations of how each of the options in the CFANN DSS can be implemented are provided in the user's manual (appendix 6).

The top of the CFANN DSS control panel (fig. 51) shows the simulation period, output options, and user-setting buttons that allow specifications of changes to rainfall and groundwater-use values. The CFANN DSS also shows streaming graphics on the control panel while a simulation is running. The graphs display the historical measured data, simulated historical conditions (to show model accuracy), the simulated output using the input option set by using the GUI controls or an input file, and the difference in simulated outputs between user-specified and historical conditions (fig. 51).

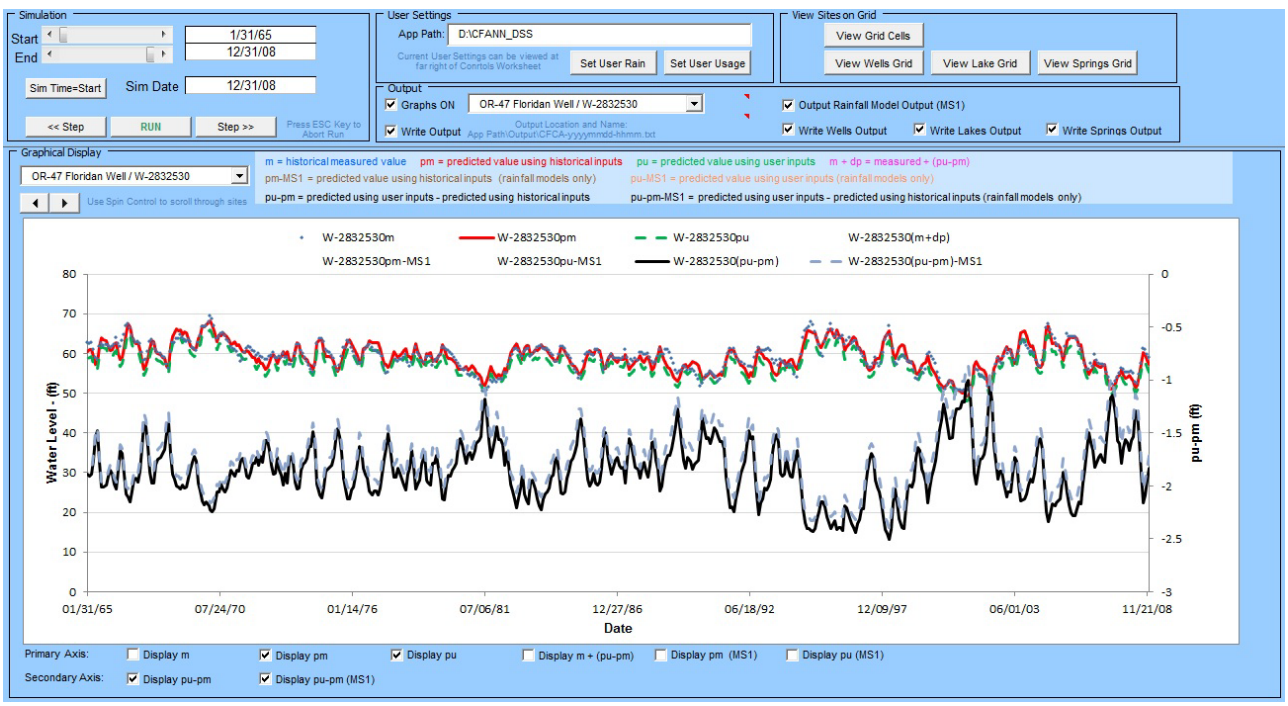


Figure 51. DiagramSimulator controls used to run a simulation of the Central Florida Artificial Neural Network Decision Support System and the associated graphical display.

Comparison of Rainfall and Groundwater-Use Effects

To help understand the effects of rainfall and groundwater use on hydrologic response in central Florida, the CFANN DSS can be used to facilitate a sensitivity analysis and as a component of a vulnerability assessment. In the following sections, the results of sensitivity analyses are described and the potential for application of the forcing-response behavior represented by the CFANN DSS to assessments of vulnerability is discussed

Sensitivity Analysis

The CFANN DSS was used to simulate four scenarios, representing hypothetical changes in historical rainfall and groundwater use, to determine changes in lake water level, groundwater level, and spring flow simulated by the ANN models. The scenarios consist of using time series of rainfall and groundwater use at 70 and 130 percent of historical values. These adjustments were effected by using the appropriate controls in the CFANN DSS to apply a constant multiplier to uniformly increase or decrease historical rainfall and groundwater-use time series by 30 percent (see sections 3.3 and 3.4.3 in appendix 6). The resulting simulated hydrologic response is indicative of the sensitivity of the model for each site to equal percentage changes in rainfall and groundwater use that are constant over time. Actual changes in rainfall and groundwater use are not likely to be equal even on a percentage basis and likely would vary over a range of short- and long-term time scales; thus, these simulations should not be interpreted as representing expected changes. Analysis of hypothetical scenarios, however, can be an effective tool for understanding model sensitivity to changes in model inputs, and the user can interpret results accordingly to determine if the model is appropriate for a particular application.

Results of the rainfall and groundwater-use scenarios indicate a wide range of responses in water levels at lakes and wells and flows at springs (figs. 52, 53). Most sites show a greater response to a 30-percent change in rainfall than a 30-percent change in groundwater use (for both maximum and minimum changes in water level or flow). Lakes and wells generally show a larger response to rainfall than groundwater use, whereas springs generally show a mixed response, with some springs showing a larger response to rainfall and some springs showing a larger response to groundwater use. For lakes, the median maximum changes in water level were a 5.07 ft increase for 130 percent rainfall, a 6.33 ft decrease for 70 percent rainfall, a 0.95 ft increase for 70 percent groundwater use, and a 1.22 ft decrease for 130 percent groundwater use. For wells, the median maximum changes in water level were a 2.39 ft increase for 130 percent rainfall, a 3.23 ft decrease for 70 percent rainfall, a 0.54 ft increase for 70 percent groundwater use, and a 0.51 ft decrease for 130 percent groundwater use. For springs, the median maximum changes in flow were a 3.57 ft^3/s increase for 130 percent rainfall, a 4.80 ft^3/s decrease for 70 percent rainfall, a 4.54 ft^3/s increase for 70 percent groundwater use, and a 3.60 ft^3/s decrease for 130 percent groundwater use.

Site specific differences exist, whereby some individual lakes, wells, and springs show a greater response to a 30-percent change in groundwater use than a 30-percent change in rainfall. Comparison of the ratios between maximum change in water level or flow due to increases or decreases in groundwater use and maximum change due to increases or decreases in rainfall indicates that these ratios typically exceed 1.0 for one lake (Lake Apopka, site 170), one well (Cocoa P Floridan well, site 622) and three springs (Wekiwa, Rock, and Palm Springs; sites 308, 309, and 313, respectively). Some variability in these ratios occurs because of asymmetric sensitivity where positive responses are not equal in absolute value to negative responses. For example, a 30-percent increase in rainfall generally caused smaller water-level increases for individual lakes and wells than the water-level decreases caused by a 30-percent decrease in rainfall (figs. 52*A*, 53*A*). Asymmetric sensitivity is common in complex natural systems and is simply the manifestation of nonlinearity in the input-output dynamics of the hydrologic system in central Florida. As a result of these nonlinearities, ratios of groundwater-use response to rainfall response more often exceeded 1.0 for rainfall increases than for rainfall decreases. The following 12 sites show a greater response to a 30-percent increase in groundwater use than a 30-percent increase in rainfall: Lake Clinch (site 50), Lake Deeson (site 77), Lake Apopka (site 170), Lake Howell (site 207), Lake Anderson (site 221), Lake Sue (site 297), Bay Lake Floridan well (site 602), Cocoa P Floridan well (site 622), Mascotte Surficial well (site 643), Wekiwa Springs (site 308), Rock Springs (site 309), and Palm Springs (site 313). The following nine sites show a greater response to a 30-percent decrease in groundwater use than a 30-percent increase in rainfall: Lake Apopka (site 170), Lake Howell (site 207), Lake Anderson (site 221), Cocoa P Floridan well (site 622), Lake Sawyer Floridan well (site 630), Taft Surficial well (site 712), Wekiwa Springs (site 308), Rock Springs (site 309), and Palm Springs (site 313). Of the sites showing a larger relative response, Lake Howell and the Taft Surficial well have small absolute responses to groundwater use where the maximum water-level increase or decrease was less than 1 ft (fig. 52*B*).

Before drawing specific inferences on rainfall and groundwater-use effects for individual sites, the CFANN DSS should be used to examine the simulated time series for the sites of interest. Two sites were selected for illustrative purposes: the ROMP 60 Floridan well (site 472), which shows a much larger response to a 30-percent change in rainfall compared to the same percentage change in groundwater use, and the Cocoa P Floridan well (site 622), which shows a larger response to a 30-percent change in groundwater use compared to the same percentage change in rainfall (figs. 52, 53). Discussion of the results from the two selected sites illustrates

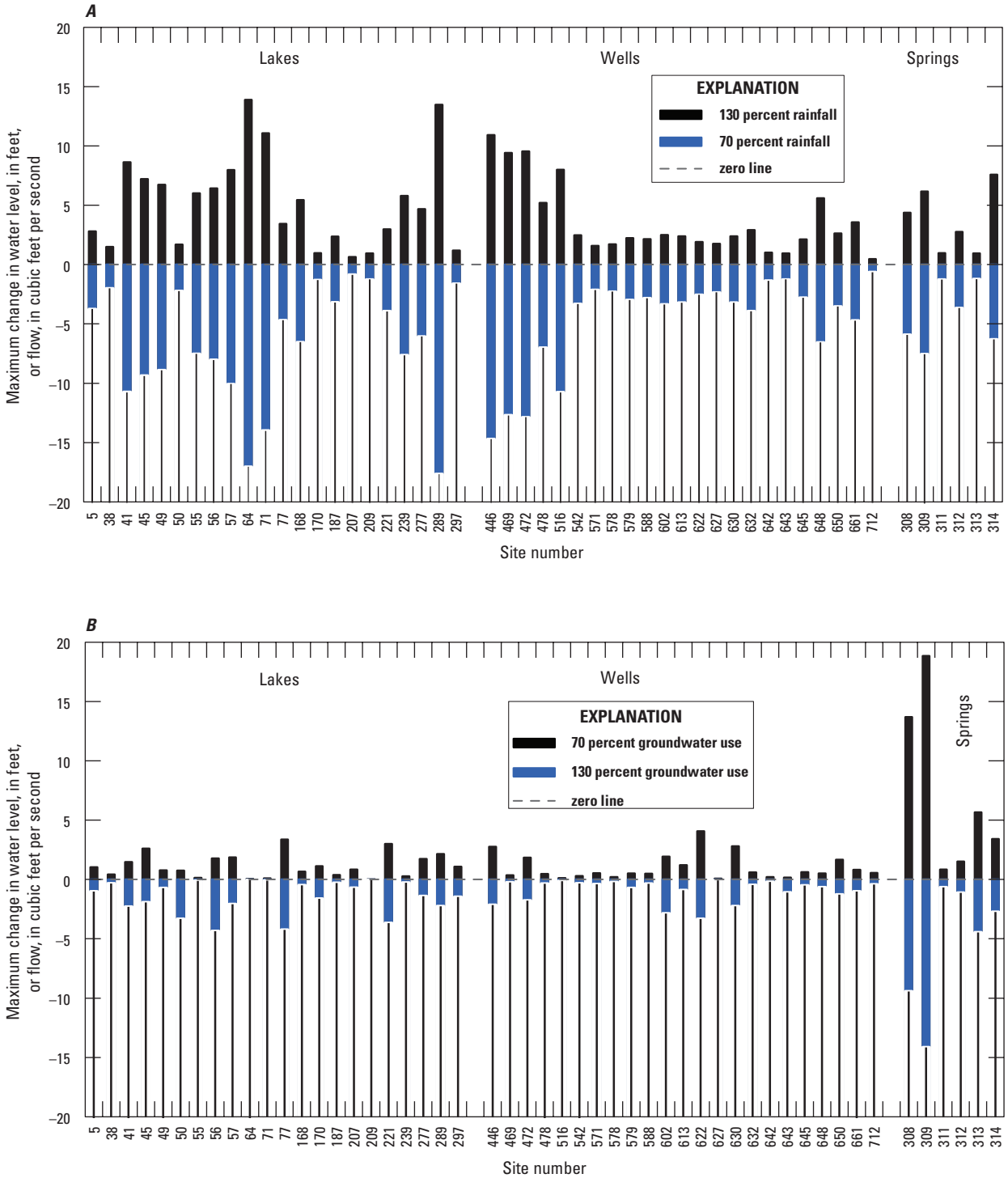


Figure 52. Sensitivity of lake water-level, groundwater-level, and spring-flow models as indicated by the maximum change in water level or flow simulated during 1965 to 2008 caused by changes in *A*, rainfall, and *B*, groundwater use. Rainfall and groundwater-use changes were simulated separately using the Central Florida Artificial Neural Network Decision Support System and specifying values of 130 percent for the increase and 70 percent for the decrease, which were applied to all historical values. Site information and locations are shown in table 2 and figure 32.

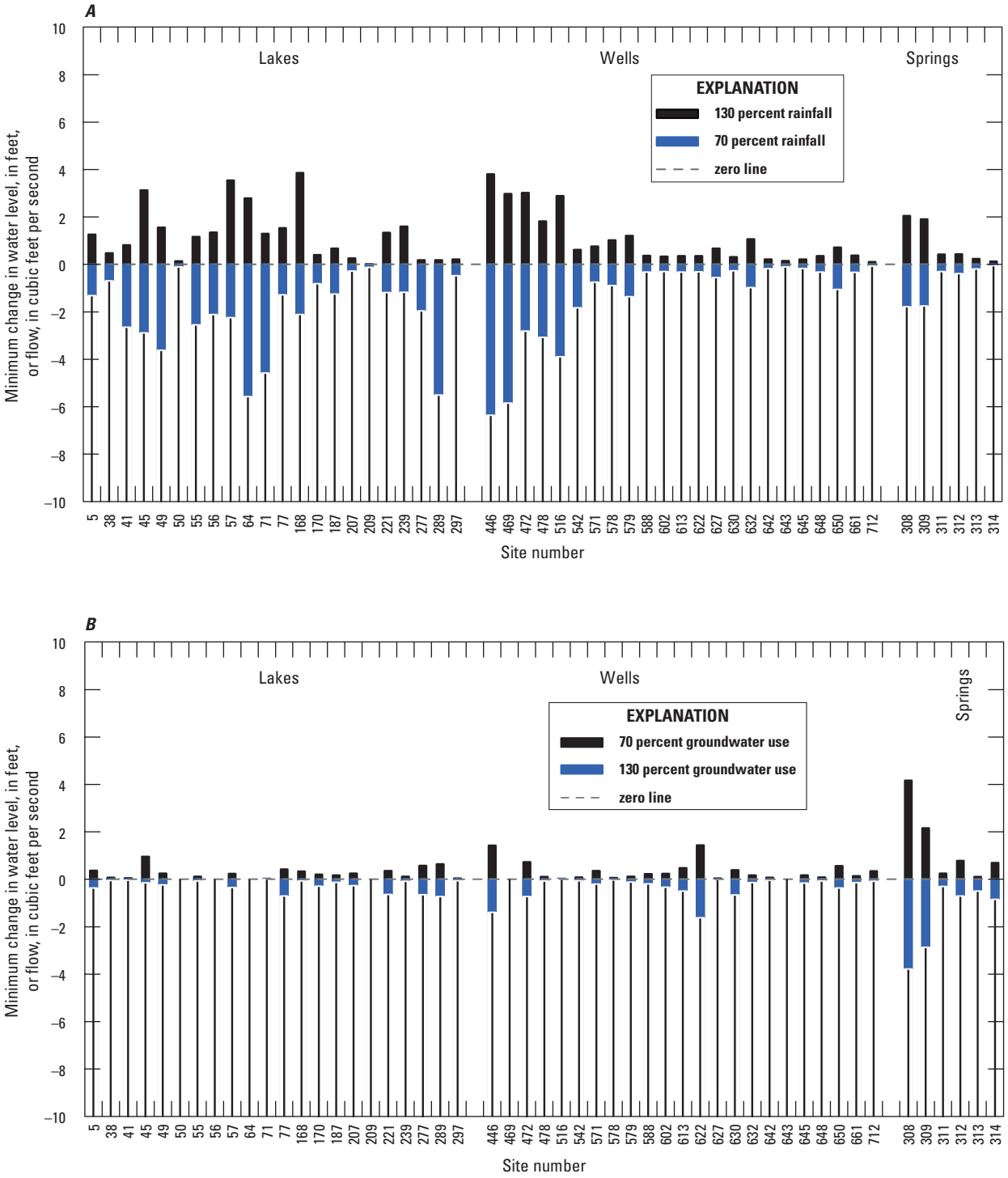


Figure 53. Sensitivity of lake water-level, groundwater-level, and spring-flow models as indicated by the minimum change in water level or flow simulated during 1965 to 2008 caused by changes in *A*, rainfall, and *B*, groundwater use. Rainfall and groundwater-use changes were simulated separately using the Central Florida Artificial Neural Network Decision Support System and specifying values of 130 percent for the increase and 70 percent for the decrease, which were applied to all historical values. Site information and locations are shown in table 2 and figure 32.

appropriate interpretation, and potential limitations, of the hypothetical rainfall and groundwater-use scenarios.

Groundwater level at the ROMP 60 Floridan well was highly variable, exhibiting fluctuations exceeding 10 ft over time scales ranging from seasonal to multidecadal (fig. 54*A*). In fact, many wells in southwest Polk County show large ranges in water level (fig. 10), and the five wells for which groundwater-level models were developed that show the largest responses to rainfall (sites 446, 469, 472, 478, and 516; figs. 52*A*, 53*A*) are located in this area (fig. 32). The history of the phosphate industry (both the mining and processing of ore) in southwest Polk County (fig. 6) and associated groundwater withdrawals (fig. 29) have been suggested previously as a factor affecting historical groundwater-level fluctuations in the area (Spechler and Kroening, 2007, p. 46). Variations in rainfall and agricultural groundwater use also have been suggested as contributing to reported groundwater-level trends (Spechler and Kroening, 2007, p. 46).

Linear correlation analyses indicated that correlations between rainfall and groundwater level at these five wells tended to be stronger (R ranged from 0.19 to 0.76 with an average of 0.54, fig. 35) than correlations with total groundwater use (R ranged from –0.55 to 0.49 with an average of –0.20, fig. 38). Consideration of spatial variations in groundwater use, which are not realized by using data for total groundwater use, may contribute to the differences in correlations. To investigate this possibility further, a linear correlation analysis was performed using the 12-month MWA of groundwater use (the sum of all categories) for each GUDV grid cell (fig. 28) and hydrologic response at each site. For GUDV grid cells in Polk County, the strongest correlations with groundwater-level for the five wells in question (sites 446, 469, 472, 478, and 516; fig. 32) were with groundwater use in grid cells RC70, RC71, and RC81 (fig. 28) (R ranged from –0.40 to –0.52 with an average of –0.46). These correlations generally are slightly stronger than those for total groundwater use on which the groundwater-use ANN models were developed (R between groundwater level and 12-month MWA of total groundwater use at these five wells ranged from –0.48 to 0.06 with an average of –0.30).

The relatively small responses to changes in groundwater use predicted by the groundwater-level models for the ROMP 60 Floridan well (fig. 54*B*) and the other four wells in this area (sites 446, 469, 478, and 516; figs. 52, 53) may be at least partly attributable to spatial variability not accounted for in the groundwater-use ANN models. Therefore, not accounting for spatial variations in groundwater use could result in underestimation of the effects of phosphate industry groundwater withdrawals. Groundwater use in the grid cells with stronger correlations (RC70, RC71, and RC81; fig. 28) primarily consists of phosphate industry withdrawals. Time series of groundwater use in grid cells RC71 and RC81 generally resemble the pattern of total phosphate industry groundwater use with an upward trend from 1950 to the mid-1960s followed by a more gradual downward trend to 2008 (fig. 29), whereas for grid cell RC70 the upward trend in groundwater use extended to the mid-1970s followed by a downward trend to 2008. In contrast, time series of groundwater use for other GUDV grid cells in the phosphate mining area (RC60, RC61, RC72, RC80, RC82; fig 28) show different patterns reflecting contributions from public supply and agricultural (predominantly citrus) groundwater withdrawals. In summary, groundwater-level fluctuations at the ROMP 60 Floridan well (site 472), and likely at the other four wells in southwest Polk County (sites 446, 469, 478, and 516; fig. 32), are attributable to a combination of fluctuations in rainfall, phosphate industry groundwater use, and other categories of groundwater use.

Groundwater levels at the Cocoa P Floridan well (site 622) show a slight but continuous downward trend from the beginning of the simulation period (1965) to the early 1980s, followed by a period of relatively flat trend to the end of the simulation period (fig. 55*A*). Superimposed on these long-term patterns are shorter multiyear periods of decline and recovery in the late 1960s, early 1980s, late 1980s to early 1990s, late 1990s to early 2000s, and the late 2000s (fig. 55*A*). These groundwater-level fluctuations correlate well with long-term trends and high frequency variability in both rainfall and groundwater-use. As a result, the groundwater-level model shows comparable responses to rainfall and groundwater use, with a slightly greater sensitivity to a 30-percent change in groundwater use, exhibiting a 1.43- to 4.07-ft increase or a 1.65- to 3.32-ft decrease in groundwater level, compared to a 30-percent change in rainfall, exhibiting a 0.35- to 1.91-ft increase or a 0.34- to 2.56-ft decrease in groundwater level (fig. 55*B*). The Cocoa P Floridan well is in east-central Orange County (site 622, fig. 32) in the expanding Orlando metropolitan area, which has experienced substantial urbanization (fig. 7) and the increases in groundwater withdrawals that accompany such population growth. When interpreting the time series of hydrologic responses from the hypothetical scenarios, it is important to note that applying constant percentage changes to rainfall and groundwater use results in variable absolute changes. For example, for the Cocoa P Floridan well, the smaller responses to rainfall changes in 2000 are at least partly attributable to the low rainfall totals that year (fig. 33*B*), and the long-term upward and downward responses to groundwater-use changes are at least partly attributable to long-term trends in groundwater use (fig. 34*A*).

Overall, results of the simulations of the hypothetical scenarios indicate that the flow of springs tends to be more sensitive to groundwater use than are the water levels of lakes or wells when compared on the basis of equal percentage changes in rainfall. Interestingly, the predominant sources of inflow to Lake Apopka (site 170) are surface-water runoff (average 1995 rate of 61 ft^3/s simulated by the ECFT model; Nicasio Sepúlveda, U.S. Geological Survey, written commun., 2013) and Apopka Spring (1971–2009 mean flow of 28 ft^3/s), and the hypothetical scenarios show a lake water-level response to groundwater use comparable to that of rainfall. Even though the absolute simulated water-level changes are relatively small (0.20 to 1.62 ft; figs. 52, 53) for Lake Apopka, the similar relative sensitivity to rainfall and groundwater-use

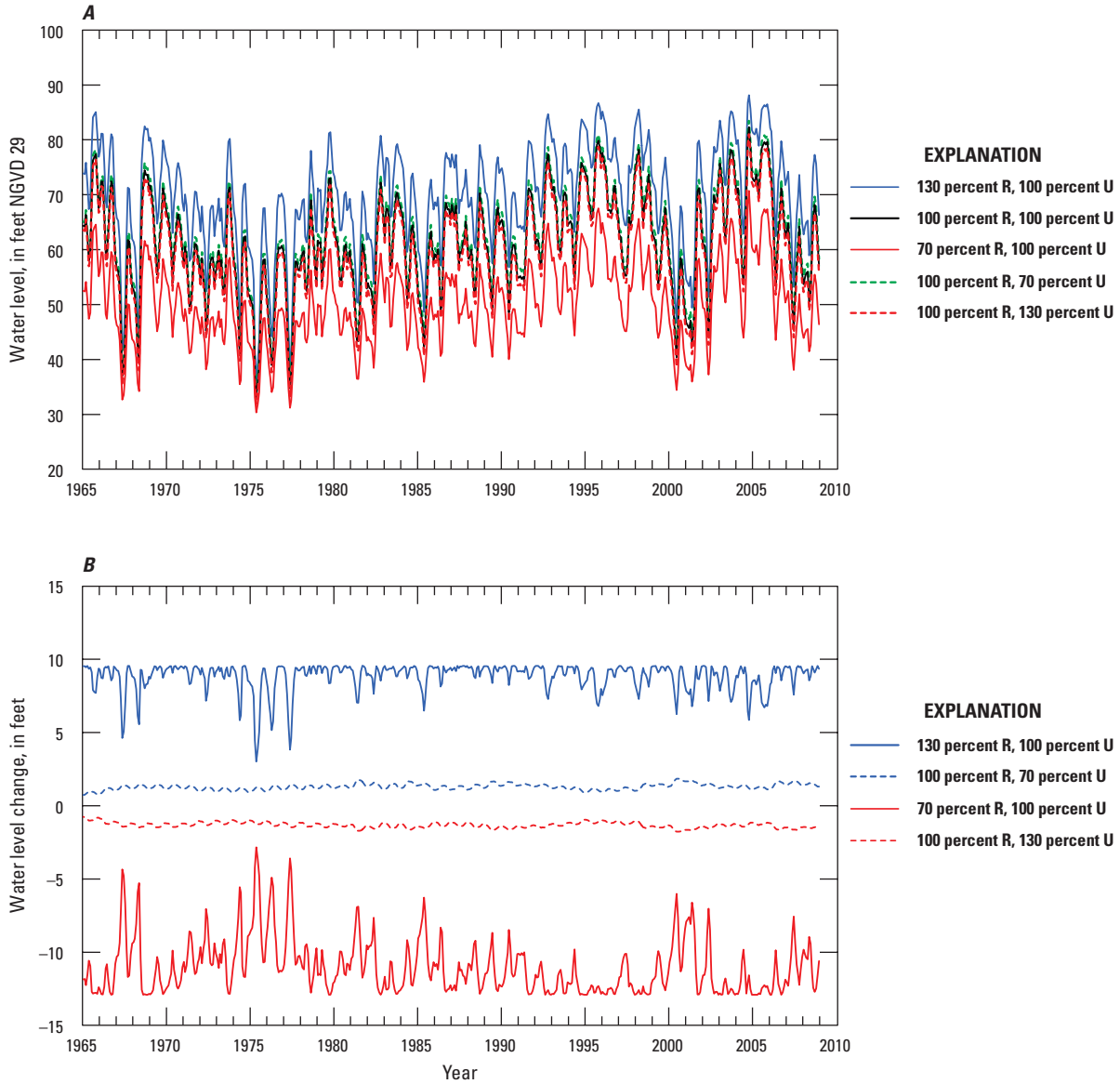


Figure 54. Simulated water levels at the ROMP 60 Floridan well (site 472; table 2; fig. 32) for historical conditions and hypothetical rainfall and groundwater-use scenarios showing *A*, water-level altitude for historical and scenario conditions, and *B*, the change in water level between historical and scenario conditions. Historical conditions are represented by using 100 percent rainfall (R) and 100 percent groundwater use (U). Scenarios are represented by increases and decreases of 30 percent in historical values of rainfall and groundwater use.

changes and the presence of Apopka Spring inflow suggest that the sensitivity of lake water-level response may be affected by changes in flow from the spring. Springs are commonly considered to reflect the integrated effects of hydrologic conditions over large areas. Additionally, from a long-term, steady-state, water-balance perspective, increases in groundwater withdrawals from an aquifer system are balanced by increases in aquifer recharge or decreases in aquifer discharge, with reductions in natural discharge typically being the most important (Bredehoeft, 1997).

In central Florida, spring flow represents a substantial outflow from the aquifer system, totaling 68 percent of groundwater withdrawals for 1995 (ECF model; McGurk and Presley, 2002) and 38 percent of groundwater withdrawals for 1995 to 2006 (ECFT model; Sepúlveda and others, 2012) (fig. 41*B*). Accordingly, increases in groundwater withdrawals over time are likely to eventually result in some degree of decreases in spring flows as the hydrologic system moves toward a long-term balance, and the sensitivities of the spring-flow ANN models to the hypothetical rainfall and

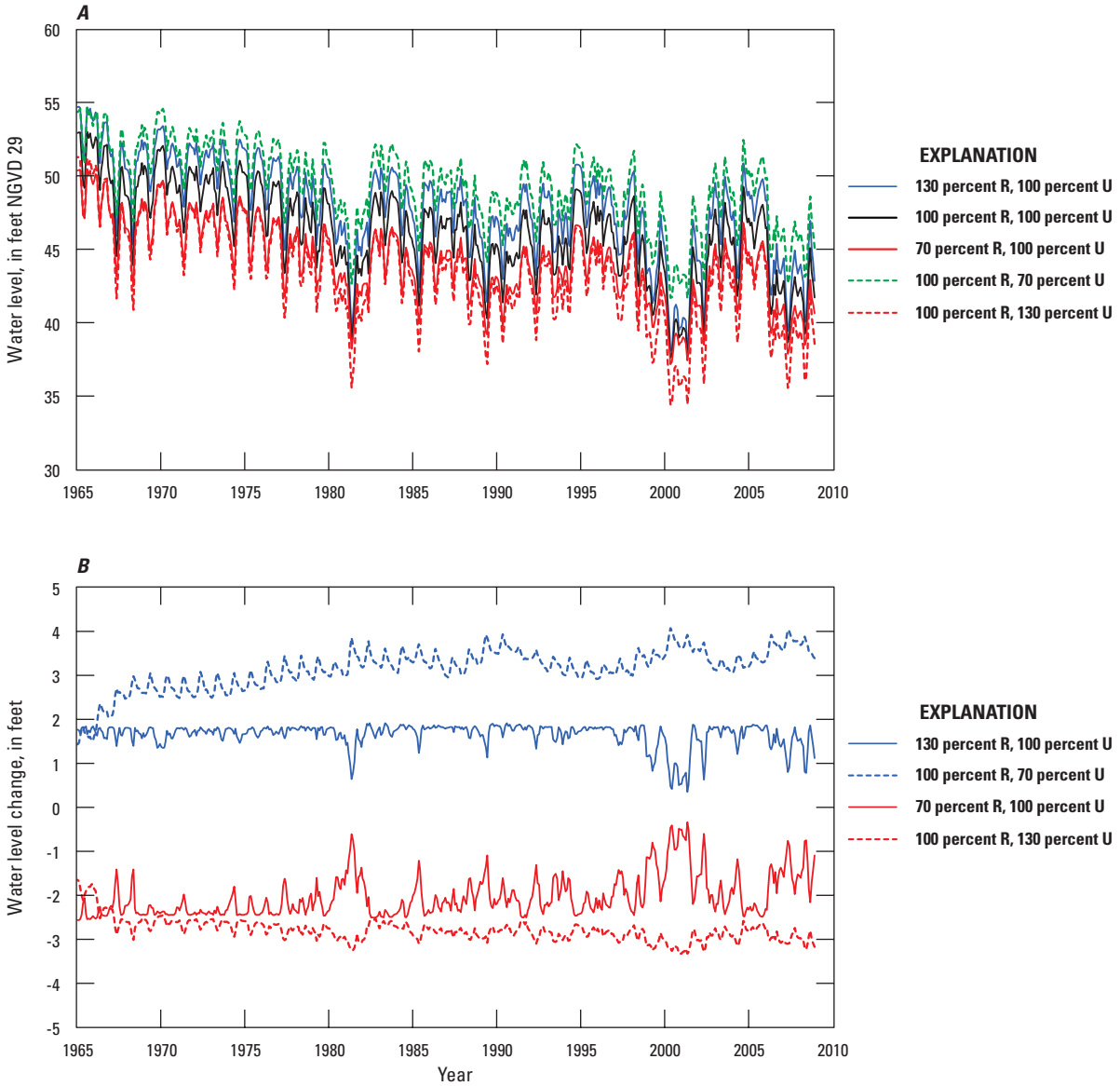


Figure 55. Simulated water levels at the Cocoa P Floridan well (site 622; table 2; fig. 32) for historical conditions and hypothetical rainfall and groundwater-use scenarios showing *A*, water-level altitude for historical and scenario conditions, and *B*, the change in water level between historical and scenario conditions. Historical conditions are represented by using 100 percent rainfall (R) and 100 percent groundwater use (U). Scenarios are represented by increases and decreases of 30 percent in historical values of rainfall and groundwater use.

groundwater-use scenarios are at least qualitatively consistent with this equilibrium-seeking process. Compared to the springs, the lower sensitivities of most lakes and wells to the same hypothetical rainfall and groundwater-use scenarios are likely attributable in part to the simple hydraulic principle that a small change in water level can cause a large change in flow rate if such a water-level change results in a large percentage change in hydraulic gradient. Thus, modulation by the hydrologic system of the relations between water-level changes and flow rates determines the degree to which the water balance responds to changes in the primary system forcings of rainfall and groundwater use.

Comparison of the predictive accuracies of the rainfall and groundwater-use ANN models indicates that the hydrologic system in central Florida generally is affected by groundwater use differently during wet periods than during dry periods. During wet periods, when little or no system storage is available (high water levels), groundwater-use effects are less evident. By contrast, during dry periods, when there is excess system storage (lower water levels), groundwater-use effects

are clearly noticeable in the historical record of hydrologic response. The residual error time series of the final water-level and spring-flow models (combined rainfall and groundwater-use ANN models) were compared to the residual error time series of the respective rainfall ANN models—a reduction in residual error of the final combined model compared to the rainfall model indicates the improvement in predictive accuracy gained by addition of the groundwater-use model. This improvement in predictive accuracy varied with the magnitude of the hydrologic response, where a larger reduction in residual error coincided with periods of lower water levels or flows (bottom 25 percent of observed 1965–2008 data) and a smaller reduction in residual error coincided with periods of higher water levels or flows (upper 25 percent of observed 1965–2008 data). The reduction in residual error generally was greater for months with lower than higher water levels or spring flows, for example: 0.20 ft for Prevatt Lake (site 5), 0.087 ft for the OR-47 Floridan well (site 648), and 0.16 ft^3/s for Sanlando Springs (site 312). Extended periods of lower water levels or spring flows occur during drought periods, such as experienced throughout central Florida in the early 1980s and late 1990s to early 2000s (for example, figs. 45*A*, 46*A*, 49*A*). Thus, by driving the overall behavior of the system, rainfall indirectly influences the degree to which groundwater use will drive persistent trends in water levels and flows, with groundwater-use effects more prevalent during periods of low water levels and spring flows caused by low rainfall and less prevalent during periods of high water levels and spring flows caused by high rainfall. Differences in the magnitudes of rainfall and groundwater use during wet and dry periods also are important determinants of hydrologic response.

Potential Application to Vulnerability Assesment

Assessment of vulnerability of the hydrologic system in central Florida to environmental change, whether natural or human derived, is a potential application of the forcing-response behavior learned through the data-mining analyses. Vulnerability has many working definitions, but Füssel (2007) describes four fundamental dimensions needed to characterize a vulnerable situation: system of concern, valued attributes of the system, hazards to which the system is subject, and time period for which assessment is desired. Luers and others (2003) describe vulnerability as "...the degree to which human and environmental systems are likely to experience harm due to a perturbation or stress...," and they developed a vulnerability metric as the ratio of the system sensitivity to the system state relative to a threshold below which harm would occur times the occurrence probability of the forcings of concern. Adger (2006) notes that in the framework of Luers and others (2003), sensitivity may be a physical parameter of the system or a social parameter that measures well being. Thus, the quantification of system sensitivity factors prominently into an assessment of vulnerability. The CFANN DSS provides an easy-to-use tool for sensitivity analysis that could facilitate vulnerability assessments. Results of the CFANN DSS could be incorporated either directly into vulnerability assessments based on water-level or flow conditions or indirectly if relations between predicted water-level or flow sensitivities and some measure of system well being (for example, social or economic effect) are established. On the basis of a range of scenarios of interest to water-resource managers, sensitivities derived from the CFANN DSS could be used to identify the contribution of rainfall and groundwater use to the hydrologic system's overall vulnerability. Considerable uncertainty can exist in quantifying the forcing-response behavior of a system, and results may need to be interpreted somewhat qualitatively. Luers (2005) recognized that some systems cannot be definitively quantified and presented a framework whereby the relative vulnerability of a system to forcings of concern under different scenarios can be identified; such an approach may prove beneficial to water-resource management decision making.

Limitations of Artificial Neural Network Models

Several factors potentially limit the application of the ANN models: confounding effects of correlation between rainfall and groundwater use, quality and completeness of the historical databases, and spatial variations in groundwater use. These factors should be considered when interpreting results of simulations made with the ANN models and are discussed in more detail below.

Often, explanatory variables share information about the behavior of a response variable. It is difficult, if not impossible, to distinguish the individual effects of these variables (sometimes known as confounded or correlated variables) on a response variable. For the ANN models, rainfall and groundwater use are correlated to some degree because rainfall patterns partly determine when and how much groundwater is withdrawn by utilities, industry, and agriculture. Correlation analysis of rainfall and groundwater use and results herein indicate that groundwater use tends to be negatively correlated with rainfall, with R values ranging from –0.76 to 0.71 and an average of –0.14 (fig. 39). Application of the two-stage modeling architecture, whereby hydrologic response was first modeled with rainfall and then with groundwater use (fig. 40), allowed the possibility during the model training process that the correlation between rainfall and groundwater use was incorporated into the rainfall ANN model, causing aliasing of this interaction between groundwater-use and rainfall effects by the rainfall model inputs. Aliasing is caused by confounded variables, which may lead the analyst to attribute the effects to a particular variable or combination of variables when the actual causative behavior is due to another variable or combination of variables. It was necessary to use this two-stage approach for the reasons described in the Input-Output Mapping and Problem Representation section, with the

primary reasons being the limited availability and quality of the groundwater-use data and the findings of the current and previous studies that indicate rainfall was the predominant factor affecting the water budget and the variability of hydrologic response. For the rainfall ANN models, the likelihood of aliasing increases as the model is more closely trained to match the observed data. Therefore, overfitting of the ANN models was mitigated by using a conservative training process, namely by limiting the ANN to one hidden-layer neuron, culling inputs with input-output sensitivities at odds with principles of mass conservation, retaining only those inputs that are not highly correlated with each other, and using 2-fold cross-validation where possible. An important characteristic of aliasing is that, whereas the individual effects of the confounded variables cannot be unambiguously separated, the combined effect can be accurately modeled. Although, significant changes in the underlying causative relations between the confounded variables, such as a change in water-management practices that affect water-use patterns, may result in poor performance of the model for predictive scenario simulations. Stronger aliasing causes greater ambiguity in separating the effects, and thus the higher the correlation between rainfall and groundwater use, the greater will be the potential for rainfall and groundwater-use interactions to alias rainfall effects.

Empirical models, such as ANNs, are especially sensitive to the quality and completeness of the data from which they were developed. Much of the groundwater-use data was estimated owing to a paucity of historical measurement data. Additionally, groundwater return flows (return of groundwater withdrawals to the aquifer system) are difficult to quantify, but available data indicate they could be important to an understanding of the water resources of central Florida. For example, a total of 151 Mgal/day of treated wastewater (from municipal and privately owned treatment facilities) was disposed of by land application in Lake, Orange, Osceola, Polk, and Seminole Counties in 2000, accounting for 86 percent of the total wastewater discharged by municipal and private systems in these counties (Marella, 2004, p. 37). For these same five counties in 2000, 424 Mgal/day of groundwater was withdrawn for public supply (from a publicly or privately owned community water system) (Marella, 2004, p. 14). As a result, assuming all municipal and privately owned wastewater-treatment facilities receive water originally supplied by a publicly or privately owned facility, about 36 percent of public-supply groundwater withdrawals in 2000 were returned by land application of treated wastewater. Although some of this water is lost to ET, much of it likely recharges the groundwater system depending on the particular land application methods used. Additionally, 25 to 75 percent of domestic public-supply water typically is used for irrigation (Marella, 2009, p. 15), which provides additional potential groundwater recharge. The lack of quantitative historical data on groundwater return flows could have affected correlations identified with the groundwater-use data, because hydrologic response is affected by the net effects of anthropogenic withdrawals and inputs, which for the current study were all groundwater withdrawals with the exception of drainage well recharge.

By contrast, the NOAA rain gage data were of high quality with few missing data, but only 18 gages were within or in the vicinity of the study area. Cluster analysis of the NOAA rain gage data showed that rainfall is highly variable in time and space (figs. 24, 25), and if additional long-term rainfall measurement sites had been available they likely would have provided valuable information to further define the temporal and spatial variability of rainfall. Missing hydrologic response data, which exceeded 25 percent for 22 of the sites (appendix 5), can lead to model inaccuracy owing to a lack of information on which to train the model (for example, the model for Prevatt Lake; fig. 45). In addition to incomplete records, relatively short periods of record for both groundwater-use data as well as hydrologic response data for some sites necessitated limitation of the simulation period to 43 years (1965 to 2008). Despite this multidecadal dataset, the hydrologic system was subject to only a limited number of large impact events, such as droughts and El Niño/La Niña periods, onto which the long process of human development is superimposed. Large impact events often provide the highest value input-output responses for evaluation. Increasing population and the associated land development and groundwater withdrawals in central Florida clearly predate 1965, but data during a period of predevelopment or modest development were simply too sparse or nonexistent to develop models representing a larger range in anthropogenic influences. For models to represent the fullest ranges of variability, they must be trained on the most complete and longest-term datasets available that encompass a comprehensive range of field conditions.

The effects of the spatial variability in groundwater use warrant further analysis to determine if sufficient information is present in the historical groundwater-use and hydrologic response data to quantify such effects. Examination of historical groundwater-use data using the GUDV shows that spatial variations are present (figs. 30, 31). Likewise, spatial variations in the hydraulic properties of the Floridan aquifer system are common, such as areas of relatively low transmissivity for the Upper Floridan aquifer in east Polk County and south Osceola County (Kuniansky and others, 2012) and may accentuate hydrologic responses to spatial variations in groundwater use. For some sites, such as the five wells in southwest Polk County (sites 446, 469, 472, 478, and 516; fig. 32) discussed in the previous section, linear correlations with groundwater use from individual GUDV grid cells were slightly stronger than correlations with total groundwater use. Overall, however, linear correlations with the MWAs of total groundwater use, which were the inputs used in the groundwater-use ANN models, were similar in magnitude to correlations with groundwater use for individual GUDV grid cells. Also, uncertainty exists in the locational information for some categories of groundwater use. Information on the historical spatial distributions of citrus irrigation, non-citrus agricultural irrigation, and golf course irrigation were limited, and these data should be interpreted only on a semi-regional scale (McLeod and Munch, 2012). Trichakis and others (2011), in a study of groundwater-level fluctuations in the Edward's

aquifer in Texas using ANNs, noted the importance of accurate and complete groundwater-use data for the prediction of abrupt changes in groundwater levels characteristic of karstic aquifers. Limitations on the availability of such groundwater-use data during the current study and uncertainty in groundwater return flows that may offset some of the groundwater withdrawals contributed to the need to model groundwater use on the basis of a single (spatially totaled) time series for the study area.

Given the limitations that have been described, care should be exercised to interpret the results of the ANN models carefully within the context of the assumptions, approach, and analyses documented herein. The final lake water-level, groundwater-level, and spring-flow models incorporated in the CFANN DSS should not be used alone to quantify the individual effects of rainfall and groundwater use for water-supply planning or management purposes. Examination of other lines of evidence, including physics-based models and other data such as quantified land-use and land-cover changes not examined during the present study, is not only warranted but may be indispensable. The knowledge gained from data-driven modeling and the results from physics-based modeling, when compared and used in combination, can yield a more comprehensive assessment and a more robust understanding of the hydrologic system than either of the approaches used separately.

Comparison of Artificial Neural Network Models With a Physics-Based Model

The recent completion of the ECFT groundwater flow model (Sepúlveda and others, 2012), which encompasses nearly the entire CFCA with the exception of a small portion of western Polk County, provides an opportunity to compare results between this physics-based model and the data-driven ANN models. Of the 51 sites for which ANN models were developed and incorporated in the CFANN DSS, 48 sites are simulated in the ECFT model. The three remaining sites—Island Lake at Longwood, Lake Anderson at Orlando, and Crooked Lake at Orlando (sites 209, 221, and 239, respectively; fig. 32)—were not included in the ECFT model. For the 48 sites common to the ANN and ECFT models, a Microsoft Excel® spreadsheet called the Model Data Viewer (MDV) was developed to facilitate comparison of model performance for both models. The MDV provides tabular and graphical displays of descriptive statistics for measured, ANN-simulated, and ECFT-simulated data and model-fit statistics for ANN- and ECFT-simulated data. Further details on the functionality and application of the MDV are provided in appendix 7.

The MDV was used to compute model-fit statistics (R^2, NSCE, RMSE, and PME) for the overlapping simulation period of the ANN and ECFT models, consisting of 144 monthly values from January 1995 to December 2006 (figs. 56, 57). Additionally, a complete tabular listing derived from the MDV of descriptive and model-fit statistics for measured and simulated data is provided in appendix 8. The ANN models generally had better model-fit statistics than the ECFT model, indicating that the ANN models were able to more closely match the measured data for most sites. The greatest differences between the ANN and ECFT models generally are present in water levels for the wells and to a lesser degree the lakes. The larger error for the ECFT model is attributable to underrepresenting variability in the data (as indicated by, for example, R^2), but often to a larger degree the error also is caused by larger absolute error (as indicated by, for example, PME and ME) compared to the ANN models. For example, for the Lake Oliver Surficial well (site 579), R^2 is 0.848 for the ECFT model and 0.864 for the ANN model (figs. 57*A*, 56*A*, respectively), whereas the ME is 2.62 ft for the ECFT model and 0.11 ft for the ANN model (appendix 8). As a result, the NSCE for the ECFT model (–1.18) is much lower than the NSCE for the ANN model (0.86) for this well (figs. 57*B*, 56*B*, respectively). Such differences in absolute error are attributable in part to differences in how ANN and ECFT models are constructed. Because of the spatially distributed nature of groundwater flow models, simulated conditions and any error at individual sites may affect other sites, causing, for example, the simulated water level at a well to be consistently offset from the measured values. In contrast, the ANN models are site-specific models; therefore, simulated conditions or errors at one site cannot affect other sites.

For the springs, the ANN and ECFT models generally had similar accuracy. For example, the ECFT model had better model-fit statistics for Palm Springs (site 313) than the ANN model—although both models had difficulty accurately simulating flow at this spring (figs. 56, 57). Issues with the quality and completeness of the spring-flow data described previously likely contributed to inaccuracies in the simulation of flow at springs for the ANN models as well as the ECFT model. For some sites, the model-fit statistics for the CFANN DSS simulation period (1965 to 2008; fig. 43) differ substantially from those shown in figure 56 for the period 1995 to 2006. For example, for the spring-flow model for Palm Springs, R^2 is 0.627 for 1965 to 2008 (fig. 43*A*, appendix 5) and 8.9×10^{-5} for 1995 to 2006 (fig. 56*A*, appendix 8). These differences are simply the result of the different simulation periods and inability of the ANN model, or any model, to match all data equally well. Had the ANN and ECFT models been calibrated using only contemporaneous data, the ANN models likely would have had comparatively better model-fit statistics for all sites. The capability of ANN models to accommodate a large number of degrees of freedom (adjustable model parameters) contributes to its flexibility in replicating observed data, whereas physics-based groundwater models are limited internally to the hydraulic parameters of the hydrogeologic units.

Further insight into how simulated results for each model compare at an individual site can be obtained using the graphical display functionality of the MDV. Graphs of time-series

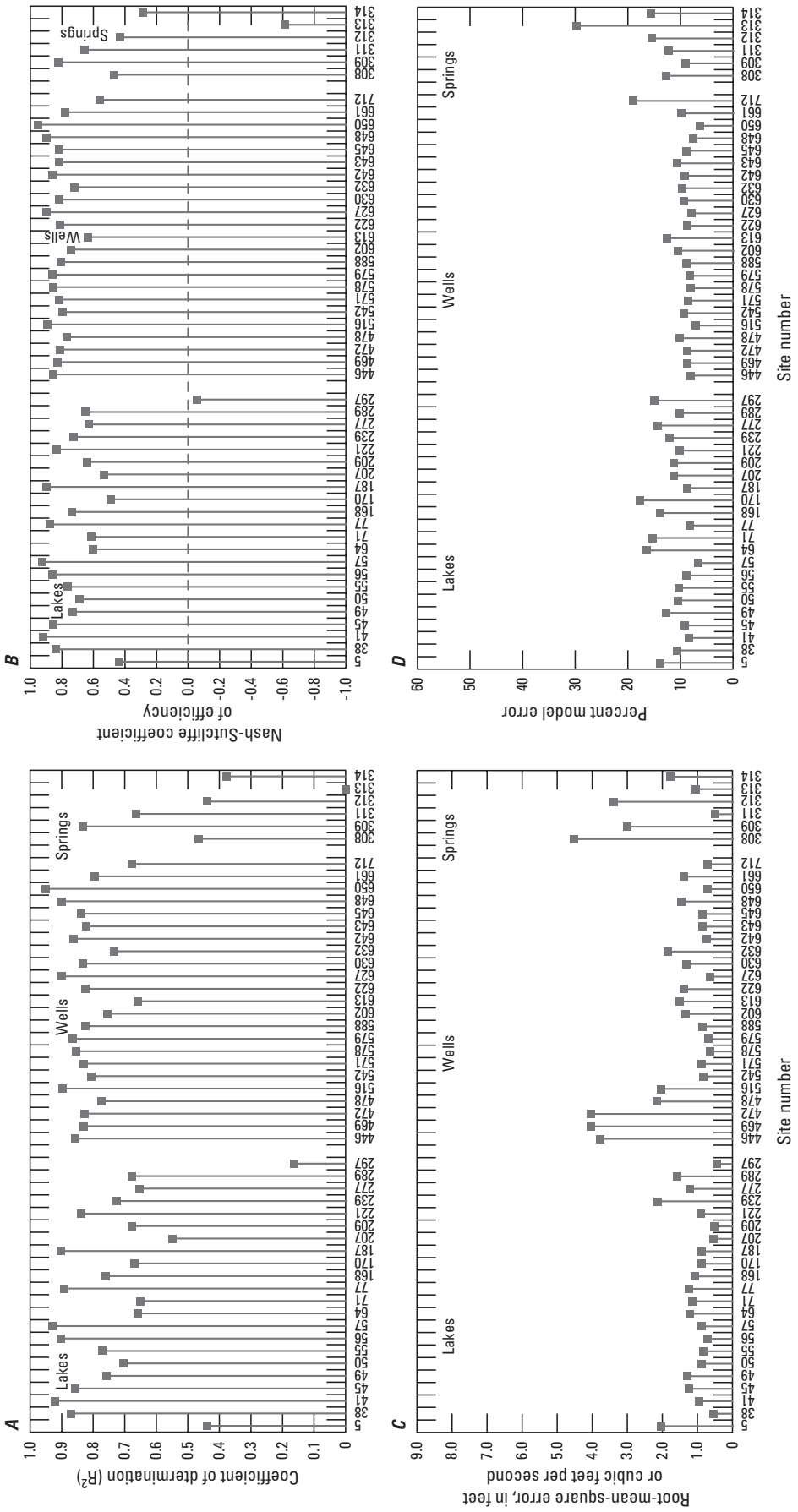


Figure 56. Model-fit statistics for the final water-level or flow models for the 51 sites incorporated in the Central Florida Artificial Neural Network Decision Support System for the 1995 to 2006 period: *A*, coefficient of determination (R^2), *B*, Nash Sutcliffe coefficient of efficiency, *C*, root-mean-square error (RMSE), and *D*, percent model error (RMSE divided by the range of data). Site information and locations are shown in table 2 and figure 32.

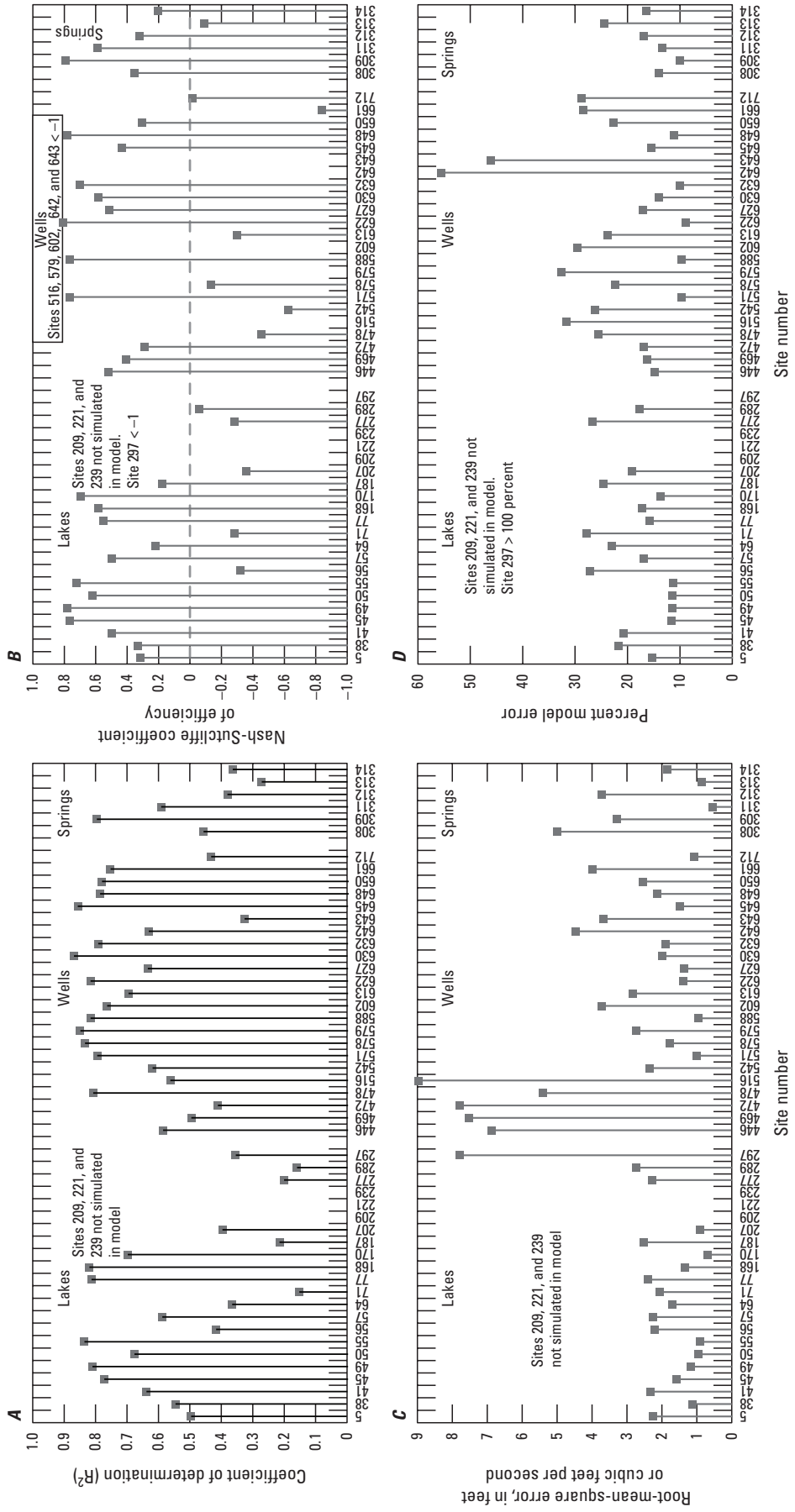


Figure 57. Model-fit statistics for 48 sites in the east-central Florida transient (ECFT) model (Sepúlveda and others, 2012) for the 1995 to 2006 period: *A*, coefficient of determination (R^2), *B*, Nash Sutcliffe coefficient of efficiency, *C*, root-mean-square error (RMSE), and *D*, percent model error (RMSE divided by the range of data). Site information and locations are shown in table 2 and figure 32.

data, cumulative z-scores, percentile rank curves, and scatter plots of measured data and simulated values from the ANN and ECFT models are displayed by the MDV (appendix 7). For example, for Crooked Lake near Babson Park (site 49), percentile rank curves indicate that the ANN model underpredicts low altitude water levels (the bottom 25 percent of data), whereas the ECFT model overpredicts low altitude water levels (the bottom 15 percent of data) (appendix 7). In contrast, both the ANN and ECFT models underpredict high altitude water levels by 1 ft or more for the upper 5 percent of data.

Even though the ANN models generally replicate the observed historical data more closely than does the ECFT model, ANN models may not more accurately predict system behavior under a different set of conditions. Each model has a unique set of assumptions and limitations, described herein for the ANN models and by Sepúlveda and others (2012) for the ECFT model. By applying machine learning artificial intelligence techniques, ANN models have a greater mathematical capability to replicate observed data than does the ECFT model. Physics-based models attempt to fit given equations based on prior definition of a conceptual model of the system. The groundwater flow equations used by the ECFT model substantially reduce the degrees of freedom available to explain head and flow variability. These limitations are not imposed in the ANN models. In short, ANNs synthesize the interactions among the model inputs and outputs to best match the observed response, whereas physics-based models prescribe such interactions based on first principles and associated approximations.

Both approaches have potential advantages and disadvantages. Where data are sparse or of poor quality, an ANN model may provide poor predictions or synthesize an incorrect mathematical input-output relation, whereas a physics-based groundwater flow model may provide poor predictions but will be constrained by the principles of mass conservation and the formulation prescribed for relating flow to the driving potential-energy gradient (most commonly a porous media model based on hydraulic head gradients and Darcy's law). Where data are plentiful and of high quality, however, an ANN model may synthesize a mathematical input-output relation that differs from that of a physics-based model in order to more closely replicate the observed data. Under such conditions, a physics-based model may be adversely constrained by the prescribed mathematical relation, resulting in a poorer capability to replicate the observed data. Some of the physics of groundwater flow in karst aquifers may not follow the commonly applied equivalent porous media concept, such as that used in the ECFT model, especially at the site scale. Under those conditions, different and more advanced models—such as dual porosity or conduit flow models (Shoemaker and others, 2008)—may be more accurate (Hill and others, 2008). The ultimate difficulty with these more advanced physics-based models for karst aquifers is that the fine-scale rock lithology and structure, which is what fundamentally controls the travel times and pathways of flowing groundwater, are effectively unmappable over any area of regional water management concern. The effects of such fine-scale features may be manifest in the observed hydrologic response and could be relatively easily simulated using ANN or other empirical modeling techniques. Hydrologic models, whether data-driven or physics-based, often more accurately predict changes in, rather than the absolute magnitude of, a simulated response, such as drawdown in groundwater level caused by increased groundwater withdrawals. Comparisons of model sensitivities and predictions under the same set of conditions would provide additional insight on the similarities and differences between the ANN and ECFT models. Hypothetical scenarios, such as those used for the ANN model sensitivity analyses, as well as more realistic projected future conditions, could be used for comparing the ANN and ECFT models. Combined applications of both physics-based and data-driven models provide another avenue for leveraging the strengths of both types of model. Improvement in the predictive accuracy of a physics-based groundwater model has been demonstrated by using data-driven models, including ANNs, to predict systematic errors in the physics-based model and then combining each model's results in an error-correcting complementary manner (Demissie and others, 2009). More robust models should yield predictions that have been derived over a wide range of conditions and are consistent with known physical principles and data on system behavior from other independent sources.

Summary and Conclusions

In central Florida, variations in both rainfall and groundwater use may affect surface and subsurface water-level and flow conditions, potentially affecting the ability of the hydrologic system to meet both human and environmental needs. Many long-term records of historical hydrologic data for the area are available in the databases of local, State, and Federal agencies, which are well suited for empirical modeling. On the basis of these data, the response of lake water levels, groundwater levels, and spring flows to changing rainfall and groundwater-use conditions over a multidecadal period was analyzed using artificial neural network (ANN) and other data-mining techniques.

The data consist of forcing (meteorological and groundwater use) and response (water level and flow) data from sites in Lake, Orange, Osceola, Polk, and Seminole Counties and the vicinity. Some meteorological data extend back to 1900 and some hydrologic response data extend back to the 1930s, whereas little groundwater-use data are available prior to 1957. A database was created to store all historical hydrologic records, which consist of daily data on rainfall (143 sites), air temperature (21 sites), potential evapotranspiration (21 sites), lake water level (307 sites), groundwater level (470 sites), and spring flow (22 sites). Reported or estimated monthly groundwater-use data were compiled for the categories public supply, citrus irrigation, non-citrus agricultural irrigation, golf course irrigation, phosphate mining, and drainage well recharge.

Historical data were evaluated using descriptive statistics, cluster analysis, and other exploratory analysis techniques to assess their suitability for more intensive data-mining analysis. A linear trend analysis was performed on the data for the sites in the hydrologic database with at least 10 measurements and a period of record of at least 10 years. Daily rainfall data from only 21 rain gages operated by the National Oceanic and Atmospheric Administration (NOAA) with 95-year or longer periods of record were analyzed owing to the presence of accumulated data (rainfall that was accumulated over multiple days and reported as a single value) in the record of other rain gages. For this initial exploratory analysis, no attempt was made to identify cyclical or nonmonotonic trends or to establish statistically significant upward or downward trends. Meteorological data indicate 67 percent of sites exhibited upward trends in air temperature over at least a 45-year period of record, whereas 76 percent exhibited downward trends in rainfall over at least a 95-year period of record. Analyses of hydrologic response data, which have varied periods of record ranging in length from 10 to 79 years, indicate that the water levels of lakes were about evenly split between upward and downward trends, whereas water levels in 69 percent of wells and flows in 68 percent of springs exhibited downward trends. For lake water levels, both closed-basin and flowthrough lakes in ridge areas generally have greater upward or downward trends in water levels than do lakes in non-ridge areas, suggesting that lakes in ridge areas are more susceptible to long-term changes in water level as a result of greater hydraulic connection with the Floridan aquifer system.

For sites with periods of record exceeding 30 years and relatively few missing data, a time-series cluster analysis was performed for each type of hydrologic response data (water levels at 121 lakes and 46 wells and flow at 9 springs) to identify dynamic similarities and differences among sites of the same type based on a correlation matrix (Pearson product-moment). For lakes, there is no apparent geographic association among the resulting six cluster analysis groups. Compared to all groups combined, however, two groups have proportionately more closed-basin lakes and two other groups have proportionately more flowthrough lakes, suggesting differences in the dynamic behavior of historical water levels in these different types of lakes. Cluster analysis results indicate that wells generally are associated geographically. Spring groups are loosely associated geographically and by magnitude of flow. Time-series cluster analyses also were performed on 1942–2008 meteorological data, consisting of rainfall at 18 NOAA gages and potential evapotranspiration computed from air temperature data at 21 NOAA gages. Results indicate four groups of rain gages with north-south and east-west geographic associations; with the exception of one gage, all gages belonging to the two groups in the western half of the study area had downward trends in rainfall. Substantial intragroup differences indicate that rainfall is highly variable in space and time. Cluster analysis of potential evapotranspiration data indicates three groups of relatively similar dynamic behavior with a generally east-west geographic association.

To enable viewing of general spatial trends in groundwater use over time, a spreadsheet application called the Groundwater-Use Data Viewer was developed that aggregates the monthly data using a grid cell size of approximately 20 × 20 kilometers. Spatial patterns and long-term trends in groundwater use are evident in both monthly and annual data for all water-use categories. Total groundwater use in the study area (excluding 1950–57 due to missing data) increased from about 250 million gallons per day (Mgal/d) in 1958 to about 590 Mgal/d in 1980 and remained relatively stable during 1981–2008, with a minimum of 559 Mgal/d in 1994 and a maximum of 773 Mgal/d in 2000. The change in trend in the early 1980s and the following period of relatively slight trend is attributable to the concomitant effects of increasing public-supply withdrawals and decreasing use of water by the phosphate industry and for citrus and non-citrus agricultural irrigation.

On the basis of the exploratory analyses of the historical data, empirical lake water-level, groundwater-level, and spring-flow models were developed for 51 sites by using ANN models and other data-mining techniques, such as signal (time series) processing methods including filtering and signal decomposition. The sites consist of 22 lakes, 23 wells, and 6 springs, with data records beginning as early as 1942 and extending to 2008. Input time series consisting of decomposed daily rainfall (1942 to 2008) and monthly total groundwater use (1957 to 2008) resulted in a hybrid signal-decomposition ANN model. Moving window averages (MWAs) having successively increasing window sizes, and differences between them called moving window differences (MWDs), were used to represent various frequencies and frequency-band components of the rainfall and groundwater-use forces that modulate water-level or spring-flow behavioral dynamics. Daily rainfall data at the 18 NOAA rain gages were used to compute MWAs for the standard rainfall (arithmetic average by time step of all gages) as well as normalized rainfall (individual gage rainfall minus the standard rainfall) at each gage using the following window sizes: 30; 90; 270; 630; 1,440; and 2,250 days. For the standard rainfall, MWDs also were computed. Likewise, monthly groundwater-use data were used to compute MWAs and MWDs for total groundwater-use flows for the study area using the following window sizes: 3, 6, 12, 24, 48, and 96 months.

To provide insight into the input-output system dynamics and to guide development of the ANN models, linear associations among rainfall, groundwater use, and hydrologic response (lake water level, groundwater level, and spring flow) were computed using the Pearson product-moment correlation coefficient (R). Results for all sites indicate that the strongest correlations between rainfall signals (270-, 630-, and 1,440-day MWAs) and hydrologic response averaged 0.66 and average correlations for all rainfall signals averaged 0.40. In contrast, the strongest correlations between groundwater-use signals (3-, 6-, 12-, 24-, 48-, and 96-month MWAs) and hydrologic response averaged –0.48 and average correlations for all groundwater-use signals averaged –0.31 for all sites.

Correlations of hydrologic response with groundwater use generally were negative because water levels and flows will decrease in response to an increase in groundwater withdrawals. The relatively strong negative correlations suggest that groundwater use is an important control on hydrologic response, although correlations of hydrologic response with groundwater use generally are weaker than those with rainfall. Additionally, long-term downward trends in rainfall were prevalent in the western half of the study area, where most of the areas of high groundwater recharge occur, and are potentially associated with downward trends in groundwater levels and spring flows. Correlations between hydrologic response and rainfall were fairly high for both nearby and distant rain gages, suggesting an "interconnectedness" of the hydrologic system in which the effects of rainfall forcing on the hydrologic system could propagate over long distances. Correlations between rainfall signals (30-, 90-, 270-, 630-, 1,440-, and 2,250-day MWAs) at the 18 NOAA gages and groundwater-use signals (3-, 6-, 12-, 24-, 48-, and 96-month MWAs) tended to be negative, where the strongest correlations averaged –0.46 and average correlations for all rainfall and groundwater-use signal combinations averaged –0.14 for all NOAA rain gage sites. These negative correlations are likely attributable to rainfall control of certain types of water use (for example, irrigation) such that groundwater withdrawals increase in response to a decrease in rainfall. Such cross correlation between rainfall and groundwater use may affect the relative explanatory strength of groundwater use or rainfall for modeling hydrologic response.

The development of the hybrid signal-decomposition ANN models for the 51 sites in the study was undertaken in two phases, resulting in a two-stage model architecture. The first phase was to train an ANN for each site to simulate the observed hydrograph using decomposed rainfall signals as inputs. The second phase was to train a second ANN for each site to simulate the residual error time series from the first model using decomposed groundwater-use signals as inputs and, at a few sites, using the predicted water level or flow from the first model as an additional input. The final simulated water level or flow is the sum of the predictions of both ANN models for the site. Final models were limited to the 1965 to 2008 period to accommodate the limited period of record of groundwater-use data and the computation of MWAs and MWDs.

Training of the ANN models was performed iteratively to select the MWAs and MWDs that were the best predictors for each output hydrograph, while culling inputs that were collinear or had input-output sensitivities at odds with principles of mass conservation. The final models explained much of the variability in observed data, with models for 43 of the 51 sites having coefficient of determination (R^2) values exceeding 0.6, and the models matched the magnitude of the observed data reasonably well, such that models for 32 of the 51 sites had percent model error (PME) less than 10 percent (PME is root-mean-square error divided by the range of the measured data). Generally, groundwater-level models, with R^2 averaging 0.774 and PME averaging 8.90 percent, performed slightly better than lake water-level models, with R^2 averaging 0.718 and PME averaging 9.96 percent. Spring-flow models had the poorest model fit, with R^2 averaging 0.546 and PME averaging 13.14 percent.

The Central Florida Artificial Neural Network Decision Support System (CFANN DSS) was developed to simulate hydrologic response to historical or user-specified rainfall and groundwater use from 1965 through 2008. The CFANN DSS integrates historical databases and the 102 site-specific ANN models, model controls, and model output into a spreadsheet application with a graphical user interface that allows the user to simulate scenarios of interest. The CFANN DSS was used to run four scenarios, representing hypothetical changes in historical rainfall and groundwater use, to determine changes in lake water level, groundwater level, and spring flow simulated by the ANN models. The scenarios consist of using time series of rainfall and groundwater use at 70 and 130 percent of historical values. Lakes and wells generally show a larger response to a 30-percent change in rainfall than a 30-percent change in groundwater use (for both maximum and minimum changes in water level or flow), whereas springs generally show a mixed response, with some springs showing a larger response to rainfall and some springs showing a larger response to groundwater use.

From a long-term, steady-state, water-balance perspective, increases in groundwater withdrawals from an aquifer system are balanced by increases in aquifer recharge or decreases in aquifer discharge, with reductions in natural discharge (for example, springs) typically most important. In central Florida, spring flow constitutes a substantial outflow from the aquifer system. Accordingly, increases in groundwater withdrawals over time are likely to eventually result in decreases in spring flows as the hydrologic system moves toward a long-term balance; such behavior would be magnified under prolonged low-recharge conditions. The sensitivity of spring-flow ANN models to the hypothetical rainfall and groundwater-use scenarios are at least qualitatively consistent with this equilibrium-seeking process. Compared to the sensitivity of the flow of springs, the lower sensitivities of the water levels of most lakes and wells to the same hypothetical rainfall and groundwater-use scenarios are likely attributable in part to the simple hydraulic principle that a small change in water level can cause a large change in flow rate if the water-level change results in a large percentage change in hydraulic gradient.

Overall, the data-mining analyses described herein indicate that the Floridan aquifer system in central Florida is a highly conductive, dynamic, open system that is strongly influenced by external forcing. The most important external forcing appears to be rainfall, which explains much of the multiyear cyclic variability and long-term downward trends observed in lake water levels, groundwater levels, and spring flows. For most sites, groundwater use explains less of the observed variability in water levels and flows than rainfall. Relative groundwater-use effects are greater, however, during

droughts, and long-term trends consistent with historical groundwater-use patterns were identified. The sensitivity of the hydrologic system to rainfall is consistent with the hydrogeology of central Florida where the Floridan aquifer system is thinly confined; the overlying confining unit (intermediate confining unit/intermediate aquifer system) is less than 130 ft thick throughout most of the study area (Sepúlveda and others, 2012). The well-drained karst terrain in central Florida largely coincides with these areas of thin confinement and facilitates relatively rapid transmission of infiltrating water from rainfall to the water table.

Previous physics-based modeling and independent water-budget analyses indicated that groundwater recharge from rainfall and artificial recharge is the largest inflow to the groundwater system in central Florida. The areally distributed nature of rainfall, as opposed to the site-specific nature of groundwater use, and the generally high transmissivity and low storativity properties of the Floridan aquifer system contribute to the prevalence of water-level and flow patterns that mimic rainfall patterns. The effects of development activities, such as groundwater withdrawals, are superimposed on the natural rainfall-driven system, which may affect both the short-term and the long-term water balance of the hydrologic system.

The dynamic nature of the groundwater system in central Florida is characterized by constant adjustments to changing hydrologic inflows and outflows. An aquifer responds to such adjustments through a combination of storage gains and losses that induce positive feedback from system boundaries—when inflows or reduced outflows from system boundaries match groundwater withdrawals, a hydrologic balance is achieved (Balleau, 2013). The effects of such behavior are reflected by changing water levels and flows. A hydrologic system with multiple forcings and positive feedbacks may have multiple stable states, that is, the system may not return to the same water-level and flow conditions (the same point in state space) after a transient disturbance (Peterson and others, 2009). This hysteretic behavior depends on the magnitude and interactions of the system inputs and outputs, and contributes to dynamic behavior of the system. In general, the data-mining analyses demonstrate that the hydrologic system in central Florida is affected by groundwater use differently during wet periods, when little or no system storage is available (high water levels), than during dry periods, when there is excess system storage (lower water levels). Thus, by driving the overall behavior of the system, rainfall indirectly influences the degree to which groundwater use will drive persistent trends in water levels and flows, with groundwater-use effects more prevalent during periods of low water levels and spring flows caused by low rainfall and less prevalent during periods of high water levels and spring flows caused by high rainfall. Differences in the magnitudes of rainfall and groundwater use during wet and dry periods also are important determinants of hydrologic response.

The data-mining analyses provided insight on salient processes influencing system behavior as gleaned from a long period of historical data representing a wide variety of hydroclimatic conditions. This insight may not be easily acquired with physics-based models, which commonly are developed by using relatively short calibration periods. The analysis methods inherent in data mining are particularly well suited for karst aquifer systems having good hydraulic communication between surface and subsurface environments, such as the Floridan aquifer system in central Florida. In contrast, for deeply buried or well-confined aquifer systems, in which surface forcing signals are substantially attenuated and system behavior evolves over a longer time period, development of empirical models may be more difficult owing to the often sparse or nonexistent data for 50- to 100-year or greater periods. An important implication of the data-mining analyses is that rainfall variability at subannual to multidecadal timescales must be considered in combination with groundwater use to provide robust system-response predictions that benefit sustainable resource management in an open karst aquifer system.

Assessment of vulnerability of the hydrologic system in central Florida to environmental change, whether natural or human derived, is a potential application of the forcing-response behavior learned through the data-mining analyses. The CFANN DSS provides an easy-to-use tool for sensitivity analysis that could facilitate vulnerability assessments. Even though considerable uncertainty can exist in quantifying the forcing-response behavior of a system, assessment of the relative vulnerability of a system to forcings of concern represented by the CFANN DSS under different scenarios may prove beneficial to water-resource management decision making.

Determining the effects of rainfall and groundwater use on hydrologic response in central Florida using empirical modeling is limited by the confounding effects of correlation between rainfall and groundwater use, quality and completeness of the historical databases, and spatial variations in groundwater use. The data-mining analyses presented herein indicate that the available historical data alone do not contain sufficient information to allow definitive quantification of cause-and-effect relations of rainfall and groundwater-use on hydrologic response. As a result, the final lake water-level, groundwater-level, and spring-flow models incorporated in the CFANN DSS should not be used alone to quantify the individual effects of rainfall and groundwater use for water-supply planning or management purposes.

Results of the ANN models were compared with those of a physics-based groundwater flow model. The ANN models were able to more closely match the measured data for most sites compared to the physics-based model. The greatest differences between the ANN and physics-based models generally are present in water levels for the wells and to a lesser degree the lakes; both types of models generally had similar accuracies for flows at springs. Both approaches have potential advantages and disadvantages. Where data are sparse or of poor quality, an ANN model may provide poor predictions or synthesize an incorrect mathematical input-output relation, whereas a physics-based groundwater flow model may provide

poor predictions but will be constrained by the principles of mass conservation and Darcy's law. The knowledge gained from data-driven modeling and the results from physics-based modeling, when compared and used in combination, can yield a more comprehensive assessment and a more robust understanding of the hydrologic system than either of the approaches used separately.

References Cited

Abarbanel, H.D.I., 1996, Analysis of observed chaotic data: New York, Springer-Verlag, 272 p.

Adamski, J.C., and German, E.R., 2004, Hydrogeology and quality of groundwater in Orange County, Florida: U.S. Geological Survey Water-Resources Investigations Report 03–4257, 113 p.

Adger, W.N., 2006, Vulnerability: Global Environmental Change, v. 16, p. 268–281.

Aly, Alaa, 2008, An assessment of interpolation methods for estimating missing daily precipitation records for rain gauges in central and south Florida: Niwot, Colo., INTERA Inc., 111 p.

Amari, S., Murata, N., Müller, K.-R., Finke, M., and Yang, H., 1996, Statistical theory of overtraining—Is cross-validation asymptotically effective?: *in* Touretsky, D.S., Mozer, M.C., and Hasselmo, M.E., eds., Advances in Neural Information Processing Systems 8, Proceedings of the 1995 Conference: Cambridge, Mass., The MIT Press, p. 176–182.

Balleau, W.P., 2013, The policy of "pumping the recharge" is out of control: Eos, v. 94, no. 1, p. 4–5.

Basso, Ron, 2003, Predicted change in hydrologic conditions along the Upper Peace River Basin due to a reduction in ground-water withdrawals: Brooksville, Southwest Florida Water Management District, 53 p.

Berry, Leonard, Bloetscher, Frederick, Hammer, N.H., Koch-Rose, Marguerite, Mitsova-Boneva, Diana, Restrepo, Jorge, Root, Tara, and Teegavarapu, Ramesh, 2011, Florida water management and adaptation in the face of climate change: Florida Climate Change Task Force, 52 p. (Also available at *http://floridaclimate.org/whitepapers/*.)

Bishop, C.M., 1995, Neural networks for pattern recognition: Oxford, Oxford University Press, 482 p.

Bradner, L.A., 1996, Estimation of recharge through selected drainage wells and the potential effects from well closure, Orange County, Florida: U.S. Geological Survey Open-File Report 96–316, 30 p.

Bredehoeft, John, 1997, Safe yield and the water budget myth: Ground Water, v. 35, no. 6, p. 929.

Chow, V.T., Maidment, D.R., and Mays, L.W., 1988, Applied hydrology: New York, McGraw Hill, 243 p.

Conrads, P.A., Feaster, T.D., and Harrelson, L.G., 2008, The effects of the Saluda dam on the surface-water and ground-water hydrology of the Congaree National Park flood plain, South Carolina: U.S. Geological Survey Scientific Investigations Report 2008–5170, 58 p.

Conrads, P.A., and Roehl, E.A., Jr., 2006, Estimating water depths using artificial neural networks, *in* Gourbesville, Philippe, Cunge, Jean, Guinot, Vincent, and Liong, Shie-Yui, eds., Proceedings of the 7th International Conference on Hydroinformatics 2006, September 4–8, 2006, Nice, France.

Conrads, P.A., and Roehl, E.A., Jr., 2007, Analysis of salinity intrusion in the Waccamaw River and the Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina, 1995–2002: U.S. Geological Survey Scientific Investigations Report 2007–5110, 41 p., 2 apps.

Conrads, P.A., and Roehl, E.A., Jr., 2010, Analysis and simulation of water-level, specific conductance, and total phosphorus dynamics of the Loxahatchee National Wildlife Refuge, Florida, 1995–2006: U.S. Geological Survey Scientific Investigations Report 2010–5244, 64 p.

Conrads, P.A., Roehl, E.A., Jr., Daamen, R.C., Chapelle, F.H., Lowery, M.A., and Mundry, U.H., 2007, Visualization and time-series analysis of ground-water data for C-Area, Savannah River Site, South Carolina, 1984–2004: U.S. Geological Survey Scientific Investigations Report 2007–5140, 33 p.

Conrads, P.A., Roehl, E.A., Jr., Daamen, R.C., and Kitchens, W.M., 2006, Simulation of water levels and salinity in the rivers and tidal marshes in the vicinity of the Savannah National Wildlife Refuge, Coastal South Carolina and Georgia: U.S. Geological Survey Scientific Investigations Report 2006–5187, 134 p.

Conrads, P.A., Roehl, E.A., Jr., and Davie, S.R., 2011, Simulation of specific conductance and chloride concentration in Abercorn Creek, Georgia, 2000–2009: U.S. Geological Survey Scientific Investigations Report 2011–5074, 46 p.

Conrads, P.A., Roehl, E.A., Jr., and Martello, W.P., 2003, Development of an empirical model of a complex, tidally affected river using artificial neural networks: Water Environment Federation, National TMDL Science and Policy 2003 Specialty Conference, November 2003, Chicago, Ill.

Coppola, Emery, Jr., Poulton, Mary, Charles, Emmanuel, Dustman, John, and Szidarovszky, Ferenc, 2003, Application of artificial neural networks to complex groundwater management problems: Natural Resources Research, v. 12, no. 4, p. 303–320.

Coppola, Emery, Jr., Szidarovszky, Ferenc, Poulton, Mary, Charles, Emmanuel, 2003, Artificial neural network approach for predicting transient water levels in a multilayered groundwater system under variable state, pumping, and climate conditions: Journal of Hydrologic Engineering, v. 8, no. 6, p. 348–360.

Daamen, R.C., Roehl, E.A., Jr., Conrads, P.A., and Kitchens, W.M., 2006, Integrating 3D hydrodynamic transport and ecological plant models of the Savannah River estuary using artificial neural network models, *in* Gourbesville, Philippe, Cunge, Jean, Guinot, Vincent, and Liong, Shie-Yui, eds., Proceedings of the 7th International Conference on Hydroinformatics 2006, September 4–8, 2006, Nice, France.

Demissie, Y.K., Valocchi, A.J., Minsker, B.S., and Bailey, B.A., 2009, Integrating a calibrated groundwater flow model with error-correcting data-driven models to improve predictions: Journal of Hydrology, v. 364, p. 257–271.

Devine, T.W., Roehl, E.A., and Busby, J.B., 2003, Virtual sensors—Cost effective monitoring: Air and Waste Management Association Annual Conference, June 2003.

Dutta, S., Wierenga, B., and Dalebout, A., 1997, Case-based reasoning systems: from automation to decision-aiding and stimulation: IEEE Transactions on Knowledge and Data Engineering, v. 9, no. 6, p. 911–922.

Environmental Simulations, Inc., 2007, Refinement of the District Wide Regulation Model for Southwest Florida Water Management District: Reinholds, Pa., Environmental Simulations, Inc., 47 p.

Faybishenko, B.A., 1995, Hydraulic behavior of quasi-saturated soils in the presence of entrapped air— Laboratory experiments: Water Resources Research, v. 31, no. 10, p. 2421–2435.

Fayer, M.J., and Hillel, D., 1986, Air encapsulation—II. Profile water storage and shallow water table fluctuations: Soil Science Society of America Journal, v. 50, no. 3, p. 572–577.

Florida Office of Economic and Demographic Research, 2012, Historical data—Florida County population census counts—1900 to 2010, accessed June 2012 at *http://edr.state.fl.us/Content/population-demographics/2010-census/data/index.cfm*.

Fry, J.A., Xian, George, Jin, Suming, Dewitz, J.A., Homer, C.G., Yang, Limin, Barnes, C.A., Herold, N.D., and Wickham, J.D., 2011, Completion of the 2006 National Land Cover Database for the conterminous United States: Photogrammetric Engineering & Remote Sensing, v. 77, no. 9, p. 858–864.

Füssel, H.-M., 2007, Vulnerability—A generally applicable conceptual framework for climate change research: Global Environmental Change, v. 17, p. 155–167.

German, E.R., and Adamski, J.C., 2005, Hydrogeology and water quality of lakes and streams in Orange County, Florida: U.S. Geological Survey Scientific Investigations Report 2005–5052, 103 p.

Harbaugh, A.W., 2005, MODFLOW-2005, The U.S. Geological Survey modular ground-water model—The ground-water flow process: U.S. Geological Survey Techniques and Methods, book 6, chap. A16, 196 p.

Harbaugh, A.W., Banta, E.R., Hill, M.C., and McDonald, M.G., 2000, MODFLOW-2000, The U.S. Geological Survey modular ground-water flow model—User guide to modularization concepts and the ground-water flow process: U.S. Geological Survey Open-File Report 00–92, 121 p.

Harbaugh, A.W., and McDonald, M.G., 1996, User's documentation for MODFLOW-96, An update to the U.S. Geological Survey modular finite-difference ground-water flow model: U.S. Geological Survey Open-File Report 96–485, 56 p.

Hargreaves, G.H., and Samani, Z.A., 1985, Reference crop evapotranspiration from temperature: Applied Engineering in Agriculture, v. 1, no. 2, p. 96–99.

Hartigan, J.A., and Wong, M.A., 1979, Algorithm AS 136—A K-means clustering algorithm: Journal of the Royal Statistical Society, Series C (Applied Statistics), v. 28, no. 1, p. 100–108.

Hartman and Associates, Inc., 2003, Drainage well inventory for Orange, Seminole, and Lake Counties, Contract Number C-12355: Prepared for South Florida Water Management District, West Palm Beach, Fla., 58 p.

Hastie, Trevor, Tibshirani, Robert, and Friedman, Jerome, 2001, The elements of statistical learning—Data mining, inference, and prediction: New York, Springer Science+Business Media LLC, 533 p.

Hill, M.E., Martin, Angel, and Stewart, M.T., 2008, Performance evaluation of the MODFLOW-2005 conduit flow process applied to a karst aquifer underlying west-central Florida, *in* Kuniansky, E.L., ed., U.S. Geological Survey Karst Interest Group Proceedings, Bowling Green, Ky., May 27–29, 2008: U.S. Geological Survey Scientific Investigations Report 2008–5023, 142 p.

Hinton, G.E., 1992, How neural networks learn from experience: Scientific American, September 1992, p. 145–151.

Intera, 2007a, Statistical modeling of spring discharge at Apopka and Bugg Springs in Lake County, Florida: Palatka, Fla., St. Johns River Water Management District Special Publication SJ2007–SP8, 94 p.

Intera, 2007b, Statistical modeling of spring discharge at Rock and Wekiva Springs in Orange County, Florida: Palatka, Fla., St. Johns River Water Management District Special Publication SJ2007–SP12, 85 p.

Jensen, B.A., 1994, Expert systems—Neural networks, instrument engineers' handbook (3d ed.): Radnor, Pa., Chilton, p. 48–54.

Kim, T.-W., and Valdés, J.B., 2003, Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks: Journal of Hydrologic Engineering, v. 8, no. 6, p. 319–328.

Kimrey, J.O., and Fayard, L.D., 1984, Geohydrologic reconnaissance of drainage wells in Florida: U.S. Geological Survey Water-Resources Investigations Report 84–4021, 67 p.

Kingston, G.B., Maier, H.R., and Lambert, M.F., 2005, Calibration and validation of neural networks to ensure physically plausible hydrological modeling: Journal of Hydrology, v. 314, p. 158–176.

Kişi, Ösgür, 2009, Neural networks and wavelet conjunction model for intermittent streamflow forecasting: Journal of Hydrologic Engineering, v. 14, no. 8, p. 773–782.

Knowles, Leel, Jr., O'Reilly, A.M., and Adamski, J.C., 2002, Hydrogeology and simulated effects of ground-water withdrawals from the Floridan aquifer system in Lake County and in the Ocala National Forest and vicinity, north-central Florida: U.S. Geological Survey Water-Resources Investigations Report 02–4207, 140 p.

Kugiumtzis, Dimitris, Lillekjendlie, Bjørn, and Christophersen, Nils, 1994, Chaotic time series—Part I. Estimation of some invariant properties in state-space: Modeling, Identification, and Control, v. 15, no. 4, p. 205–224.

Kuniansky, E.L., Bellino, J.C., and Dixon, J.F., 2012, Transmissivity of the Upper Floridan aquifer in Florida and parts of Georgia, South Carolina, and Alabama: U.S. Geological Survey Scientific Investigations Map 3204, 1 sheet, scale 1:100,000.

Legates, D.R., and McCabe, G.J., Jr., 1999, Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation: Water Resources Research, v. 35, no. 1, p. 233–241.

Lillekjendlie, Bjørn, Kugiumtzis, Dimitris, and Christophersen, Nils, 1994, Chaotic time series—Part II. System identification and prediction: Modeling, Identification, and Control, v. 15, no. 4, p. 225–243.

Luers, A.L., 2005, The surface of vulnerability—An analytical framework for examining environmental change: Global Environmental Change, v. 15, p. 214–223.

Luers, A.L., Lobell, D.B., Sklar, L.S., Addams, C.L., and Matson, P.A., 2003, A method for quantifying vulnerability, applied to the Yaqui Valley, Mexico: Global Environmental Change, v. 13, p. 255–267.

Marella, R.L., 1995, Water-use data by category, county, and water management district, 1950–90: U.S. Geological Survey Open-File Report 94–521, 114 p.

Marella, R.L., 2004, Water withdrawals, use, discharge, and trends in Florida, 2000: U.S. Geological Survey Scientific Investigations Report 2004–5151, 136 p.

Marella, R.L., 2009, Water withdrawals, use, and trends in Florida, 2005: U.S. Geological Survey Scientific Investigations Report 2009–5125, 49 p.

McCuen, R.H., Knight, Zachary, and Cutter, A.G., 2006, Evaluation of the Nash-Sutcliffe Efficiency Index: Journal of Hydrologic Engineers, v. 11, no. 6, p. 597–602.

McGurk, Brian, and Presley, P.F., 2002, Simulation of the effects of ground-water withdrawals on the Floridan aquifer system in east-central Florida—Model expansion and revision: Palatka, Fla., St. Johns River Water Management District Technical Publication SJ2002–3, 196 p.

McLeod, Jay, and Munch, Doug, 2012, Methodology for estimating historical agricultural groundwater use in the Central Florida Coordination Area (CFCA) 1957–2010: Palatka, Fla., St. Johns River Water Management District Technical Publication SJ2012–2, 126 p.

Mecikalski, J.R., Sumner, D.M., Jacobs, J.M., Pathak, C.S., Paech, S.J., and Douglas, E.M., 2011, Use of visible geostationary operational meteorological satellite imagery in mapping reference and potential evapotranspiration over Florida, *in* Labedzki, Leszek, ed., Evapotranspiration: Vienna, Austria, InTech Publishers, 446 p.

Miller, J.A., 1986, Hydrogeologic framework of the Floridan aquifer system in Florida and parts of Georgia, Alabama, and South Carolina: U.S. Geological Survey Professional Paper 1403–B, 91 p.

Motter, A.E., and Campbell, D.K., 2013, Chaos at fifty: Physics Today, v. 66, no. 5, p. 27–33.

Munch, Douglas, 2014, Groundwater withdrawals for public supply use in the Central Florida Coordination Area (CFCA) and vicinity, 1945–2010: Palatka, Fla., St. Johns River Water Management District Professional Paper SJ2014-PP1, 20 p., 2 appendixes.

Murch, R.R., and Tara, Patrick, 2010, Central Florida Coordination Area—Statistical analysis: Prepared for the St. Johns River Water Management District, Palatka, Fla., and the Southwest Florida Water Management District, Brooksville, Fla., 106 p., 2 apps.

Murray, L.C., Jr., 2010, Relations between groundwater levels and anthropogenic and meteorological stressors at selected sites in east-central Florida, 1995–2007: U.S. Geological Survey Scientific Investigations Report 2010–5046, 31 p.

Nash, J.E., and Sutcliffe, J.V., 1970, River flow forecasting through conceptual models, Part 1—A discussion of principles: Journal of Hydrology, v. 10, no. 3, p. 282–290.

National Oceanic and Atmospheric Administration, 2011, NOAA's 1981–2010 Climate normal, accessed November 2011 at *http://www.ncdc.noaa.gov/oa/climate/normals/usnormals.html.*

Obeysekera, J., Park, J., Irizarry-Ortiz, M., Trimble, P., Barnes, J., VanArman, J., Said, W., and Gadzinski, E., 2011, Past and projected trends in climate and sea level for South Florida: Hydrologic and Environmental Systems Modeling Technical Report—July 5, 2011, Interdepartmental Climate Change Group, South Florida Water Management District, West Palm Beach, Fla., 149 p.

O'Reilly, A.M., 1998, Hydrogeology and simulation of the effects of reclaimed-water application in west Orange and southeast Lake Counties, Florida: U.S. Geological Survey Water-Resources Investigations Report 97–4199, 91 p.

O'Reilly, A.M., 2007. Effects of the temporal variability of evapotranspiration on hydrologic simulation in central Florida: U.S. Geological Survey Scientific Investigations Report 2007–5100, 36 p.

O'Reilly, A.M., Spechler, R.S., and McGurk, B.E., 2002, Hydrogeology and water-quality characteristics of the Lower Floridan aquifer in east-central Florida: U.S. Geological Survey Water-Resources Investigations Report 02–4193, 60 p.

Osburn, William, Toth, David, and Boniol, Don, 2006, Springs of the St. Johns River Water Management District: Palatka, Fla., St. Johns River Water Management District Technical Publication SJ2006–3, superseded by online version, accessed May 2013 at *http://www.sjrwmd.com/springs/*

Özger, Mehmet, Mishra, A.K., and Singh, V.P., 2012, Long lead time drought forecasting using a wavelet and fuzzy logic combination model—A case study in Texas: Journal of Hydrometeorology, v. 13, no. 1, p. 284–297.

Peterson, T.J., Argent, R.M., Western, A.W., and Chiew, F.H.S., 2009, Multiple stable states in hydrological models—An ecohydrological investigation: Water Resources Research, v. 45, W03406, 17 p.

Pielke, R.A., Sr., Walko, R.L., Steyaert, L.T., Vidale, P.L., Liston, G.E., Lyons, W.A., and Chase, T.N., 1999, The influence of anthropogenic landscape changes on weather in south Florida: Monthly Weather Review, v. 127, p. 1663–1673.

Renken, R.A., Dixon, Joann, Koehmstedt, John, Ishman, Scott, Lietz, A.C., Marella, R.L., Telis, Pamela, Rodgers, Jeff, and Memberg, Steven, 2005, Impact of anthropogenic development on coastal ground-water hydrology in southeastern Florida, 1900–2000: U.S. Geological Survey Circular 1275, 77 p.

Risley, J.C., Roehl, E.A., Jr., and Conrads, P.A., 2002, Estimating water temperatures in small streams in western Oregon using neural network models: U.S. Geological Survey Water-Resources Investigations Report 02–4218, 59 p.

Robertson, A.F., Mills, L.R., and Parsons, D.C., 1978, Ground water withdrawn for municipal, industrial, and irrigation use in the upper Peace and Alafia River basins, west-central Florida, 1970–1974: U.S. Geological Survey Open-File Report 78–29, 59 p.

Roehl, E.A., Jr., and Conrads, P.A., 1999, Near real-time control for matching wastewater discharges to the assimilative capacity of a complex, tidally affected river basin: 1999 South Carolina Environmental Conference, March 15–16, 1999, Myrtle Beach, S.C.

Roehl, E.A., Jr., and Conrads, P.A., 2006, Hydrologic modeling using multivariate state space reconstruction, *in* Gourbesville, Philippe, Cunge, Jean, Guinot, Vincent, and Liong, Shie-Yui, eds., Proceedings of the 7th International Conference on Hydroinformatics 2006, September 4–8, 2006, Nice, France, p. 765–772.

Roehl, E.A., Jr., Conrads, P.A., and Cook, J.B., 2003, Discussion of using complex permittivity and artificial neural networks for contaminant prediction: Journal of Environmental Engineering, November 2003, p. 1069–1071.

Roehl, E.A., Jr., Conrads, P.A., and Daamen, R.C., 2006, Features of advanced decision support systems for environmental studies, management, and regulation, *in* Voinov, A., Jakeman, A.J., and Rizzoli, A.E., eds., Proceedings of the iEMSs Third Biennial Meeting—Summit on Environmental Modelling and Software, International Environmental Modelling and Software Society, July 2006, Burlington, Vt.

Roehl, E.A., Jr., Risley, J.C., Stewart, Jana, and Mitro, Matthew, 2006, Numerically optimized empirical modeling of highly dynamic, spatially expansive, and behaviorally heterogeneous hydrologic systems—Part 1, *in* Voinov, A., Jakeman, A.J., and Rizzoli, A.E., eds., Proceedings of the iEMSs Third Biennial Meeting—Summit on Environmental Modelling and Software, International Environmental Modelling and Software Society, July 2006, Burlington, Vt.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J., 1986, Learning internal representations by error propagation, Parallel Distributed Processing—Explorations in the microstructure of cognition, Volume 1: Cambridge, Mass., The MIT Press, p. 318–362.

Sahoo, Sasmita, and Jha, M.K., 2013, Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment: Hydrogeology Journal, v. 21, p. 1865–1887.

Schiner, G.R., 1993, Geohydrology of Osceola County, Florida: U.S. Geological Survey Water-Resources Investigations Report 92–4076, 68 p.

Sepúlveda, Nicasio, 2002, Simulation of ground-water flow in the intermediate and Floridan aquifer systems in peninsular Florida: U.S. Geological Survey Water-Resources Investigations Report 02–4009, 130 p.

Sepúlveda, Nicasio, 2009, Analysis of methods to estimate springs flows in a karst aquifer: Ground Water, v. 47, no. 3, p. 337–349.

Sepúlveda, Nicasio, Tiedeman, C.R., O'Reilly, A.M., Davis, J.B., and Burger, Patrick, 2012, Water budget and groundwater flow in the surficial and Floridan aquifer systems in east-central Florida: U.S. Geological Survey Scientific Investigations Report 2012–5161, 214 p.

Shoemaker, W.B., Kuniansky, E.L., Birk, S., Bauer, S., and Swain, E.D., 2008, Documentation of a Conduit Flow Process (CFP) for MODFLOW-2005: U.S. Geological Survey Techniques and Methods, book 6, chap. A24, 50 p.

Smith, S.K., and Rayer, Stefan, 2012, Projections of Florida population by county, 2011–2040: Gainesville, Fla., University of Florida, Bureau of Economic and Business Research Bulletin 162, 8 p.

Spechler, R.M., and Halford, K.J., 2001, Hydrogeology, water quality, and simulated effects of ground-water withdrawals from the Floridan aquifer system, Seminole County and vicinity, Florida: U.S. Geological Survey Water-Resources Investigations Report 01–4182, 116 p.

Spechler, R.M., and Kroening, S.E., 2007, Hydrology of Polk County, Florida: U.S. Geological Survey Scientific Investigations Report 2006–5320, 114 p.

Stewart, Jana, Mitro, Matthew, Roehl, E.A., Jr., and Risley, J.C., 2006, Numerically optimized empirical modeling of highly dynamic, spatially expansive, and behaviorally heterogeneous hydrologic systems—Part 2, *in* Gourbesville, Philippe, Cunge, Jean, Guinot, Vincent, and Liong, Shie-Yui, eds., Proceedings of the 7th International Conference on Hydroinformatics 2006, September 4–8, 2006, Nice, France.

Sumner, D.M., 2006, Adequacy of selected evapotranspiration approximations for hydrologic simulation: Journal of the American Water Resources Association, v. 42, no. 3, p. 699–711.

Trichakis, I.C., Nikolos, I.K., and Karatzas, G.P., 2011, Artificial neural network (ANN) based modeling for karstic groundwater level simulation: Water Resources Management, v. 25, no. 4, p. 1143–1152.

Trimble, P., Obeysekera, J., Cadavid, L., and Santee, E.R., 2006, Applications of climate outlooks for water management in south Florida, *in* Garbrecht, J.D., and Piechota, T.C., eds., Climate variations, climate change, and water resources engineering: Reston, Va., Environmental and Water Resources Institute of the American Society of Civil Engineers, p. 65–85.

U.S. Geological Survey, 1990, Land use and land cover digital data from 1:250,000- and 1:100,000-scale maps: U.S. Geological Survey Data Users Guide 4, 33 p.

U.S. Geological Survey, 2008, USGS science in the decade 2007–2017: Facing tomorrow's challenges—An overview: U.S. Geological Survey Fact Sheet 2008-3008, 4 p.

Wang, Wensheng, and Ding, Jing, 2003, Wavelet network model and its application to the prediction of hydrology: Nature and Science, v. 1, no. 1, p. 67–71.

Wei, Shouke, Yang, Hong, Song, Jinxi, Abbaspour, Karim, and Xu, Zongxue, 2013, A wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows: Hydrological Sciences Journal, v. 58, no. 2, p. 374–389.

Weiss, S.M., and Indurkhya, Nitin, 1998, Predictive data mining—A practical guide: San Francisco, Morgan Kaufmann Publishers, Inc., 228 p.

Table 2. Site information for lakes, wells, and springs for which artificial neural network models were developed.

[GW, groundwater level; LK, lake water level; SP, spring flow; FA, Floridan aquifer system; SAS, surficial aquifer system; --, not applicable]

Site number	ANN model site identification label	ECFT model site identification label	Hydrologic database site identification label	Site name used in this report	Latitude	Longitude	Data type	Aquifer	Well depth, feet	Period of record start date	Period of record end date	Period of record length, years	Number of measurements	Average number of measurements per year	Cluster analysis group number
5	L-1547081	63	15470818_M	Prevatt Lake (Apopka)	28.7152	-81.4862	LK	--	--	1/1/1960	12/3/2008	49.0	1,979	40.4	5
38	L-2263868	134	2263868	South Lake (Vineland)	28.4132	-81.5378	LK	--	--	4/9/1969	5/6/2009	40.1	14,347	357.8	6
41	L-2266239	125	2266239	Trout Lake (Clermont)	28.4514	-81.7165	LK	--	--	3/16/1970	3/28/2009	39.1	2,103	53.8	3
45	L-2268200	349	2268200	Lake Wales (Lake Wales)	27.9039	-81.5787	LK	--	--	12/31/1951	6/9/2009	57.5	3,499	60.9	2
49	L-2269200	259	2269200	Crooked Lake (Babson Park)	27.8280	-81.5530	LK	--	--	4/29/1945	6/22/2009	64.2	4,021	62.6	1
50	L-2269300	269	2269300	Lake Clinch (Frostproof)	27.7545	-81.5401	LK	--	--	1/31/1947	6/1/2009	62.4	3,636	58.3	2
55	L-2293670	211	2293670	Lake Otis (Winter Hven)	28.0170	-81.7142	LK	--	--	8/4/1954	6/23/2009	54.9	14,349	261.3	6
56	L-2293728	223	2293728	Lake Annie (Waverly)	27.9852	-81.6107	LK	--	--	8/21/1970	6/23/2009	38.9	1,875	48.2	3
57	L-2293763	236	2293763	Lake Starr (Waverly)	27.9545	-81.5923	LK	--	--	2/28/1983	6/23/2009	26.3	4,979	189.1	2
64	L-2294071	351	2294071	Lake McLeod (Eagle Lake)	27.9738	-81.7499	LK	--	--	3/13/1965	6/23/2009	44.3	2,773	62.6	1
71	L-2294442	218	2294442	Eagle Lake (Eagle Lake)	27.9767	-81.7754	LK	--	--	2/12/1966	6/23/2009	43.4	1,977	45.6	1
77	L-2310950	178	2310950	Lake Deeson (Lakeland)	28.1110	-81.9302	LK	--	--	7/30/1954	6/1/2009	54.9	11,910	217.0	6
168	L-2930258	89	2930258	Apshawa Lake (Clermont)	28.5996	-81.7733	LK	--	--	4/6/1953	12/22/2008	55.8	7,228	129.6	2
170	L-3000300	64	30003000	Lake Apopka (Apopka)	28.5627	-81.6282	LK	--	--	9/1/1942	1/5/2009	66.4	23,743	357.6	4
187	L-3980647	113	3980647	Lake Louisa (Clermont)	28.4975	-81.7338	LK	--	--	3/1/1957	12/20/2008	51.8	8,930	172.3	5

Table 2. Site information for lakes, wells, and springs for which artificial neural network models were developed.—Continued

[GW, groundwater level; LK, lake water level; SP, spring flow; FA, Floridan aquifer system; SAS, surficial aquifer system; --, not applicable]

Site number	ANN model site identification label	ECFT model site identification label	Hydrologic database site identification label	Site name used in this report	Latitude	Longitude	Data type	Aquifer	Well depth, feet	Period of record start date	Period of record end date	Period of record length, years	Number of measurements	Average number of measurements per year	Cluster analysis group number
207	L-7579_M	76	7579_M	Lake Howell (Casselberry)	28.6369	-81.3179	LK	--	--	10/10/1978	1/29/2009	30.3	313	10.3	1
209	L-7583	--	7583	Island Lake (Longwood)	28.6885	-81.3617	LK	--	--	10/9/1978	1/30/2009	30.3	314	10.4	1
221	L-ANDERSON	--	ANDERSON	Lake Anderson (Orlando)	28.4990	-81.3370	LK	--	--	1/1/1960	1/6/2009	49.0	400	8.2	6
239	L-CROOKED	--	CROOKED	Crooked Lake (Orlando)	28.5934	-81.4778	LK	--	--	12/1/1960	1/7/2009	48.1	412	8.6	2
277	L-MCCOY	66	MCCOY	Lake McCoy (Apopka)	28.6892	-81.4969	LK	--	--	3/1/1967	10/1/2008	41.6	333	8.0	5
289	L-ROSE	116	ROSE (WEKIVA RIVER)	Lake Rose (Gotha)	28.5367	-81.5044	LK	--	--	1/1/1960	10/2/2008	48.8	484	9.9	2
297	L-SUE	98	SUE	Lake Sue (Winter Park)	28.5769	-81.3544	LK	--	--	11/1/1960	10/3/2008	48.0	487	10.2	1
308	S-2234600	SP21SN135	2234600	Wekiwa Springs	28.7122	-81.4598	SP	--	--	3/8/1932	6/17/2009	77.3	1,901	24.6	3
309	S-2234610	SP16SN133	2234610	Rock Springs	28.7558	-81.4992	SP	--	--	2/5/1931	6/8/2009	78.4	2,397	30.6	3
311	S-2234650	SP14SN137	2234650	Miami Springs	28.7103	-81.4426	SP	--	--	8/8/1945	6/8/2009	63.9	1,579	24.7	2
312	S-2234991	SP17SN122	2234991	Sanlando Springs	28.6889	-81.3956	SP	--	--	11/12/1941	6/15/2009	67.6	1,665	24.6	2
313	S-2234996	SP15SN124	2234996	Palm Springs	28.6911	-81.3926	SP	--	--	11/12/1941	6/15/2009	67.6	1,593	23.6	2
314	S-2234997	SP19SN126	2234997	Starbuck Spring	28.6969	-81.3909	SP	--	--	7/25/1944	6/15/2009	64.9	1,433	22.1	2
446	W-2745470	romp_45_ap	274547081470903	ROMP 45 Avon Park well	27.7643	-81.7861	GW	FA	757	8/21/1980	6/30/2009	28.9	10,232	354.3	1
469	W-2753140	romp_59_ap	275314081514201	ROMP 59 Avon Park well	27.8839	-81.8641	GW	FA	1,048	2/2/1977	5/11/2009	32.3	11,650	360.8	1

Table 2. Site information for lakes, wells, and springs for which artificial neural network models were developed.—Continued

[GW, groundwater level; LK, lake water level; SP, spring flow; FA, Floridan aquifer system; SAS, surficial aquifer system; --, not applicable]

Site number	ANN model site identification label	ECFT model site identification label	Hydrologic database site identification label	Site name used in this report	Latitude	Longitude	Data type	Aquifer	Well depth, feet	Period of record start date	Period of record end date	Period of record length, years	Number of measurements	Average number of measurements per year	Cluster analysis group number
472	W-2753260	romp_60_fl	275326081585801	ROMP 60 Floridan well	27.8909	-81.9825	GW	FA	710	2/8/1955	6/30/2009	54.4	13,737	252.4	1
478	W-2754110	romp_57_fl	275411081372001	ROMP 57 Floridan well	27.9036	-81.6225	GW	FA	634	7/21/1981	5/11/2009	27.8	10,010	359.8	1
516	W-2759590	sanlon_ran	275959081552501	Sanlon Ranch Floridan well	28.0002	-81.9230	GW	FA	1,220	1/10/1970	5/11/2009	39.4	13,027	331.0	1
542	W-2810080	lk_alfdp2	281008081441801	Lake Alfred Floridan well	28.1686	-81.7376	GW	FA	425	7/8/1959	5/11/2009	49.9	16,105	322.9	2
571	W-2817140	lk_joel	281714081093001	Lake Joel Floridan well	28.2875	-81.1581	GW	FA	750	11/20/1969	5/12/2009	39.5	11,847	299.9	2
578	W-2822020-1	lk_oli_dp	282202081384601	Lake Oliver Floridan well	28.3675	-81.6459	GW	FA	318	2/24/1959	5/11/2009	50.2	16,897	336.3	2
579	W-2822020-2	lk_oli_sh	282202081384602	Lake Oliver Surficial well	28.3675	-81.6459	GW	SAS	38	2/10/1959	5/11/2009	50.3	14,893	296.2	2
588	W-2823410	cocoa_a	282341081040101	Cocoa A Floridan well	28.3950	-81.0667	GW	FA	516	3/9/1960	5/11/2009	49.2	17,673	359.2	2
602	W-2825280	baylake_dp	282528081340901	Bay Lake Floridan well	28.4247	-81.5690	GW	FA	223	3/1/1966	5/11/2009	43.2	14,739	341.0	3
613	W-2825310	cocoa_d	282531081095701	Cocoa D Floridan well	28.4256	-81.1656	GW	FA	300	11/10/1965	5/11/2009	43.5	15,010	344.8	3
622	W-2826230	cocoa_p	282623081153801	Cocoa P Floridan well	28.4400	-81.2603	GW	FA	439	7/6/1961	5/12/2009	47.9	12,134	253.4	3
627	W-2827170	romp_dp101	282717081553101	ROMP 101 Floridan well	28.4550	-81.9251	GW	FA	404	7/7/1977	6/1/2009	31.9	11,161	349.6	2
630	W-2827380	lk_sawyer	282738081341401	Lake Sawyer Floridan well	28.4608	-81.5704	GW	FA	178	2/22/1979	5/11/2009	30.2	9,330	308.6	3
632	W-2828350	palm_lake	282835081305201	Palm Lake Dr Floridan well	28.4778	-81.5070	GW	FA	235	2/22/1979	5/11/2009	30.2	9,109	301.3	3
642	W-2832040-1	mascotted	283204081544901	Mascotte Floridan well	28.5347	-81.9134	GW	FA	160	1/27/1959	5/11/2009	50.3	17,133	340.5	2

Table 2. Site information for lakes, wells, and springs for which artificial neural network models were developed.—Continued

[GW, groundwater level; LK, lake water level; SP, spring flow; FA, Floridan aquifer system; SAS, surficial aquifer system; --, not applicable]

Site number	ANN model site identification label	ECFT model site identification label	Hydrologic database site identification label	Site name used in this report	Latitude	Longitude	Data type	Aquifer	Well depth, feet	Period of record start date	Period of record end date	Period of record length, years	Number of measurements	Average number of measurements per year	Cluster analysis group number
643	W-2832040-2	mascottes	283204081544902	Mascotte Surficial well	28.5347	-81.9134	GW	SAS	30	1/28/1959	5/11/2009	50.3	16,858	335.0	2
645	W-2832490	or0007	283249081053201	Bithlo 1 Floridan well	28.5475	-81.0920	GW	FA	492	10/20/1960	5/11/2009	48.6	17,211	354.2	2
648	W-2832530	or_47	283253081283401	OR-47 Floridan well	28.5483	-81.4759	GW	FA	350	9/30/1930	4/28/2009	78.6	21,315	271.1	3
650	W-2833140	city_wel_r	283314081455501	Clermont Floridan Rep. well	28.5542	-81.7651	GW	FA	525	5/17/1982	5/11/2009	27.0	8,908	329.9	3
661	W-2841470	seminol125	284147081220201	Seminole 125 Floridan well	28.6967	-81.3670	GW	FA	146	10/25/1951	5/11/2009	57.6	17,959	311.9	3
712	W-5038	taft_g	5038	Taft Surficial well	28.4361	-81.3715	GW	SAS	7	6/10/1969	7/9/2004	35.1	12,629	359.8	2

Appendix 1. Description of hydrologic database

The hydrologic database was developed using Microsoft Access 2010® to store all historical hydrologic data compiled during this study and is available for download from the U.S. Geological Survey (USGS) (*http://dx.doi.org/10.3133/sir20145032*). The database comprises nine tables. The "README" Access table contains two fields, VERSION_DATE and COMMENTS, which document revisions to the database and the corresponding date the change was made. The "Station List" Access table contains site-specific information about each of the 963 sites. Table A1-1 lists the field names, type of field, and a brief description of the data in the fields for the "Station List" Access table. The STATION_ID field serves as the unique identifier by which the site record in the "Station List" Access table can be related to records in other Access tables. The other seven Access tables contain the time-series data for all sites for each type of data; the tables listed below provide the field names, type of field, and a brief description of the data in the field for each of the Access data tables.

- "GOES_ET" Access table contains potential and reference daily evapotranspiration data (extracted for the 2-kilometer pixel overlying each of the 21 National Oceanic and Atmospheric Administration (NOAA) air temperature measurement sites) from the USGS Statewide Evapotranspiration dataset (*http://fl.water.usgs.gov/et/*) (Mecikalski and others, 2011) (table A1-2);
- "Hargreaves_PET" Access table contains daily potential evapotranspiration data (at each of the 21 NOAA air temperature measurement sites) computed by using the Hargreaves equation (Hargreaves and Samani, 1985) (table A1-3);
- "Lakes" Access table contains daily lake water-level data (table A1-4);
- "Rainfall" Access table contains daily rainfall data, with the caveat that sites not operated by NOAA may contain data accumulated over periods from 2 days to a month or more (table A1-5);
- "Springs" Access table contains daily spring-flow data (table A1-6);
- "Temperature" Access table contains daily minimum, maximum, and mean air temperature data at 21 NOAA sites (table A1-7); and
- "Wells" Access table contains daily groundwater-level data (table A1-8).

Codes denoting the quality of the data were assigned to many individual data values by the St. Johns River Water Management District, South Florida Water Management District, and Southwest Florida Water Management District and are defined in tables A1-9, 1-10, and 1-11, respectively). Separate quality codes are provided for rainfall data as explained in table A1-5 .

Data for duplicate sites (sites monitored by more than one agency) were merged, where consistency in data quality among the collecting agencies could be verified. When data were available for the same date at duplicate sites, the data value for the site with the longer period of record was retained. As a result, a single time series of data was created for each site. In the STATION_ID field, sites with an "_M" appended at the end of the station identifier denote duplicate sites with different station identifiers that were merged and assigned the station identifier for the station with the longest period of record. Merged sites with the same station identifier are not denoted with an "_M" and use the original station identifier.

Two Access queries are included in the hydrologic database for updating the site-specific statistics in the Station List Access table. First, the "Build Station List (Step 1)" query is executed to delete all existing records in the Station List Access table. Second, the "Build Station List (Step 2)" query is executed to compute all statistics on the basis of the data currently stored in Lakes, Rainfall, Springs, Temperature, and Wells Access tables and repopulate all the fields in the Station List Access table. In this way, should it be necessary to add or remove data records, the Station List Access table can easily be updated.

Table A1-1 . Description of fields in the "Station List" table of the hydrologic database.

[ft, feet; cfs, cubic feet per second; in, inches, deg F, degrees Fahrenheit; yr, year]

Field name	Data type	Description
STATION_ID	Text	Unique identifier assigned by collecting agency
STATION_NM	Text	Station name assigned by collecting agency
AGENCY	Text	Collecting agency
TYPE	Text	Data type: GW, groundwater level, ft; LK, lake level, ft; SP, spring flow, cfs; RN, rainfall, in; TM, minimum, maximum, and mean daily air temperature, deg F
AQUIFER	Text	Aquifer tapped by open interval of well; SAS, surficial aquifer system; IM, intermediate aquifer system; FA, Floridan aquifer system; UFA, Upper Floridan aquifer; LFA, Lower Floridan aquifer
DEPTH	Number	Total depth of well, in feet
POR_START	Date/Time	Start date of period of record
POR_END	Date/Time	End date of period of record
YEARS	Number	Length of period of record, in years
COUNT	Number	Number of measurements (total)
COUNTS_PER	Number	Number of measurements per year
MIN	Number	Minimum measurement value; For TM, statistic computed on Mean Temperature
MAX	Number	Maximum measurement value; For TM, statistic computed on Mean Temperature
RANGE	Number	Maximum minus minimum measurement value; For TM, statistic computed on Mean Temperature
MEAN	Number	Arithmetic mean of measurement value; For TM, statistic computed on Mean Temperature
SLOPE	Number	Linear trend: in/yr for GW, LK, RN; cfs/yr for SP; deg F/yr for TM
R	Number	Pearson product-moment correlation coefficient for linear trend
R_SQUARED	Number	Coefficient of determination for linear trend
T	Number	Student's t statistic computed from COUNT and R
LAT_DD	Number	Latitude, decimal degrees
LON_DD	Number	Longitude, decimal degrees
PRIORITY	Text	Classification assigned by SJRWMD and SWFWMD

Table A1-2. Description of fields in the "GOES_ET" table of the hydrologic database.

[km, kilometer; mm/d, millimeters per day]

Field name	Data type	Description
Station_ID	Text	Unique identifier assigned by collecting agency; for GOES_ET sites this is equal to the NOAA identification number followed by a "G" for the 2-km pixel overlying the NOAA station
Measurement_Date	Date/Time	Date measurement was made
PET_Value	Number	Potential evapotranspiration, in mm/d
RET_Value	Number	Reference evapotranspiration, in mm/d

Table A1-3. Description of fields in the "Hargreaves_PET" table of the hydrologic database.

[mm/d, millimeters per day]

Field name	Data type	Description
Station_ID	Text	Unique identifier assigned by collecting agency; for Hargreaves_PET sites this is equal to the Station_ID for the corresponding air temperature site
Measurement_Date	Date/Time	Date measurement was made
PET_Value	Number	Potential evapotranspiration, in mm/d

Table A1-4. Description of fields in the "Lakes" table of the hydrologic database.

Field name	Data type	Description
Station_ID	Text	Unique identifier assigned by collecting agency
Measurement_Date	Date/Time	Date measurement was made
Lake_Value	Number	Lake water level, in feet
Code	Text	Quality code assigned by collecting agency (see tables A1-9, A1-10, and A1-11); blank if no quality code specified by collecting agency

Table A1-5. Description of fields in the "Rainfall" table of the hydrologic database.

[NOAA, National Oceanic and Atmospheric Administration; SJRWMD, St. Johns River Water Management Distric; SWFWMD, Southwest Florida Water Management District]

Field name	Data type	Description
Station_ID	Text	Unique identifier assigned by collecting agency
Measurement_Date	Date/Time	Date measurement was made
Rainfall_Value	Number	Rainfall total, in inches
Error	Number	An error value computed by the algorithm developed by Aly (2008) for NOAA records coded as "E"; a value of 0.00 indicates a rainfall data value used by Aly (2008) that was coded as "G"; blank if rainfall data value not used by Aly (2008)
Code	Text	Quality code assigned by collecting agency: A, accumulated data at NOAA station disaggregated to daily data by SJRWMD; E, estimated data at NOAA station computed by Aly (2008); G, good data reported by NOAA or accumulated data disaggregated to daily data by SWFWMD; M, missing data at NOAA station estimated by SJRWMD; blank if no quality code specified by collecting agency

Table A1-6. Description of fields in the "Springs" table of the hydrologic database.

[cfs, cubic feet per second]

Field name	Data type	Description
Station_ID	Text	Unique identifier assigned by collecting agency
Measurement_Date	Date/Time	Date measurement was made
Springflow_Value	Number	Spring flow, in cfs
Code	Text	Quality code assigned by collecting agency (see tables A1-9, A1-10, and A1-11); blank if no quality code specified by collecting agency

Table A1-7. Description of fields in the "Temperature" table of the hydrologic database.

[deg F, degrees Fahrenheit]

Field name	Data type	Description
Station_ID	Text	Unique identifier assigned by collecting agency; for temperature sites this is equal to the NOAA identification number followed by a "T"
Measurement_Date	Date/Time	Date measurement was made
Tmax	Number	Maximum temperature, in deg F
Tmin	Number	Minimum temperature, in deg F
Tmean	Number	Mean temperature, in deg F

Table A1-8. Description of fields in the "Wells" table of the hydrologic database.

Field name	Data type	Description
Station_ID	Text	Unique identifier assigned by collecting agency
Measurement_Date	Date/Time	Date measurement was made
Groundwater_Value	Number	Groundwater level, in feet
Code	Text	Quality code assigned by collecting agency (see tables A1-9, A1-10, and A1-11); blank if no quality code specified by collecting agency

Table A1-9. Description of quality codes assigned by the St. Johns River Water Management District for data obtained from their database.

Code	Description
1	Good quality records
2	Good quality, edited data
3	Good quality data obtained from USGS
5	Good quality data obtained from SWFWMD
72	Correlation w/other station - high confidence
74	Correlation w/ other station - see documentation regarding confidence
79	Records partly estimated
85	Estimated by contractor high confidence see documentation
111	Estimate based on correlation with neighboring station
112	Estimated Discharge
130	Unverifiable—quality unknown
140	Provisional data—quality unknown

Table A1-10. Description of quality codes assigned by the South Florida Management District for data obtained from their database.

Code	Description
E	Estimated
P	Summary computed from partial record

Table A1-11. Description of quality codes assigned by the Soutwest Florida Water Management District for data obtained from their database.

Code	Description
1	Good continuous record
2	Good quality edited data
79	Records partly estimated
95	Estimated
96	Override
97	Surveyed

Aly, Alaa, 2008, An assessment of interpolation methods for estimating missing daily precipitation records for rain gauges in central and south Florida: Niwot, Colo., INTERA Inc., 111 p.

Hargreaves, G.H., and Samani, Z.A., 1985, Reference crop evapotranspiration from temperature: Applied Engineering in Agriculture, v. 1, no. 2, p. 96–99.

Mecikalski, J.R., Sumner, D.M., Jacobs, J.M., Pathak, C.S., Paech, S.J., and Douglas, E.M., 2011, Use of visible geostationary operational meteorological satellite imagery in mapping reference and potential evapotranspiration over Florida, *in* Labedzki, Leszek, ed., Evapotranspiration: Vienna, Austria, InTech Publishers, 446 p.

Appendix 2 . Description of Groundwater-Use Data Viewer

The Groundwater-Use Data Viewer (GUDV) is a Microsoft Excel 2010® workbook created as part of this study to allow graphical review of the groundwater-use data compiled during the study and is available for download from the U.S. Geological Survey (*http://dx.doi.org/10.3133/sir20145032*). Data were aggregated into a 72-cell gridded representation of the study area using ArcGIS® and point and polygon geographic information system (GIS) data. Each grid cell is 19.67 kilometers (east to west) by 20.90 kilometers (north to south) (64,544 by 68,578 feet) and the 72-cell grid encompasses the groundwater-use data in the Central Florida Coordination Area and vicinity (see figure 28 of the main report). Groundwater use consists of six types: public supply, phosphate mining, citrus irrigation, non-citrus agricultural irrigation, golf course irrigation, and drainage well recharge. The GUDV grid and associated groundwater-use data were used to guide development of the artificial neural network models and the Central Florida Artificial Neural Network Decision Support System.

The workbook consists of 10 worksheets containing the groundwater-use data, summary results, maps, and the interactive graphical display. Additionally, the "Info" worksheet contains a history of GUDV version releases and updates. Six worksheets contain monthly groundwater withdrawal or inflow data in million gallons per day reported for the study area: "AG_Summed" contains the non-citrus agricultural irrigation withdrawal data, "Citrus_Summed" contains the citrus irrigation withdrawal data, "Golf_Summed" contains the golf course irrigation withdrawal data, "Phosphate_Summed" contains the phosphate mining withdrawal data (comprising water used for ore extraction and chemical plants), "Drainage_Summed" contains the drainage well recharge (inflow) data, and "WU_Summed" contains the public-supply (water utility) withdrawal data. Withdrawals are reported as positive values and inflows are reported as negative values. The reported period of record varies for each groundwater-use type; however, for consistency, each type was tabulated in these worksheets for months between January 1950 and December 2008. Data not reported for a given month are indicated by a cell with contents equal to "#N/A." The "ALL_SUMMED" worksheet collates the data for all the types of groundwater use and includes a summation of all types in a format that can be easily graphed.

The "Summary" worksheet presents summary descriptive statistics for the summation of all groundwater-use types for each cell in the GUDV. The descriptive statistics include the count, minimum, maximum, and summed values of monthly reported groundwater use per cell. The statistics are shown in four graphs and an associated table on a per-cell basis.

The "Maps" worksheet contains 11 maps of the study area showing the GUDV grid and groundwater-use distribution displayed by type. The maps show that non-citrus agricultural, citrus, golf, and phosphate mine locations were polygon GIS data. The remaining maps show that the other datasets—public-supply water-treatment plants in SJRWMD and SFWMD compiled by SJRWMD, public-supply water-treatment plants in SWFWMD, drainage wells, and phosphate chemical plants—were point GIS data.

The "Viewer" worksheet is used interactively to display/review the groundwater-use data on a type and monthly basis (see figures 30 and 31 of the main report). The worksheet contains six grids of the different groundwater-use types that show the groundwater use per cell for a given month and are overlain with a simple base map. Additionally, a separate map showing the location of the grid on a base map is included to orient the user to the grid cell numbers and the pumping distribution in the study area. The grid in the center of the worksheet is a summation by cell of groundwater use for all types for the given month. A toggle button located to the left of this "SUM" grid allows the user to change the month forward or backward by 1-month increments. The selected date (month and year) is displayed below the toggle button. Months are numbered sequentially starting with January 1950, and the corresponding "month column number" is displayed below the date. The user can "jump" to a specific month and year by entering the month column number directly. Month column numbers range from 2 for January 1950 to 709 for December 2008. Grid cells are automatically shaded using a color ramp of green to red to indicate relative water use in the study area for the selected month. With the exception of drainage well inflows, dark red colors indicate the highest reported water use and dark green colors indicate the lowest reported water use. Because drainage well inflows are reported as negative values, dark red indicates the lowest inflow and dark green indicates the highest inflow. If no groundwater use was reported for a cell, the cell is not colored.

Appendix 3. Summary of artificial neural network models

The rainfall artificial neural network (ANN) models were developed using daily rainfall, lake water levels, groundwater levels, and spring flows. The dataset includes daily data from 1942 to 2008. Below are nomenclature examples for input variable names for the rainfall models (a complete list of all input variable names with descriptions is provided in appendix 4, table A4-1):

- RS-FA630 = standard rainfall (RS) from gap-filled data (F) that has been moving window averaged (A) over specified number of days (630 days in this example);
- RS-FA90D = RS-FA90 – RS-FA270, where "D" denotes a difference from the moving window average (MWA) of the next larger window size (270 days in this example); the MWAs used are 30; 90; 270; 630; 1,440; and 2,250 days; and
- R5973-FA1440DS = R5973-FA1440 – RS-FA1440, where "DS" denotes the difference between rainfall (R) at the National Oceanic and Atmospheric Administration (NOAA) gage (number 5973 in this example) and the standard rainfall to which a MWA of the same window size (1,440 days in this example) has been applied.

Groundwater-use ANN models were developed using monthly groundwater-use estimates and monthly water-level and flow residuals. The daily predictions and residuals were converted to monthly values by calculating the monthly average. Below are nomenclature examples for input variable names for the groundwater-use models (a complete list of all input variable names with descriptions is provided in appendix 4, table A4-2):

- USE-TOT_A12 = Summation of all monthly groundwater-use data (USE-TOT) that has been moving window averaged (A) over specified number of months (12 months in this example);
- USE-TOT_D12 = USE-TOT_A12 – USE-TOT_A24, where "D" denotes a difference from the moving window average (MWA) of the next larger window size (24 months in this example); the MWAs used are 3, 6, 12, 24, 48, and 96 months; and
- P_W-2759590-M = predicted (P) values from the specified rainfall model (W-2759590 in this example) converted to monthly (M) values.

Input variable names, statistics, and other notes (for example, whether raw data were interpolated and how standard rainfall was computed) for both the rainfall and groundwater-use ANN models for each of the 51 sites are included below. Statistics for the rainfall ANN models include the period 1942 to 2008. Statistics for the groundwater-use ANN models include the period 1950 to 2008. Statistics for the final models (combined rainfall and groundwater-use ANN models) include the period 1965 to 2008. For two lakes (L-1547081, L-7583) and four springs (S-2234600, S-2234650, S-2234991, and S-2234997), rainfall and groundwater-use ANN model outputs exhibit considerably lower coefficients of determination (R^2) when converted to monthly averaged data. These six sites all have sparse data for the period of record up to and including 2003 and exhibit considerably higher R^2 for the period 2004 to 2008. Detailed information about the ANN models for each is provided in table A3-1 [accessible from the publication Web site (*http://dx.doi.org/10.3133/sir20145032*)].

Appendix 4. Input variables used in artificial neural network models

For the rainfall artificial neural network (ANN) models, a total of 119 rainfall input variables were computed, consisting of 6 moving window averages (MWAs) of the standard rainfall, 5 moving window differences (MWDs) of the standard rainfall, and 6 MWAs for normalized rainfall at each of the 18 National Oceanic and Atmospheric Administration (NOAA) rain gages. For the groundwater-use ANN models, a total of 13 groundwater-use input variables were computed, consisting of 6 MWAs, 6 MWDs, and the original monthly data. For some of the groundwater-use ANN models, the predicted water level or flow from the respective rainfall ANN model also was used as an input variable. Listings of the variables retained in the rainfall and groundwater-use models after training of the ANNs are provided in table A4-1 and A4-2.

Table A4-1. Input variables used in rainfall artificial neural network models.

[MWA, moving window average; NOAA, National Oceanic and Atmospheric Administration]

Input variable	Description
RS-FA30	30-day MWA of standard rainfall
RS-FA90	90-day MWA of standard rainfall
RS-FA270	270-day MWA of standard rainfall
RS-FA630	630-day MWA of standard rainfall
RS-FA1440	1440-day MWA of standard rainfall
RS-FA2250	2250-day MWA of standard rainfall
RS-FA30D	Moving window difference: 30-day MWA minus 90-day MWA of standard rainfall
RS-FA90D	Moving window difference: 90-day MWA minus 270-day MWA of standard rainfall
RS-FA270D	Moving window difference: 270-day MWA minus 630-day MWA of standard rainfall
RS-FA630D	Moving window difference: 630-day MWA minus 1440-day MWA of standard rainfall
RS-FA1440D	Moving window difference: 1440-day MWA minus 2250-day MWA of standard rainfall
R369-FA30DS	30-day MWA of normalized rainfall at NOAA gage 369
R369-FA270DS	270-day MWA of normalized rainfall at NOAA gage 369
R369-FA630DS	630-day MWA of normalized rainfall at NOAA gage 369
R369-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 369
R369-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 369
R478-FA90DS	90-day MWA of normalized rainfall at NOAA gage 478
R478-FA270DS	270-day MWA of normalized rainfall at NOAA gage 478
R478-FA630DS	630-day MWA of normalized rainfall at NOAA gage 478
R478-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 478
R478-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 478
R1163-FA30DS	30-day MWA of normalized rainfall at NOAA gage 1163
R1163-FA90DS	90-day MWA of normalized rainfall at NOAA gage 1163
R1163-FA630DS	630-day MWA of normalized rainfall at NOAA gage 1163
R1163-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 1163
R1163-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 1163
R1641-FA30DS	30-day MWA of normalized rainfall at NOAA gage 1641

Table A4-1. Input variables used in rainfall artificial neural network models.—Continued

[MWA, moving window average; NOAA, National Oceanic and Atmospheric Administration]

Input variable	Description
R1641-FA90DS	90-day MWA of normalized rainfall at NOAA gage 1641
R1641-FA270DS	270-day MWA of normalized rainfall at NOAA gage 1641
R1641-FA630DS	630-day MWA of normalized rainfall at NOAA gage 1641
R1641-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 1641
R2229-FA270DS	270-day MWA of normalized rainfall at NOAA gage 2229
R2229-FA630DS	630-day MWA of normalized rainfall at NOAA gage 2229
R2229-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 2229
R2229-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 2229
R3137-FA90DS	90-day MWA of normalized rainfall at NOAA gage 3137
R3137-FA270DS	270-day MWA of normalized rainfall at NOAA gage 3137
R3137-FA630DS	630-day MWA of normalized rainfall at NOAA gage 3137
R3137-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 3137
R4502-FA30DS	30-day MWA of normalized rainfall at NOAA gage 4502
R4502-FA90DS	90-day MWA of normalized rainfall at NOAA gage 4502
R4502-FA270DS	270-day MWA of normalized rainfall at NOAA gage 4502
R4502-FA630DS	630-day MWA of normalized rainfall at NOAA gage 4502
R4502-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 4502
R4502-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 4502
R4625-FA30DS	30-day MWA of normalized rainfall at NOAA gage 4625
R4625-FA90DS	90-day MWA of normalized rainfall at NOAA gage 4625
R4625-FA270DS	270-day MWA of normalized rainfall at NOAA gage 4625
R4625-FA630DS	630-day MWA of normalized rainfall at NOAA gage 4625
R4625-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 4625
R4625-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 4625
R5076-FA90DS	90-day MWA of normalized rainfall at NOAA gage 5076
R5076-FA270DS	270-day MWA of normalized rainfall at NOAA gage 5076
R5076-FA630DS	630-day MWA of normalized rainfall at NOAA gage 5076

Table A4-1. Input variables used in rainfall artificial neural network models.—Continued

[MWA, moving window average; NOAA, National Oceanic and Atmospheric Administration]

Input variable	Description
R5076-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 5076
R5076-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 5076
R5612-FA90DS	90-day MWA of normalized rainfall at NOAA gage 5612
R5612-FA270DS	270-day MWA of normalized rainfall at NOAA gage 5612
R5612-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 5612
R5612-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 5612
R5973-FA30DS	30-day MWA of normalized rainfall at NOAA gage 5973
R5973-FA90DS	90-day MWA of normalized rainfall at NOAA gage 5973
R5973-FA270DS	270-day MWA of normalized rainfall at NOAA gage 5973
R5973-FA630DS	630-day MWA of normalized rainfall at NOAA gage 5973
R5973-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 5973
R5973-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 5973
R6414-FA30DS	30-day MWA of normalized rainfall at NOAA gage 6414
R6414-FA90DS	90-day MWA of normalized rainfall at NOAA gage 6414
R6414-FA270DS	270-day MWA of normalized rainfall at NOAA gage 6414
R6414-FA630DS	630-day MWA of normalized rainfall at NOAA gage 6414
R6414-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 6414
R6414-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 6414
R6628-FA30DS	30-day MWA of normalized rainfall at NOAA gage 6628
R6628-FA270DS	270-day MWA of normalized rainfall at NOAA gage 6628
R6628-FA630DS	630-day MWA of normalized rainfall at NOAA gage 6628
R7205-FA30DS	30-day MWA of normalized rainfall at NOAA gage 7205
R7205-FA90DS	90-day MWA of normalized rainfall at NOAA gage 7205
R7205-FA270DS	270-day MWA of normalized rainfall at NOAA gage 7205
R7205-FA630DS	630-day MWA of normalized rainfall at NOAA gage 7205
R7205-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 7205
R7205-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 7205

Table A4-1. Input variables used in rainfall artificial neural network models.—Continued

[MWA, moving window average; NOAA, National Oceanic and Atmospheric Administration]

Input variable	Description
R7982-FA30DS	30-day MWA of normalized rainfall at NOAA gage 7982
R7982-FA90DS	90-day MWA of normalized rainfall at NOAA gage 7982
R7982-FA270DS	270-day MWA of normalized rainfall at NOAA gage 7982
R7982-FA630DS	630-day MWA of normalized rainfall at NOAA gage 7982
R7982-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 7982
R7982-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 7982
R8942-FA90DS	90-day MWA of normalized rainfall at NOAA gage 8942
R8942-FA270DS	270-day MWA of normalized rainfall at NOAA gage 8942
R8942-FA630DS	630-day MWA of normalized rainfall at NOAA gage 8942
R8942-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 8942
R8942-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 8942
R9401-FA90DS	90-day MWA of normalized rainfall at NOAA gage 9401
R9401-FA270DS	270-day MWA of normalized rainfall at NOAA gage 9401
R9401-FA630DS	630-day MWA of normalized rainfall at NOAA gage 9401
R9401-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 9401
R9401-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 9401
R9707-FA90DS	90-day MWA of normalized rainfall at NOAA gage 9707
R9707-FA270DS	270-day MWA of normalized rainfall at NOAA gage 9707
R9707-FA630DS	630-day MWA of normalized rainfall at NOAA gage 9707
R9707-FA1440DS	1440-day MWA of normalized rainfall at NOAA gage 9707
R9707-FA2250DS	2250-day MWA of normalized rainfall at NOAA gage 9707

Table A4-2 . Input variables used in groundwater-use artificial neural network models.

[MWA, moving window average]

Input variables	Description
USE-TOT_A3	3-month MWA of total groundwater use
USE-TOT_A6	6-month MWA of total groundwater use
USE-TOT_A12	12-month MWA of total groundwater use
USE-TOT_A24	24-month MWA of total groundwater use
USE-TOT_A48	48-month MWA of total groundwater use
USE-TOT_A96	96-month MWA of total groundwater use
USE-TOT_D1	Moving window difference: 1-month MWA minus 3-month MWA of total groundwater use
USE-TOT_D3	Moving window difference: 3-month MWA minus 6-month MWA of total groundwater use
USE-TOT_D6	Moving window difference: 6-month MWA minus 12-month MWA of total groundwater use
USE-TOT_D12	Moving window difference: 12-month MWA minus 24-month MWA of total groundwater use
USE-TOT_D24	Moving window difference: 24-month MWA minus 48-month MWA of total groundwater use
USE-TOT_D48	Moving window difference: 48-month MWA minus 96-month MWA of total groundwater use
P_L-2310950-M	Predicted water levels from the L-2310950 rainfall model converted to monthly values
P_L-ANDERSON-M	Predicted water levels from the L-ANDERSON rainfall model converted to monthly values
P_L-ROSE-M	Predicted water levels from the L-ROSE rainfall model converted to monthly values
P_L-SUE-NEW-M	Predicted water levels from the L-SUE rainfall model converted to monthly values
P_S-2234600-M	Predicted flows from the S-2234600 rainfall model converted to monthly values
P_S-2234610-M	Predicted flows from the S-2234610 rainfall model converted to monthly values
P_S-2234650-M	Predicted flows from the S-2234650 rainfall model converted to monthly values
P_S-2234991-M	Predicted flows from the S-2234991 rainfall model converted to monthly values
P_W-2759590-M	Predicted water levels from the W-2759590 rainfall model converted to monthly values
P_W-2810080-M	Predicted water levels from the W-2810080 rainfall model converted to monthly values
P_W-2826230-M	Predicted water levels from the W-2826230 rainfall model converted to monthly values

Appendix 5. Summary statistics for lake water-level, groundwater-level, and spring-flow models

The data presented in table A5-1 are summary statistics for both measured and simulated data for the final lake water-level, groundwater-level, and spring-flow models. Statistics were computed using monthly data from January 1965 through December 2008; a measured time series consists of arithmetic averages of available measured daily data for each month and a simulated time series consists of the final predicted values from the two-stage model (combined rainfall and groundwater-use artificial neural network models). Values of statistics reported for the individual model summaries provided in appendix 3 will be different than those reported here because of different periods of record and time step used for the rainfall artificial neural network models and (or) the inclusion of 60-day interpolated estimates in the training datasets for some sites that also were used in the computation of model-fit statistics in appendix 3.

Table A5-1 . Summary statistics for measured and simulated data for the lake water-level, groundwater-level, and spring-flow models, 1965–2008.

[Meas, measured; Sim, simulated; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing measured data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	$PN_{missing}$	R^2	NSCE	ME	SE	RMSE	PME
5	Prevatt Lake (Apopka)	Meas	46.87	61.59	54.17	54.19	14.72	2.38	302	--	--	--	--	--	--	--
		Sim	49.58	58.52	53.79	53.70	8.94	1.64	528	42.8%	0.432	0.431	0.03	1.25	1.80	12.20%
38	South Lake (Vineland)	Meas	89.20	94.35	92.75	93.01	5.15	1.13	474	--	--	--	--	--	--	--
		Sim	88.74	94.32	92.76	93.04	5.58	1.07	528	10.2%	0.820	0.814	-0.00	0.47	0.49	9.47%
41	Trout Lake (Clermont)	Meas	86.01	98.26	93.03	93.14	12.26	2.82	462	--	--	--	--	--	--	--
		Sim	85.34	98.56	93.14	93.23	13.22	2.51	528	12.5%	0.791	0.790	-0.01	1.20	1.29	10.56%
45	Lake Wales (Lake Wales)	Meas	97.73	111.66	103.28	103.20	13.93	2.90	509	--	--	--	--	--	--	--
		Sim	98.57	110.05	103.29	103.03	11.48	2.59	528	3.6%	0.778	0.778	-0.00	1.23	1.37	9.83%
49	Crooked Lake (Babson Park)	Meas	106.24	123.42	114.49	115.28	17.18	3.70	528	--	--	--	--	--	--	--
		Sim	105.77	121.42	114.47	115.21	15.65	3.79	528	0.0%	0.887	0.881	0.02	1.27	1.28	7.44%
50	Lake Clinch (Frostproof)	Meas	100.34	109.41	103.85	104.02	9.06	1.72	507	--	--	--	--	--	--	--
		Sim	100.39	107.06	103.68	103.86	6.67	1.50	528	4.0%	0.761	0.754	0.14	0.73	0.85	9.38%
55	Lake Otis (Winter Haven)	Meas	120.02	128.43	124.44	124.48	8.41	1.82	528	--	--	--	--	--	--	--
		Sim	120.30	127.68	124.43	124.42	7.38	1.58	528	0.0%	0.752	0.752	0.02	0.79	0.90	10.75%
56	Lake Annie (Waverly)	Meas	108.40	117.56	112.32	111.93	9.16	2.22	458	--	--	--	--	--	--	--
		Sim	108.55	117.44	112.32	111.98	8.89	1.94	528	13.3%	0.853	0.853	0.01	0.78	0.85	9.29%

Table A5-1 . Summary statistics for measured and simulated data for the lake water-level, groundwater-level, and spring-flow models, 1965–2008.—Continued

[Meas, measured; Sim, simulated; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing measured data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	$PN_{missing}$	R^2	NSCE	ME	SE	RMSE	PME
57	Lake Starr (Waverly)	Meas	96.38	109.74	102.04	101.63	13.37	2.85	297	--	--	--	--	--	--	--
		Sim	97.12	109.32	101.95	101.68	12.20	2.26	528	43.8%	0.879	0.878	-0.01	0.97	1.00	7.46%
64	Lake McLeod (Eagle Lake)	Meas	115.66	131.98	123.70	122.88	16.32	4.12	475	--	--	--	--	--	--	--
		Sim	117.14	131.99	123.51	122.72	14.85	3.82	528	10.0%	0.894	0.894	0.00	1.28	1.34	8.21%
71	Eagle Lake (Eagle Lake)	Meas	118.91	131.50	124.90	124.41	12.59	3.07	470	--	--	--	--	--	--	--
		Sim	119.55	130.92	124.74	124.19	11.37	2.83	528	11.0%	0.885	0.885	0.00	0.97	1.04	8.26%
77	Lake Deeson (Lakeland)	Meas	117.67	132.94	125.67	125.82	15.27	2.76	524	--	--	--	--	--	--	--
		Sim	119.99	132.00	125.76	125.91	12.01	2.29	528	0.8%	0.615	0.611	-0.07	1.43	1.72	11.25%
168	Apshawa Lake (Clermont)	Meas	80.24	88.08	84.46	84.64	7.84	1.67	500	--	--	--	--	--	--	--
		Sim	81.24	88.88	84.40	84.35	7.64	1.56	528	5.3%	0.679	0.669	0.02	0.87	0.96	12.24%
170	Lake Apopka (Apopka)	Meas	62.83	68.47	66.53	66.76	5.64	0.91	524	--	--	--	--	--	--	--
		Sim	63.75	68.07	66.49	66.62	4.32	0.81	528	0.8%	0.553	0.531	0.03	0.54	0.62	11.01%
187	Lake Louisa (Clermont)	Meas	88.08	98.30	95.25	95.53	10.23	1.74	524	--	--	--	--	--	--	--
		Sim	88.17	97.65	95.26	95.65	9.48	1.64	528	0.8%	0.793	0.790	-0.01	0.75	0.80	7.77%
207	Lake Howell (Casselberry)	Meas	51.05	56.58	53.04	53.10	5.53	0.80	301	--	--	--	--	--	--	--
		Sim	50.78	55.22	53.27	53.29	4.44	0.79	528	43.0%	0.654	0.648	-0.01	0.42	0.48	8.62%

Table A5-1 . Summary statistics for measured and simulated data for the lake water-level, groundwater-level, and spring-flow models, 1965–2008.—Continued

[Meas, measured; Sim, simulated; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing measured data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	$PN_{missing}$	R^2	NSCE	ME	SE	RMSE	PME
209	Island Lake (Longwood)	Meas	78.62	84.17	81.54	81.69	5.55	0.73	301	--	--	--	--	--	--	--
		Sim	78.91	82.16	81.46	81.63	3.25	0.59	528	43.0%	0.564	0.563	0.02	0.37	0.48	8.72%
221	Lake Anderson (Orlando)	Meas	74.03	86.16	79.84	79.96	12.13	2.27	354	--	--	--	--	--	--	--
		Sim	75.22	83.71	80.22	80.52	8.49	1.78	528	33.0%	0.660	0.660	0.03	1.05	1.32	10.90%
239	Crooked Lake (Orlando)	Meas	54.69	73.60	64.34	64.65	18.91	3.70	351	--	--	--	--	--	--	--
		Sim	56.56	71.84	63.97	63.80	15.28	3.14	528	33.5%	0.630	0.629	0.12	1.82	2.25	11.91%
277	Lake McCoy (Apopka)	Meas	52.43	62.01	59.48	60.23	9.58	2.09	321	--	--	--	--	--	--	--
		Sim	51.49	61.82	59.54	60.09	10.33	1.97	528	39.2%	0.640	0.629	0.01	1.14	1.27	13.29%
289	Lake Rose (Gotha)	Meas	67.78	86.62	79.17	80.42	18.84	3.29	415	--	--	--	--	--	--	--
		Sim	67.29	82.99	79.32	80.04	15.70	2.91	528	21.4%	0.649	0.641	-0.26	1.66	1.97	10.48%
297	Lake Sue (Winter Park)	Meas	69.60	73.40	71.46	71.48	3.80	0.64	423	--	--	--	--	--	--	--
		Sim	70.37	72.65	71.44	71.43	2.28	0.49	528	19.9%	0.635	0.633	0.03	0.31	0.39	10.19%
446	ROMP 45 Avon Park well	Meas	34.43	83.66	62.96	64.50	49.23	9.82	334	--	--	--	--	--	--	--
		Sim	38.15	81.08	64.26	65.47	42.93	8.71	528	36.7%	0.815	0.814	-0.04	3.97	4.23	8.60%
469	ROMP 59 Avon Park well	Meas	35.69	85.19	63.41	64.66	49.50	9.68	381	--	--	--	--	--	--	--
		Sim	40.44	81.57	63.69	64.76	41.13	8.68	528	27.8%	0.793	0.792	-0.07	4.04	4.42	8.92%

Table A5-1 . Summary statistics for measured and simulated data for the lake water-level, groundwater-level, and spring-flow models, 1965–2008.—Continued

[Meas, measured; Sim, simulated; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing measured data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	$PN_{missing}$	R^2	NSCE	ME	SE	RMSE	PME
472	ROMP 60 Floridan well	Meas	27.36	83.20	59.78	60.21	55.84	9.98	516	--	--	--	--	--	--	--
		Sim	33.17	82.27	60.95	60.90	49.10	9.15	528	2.3%	0.720	0.703	-1.17	4.80	5.44	9.75%
478	ROMP 57 Floridan well	Meas	89.03	111.90	102.40	102.70	22.87	4.47	325	--	--	--	--	--	--	--
		Sim	90.50	111.10	102.05	102.26	20.60	3.96	528	38.4%	0.733	0.728	-0.06	2.14	2.33	10.19%
516	Sanlon Ranch Floridan well	Meas	67.79	104.41	88.96	89.42	36.62	6.73	461	--	--	--	--	--	--	--
		Sim	70.46	102.43	89.84	90.12	31.97	6.01	528	12.7%	0.842	0.837	-0.46	2.42	2.72	7.42%
542	Lake Alfred Floridan well	Meas	120.88	130.98	126.98	127.07	10.10	1.72	498	--	--	--	--	--	--	--
		Sim	122.08	129.98	127.01	127.20	7.90	1.50	528	5.7%	0.710	0.709	-0.02	0.81	0.93	9.20%
571	Lake Joel Floridan well	Meas	36.68	47.15	43.38	43.61	10.47	1.84	386	--	--	--	--	--	--	--
		Sim	37.63	46.32	43.71	44.14	8.69	1.74	528	26.9%	0.819	0.813	-0.04	0.77	0.79	7.59%
578	Lake Oliver Floridan well	Meas	103.74	111.66	108.39	108.51	7.92	1.35	517	--	--	--	--	--	--	--
		Sim	104.41	110.89	108.40	108.51	6.48	1.32	528	2.1%	0.763	0.754	-0.01	0.64	0.67	8.48%
579	Lake Oliver Surficial well	Meas	106.23	114.64	110.92	111.06	8.41	1.52	450	--	--	--	--	--	--	--
		Sim	106.81	113.72	110.90	111.13	6.91	1.43	528	14.8%	0.834	0.834	-0.00	0.57	0.62	7.36%
588	Cocoa A Floridan well	Meas	29.55	40.28	35.91	36.11	10.73	1.86	520	--	--	--	--	--	--	--
		Sim	30.96	40.49	35.93	36.05	9.53	1.86	528	1.5%	0.789	0.775	-0.03	0.86	0.88	8.20%

Table A5-1 . Summary statistics for measured and simulated data for the lake water-level, groundwater-level, and spring-flow models, 1965–2008.—Continued

[Meas, measured; Sim, simulated; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing measured data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	$PN_{missing}$	R^2	NSCE	ME	SE	RMSE	PME
602	Bay Lake Floridan well	Meas	78.01	96.85	88.07	88.15	18.84	3.37	482	--	--	--	--	--	--	--
		Sim	79.82	95.82	88.43	88.16	16.00	3.61	528	8.7%	0.798	0.757	-0.30	1.63	1.66	8.82%
613	Cocoa D Floridan well	Meas	27.67	43.68	35.73	35.64	16.01	3.05	480	--	--	--	--	--	--	--
		Sim	29.17	42.83	36.01	35.82	13.66	3.07	528	9.1%	0.819	0.807	-0.17	1.30	1.34	8.35%
622	Cocoa P Floridan well	Meas	34.80	52.00	45.32	45.42	17.20	2.97	387	--	--	--	--	--	--	--
		Sim	37.62	53.01	46.19	46.42	15.39	3.14	528	26.7%	0.793	0.791	-0.09	1.25	1.36	7.90%
627	ROMP 101 Floridan well	Meas	92.50	100.54	97.49	97.87	8.04	1.66	363	--	--	--	--	--	--	--
		Sim	92.29	100.30	97.38	97.64	8.01	1.46	528	31.3%	0.800	0.800	0.01	0.68	0.74	9.26%
630	Lake Sawyer Floridan well	Meas	73.42	87.52	81.63	81.83	14.10	2.75	298	--	--	--	--	--	--	--
		Sim	73.27	88.89	82.72	82.94	15.62	3.06	528	43.6%	0.796	0.787	0.03	1.23	1.27	9.00%
632	Palm Lake Dr Floridan well	Meas	58.02	77.42	67.37	67.45	19.40	3.24	292	--	--	--	--	--	--	--
		Sim	59.46	74.87	67.60	67.78	15.41	2.87	528	44.7%	0.686	0.677	0.03	1.69	1.84	9.48%
642	Mascotte Floridan well	Meas	94.16	102.29	99.69	99.91	8.13	1.41	494	--	--	--	--	--	--	--
		Sim	94.24	101.62	99.62	99.80	7.38	1.22	528	6.4%	0.735	0.734	0.05	0.63	0.73	8.93%
643	Mascotte Surficial well	Meas	95.07	103.16	100.35	100.47	8.09	1.45	492	--	--	--	--	--	--	--
		Sim	95.30	102.20	100.35	100.53	6.90	1.14	528	6.8%	0.659	0.658	-0.00	0.68	0.84	10.44%

Table A5-1 . Summary statistics for measured and simulated data for the lake water-level, groundwater-level, and spring-flow models, 1965–2008.—Continued

[Meas, measured; Sim, simulated; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing measured data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	$PN_{missing}$	R^2	NSCE	ME	SE	RMSE	PME
645	Bithlo 1 Floridan well	Meas	29.20	39.96	35.56	35.74	10.76	1.87	515	--	--	--	--	--	--	--
		Sim	30.07	39.94	35.60	35.65	9.87	1.88	528	2.5%	0.804	0.793	-0.01	0.83	0.85	7.93%
648	OR-47 Floridan well	Meas	49.28	68.80	59.14	59.23	19.52	3.71	495	--	--	--	--	--	--	--
		Sim	49.20	68.23	59.05	58.84	19.03	3.62	528	6.3%	0.840	0.836	0.02	1.46	1.50	7.69%
650	Clermont Floridan Rep. well	Meas	74.76	86.01	82.51	83.31	11.25	2.53	286	--	--	--	--	--	--	--
		Sim	74.63	88.39	83.12	83.59	13.76	2.48	528	45.8%	0.895	0.894	0.05	0.80	0.83	7.33%
661	Seminole 125 Floridan well	Meas	32.61	49.44	42.02	42.29	16.83	3.16	497	--	--	--	--	--	--	--
		Sim	33.34	49.57	42.00	42.00	16.23	3.07	528	5.9%	0.808	0.801	0.08	1.36	1.41	8.38%
712	Taft Surficial well	Meas	92.45	97.16	94.60	94.65	4.71	1.08	414	--	--	--	--	--	--	--
		Sim	92.57	96.55	94.62	94.64	3.98	0.87	528	21.6%	0.543	0.540	-0.03	0.57	0.73	15.53%
308	Wekiwa Springs	Meas	51.53	89.85	66.52	66.40	38.32	7.21	281	--	--	--	--	--	--	--
		Sim	50.68	84.04	67.98	67.99	33.36	5.78	528	46.8%	0.563	0.563	0.07	3.58	4.76	12.43%
309	Rock Springs	Meas	38.36	76.90	57.38	57.44	38.54	7.22	297	--	--	--	--	--	--	--
		Sim	42.24	82.11	59.30	59.57	39.87	6.97	528	43.8%	0.725	0.724	-0.08	3.39	3.79	9.85%
311	Miami Springs	Meas	2.90	7.28	5.20	5.22	4.38	0.97	170	--	--	--	--	--	--	--
		Sim	3.82	6.73	5.31	5.30	2.91	0.61	528	67.8%	0.575	0.541	-0.17	0.46	0.66	15.02%

Table A5-1 . Summary statistics for measured and simulated data for the lake water-level, groundwater-level, and spring-flow models, 1965–2008.—Continued

[Meas, measured; Sim, simulated; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing measured data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	$PN_{missing}$	R^2	NSCE	ME	SE	RMSE	PME
312	Sanlando Springs	Meas	8.99	32.90	19.31	18.78	23.91	4.50	190	--	--	--	--	--	--	--
		Sim	12.43	25.97	19.29	19.37	13.54	2.64	528	64.0%	0.376	0.375	0.05	2.18	3.56	14.88%
313	Palm Springs	Meas	2.75	12.00	6.52	6.09	9.25	1.74	188	--	--	--	--	--	--	--
		Sim	4.24	15.62	7.85	6.68	11.38	2.87	528	64.4%	0.627	0.610	-0.00	0.99	1.09	11.73%
314	Starbuck Spring	Meas	8.30	19.57	13.75	13.67	11.27	2.08	185	--	--	--	--	--	--	--
		Sim	10.90	24.95	14.20	13.87	14.05	2.16	528	65.0%	0.405	0.345	0.01	1.42	1.69	14.96%

Appendix 6. User's manual for the Central Florida Artificial Neural Network Decision Support System (CFANN DSS)

Contents

1. Introduction....124
2. CFANN DSS Installation, Removal, and Technical Assistance....124
    2.1 Installation....124
    2.2 Removal....124
    2.3 Technical Assistance....124
3. CFANN DSS Features and Operation....124
    3.1 "Info" Worksheet....125
    3.2 "Controls" Worksheet....125
        3.2.1 Simulation Section....125
        3.2.2 User Settings Section....125
        3.2.3 Output Section....125
        3.2.4 View Sites on Grid Section....125
        3.2.5 Graphical Display Section....128
        3.2.6 Tabular Display Section....128
        3.2.7 Current User Settings Section....128
    3.3 "Rain_UserSetpoints" Worksheet....128
    3.4 "Usage_UserSetpoints" Worksheet....131
        3.4.1 SELECT USER OPTION Section....131
        3.4.2 USAGE BY TYPE/CELL Section....131
        3.4.3 TOTAL USAGE (U-TOT) Section....133
        3.4.4 USER TIME SERIES Section....133
    3.5 "Database-Rain-M," "Database-Hyd-M," and "Database_UsageCells" Worksheets....135
    3.6 "Output-MS2," "Output-MS1," and "Output-UserSettings" Worksheets....135
    3.7 "ReleaseNotes" Worksheet....136
4. Quick Start Guide for the CFANN DSS....136
5. Installing ADMQuestRT Dynamic Link Library....137

Figures

A6-1. "Info" worksheet. 126

A6-2. Simulation section of the "Controls" worksheet used to set the "Start" and "Stop" dates for simulation, initiate a run, or step forward and backward through the data. .126

A6-3. Simulation section of the "Controls" worksheet showing the message that appears when a simulation is interrupted by pressing the "Esc" key twice....126

A6-4. User Settings section of the "Controls" worksheet used to navigate to the worksheets for making user-specified adjustments to rainfall and groundwater use inputs. 127

A6-5. Output section of the "Controls" worksheet used to set graphing and output options. 127

A6-6. View Sites on Grid section of the "Controls" worksheet used to view images of site locations within the Groundwater Use Data Viewer (GUDV) grid displayed when "View selected". 127

A6-7. Graphical Display section of the "Controls" worksheet used to display measured and predicted hydrographs. The site shown in the drop-down box is the site currently displayed. 129

A6-8. Tabular Display section of the "Controls" worksheet showing measured and predicted data for the current simulation time. 129

A6-9. Current User Settings section of the "Controls" worksheet showing user settings for rainfall and usage. 130

A6-10. "Rain_UserSetpoints" worksheet used to make user-specified adjustments to rainfall model inputs. 130

A6-11. The SELECT USER OPTION section of the "Usage_UserSetpoints" worksheet is always visible and is used to select one of five user usage options. 132

A6-12. Partial view of the USAGE BY TYPE/CELL section of the "Usage_UserSetpoints" worksheet used for setting the %-type or bias-type option. 132

A6-13. The TOTAL USAGE (U-TOT) section of the "Usage_UserSetpoints" worksheet used for setting the %-utot and bias-utot options. 134

A6-14. Partial view of USER TIME SERIES section of the "Usage_UserSetpoints" worksheet used for setting the user-ts option. 134

A6-15. Graph of historical and user-specified usage time series displayed using the "Display Usage Chart" button in the USER TIME SERIES section of the "Usage_ UserSetpoints" worksheet. 135

A6-16. Download vcredist_x86.exe. 138

A6-17. If the above message is received, select Yes and continue with the installation. 138

A6-18. If the above message is received, select Repair and continue with the installation ...138

A6-19. Screenshot of command used to register the ADMQuestRT dynamic link library The message in the lower right corner of the above screenshot should appear (with the message title "RegSvr32') when the DLL has been successfully registered 139

1. Introduction

This document describes how to install and operate the Central Florida Artificial Neural Network Decision Support System (CFANN DSS), which is available for download from the U.S. Geological Survey (*http://dx.doi.org/10.3133/sir20145032*). The CFANN DSS is a decision support system (DSS) built around a suite of empirical hydrologic models for the simulation of lake water levels, groundwater levels, and spring flow at discrete sites in central Florida.

2. CFANN DSS Installation, Removal, and Technical Assistance

NOTE: *The CFANN DSS requires a 64-bit Microsoft Windows® operating system. The CFANN DSS was developed using Microsoft Excel 2010® (32-bit installation) and may not work with other versions.*

2.1 Installation

1. Copy the CFANN_DSS folder to your computer hard drive. The user may rename this folder, which contains the following subdirectories and files:
   - ANN-Lakes – a folder containing all ANN models used by CFANN DSS to predict lake water levels;
   - ANN-Springs – a folder containing all ANN models used by CFANN DSS to predict spring flows;
   - ANN-Wells – a folder containing all ANN models used by CFANN DSS to predict well water levels;
   - Images – a folder containing all images used by CFANN DSS to display site locations and Groundwater-Use Data Viewer (GUDV) grid cells;
   - Output – a folder containing all output files written by CFANN DSS;
   - ADMQuestRT.dll – a custom Microsoft Excel® Add-in required to execute the ANN models;
   - CFANN_DSS-yyyymmdd.xlsm[1] – a Microsoft Excel® workbook application; and
   - CFANN_DSS_UserManual-yyyymmdd.pdf – an Adobe® Portable Document Format document of the appendix you are reading now.
2. Install ADMQuestRT.dll by following the directions in section 5 of this user's manual (appendix 6 of the main report).
3. Set the macro security level of Microsoft Excel® to "Disable all macros with notification" using the Developer > Macro Security > Macro Settings menu. The CFANN DSS uses Microsoft Visual Basic for Applications® (VBA) macros for a variety of purposes and must be able to execute them to operate correctly.
4. Open CFANN_DSS-yyyymmdd.xlsm to run simulations. When the Microsoft Excel® security warning reports macros have been disabled, click "Enable Content," otherwise CFANN DSS will not operate properly.

2.2 Removal

Uninstall the ADMQuestRT dynamic link library by entering the following command at the command (DOS) prompt:

regsvr32 /u c:admquestrt\admquestrt.dll

Delete the folder containing the application and its contents. Consider removing the add-ins and reverting to the default Microsoft Excel® security settings.

2.3 Technical Assistance

Please contact Andy O'Reilly of the U.S. Geological Survey at *aoreilly@usgs.gov*, if you have questions or problems with the CFANN DSS. If unavailable, you may contact the U.S. Geological Survey Florida Water Science Center using the "Contact" webpage at *http://fl.water.usgs.gov/*.

3. CFANN DSS Features and Operation

The CFANN DSS runs monthly simulations of the effects of rainfall and groundwater use on water levels at 22 lakes and 23 wells and flows at 6 springs within the Central Florida Coordination Area (CFCA) for the period 1965–2008. Each site is modeled using artificial neural networks (ANNs). A two-stage model architecture is used with the first ANN model set (MS1) predicting a water level or flow using rainfall inputs from rainfall measured at 18 National Oceanic and Atmospheric Administration (NOAA) sites. The residual (difference between the measured and predicted output from MS1) is then modeled using groundwater-use (usage) inputs. Usage inputs—as collated by type, month, and million gallons per day in the GUDV (see appendix 2 of main report)—are calculated from the summation of all usages in the GUDV grid (denoted as USETOT). The two submodels are then summed and denoted as ANN model set 2 (MS2).

[1]yyyymmdd is the version date of the CFANN DSS image to be installed.

Users can make changes to the rainfall and (or) usage inputs to run "what-if" scenarios. Rainfall can be set for each of the 18 NOAA sites as a percentage of historical, or bias (constant offset) in inches per year. Although usage is only incorporated in the ANNs as USETOT, the user can adjust the usage in the following ways:

- Percentage of historical or a constant monthly bias in million gallons per day (MGD/month) for USETOT;
- Percentage of historical or a constant monthly bias in million gallons per day (MGD/month) for each usage type and GUDV grid cell; or
- User-specified time series for any GUDV grid cell or usage type.

The CFANN DSS is opened like any standard Microsoft Excel® workbook. Simply open the CFANN_DSS-yyyymmdd.xlsm file and begin. When the workbook is closed, it may be saved to keep any changes made to the simulation control settings, such as rainfall or usage adjustments made for a particular scenario; otherwise select "Don't Save" to retain the original simulation control settings. The CFANN DSS and its graphical user interface (GUI) comprise a number of worksheets. Detailed descriptions of each worksheet are provided below in section 3, followed in section 4 by an outline of the basic steps, or "Quick Start" guide, to facilitate running a simulation.

3.1 "Info" Worksheet

The "Info" worksheet is automatically displayed when the CFANN DSS is first loaded (fig. A6-1). The worksheet shows a map of the study area and gives the application's version date and the contact information of its developers.

3.2 "Controls" Worksheet

The "Controls" worksheet is the GUI component that lets the user set up and run simulations. The worksheet provides numerical and streaming graphical information that can be observed during simulations or when incrementally stepping through time. This allows the user to examine in detail specific periods and behaviors of interest. The worksheet is divided into multiple sections that are described below.

3.2.1 Simulation Section

The Simulation section (fig. A6-2) is used to set "Start" and "Stop" dates for a simulation using the scroll bars. The scroll bars can change the date in monthly (clicking on arrows) or yearly (clicking inside the scroll bar) increments. The end date must be more recent than the start date. The "Sim Date" text box indicates the time stamp that is providing the current input values to the ANN models. The "<<Step" and "Step>>" buttons move the current time stamp backward or forward one time step with each click. The "Sim Time=Start" button sets the current time stamp to the simulation "Start" date. The "RUN" button will start and run a simulation between the dates indicated by the simulation "Start" and "End" dates. A simulation may be stopped at any time during an execution by pressing the "Esc" key twice, after which a message will be written in the Simulation section (fig. A6-3).

3.2.2 User Settings Section

The User Settings section (fig. A6-4) contains the "App Path" textbox that displays the location of the CFANN DSS application files. The "Set User Rain" and "Set User Usage" buttons take the user to the appropriate worksheet for setting the user-specified values.

3.2.3 Output Section

The Output Section (fig. A6-5) is used to turn on and off the display of graphs and writing of output files. Additional details on graph and output options are provided below.

1. Graph Options: If graphs are turned off ("Graphs ON" check box is unchecked), no graphical display of the output is available within the application during a simulation run or after run completion. If graphs are turned on ("Graphs ON" check box is checked), the user can select a site to be dynamically displayed during the simulation run. In addition, the user can choose to display model set 1 (MS1), which are the rainfall models, output on the graph. Additional details on graphical display are provided in section 3.2.5.
2. Output Options: If "Write Output" is turned off, no output will be saved for the run. If turned on, the user also can select which type of sites to output. The options are all wells, all lakes, and all springs. The default setting is for all sites to be included in the output.

3.2.4 View Sites on Grid Section

The View Sites on Grid section (fig. A6-6) allows users to view an image of the GUDV grid and the locations of the wells, lakes, and springs within the grid by selecting the associated button. The image will display in a separate window that can be moved around the screen and closed by clicking on the X in the top right corner.

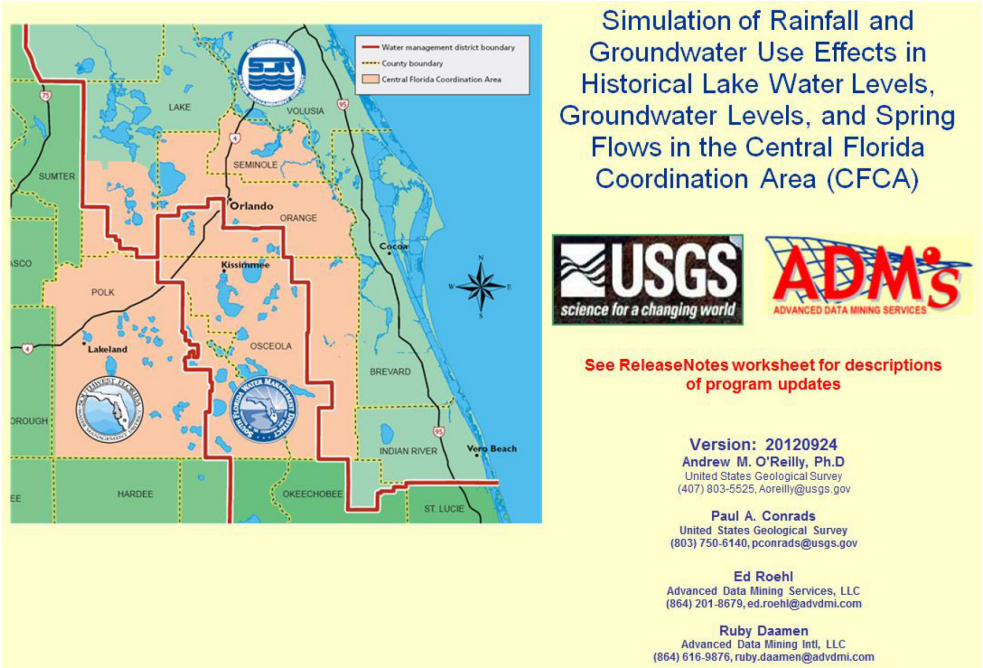


Figure A6-1. "Info" worksheet.

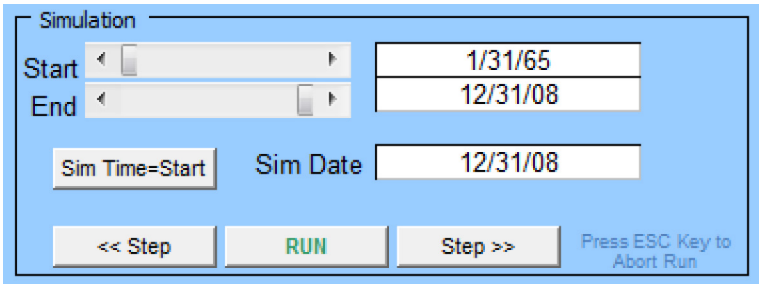


Figure A6-2. Simulation section of the "Controls" worksheet used to set the "Start" and "Stop" dates for simulation, initiate a run, or step forward and backward through the data.

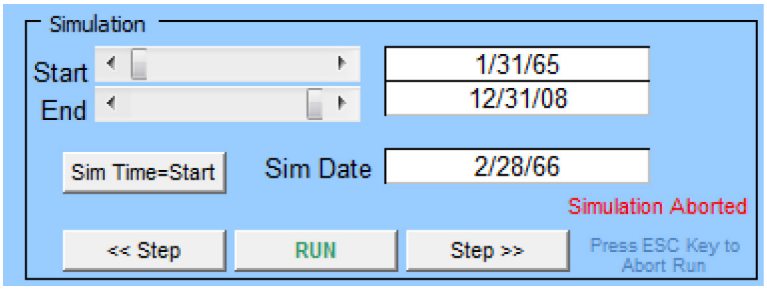


Figure A6-3. Simulation section of the "Controls" worksheet showing the message that appears when a simulation is interrupted by pressing the "Esc" key twice.

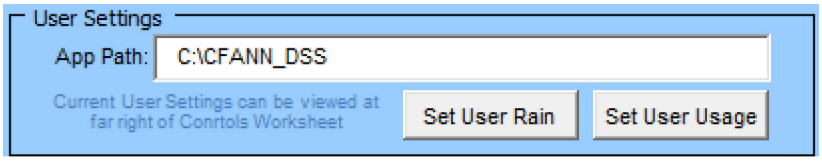


Figure A6-4. User Settings section of the "Controls" worksheet used to navigate to the worksheets for making user-specified adjustments to rainfall and groundwater use inputs.

Figure A6-5. Output section of the "Controls" worksheet used to set graphing and output options.

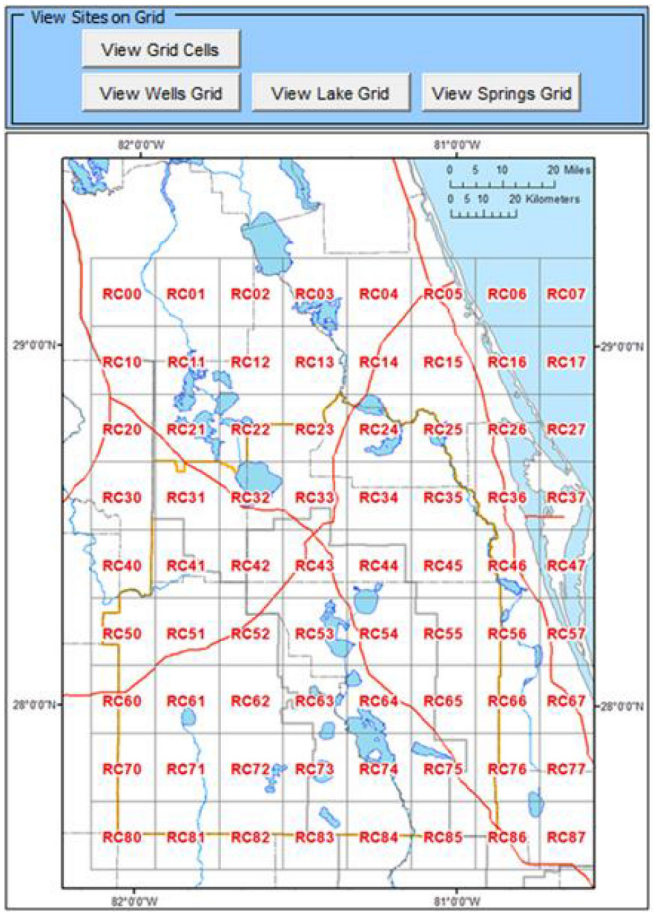


Figure A6-6. View Sites on Grid section of the "Controls" worksheet used to view images of site locations within the Groundwater Use Data Viewer (GUDV) grid displayed when "View selected".

3.2.5 Graphical Display Section

The Graphical Display section (fig. A6-7) is used to create hydrographs for a site. If "Graphs ON" is checked in the Output section (see section 3.2.3) the focus will move to the Graphical Display section when a simulation is run. The user-selected site displayed in the drop-down box will be dynamically displayed during a simulation. Upon simulation completion the user can scroll forward and backward through the sites using the spin control or select specific sites using the drop-down box. The time series displayed on the graph can be selected by using the check boxes located below the graph:

1. Primary Axis
   - Display m: Monthly averaged historical measured value
   - Display pm: Predicted output using historical measured inputs
   - Display pu: Predicted output using user-specified inputs
   - Display m + (pu – pm): Sum of measured value and the difference between user-specified and measured output
   - Display pm (MS1): Predicted output using historical inputs for model set 1 (rainfall models). MS1 outputs will be displayed if the user has selected option to include them in the Output section (see section 3.2.3).
   - Display pu (MS1): Predicted output using user inputs for model set 1 (rainfall models)
2. Secondary Axis
   - Display pu – pm: Difference between user-specified and measured output
   - Display pu – pm (MS1): Difference between user-specified and measured output for model set 1 (rainfall models)

3.2.6 Tabular Display Section

The Tabular Display section (fig. A6-8) has three tables showing measured and predicted output for wells, lakes, and springs. Moving the mouse over the red caret immediately above and to the right of a header in any of these tables will provide a description of the header variable. Similarly, moving the mouse over the red caret immediately above and to the right of the ANN model label will provide a short name of the well, lake, or spring. In addition, double clicking on any ANN model label will cause a pop-up window to appear displaying an image of the site location within the GUDV grid. To facilitate locating individual sites, images shown in this pop-up window provide a close-up view compared to the study area view shown in the View Sites on Grid section (see section 3.2.4). If "Graphs ON" is unchecked in the Output section (see section 3.2.3), the focus will move to the Tabular Display section when a simulation is run. Each table dynamically displays the following data for the current simulation time:

- meas (m): Monthly averaged historical measured value
- pred-meas (pm): Predicted output using historical measured inputs
- pred-user (pu): Predicted output using user-specified inputs
- pu – pm: Difference between user-specified and measured output
- MEAS + d(pu –pm): Sum of measured value and the difference between user-specified and measured output

3.2.7 Current User Settings Section

The Current User Settings section (fig. A6-9) displays the current user-specified settings for rainfall and usage. The user-specified values are listed in the columns labeled "User SP" followed by a column displaying the appropriate units. No changes can be made in this area. The user must go to the User Settings section to make any changes to the settings (see section 3.2.2).

3.3 "Rain_UserSetpoints" Worksheet

All adjustments to user rainfall settings are made on the "Rain_UserSetpoints" worksheet (fig. A6-10). Each of the 18 NOAA sites can be adjusted individually. In addition, all sites can be adjusted globally to the same user setting. Follow the steps below to adjust sites individually or globally.

1. Select User Option using the appropriate "User Opt" scroll bar:
   - %: Percentage of historical value
   - bias: Constant offset added or subtracted from historical value
2. Set user setting using the "%" scroll bar or the "bias in/yr" scroll bar based on "User Opt" selected in step 1. To make a global change to all sites, use the "Set ALL NOAA SITES" scroll bars followed by clicking the "Set ALL NOAA Sites" button to update the setting for each site. To make changes to individual sites, use the scroll bars beneath the respective NOAA site numbers. Refer to the map to the right of the scroll bars for site locations, which are grouped according to a time-series cluster analysis based on

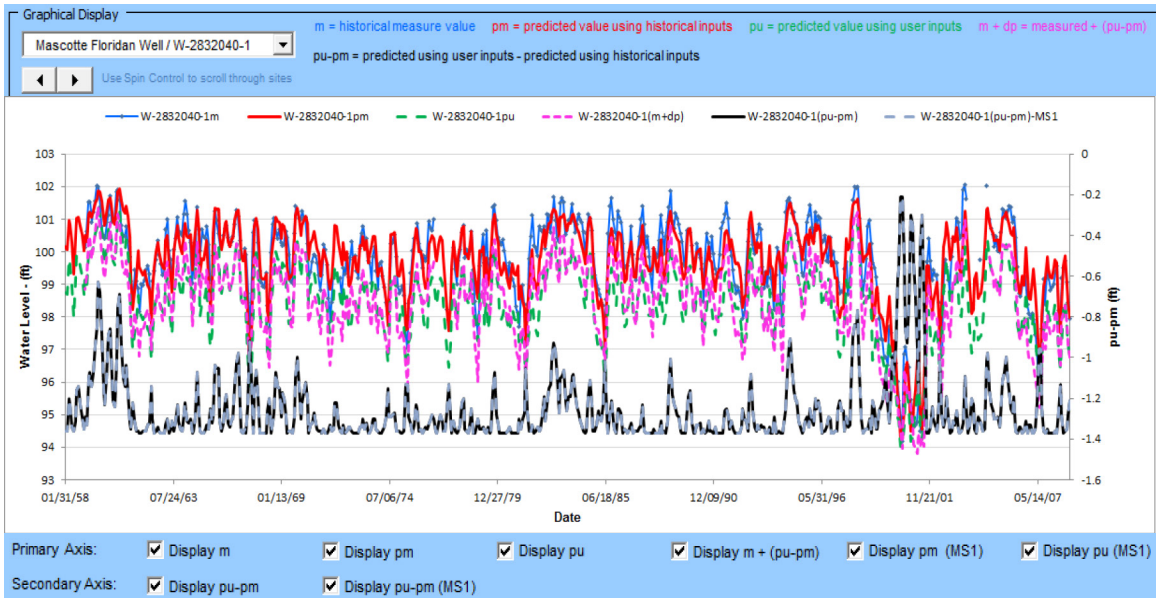
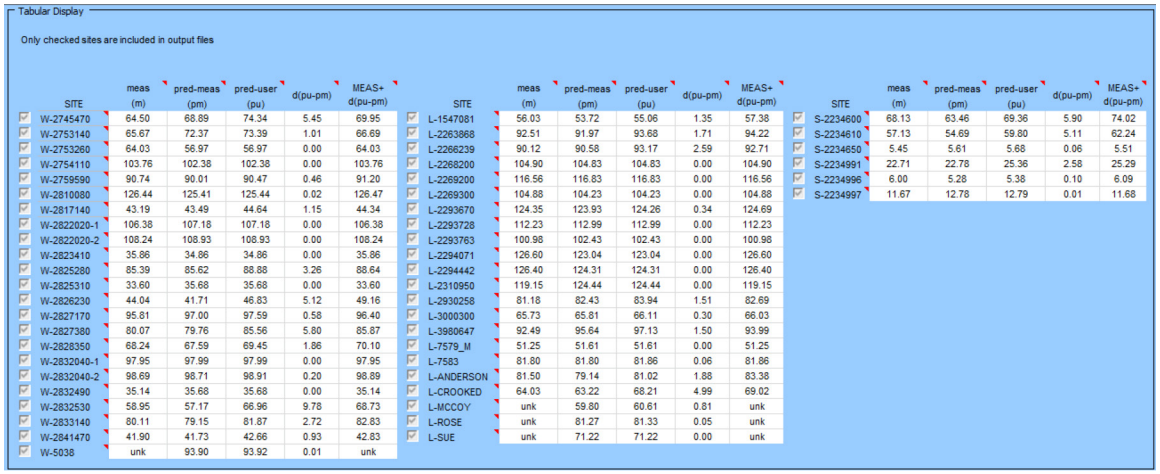


Figure A6-7. Graphical Display section of the "Controls" worksheet used to display measured and predicted hydrographs. The site shown in the drop-down box is the site currently displayed.



Tabular Display

Only checked sites are included in output files

SITE	meas (m)	pred-meas (pm)	pred-user (pu)	d(pu-pm)	MEAS+ d(pu-pm)
W-2745470	64.50	68.89	74.34	5.45	69.95
W-2753140	65.67	72.37	73.39	1.01	66.69
W-2753260	64.03	56.97	56.97	0.00	64.03
W-2754110	103.76	102.38	102.38	0.00	103.76
W-2759590	90.74	90.01	90.47	0.46	91.20
W-2810080	126.44	125.41	125.44	0.02	126.47
W-2817140	43.19	43.49	44.64	1.15	44.34
W-2822020-1	106.38	107.18	107.18	0.00	106.38
W-2822020-2	108.24	108.93	108.93	0.00	108.24
W-2823410	35.86	34.86	34.86	0.00	35.86
W-2825280	85.39	85.62	88.88	3.26	88.64
W-2825310	33.60	35.68	35.68	0.00	33.60
W-2826230	44.04	41.71	46.83	5.12	49.16
W-2827170	95.81	97.00	97.59	0.58	96.40
W-2827380	80.07	79.76	85.56	5.80	85.87
W-2828350	68.24	67.59	69.45	1.86	70.10
W-2832040-1	97.95	97.99	97.99	0.00	97.95
W-2832040-2	98.69	98.71	98.91	0.20	98.89
W-2832490	35.14	35.68	35.68	0.00	35.14
W-2832530	58.95	57.17	66.96	9.78	68.73
W-2833140	80.11	79.15	81.87	2.72	82.83
W-2841470	41.90	41.73	42.66	0.93	42.83
W-5038	unk	93.90	93.92	0.01	unk

SITE	meas (m)	pred-meas (pm)	pred-user (pu)	d(pu-pm)	MEAS+ d(pu-pm)
L-1547081	56.03	53.72	55.06	1.35	57.38
L-2263868	92.51	91.97	93.68	1.71	94.22
L-2266239	90.12	90.58	93.17	2.59	92.71
L-2268200	104.90	104.83	104.83	0.00	104.90
L-2269200	116.56	116.83	116.83	0.00	116.56
L-2269300	104.88	104.23	104.23	0.00	104.88
L-2293670	124.35	123.93	124.26	0.34	124.69
L-2293728	112.23	112.99	112.99	0.00	112.23
L-2293763	100.98	102.43	102.43	0.00	100.98
L-2294071	126.60	123.04	123.04	0.00	126.60
L-2294442	126.40	124.31	124.31	0.00	126.40
L-2310950	119.15	124.44	124.44	0.00	119.15
L-2930258	81.18	82.43	83.94	1.51	82.69
L-3000300	65.73	65.81	66.11	0.30	66.03
L-3980647	92.49	95.64	97.13	1.50	93.99
L-7579_M	51.25	51.61	51.61	0.00	51.25
L-7583	81.80	81.80	81.86	0.06	81.86
L-ANDERSON	81.50	79.14	81.02	1.88	83.38
L-CROOKED	64.03	63.22	68.21	4.99	69.02
L-MCCOY	unk	59.80	60.61	0.81	unk
L-ROSE	unk	81.27	81.33	0.05	unk
L-SUE	unk	71.22	71.22	0.00	unk

SITE	meas (m)	pred-meas (pm)	pred-user (pu)	d(pu-pm)	MEAS+ d(pu-pm)
S-2234600	68.13	63.46	69.36	5.90	74.02
S-2234610	57.13	54.69	59.80	5.11	62.24
S-2234650	5.45	5.61	5.68	0.06	5.51
S-2234991	22.71	22.78	25.36	2.58	25.29
S-2234996	6.00	5.28	5.38	0.10	6.09
S-2234997	11.67	12.78	12.79	0.01	11.68

Figure A6-8. Tabular Display section of the "Controls" worksheet showing measured and predicted data for the current simulation time.

Display only of current rain and usage user settings. To change current settings click on "Set User Rain" or "Set User Usage"

View Usage Grid

U-TOT			
User SP	100.00	Units	%

Current User Rain Settings

Site	User Opt	User SP	
369	%	70	%
478	%	70	%
1163	%	70	%
1641	%	70	%
2229	%	70	%
3137	%	70	%
4502	%	70	%
4625	%	70	%
5076	%	70	%
5612	%	70	%
5973	%	70	%
6414	%	70	%
6628	%	70	%
7205	%	70	%
7982	%	70	%
8942	%	70	%
9401	%	70	%
9707	%	70	%

Current User Usage Settings

Row/Column	Agriculture		Citrus		Golf		Water User		Drainage		Phosphate	
	User SP	Units	User Opt	User SP	User Opt	User SP	User Opt	User SP	User Opt	User SP	User Opt	User SP
RC00												
RC01												
RC02	%-utot	MGD										
RC03	%-utot	MGD										
RC04												
RC05												
RC06												
RC07												
RC10	%-utot	MGD	%-utot	MGD	%-utot	MGD						
RC11	%-utot	MGD	%-utot	MGD	%-utot	MGD						
RC12	%-utot	MGD	%-utot	MGD	%-utot	MGD						
RC13	%-utot	MGD	%-utot	MGD								
RC14							%-utot	MGD				
RC15												
RC16												
RC17												
RC20	%-utot	MGD	%-utot	MGD	%-utot	MGD	%-utot	MGD				
RC21	%-utot	MGD	%-utot	MGD	%-utot	MGD	%-utot	MGD				

Figure A6-9. Current User Settings section of the "Controls" worksheet showing user settings for rainfall and usage.

NOTES

Use controls below to set the user rainfall for each site. Rainfall can be set for all sites or for each site. User options are:

1. % historical - allowable range is +/- 50% of historical values

OR

2. Biased up to +/- 15 in/yr

Return to Controls Worksheet

Set All NOAA SITES	
User Opt	%
%	70
bias in/yr	0.0

Set ALL NOAA Sites

NOAA SITES

Group 1	2229		6628		7982		8942	
User Opt		%		%		%		%
%		70		70		70		70
bias in/yr		0.0		0.0		0.0		0.0
Group 2	**3137**		**4502**		**4625**		**5612**	
User Opt		%		%		%		%
%		70		70		70		70
bias in/yr		0.0		0.0		0.0		0.0
Group 3	**1163**		**1641**		**5076**		**6414**	
User Opt		%		%		%		0
%		70		70		70		70
bias in/yr		0.0		0.0		0.0		0.0
Group 4	**369**		**478**		**5973**		**7205**	
User Opt		%		%		%		%
%		70		70		70		70
bias in/yr		0.0		0.0		0.0		0.0

9401		9707	
	%		%
	70		70
	0.0		0.0

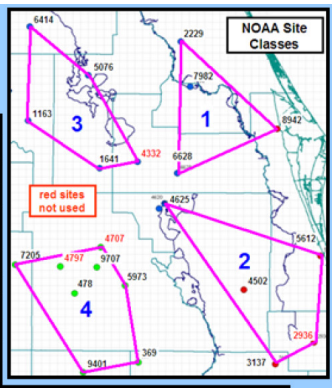


Figure A6-10. "Rain_UserSetpoints" worksheet used to make user-specified adjustments to rainfall model inputs.

1,440-day moving window averages (see figure 24 of the main report). Allowable ranges are

- For %: minimum of –50%, maximum of 50%. The scroll bars can change the percentage in 1% (clicking on arrows) or 10% (clicking inside the scroll bar) increments.
- For bias: minimum of –15 in/yr, maximum of 15 in/yr. The scroll bars can change the bias in 1 in/yr (clicking on arrows) or 5 in/yr (clicking inside the scroll bar) increments.

3. Once all rainfall user-specified settings are complete, select the "Return to Controls Worksheet" button to return control to the "Controls" worksheet. User-specified settings made in step 2 will be displayed in Current User Settings section of the "Controls" worksheet (see section 3.2.7). It is important to note that only the settings for the user option selected in step 1 will be displayed in this section and used by the CFANN DSS.

3.4 "Usage_UserSetpoints" Worksheet

All adjustments to the usage user settings are made on the "Usage_UserSetpoints" worksheet. The worksheet is divided into four sections. A detailed description of each worksheet section is provided below.

3.4.1 SELECT USER OPTION Section

The SELECT USER OPTION section on the far left of the "Usage_UserSetpoints" worksheet remains visible and is used to set the user usage option (User Opt) by using the drop-down menu as well as to view the GUDV (usage) grid and return to the "Controls" worksheet (fig. A6-11). Once a user usage selection is made, focus will move to one of the three sections in the right portion of the worksheet. Each of the four sections of the worksheet contains important notes that should be consulted on how to make user-specified adjustments. The user usage options and their corresponding worksheet section are listed below.

- %-type: USAGE BY TYPE/CELL
- bias-type: USAGE BY TYPE/CELL
- %-utot: TOTAL USAGE (U-TOT)
- bias-utot: TOTAL USAGE (U-TOT)
- user-ts: USER TIME SERIES

It is important to note that while the USAGE BY TYPE/CELL and USER TIME SERIES sections allow the user to adjust usage inputs by type or individual GUDV cell, all usage data are summed into a single total usage time series for signal decomposition and input to the groundwater-use ANN models. The ability to adjust usage spatially is provided only to facilitate conceptualization and implementation of user-specified changes. Identical predictions can be made globally by adjusting total summed usage using the TOTAL USAGE (U-TOT) section, although doing so for more complex spatial adjustments would require preprocessing of the appropriate %-utot and bias-utot values. The temporal adjustment of usage, whether made globally or only for specific GUDV cells, must be made using the USER TIME SERIES section.

After all user-specified usage settings have been made in one of the three sections described below (see sections 3.4.2, 3.4.3, and 3.4.4), select the "Return to Controls Worksheet" button to return control to the "Controls" worksheet. User-specified settings will be displayed in the Current User Settings section of the "Controls" worksheet (see section 3.2.7). It is important to note that only the settings for one of the five user options selected in the SELECT USER OPTION section will be displayed in the Current User Settings section of the "Controls" worksheet and used by the CFANN DSS.

3.4.2 USAGE BY TYPE/CELL Section

The USAGE BY TYPE/CELL section (fig. A6-12) is used to adjust groundwater-use model inputs. This section contains a table that is organized into rows and columns corresponding to GUDV grid cells and usage types, respectively. Each GUDV grid cell is listed as a separate row and labeled with a unique identifier, RC##, where the first # is the GUDV grid row number (from 0 to 8) and the second # is the GUDV grid column number (from 0 to 7). Click on the "View Usage Grid" button in the SELECT USER OPTION section (fig. A6-11) to display an image of the GUDV grid. The table contains six columns with the following headings representing specific usage types:

- AG Grid Cells: Groundwater withdrawals for non-citrus agricultural irrigation
- Citrus Grid Cells: Groundwater withdrawals for citrus irrigation
- Golf Grid Cells: Groundwater withdrawals for golf course irrigation
- WU Grid Cells: Groundwater withdrawals public-supply water utilities
- Drain Grid Cells: Groundwater recharge for drainage wells
- PHOS Grid Cells: Groundwater withdrawals for phosphate mining (ore extraction and chemical plants)

The table has a blank gray background for those GUDV grid cells that contain no historical usage for a given usage type. The table has scroll bars for each usage type that has historical

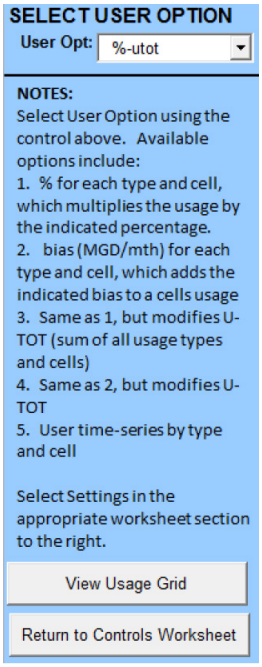


Figure A6-11. The SELECT USER OPTION section of the "Usage_UserSetpoints" worksheet is always visible and is used to select one of five user usage options.

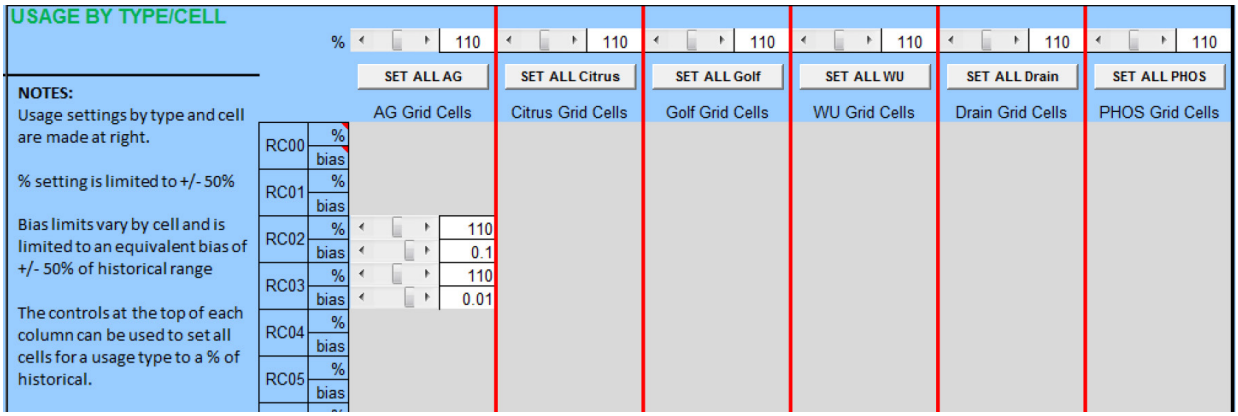


Figure A6-12. Partial view of the USAGE BY TYPE/CELL section of the "Usage_UserSetpoints" worksheet used for setting the %-type or bias-type option.

usage data for a given GUDV grid cell. Adjustments can be a user percentage or bias specified in the SELECT USER OPTION section (see section 3.4.1) and may be made by usage type or for individual GUDV grid cells as explained below.

1. %-type: Percentage of historical value. To set all cells for a given usage type, first use the scroll bar at the top of each column to change the percentage in 1% (clicking on arrows) or 10% (clicking inside the scroll bar) increments, and second click the "SET ALL" button below the scroll bar to apply the change to the cells below (fig. A6-12). To set an individual cell, use the scroll bar in the row corresponding to the desired GUDV grid cell identifier and the column corresponding to the desired usage type. The percentage can be changed in 1% (clicking on arrows) or 10% (clicking inside the scroll bar) increments. Note that individual adjustments can only be made for those GUDV cells with historical data for a given type.

2. bias-type: Constant offset added or subtracted from historical value. To set an individual cell, use the scroll bar in the row corresponding to the desired GUDV grid cell identifier and the column corresponding to the desired usage type. Because the magnitude of usage varies substantially from cell to cell, the bias adjustment increments vary by cell, and changes in bias to all cells of the same usage type are not permitted. The scroll bar effects a change in bias in increments equivalent to 5% of historical range (clicking on arrows) or 50% of historical range (clicking inside the scroll bar). The minimum and maximum bias limits are equivalent to –50% and 50%, respectively, of the historical range (maximum usage minus minimum usage) for that cell.

3.4.3 TOTAL USAGE (U-TOT) Section

Regardless of which worksheet section is used to make usage adjustments, all usage data are summed into a single total usage time series (U-TOT) for signal decomposition and input to the groundwater-use ANN models. The TOTAL USAGE (U-TOT) section (fig. A6-13) is used for making direct adjustments to this total summed usage time series as described below.

1. %-type: Percentage of historical value. Use the scroll bar to change the percentage in 1% (clicking on arrows) or 10% (clicking inside the scroll bar) increments. The minimum and maximum percentage limits are –50% and 50%, respectively.

2. bias-type: Constant offset added or subtracted from historical value. Use the scroll bar to change the bias in increments equivalent to 5% of historical (clicking on arrows) or 50% of historical (clicking inside the scroll bar). The minimum and maximum bias limits are equivalent to –50% and 50%, respectively, of the historical range (maximum U-TOT minus minimum U-TOT).

3.4.4 USER TIME SERIES Section

The USER TIME SERIES section (fig. A6-14) is used to input any user-specified usage data. This section contains a table with rows representing months (708) from January 1950 through December 2008 and columns representing each usage type (6) for each GUDV grid cell (72), for a total of 305,856 data values. Table headings are labeled consistent with the naming convention used for the USAGE BY TYPE/CELL section (see section 3.4.2). For example, the column labeled "RC34_Drain" contains drainage well recharge data for row 3 and column 4 of the GUDV grid. Date stamps used in the table correspond to the last day of each month, indicating that data represent monthly values in million gallons per day per month. Groundwater withdrawals are represented by positive values and groundwater recharge is represented by negative values. It is important to note that date stamps are provided in this table from January 1950 through December 1956, but these data are not currently used by the CFANN DSS. Data for citrus irrigation, non-citrus agricultural irrigation, and golf course irrigation prior to 1957 were not available for this study, and estimates for public supply and phosphate mining become increasingly uncertain for older data during the earlier historical period; therefore, the current version of the CFANN DSS uses usage data from only 1957 through 2008.

Unlike the USAGE BY TYPE/CELL section, by using the user-ts option (see section 3.4.1) all cells can be populated in the USER TIME SERIES section. A column header with a gray background indicates no historical usage was available for that GUDV grid cell, but users can input usage data in these cells and these data will be used by the ANN models. Data can be entered by typing or by copying and pasting from another spreadsheet. To restore historical usage data for a given usage type, select the appropriate command button. For example, "Copy Historical AG" will replace any user time series in the AG cells with the historical data (fig. A6-14). This allows for the use of a combination of historical and user time-series data. It is important to note that while much latitude is provided to enter data in this section, it is the responsibility of the user to check the correctness of the user-defined input data and its synchronization with the given date stamps.

Time-series graphs of historical and user-specified usage can be displayed in a pop-up window by clicking the "Display Usage Chart" button (fig. A6-15). The graphical display is blank until a view option is selected. The CFANN DSS currently uses only the total summed usage (USETOT) for input to the groundwater-use ANN models, but plots can be displayed of measured and user-specified usages by type. The plots apply only to the data entered in the USER TIME

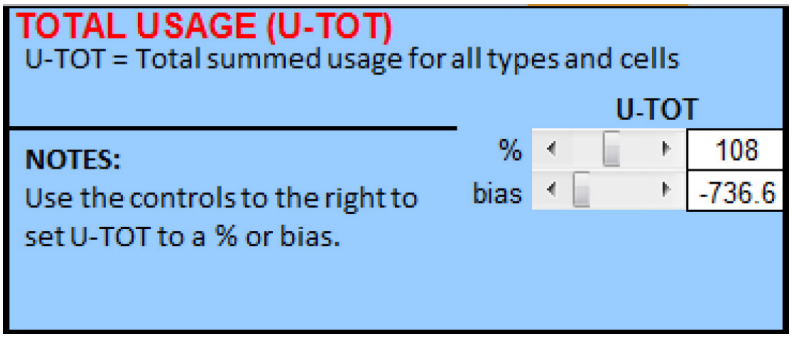
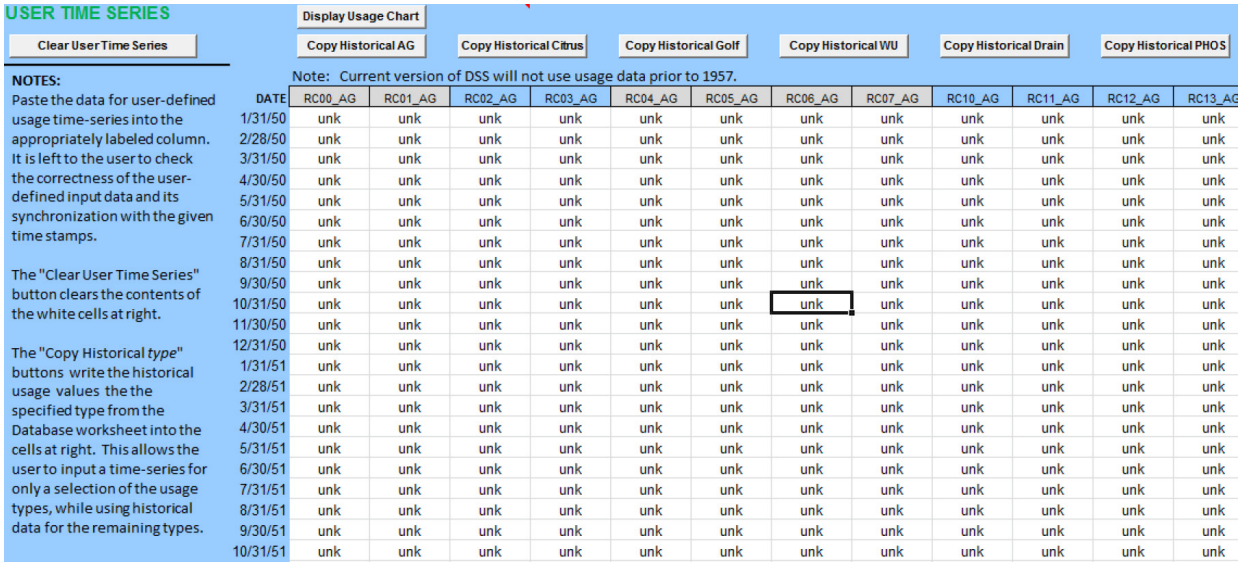


Figure A6-13. The TOTAL USAGE (U-TOT) section of the "Usage_UserSetpoints" worksheet used for setting the %-utot and bias-utot options.



DATE	RC00_AG	RC01_AG	RC02_AG	RC03_AG	RC04_AG	RC05_AG	RC06_AG	RC07_AG	RC10_AG	RC11_AG	RC12_AG	RC13_AG
1/31/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
2/28/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
3/31/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
4/30/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
5/31/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
6/30/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
7/31/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
8/31/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
9/30/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
10/31/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
11/30/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
12/31/50	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
1/31/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
2/28/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
3/31/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
4/30/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
5/31/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
6/30/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
7/31/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
8/31/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
9/30/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk
10/31/51	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk	unk

Figure A6-14. Partial view of USER TIME SERIES section of the "Usage_UserSetpoints" worksheet used for setting the user-ts option.

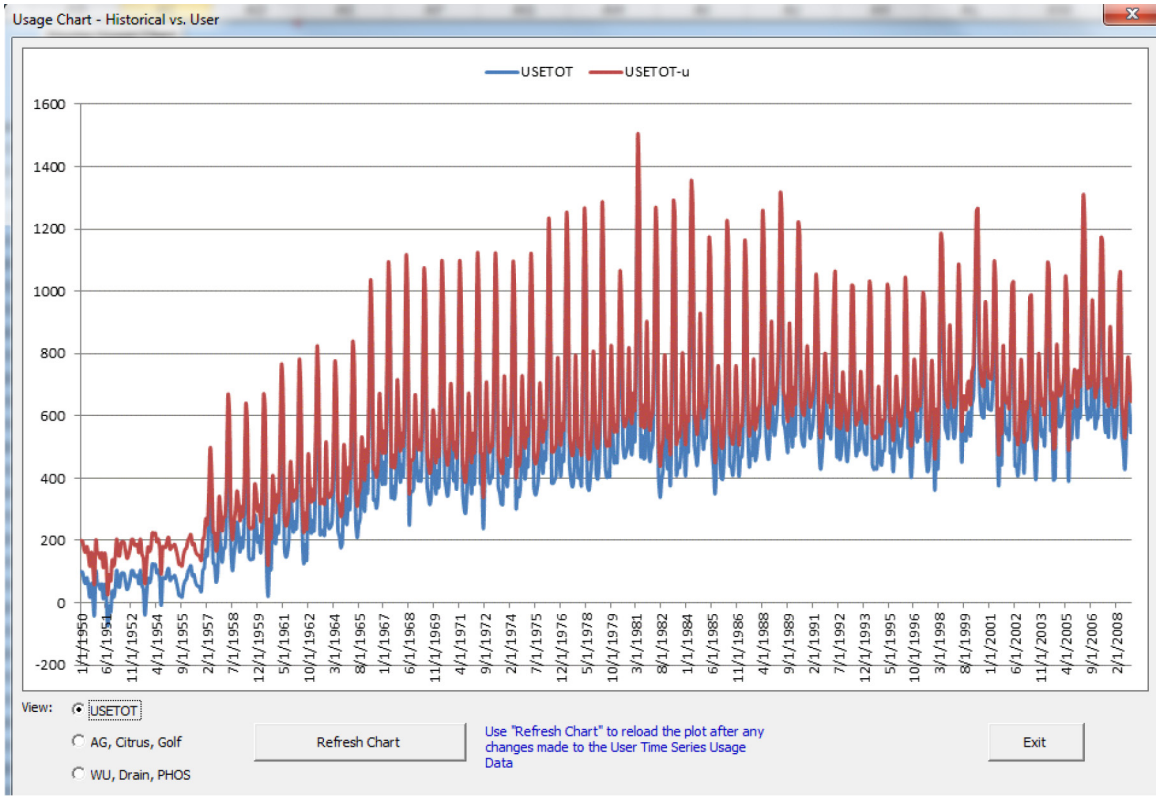


Figure A6-15. Graph of historical and user-specified usage time series displayed using the "Display Usage Chart" button in the USER TIME SERIES section of the "Usage_UserSetpoints" worksheet.

SERIES section. If changes are made to the time series, select the "Refresh Chart" button to update the graphical display. Select the "Exit" command to close the graphical display window.

3.5 "Database-Rain-M," "Database-Hyd-M," and "Database_UsageCells" Worksheets

The "Database-Rain-M," "Database-Hyd-M," and "Database_UsageCells" worksheets contain the monthly historical data for rainfall at the 18 NOAA sites, water level or flow at the 51 study sites, and usage by type and GUDV grid cell, respectively. Rainfall (in inches per day), water level (in feet NGVD 29), and flow (in cubic feet per second) data were converted from daily to monthly values by calculating monthly averages. Monthly average rainfall is included for each gage for 30-; 90-; 270-; 630-; 1,440-; and 2,250-day moving window averages of the filled, daily data. Usage data (in million gallons per day) was originally available in monthly intervals and was stored as such after spatial aggregation using the GUDV grid.

3.6 "Output-MS2," "Output-MS1," and "Output-UserSettings" Worksheets

The "Output-MS2," "Output-MS1," and "Output-UserSettings" worksheets contain simulation time-series data from each run of the CFANN DSS when the "Write Output" check box is selected (see section 3.2.3). Data are written from the "Start" date to the "End" date specified in the Simulation

section (see section 3.2.1). Data in each worksheet are also written to three corresponding ASCII files stored in the "Output" folder (see section 2.1). Data in each of these worksheets are described below.

1. The "Output-MS2" worksheet contains the output for the combined rainfall and groundwater-use models for each site. The corresponding ASCII file created will be named "CFCA-yyyymmdd-hhmm.txt" where "yyyymmdd-hhmm" is the year, month, day, hour, and minute when the file was created. Each column heading consists of the ANN model label (for example, W-2745470) denoting the site concatenated with one of the following suffixes denoting the type of data [for example, W-2745470(m+dp)]:
   - m: Monthly averaged historical measured value
   - pm: Predicted output using historical measured inputs
   - pu: Predicted output using user-specified inputs
   - (pu–pm): Difference between user-specified and measured output
   - (m+dp): Sum of measured value and the difference between user-specified and measured output
2. The "Output-MS1" worksheet contains the output for model set 1 (MS1), which are the rainfall models, for each site. The corresponding ASCII file created will be named "CFCA-yyyymmdd-hhmm-MS1.txt" where "yyyymmdd-hhmm" is the year, month, day, hour, and minute when the file was created. Each column heading consists of the ANN model label (for example W-2745470) denoting the site concatenated with one of the following suffixes denoting the type of data (for example, W-2745470(m+dp)-MS1):
   - m-MS1: Monthly averaged historical measured value
   - pm-MS1: Predicted output using historical measured inputs for rainfall model
   - pu-MS1: Predicted output using user-specified inputs for rainfall model
   - (pu–pm)-MS1: Difference between user-specified and measured output for rainfall model
   - (m+dp)-MS1: Sum of measured value and the difference between user-specified and measured output for rainfall model
3. The "Output-UserSettings" worksheet contains user setting information for rainfall and usage. The corresponding ASCII file created will be named "CFCA-yyyymmdd-hhmm-UserSettings.txt" where "yyyymmdd-hhmm" is the year, month, day, hour, and minute when the file was created. The specific rainfall and usage information are as follows:
   - For each NOAA rain gage, monthly average of 30-day moving window average of daily historical measured data (from the "Database-Rain-M" worksheet) and user-specified rainfall data are provided. Each column heading contains the NOAA site number with an "m" suffix denoting historical measured data or a "u" suffix denoting user-specified data (for example, R369-FA30M-m).
   - For total summed usage (USETOT), historical measured total summed usage (column heading USETOT-m) and user-specified total summed usage (column heading USETOT-u) are provided.

3.7 "ReleaseNotes" Worksheet

The "ReleaseNotes" worksheet maintains a history of CFANN DSS updates.

4. Quick Start Guide for the CFANN DSS

The following steps outline the basic procedure to run a CFANN DSS simulation. Please refer to section 3 for more detailed instructions.

1. Select the "Controls" worksheet.
2. Use the scroll bars on the "Controls" worksheet to set a start date and end date for the simulation run.
3. Use the "Set User Rain" command button to go to the "Rain_UserSetpoints" worksheet. Use the scroll bars to set the user rain as
   a. percentage of historical; or
   b. bias (constant offset in inches per year)

Each rain site can be changed individually OR all sites can be set to a common value. Return to the Controls worksheet by selecting the "Return to Controls Worksheet" command button.

4. Use the "Set User Usage" command button to go to the "Usage_UserSetpoints" worksheet and set user usage values.
   a. Use the drop-down combo box to select a user option. Once selected the screen focus will move to that area of the worksheet. Options are
      i. %-type (percentage of historical by usage type);

ii. bias-type (constant offset in million gallons per day by usage type);

iii. %-utot (percentage of historical for total usage);

iv. bias-utot (constant offset in million gallons per day for total usage); or

v. user-ts (user-specified time series in million gallons per day per month).

b. Use the scroll bars to set usage values for options i–iv. For option v, paste the time series in the appropriate columns; a combination of historical time series and user-specified time series can be used.

Return to the Controls worksheet by selecting the "Return to Controls Worksheet" command button.

5. Turn graphs On or Off using the "Graphs ON" check box. The drop-down combo box to the right of the check box can be used to select which site to display during the simulation if graphs are turned on.

6. Turn Write Output On or Off using the "Write Output" checkbox. Use the "Write Wells Output," "Write Lakes Output," and "Write Springs Output" check boxes to select which sites to include in the output.

7. Start run by selecting the "Run" Command button.

8. Upon run completion, three files will be saved to the Output subdirectory if Write Output is turned On. The three files are

a. CFCA-yyyymmdd-hhmm.txt – Output for each location selected for output which includes for each monthly time stamp

i. Monthly average historical measured data (m)

ii. Predicted using measured inputs (pm)

iii. Predicted using user inputs (pu)

iv. Predicted measured minus predicted user (pu–pm)

v. Measured + pu-pm

b. CFCA-yyyymmdd-hhmm-MS1.txt – Same as above, but only includes the rainfall models (MS1) predictions.

c. CFCA-yyyymmdd-hhmm-UserSettings.txt

i. Measured (m) and user-specified (u) monthly average of 30-day moving window average of daily historical measured data for each NOAA rain station

ii. Measured (m) and user-specified (u) USETOT (sum of all usages for a given month)

5. Installing ADMQuestRT Dynamic Link Library

Follow the steps below to install the ADMQuestRT dynamic link library (DLL).

1. Install vcredist_x86.exe by downloading (fig. A6-16) from *http://www.microsoft.com/download/en/details.aspx?id=26999*. Click Yes if you get the prompt, "Do you want to allow the following program to make changes to this computer?" (fig. A6-17). This ensures that the proper runtime components of Visual C++ 2010 are installed.

2. If the install prompts to repair or remove, select repair; this will re-copy the 32 bit DLLs (fig. A6–18).

3. Create the following folder on your c: drive: c:\ADMQuestRT

4. Copy the file ADMQuestRT.dll into the newly create folder.

5. Run the following command as administrator (fig. A6-19): regsvr32 /v c:\admquestrt\admquestrt.dll To run as administrator, right click on the Command Prompt (found in All Programs/Accessories) and select "Run as administrator."

6. There should be no errors from the previous step. If an error occurs, take a screenshot of the error message and email to *ruby.daamen@advdmi.com*

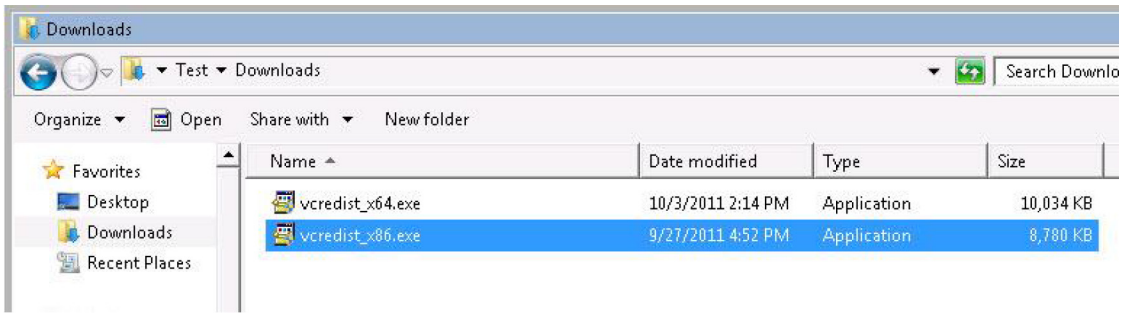


Figure A6-16. Download vcredist_x86.exe.

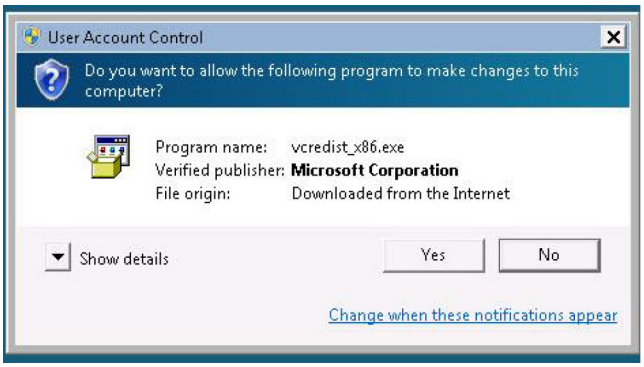


Figure A6-17. If the above message is received, select Yes and continue with the installation.

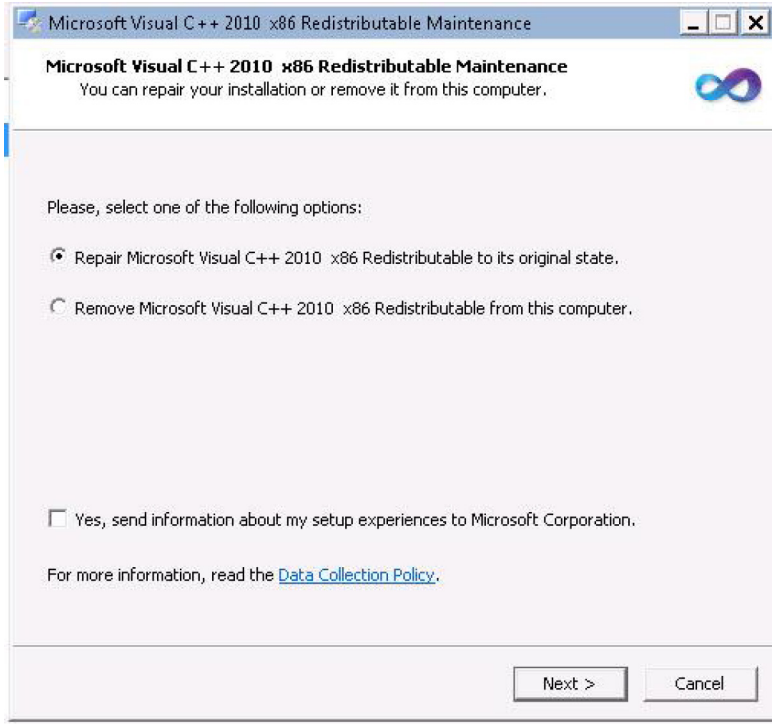


Figure A6-18. If the above message is received, select Repair and continue with the installation.

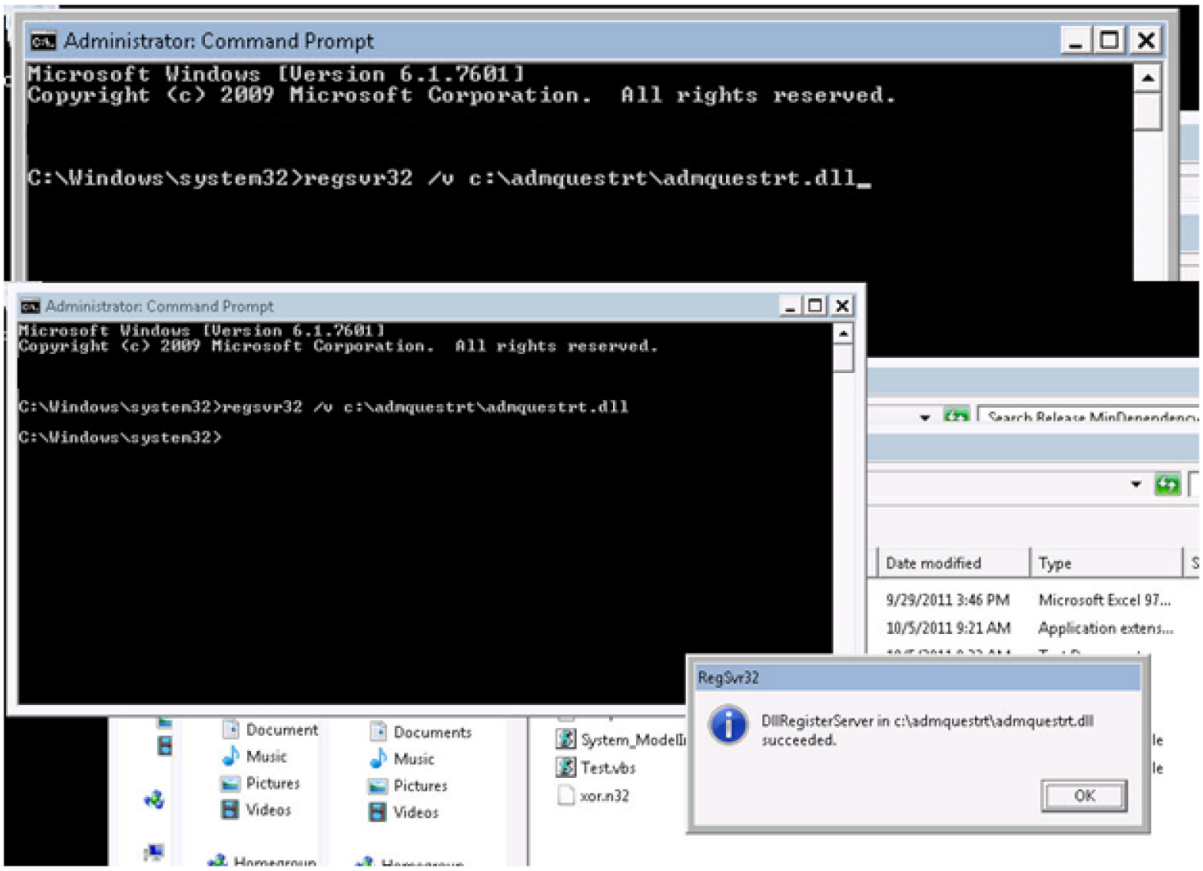


Figure A6-19. Screenshot of command used to register the ADMQuestRT dynamic link library (DLL). The message in the lower right corner of the above screenshot should appear (with the message title "RegSvr32') when the DLL has been successfully registered.

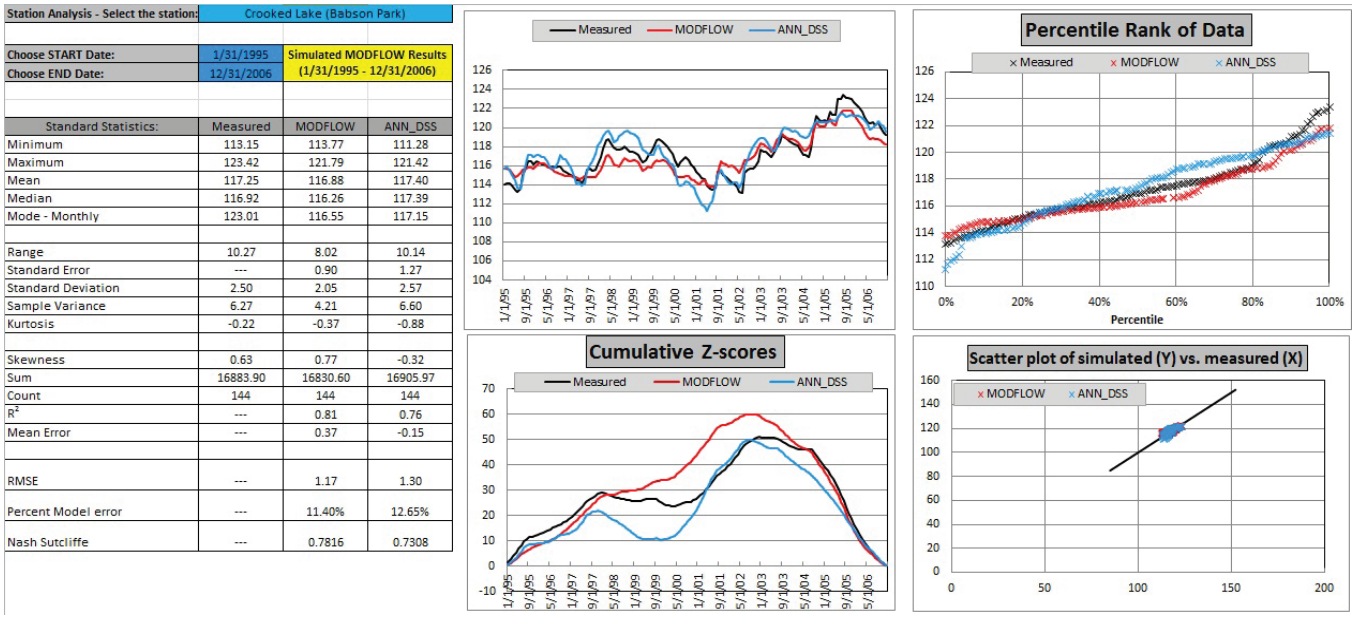
Appendix 7. Description of Model Data Viewer

The Model Data Viewer (MDV) is a Microsoft Excel 2010® workbook created to allow quick review and analysis of data and results collated during this study, consisting of measured data, data simulated by the east-central Florida transient (ECFT) MODFLOW model (Sepúlveda and others, 2012), and data simulated by the Central Florida Artificial Neural Network Decision Support System (CFANN DSS) described in appendix 6. The MDV is available for download from the U.S. Geological Survey (*http://dx.doi.org/10.3133/sir20145032*). The workbook consists of six worksheets that contain the measured and simulated data, provide statistical and graphical output for one to six sites, and allow user adjustment of certain specifications that control processing of the data. The worksheets titled "Setup," "Measured," "MODFLOW," "ANN_DSS," "Station_analysis," and "Multi-Station_analysis" are used interactively to review, compare, and analyze measured and simulated data. Each worksheet is described in more detail below. Additionally, the "Info" worksheet contains a history of MDV version releases and updates.

The "Setup" worksheet consists of six columns listing the naming convention for the 51 sites included in the MDV, which are the same 51 sites included in the CFANN DSS. The MDV uses the column titled "COMMON_NAME" as the name that is populated in the output worksheets. In addition, the "Setup" worksheet contains four columns of data that are used during the interactive process to populate the output worksheets. The "Data_type" column (spreadsheet column K) describes how the MDV processes the measured data in the viewer. The "Data_type" can be set to "Average" to calculate monthly averaged data or "Last Day" to use the last day of every month to populate the output worksheets. In general, if the user agrees with the current setting of the "Data_type" column, the user will not need to make any modifications to the "Setup" worksheet to use the MDV. The user, however, can view the "Setup" worksheet to review the various sites included in the MDV. To make any changes to the "Setup" worksheet, the user must first "unprotect" the worksheet by selecting that option under the Microsoft Excel® "Review" tab.

The "Measured," "MODFLOW," and "ANN_DSS" worksheets contain the data and results for the 51 sites. The "Measured" worksheet contains daily observed data from 9/30/1930 to 12/31/2009. The "MODFLOW" and "ANN_DSS" worksheets contain monthly data listed for the last day of each month for the ECFT model and CFANN DSS, respectively. The "MODFLOW" data were simulated from 1/31/1995 to 12/31/2006. The "ANN_DSS" data were simulated from 1/31/1965 to 12/31/2008. Three sites were not simulated by the ECFT model—Island Lake (Longwood), Lake Anderson (Orlando), and Crooked Lake (Orlando)—therefore, if these sites are selected, no statistical or graphical output will be displayed for the "MODFLOW" data.

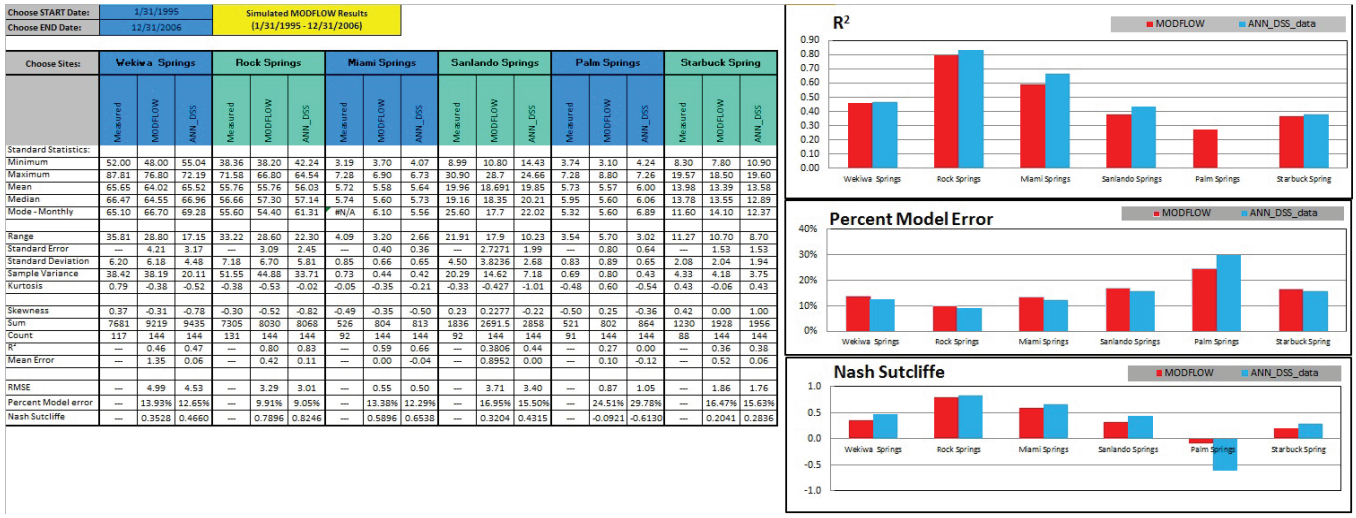
The "Station_analysis" and "Multi-Station_analysis" worksheets contain the statistical and graphical output and require user interaction to select the site(s) and dates of interest. The period of MODFLOW simulation was from 1/31/1995 to 12/31/2006, so users should always be attentive to date ranges when comparing MODFLOW results to Measured or ANN_DSS data that extend outside the MODFLOW simulation period. The "Station_analysis" worksheet consists of four graphs showing comparisons of the sets of data including time-series data, cumulative z-scores, percentile rank curves, and a scatter plot of the MODFLOW and ANN_DSS data (y-axis) versus the measured data on the x-axis (fig. A7-1). The cumulative z-score of a time series is calculated by subtracting the mean of the time series from each data point, dividing the resulting dividend by the standard deviation of the time series, and computing a running sum of this quotient starting with the first data point. Breaks in slope in a time-series graph of cumulative z-score indicate points in time when changes in hydrologic behavior occurred, such as a change from falling to rising water levels indicating the beginning of a drought recovery. The site displayed on the graphs in the "Station_analysis" worksheet may be selected using the drop-down menu to the right of the cell labeled "Station Analysis – Select the station:" located at the top left corner of the worksheet. The start and end dates for the graphs and statistical results may be selected using the drop-down menu to the right of the cells labeled "Choose START Date:" and "Choose END Date:" located below the site selection menu. The "Multi-Station_analysis" worksheet consists of bar graphs for up to six sites showing the fit of the simulated data to the measured data by graphing the coefficients of determination (R^2), percent model errors, and the Nash-Sutcliffe coefficients of efficiency (fig. A7-2). The start and end dates for the graphs and statistics may be selected using the drop-down menu to the right of the cells labeled "Choose START Date:" and "Choose END Date:" located in the top left corner of the worksheet. The six sites displayed on the graphs may be selected using the six drop-down menus located to the right of the cell labeled "Choose Sites:". Both the "Station_analysis" and "Multi-Station_analysis" worksheets present the following statistical results in tabular form: minimum, maximum, mean, median, mode, range, standard error, standard deviation, sample variance, kurtosis, skewness, sum, count, R^2, mean error, root mean square error (RMSE), percent model error, and the Nash-Sutcliffe coefficient of efficiency (Nash and Sutcliffe, 1970).



Station Analysis - Select the station:	Crooked Lake (Babson Park)	
Choose START Date:	1/31/1995	Simulated MODFLOW Results (1/31/1995 - 12/31/2006)
Choose END Date:	12/31/2006	

Standard Statistics:	Measured	MODFLOW	ANN_DSS
Minimum	113.15	113.77	111.28
Maximum	123.42	121.79	121.42
Mean	117.25	116.88	117.40
Median	116.92	116.26	117.39
Mode - Monthly	123.01	116.55	117.15
Range	10.27	8.02	10.14
Standard Error	---	0.90	1.27
Standard Deviation	2.50	2.05	2.57
Sample Variance	6.27	4.21	6.60
Kurtosis	-0.22	-0.37	-0.88
Skewness	0.63	0.77	-0.32
Sum	16883.90	16830.60	16905.97
Count	144	144	144
R^2	---	0.81	0.76
Mean Error	---	0.37	-0.15
RMSE	---	1.17	1.30
Percent Model error	---	11.40%	12.65%
Nash Sutcliffe	---	0.7816	0.7308

Figure A7-1. Screen capture showing the "Station_analysis" worksheet of the Model Data Viewer.



Choose START Date:	1/31/1995	Simulated MODFLOW Results (1/31/1995 - 12/31/2006)
Choose END Date:	12/31/2006	

Choose Sites:	Wekiwa Springs			Rock Springs			Miami Springs			Sanlando Springs			Palm Springs			Starbuck Spring		
	Measured	MODFLOW	ANN_DSS	Measured	MODFLOW	ANN_DSS	Measured	MODFLOW	ANN_DSS	Measured	MODFLOW	ANN_DSS	Measured	MODFLOW	ANN_DSS	Measured	MODFLOW	ANN_DSS
Standard Statistics:																		
Minimum	52.00	48.00	55.04	38.36	38.20	42.24	3.19	3.70	4.07	8.99	10.80	14.43	3.74	3.10	4.24	8.30	7.80	10.90
Maximum	87.81	76.80	72.19	71.58	66.80	64.54	7.28	6.90	6.73	30.90	28.7	24.66	7.28	8.80	7.26	19.57	18.50	19.60
Mean	65.65	64.02	65.52	55.76	55.76	56.03	5.72	5.58	5.64	19.96	18.691	19.85	5.73	5.57	6.00	13.98	13.39	13.58
Median	66.47	64.55	66.96	56.66	57.30	57.14	5.74	5.60	5.73	19.16	18.35	20.21	5.95	5.60	6.06	13.78	13.55	12.89
Mode - Monthly	65.10	66.70	69.28	55.60	54.40	61.31	#N/A	6.10	5.56	25.60	17.7	22.02	5.32	5.60	6.89	11.60	14.10	12.37
Range	35.81	28.80	17.15	33.22	28.60	22.30	4.09	3.20	2.66	21.91	17.9	10.23	3.54	5.70	3.02	11.27	10.70	8.70
Standard Error	---	4.21	3.17	---	3.09	2.45	---	0.40	0.36	---	2.7271	1.99	---	0.80	0.64	---	1.53	1.53
Standard Deviation	6.20	6.18	4.48	7.18	6.70	5.81	0.85	0.66	0.65	4.50	3.8236	2.68	0.83	0.89	0.65	2.08	2.04	1.94
Sample Variance	38.42	38.19	20.11	51.55	44.88	33.71	0.73	0.44	0.42	20.29	14.62	7.18	0.69	0.80	0.43	4.33	4.18	3.75
Kurtosis	0.79	-0.38	-0.52	-0.38	-0.53	-0.02	-0.05	-0.35	-0.21	-0.33	-0.427	-1.01	-0.48	0.60	-0.54	0.43	-0.06	0.43
Skewness	0.37	-0.31	-0.78	-0.30	-0.52	-0.82	-0.49	-0.35	-0.50	0.23	0.2277	-0.22	-0.50	0.25	-0.36	0.42	0.00	1.00
Sum	7681	9219	9435	7305	8030	8068	526	804	813	1836	2691.5	2858	521	802	864	1230	1928	1956
Count	117	144	144	131	144	144	92	144	144	92	144	144	91	144	144	88	144	144
R^2	---	0.46	0.47	---	0.80	0.83	---	0.59	0.66	---	0.3806	0.44	---	0.27	0.00	---	0.36	0.38
Mean Error	---	1.35	0.06	---	0.42	0.11	---	0.00	-0.04	---	0.8952	0.00	---	0.10	-0.12	---	0.52	0.06
RMSE	---	4.99	4.53	---	3.29	3.01	---	0.55	0.50	---	3.71	3.40	---	0.87	1.05	---	1.86	1.76
Percent Model error	---	13.93%	12.65%	---	9.91%	9.05%	---	13.38%	12.29%	---	16.95%	15.50%	---	24.51%	29.78%	---	16.47%	15.63%
Nash Sutcliffe	---	0.3528	0.4660	---	0.7896	0.8246	---	0.5896	0.6538	---	0.3204	0.4315	---	-0.0921	-0.6130	---	0.2041	0.2836

Figure A7-2. Screen capture showing the "Multi-Station_analysis" worksheet of the Model Data Viewer.

Nash, J.E., and Sutcliffe, J.V., 1970, River flow forecasting through conceptual models, Part 1—A discussion of principles: Journal of Hydrology, v. 10, no. 3, p. 282–290.

Sepúlveda, Nicasio, Tiedeman, C.R., O'Reilly, A.M., Davis, J.B., and Burger, Patrick, 2012, Water budget and groundwater flow in the surficial and Floridan aquifer systems in east-central Florida: U.S. Geological Survey Scientific Investigations Report 2012–5161, 214 p.

Appendix 8. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data

The data presented in table A8-1 are summary statistics for measured data; simulated data for the final lake water-level, groundwater-level, and spring-flow artificial neural network (ANN) models executed by the Central Florida Artificial Neural Network Decision Support System described in this report; and simulated data for the east-central Florida transient (ECFT) MODFLOW model developed by Sepúlveda and others (2012). Statistics were computed using monthly data from January 1995 through December 2006, where a measured time series consists of arithmetic averages of available measured daily data for each month, and simulated ANN and ECFT data consist of the final predicted monthly values from the respective models. Statistics were computed using the Model Data Viewer described in appendix 7.

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
5	Prevatt Lake (Apopka)	Meas	46.87	61.59	54.96	55.66	14.72	2.72	116	19.4%	--	--	--	--	--	--
		DSS	49.68	58.52	54.48	54.74	8.84	2.03	144	0.0%	0.438	0.435	-0.07	1.27	2.05	13.89%
		MOD	52.39	56.79	55.09	55.22	4.40	1.17	144	0.0%	0.496	0.313	-0.51	0.62	2.26	15.32%
38	South Lake (Vineland)	Meas	89.20	94.35	92.29	92.56	5.15	1.37	142	1.4%	--	--	--	--	--	--
		DSS	88.74	94.32	92.50	92.87	5.58	1.41	144	0.0%	0.870	0.839	-0.20	0.51	0.55	10.66%
		MOD	88.52	94.80	92.54	93.03	6.28	1.59	144	0.0%	0.546	0.333	-0.23	1.08	1.12	21.70%
41	Trout Lake (Clermont)	Meas	87.05	98.26	94.65	95.82	11.22	3.29	142	1.4%	--	--	--	--	--	--
		DSS	85.34	98.56	94.59	95.83	13.22	3.37	144	0.0%	0.921	0.917	0.04	0.95	0.95	8.45%
		MOD	90.78	97.83	94.23	94.51	7.05	1.46	144	0.0%	0.638	0.499	0.41	0.89	2.33	20.76%
45	Lake Wales (Lake Wales)	Meas	98.11	111.66	104.58	104.58	13.55	3.26	144	0.0%	--	--	--	--	--	--
		DSS	98.62	110.05	104.50	104.64	11.43	3.07	144	0.0%	0.857	0.856	0.08	1.17	1.24	9.13%
		MOD	99.26	111.08	104.32	103.93	11.82	2.80	144	0.0%	0.772	0.765	0.26	1.34	1.58	11.65%
49	Crooked Lake (Babson Park)	Meas	113.15	123.42	117.25	116.92	10.27	2.50	144	0.0%	--	--	--	--	--	--
		DSS	111.28	121.42	117.40	117.39	10.14	2.57	144	0.0%	0.759	0.731	-0.15	1.27	1.30	12.65%
		MOD	113.77	121.79	116.88	116.26	8.02	2.05	144	0.0%	0.810	0.782	0.37	0.90	1.17	11.40%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
50	Lake Clinch (Frostproof)	Meas	101.00	109.41	104.92	104.96	8.41	1.57	136	5.6%	--	--	--	--	--	--
		DSS	101.53	107.06	104.67	104.76	5.53	1.26	144	0.0%	0.704	0.686	0.19	0.67	0.88	10.47%
		MOD	102.12	107.00	104.51	104.62	4.88	1.23	144	0.0%	0.676	0.621	0.36	0.69	0.97	11.51%
55	Lake Otis (Winter Haven)	Meas	120.44	128.43	125.44	125.52	7.99	1.69	144	0.0%	--	--	--	--	--	--
		DSS	120.88	127.68	125.47	126.01	6.80	1.63	144	0.0%	0.772	0.764	-0.03	0.78	0.82	10.28%
		MOD	122.90	127.49	125.75	125.80	4.59	1.06	144	0.0%	0.835	0.720	-0.30	0.43	0.90	11.21%
56	Lake Annie (Waverly)	Meas	109.46	117.56	114.48	114.75	8.10	1.91	144	0.0%	--	--	--	--	--	--
		DSS	109.50	117.44	114.26	114.43	7.94	2.13	144	0.0%	0.901	0.860	0.21	0.68	0.71	8.82%
		MOD	108.51	118.59	113.41	112.98	10.08	2.48	144	0.0%	0.416	-0.322	1.06	1.91	2.20	27.11%
57	Lake Starr (Waverly)	Meas	96.38	109.74	103.38	103.63	13.37	3.17	144	0.0%	--	--	--	--	--	--
		DSS	97.12	109.32	103.34	103.73	12.20	3.30	144	0.0%	0.929	0.923	0.04	0.88	0.88	6.57%
		MOD	97.54	110.74	103.12	102.05	13.20	3.34	144	0.0%	0.588	0.498	0.27	2.16	2.25	16.83%
64	Lake McLeod (Eagle Lake)	Meas	124.58	131.98	128.78	129.02	7.40	1.92	143	0.7%	--	--	--	--	--	--
		DSS	123.09	131.99	128.54	128.65	8.90	1.93	144	0.0%	0.657	0.602	0.24	1.14	1.21	16.41%
		MOD	125.01	132.13	128.35	127.98	7.12	1.76	144	0.0%	0.365	0.220	0.42	1.42	1.70	22.96%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
71	Eagle Lake (Eagle Lake)	Meas	124.06	131.50	128.56	128.56	7.44	1.82	142	1.4%	--	--	--	--	--	--
		DSS	123.78	130.92	128.28	128.40	7.14	1.67	144	0.0%	0.649	0.611	0.29	1.00	1.14	15.27%
		MOD	125.31	132.24	128.22	127.53	6.93	1.86	144	0.0%	0.152	-0.285	0.35	1.73	2.07	27.76%
77	Lake Deeson (Lakeland)	Meas	117.67	132.94	126.45	127.75	15.27	3.58	144	0.0%	--	--	--	--	--	--
		DSS	120.54	132.00	126.75	127.39	11.46	3.07	144	0.0%	0.892	0.878	-0.29	1.02	1.25	8.20%
		MOD	123.97	129.91	127.30	127.59	5.94	1.60	144	0.0%	0.813	0.551	-0.85	0.70	2.40	15.70%
168	Apshawa Lake (Clermont)	Meas	80.24	88.01	85.22	85.63	7.77	2.07	119	17.4%	--	--	--	--	--	--
		DSS	81.26	88.88	85.17	85.31	7.62	2.00	144	0.0%	0.759	0.735	-0.32	0.92	1.07	13.74%
		MOD	80.61	86.68	84.03	84.35	6.07	1.51	144	0.0%	0.820	0.581	0.86	0.58	1.34	17.26%
170	Lake Apopka (Apopka)	Meas	62.83	67.82	66.16	66.64	4.99	1.24	142	1.4%	--	--	--	--	--	--
		DSS	63.75	68.07	66.65	66.98	4.32	1.08	144	0.0%	0.670	0.490	-0.52	0.61	0.88	17.69%
		MOD	64.01	68.04	66.16	66.27	4.03	1.02	144	0.0%	0.698	0.697	-0.02	0.56	0.68	13.64%
187	Lake Louisa (Clermont)	Meas	88.08	98.30	94.85	95.91	10.23	2.77	141	2.1%	--	--	--	--	--	--
		DSS	88.17	97.65	94.80	95.93	9.48	2.75	144	0.0%	0.902	0.898	0.11	0.87	0.88	8.63%
		MOD	91.91	99.08	94.32	94.07	7.17	1.34	144	0.0%	0.215	0.176	0.54	1.20	2.51	24.54%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
207	Lake Howell (Casselberry)	Meas	51.80	56.58	53.02	52.87	4.78	0.79	120	16.7%	--	--	--	--	--	--
		DSS	51.38	54.33	52.94	52.93	2.95	0.64	144	0.0%	0.549	0.532	0.05	0.45	0.54	11.25%
		MOD	50.93	56.58	52.95	52.76	5.65	1.15	144	0.0%	0.396	-0.360	0.01	0.92	0.92	19.19%
209	Island Lake (Longwood)	Meas	78.62	83.12	81.41	81.59	4.49	0.84	101	29.9%	--	--	--	--	--	--
		DSS	79.14	82.16	81.51	81.80	3.02	0.64	144	0.0%	0.678	0.638	-0.16	0.38	0.51	11.28%
		MOD	--	--	--	--	--	--	--	--	--	--	--	--	--	--
221	Lake Anderson (Orlando)	Meas	74.03	82.89	80.07	80.67	8.86	2.19	133	7.6%	--	--	--	--	--	--
		DSS	75.22	82.97	80.03	80.69	7.75	2.06	144	0.0%	0.837	0.833	0.10	0.86	0.90	10.12%
		MOD	--	--	--	--	--	--	--	--	--	--	--	--	--	--
239	Crooked Lake (Orlando)	Meas	55.97	73.60	65.50	66.14	17.63	4.04	109	24.3%	--	--	--	--	--	--
		DSS	56.56	71.84	64.55	64.73	15.28	4.00	144	0.0%	0.725	0.724	-0.05	1.73	2.12	12.04%
		MOD	--	--	--	--	--	--	--	--	--	--	--	--	--	--
277	Lake McCoy (Apopka)	Meas	53.44	62.01	59.81	60.71	8.57	2.01	117	18.8%	--	--	--	--	--	--
		DSS	51.49	61.82	59.48	60.35	10.33	2.40	144	0.0%	0.653	0.631	-0.08	1.14	1.22	14.29%
		MOD	56.43	64.18	60.60	61.01	7.75	1.85	144	0.0%	0.200	-0.284	-1.20	1.46	2.28	26.64%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
289	Lake Rose (Gotha)	Meas	71.16	86.62	80.12	80.88	15.46	2.65	121	16.0%	--	--	--	--	--	--
		DSS	68.44	82.99	80.00	80.92	14.55	3.09	144	0.0%	0.677	0.652	-0.37	1.37	1.57	10.12%
		MOD	76.25	84.54	79.99	79.07	8.29	2.28	144	0.0%	0.159	-0.058	-0.26	2.10	2.73	17.64%
297	Lake Sue (Winter Park)	Meas	69.74	72.71	70.91	70.87	2.97	0.43	120	16.7%	--	--	--	--	--	--
		DSS	70.37	71.69	71.03	71.07	1.32	0.31	144	0.0%	0.163	-0.057	-0.15	0.29	0.45	15.04%
		MOD	77.67	79.89	78.67	78.65	2.22	0.45	144	0.0%	0.354	-319.100	-7.77	0.36	7.78	261.72%
446	ROMP 45 Avon Park well	Meas	36.99	83.66	66.71	68.69	46.67	9.94	142	1.4%	--	--	--	--	--	--
		DSS	40.91	80.95	66.79	68.43	40.04	9.05	144	0.0%	0.856	0.856	-0.10	3.47	3.78	8.09%
		MOD	58.43	79.17	68.58	67.51	20.74	5.92	144	0.0%	0.585	0.521	-1.90	3.86	6.88	14.74%
469	ROMP 59 Avon Park well	Meas	38.93	85.19	67.93	69.78	46.26	9.75	144	0.0%	--	--	--	--	--	--
		DSS	42.26	81.57	67.94	69.41	39.31	8.90	144	0.0%	0.830	0.830	-0.02	3.70	4.02	8.70%
		MOD	60.68	81.59	70.81	69.29	20.91	6.52	144	0.0%	0.495	0.407	-2.88	4.66	7.51	16.24%
472	ROMP 60 Floridan well	Meas	37.22	83.20	66.55	68.23	45.98	9.24	141	2.1%	--	--	--	--	--	--
		DSS	40.41	82.27	65.51	67.27	41.86	9.71	144	0.0%	0.828	0.810	0.73	3.94	4.03	8.76%
		MOD	60.51	81.18	69.31	67.46	20.67	7.39	144	0.0%	0.411	0.290	-2.84	5.73	7.79	16.94%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
478	ROMP 57 Floridan well	Meas	90.72	111.90	104.18	104.84	21.18	4.48	139	3.5%	--	--	--	--	--	--
		DSS	92.60	111.10	104.32	105.20	18.50	4.16	144	0.0%	0.773	0.768	-0.15	2.03	2.16	10.20%
		MOD	90.55	106.55	99.16	99.11	16.00	3.74	144	0.0%	0.807	-0.452	5.02	1.66	5.40	25.49%
516	Sanlon Ranch Floridan well	Meas	76.04	104.41	93.18	94.41	28.37	6.14	141	2.1%	--	--	--	--	--	--
		DSS	77.15	102.43	93.61	94.88	25.28	5.54	144	0.0%	0.896	0.891	-0.27	1.78	2.03	7.14%
		MOD	76.85	93.58	85.35	84.48	16.73	4.92	144	0.0%	0.562	-1.130	7.99	3.23	8.97	31.60%
542	Lake Alfred Floridan well	Meas	121.99	130.98	127.55	128.06	8.99	1.85	139	3.5%	--	--	--	--	--	--
		DSS	122.44	129.98	127.44	127.78	7.54	1.65	144	0.0%	0.805	0.797	0.16	0.74	0.83	9.26%
		MOD	125.43	132.22	129.64	130.02	6.79	1.55	144	0.0%	0.619	-0.625	-2.06	0.97	2.36	26.22%
571	Lake Joel Floridan well	Meas	36.68	47.04	43.01	43.52	10.36	2.07	138	4.2%	--	--	--	--	--	--
		DSS	37.63	46.32	43.18	43.71	8.69	2.11	144	0.0%	0.830	0.816	-0.10	0.88	0.89	8.57%
		MOD	37.28	46.44	42.73	43.18	9.16	1.93	144	0.0%	0.794	0.766	0.34	0.89	1.00	9.66%
578	Lake Oliver Floridan well	Meas	103.74	111.66	108.50	108.79	7.92	1.66	141	2.1%	--	--	--	--	--	--
		DSS	104.41	110.89	108.44	108.82	6.48	1.61	144	0.0%	0.854	0.852	0.06	0.61	0.64	8.06%
		MOD	102.71	111.66	106.92	106.98	8.95	1.86	144	0.0%	0.834	-0.132	1.59	0.76	1.76	22.27%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
579	Lake Oliver Surficial well	Meas	106.23	114.64	111.22	111.57	8.41	1.86	140	2.8%	--	--	--	--	--	--
		DSS	106.81	113.72	111.14	111.61	6.91	1.69	144	0.0%	0.864	0.860	0.11	0.63	0.69	8.26%
		MOD	104.22	114.26	108.62	108.67	10.04	2.02	144	0.0%	0.848	-1.176	2.62	0.80	2.74	32.60%
588	Cocoa A Floridan well	Meas	29.55	39.28	35.62	36.02	9.73	1.96	141	2.1%	--	--	--	--	--	--
		DSS	30.96	39.54	35.61	35.94	8.58	2.03	144	0.0%	0.826	0.808	-0.01	0.85	0.86	8.79%
		MOD	30.46	39.53	36.04	36.23	9.07	1.88	144	0.0%	0.816	0.765	-0.42	0.81	0.95	9.73%
602	Bay Lake Floridan well	Meas	78.01	90.59	85.60	85.96	12.58	2.60	140	2.8%	--	--	--	--	--	--
		DSS	79.82	89.95	85.45	85.76	10.13	2.54	144	0.0%	0.754	0.741	0.10	1.27	1.32	10.53%
		MOD	76.61	86.97	82.08	82.11	10.36	2.43	144	0.0%	0.763	-1.037	3.49	1.20	3.71	29.51%
613	Cocoa D Floridan well	Meas	27.67	39.55	34.20	34.46	11.88	2.48	137	4.9%	--	--	--	--	--	--
		DSS	29.17	38.75	34.27	34.59	9.58	2.33	144	0.0%	0.657	0.637	-0.14	1.38	1.50	12.61%
		MOD	29.08	43.09	36.58	36.93	14.01	2.48	144	0.0%	0.695	-0.300	-2.44	1.39	2.83	23.84%
622	Cocoa P Floridan well	Meas	34.80	50.64	43.99	44.38	15.84	3.21	117	18.8%	--	--	--	--	--	--
		DSS	37.62	49.28	44.39	45.05	11.66	2.85	144	0.0%	0.825	0.812	-0.37	1.22	1.39	8.78%
		MOD	36.51	50.88	43.69	43.90	14.37	2.79	144	0.0%	0.816	0.810	0.21	1.20	1.40	8.83%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
627	ROMP 101 Floridan well	Meas	92.50	100.54	97.44	98.08	8.04	1.96	136	5.6%	--	--	--	--	--	--
		DSS	92.29	100.30	97.47	98.11	8.01	1.89	144	0.0%	0.899	0.898	-0.01	0.62	0.63	7.80%
		MOD	94.34	100.60	98.11	98.24	6.26	1.35	144	0.0%	0.634	0.514	-0.65	0.83	1.37	17.01%
630	Lake Sawyer Floridan well	Meas	73.42	87.52	81.39	81.82	14.10	3.06	126	12.5%	--	--	--	--	--	--
		DSS	73.27	86.76	81.07	81.78	13.49	3.17	144	0.0%	0.833	0.816	0.13	1.31	1.31	9.30%
		MOD	75.88	87.99	82.91	83.03	12.11	2.84	144	0.0%	0.869	0.580	-1.64	1.05	1.98	14.05%
632	Palm Lake Dr Floridan well	Meas	58.43	77.42	68.06	68.05	18.99	3.48	121	16.0%	--	--	--	--	--	--
		DSS	59.46	74.87	67.68	67.99	15.41	3.35	144	0.0%	0.733	0.719	0.14	1.76	1.84	9.71%
		MOD	62.47	75.95	69.00	68.92	13.48	2.74	144	0.0%	0.792	0.702	-0.98	1.26	1.90	9.99%
642	Mascotte Floridan well	Meas	94.16	102.19	99.16	99.58	8.03	1.98	128	11.1%	--	--	--	--	--	--
		DSS	94.24	101.62	99.15	99.68	7.38	1.75	144	0.0%	0.861	0.860	-0.01	0.66	0.74	9.23%
		MOD	90.68	98.12	95.01	95.09	7.44	1.66	144	0.0%	0.631	-4.053	4.29	0.99	4.46	55.49%
643	Mascotte Surficial well	Meas	95.07	103.04	99.90	100.34	7.97	1.99	128	11.1%	--	--	--	--	--	--
		DSS	95.30	102.20	99.97	100.40	6.90	1.63	144	0.0%	0.820	0.816	-0.06	0.72	0.85	10.70%
		MOD	99.96	104.72	103.26	103.54	4.76	1.28	144	0.0%	0.326	-2.416	-3.29	1.08	3.68	46.13%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
645	Bithlo 1 Floridan well	Meas	29.20	38.82	35.07	35.50	9.62	1.98	141	2.1%	--	--	--	--	--	--
		DSS	30.07	38.87	34.95	35.30	8.80	2.06	144	0.0%	0.838	0.815	0.14	0.84	0.85	8.85%
		MOD	28.09	37.76	33.79	34.02	9.67	1.91	144	0.0%	0.857	0.433	1.29	0.73	1.49	15.50%
648	OR-47 Floridan well	Meas	49.28	68.53	58.89	58.82	19.25	4.52	135	6.3%	--	--	--	--	--	--
		DSS	49.20	67.04	59.09	59.03	17.84	4.44	144	0.0%	0.899	0.895	-0.24	1.42	1.46	7.61%
		MOD	51.24	70.19	58.92	58.97	18.95	3.52	144	0.0%	0.786	0.778	-0.03	1.68	2.13	11.06%
650	Clermont Floridan Rep. well	Meas	74.76	86.01	81.96	82.80	11.25	3.05	124	13.9%	--	--	--	--	--	--
		DSS	74.63	85.47	81.81	82.68	10.84	2.83	144	0.0%	0.950	0.948	0.14	0.67	0.70	6.20%
		MOD	78.52	89.31	83.92	83.82	10.79	2.36	144	0.0%	0.780	0.302	-2.09	1.14	2.55	22.65%
661	Seminole 125 Floridan well	Meas	32.90	46.96	40.88	40.68	14.06	2.94	144	0.0%	--	--	--	--	--	--
		DSS	33.34	46.06	40.78	41.05	12.72	2.97	144	0.0%	0.795	0.780	0.10	1.35	1.38	9.81%
		MOD	30.29	44.40	37.17	37.17	14.11	2.39	144	0.0%	0.752	-0.837	3.71	1.20	3.99	28.38%
712	Taft Surficial well	Meas	92.45	96.17	94.17	94.24	3.72	1.06	111	22.9%	--	--	--	--	--	--
		DSS	92.63	96.25	94.51	94.59	3.62	0.96	144	0.0%	0.677	0.562	-0.36	0.54	0.70	18.95%
		MOD	90.77	94.66	93.61	93.91	3.89	1.01	144	0.0%	0.433	-0.015	0.62	0.80	1.07	28.85%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
308	Wekiwa Springs	Meas	52.00	87.81	65.65	66.47	35.81	6.20	117	18.8%	--	--	--	--	--	--
		DSS	55.04	72.19	65.52	66.96	17.15	4.48	144	0.0%	0.466	0.466	0.06	3.17	4.53	12.65%
		MOD	48.00	76.80	64.02	64.55	28.80	6.18	144	0.0%	0.456	0.353	1.35	4.21	4.99	13.93%
309	Rock Springs	Meas	38.36	71.58	55.76	56.66	33.22	7.18	131	9.0%	--	--	--	--	--	--
		DSS	42.24	64.54	56.03	57.14	22.30	5.81	144	0.0%	0.832	0.825	0.11	2.45	3.01	9.05%
		MOD	38.20	66.80	55.76	57.30	28.60	6.70	144	0.0%	0.796	0.790	0.42	3.09	3.29	9.91%
311	Miami Springs	Meas	3.19	7.28	5.72	5.74	4.09	0.85	92	36.1%	--	--	--	--	--	--
		DSS	4.07	6.73	5.64	5.73	2.66	0.65	144	0.0%	0.665	0.654	-0.04	0.36	0.50	12.29%
		MOD	3.70	6.90	5.58	5.60	3.20	0.66	144	0.0%	0.591	0.590	-0.00	0.40	0.55	13.38%
312	Sanlando Springs	Meas	8.99	30.90	19.96	19.16	21.91	4.50	92	36.1%	--	--	--	--	--	--
		DSS	14.43	24.66	19.85	20.21	10.23	2.68	144	0.0%	0.437	0.431	-0.00	1.99	3.40	15.50%
		MOD	10.80	28.70	18.69	18.35	17.90	3.82	144	0.0%	0.381	0.320	0.90	2.73	3.71	16.95%
313	Palm Springs	Meas	3.74	7.28	5.73	5.95	3.54	0.83	91	36.8%	--	--	--	--	--	--
		DSS	4.24	7.26	6.00	6.06	3.02	0.65	144	0.0%	0.000	-0.613	-0.12	0.64	1.05	29.78%
		MOD	3.10	8.80	5.57	5.60	5.70	0.89	144	0.0%	0.272	-0.092	0.10	0.80	0.87	24.51%

Table A8-1. Summary statistics for measured, ANN-simulated, and MODFLOW-simulated data for the 51 sites, 1995–2006.—Continued

[Meas, measured; DSS, simulated by Central Florida Artificial Neural Network Decision Support System; MOD, simulated by east-central Florida transient MODFLOW model; Min, minimum; Max, maximum; SD, standard deviation; N, number of data points; $PN_{missing}$, percentage of missing data points; R^2, coefficient of determination; NSCE, Nash-Sutcliffe coefficient of efficiency; ME, mean error; SE, standard error; RMSE, root-mean-square error; PME, percent model error computed by dividing the RMSE by the range of the measured data; --, not applicable; for lakes and wells, dimensional statistics are in feet; for springs, dimensional statistics are in cubic feet per second]

Site number	Site name	Data type	Min	Max	Mean	Median	Range	SD	N	PN missing	R^2	NSCE	ME	SE	RMSE	PME
314	Starbuck Spring	Meas	8.30	19.57	13.98	13.78	11.27	2.08	88	38.9%	--	--	--	--	--	--
		DSS	10.90	19.60	13.58	12.89	8.70	1.94	144	0.0%	0.378	0.284	0.06	1.53	1.76	15.63%
		MOD	7.80	18.50	13.39	13.55	10.70	2.04	144	0.0%	0.364	0.204	0.52	1.53	1.86	16.47%

Sepúlveda, Nicasio, Tiedeman, C.R., O'Reilly, A.M., Davis, J.B., and Burger, Patrick, 2012, Water budget and ground-water flow in the surficial and Floridan aquifer systems in east-central Florida: U.S. Geological Survey Scientific Investigations Report 2012–5161, 214 p.

Prepared by the Raleigh and West Trenton Publishing Service Centers

For additional information, write to:

Director
U.S. Geological Survey
Florida Water Science Center
4446 Pet Lane, Suite 108
Lutz, FL 33559

or visit our Web site at:
http://fl.water.usgs.gov

ISSN 2328-0328 (online)
http://dx.doi.org/10.3133/sir20145032