Hyper-Resolution Global Land Surface Model at Regional-to-Local Scales with observed Groundwater data assimilation

By

Raj Shekhar Singh

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Geography

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Norman L Miller, Chair
Professor Dennis D Baldocchi
Professor John Chiang
Professor Yoram Rubin

Spring 2014

UMI Number: 3686454

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI 3686454

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

©Copyright by Raj Shekhar Singh 2014
All Rights Reserved

Abstract

Hyper-Resolution Global Land Surface Model at Regional-to-Local Scales with observed Groundwater data assimilation

by

Raj Shekhar Singh

Doctor of Philosophy in Geography

University of California Berkeley

Professor Norman L Miller, Chair

Modeling groundwater is challenging: it is not readily visible and is difficult to measure, with limited sets of observations available. Even though groundwater models can reproduce water table and head variations, considerable drift in modeled land surface states can nonetheless result from partially known geologic structure, errors in the input forcing fields, and imperfect Land Surface Model (LSM) parameterizations. These models frequently have biased results that are very different from observations. While many hydrologic groups are grappling with developing better models to resolve these issues, it is also possible to make models more robust through data assimilation of observation groundwater data. The goal of this project is to develop a methodology for high-resolution land surface model runs over large spatial region and improve hydrologic modeling through observation data assimilation, and then to apply this methodology to improve groundwater monitoring and banking.

The high-resolution LSM modeling in this dissertation shows that model physics performs well at these resolutions and actually leads to better modeling of water/energy budget terms. The overarching goal of assimilation methodology is to resolve the critical issue of how to improve groundwater modeling in LSMs that lack sub-surface parameterizations and also run them on global scales. To achieve this, the research in this dissertation has been divided into three parts. The first goal was to run a commonly used land surface model at *hyper resolution* (1 km or finer) and show that this improves the modeling results without breaking the model. The second goal was to develop an observation data assimilation methodology to improve the high-resolution model. The third was to show real-world applications of this methodology.

The need for improved accuracy is currently driving the development of *hyper-resolution* land surface models that can be implemented at a continental scale with resolutions of 1 km or finer. In Chapter 2, I describe our research incorporating fine-scale grid resolutions and surface data into the National Center for Atmospheric Research (NCAR) Community Land Model (CLM v4.0) for simulations at 1 km, 25 km, and 100 km resolution using 1 km soil and topographic information. Multi-year model runs were performed over the southwestern United States, including the entire state of California and the Colorado River basin. Results show changes in the total amount of CLM-modeled water storage and in the spatial and temporal distributions of water in snow and soil reservoirs, as well as in surface fluxes and energy balance. We also demonstrate the critical scales at which important hydrological processes—such as snow water equivalent, soil moisture content, and runoff—begin to more accurately capture the magnitude of the land water balance for the entire domain. This proves that grid resolution itself is also a critical component of accurate model simulations, and of hydrologic budget closure.

To inform future model progress, we compare simulation outputs to station and gridded observations of model fields. Although the higher grid resolution model is not driven by high-resolution forcing, grid resolution changes alone yield significant reductions in the Root Mean Square Error (RMSE) between model outputs and actual observations: the RMSE decreases by more than 35% for soil moisture, 36% for terrestrial water storage anomaly, 34% for sensible heat, and 12% for snow water equivalent. The results of a 100 m resolution simulation over a spatial sub-domain indicate that parameters such as drainage, runoff, and infiltration are significantly impacted when hillslope scales of ~100 meters or finer are considered. We further show how limitations in the current model physics, including no lateral flow between grid cells, can affect model simulation accuracy.

The results presented in Chapter 2 are encouraging, but also highlight the limitations in improving an LSM by only increasing spatial resolution of the model and the surface datasets. As was shown with the water table depth analysis, increasing model resolution cannot compensate for parameterization errors and lack of sub–surface information in CLM. However, this problem can be solved by providing additional information to the model in the form of water table depth via data assimilation.

In Chapter 3, I discuss the development and verification of a methodology for assimilating observed groundwater depth measurements from multiple wells into the high spatial resolution LSM. A kriging-based interpolation technique is employed to overcome the problem of spatially and temporally sparse observations, and the interpolated data is assimilated into the CLM4.0 at 1 km resolution in a test region in northern California. Direct insertion and Ensemble Adjusted Kalman Filter (EAKF) based techniques are used for assimilation with direct insertion, producing better results and demonstrating major improvement in the simulation of water table depth. The Linear Pearson correlation coefficient between the observed well data and the assimilated model is 0.810, as opposed to only 0.107 for the non-

assimilated model. This improvement is most significant where the water table depth is greater than 5 m. Assimilation also improves soil moisture content, especially in the dry season when the water table is at its lowest. Other variables, including sensible heat flux, ground evaporation, infiltration, and runoff are also analyzed in order to quantify the effect of this assimilation methodology. Though the changes in these variables seem small, they can be very important in coupled models, and the improved simulation of groundwater in the assimilated model can explain the changes in these results.

This assimilation technique has been designed for use in regions with sparse and varied observation data, and it can be easily adapted to work in LSMs with a functional groundwater component. This gives us the capability to better model groundwater for the recent past and present, and also the potential to apply climate projections to probabilistically predict groundwater for future climate-change scenarios.

We have collaborated with Wellintel Inc. to implement our methodology on the ground using their observation data. We are in the process of setting up our model over a large region along the central California coast, where for the past few months Wellintel has implemented a pilot with measurements based on its water table depth measuring devices. The aim of this collaboration is to provide users with actionable water table depth data in and around their properties for the past, present, and near future. We are combining this methodology with Wellintel data to create a groundwater-management and groundwater-banking monitoring tool.

This is the first time that groundwater assimilation has been simulated in a high-resolution LSM, and as such this project provides proof-of-concept and application of a unique methodology that can be run at hyper resolution with data assimilation. The assimilation method is a very powerful tool that researchers can now apply to model land surface parameters much better than previously.

Dedication

To my parents for their support, patience, encouragement and believing in me.

To my advisor Prof. Norman L. Miller, for being much more than just an advisor and a guardian in this journey.

To my sisters Lipi and Leena, who have always supported and encouraged me to achieve my targets.

To Nishank, Dibyendu and other friends for thinking of me.

My deepest appreciation and love for my wife Momo Zheng Singh, for pushing me further than I ever thought I could reach, for being my teacher, partner and friend in this journey. I could never have made it this far without her.

Finally I dedicate this dissertation to my soon-to-be-born son, Shiv Huolong Singh.

Acknowledgement

My deepest thanks to my advisor, Prof. Norman Miller, for supporting me at every step of the way and making sure that I always have his support when needed. His guidance and faith in my abilities have always meant a great deal to me and are the reason why I am able to complete my degree.

I would like to thank my committee members. Prof. John Chiang for teaching me atmospheric physics and dynamics, Prof. Dennis Baldocchi for teaching me the importance of observation data in science and Prof. Yoram Rubin for kindling my interest in groundwater hydrology and geostatistics.

My sincere thanks goes out to Prof. James Familglietti for his support and for giving me a chance to work with his outstanding lab group at UC Irvine. I would like to thank his UC Irvine research group for hosting me for many weeks, especially JT Reager for his advice and support while collaborating with me.

I would like to thank the Department of Geography for providing me a really unique amalgam of social and physical sciences resulting in a great learning experience. I would also like to thank the Department for the McCone fellowship and Graduate Student Instructor positions. I would like to thank the departmental staff Marjorie, Nat, Dan, Deborah and Delores for answering all my questions, providing advice and taking care of all the formalities.

I would like to thank Shail Kumar for all the help and support throughout the program and before during the summer internship prior to enrolling. I would like to thank Mr. PK Sinha for his grant that provided me partial support for three years and the Energy Biosciences Institute for supporting me during my first year.

Finally I would like to thank my friends and cohorts who made my life at Berkeley such a pleasant experience.

Table of Contents

List of Figures

Figure 2.16: The 100-m resolution daily water table depth in meters for January 30, 2003 (left panel). The water table depth spatial gradients for the same model day, in meters (right panel). Gradients greater than 0.3 meters are shown in red, and some gradients (<1%) exceeded 1 meter.

List of tables

Nomenclature

Symbols

$\alpha^{\mu}_{g,\Lambda}$ - Direct beam albedo

$\alpha_{g,\Lambda}$ - Diffuse ground albedo

A - upstream or contributing area per unit contour length

f_{drai} - decay factor for drainage (m^{-1})

$f_{frz,1}$ - impermeable area fraction in frozen soil for the top layer

f_{imp} - fraction of impermeable area determined from the ice content of the soil;

f_{max} - maximum saturated fraction of the grid cell

f_{over} - decay factor for runoff (m^{-1})

f_{sat} - fractional saturated/impermeable area

f_{sno} - Fraction of the ground with snow cover

h – Lag; distance between data pairs while calculating Semi-variogram

n - Number of observations (Equations 3.2 & 3.4)

q_{drai} – Drainage ($kg\ m^{-2}\ s^{-1}$)

$q_{drai,max}$ - maximum drainage capacity ($kg\ m^{-2}\ s^{-1}$)

q_{infl} – Infiltration ($kg\ m^{-2}\ s^{-1}$)

$q_{infl,max}$ is the maximum soil infiltration capacity ($kg\ m^{-2}\ s^{-1}$)

q_{over} – Surface runoff ($kg\ m^{-2}\ s^{-1}$)

$q_{recharge}$ - Unconfined aquifer recharge rate ($kgm^{-2}s^{-1}$)

r – Correlation Coefficient

r_{sp} - Spearman rank correlation

S_y – Specific yield

T_{atm} – Air temperature (K)

W_a – Water stored in the unconfined aquifer (mm)

W_t – Total groundwater (mm)

Y_e - Model estimate

Y_o - Observation

ZΔ - water table depth (m)

β - grid cell topographic slope angle

$\gamma(h)$ - Semi-Variance at lag h

Γ_d - negative of the dry adiabatic lapse rate (Km^{-1})

θ_{atm}- Atmospheric potential temperature (K)

λ_α - Kriging weight

μ - Lagrange multiplier (Equation 1.11)

Acronyms

CAM - Community Atmospheric Model

CCV - California Central Valley

CDF - Cumulative Distribution Function

CESM - Community Earth System Model

CLM - Community Land Model

CSR - Center for Space Research

DA – Data Assimilation

DAReS - Data Assimilation Research Section

DART - Data Assimilation Research Testbed

DWR - Department of Water Resources (California)

EAKF - Ensemble Adjusted Kalman Filter

EnKF - Ensemble Kalman Filter

GCM - Global Circulation Model

GIS - Geographic Information System

GRACE – Gravity Recovery and Climate Experiment

GUI - Graphic User Interface

LIS- Land Information System

LSM - Land Surface Model

MSL - Mean Sea Level

NCAR - National Center for Atmospheric Research

NED- National Elevation Dataset

NLDAS - North American Land Data Assimilation System

NSIDC - National Snow and Ice Data Center

OK - Ordinary Kriging

RMSE - Root Mean Square Error

RTM - River Transport Model

SCAN - Soil Climate Analysis Network

SFREC - Sierra Foothill Research and Extension Center

SNICAR- SNow and ICe Aerosol Radiation

SNODAS - Snow Data Assimilation System

SSURGO - Soil Survey Geographic database

STATSGO- State Soil Geographic database

SWE - Snow Water Equivalent

TI - Topographic Index/Wetness Index

TWS - Terrestrial Water Storage

USDA-ARS - United States Department of Agriculture–Agricultural Research Service

USGS - United States Geological Survey (USGS)

Chapter 1: Introduction

Terrestrial energy and water budgets can have a profound influence on the overall behavior of the climate system. Monitoring and simulating these processes is key for a number of important applications, including agricultural productivity forecasting, flood and drought forecasting, and water resources management. However, to address critical modeling challenges associated with these processes there is a need for Land Surface Models (LSMs) that can be implemented either globally or over very large domains with a resolution of 1 km or finer, and for techniques through which observation data can be assimilated to improve model physics and output. Development of such high-resolution LSMs—sometimes referred to as *hyper-resolution models*—has been described as the grand challenge for the hydrologic and land surface modeling community [*Wood et al.*, 2011]. When combined with assimilation techniques [*Effort et al.*, 2012; *NRC*, 2000], these hyper-resolution LSMs would allow for a better representation of processes that are sub-grid to the current generation of models, and potentially enable more realistic process-level simulations. Developing high-resolution (<1 km) LSMs that can also correctly represent sub-surface hydrologic processes is also one of the big challenges currently facing the land surface modeling community.

The goal of this dissertation is to run a global land surface model at high resolution and to develop new assimilation techniques and constraints for bridging high-resolution LSM simulations at regional-to-local scales with reduced uncertainty for the study of terrestrial water storage variations, energy budgets, and water budgets. I have developed a methodology to adapt a more established and widely used LSM to high resolution and combine it with data assimilation in order to improve model results for monitoring water resources and groundwater banking. To this end, I have collaborated with the groundwater monitoring company Wellintel Inc. in developing our model-assimilation methodology to help average users get more information about groundwater in and around their property.

1.1 Background

Many important fine-scale land surface characteristics are approximated empirically in LSMs, including heterogeneous topography, soil texture, and vegetation [*Lawrence et al.*, 2011; *Oleson et al.*, 2010b; *Oleson et al.*, 2008]. As a result, there are important underrepresented hydrological processes in coarser LSMs, which include the effects of slope and aspect on runoff, infiltration, drainage, and groundwater storage, as well as soil moisture redistribution and evapotranspiration, and consequent effects on the surface energy budget [*Ivanov et al.*, 2004; *VanderKwaak and Loague*, 2001]. A number of studies have quantified the simulation improvements of LSMs with increased horizontal resolution [*A Kumar et al.*, 2006; *Meissner and Gerd*, 2009; *Wood et al.*, 2011], and higher-resolution model processes are generally expected to be better resolved with more realistic process

representations [*Kollet and Maxwell*, 2008]. High-resolution or hyper-resolution models could also be coupled with regional climate models to provide better-resolved forecasts and predictions for local management [*Wood et al.*, 2011]. Because the grid size in distributed models has a direct effect on information content and on the accuracy of simulation output [*Kuo et al.*, 1999], studying the effects of model resolution on model responses is an essential step toward further improving our community modeling efforts [*Famiglietti and Wood*, 1994]; already, higher-resolution grid size in LSMs has been shown to improve topographic characteristics, wetness index, and outflow [*Wolock and Price*, 1994; *Wolock and McCabe*, 2000; *Zhang and Montgomery*, 1994]. High spatial resolution models also help improve simulated runoff and other hydrologic parameters by incorporating different runoff mechanisms at varying scales [*Haddeland et al.*, 2002; *Kuo et al.*, 1999; *VanderKwaak and Loague*, 2001]; [*Vivoni et al.*, 2005]. Further, high-resolution LSMs improve urban area simulations, as well as snow water equivalent and snow-covered area variations in mountainous regions [*Christensen et al.*, 1998; *Meierdiercks et al.*, 2010]. To adequately address critical energy balance and water cycle applications, a high spatial resolution approach with an LSM on the order of 100 m to 1 km would be useful [*Wood et al.*, 2011].

Recent studies applying LSMs at resolutions approaching 1 km have been performed at regional scales, including COSMO-CLM [*Meissner and Gerd*, 2009] and the NASA Land Information System (LIS) [*A Kumar et al.*, 2006], and at catchment scales approaching even higher resolutions [*Christensen et al.*, 1998; *Giorgi*, 1990; *Jin et al.*, 2010; *Jones et al.*, 1995; *Meissner and Gerd*, 2009; *Rigon et al.*, 2006; *Skamarock et al.*, 2005]. Still, increasing the spatial resolution of LSMs to reach finer scales for large domains remains a priority. Recent developments have allowed us to increase the spatial resolution of global LSMs to 10–50 km, yet this might be insufficient for resolving the complex terrain and slope information needed for significant model improvements. The primary challenges of modeling at high resolution are the lack of correct model parameterization at this resolution, the required computational resources and the lack of input and of forcing datasets at this resolution [*Famiglietti et al.*, 2009; *Kollet et al.*, 2010; *A Kumar et al.*, 2006]. Global LSMs are all highly parameterized, and most are lumped single-column models that operate outside the spatial range for which the governing equations were derived. This is done with the underlying assumption that the equations still capture the basic behavior of the system, but in fact many of these parameterizations and assumptions might not be suitable for accurate high-resolution modeling, and there is a need for further study to quantify their effects.

Even for high-resolution models, modeling hydrologic parameters has been challenging; the models still do not incorporate many important physical features, such as sub-surface stratigraphy, lateral movement of ground water, and so on. Historically, hydrologic modeling has been carried out either through complex multi-dimensional fine-scale process descriptions or through lumped water balance models that are applied at only global and regional scales. In spite of notable progress in the development of advanced hydrologic models and data resources

over the past two decades, the current class of physically based hydrologic models falls short of providing much needed regional-to-local information on water availability for addressing emerging societal needs [*A Kumar et al.*, 2006; *Wood et al.*, 2011]. Instead, a high spatial resolution system approaching the order of 100 m to 1 km is required [*Singh et al.*, 2014b; *Wood et al.*, 2011]; such a model is expected to be better resolved with more realistic outcomes [*Kollet and Maxwell*, 2008]. In addition to increased resolution, interactive groundwater dynamics have been added to LSMs [*Fan*, 2007; *Liang et al.*, 2003; *Lo et al.*, 2008; *Lo et al.*, 2010; *Maxwell and Miller*, 2005; *Miguez-Macho et al.*, 2007; *Xie et al.*, 2007; *Yeh and Eltahir*, 2005a; b]. These advances indicate the importance of representing shallow groundwater variations and interaction with soil moisture.

1.2 Community Land Model

This study is based on simulations using the Community Land Model version 4.0 (CLM4.0), the land component of the National Center for Atmospheric Research (NCAR) Community Earth System Model (CESM 1.0.4) [*Oleson et al.*, 2010b]. CLM is one of the most widely used models and has been well studied; it is under constant development, and new features and improvements are added regularly. As a part of the CESM, CLM4.0 is easy to run in a coupled mode with the atmosphere, ice, and ocean models, and it also performs well in an offline mode with atmospheric forcings. The model is well explained in technical notes and user guides [*Kluzek*, 2012; *Oleson et al.*, 2010b], so here I discuss only the aspects of the model that are of primary importance for my analysis in this study.

The hydrologic cycle over land in CLM4.0 includes representations of interception of precipitation by plant foliage and woodstems, throughfall and stemflow, transpiration, soil evaporation, canopy evaporation, infiltration, runoff, soil water, aquifer recharge, and snow. These are directly linked to the biogeophysics of the model, and also affect temperature, precipitation, and runoff. Total runoff (both surface and sub-surface) is routed downstream to oceans using a River Transport Model (RTM) [*Branstetter*, 2001; *Gent et al.*, 2010] that is synchronously coupled to CLM4.0 for hydrological applications and for improved land-ocean-sea ice-atmosphere coupling in the CESM. The hydrology scheme has been updated from the earlier CLM3.5 and now includes a revised numerical solution of Richard's Equation [*Decker and Zeng*, 2009; *Zeng and Decker*, 2009], a revised soil evaporation parameterization that removes the soil resistance term introduced in CLM3.5, and an increase in the number of sub-surface layers to 15. The top ten layers (0–3.8 m) are hydrologically active, and the lower five (3.8–42 m) are described as thermal slabs that are hydrologically inactive and modeled as an unconfined aquifer (Figure 1.1; from CESM). The 3.8 m depth for the hydrologically active soil layer is assumed to be constant throughout the simulated region.

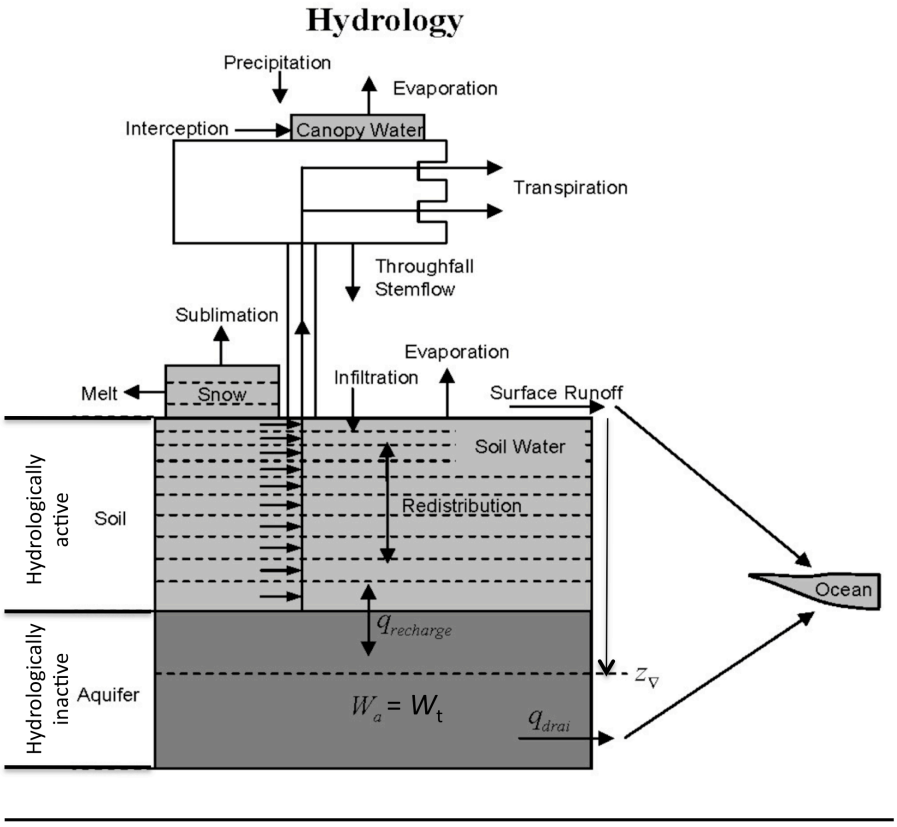

Figure 1.1: Schematic of Hydrology in CLM4.0 showing the different subsurface layers and treatment of major groundwater components.

CLM4.0 addresses several elements that enable the study of two-way interactions between the climate and human activities at the land surface, including land cover/land use change, agricultural practices, and urbanization. An irrigation model was added [*Levis et al.*, 2012], as well as an urban land unit type and associated urban canyon model [*Oleson et al.*, 2010b] for the study of urban climate and heat islands [*Oleson et al.*, 2010b; *Oleson et al.*, 2008], which improves CLM's potential as a high-resolution land surface modeling tool (Figure 1.2; from CESM). The urban environment as simulated in CLM4.0 is based on the *urban canyon* concept, and allows for the study of how climate change affects urban energy balance and the evaluation of possible urban planning and design strategies to mitigate warming (e.g., white roofs). The canyon system consists of roofs, walls (shaded and sunlit), and a canyon floor. The canyon floor is divided into pervious (e.g., residential lawns and parks) and impervious (e.g., roads, parking lots, sidewalks) fractions. The heat and moisture fluxes from each surface interact through a bulk air mass that represents air in the urban canopy layer, for which specific humidity, wind, and temperature are prognosed. The urban canopy air temperature can be compared with the temperature of surrounding vegetated/soil (rural) surfaces to determine heat island characteristics.

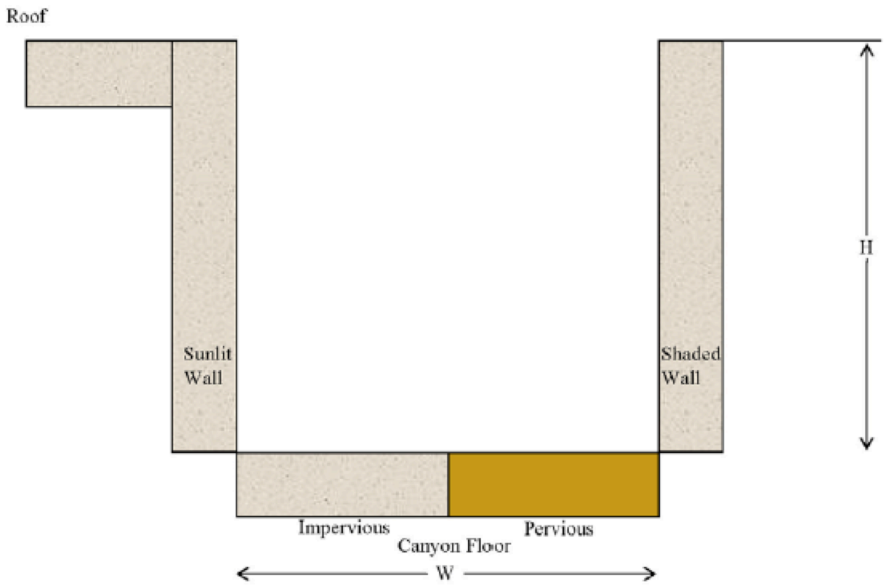

Figure 1.2: Schematic showing the different aspects of the urban model in CLM4.0.

The snow model in CLM4.0 has been significantly modified to incorporate SNICAR (SNow and ICe Aerosol Radiation) data, which represents the effect of aerosol deposition (e.g., black and organic carbon and dust) on albedo, introduces a grain-size dependent snow-aging parameterization, and permits vertically resolved snowpack [*Flanner and Zender*, 2005; *Flanner et al.*, 2007]. The snow model now also includes a new density-dependent snow cover fraction parameterization [*Niu et al.*, 2007], a revised snow burial fraction over short vegetation [*Wang and Zeng*, 2009], and corrections to snow compaction [*Oleson et al.*, 2008]. These changes are explained in detail in the CLM4.0 technical report [*Oleson et al.*, 2010b] and user guide [*Kluzek*, 2012].

Fundamental to the CLM4.0 hydrology is the fractional saturated/impermeable area (f_{sat}) estimation, which is determined by the topographic characteristics and soil moisture state of a grid cell [*Niu and Yang*, 2006]:

$$f_{sat} = (1 - f_{frz,1}) f_{max} \exp(-0.5 f_{over} Z\Delta) + f_{frz,1} \quad (1.1)$$

where f_{max} is the maximum saturated fraction of the grid cell with respect to soil moisture, f_{over} is the decay factor (m^{-1}), $Z\Delta$ (m) is the water table depth, and $f_{frz,1}$ is the impermeable area fraction in frozen soil for the top layer. The maximum saturated fraction f_{max} is defined as the discrete cumulative distribution function (CDF) of the topographic index/wetness index (TI) when the grid cell mean water table depth is zero [*Niu et al.*, 2005]. It is calculated as the percentage of a grid cell where TI is larger than or equal to the grid cell mean TI. Topographic Index in this

study has been calculated explicitly for each grid cell at the resolution of the model using the USGS 1/3 arc second (~10 m) National Elevation Dataset (NED, Data available from USGS) through the process described by [*Quinn et al.*, 1995; *Wolock and McCabe*, 2000]. TI is defined as the following:

$$TI = \ln (A/\tan\beta) \quad (1.2)$$

where A is the upstream or contributing area per unit contour length, and β is the grid cell topographic slope angle [*Beven and Kirkby, 1979*]. TI is less for steeper slopes and more for flat regions, which results in f_{max} being smaller for more hilly grid cells and larger for grid cells with flat topography. Calculation of TI for studies in this dissertation has been described in detain in Appendix A.

Drainage or sub-surface runoff (q_{drai}, kg m^{-2} s^{-1}) is calculated using the SIMTOP scheme [*Niu et al.*, 2005] with a modification to account for reduced drainage in frozen soils:

$$q_{drai}=(1-f_{imp})q_{drai,max}\, exp(-f_{drai}\, Z\Delta) \quad (1.3)$$

where f_{imp} is the fraction of impermeable area determined from the ice content of the soil, f_{drai} is the decay factor (m^{-1}), and $q_{drai,max}$= 5.5x10^{-3} kg m^{-2} s^{-1} is the maximum drainage when the water table depth is at the surface. This maximum drainage is a global constant determined through sensitivity analysis and comparison with observed runoff [*Oleson et al.*, 2010b]. $Z\Delta$ (m) is the water table depth.

Surface runoff (q_{over}, kg m^{-2} s^{-1}) consists of overland flow due to saturation excess (Dunne runoff) and infiltration excess (Hortonian runoff), and the maximum soil infiltration capacity is determined from soil texture and soil moisture [*Entekhabi and Eagleson*, 1989]:

$$q_{over} = f_{sat}q_{liq,0} + (1-f_{sat})max(0,q_{liq,0} - q_{infl,max}) \quad (1.4)$$

Infiltration (q_{infl}, kg m^{-2} s^{-1}) into the surface soil layer is calculated as the residual of the surface water balance:

$$q_{infl} = q_{liq,0} - q_{over} \quad (1.5)$$

$q_{liq,0}$ is the total liquid precipitation reaching ground plus snow melt; $q_{infl,max}$ is the maximum soil infiltration capacity (kg m^{-2} s^{-1}), which is determined from the soil texture and soil moisture values; and f_{max} is the maximum saturated fraction of the grid cell.

The saturated hydraulic conductivity, volumetric water content at saturation, Clapp and Hornberger exponent, and saturated soil matric potential are determined using soil texture values as described by [*Clapp and Hornberger*, 1978; *Cosby et al.*, 1984;

Niu et al., 2007; *Oleson et al.*, 2010b]. The high-resolution soil texture dataset needed for this purpose was produced using the CONUS-SOIL dataset at 30 arc second (~1 km) resolution [*Miller and White*, 1998].

Determination of water table depth ZΔ is via [*Niu et al.*, 2007], where a groundwater component is added in the form of an unconfined aquifer lying below the hydrologically active upper layers in the soil column (Figure 1.1). The solution for ZΔ (m) is dependent on whether the water table is within or below the active soil column layers, and the active and inactive water storage terms are used to account for these conditions. The first water storage term W_a (mm) is the water stored in the unconfined aquifer, and varies with the change in water table depth when the water table is below the lower boundary of the hydrologically active soil column. The second water storage term W_t (mm) is the total groundwater, which includes water both within the soil column and in the unconfined aquifer. When the water table is below the soil column then $W_t = W_a$ (Figure 1.1), and when the water table is within the soil column W_a is constant (5000mm). This is because there is no water exchange between the soil column and the underlying aquifer, while W_t varies with soil moisture conditions in different hydrologically active layers. These two water stores are updated as the water table changes within the active soil column or the inactive soil layers[*Oleson et al.*, 2010b].

There is an unconfined aquifer at the bottom (below 3.8 m) of the soil column (Figure 1.1). The depth of the water table is ZΔ (m), and changes in aquifer water content W_a (mm) and W_t (mm) are controlled by the balance between drainage from the aquifer q_{drai} and the unconfined aquifer recharge rate $q_{recharge}$ ($kgm^{-2}s^{-1}$) (defined as positive from soil to aquifer). The water table depth is calculated from the aquifer water storage scaled by the average specific yield S_y, where $S_y = 0.2$ is the fraction of water volume that can be drained by gravity in an unconfined aquifer [*Niu et al.*, 2007; *Oleson et al.*, 2010b], with the assumption that the initial amount of water in the aquifer is 4800 mm and the corresponding water table depth is one meter below the bottom of the active soil layer. For the case where the water table is within the soil column, there is no water exchange between the soil column and the underlying aquifer and the water table depth is calculated accordingly [*Oleson et al.*, 2010b].

Atmospheric potential temperature θ_{atm} (K) is an important parameter affected by the high resolution topographic information, and is defined as follows:

$$\theta_{atm} = T_{atm} + \Gamma_d z_{atm,h} \tag{1.6}$$

where T_{atm} (K) is the air where T_{atm} is the temperature at height $z_{atm,h}$, and Γ_d = 0.0098 km^{-1} is the negative of the dry adiabatic lapse rate [*Oleson et al.*, 2010b].

1.3 High Resolution Modeling

Performing high-resolution LSM simulations globally is one of the grand challenges facing the hydrologic community. The problem can be approached from multiple directions and needs to be broken down into smaller pieces so that it can be solved efficiently. While some groups are working to develop and improve new models that perform better at higher resolutions, my research focuses instead on adapting a more established and widely employed global LSM to high resolution then combining this model with data assimilation techniques to improve model results. In Chapter 2 of this dissertation I describe our approach to this issue in detail.

Running models at high resolution requires high-resolution input and forcing data, which is frequently very difficult to create—especially for high-resolution or hyper-resolution models running over very large spatial domains. We therefore prioritized which data we could possibly attain or derive at high resolution based on availability and the importance or sensitivity of such data in the model.

Topography is one of the most important spatial fields for determining surface water flow, and is a key control on soil moisture variability at high resolution [*Famiglietti et al.*, 1998; *Jana and Mohanty*, 2012a; b]. A previous analysis of TOPMODEL results has shown that model predictions of the depth of the water table, of the ratio of overland flow to total flow, peak flow, and variance, and of the skew of predicted stream flow were all affected by the digital elevation model (DEM) resolution [*Wolock and Price*, 1994]. When used to calculate TI (Equation 1.2), high-resolution topography affects the amount of information relayed to the model in the form of maximum saturated fraction (f_{max}). This results in significantly improved fractional saturated/impermeable area (f_{sat}) estimation, which itself improves drainage, runoff, infiltration, and soil moisture calculations. High-resolution topography improves the representation of air temperature and the atmospheric potential temperature in CLM4.0. As such, more refined topography will result in a more precise representation of temperature, snow, and evapotranspiration variability within the model domain. Alternatively, spatially averaged topography will result in smoothed wider and lower elevation plateaus that might contain different total snow amounts and an effectively different water balance. Spatial variation in elevation over steep terrain is more faithfully represented at high resolution, and will result in more accurate predictions of snow accumulation and depletion.

Soil texture plays an important role in the calculation of soil conductivity parameters, and studies have shown that improving soil depth information improves land surface modeling results [*Decker and Zeng*, 2009; *Gochis et al.*, 2010]. The saturated hydraulic conductivity, volumetric water content at saturation, Clapp and Hornberger exponent, and saturated soil matric potential are determined using soil texture values as described by [*Clapp and Hornberger*, 1978; *Cosby et al.*, 1984; *Niu et al.*, 2007; *Oleson et al.*, 2010b]. Soil texture also influences the partitioning of moisture and energy fluxes at the land surface, and offers important feedback to the

energy and water cycles over timescales ranging from hourly to inter-annual [*Reichle et al.*, 2002]. A more refined soil texture dataset will therefore improve the model outputs.

1.4 Kriging and Geostatistics Tools

Geostatistics offers a way to describe the spatial continuity of natural phenomena, and provides adaptations of classical regression techniques to take advantage of this continuity [*Isaaks and Srivastava*, 1990]. Geostatistics is used to analyze auto-correlated data, and in my research I employ these tools to compensate for gaps in sparse observation data before data assimilation. Due to the unique nature of groundwater (as explained in chapter 3), the geostatistics tools are shown to work very well, yielding spatially interpolated values from observation data that were used for assimilation. Geostatistical methods are optimal when the data are normally distributed and stationary. Here I use two components of geostatistics: semi-variogram analysis and kriging.

1.4.1 Semi-Variogram

The variogram characterizes the spatial continuity or roughness of a dataset**.** Ordinary one-dimensional statistics for two datasets might be nearly identical, but the spatial continuity can be quite different**.** Variogram analysis utilizes the experimental variogram that is calculated from the observation data and the variogram model, which is fitted to the observation data**.** The experimental variogram is calculated by averaging one half of the squared difference over all pairs of observations with the specified separation distance and direction, plotted as a two-dimensional graph. The variogram model is chosen from a set of mathematical functions that describe spatial relationships by matching the shape of the curve of the experimental variogram to the shape of the curve of the mathematical function.

The experimental semi-variogram is calculated for the data using the following equations:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{N(h)} [Z(u) - Z(u+h)]^2 \tag{1.7}$$

where $\gamma(h)$ is the value of the semi-variance at lag h, the separation distance between the data pairs. $N(h)$ is the number of data pairs found for the specified lag vector h, and $Z(u)$ and $Z(u+h)$ are attribute values at location u and location $u+h$, respectively. Lag h is defined as the separation distance between the data pairs.

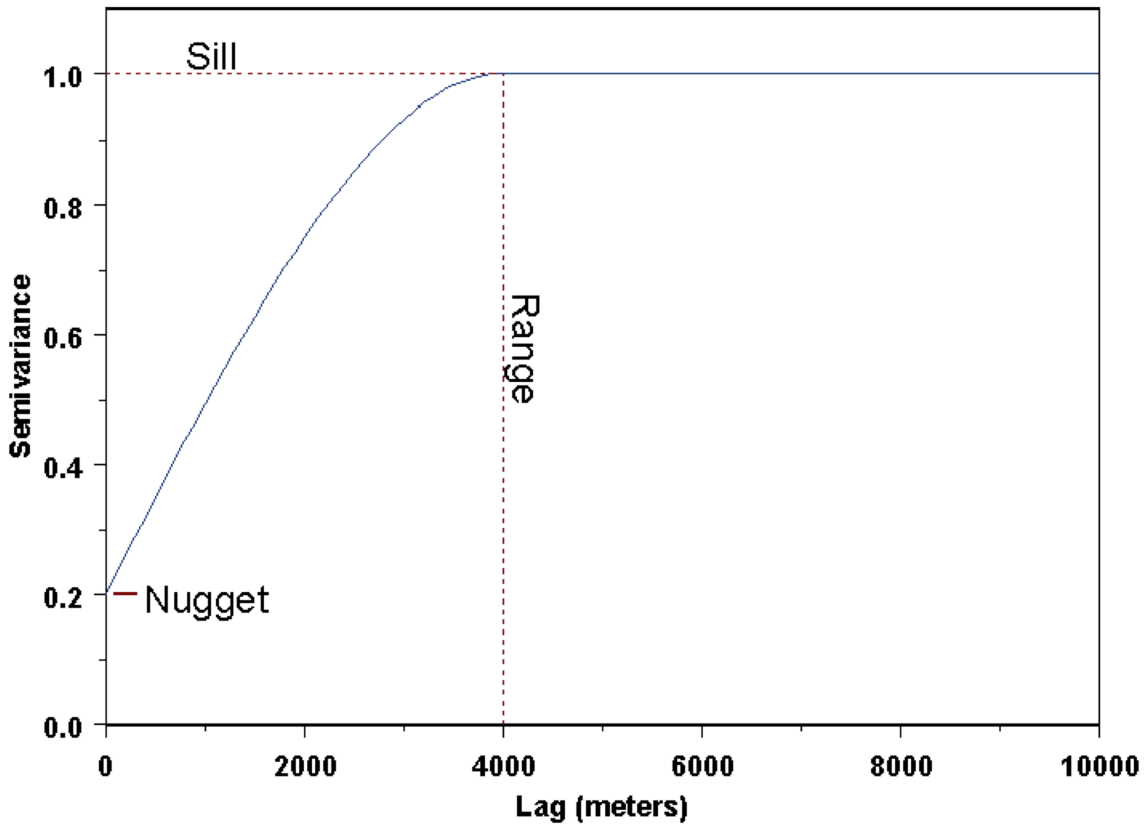

Figure 1.3: Characteristics of the Semi-variogram

The resulting semi-variogram has the following important characteristics.

Nugget: In theory, the semi-variogram value at the origin (0 lag) should be zero. If it is significantly different from zero for lags very close to zero, then this semi-variogram value is referred to as the *nugget*. The nugget represents variability at distances smaller than the typical sample spacing, including measurement error.

Sill: This is the semi-variance value at which the variogram levels off. It also refers to the "amplitude" of a certain component of the semi-variogram. For the plot above, the "sill" could refer to the overall sill (1.0) *or* to the difference (0.8) between the overall sill and the nugget (0.2).

Range: This is the lag distance at which the semi-variogram (or semi-variogram component) reaches the sill value. Presumably, autocorrelation is essentially zero beyond the range.

For the sake of kriging, we need to replace the empirical semi-variogram with an acceptable semi-variogram model. Part of the reason for this is that the kriging algorithm requires access to semi-variogram values for lag distances other than those used in the empirical semi-variogram. More importantly, however, the semi-

variogram models used in the kriging process must obey certain numerical properties in order for the kriging equations to be solvable. Therefore, we choose from a palette of acceptable semi-variogram models.

1.4.2 Ordinary Kriging

Kriging is a spatial interpolation based on regressing against observed values of surrounding data points, weighted according to spatial covariance values. The technique is used for generating optimal, unbiased estimates of regionalized variables at un-sampled locations using the structural properties of a semi-variogram and the initial set of data values. Kriging also provides the variance of the estimates at every point [*Isaaks and Srivastava*, 1990; *Leuangthong*, 2008; *Wackernagel*, 1999]. It takes into consideration the spatial structure of the parameter and therefore outperforms other methods, such as the arithmetic mean, nearest neighbor, distance weighted, and polynomial interpolation methods. Kriging helps to compensate for the effects of data clustering by assigning less weight to individual points within a cluster than to isolated data points. Kriging also provides the variance at every point, which is an indicator of the accuracy of the estimated value; this is the major advantage of kriging over other estimation techniques [*V Kumar and Remadevi*, 2006].

Ordinary Kriging (OK) is the most widely employed kriging method, and also the best unbiased linear estimator. It estimates values at unsampled locations between known data points using a linear estimation procedure in a region for which the variogram is known. This technique is more appropriate than other kriging procedures when the variable being estimated does not exhibit a strong spatial trend in any particular direction [*Chung and Rogers*, 2011; *Wackernagel*, 1999]; the estimation is unbiased in the linear sense and results in minimum error variance. The prerequisite for Ordinary Kriging is the assumption of stationarity and the existence of the variogram model.

In this study, Ordinary Kriging is used to estimate water table depth and to support data assimilation when very sparse observations are available. The slowly spatially and temporally varying nature of water table depth helps in this assumption. This kriging methodology has been well described and documented [*Isaaks and Srivastava*, 1990; *Leuangthong*, 2008; *Wackernagel*, 1999] and only a cursory review is provided here. All kriging estimators are variants of the basic linear regression estimator $Z^*(u)$, which is defined by[*Goovaerts*, 1997]:

$$Z^*(u) - m(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha [Z(u_\alpha) - m(u_\alpha)] \tag{1.8}$$

where u are location vectors for an estimation point and one of its neighboring data points. Vector u is indexed by α, where α ranges from 1 to n(u), n(u) is the number of data points in the local neighborhood used for estimation of $Z^*(u)$, m(u) and $m(u_\alpha)$ are the expected values of Z(u) and $Z(u_\alpha)$, and $\lambda_\alpha(u)$ is the kriging weight assigned to datum $Z(u_\alpha)$ for estimation location u. The same datum will receive a different weight for different estimation locations.

Z(u) is treated as a random field with trend component m(u) and residual component R(u) = Z(u)- m(u). Kriging estimates the residual at u as a weighted sum of residuals at surrounding data points. Weights λ_α are derived from the covariance function or semi-variogram, which should characterize the residual component. The goal is to determine weights λ that minimize the variance of the estimator, which under unbiased constraints is the following:

$$E[Z^*(u) - Z(u)] = 0 \tag{1.9}$$

For Ordinary Kriging, rather than assuming that the mean is constant over the entire domain we assume that it is constant in the local neighborhood of each estimation point—that is, $m(u_\alpha) = m(u)$ for each nearby data value $Z(u_\alpha)$ that we use to estimate Z(u). The interpolation value Z^* at any location x_0 is this:

$$Z^*(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) \qquad \text{i=1,2,3...N} \tag{1.10}$$

where λ_i is the weight for the observation Z at location x_i. In kriging, the weights λ_i are calculated by the following equations so that $Z^*(x_0)$ is unbiased and optimal:

$$\sum_{j=1}^{N} \lambda_j \gamma(x_i, x_j) + \mu = \gamma(x_i, x_0) \qquad \text{i=1,2,3...N} \tag{1.11}$$

$$\sum_{j=1}^{N} \lambda_j = 1 \tag{1.12}$$

where μ is the Lagrange multiplier and $\gamma(x_i, x_j)$ is the semi-variogram between the two points x_i and x_j, where i is not equal to j.

1.5 DART - Data Assimilation Research Testbed

The Data Assimilation Research Testbed (DART) is a community facility for ensemble data assimilation, developed and maintained at the NCAR Data Assimilation Research Section (DAReS). DART offers modelers, observational scientists, and geophysicists powerful, flexible data assimilation (DA) tools that are easy to implement and use and can be customized to support efficient operational DA applications. The DART software environment makes it easy to explore a variety of data assimilation methods and observations with different numerical models, and is it designed to facilitate the combination of assimilation algorithms, models, and real (as well as synthetic) observations.

DART employs a modular programming approach to apply an Ensemble Kalman Filter (EnKF) that nudges the underlying model toward a state more consistent with observations. This method requires running multiple instances of a model to generate an ensemble of states. A forward operator appropriate for the type of observation being assimilated is applied to each state to generate the model's estimate of the observation. DART algorithms are designed such that incorporating new models and new observation types requires only minimal coding within a small set of interface routines, and no modification to the existing model code.

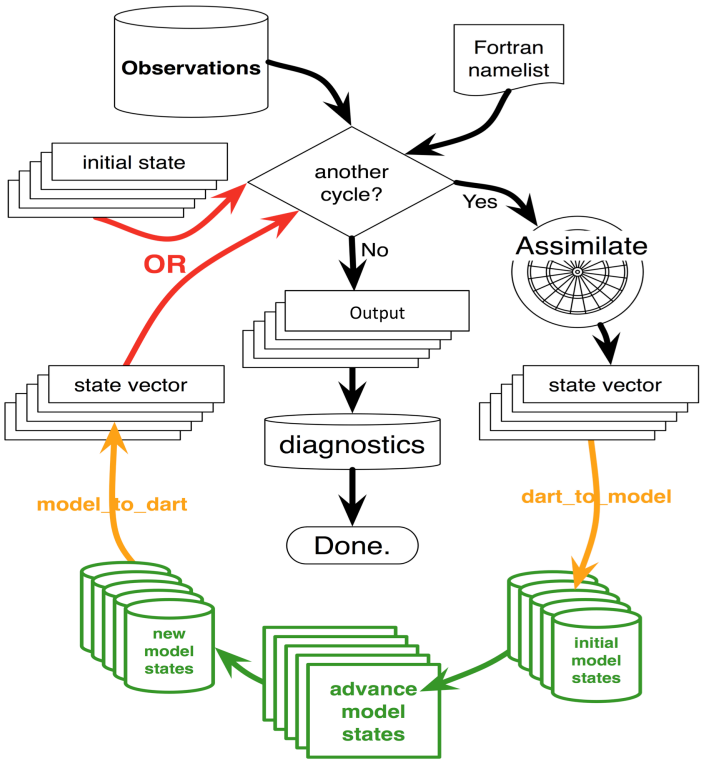

Figure 1. 4: Schematic diagram showing the DART assimilation methodology (from the NCAR DAReS website)

Figure 1.4 describes the DART view of ensemble data assimilation for models that run as separate executables. Everything is driven by a Fortran namelist and the presence or absence of observations. A Fortran executable named "filter" reads a namelist, an initial state for the ensemble, and a file containing observations, and then goes to work. Given the observations and an initial state, filter assimilates the observations and then determines how far to advance the model (using information from the namelist and the observation file). Filter forks a shell script to the system that is responsible for three things: 1) converting the DART state vectors and "advance_to_time" to the format required by the underlying model, 2) advancing the model, and 3) converting the model output into a form suitable for filter. The model advances each ensemble member (either in turn or all at once) and the model output is converted to the input format expected by filter. The shell script then finishes and signals filter to continue. This returns the assimilation procedure back to the beginning and the cycle continues until either there are no further observations to assimilate or the control information in the Fortran namelist is met. When that happens a set of restart files that are suitable for continuing an experiment with more observations and diagnostic files are written. These diagnostic files allow for exploration of the assimilation before and after each assimilation step and in "observation space"; each real observation is paired with the estimates of the observation from each of the ensemble members (if desired). Minimally, the ensemble-mean estimate of the observation and the ensemble spread of the estimates are recorded.

1.6 Motivation and Outline

My dissertation work aims to be a part of the solution to a grand challenge: running hyper-resolution LSMs at scales for a global domain with reduced uncertainty and applications for groundwater management. To this end, I employed a commonly used LSM and developed a methodology for simulations at very high resolutions using fine-scale topographic and soil texture data; hydrologic modeling is then improved by data assimilation of observed water table depth measurements. I show that the local energy and water budget terms can be calculated with improved accuracy through this application of assimilation methods within a high spatial resolution model. I also developed a kriging-based interpolation method to solve the problem of sparse observation datasets for data assimilation. I have partnered with Wellintel Inc. to apply this new methodology to modeling groundwater at high spatio-temporal resolution in order to facilitate groundwater management and banking.

Below is a brief outline of the work I present here as my dissertation.

Chapter 1 above introduces the problem and explores the literature to understand the previous research and ongoing challenges in this area. I provide brief

information about the NCAR Community Land Model (CLM4.0) used in the study, and how it is set up. I describe and explain the high-resolution model, creation of high-resolution surface datasets, atmospheric forcing datasets, and terrestrial observation data. I also discuss the kriging methodology employed for interpolation of data, and the data assimilation technique (including DART and direct insertion).

In Chapter 2 I demonstrate high-resolution land surface simulations over a large domain and examine the effects of increased spatial resolution and high-resolution topographic and soil texture information on water and energy-cycle variables in CLM4.0. I run CLM4.0 at a 0.01° (~1 km) resolution over the southwestern United States and at a 100 m resolution for a nested sub-domain located along the California Central Valley and Sierra Nevada foothills. A secondary aim of this work is to develop and test a 1 km resolution LSM over the state of California, which can be used for future research applications. In this chapter I attempt to resolve the inherent data and computational limitations by creating high-resolution datasets and by focusing on a single aspect of model development: how sub-grid scale parameterizations affect model results and accuracy at increasingly finer scales. This chapter is based on the manuscript [*Singh et al.*, 2014b], which has been submitted for publication.

In Chapter 3 I take the high-resolution model developed in earlier chapters and develop a methodology to assimilate observed water table depth data into CLM4.0. The goal is to evaluate the improvement in hydrologic predictions (i.e., water table depth, surface soil moisture, root zone soil moisture, latent heat, ground evaporation, surface runoff, drainage, and infiltration) of a high spatial resolution (100 m) version of CLM4.0 within a 1° by 1° domain in the Sierra Nevada foothills in Northern California through assimilation of observed groundwater depth measurements from multiple wells in the region. The study shows that the water table depth can be calculated with improved accuracy through the application of kriging-based assimilation methods in a high spatial resolution land surface model. We performed evaluations to determine the best method for groundwater assimilation and also analyzed the effect of assimilation on other water and energy budget terms in CLM4.0, including runoff, soil moisture, ground evaporation, and sensible heat. I show that assimilation successfully yields a good approximation of water table depth in CLM4.0 throughout the region across all seasons. This chapter is based on a manuscript [*Singh et al.*, 2014a], which has been submitted for publication and is now under revision.

Chapter 4 describes the collaboration with the private company Wellintel Inc. to implement our modeling methodology with observation data assimilation on a region in central California centered near Paso Robles. Wellintel has already installed a dozen pilot wells and is installing thousands of low-cost and very reliable water table depth measuring devices, which gives them one of the best observational datasets on local groundwater variations. This chapter details how we assimilate Wellintel's observation data into our high-resolution model through the

methodology described in earlier chapters. This gives us the capability to better model groundwater for the recent past and present, and also the potential to force our model with climate projections to probabilistically predict groundwater for future climate-change scenarios. The project described in this chapter is ongoing as we await installation of more Wellintel sensor devices and the finalization of our near-real-time CLM+DART setup. This chapter is also being turned into a manuscript for publication.

In Chapter 5 I summarize the results of this dissertation and offer my concluding remarks and recommendations. This section also describes future work that could be undertaken to improve upon the results from this research and take it forward.

1.7 Summary of Contribution

The work presented here aims to increase our understanding of high-resolution LSM simulations at large spatial domains that assimilate observed groundwater data resulting in improved model predictability with reduced uncertainties.

Key contributions of this work include the following:

1. Running CLM4.0 simulations at high-to-hyper resolutions of 1 km and 100 m over a very large spatial domain with high-resolution input data, while demonstrating that the process can be easily expanded to global scales.

2. Quantifying the effects of changing model grid resolution on CLM4.0 physics, and demonstrating how sub-grid scale parameterizations affect model results at increasingly finer scales. This shows the resolution at which parameters work well, and so highlights the need to develop either robust parameterizations at required resolutions or (even better) scalable parameterizations.

3. Proving that grid resolution itself is also critical for accurate model simulations and for hydrologic budget closure, but also that there are no improvements in modeling when the model resolution increases from ~100 km to ~25 km, as this is too coarse. Parameter calibration efforts are potentially fruitless (or at least less effective) if an appropriate model resolution is not achieved first.

4. Demonstrating that the higher-resolution model fails to show any improvement in the simulated water table depth due to faulty model parameterizations for this term.

5. Developing a kriging-based interpolation scheme to solve the problem of spatially and temporally sparse observation data. These data are necessary for generating input data for assimilation.

6. Developing the DART+CLM groundwater assimilation methodology with help from the NCAR DaRES group and the direct insertion assimilation methodology in CLM4.0. I also demonstrate that the methodology can be expanded to other LSMs with groundwater components. This method is most useful in areas with sparse observation networks, significant groundwater dependence, and inadequate recharge.

7. Demonstrating that groundwater assimilation in the high-resolution CLM4.0 significantly improves modeling results for water table depth. The improvement is more significant in regions with deep water table profiles, as CLM4.0 structurally performs very poorly in regions with water table depths of more than 5 meters.

8. Collaborating with a private company to complement their observation data resources with my high-resolution model using the new assimilation methodology, and so providing users with a better assessment of groundwater levels (both current and projected) at various spatial resolutions.

Chapter 2: Towards hyper-resolution land surface modeling: The effects of model grid resolution on simulations of the Southwestern US

2.1 Introduction

Hyper-resolution land surface modeling has been stated to be one of the 'grand challenges' for the land surface modeling community. One way to approach the challenge of better global modeling is to adapt a more established and widely used global LSM for high-resolution simulation and to draw lessons from the results. In this study, we examine the effects of increasing spatial resolution in a popular LSM with a particular focus on hydrologic and energy budget processes. We have tried to work around the inherent data and computational limitations by creating our own high-resolution datasets, and by focusing on a single aspect of model development: how sub-grid scale parameterizations affect model results at increasingly finer scales. We hope that these regional-scale results will yield information that can be easily expanded to global scales.

Here, we apply NCAR's Community Land Model version 4 (CLM4.0) [*Oleson et al.*, 2010b] at 0.01-degree (~1km) resolution over the southwestern United States, and at 100m resolution for a nested sub-domain located along the California Central Valley and Sierra Nevada foothills. We test the effects of 1-km topographic and soil texture information in CLM4.0, and the spatial sensitivity of the water and energy cycle variables that depend on these fields. The resulting model outputs are analyzed to understand how model physics are affected by changing model grid resolutions. Model outputs are also compared to regional observations, providing some qualitative direction for model improvement efforts, and offering information about the necessary grid-resolutions and surface data sets for improved model accuracy.

2.2 Methods

CLM4.0 was run in off-line mode over the southwestern United States at three spatial resolutions; "hyper-resolution scale", or ~1km grid cells (0.01° x 0.01°), "present high-resolution global model scale", or ~25km grid cells (0.23° x 0.31°) and "normal global model scale", or ~100km (0.9° x 1.25°). The model was run for six years (2000-2005), for which good forcing and observation datasets were available, and for which the surface datasets in CLM4.0 are optimized.

Additionally, a "hillslope scale" 100m-resolution, 1-year simulation from 1 January 2003 to 31 December 2003 was conducted within a small test area to evaluate the model physics and spatial sensitivity of the model outputs at such high resolutions.

2.2.1 Model description

This study is based on simulations of CLM4.0, the land component of the National Center for Atmospheric Research Community (NCAR) Earth System Model (CESM 1.0.4) [*Oleson et al.*, 2010b]. CLM4.0 is one of the most commonly used LSMs and has deep sub-surface component combined with a complex above surface description of processes. CLM has been used on regional and global scales and as a part of the CESM system it is well suited to be readily combined with atmospheric, ocean and sea ice models. CLM4.0 has been explained in detail in the chapter one. It has been setup to run at resolutions of 1km and 100m and run in offline mode with atmospheric forcing data for this study.

2.2.2 Study area

The study area is the southwestern United States, and includes California, Nevada, and parts of Oregon, Idaho, Utah and Arizona. The area contains large climatic variations, has major river basins and has been an ongoing focus area for hydrologic studies by the authors. The coordinates for the 1km resolution model domain are 113.3^{o}E-124.5^{o}E x 31.4^{o}N-43.4^{o}N, and it is divided into 1200 x 1120 grid cells, each 0.01 degrees (Figure 2.1). The 100-m resolution test region has coordinates 39.191^{o}N-39.426^{o}N x 238.90625^{o}E -238.59375^{o}E. This sub-domain is located in the Sierra Nevada foothills and is divided into 280 x 370 grid cells (Figure 2.1 inset). This sub-domain includes the Sierra Foothills Research and Extension Center (SFREC) research site run by University of California. The larger domain contains two Fluxnet sites, one Soil Climate Analysis Network (SCAN) site, and one United States Department of Agriculture – Agricultural Research Service (USDA-ARS) site, with data in the required time period and which were used to obtain data for snow water content, soil moisture content, sensible heat flux and latent heat flux for evaluation of model output.

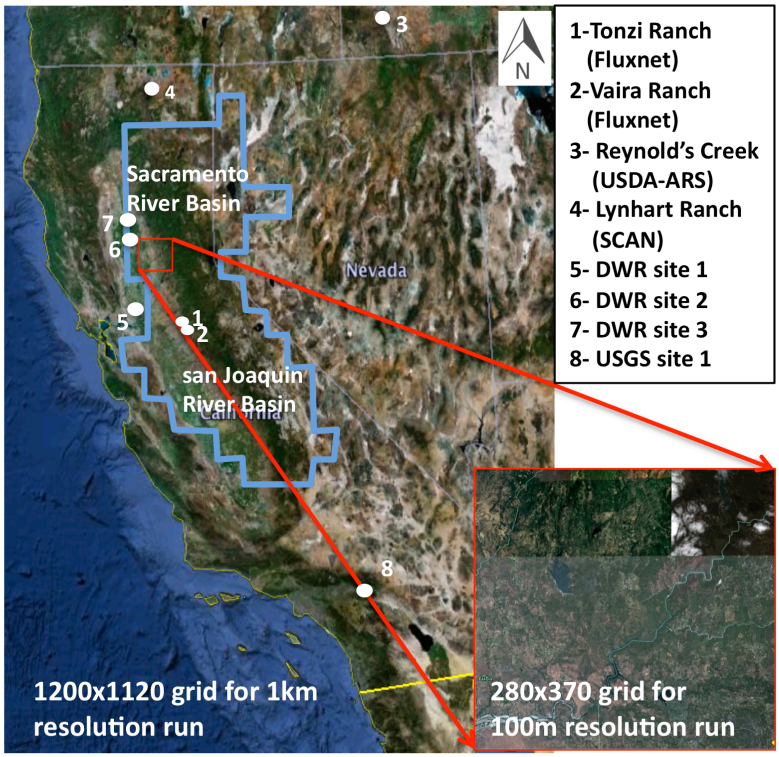

Figure 2.1: Study area over which the model was run, showing the 1-degree grid cells used in masking the GRACE observations and the observation data stations.

2.2.3 Data

2.2.3.1 Model surface datasets

Creating high-resolution surface datasets is one of the biggest challenges associated with hyper-resolution land surface modeling. Topography is one of the most important parameters in determining surface water flows, and is a key control on soil moisture variability at high resolution [*Famiglietti et al.*, 1998; *Jana and Mohanty*; b]. A previous analysis of TOPMODEL results showed that model predictions of the depth to the water table, the ratio of overland flow to total flow, peak flow, and variance and skew of predicted stream flow were all affected by the DEM resolution [*Wolock and Price*, 1994]. Soil texture also plays an important role in the calculation of soil conductivity parameters, and studies have shown that improving soil depth information improves land surface modeling results [*Gochis et al.*, 2010].

Following [*Oleson et al.*, 2010a], surface datasets were created at each spatial resolution for topography and soil texture. In previous studies, use of a 1/3 arc-second DEM dataset is viewed as a ‘‘reasonable compromise’’ between the need for

fine-grained accuracy and the demands of a high data volume [*Wolock and McCabe*, 2000; *Zhang and Montgomery*, 1994]. The topographic data for the model runs in this study was generated at 1-km and 100-m resolutions using 1/3 arc-second (~10m) resolution data available from NED USGS. The Topographic Index (TI) values were calculated from 1/3 arc-second DEM using the ArcGIS™ software following the method described in [*Quinn et al.*, 1995]. The resulting 1/3 arc-second resolution TI was used to calculate f_{max} values at 0.01° (~1km) resolution using the method described by [*Niu et al.*, 2005] (Figure 2.2).

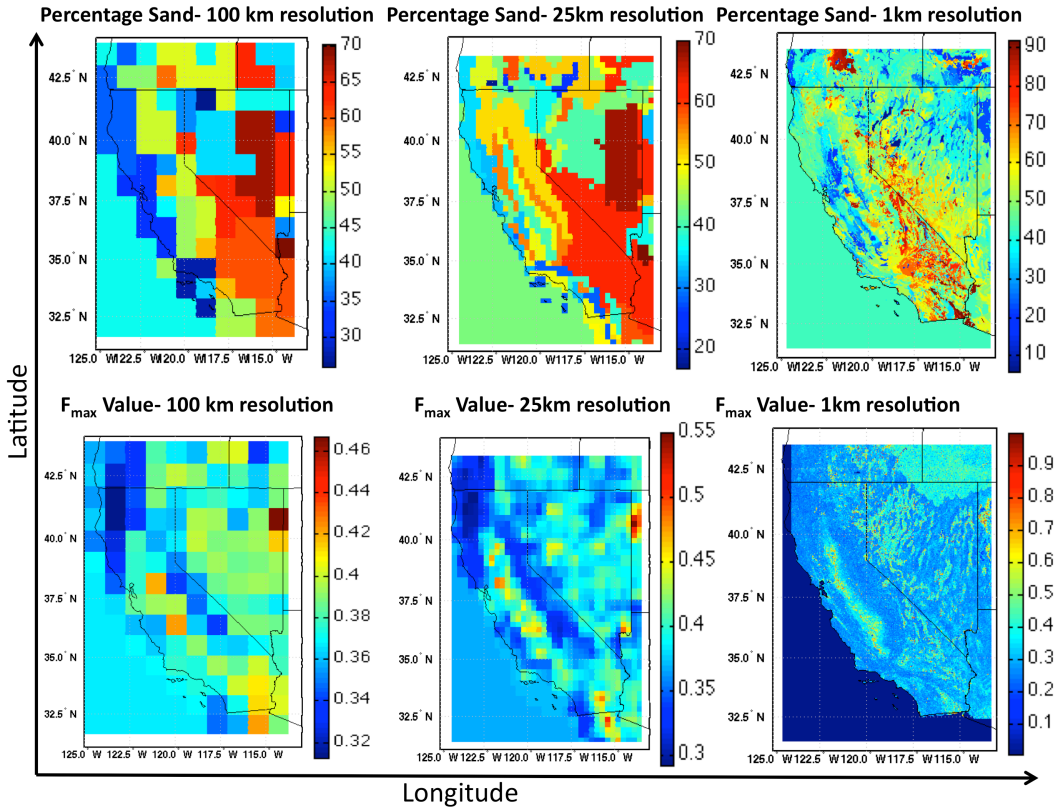

Figure 2.2: Comparing between ~100km, ~25km and ~1km resolution data sets for soil sand fraction and F_{max} for the test region

The distribution plot of f_{max} values at various resolutions (Figure 2.3) shows that the 1km resolution values have a broader and more continuous range, and represents a distribution more comparable with reality. The mean values of the 1km resolution topography are also lower than those of the coarser resolution topography, as a larger number of steep slopes are captured in the high-resolution data. For steeper slopes, the TI calculated tends to be lower, and thus the f_{max} values are lower.

The high-resolution soil texture dataset was produced using the United States Geological Survey (USGS) State Soil Geographic (STATSGO) [*Miller and White*, 1998] dataset at a 30 arc-second resolution. Higher (~100m) resolution Soil Survey Geographic database (SSURGO) datasets were not available for the whole domain,

so a nearest neighbor interpolation was used to create soil texture data at 100m resolution. Figure 2.2 shows the percentage sand value and f_{max} calculated at 1km resolution compared to coarser resolution data. For the 100m-resolution simulation f_{max} was calculated at 3 arc-second (~100m) resolution for the smaller region (Figure 2.1, inset), while soil texture remained at the 30 arc-second resolution.

Except for the topographic and soil texture data, which were calculated at 1km resolution by the authors, all other surface and aerosol input data were provided by the NCAR CESM forcing dataset library at 0.23x0.31 degree resolution .

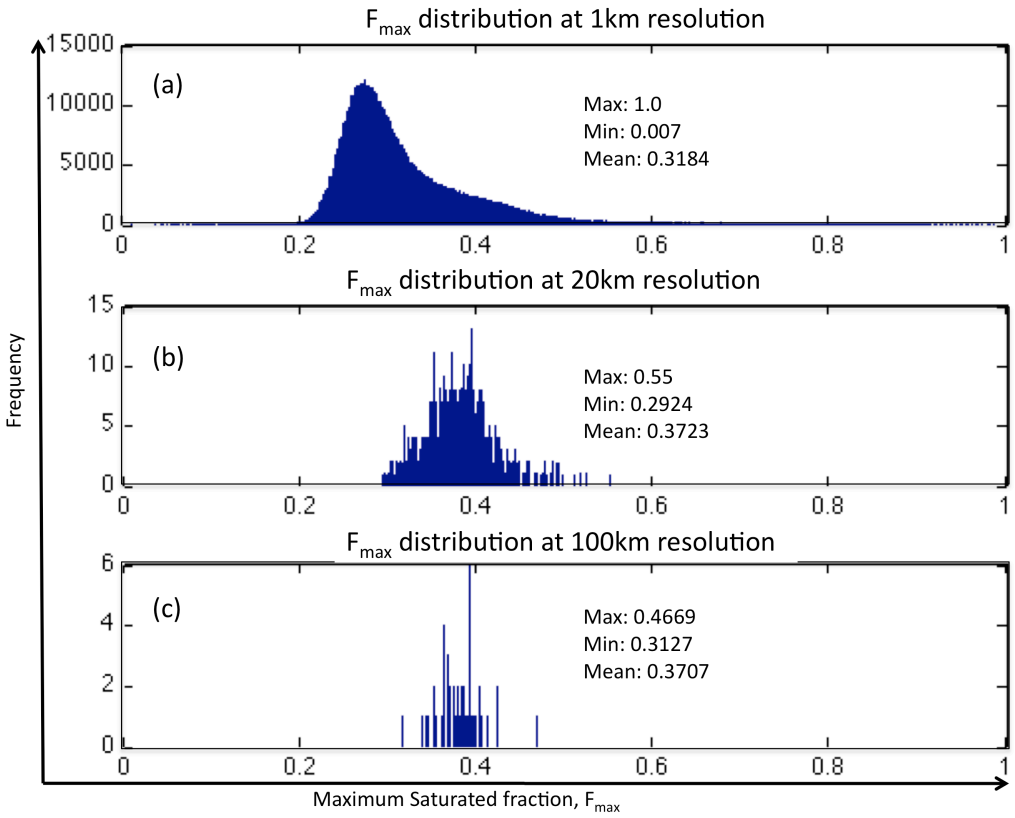

Figure 2.3: Distribution of maximum saturated fraction (f_{max}) across the region at various resolutions.

2.2.3.2 Model forcing datasets

Simulations at each resolution were forced with the 1/8-degree (0.125° x .125°) resolution, 01 Jan 1979 to 31 Dec 2010 hourly atmospheric forcing data from North American Land Data Assimilation System (NLDAS-2) atmospheric data, [*Mitchell et al.*, 2004b] rather than the usual T62 resolution NCAR provided forcing data [*Qian et al.*, 2006]. The NLDAS-2 domain, spatial resolution, computational grid, terrain height, and land mask in NLDAS-2 are identical to NLDAS-1 [*Mitchell et al.*, 2004a]. The initialization files were created after spin up of the model from bare soil at each

resolution to reach thermal and hydrologic equilibrium [*Lo et al.*, 2008]; in this case the models were spun-up for 21 years from 1979 to 1999.

2.2.3.3 Observation datasets

The model outputs were compared with observation data to test the improvements in the higher resolution model. Soil moisture observation data was obtained from the US Department of Agriculture-Agriculture Research Service [*Jackson et al.*, 2010] site at Reynolds Creek, the Soil Climate Analysis Network (SCAN) site at Lynhart Ranch, and the FLUXNET sites at Tonzi Ranch and Vaira Ranch [*Baldocchi*, 2011].

Sensible heat and Latent heat observation values were obtained from the FLUXNET sites at Tonzi Ranch and Vaira Ranch. Groundwater observation data was obtained from various wells maintained by the California Department of Water Resources (DWR) and the US Geological Survey (USGS). DWR has hundreds of monitoring wells in California and USGS has a few sites in this region too. Most wells did not have enough observations during the model run time period and were excluded from our analysis.

Snow water equivalent data was obtained from the National Snow and Ice Data Center (NSIDC) Snow Data Assimilation System (SNODAS) data base [*Center*, 2004]. Terrestrial Water Storage Anomaly (ΔTWS) was calculated from NASA's Gravity Recovery and Climate Experiment (GRACE) data at the University of Texas at Austin, Center for Space Research (CSR) [*Wahr et al.*, 2004], and the accompanying error estimates, were constructed from level-3, Release 05 (RL-05) GRACE gridded land solutions, available on the GRACE Tellus website (www.grace.jpl.nasa.gov).

2.2.4 Computational considerations

There were considerable computational and storage considerations for the 1km resolution model to run over the southwestern US region. We used the parallel computing clusters at Lawrence Berkeley National Lab for the model runs. For this study we used computing 360 nodes to run the model, which took 2 days wall clock time to complete 1 year of model simulation with 500GB of output data. A global run at this resolution or higher would require very high computational resources and is one of the main reasons why we wanted to test the model at this fine scale over the southwestern US only.

2.3 Results

Model output from the 1km-resolution simulations were compared with the 25km and 100km resolution simulations and also with observations. We used the mean values (Table 2.1), Pearson's correlation coefficient (*r*) (Table 2.2), Root Mean Square Error (RMSE) (Table 2.3), and model Bias (Table 2.4) to show agreement across simulations and to quantify reduced errors associated with higher resolution simulations. Results from the 100m-resolution simulations were used to test selected variables that are more sensitive to slope, such as surface runoff, drainage and infiltration. Individual results are described below.

Table 2.1: Mean soil moisture content, infiltration, surface runoff, and sub-surface runoff at 1km, 20km, and 100km resolution.

Spatial mean	1km resolution	20km resolution	100 km resolution
Soil Moisture (mm^3/ mm^3)	0.1430	0.1364	0.1370
Infiltration (mm/day)	0.4197	0.3872	0.3906
Surface Runoff (mm/day)	0.1786	0.3419	0.4169
Sub-Surface Runoff (mm/day)	0.3996	0.7535	0.9625

Table 2.2: Average Correlation Coefficient (r) between observation data and model outputs at various resolutions.

Correlation Coefficient (r)	Soil Moisture (mm^3/ mm^3)	ΔTWS (mm)	Latent Heat (W/m2)	Sensible Heat (W/m2)	SWE (mm)
Observation-1km CLM	0.807	0.735	0.718	0.914	0.880
Observation-25km CLM	0.651	0.525	0.5831	0.830	0.650
Observation-100km CLM	0.640	0.454	0.616	0.830	0.360

Table 2.3: Average Root Mean Square Error (RMSE) between observation data and model outputs at various resolutions.

RMSE	Soil Moisture (mm^3/ mm^3)	ΔTWS (mm)	Latent Heat (W/m2)	Sensible Heat (W/m2)	SWE (mm)
Observation- 1km CLM	0.077	214.130	22.724	20.737	0.081
Observation-25km CLM	0.104	290.156	26.160	27.850	0.091
Observation-100km CLM	0.100	262.589	23.356	30.093	0.117

Table 2.4: Average Bias between observation data and model outputs at various resolutions.

RMSE	Soil Moisture (mm^3/ mm^3)	ΔTWS (mm)	Latent Heat (W/m2)	Sensible Heat (W/m2)	SWE (mm)
Observation- 1km CLM	0.0439	20.04	12.843	6.397	-0.74
Observation- 25km CLM	0.0788	17.1	12.420	6.146	0.99
Observation- 100km CLM	0.0111	16.77	4.396	10.783	1.40

Table 2.5: CLM snow water equivalent statistics for three model grid resolutions.

Model SWE	1km	25km	100km
Mean (mm)	4.24	3.58	3.98
Standard dev. (mm)	21.47	22.56	23.39
Maximum (mm)	354.47	340.74	206.7
Total volume (mm^3)	6.33x1020	2.72x1020	3.01x1020

2.3.1 Soil Moisture

At higher resolution the maximum saturated area (f_{max}) decreases, this should lead to decreases in the fractional saturated area (f_{sat}). Figure 2.3 shows a 17% decrease in mean f_{max} value and Figure 2.4 shows a 10% decrease in the mean f_{sat} value at 1km resolution, as compared to at 20km resolution simulation. It also leads to a decrease in runoff, increased mean infiltration values, and thus an increase in mean soil moisture content (Table 2.1). Increasing resolution also increases the distribution range of f_{max} and f_{sat} values seen in Figure 2.3 and Figure 2.4, respectively, as a wider variety of slopes are taken into consideration. The water table depth at a location also affects the f_{sat} distribution, and the regions with deep water table and the regions with shallow water table depth are reflected by the bimodal shape of the histogram. The f_{sat} distribution affects the soil moisture content and a similar bimodal distribution is found for the surface (upper layer) soil moisture content distribution (Figure 2.5), the bimodal distribution here is reflective of the wet and dry regions and also of the seasons.

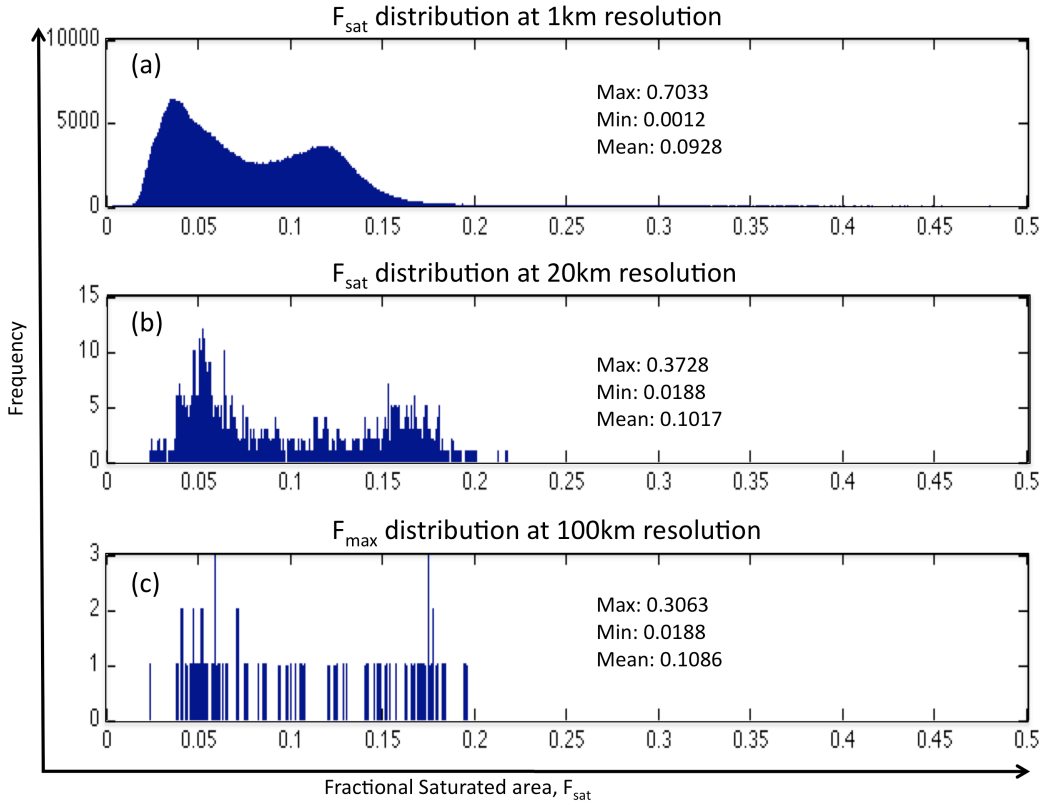

Figure 2.4: Comparison of the distribution of fractional saturated/impermeable area (f_{sat}) at various resolutions over the test region. (a) 1km resolution CLM (b) 20 km resolution CLM (c) 100 km resolution CLM

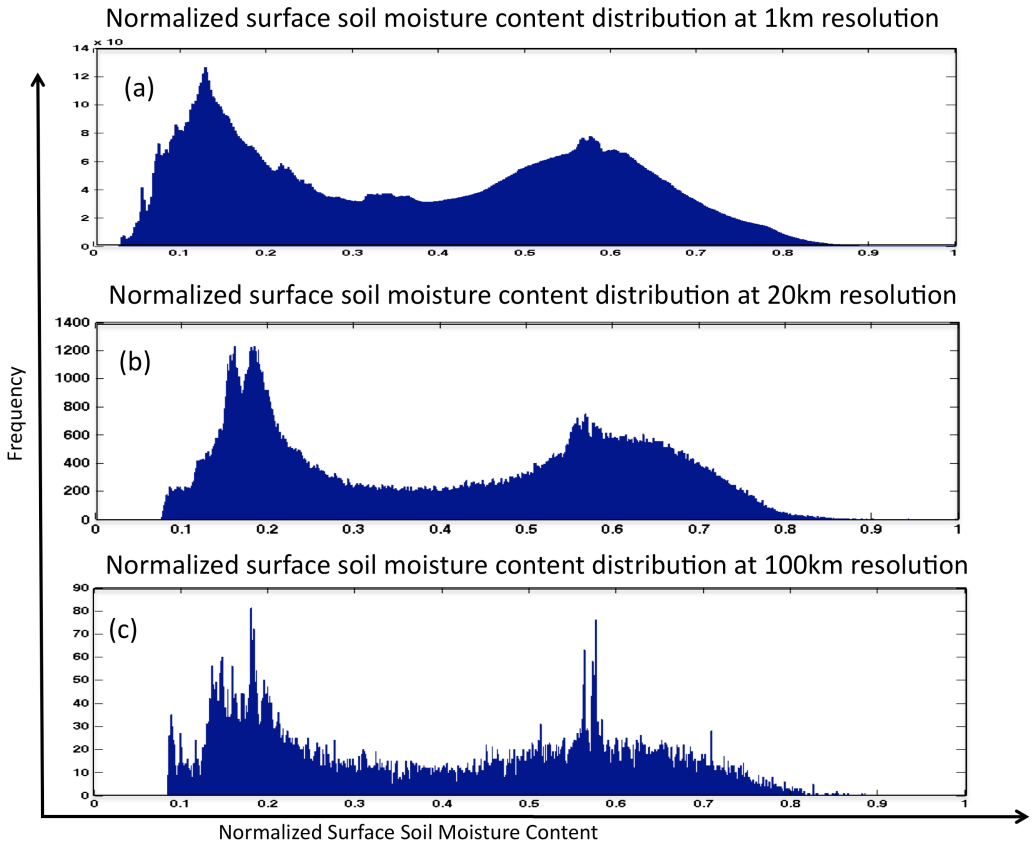

Figure 2.5: Distribution of normalized surface soil moisture content at various resolutions over the test region. (a) 1km resolution CLM (b) 20 km resolution CLM (c) 100 km resolution CLM

Model simulations at 100m, 1km, and 100km for 2003-2005 are compared to observations at three FLUXNET sites: Tonzi Ranch, Vaira Ranch, and Reynolds Creek (Figure 2.6). All simulations follow the observed seasonal trend fairly well, but the 1km resolution simulation shows marked improvement over the 25km and 100km resolution simulations especially when short term trends and sudden variations in soil moisture are taken into account. The model-to-observation correlation coefficients for the Tonzi Ranch FLUXNET site are 0.9291 for 1km, 0.7761 for 25km, and 0.7763 for 100km, thus showing an improvement of 19.7% over the 1km and the 25km resolution simulations (Figure 2.6, a). Improvements in the correlation coefficients of 15.5% and 43.2% respectively were seen in the Vaira Ranch site and the Reynold's Creek site (Figure 2.6 b and c). Comparing mean values for all sites the 1km resolution simulation showed an improvement of 23% in correlation coefficient (Table 2.2) and a 35% improvement in RMSE (Table 2.3) compared to the 20km resolution simulation.

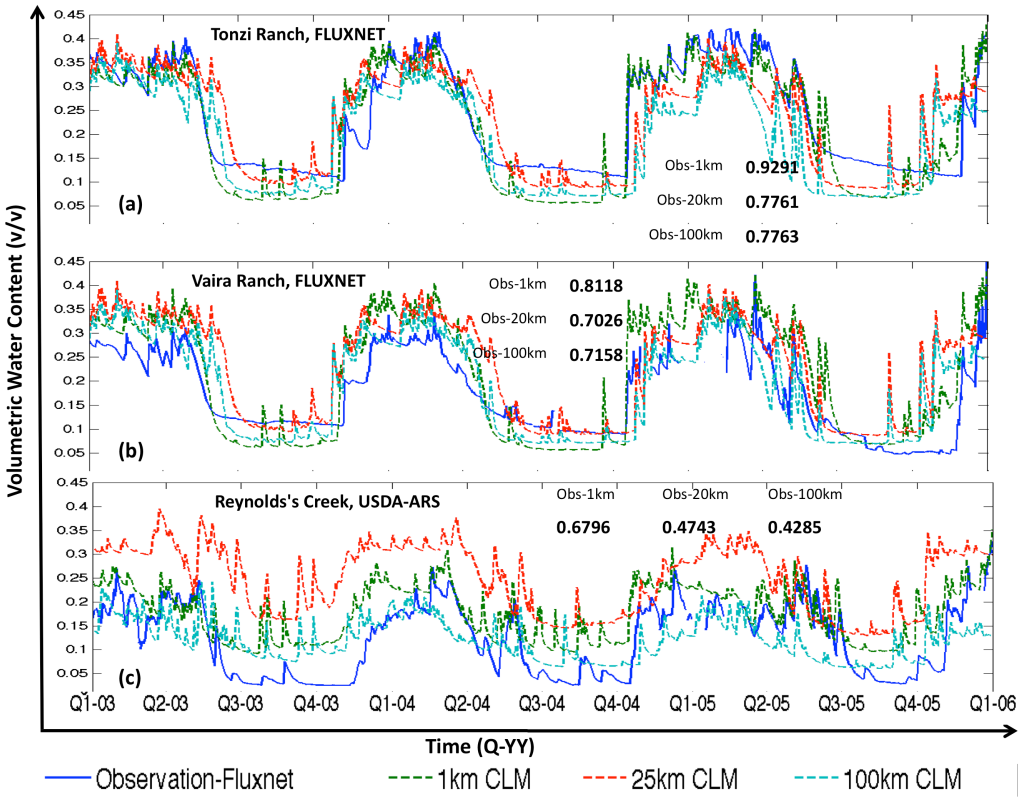

Figure 2.6: Comparison of surface Soil Moisture model outputs with observation (a) FLUXNET, Tonzi Ranch site, (b) Fluxnet, Vaira ranch site, and (c) USDA-ARS, Reynold's Creek site. Correlation coefficient values in table.

2.3.2 Snow Water Equivalent

Maps of time-averaged Snow Water Equivalent (SWE) for the entire comparison period (Figure 2.7) show an increasingly complex spatial structure in SWE with increasing model resolution. The increased heterogeneity includes higher maxima in individual grid-cell SWE amounts with higher variability across the domain, as well as a higher total SWE amount across the domain (Table 2.5). There is a scale dependent difference in the CLM4.0-derived timing of snow pack accumulation and depletion rates resulting in changes to the seasonality of snow storage (Figure 2.8).

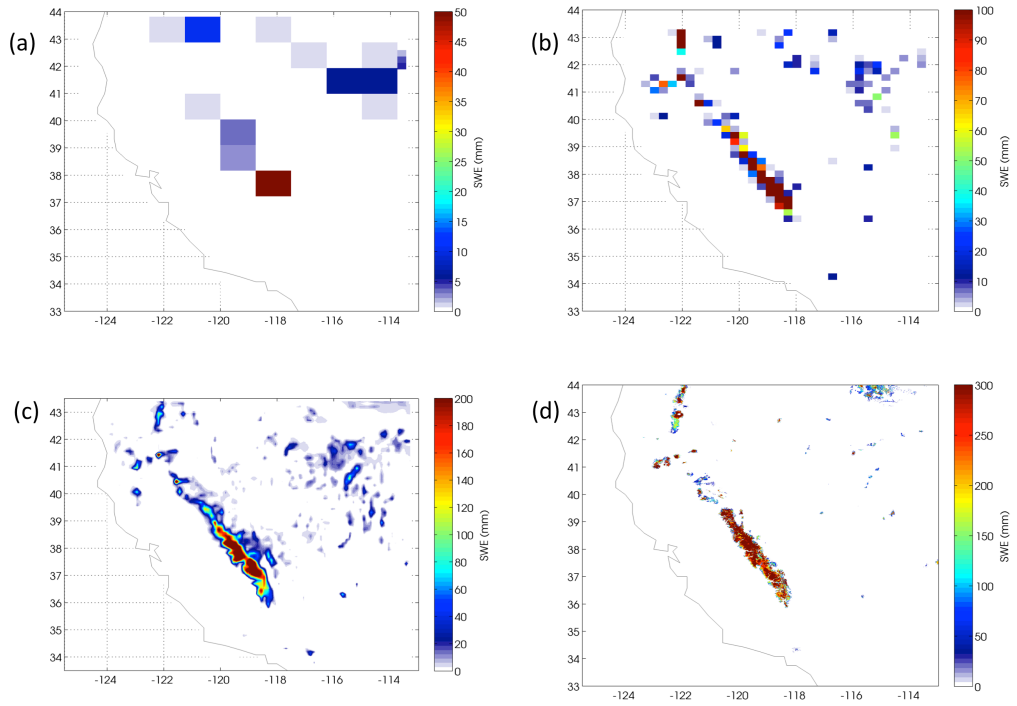

Figure 2.7: Comparison of the time-mean (January 2003 – December 2005) snow water equivalents over the domain for: (a) ~100km resolution simulation; (b) ~25km resolution simulation; and (c) ~1km resolution simulation. These are compared with (d) 2003-2005 SNODAS SWE at 1 km resolution.

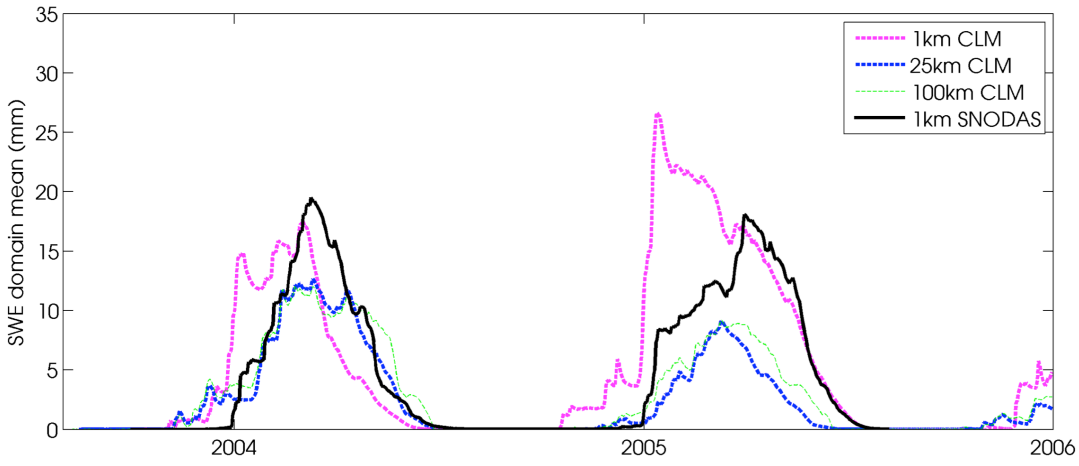

Figure 2.8: Time series of the spatial-mean snow water equivalents over the domain for: ~100km resolution simulation (green); ~25km resolution simulation (blue);

and ~1km resolution simulation (magenta). These are compared with 2000-2005 SNODAS SWE at 12.5 km resolution (black). All time series have been smoothed with a 21-day boxcar filter.

CLM4.0-simulated SWE for the southwestern US is compared with the gridded 1km resolution SNODAS (NOHRSC, 2004) daily data from 1 September 2003 to 31 December 2005 (Figure 2.8). The correlation coefficients (r) between the SNODAS and CLM4.0 domain-total snow water equivalent time series are 0.88 at 1km resolution, 0.65 at 25km resolution, and 0.36 at 100km resolution (Table 2.2). The RMSE between the SNODAS and CLM4.0 is 0.081 m at 1km resolution, 0.091 m at 25km resolution, and 0.117 at 100km resolution (Table 2.3). This shows that there is some improvement in the timing of snow accumulation and snowmelt with the higher resolution model simulation, and a low bias in snow amount for the coarser model resolutions.

Figure 2.9 shows the distribution of the time-mean snow water equivalent within the domain for all three resolutions. The 1km resolution simulation shows a more continuous spread of snow amount across the domain, indicating more heterogeneity in grid cell snow compared to the 25 km and 100 km simulations. This is due to the improved distribution of temperatures across the domain, whose time-mean distribution is also shown in Figure 2.9. The dependence of snow cover on temperature affects the total amount of snow within the domain, due to the heterogeneous topography in the 1km simulation allowing for a higher number of cold grid cells.

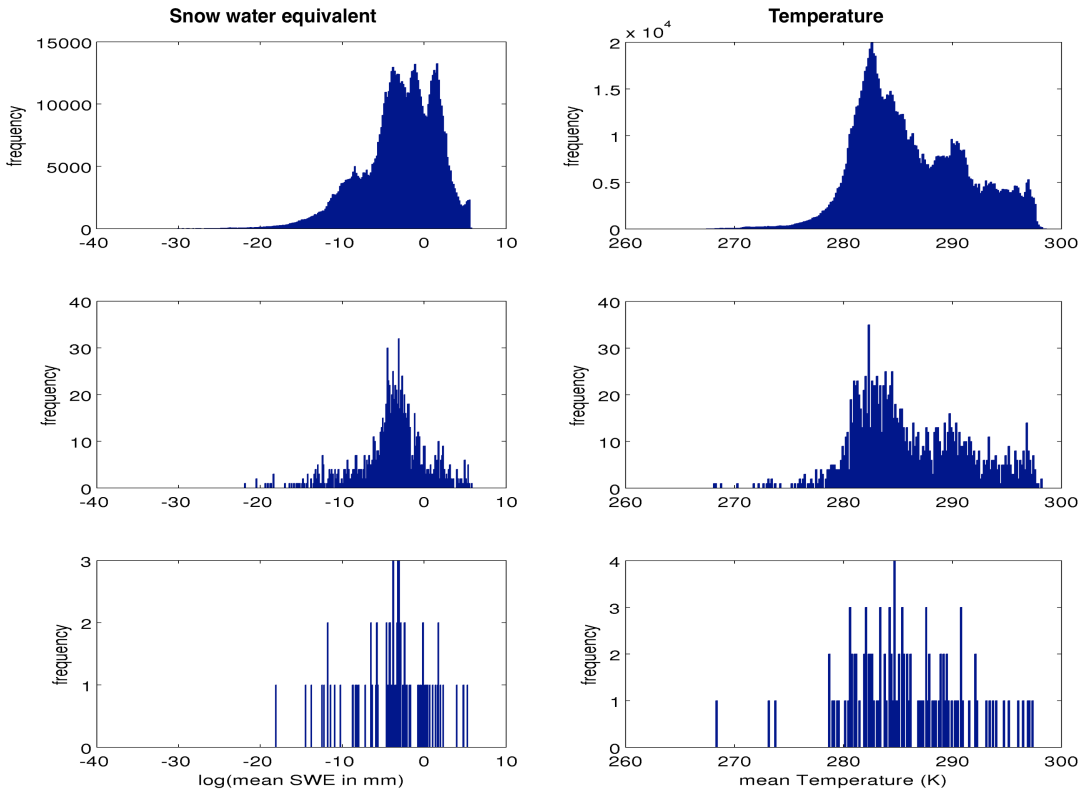

Figure 2.9: Time-mean snow water equivalent distributions across the domain for the three model resolutions (left column). Time mean 2m air temperature distributions across the domain for the three model resolutions (right column). The top row contains the 1-km simulations, the middle row contains the 25-km simulations and the bottom row contains the 100-km simulations.

2.3.3 Water Table depth

The water table depth (ZΔ) values obtained from the models are compared with observation data obtained from dozens of well sites with daily measurement data maintained by the California Department of Water Resources and the USGS. The ZΔ calculation in CLM4.0 has a very high level of parameterization and thus increasing resolution has a minimal effect, as seen in Figure 2.6, where 1km, 25km, and 100km resolution simulations yield similar results. CLM4.0 reasonably well predicts when there is a shallow (< 4m) water table within the region, as it calculates the amount of water in the top ten active soil layers to approximately 4 meters below the surface. However, the model has difficulty in deeper regions.

The ZΔ is estimated by the water balance through a reservoir system [*Oleson et al.*, 2010b], because of this, the ZΔ in CLM4.0 does not vary substantially from the initial

value. To illustrate this impact, outputs from four regions with different ZΔ profiles from shallow to very deep are plotted in Figure 2.10. The ZΔ in CLM4.0 remains constant at a very shallow level of 3-5 meters and has significantly less variation throughout the region. The results indicate that regions where ZΔ is shallow (Figure 2.10a) the CLM4.0-calculated ZΔ is closer to observation values, but in places where ZΔ is below 5-6 meters, the CLM4.0 calculated ZΔ completely misses the observed trend (Figures 2.10 b, c, and d).

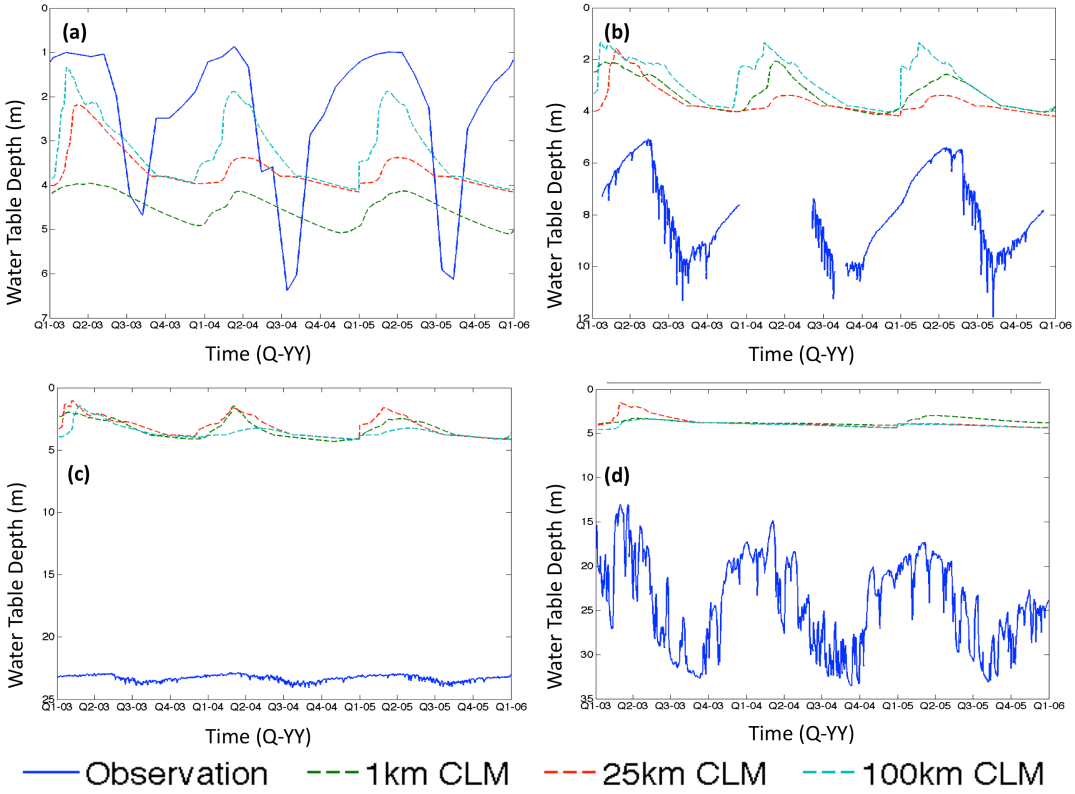

Figure 2.10: Comparison of water table depth at 1km, 25km, and 100km resolution with Observation data. (a) DWR site 1 with shallow water table depth. (b) DWR site 2 with medium water table depth. (c) DWR site 3 with Deep water table depth. (d) USGS site 1 with deep water table depth.

2.3.4 Terrestrial Water Storage

Monthly Terrestrial Water Storage (TWS) in the combined Sacramento and San Joaquin River Basins was calculated at the three resolutions using CLM4.0 output. A river basin mask was created for each model resolution to match the GRACE resolution basins for extracting the output data. Mean-monthly CLM4.0 data was then used to generate the TWS time series for these river basins from 1 January 2003 to 31 December 2005 by summing the model-based groundwater, soil

moisture content and snow water equivalent at each time step.

Generally, our results show that as model resolution increases the magnitude of TWS and its variability increases. This is due to the decrease in runoff as described in section 3.1, and also to changes in SWE as described in section 3.2. The relatively little change in $Z\Delta$ across resolutions does not contribute to TWS differences between resolutions.

We calculated the model terrestrial water storage anomaly (ΔTWS) and compared these results with the Gravity Recovery and Climate Experiment (GRACE)-derived monthly observation of ΔTWS (Figure 2.11), where ΔTWS was calculated as the difference between the monthly mean values and the mean over the years 2003-2005. The GRACE observations and their accompanying error estimates were obtained from the latest release version (RL05) of the GRACE dataset, and were subset to match the time period of the model simulations.

The correlation coefficients between CLM4.0 and GRACE ΔTWS are 0.735 for 1km, 0.525 for 25km, and 0.4535 for 100km resolution simulations (Table 2.2). The ΔTWS RMSE values also show significant improvement at higher resolution, where 1km is 210 mm, 25km is 290 mm and 100km is 260mm. The 1km resolution shows an improvement of 40% in the correlation coefficient as compared to the 25km resolution, and it shows an improvement of 35% in RMSE values when compared to the 100km resolution (Table 2.3). Even though the 1km RMSE and correlation coefficient shows significant improvement, the RMSE values are still very high. This may be attributed to less than satisfactory groundwater processes within CLM4.0, as groundwater variation is by far the largest contributor to the ΔTWS calculation.

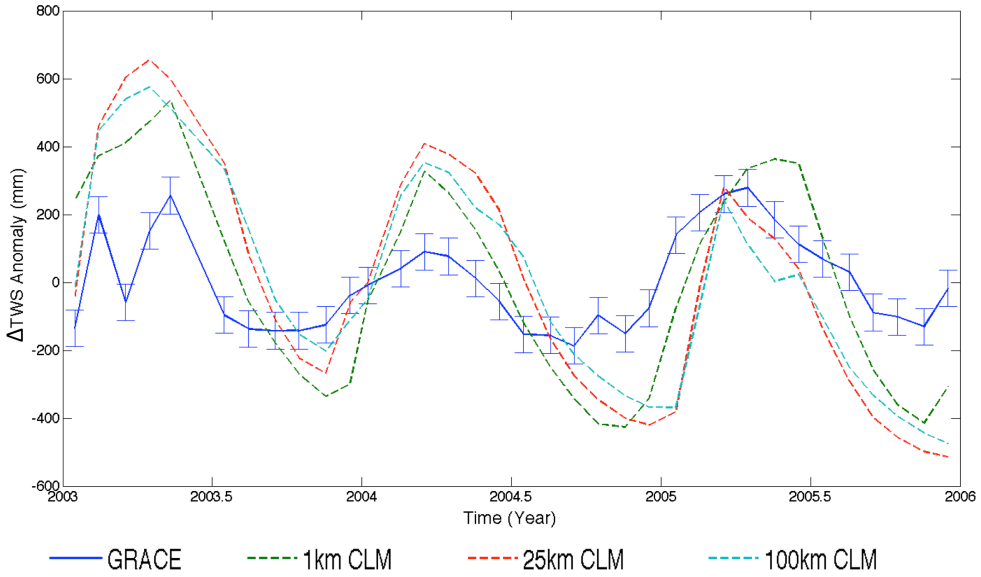

Figure 2.11: Comparison of ΔTWS from GRACE observations (with error bars) with ΔTWS calculated from model outputs at various resolutions for Sacramento-San Joaquin River Basin.

2.3.5 Sensible and Latent heat fluxes

Increasing the model grid resolution and adding more realistic topographic detail generally improves model temperature simulation by creating more realistic elevations and surface slopes. The change in runoff driven by changes in f_{max} affects the amount of moisture available for evaporation, which in turn affects the division of sensible and latent heat fluxes. In regions with relatively more water and sufficient energy input, the higher resolution model consumes more energy in latent heating.

The sensible heat and latent heat calculated in the CLM4.0 simulations were compared with observations obtained from the FLUXNET sites at Tonzi Ranch and Vaira Ranch for the period 2003-2005. The observation to model correlation coefficients for sensible heat show the 1km resolution simulation has approximately a 10% improvement over the 25km and 100km resolutions for both the Tonzi and Vaira Ranch sites (Figures 2.12 and 2.13). Similarly for the latent heat values the 1km resolution simulation shows approximately 20% improvement in correlation coefficient compared to coarser resolution outputs (Table 2.2).

RMSE values calculated for sensible heat between the observation and model at 1km show 48% and 24% error reductions for the Tonzi ranch and Vaira ranch sites respectively. For latent heat the 1km resolution simulation shows an improvement of 23% and 11% for the Tonzi ranch and Vaira ranch sites, respectively (Table 2.3). Observation to model Bias (Table 2.4) suggests no significant trend between the model bias and model resolution, thus implying that the increase in spatial resolution does not lead to any systematic errors in CLM4.0. The improved 1km RMSE can be attributed to better resolved surface heterogeneities such as slope and height, which are not captured at 100 km or 25 km resolution. The CLM4.0-simulated sensible and latent heat values at 1km, 25km, and 100km follow the observed seasonal variation closely thus showing that the model physics at seasonal scale are the same at various resolutions and the main difference between the outputs occurs at higher temporal resolution and spatial resolution.

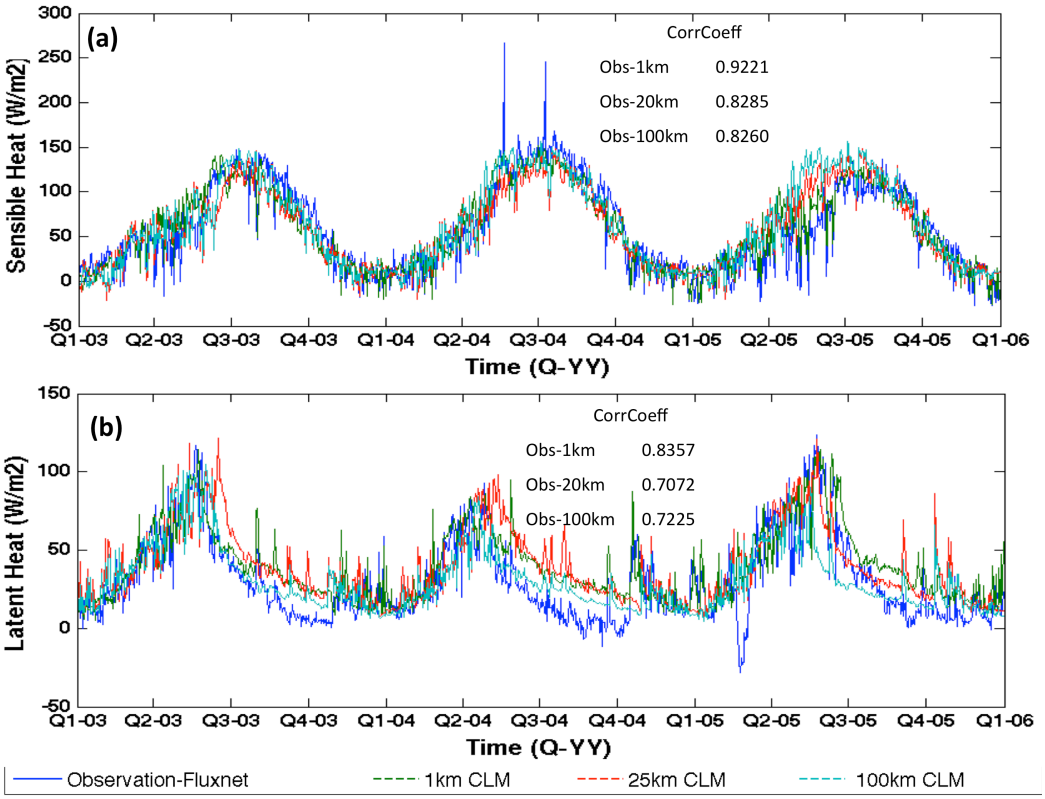

Figure 2.12: Comparison between observed Sensible and Latent heat fluxes with model outputs at Tonzi Ranch site (FLUXNET). (a) Time series of Sensible heat observation and CLM4.0 predictions from 2003-2005 with Correlation Coefficient between observation and predictions in the table (b) Time series of Latent heat observations and CLM4.0 predictions from 2003-2005 with Correlation Coefficient between observation and predictions in the table.

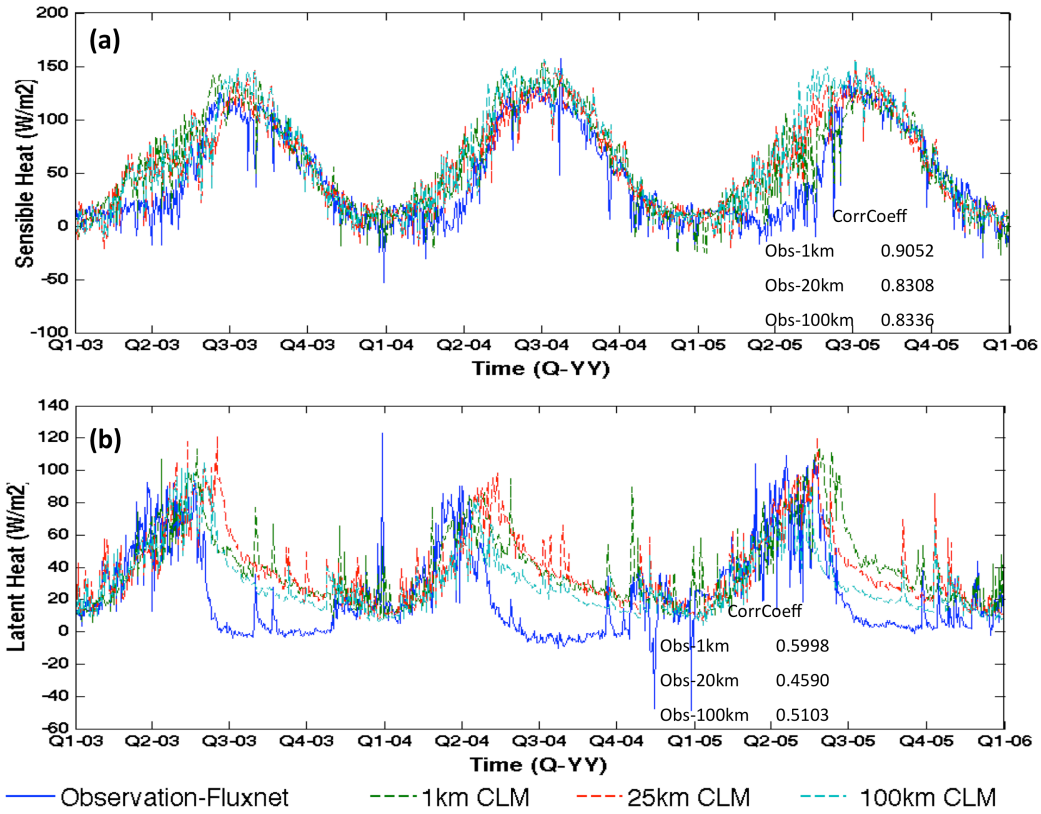

Figure 2.13: Comparison between observed Sensible and Latent heat fluxes with model outputs at Vaira Ranch site (FLUXNET). (a) Time series of Sensible heat Observation and CLM4.0 predictions from 2003-2005 with Correlation Coefficient between observation and predictions in the table. (b) Time series of Latent heat Observation and CLM4.0 predictions from 2003-2005 with Correlation Coefficient between observation and predictions in the table.

2.3.6 Representative hillslope simulation at 100m resolution

As an additional experiment, CLM4.0 was run for a small test region (Figure 2.1, inset) at 3-arc-second (~100m) resolution from 1 January 2003 to 31 December 2003. At this resolution, CLM4.0 approaches the hillslope scale and topographic features, especially extremely steep or flat terrain, are more realistically represented by the TI values. Topography tends to be spatially smoothed by averaging at coarser resolutions. This can be seen in the histogram of the distribution of the fractional saturated area (f_{sat}) values in Figure 2.14, in which the 100m-resolution histogram has longer and smoother tails. The extreme values in the 100m-resolution histogram are representative of the steep slopes not captured at 1km. The effect of scale on topographic smoothing has a significant impact on model outputs, especially runoff, infiltration, and drainage. For comparison between resolutions, we define spatial bias here as the difference between the 1km-resolution output and the interpolated values of the 100m-resolution outputs

aggregated to 1km over the same grid cells within the sub-domain. The biases are plotted against slope as represented by f_{sat} at 1km and 100m (Figure 2.15).

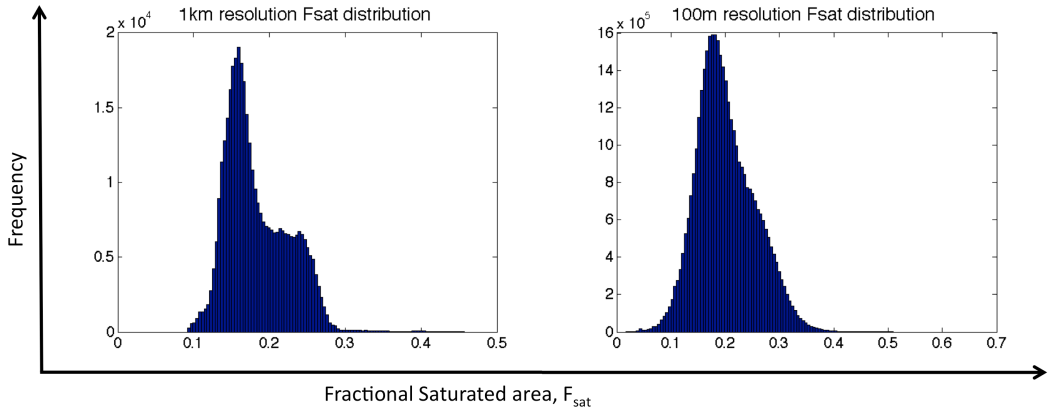

Figure 2.14: Histogram showing the distribution of fsat values in 1km and 100m resolution models and its impact on the Drainage, Runoff and Infiltration values.

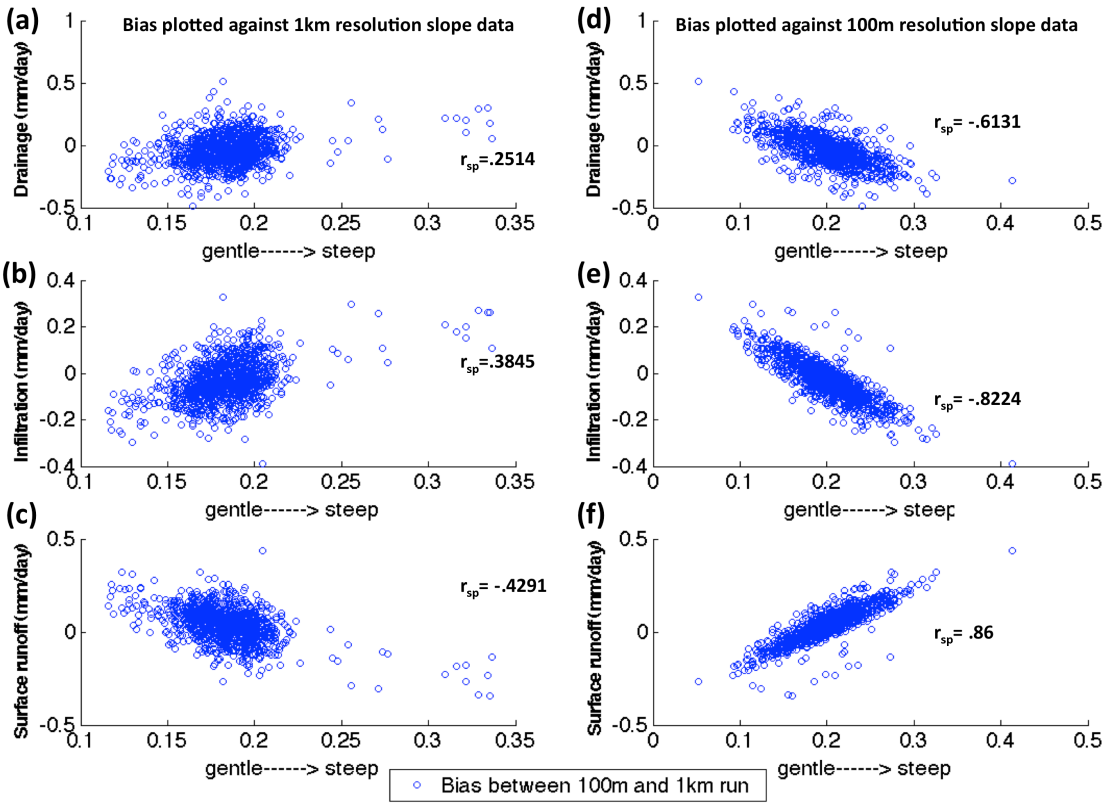

Figure 2.15: Bias between the 1km and 100m resolution runs plotted against slope data, which is represented by the fractional saturated area. 'r_{sp}' is the Spearman rank correlation represents a significant correlation if $r_{sp} > 0.35$. (a) Sub-surface drainage bias plotted against f_{sat} calculated at 1km and (d) 90m resolutions. (b)

Infiltration bias plotted against $\underline{f_{sat}}$ calculated at 1km and (e) 90m resolutions. (c) Surface runoff bias plotted against $\underline{f_{sat}}$ calculated at 1km and (f) 90m resolutions

In comparing the 100m resolution with the 1km resolution, the sub-surface drainage and infiltration rates should decrease and the surface runoff rate should increase with steeper slopes, hence the biases in these rates should also increase as we move toward steeper slopes. It can be seen in Figure 2.15 d, e, and f that the bias does become stronger as we move from gentle terrain to steeper slopes. It's also important to note that when the biases are plotted against the 1km-resolution slope data there is no clear trend in the biases (Figure 2.15 a, b, and c). They show a positive correlation coefficient between sub-surface drainage and infiltration, and a negative correlation for surface runoff values.

The Spearman correlation values (r_{sp}) were also calculated and a statistically significant relationship ($\alpha = 0.01$) between the model bias and slope is indicated by Spearman correlation values greater than 0.35. It can be can seen in Figures 2.15 d, e, and f that the relationship between the bias in model outputs and 100m resolution topographic data is robustly significant, while in Figures 2.15 a, b, and c the 1km-resolution topographic data do not show such a relationship. An analysis between the coarser resolution simulations over the entire southwestern US domain with the coarser topographic data did not indicate such a relationship (not shown). This demonstrates that the slope information is not completely relayed even at the 1km-resolution to CLM4.0 for these variables and the full effect of realistic topography on runoff, infiltration, and drainage can be improved by reaching the true hillslope scale of 100m or less.

The CLM4.0 model physics do not include lateral subsurface flow. To test the effects of this assumption on model hydrology, we examine the distribution of simulated water table depths for a single day in the winter season, 30 January 2003, when the water table was at its highest annual level (Figure 2.16). The water table depths show spatial patterning consistent with the spatial distribution of soil textures and with gradients in topography. Since soil texture information was included at a coarser resolution of 30 arc-second (~1km), which is the highest resolution currently available for simulations with 10-m topographic information, and because lateral flow is not included, there is spatial heterogeneity in water table depths that would likely not be sustainable with lateral flow included.

The map of water table depth gradients [m/100m; calculated as $(dZ/dx^2 + dZ/dy^2)^{1/2}$, where Z is water table depth [m], shows regions where there is a 100m grid cell gradient of greater than 30cm depth in red (Figure 2.16). Approximately 21% of model grid cells showed a significant gradient in water table heights (>0.3%) during the wettest period, while certain regions (<1% of the domain) showed a gradient – or more accurately, a discontinuity – in water table depths of more than 1m. Even with higher resolution soil texture information, these gradients in the domain during the wet season could result in the lateral redistribution of water for a 100m

resolution simulation. The fraction of grid cells showing gradients greater than 0.3% dropped significantly after the wet season, with a mean of approximately 2% through the entire yearlong simulation.

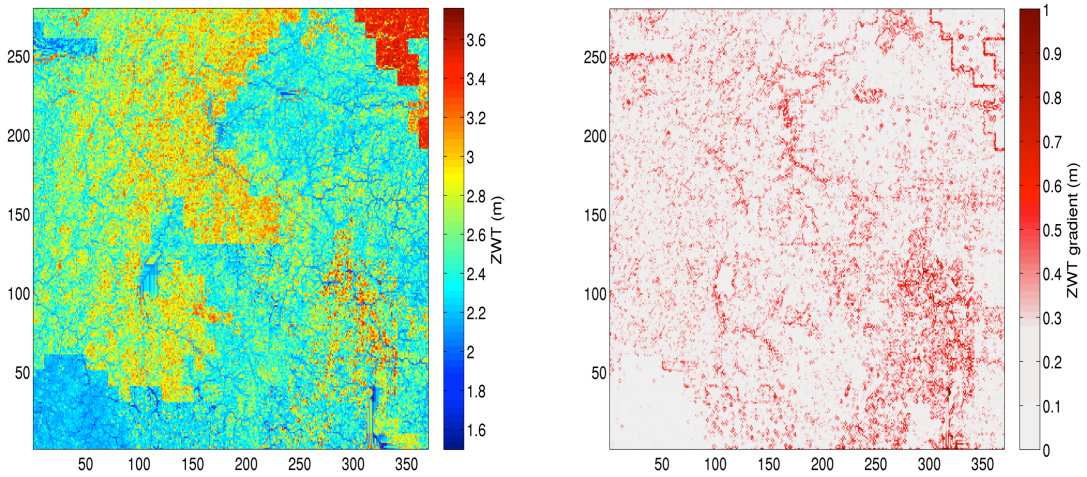

Figure 2.16: The 100m resolution daily water table depth in meters for January 30, 2003 (left panel). The water table depth spatial gradients for the same model day, in meters (right panel). Gradients (per 100m) greater than 0.3 meters are shown in red, and some gradients (<1%) exceeded 1 meter.

2.4 Discussion

The scientific aim of this study is to show high-resolution simulations over a large domain where the accompanying high-resolution topographic and soil texture data differ from coarser resolution simulations using a popular LSM, CLM4.0. A secondary aim of this work was to develop and test a 1km resolution model over the state of California, which can be used for future research applications.

Our results help to identify several scale-based and non-scale-based problems in CLM4.0 model parameterizations. For a number of parameterizations (those used in drainage, runoff and infiltration), the change in resolution has a significant impact, especially when hillslope scales of ~100m are considered. For other parameterizations (those used in the calculation of water table depth), simulation results were not improved by a high-resolution grid and surface data.

Model outputs also do not show much improvement in correlation values or RMSE with observations as the model resolution increases from ~100km to ~25 km. This is because there is little difference in topographic information between smoothened 25 km and 100km resolutions, as these scales are still too coarse. The biases

between model outputs and observations were also calculated to check if the change in CLM4.0 resolution introduces any systematic errors. However, the biases do not show any trends relative to model resolution. This leads us to the conclusion that the increase in resolution of CLM4.0 does not introduce error and the model generally works well for higher resolution applications.

Soil moisture content is calculated by Richard's equation in CLM4.0 and represents an important test of model resolution on the hydrologic cycle, as it is heavily influenced by the resolution of slope and soil texture. Near-surface soil moisture is a key variable within coupled atmospheric and terrestrial hydrologic models used for climate simulations, as it represents a residual of the fluxes in the water balance. Since soil moisture influences the partitioning of water and energy fluxes at the land surface, it has important feedbacks on the energy and water cycles over timescales ranging from hourly to inter-annual [*Reichle et al.*, 2002].

The results show improvement in the CLM4.0 simulation at 1km resolution for surface parameters such as the soil moisture content and sensible heat when compared to observations. These variables are highly affected by the soil texture and surface topography and improve when higher resolution topographic and soil texture data are inputted because of changes in the topography represented in f_{max}. At the resolutions discussed, all simulations follow the seasonal patterns closely with observations. This is expected, as the atmospheric forcing data, which drives the seasonal to annual patterns, is the same for all cases. This result indicates that the model physics perform similarly at all resolutions, and that improvement in the correlation at higher resolution with observations is likely due to the resolution of the input topography and soil texture.

The simulation of snow water equivalent and snow cover area in CLM4.0 is highly sensitive to spatial resolution, and topographic effects could contribute substantially to the water budget through snow storage. For the 2005 simulation there was 74% more snow across the study domain at 1km, as compared to the 25 km resolution. The effect of topography on snow alters the timing and magnitude of snowmelt runoff generation, and has immediate implications for CLM4.0-based water resource applications. In coarser resolution CLM4.0 simulations, where complex mountain terrain is represented as a broad plateau, altered snowmelt and runoff generation leads to a change in the duration of the snow season and a bias in total snow amount across the domain. This supports previous work showing that sub-grid heterogeneity in surface characterizations play a major role in snow formation and melt [*Jin and Miller*, 2007]. Consequently, the CLM4.0 density-dependent snow cover fraction may require significant calibration to be effectively employed at coarse scales (e.g. 0.25 degrees).

Snow cover can have a cumulative impact on long-range coupled climate simulations through land-surface and atmosphere feedbacks and changes in the timing and magnitude of freshwater input to the oceans. Because of the CLM4.0 model design, the presence of snow in a grid cell can have feedbacks on the coupled

climate system through changes in surface albedo. The overall direct beam ($\alpha^{\mu}_{g,\Lambda}$) and diffuse ground ($\alpha_{g,\Lambda}$) albedos in the CLM4.0 are weighted combinations of "soil" and snow albedos.

$$\alpha^{\mu}_{g,\Lambda} = \alpha^{\mu}_{soi,\Lambda}\left(1 - f_{sno}\right) + \alpha^{\mu}_{sno,\Lambda} f_{sno} \qquad 2.1$$

$$\alpha_{g,\Lambda} = \alpha_{soi,\Lambda}\left(1 - f_{sno}\right) + \alpha_{sno,\Lambda} f_{sno} \qquad 2.2$$

where f_{sno} is the fraction of the ground with snow cover (Niu and Yang 2007). These directly affect the direct and diffuse radiative fluxes absorbed by the vegetation and land surface. In order to obtain any reasonable accuracy in higher order processes (such as the effects of black carbon or grain-size snow aging), these albedos need to be faithfully approximated to achieve the correct base line (first order) snow amount for the grid cells within the study region.

The higher resolution model fails to show any improvement in the simulated water table depth. This can be explained by the water table depth calculation in CLM4.0, which is highly parameterized with a specific yield fixed at 0.2 throughout, an initial amount of water in the aquifer is fixed at 4800mm, and the water table is initialized at 1m below the bottom active soil layer. This parameterization, combined with a lack of lateral groundwater flow and lack of information on subsurface stratigraphy in CLM4.0 makes it very difficult to accurately calculate water table depth variations. It also forces the water table depth to remain mostly shallow throughout the region.

Increased resolution in CLM4.0 significantly improves the terrestrial water storage anomaly, ΔTWS. The ΔTWS value in the 1km-resolution simulation shows significant improvement when compared to GRACE satellite observation data for the Sacramento-San Joaquin Basin. This can be attributed to changes in runoff, snow and soil moisture at this scale.

Differences between the 100m and 1km resolution runoff, infiltration and drainage are mainly due to the change in the topographic surface data. However, the CLM4.0 simulation for the southwestern US at 100m resolution requires significant computational resources for the creation of surface data at 100m, and Terabytes of output data that require storage. Such limitations made this simulation a significant challenge. For these reasons, the 100m simulation was restricted to a small test domain within the southwestern US study region.

The topographic slope data in the 100m-resolution simulation has a significant impact on the division of surface water between runoff, infiltration, and drainage. These values were negligibly impacted by the increase in model resolution from

~100km to 1km, as the hill slopes that affect them are smoothened out, even at 1km resolution topography. Between the 1km and 100m resolution model outputs, the biases are significantly correlated with the topographic data resolution. This result validates the initial hypothesis that the hillslope scale heterogeneities affecting these variables can only be captured at spatial scales of 100m-resolution and finer. While the 1km resolution version of CLM4.0 works well for surface parameters, such as sensible heat and soil moisture, the resolution needs to be at 100m or finer to actually see the effect of slope on variables such as runoff, drainage, and infiltration. Ideally, 100m would be a more preferred scale to capture the hillslope scale hydrologic processes, though there remain limitations in obtaining input data at this resolution. Our results indicate that lateral flow between grid cells is an important missing process needed for simulations at 100m resolution or finer during the wettest portion of the year.

2.5 Conclusion

Many processes in the CLM4.0 are represented in an empirical fashion, and increasing model grid resolution, while proven here to be helpful, is ultimately only one important component of model improvement. More accuracy in model forcing at the required simulation resolutions is also a critical step toward producing accurate results. Also, active calibration of the model using reliable and consistent observations of model states and fluxes is important. The results we show help to demonstrate the critical scales for which important hydrological processes, such as snow water equivalent, soil moisture content, and runoff begin to more accurately capture the magnitude of the land water balance for the entire domain. This proves that grid resolution itself is also a critical component of accurate model simulations, and for hydrologic budget closure. Parameter calibration efforts may be fruitless (or at least less effective) if an appropriate model resolution is not achieved first.

Correct implementations of surface flow in hyper-resolution hydrologic models will also require much better representation of the subsurface, including lateral flow. The importance of subsurface and surface water dynamics for land surface and land atmosphere exchanges has been addressed by various studies [*Bierkens and Van den Hurk*, 2007]. These studies suggest that there exists a strong linkage between the mass, energy, and momentum balances of the subsurface and the land surface, which require integration of two different paradigms.

The results presented in this numerical experiment are encouraging, but also point out the limitations in improving an LSM by just increasing spatial resolution and surface datasets. As was shown with the water table depth analysis, there is a need to develop parameterizations at the required resolution and to improve the way variables, such as the water table depth, are calculated in the CLM4.0 physics. As we increase the model resolution, correct implementations of surface flow will also require much better representation of the subsurface soil texture and stratigraphy.

Chapter 3: Improving Land Surface Model predictions at high resolution via assimilation of groundwater observation data.

3.1 Introduction

Groundwater resource management is very important for a sustainable future. Over 2.5 billion people worldwide rely on groundwater as their primary source of drinking water and for crop irrigation. There are no comprehensive national or global groundwater level networks in existence with uniform coverage at high resolution of major aquifers, climate zones and land use types [*Hutson*, 2004; *Shah et al.*, 2001]. To better assess and manage groundwater supplies, there is a recognized need to improve monitoring of these resources, especially at regional scales through use of better models and assimilation techniques[*Effort et al.*, 2012; *NRC*, 2000] it also provides a key towards sustainable management of water resources and better understanding of the local impacts of climate and land use change.

Groundwater modeling is challenging since groundwater variations are not readily visible and it is difficult to measure spatially, with limited sets of observations available. Even though groundwater models can reproduce water table and head variations, partially known geologic structure, errors in the input forcing fields and imperfect LSM parameterizations can lead to considerable drift in modeled land surface states. These models frequently tend to have biased results that are very different from observations. As we have shown in the previous chapter that even increasing the model resolution to 1km or 100 meters is not of much help in terms of better groundwater modeling due to systematic errors in the parameterization of water table depth calculations, lack of subsurface stratiagraphy information and absence of lateral water flow between grid cells. While many hydrologic groups are working to develop better models that can solve some of these issues, there are methods to make the models more robust through data assimilation of observed groundwater data.

During the past decade a range of data assimilation techniques have been developed to optimally merge coarse-resolution LSM estimates with satellite observations to reduce modeling errors arising from various sources. At their core, these approaches provide a methodology for properly updating error-prone model predictions with incomplete and uncertain observations of model states [*Chen et al.*, 2011; *Moradkhani*, 2008; *Reichle et al.*, 2002]. With the availability of new and better observation datasets, such as the remote-sensed Gravity Recovery and Climate Experiment (GRACE) observations, new methodologies are being developed [*Zaitchik et al.*, 2008], but with partial success due to inherent model biases [*Chen et al.*, 2011]. There have been very few studies on the assimilation of groundwater measurements into LSMs, even though such models poorly simulate deep groundwater as we have shown in chapter 2. As deep unconfined aquifer dynamics

are now being simulated in LSM's, for example, the NCAR's Community Land Model version 4.0 (CLM4.0)[*Oleson et al.*, 2010b; *Zeng and Decker*, 2009], it provides us with an opportunity to improve the groundwater modeling using assimilation of well level observation data.

The aim of this study is to evaluate the improvements in hydrologic prediction (i.e. water table depth, surface soil moisture, root zone soil moisture, latent heat, ground evaporation, surface runoff, drainage and infiltration) of a high spatial resolution version of CLM4.0 within a 1° by 1° domain in the Sierra Nevada foothills in Northern California through assimilation of observed groundwater depth measurements from multiple wells in the region. The primary hypothesis is that the local water budget terms can be calculated with improved accuracy through the application of groundwater kriging based assimilation methods within a high spatial resolution model. In this study the NCAR Community Land Model version 4 (CLM4.0) is run at 0.01-degree (~1km) resolution over the test region. Two methods for data assimilation; Ensemble Kalman Filter (EnKF) based assimilation and direct insertion are used to improve model predictions and the outputs are compared with predictions from CLM4.0 without assimilation and observation data. The model setup for this study is similar to the one in chapter 2. The assimilation methodologies, datasets used and relevant model dynamics are explained in Section 3.2. Results are presented and evaluated in section 3.3 and these results and their consequences are discussed in Section 3.4 of this chapter.

3.2 Methods

The CLM4.0 is run offline without assimilation and with assimilation using the direct insertion method from 1 April 2003 to 30 March 2004 at a 30 arc-second (~1km) resolution over the test region.

3.2.1 Model description

The model used in this work for advancing hyper-resolution terrestrial simulations is CLM4.0, the land component of the NCAR Community Earth System Model. Fundamental to the CLM4.0 hydrology is the fractional saturated area (f_{sat}) and Topographic Index that have been discussed earlier in Chapter 1. The new version of CESM now supports multi instance runs and was used for the CLM+DART assimilation runs, the land surface part remains the same in both versions of CESM.

The saturated hydraulic conductivity, volumetric water content at saturation, the Clapp and Hornberger exponent and the saturated soil matric potential are determined using soil texture values [*Clapp and Hornberger*, 1978; *Cosby et al.*, 1984; *Niu et al.*, 2007; *Oleson et al.*, 2010b]. The high-resolution soil texture dataset for this purpose was produced using the CONUS-SOIL dataset at 30 arc second (~1km) resolution[*Miller and White*, 1998].

Determination of water table depth, ZΔ (m) in CLM4.0 is based on the work by [*Niu et al.*, 2007]. In their approach, a groundwater component is added in the form of an unconfined aquifer lying below the hydrologically active upper layers in the soil column (Figure 1.1). The solution for ZΔ is dependent on whether the water table is within or below the active soil column layers. The active and inactive water storage terms are used to account for these conditions. The first water storage term, W_a (mm), is the water stored in the unconfined aquifer and it varies with the change in water table depth when the water table is below the lower boundary of the hydrologically active soil column. The second water storage term, W_t (mm) *i*s the total groundwater, which includes water within the soil column and water in the unconfined aquifer. When the water table is below the soil column then $W_t = W_a$ and when the water table is within the soil column, W_a is constant (5000mm) because there is no water exchange between the soil column and the underlying aquifer, while W_t varies with soil moisture conditions. These two water stores are updated as the water table changes within the active soil column or in the inactive soil layers [*Oleson et al.*, 2010b]. The water table depth is calculated from the aquifer water storage scaled by the average specific yield (S_y), where S_y = 0.2 is the fraction of water volume that can be drained by gravity in an unconfined aquifer [*Niu et al.*, 2007]; [*Oleson et al.*, 2010b], with the assumptions that the initial amount of water in the aquifer is 4800 mm and the corresponding water table depth is one meter below the bottom of the active soil layer. For the case when the water table is within the soil column, there is no water exchange between the soil column and the underlying aquifer and the water table depth is calculated accordingly [*Oleson et al.*, 2010b]. There is an unconfined aquifer at the bottom of the soil column (Figure 1.1). The depth to the water table is ZΔ and changes in aquifer water content W_a and W_t are controlled by the balance between drainage from the aquifer water q_{drai} and the unconfined aquifer recharge rate $q_{recharge}$ ($kgm^{-2}s^{-1}$) (defined as positive from soil to aquifer) . Other aspects of the CLM4.0 affecting this study have been discussed in detail in chapter 1.

3.2.2 Study area:

The study area presented here is a 1^o x 1^o area in northern California that extends from the eastern part of the northern California Central Valley (CCV) to the Sierra Nevada foothills with coordinates -121.00E 39.00N x -122.00E 40.00N and is divided into 120x120 grid cells, each 30 arc-second (~1km), for grid cell based simulations using CLM4.0 (Figure 3.1). This study area contains the Sierra Foothills Research and Extension Center (SFREC), a UC field research station, a number of natural stream and managed irrigation source wetlands, and the Yuba River. It has strong topographic variation from east to west, where the western part of the study region lies in the CCV and is nearly flat with a mean elevation of approximately 20m-30m above mean sea level and the eastern part of the region is in the foothills and rises to 1000m in elevation.

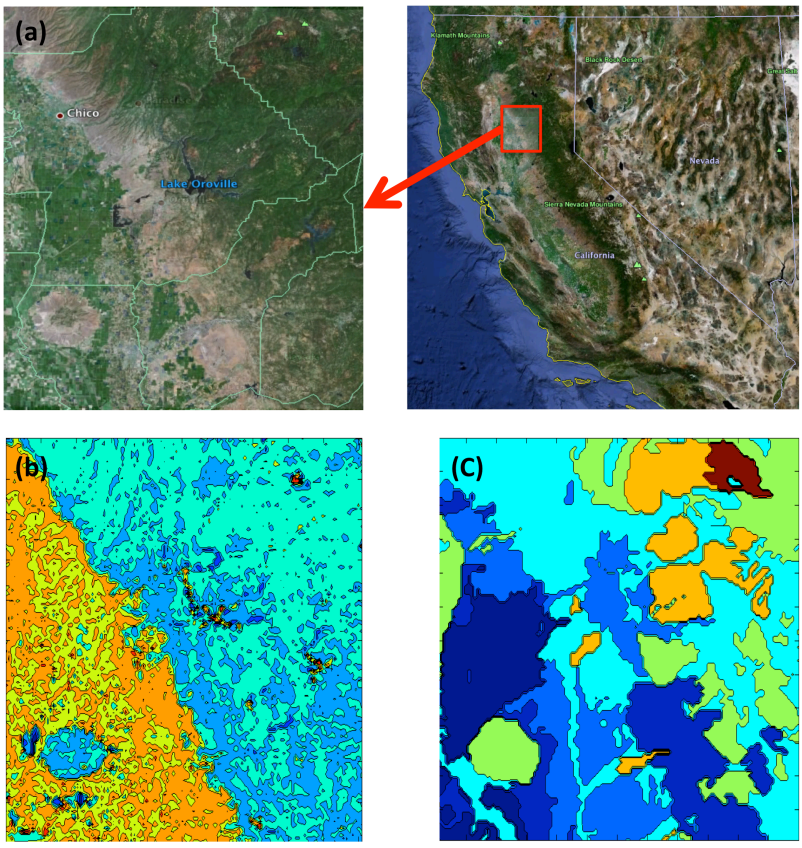

Figure 3.1: (a) The CLM4.0 model test domain, (b) Fraction of saturated area, F_{max}, and (c) Percentage Sand data at 30 arc-second (~1km) resolution.

3.2.3 Data

This study required near surface meteorological data and surface data to force the model and groundwater observation data used for a kriging-based interpolation leading to assimilation. The Digital Elevation Model (DEM) data for the topography was obtained from the 1/3 arc-second (~10m) resolution USGS dataset.

3.2.3.1 Model surface datasets

CLM4.0 was forced using 30 arc-second (~1km) resolution input data files.

The Topographic Index (TI) values were calculated from 1/3 arc-second (~10m) resolution DEM datasets provided by the USGS using the ArcGIS™ software at each pixel following the method described in [*Quinn et al.*, 1995]. The resulting TI at 1/3 arc-second resolution was then used to calculate f_{max} values at 30 arc-second resolution using the method described in [*Niu et al.*, 2005].

The high-resolution soil texture dataset was produced using the CONUS-SOIL dataset at a 30 arc-second (~1km) resolution. Other surface and aerosol input data

were provided by the NCAR 0.23 x 0.31 datasets and were held constant over the study area.

The initialization files were created after spin up of the model from bare soil at each resolution to reach thermal and hydrologic equilibrium [*Lo et al.*, 2008]; in this case the models were spun-up for 21 years from 1982 to March, 2003. A 40-member ensemble was created by a similar 21 year spin up, to initialize the multi instance CESM code used for assimilation using Ensemble Kalman Filter (EnKF).

3.2.3.2 Model forcing datasets

Offline CLM4.0 simulations were forced with the 1/8-degree (0.125° x .125°) resolution, 01 Jan 1979 to 31 Dec 2010 hourly atmospheric forcing data from the North American Land Data Assimilation System version 2 (NLDAS-2) atmospheric data, [*Mitchell et al.*, 2004b] rather than the usual T62 resolution NCAR provided forcing data [*Qian et al.*, 2006]. The NLDAS-2 domain, spatial resolution, computational grid, terrain height, and land mask in NLDAS-2 are identical to NLDAS-1 [*Mitchell et al.*, 2004a]. Earlier the atmospheric input forcing used was the global 1948-2004, 3-hour, T62 resolution (1.875°x1.914°) dataset derived by combining observation-based analyses of monthly precipitation and surface air temperature with intra-monthly variations from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) Reanalysis. We used an 80 ensemble member atmospheric forcing dataset at 2.5°x 1.875° resolution created by the NCAR-CAM (Community Atmospheric Model) runs with data assimilation and obtained from NCAR for the EnKF-based CLM4.0 assimilation run.

3.2.3.3 Groundwater data

Groundwater measurement data is collected from datasets provided by the California Department of Water Resources (DWR). The observation wells are located mainly in the Central Valley part of the study domain. Lack of well data in the mountains is explained by the assumption that the Sierra Nevada mountains have a limited capacity to store groundwater [*Famiglietti et al.*, 2011]. DWR has more than 300 well sites available but many of the well measurements are not continuous and for a particular month the number of sites with observation varies from 60 to 150. For every month there are a sufficient number of well observations to apply an ordinary kriging methodology to interpolate water table depth over the whole region, we also perform cross validation tests to make sure kriging is giving us consistent results. The observed data is used to perform kriging for the start of every month, groundwater is temporally and spatially slowly varying with time and this helps to justify this assumption [*V Kumar and Remadevi*, 2006]. A proper data quality check is maintained so that any biased data is not used in the kriging process. Thgis is done by first filtering data on the basis of DWR listed codes for different disruptions such as leakages or pumping and secondly by not using any data values

which significantly differs from other observations in the region. The DWR well data measurements are given in feet and were converted to meters for the assimilation, as all measurements in the model are in SI units.

Another set of continuous water table depth data is also retrieved from DWR that is only used for comparison and validation purposes, and not for the kriging. Continuous well data is only available at fifteen sites in the test region. These wells give good quality daily measurements but most sites have some kind of disruption for some times in the year.

3.2.4 Hydrologic data assimilation

Data assimilation is the method of making models utilize the information from observations of the system being modeled. Good assimilation makes the modeled state more consistent with the observations, particularly future observations. Effective data assimilation systems tend to make forecasts more accurate – within the ability of the model, naturally – and tend to make 'hindcasts' (the model state immediately after the observations have been assimilated) to more accurately reflect the state of the system.

The observed groundwater well data obtained for assimilation are too sparse so the data from DWR was used to create a kriging setup to interpolate groundwater depth in the test region. Kriging is then used to provide the observation data at the spatial and temporal resolution we need for the assimilation process. Kriging variance obtained is used to assign measurement errors in an Ensemble Adjusted Kalman Filter (EAKF)-based assimilation process. The kriged data is assimilated into the model at a monthly time step using two very distinct techniques, Ensemble Kalman Filter and Direction Insertion. These two methodologies, which are slightly different from those described in chapter1, are explained in detail below.

3.2.4.1 Kriging methodology

Kriging is a technique for generating optimal, unbiased estimates of regionalized variables at un-sampled locations using the structural properties of a semi-variogram and the initial set of data values. Ordinary Kriging is the most widely used kriging method and is also designated as the best-unbiased linear estimator. It estimates values at un-sampled locations between known data points using a linear estimation procedure, in a region for which the variogram is known. The estimation is unbiased in the linear sense and results in minimum error variance. Ordinary Kriging was used for the purpose of this study.

Getting quality observation data at sufficient locations was a challenge but it is extremely important for this study. The water table data from the DWR was checked for any anomalies or errors; error codes provided by DWR were used as a guide for this purpose. The first data of every month at each available well location is used for kriging at any given month. This provides a sufficient number of data locations for a

meaningful kriging, the slow spatially and temporally varying nature of water table depth helps in this assumption. The kriging methodology used here has been well described and documented in chapter 1 and only a cursory review of this specific approach is provided here, more details have been explained by [*Isaaks and Srivastava*, 1990; *Leuangthong*, 2008; *Wackernagel*, 1999].

Stationarity and de-trending of observation data is crucial to get good interpolation results through the kriging methodology. The groundwater level height from Mean Sea Level (MSL) at the corresponding observation location was calculated by subtracting the groundwater depth from the reference height. This step helps to remove anomalies that may affect the kriging process because of any sudden change in surface features. The ordinary kriging method works much better when the topographic changes are removed [*Chung and Rogers*, 2011]. A three dimensional plane is fitted to the dataset and the residuals are calculated at each site to remove any systematic trend in the data. These calculated residuals are normally distributed and stationary and thus suitable for kriging (Figure 3.2a).

The experimental semi-variogram is calculated from this residual data and a theoretical variogram was calculated for modeling purposes. The semi-variogram for every month was calculated. The lag distances were grouped into 20 bins with lag distance varying from 4000 feet to 8000 feet in various months of the year, the tolerance was fixed at half the lag distance. These lag distances and tolerances were used for calculation of the semi-variogram (Figure 3.2b). Determining the correct theoretical variogram model is important and Spherical, Gaussian and exponential theoretical variogram models were all fitted to the experimental variogram. For each theoretical variogram I calculated the standard error and the Gaussian model gave the minimum standard error for most months and thus was selected for kriging purpose.

The residuals calculated were then kriged to get values at the nodes of a 120x120 grid that exactly correspond to the CLM4.0 output resolution grid. Ordinary kriging was used to calculate the interpolated value at every location as explained in section 1.4.2. The kriged output values of the residuals thus obtained are added to the fitted plane values at each node to get groundwater level values (Figure 3.2c). The ground water levels are subtracted from the topographic height at the location obtained from USGS DEM dataset to calculate the groundwater depth at the respective nodes.

Groundwater depth is calculated at every grid cell location in the test region; in this case in a 120x120 grid matrix. We know by previous studies that groundwater in the mountainous region is very deep and cannot be modeled by CLM4.0 hydrology also there are no observational groundwater stations in the mountains thus the water table in mountainous regions is capped at 50 meters depth, as we must specify water table depth values for all points Figure 3.2(d).

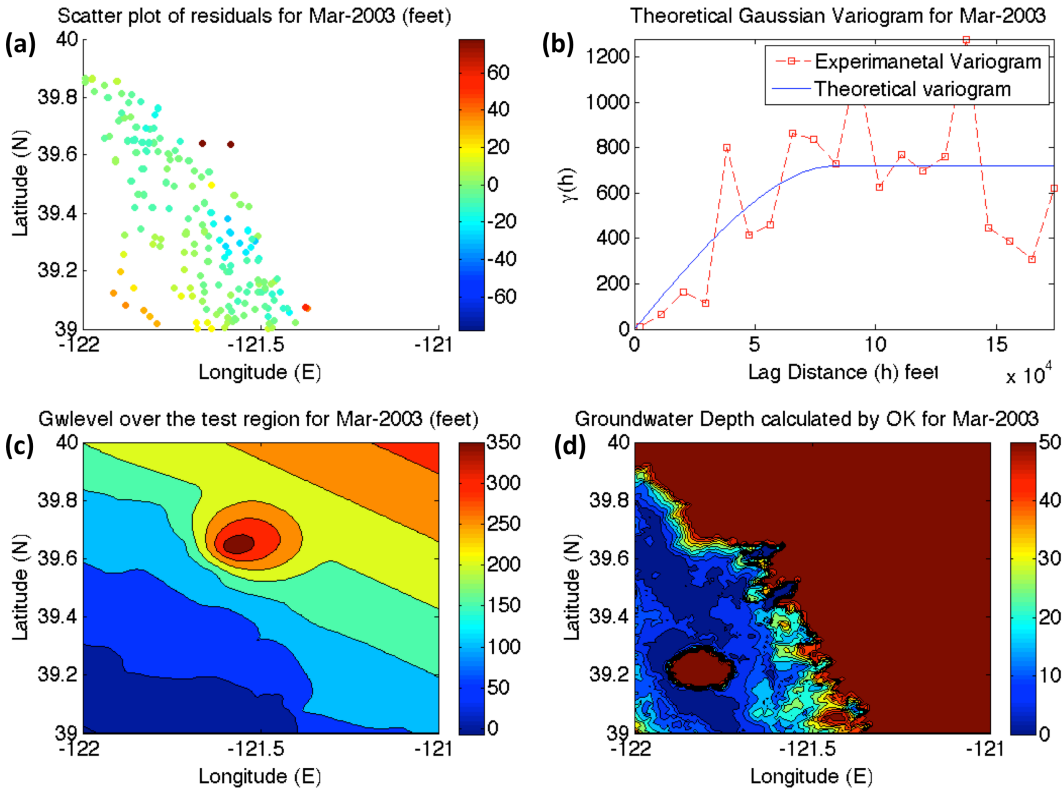

Figure 3.2: Kriging plots for March-2003 (a) Residuals at groundwater measurement well locations. (b) Experimental and theoretically fitted spherical variogram for the corresponding data. (c) Groundwater level from mean sea level calculated by ordinary kriging (OK). (d) Water table depth as calculated by the kriging method and DEM dataset, the maximum water table depth is assumed to be 50 meters.

3.2.4.2 DART- Data Assimilation Research Testbed

The Data Assimilation Research Testbed (DART) is a community facility for ensemble data assimilation. It has been developed and is maintained by the Data Assimilation Research Section (DAReS) at the National Center for Atmospheric Research (NCAR). In our study a 40-member CLM4.0, the land surface part of CESM multi instance version of CESM, is used for data assimilation. DART+CLM is set up in such a way that no changes are necessary in the CLM4.0 code itself. CLM4.0 stops and writes restart and history files at the end of the run, these files are used to extract the DART state vectors. Using the kriged observation data provided, increments are calculated and applied to the DART state vector. CLM4.0 restart files are updated with the new adjusted DART state vectors and these restart files are used by the post run script to start the new runs. The initial 40-member CLM4.0 ensemble of initialization data is created after a 21-year spin up run. The data is

forced using an 80 member atmospheric forcing dataset created from CAM runs and data assimilation by NCAR.

3.2.4.3 Assimilation methodology

Assimilation is carried out using two separate methodologies. First is the direct insertion of kriged groundwater data into restart files at the start of each month. The second method is an ensemble based methodology devised by DART that uses an Ensemble Adjusted Kalman Filter (EAKF) for data assimilation [*Anderson*, 2001].

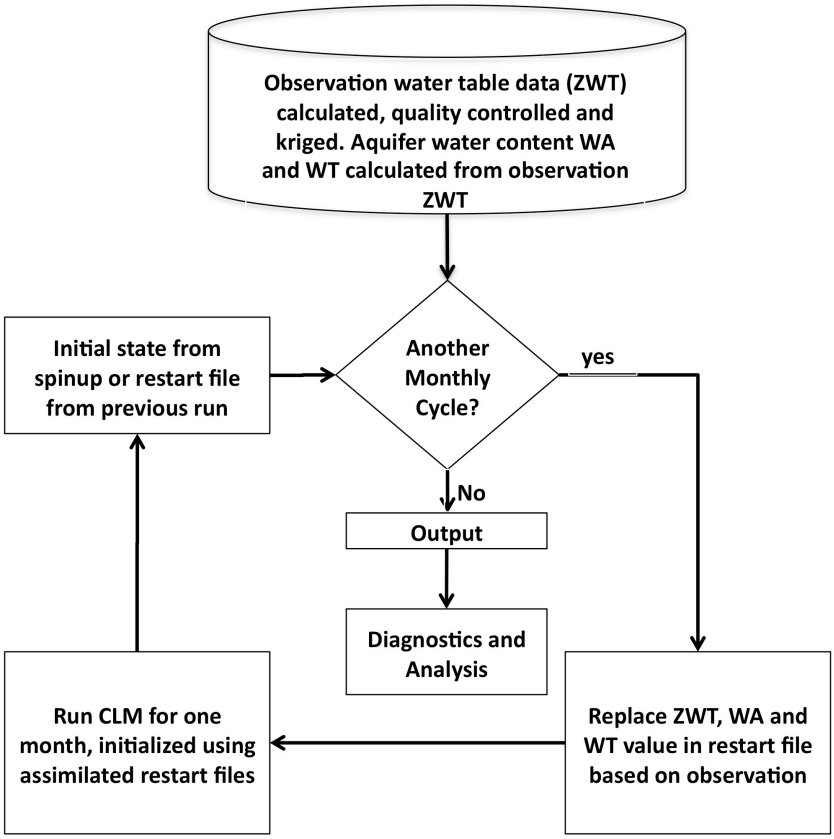

Figure 3.3: Schematic Diagram showing the direct insertion based assimilation methodology.

The processed groundwater data from DWR is kriged to calculate water table depth (ZWT_r) at each grid location by the method as described in the section 2.4.1.

Direct insertion is one of the earliest and most simplistic approaches to data assimilation in which the forecast model states are directly replaced with the observations (Figure 3.3). This approach makes the explicit assumption that the model is wrong and that the observations are right, which both disregards important information provided by the model and preserves observation errors. A key disadvantage of this approach is that model physics are solely relied upon to

propagate the information to unobserved parts of the system [*Houser et al.*, 1998; *Walker and Houser*, 2001; *Walker and Houser*, 2005]

For the direct insertion methodology, if the calculated water table depth (ZWT_r) is in the hydrologically active or inactive soil regions (Figure 1.1), then aquifer water content variables W_a_r and W_t_r are calculated for each grid location [*Oleson et al.*, 2010b]. The assimilation methodology used here is based on the principle of minimum change in the native CLM4.0 code, so we assimilate the kriged data by replacing the variables only in the restart files. The initial restart file is created after the spin-up run and subsequent restart files are created at the end of every month. The variables water table depth (ZWT), aquifer water content (W_a) and total groundwater content (W_t) in the restart file are replaced by the kriged/calculated variables; ZWT_r, W_a_r and W_t_r. The assimilated restart file is then used as an initialization file for the subsequent offline CLM4.0 simulation.

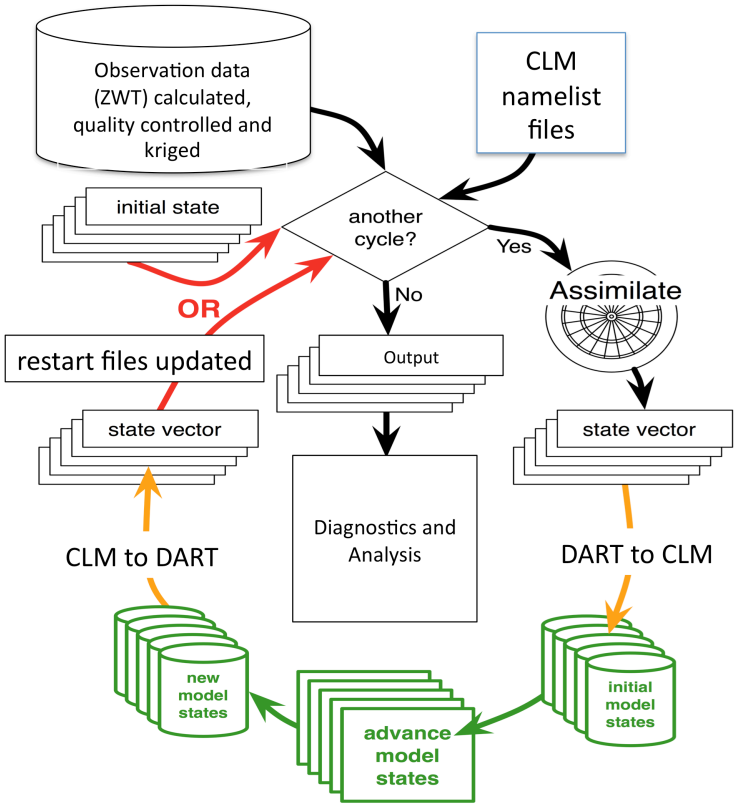

Figure 3.4: Schematic Diagram showing the DART+CLM assimilation methodology setup.

In DART+CLM assimilation methodology (Figure 3.4; modified from figure in chapter 1) observation sequence files are created using the kriged water table depth at the first day of every month. The kriging variance is calculated at each location and used to calculate the measurement error at each grid point. DART+CLM runs for

a month then assimilates the observation data provided for the month using the DART algorithm. The restart files are automatically updated and used to initialize the run for next month. The assimilation uses an Ensemble Kalman Filter to update the restart files with the new observation values. The exact methodology is shown in Figure 3.4 and more information about DART has been provided in chapter 1.

The spatial resolution of the observation groundwater data is not enough to give us a meaningful assimilation result at high resolution; therefore ordinary kriging is used to interpolate water table depth values at the start of each month at every grid point location. Kriging has been previously used successfully to predict water table depth at a coarser spatial and temporal scales [*Godin*, 2012; *V Kumar and Remadevi*, 2006]. Here kriging is used to provide the data at the spatial and temporal resolution we need for the assimilation, but kriging alone cannot give us outputs at the temporal scale we require, nor does it let us compare the effect of assimilation on other model parameters. Various assimilation methodologies can be used after kriging, such as the Ensemble Kalman smoother [*Zaitchik et al.*, 2008], Ensemble Kalman filter [*Reichle et al.*, 2009], or unscented Kalman filter [*Tian*, 2008]. Here DART uses the Ensemble Adjusted Kalman Filter (EAKF) which has been shown to give better results than the traditional Ensemble Kalman Filter [*Anderson*, 2001].

3.3 Results

The CLM4.0 is run offline without assimilation, with assimilation using direct insertion methodology and with assimilation using DART based EAKF from 1 April 2003 to 30 March 2004 at a 30 arc-second (~1km) resolution. The outputs are archived at a daily time step. The water table depth outputs are plotted and compared with each other and with observation data for all the runs. We performed evaluations to determine the best method for groundwater assimilation and also analyzed the effect of assimilation on other water and energy budget terms in the model, including runoff, soil moisture, ground evaporation and sensible heat.

The equations used to calculate the correlation coefficient (r); RMSE, skill and bias are described as:

$$r = \frac{COV(Y_o, Y_e)}{\sigma_X \sigma_Y} \tag{3.1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{1}^{n}(Y_o - Y_e)^2} \tag{3.2}$$

$$\text{Skill} = 1 - \frac{RMSE_a}{RMSE_0} \tag{3.3}$$

$$\text{Bias} = \frac{1}{n}\sum_{1}^{n}(Y_e - Y_o) \tag{3.4}$$

where Y_e is the model estimate, Y_o is the observation and n is the number of observations. $RMSE_a$ and $RMSE_o$ are root-mean square errors for the model with assimilation and model without assimilation, respectively

3.3.1 Assimilation using EAKF based DART

The Ensemble Kalman Filter based assimilation method fails to give good results. It ignores most of the observation data, as there is a huge difference between the ensemble of model calculated water table depth and the observed water table depth. The ensemble spread also remains very small and significantly different from the observation values. We tried to increase initial ensemble spread by using a longer spin-up time, but the spread stabilizes after 15 years and does not change much after that. This difference between model calculated water table depth and observation values can be attributed to the fact that the CLM4.0 groundwater model assumes a uniform 3.8m depth aquifer throughout the whole region and calculates water table depth with this assumption. The water table depth calculated has a bias towards being shallow throughout the whole region year round. The observations on the other hand are spread across the region where the water table not only significantly varies through out the region, but also during different seasons. This causes the posterior to not vary much from the prior and the ensemble spread remains very small throughout the run. We did, however, see a noticeable increase in the spread during wet winter months, as the water table rises and the model state agrees more with the observation values. This can be seen in Figure 3.5, where the spread of the prior and the posterior at the start of each assimilation cycle are shown.

Figure 3.6 shows the innovation (Posterior – Prior values) plots at the start of each assimilation cycle. It clearly shows that assimilation does not have much impact on the posterior except for some places in the wet winter months when the water table depth rises and the model state and the observation values are more in agreement.

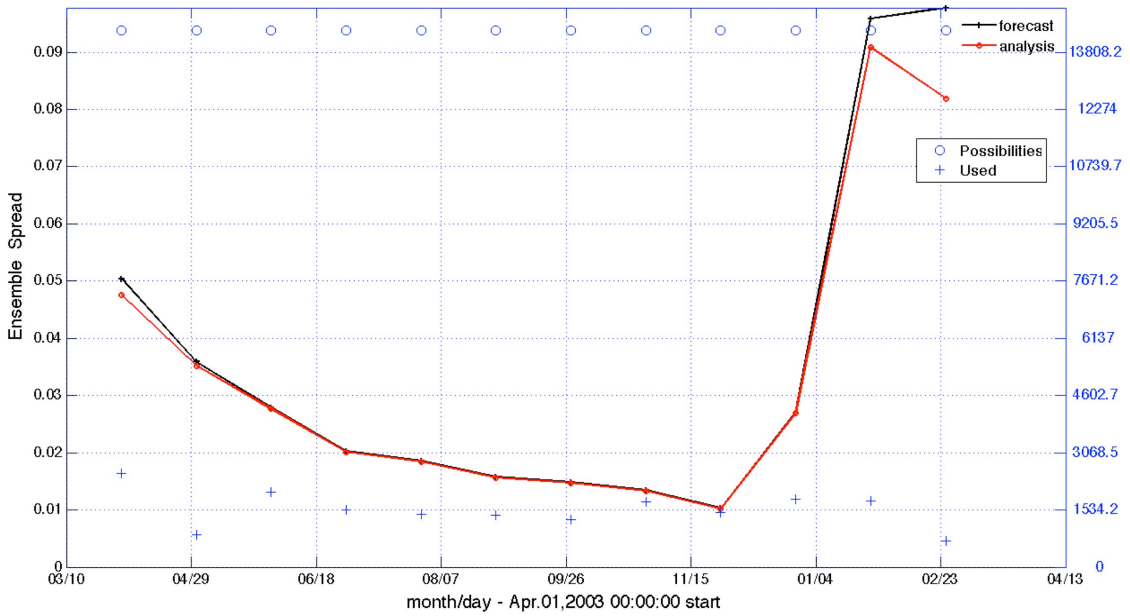

Figure 3.5: Ensemble spread of forecast (prior) and analysis (posterior) ensembles.

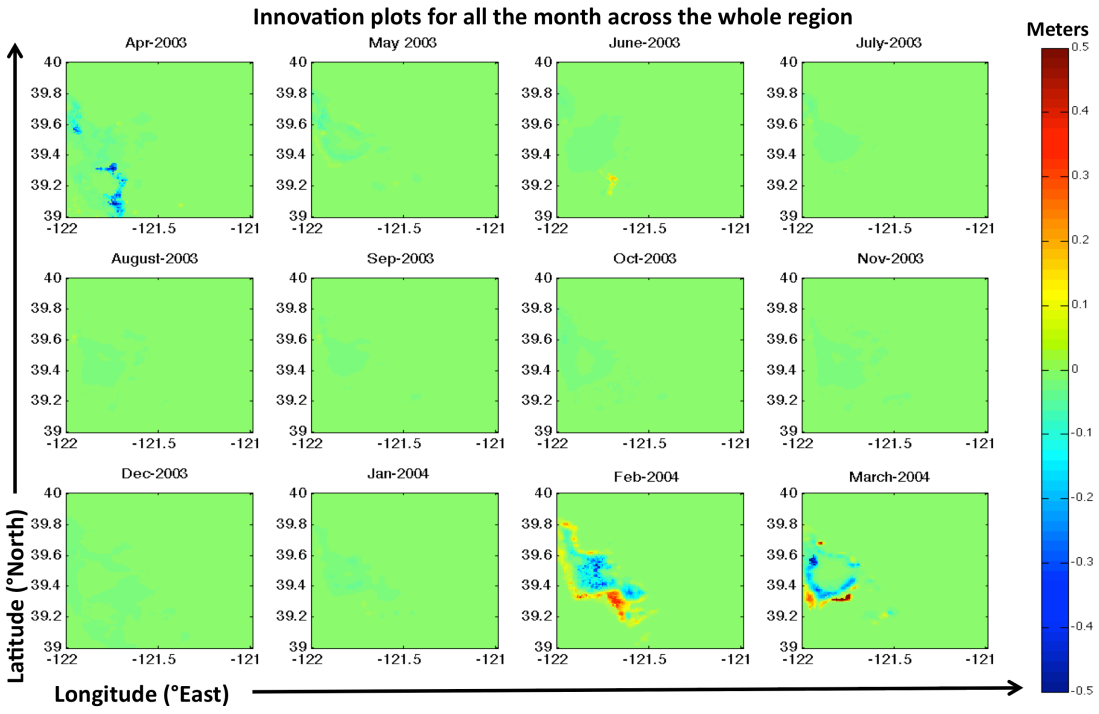

Figure 3.6: Spatial plots for the difference between Posterior and Prior at the start of each months assimilation cycle

Figure 3.7 shows the comparison of model outputs to observations at three different well locations with different water table (WT) profiles, deep (WT>10m), medium (3m<WT<10m) and shallow (WT<5m). Model outputs from the two different assimilation runs and a control run without any assimilation is compared to observation data from the continuous observations datasets obtained at these sites. The plots clearly show that the assimilation using EAKF is not having much impact on the output water table depth values and the output plot in those cases clearly looks very similar to the output from the run without assimilation. It also shows that assimilation using EAKF does not give us good output when compared with the observed water table depths at these sites. For other hydrologic fluxes too here is not much impact to be seen due to the assimilation process.

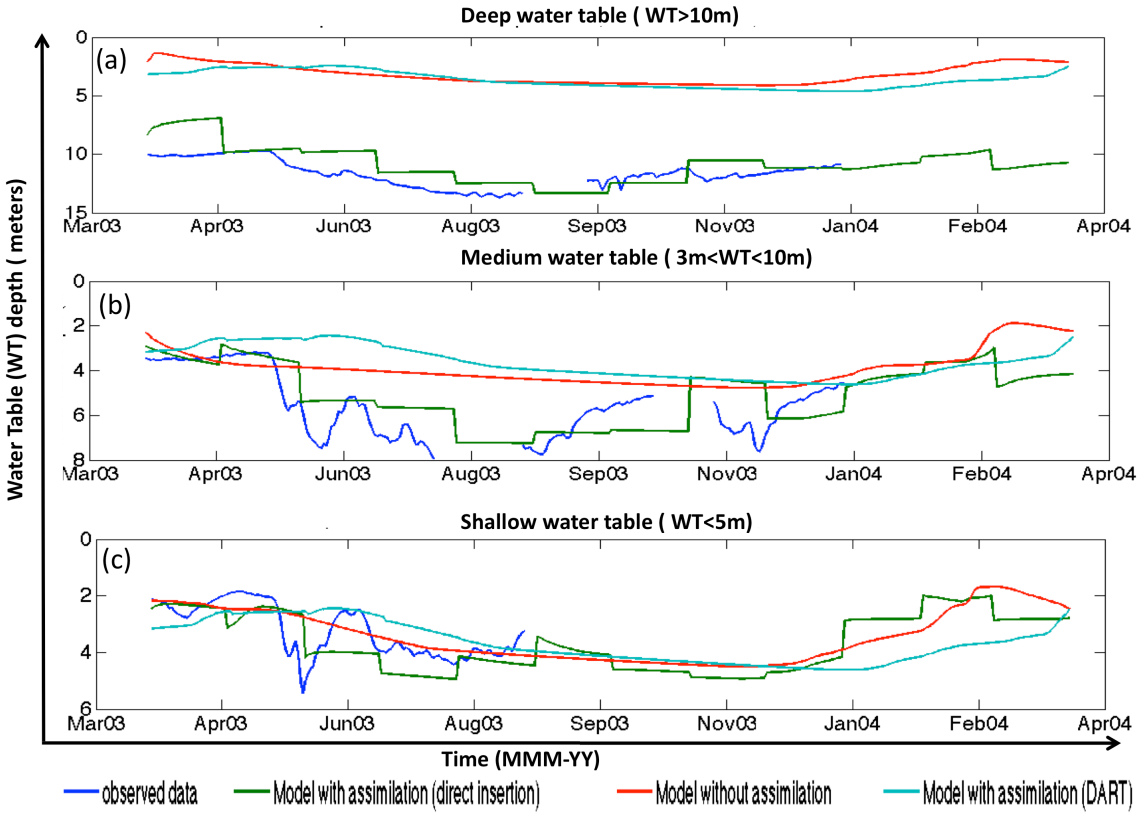

Figure 3.7: Comparison of water table depth for three well sites. (a) Deep water table profile (WT >10m) (39.5792N-121.697E) (b) Medium water table profile (3m<WT<10m) (39.5446N-121.6857E) (b) Shallow water table profile (WT <5m) (39.583 N-121.754E)

3.3.2 Assimilation using direct insertion method

The direct insertion assimilation method gives us much better results. The water table depth values at the start of every month are updated to be that of the values

obtained after ordinary kriging. The daily outputs then are compared with the observation values to check the effectiveness of the assimilation method.

3.3.2.1 Water Table depth

The SFREC test region has distinct topography that is reflected in the output water table depth. The upper right portion of the region is the Sierra Nevada Mountains, which is assumed to have little or very deep groundwater and is reflected by a water table depth in the region of more than 25 meters (figure 3.8). We can also see the Sutter Buttes at 39.2N121.8W, which is a volcanic rock formation abruptly rising to more than 600 meters, and is also called the "world's smallest mountain range".

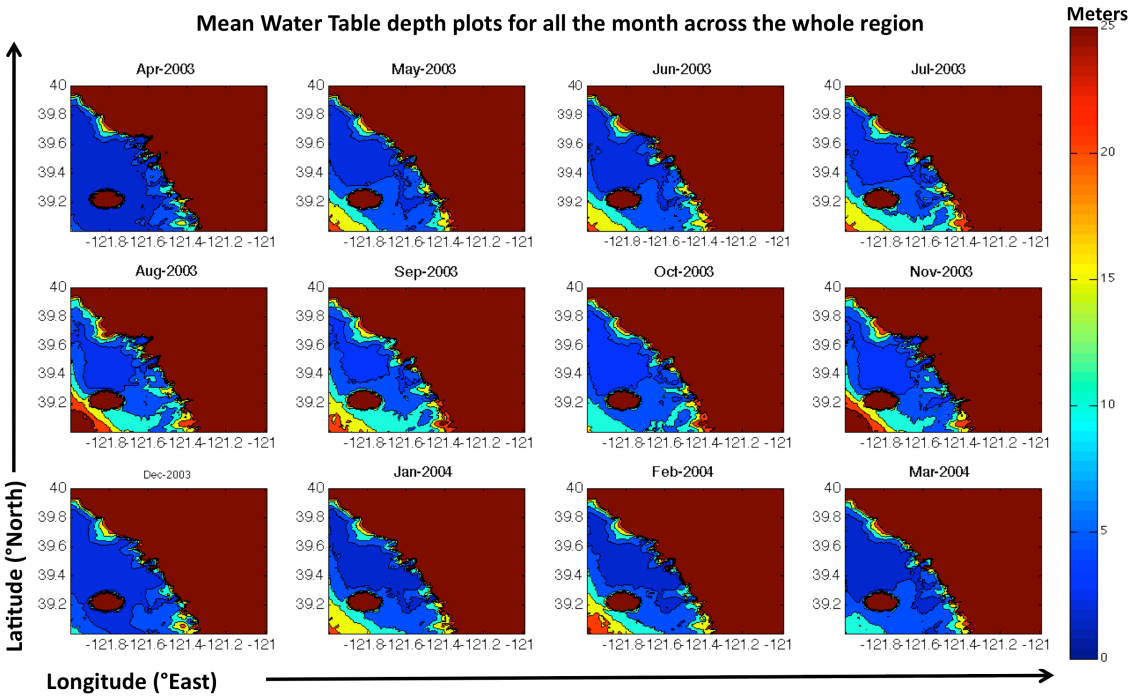

Figure 3.8: Mean monthly plots of water table depth below the surface in meters across the SFREC test region for the period April 2003 through March 2004. Sutter Buttes is seen as the quasi-circular shape in red (>25m depth).

The plots in Figure 3.8 reflect the seasonal variations of groundwater as we move from the rainy winter season to the dry summer season. As expected the lowest water table depths in this domain are in the foothill where the water table meets the surface at many places and where wetlands are formed. There are also significant numbers of observed wetlands in this region. The lower left corner of the plots show an anomalously large water table depth, which is an error due to lack of data in that region during a few months (especially August 2003 and November 2003). Overall the plots look good and further analysis and comparisons with observed values in

later sections prove that the assimilation is quite successful in giving us a good approximation of water table depth in CLM4.0 throughout the region across all seasons.

Figure 3.9 shows a time series of mean-area water table comparison for CLM4.0 with assimilation, CLM4.0 without assimilation, and observation data from 1 April 2003 to 30 March 2004. The mean-area water table observations are based on all data available for each day and the corresponding values in the models at respective measurement sites. The assimilated model follows the observation data much more closely while the model without assimilation almost always remains at a shallow level. The linear Pearson correlation coefficient of the observation data is 0.81 with the assimilated model and 0.1065 with the non-assimilated model. The RMSE of the observation data with the model with assimilation is 3.1613 while the RMSE for the model without assimilation is 8.2755 (Table 3.1). The results clearly show the vast difference between the observation and model values with assimilation and without assimilation, especially at locations where the water table is deep.

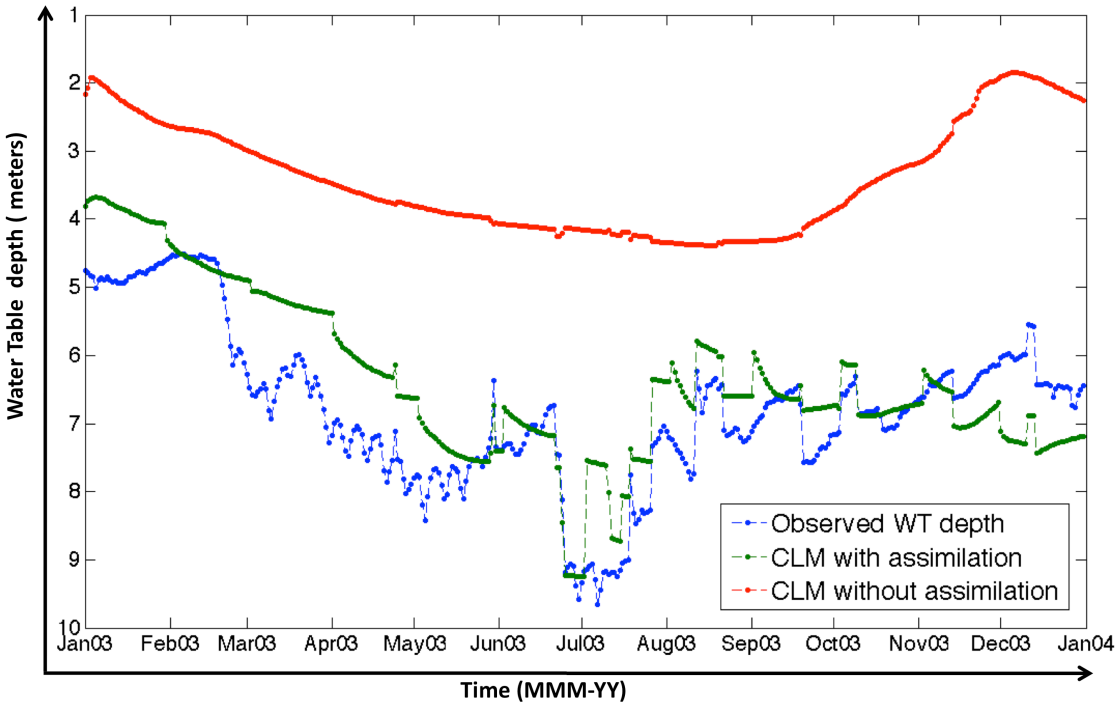

Figure 3.9: Comparison of model mean-area Water Table depth with observation well data from DWR. Mean of all available observation well data and corresponding grid cells in models for a particular day is taken for comparison.

Figure 3.10 shows plots comparing individual continuous well data with model water table depth (WT) data at three locations (a) Deep water table profile (WT>10m) 39.579 N-121.697 E, (b) Medium water table profile (3m<WT<10m) 39.544 N-121.687 E, and (c) Shallow water table profile (WT <5m) 39.583 N-

121.754 E, with the corresponding grid cell for CLM4.0 with and without assimilation from 1st April 2003 to 30th March 2004. The well site locations are all near the center of the test region and these observations are not used in the kriging interpolation scheme. CLM4.0 with assimilation shows an excellent correlation with the observation data and captures the monthly variations much better than CLM4.0 without assimilation. The water table depth in CLM4.0 without assimilation stays quite constant at a shallow water table depth. This comparison has the limitation that the model results each represent a 1km^2 grid cell, while the observation well represents a much smaller spatial footprint. The observation dataset is sometimes intermittently unavailable as in Figure 3.10(c), where it is not available after August 2003. The assimilated model also captures the annual trend very well at all three sites.

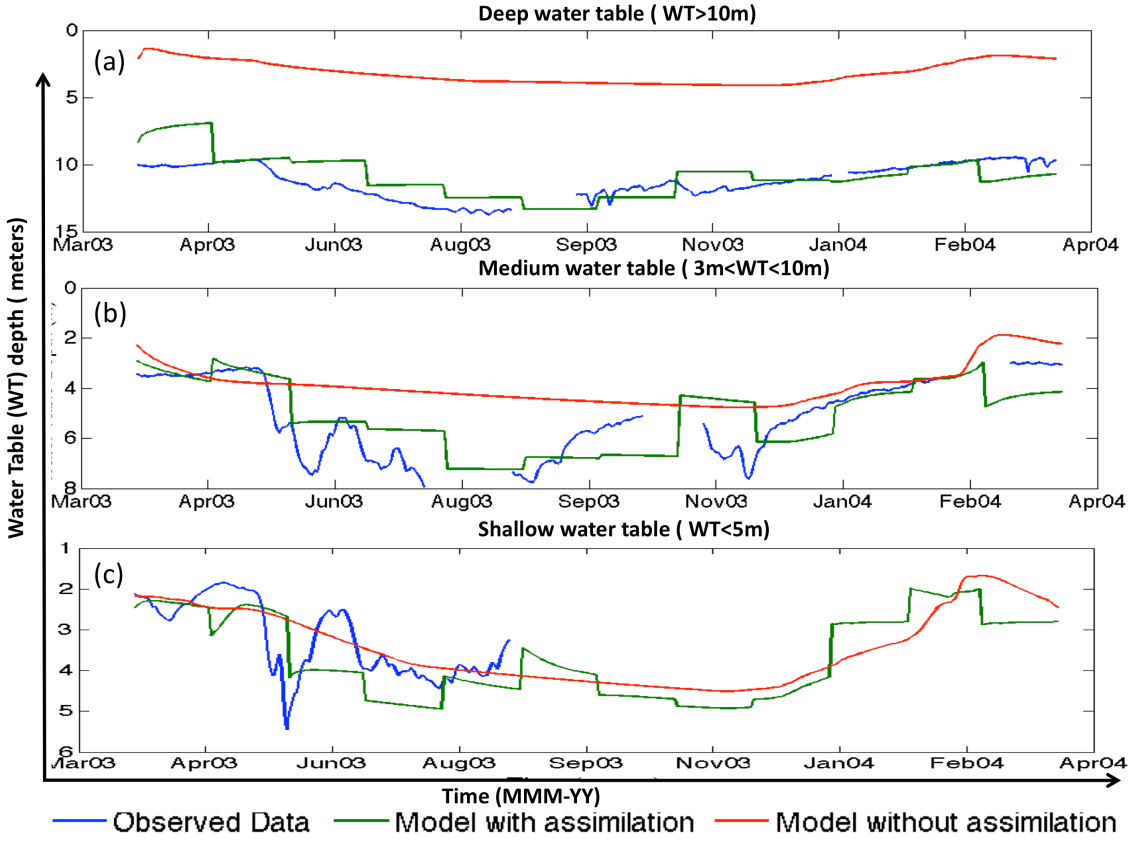

Figure 3.10: Comparison of water table depth for three well sites with continuous data. . (a) Deep water table profile (WT >10m) (39.5792 N-121.697E) (b) Medium water table profile (3m<WT<10m) (39.5446 N-121.687 E) (c) Shallow water table profile (WT<5m) (39.583 N-121.754 E)

Table 3.1: Evaluation of water table depth estimates from model with and without assimilation:

Plots	Obs-CLM4 (assimilated)	Obs-CLM4	
	RMSE	RMSE	Skill
mean-area (Figure 3.8)	0.8304	3.4641	0.76
well-to-grid at *39.544 N -121.687 E*	1.0886	1.6427	0.34
well-to-grid at *39.579 N -121.697 E*	1.3550	8.1450	0.83
well-to-grid at *39.583 N -121.754 E*	0.8035	0.6472	-0.24

Table 3.2: Evaluation of soil moisture, sensible heat and ground evaporation estimates from model with and without assimilation:

Variables	Obs-CLM4 (assimilated)	Obs-CLM4	
	RMSE	RMSE	Skill
Soil Moisture; 0-10cm	0.0237	0.0251	0.06
Soil Moisture; 0-100cm	0.0356	0.0424	0.16
Soil Moisture; 0-200cm	0.0307	0.0391	0.22
Sensible Heat; W/m²	13.1613	17.4382	0.25
Ground Evaporation; W/m²	3.2348	3.7226	0.13

When the water table depth is shallow; i.e. between 2 - 5 m (Figure 3.10c), then water table depth calculated through CLM4.0 without assimilation gives good results with high correlation and low RMSE with the observation data. Sites where the water table depth is below 5 meters the water table depth calculated by CLM4.0 without assimilation misses the trend as can see from the plot in Figure 3.10b, and the RMSE values are much higher. Table 3.1 shows the RMSE values at these three different locations and also over the entire domain, highlighting the improvements in results from CLM4.0 with assimilation compared to normal offline CLM4.0 runs.

3.3.2.2 Soil moisture

In addition to the evaluation of groundwater level variation with and without assimilation, we provide an analysis of other terrestrial water and energy budget fields such as soil moisture content, sensible heat, ground evaporation, runoff and infiltration. The effect of assimilation of water table depth values on other model variables and fluxes needs to be assessed to fully understand the impact of direct insertion based assimilation technique on CLM4.0 simulation outputs.

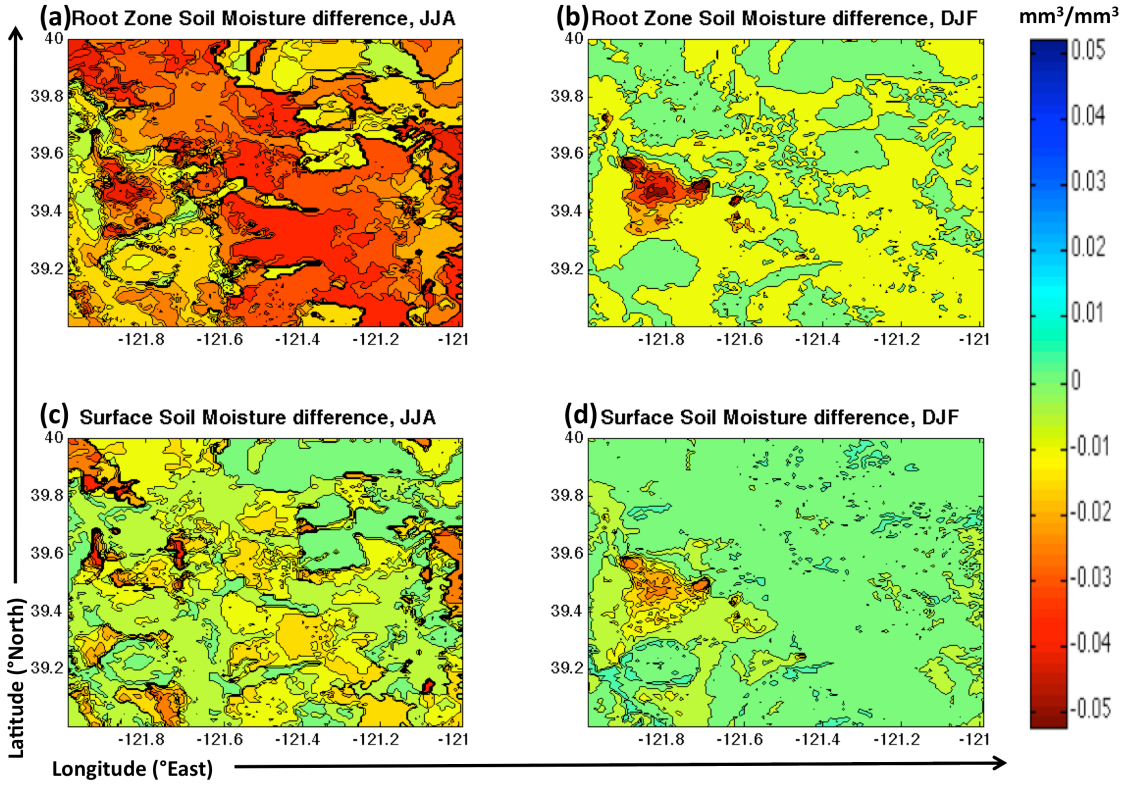

Figure 3.11: Difference in soil moisture content (mm^3/mm^3) between model with assimilation and model without assimilation for (a) the root zone top 8 layers of soil (0-100 cm) during summer (b) during winter and for surface soil moisture, top layer (0-1.7cm) (c) during summer (d) during winter.

Differences in soil moisture content across the whole region is calculated and compared for the root zone soil moisture, top 8 layers in CLM4.0 (0-100 cm), top 200 cm of soil layer and for the surface layer (0-1.7cm). Table 3.2 shows the RMSE values between the two models for these three regions, indicating a significant reduction in RMSE of CLM4.0 with assimilation compared to CLM4.0 without assimilation for all three soil regions. There is also a detectable effect of the assimilation on soil moisture across the whole region as seen in Figure 3.11. During the June-August (JJA) summer months, soil moisture content in CLM4.0 is reduced due to assimilation for both the surface and the root zone layers throughout the region (Figure 3.11 a,c). This effect can be attributed to the fact that CLM4.0 without assimilation has a very uniform shallow water table across the region, but in CLM4.0 with assimilation the water table depth becomes more variable and is in general deeper during summer months. This affects the root zone moisture and surface moisture [*Lo and Famiglietti*, 2010]. The effect in winter months is much weaker, since the water table is shallower during the wet winter season (Figure 3.11 b, d). CLM4.0 with assimilation and without assimilation have similar water table depth profiles in this study domain during wet seasons, thus the soil moisture values are also similar and the difference is quite small. The effect of groundwater on soil

moisture in a particular soil layer decreases, as the distance between that layer and water table increases, therefore the difference in the root zone soil moisture is greater than the surface soil moisture.

Surface soil moisture and root zone soil moisture are very important parameters in LSM's and affect the surface energy budget terms, evapotranspiration, boundary layer development, plant water usage and recycled precipitation. Even though the maximum value of soil moisture difference is in the range of around 5% in summers it can have significant impacts, as similar small changes in surface soil moisture have been known to affect these variables quite significantly [*Lo and Famiglietti*, 2011].

3.3.2.3 Sensible heat and Ground evaporation

The effect of groundwater assimilation on sensible heat and ground evaporation are compared for dry summer (JJA) and wet December-February (DJF) winter months in Figure 3.12. Table 3.2 calculates the RMSE values between the two models over the region, which shows significant reduction in RMSE in CLM4.0 with assimilation compared to CLM4.0 without assimilation. The differences between the sensible heat values in CLM4.0 with and without the assimilation, especially during the summer months can be explained by the fact that the water table depth in the assimilated model is lower in summer months, which reduces the surface soil moisture content during these months, causing the sensible heat to increase and ground evaporation to decrease across the whole region (Figure 3.12 a, b). The effect is negligible during winter months, as the water table depth is shallow and surface soil moisture difference is quite small this time of the year between the two models (Figure 3.11). Similarly ground evaporation decreases in summer due to less water in the soil and does not show any significant change in winters when the soil water content does not significantly differ between the two models (Figure 3.12 c, d).

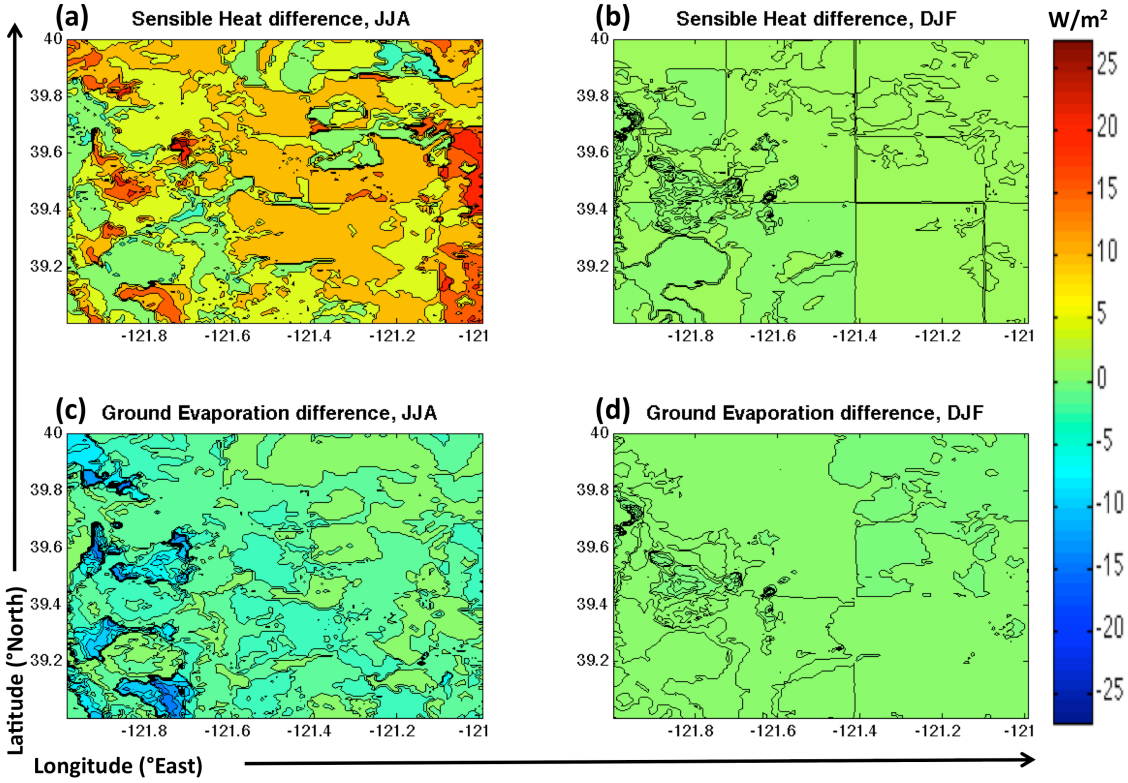

Figure 3.12: Difference in Sensible heat (W/m^2) (a) During summer and (b) During winter and Ground evaporation (W/m^2) (c) During summer, and (d) During winter across the SFREC test region.

The effect on soil moisture content, sensible heat, and ground evaporation can be explained as a result of a more accurate water table depth in the assimilated model and indicates better model results for these variables as well. The changes on these variables may seem small at around 5% for soil water content and 25 W/m^2 in sensible heat values, they can have big impacts on the boundary layer development and other variables within a coupled model [*Lo and Famiglietti*, 2011].

3.3.2.4 Surface Runoff and Infiltration

The groundwater assimilation also highlights CLM4.0 parameterization problems associated with surface runoff and infiltration in regions with deep or no groundwater (Figure 3.13). The problem highlighted is that of surface water division between runoff and infiltration. A key concept underlying this approach in CLM4.0 is the fractional saturated/impermeable area f_{sat}, which is determined by the topography, water table depth, and soil moisture content of a grid cell [*Niu et al.*, 2005; *Oleson et al.*, 2010b; *Oleson et al.*, 2008]. Grid cells where the water table depth is very deep or with no groundwater, such as the Sierra Mountains within the

test region, CLM4.0 computes very low f_{sat} values and reduces the runoff to unrealistically low values. This anomaly only affects the division between runoff and infiltration and the sum of both remains constant in both the assimilated and non-assimilated versions of CLM4.0. It is our view that this is a model-specific problem in CLM4.0 for areas with a very deep water table. The lack of a routing scheme in this offline version additionally limits these values and our ability to compare the outputs with observation data.

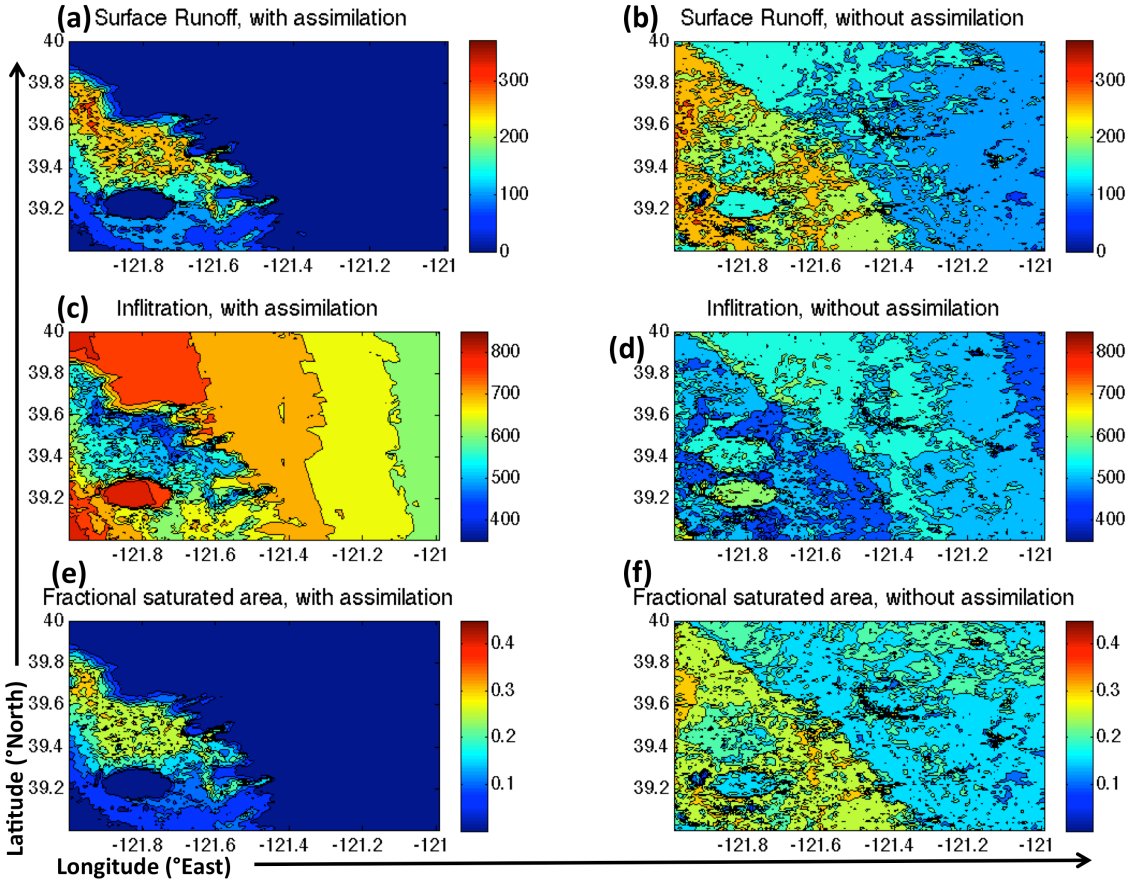

Figure 3.13: CLM4-simulated surface runoff (mm/yr) with and without assimilation (a, b); Infiltration (mm/yr) with and without assimilation (c, d); Fractional Saturated area with and without assimilation (e, f)

3.4 Discussion

Current LSMs lack appropriate groundwater dynamics for high-resolution (1km and higher) simulations. Here we have shown that groundwater dynamics, process descriptions, and parameterizations within CLM4.0 can be improved through the assimilation of groundwater observations. CLM4.0 groundwater dynamics improves

remarkably with such assimilation in the study domain that includes the eastern part of the northern California Central Valley and the Sierra Nevada foothills region.

In this study we assimilate sparse groundwater measurement data into CLM4.0 using a kriging-based interpolation technique. Kriging is necessary to generate assimilation data at higher spatial density so as to have a significant impact on the model outputs. Two different assimilation techniques are used, direct insertion and an Ensemble Kalman Filter (EnKF) based method. Though the EnKF has been widely used in the past for assimilation purposes, in this case it fails to give better results as this method relies on the prior model state to calculate the posterior values. In this experiment the CLM4.0 model states are significantly different from the observation due to inherent biases and parameterization in CLM4.0 and also the ensemble spread of the model remains small, thus rejecting most of the observation values that are significantly different from the model values.

The direct insertion method performs much better and the results indicate increased accuracy in simulated groundwater variations across the study region throughout the year. The assimilation of groundwater data in CLM4.0 does not lead to any major breakdown of model results or model parameterization. Assimilation leads to some minor changes in variables such as the soil moisture, ground evaporation and sensible heat values that can be explained to be due to improved groundwater modeling and its impact on these variables. The direct insertion assimilation technique gives good results, however at the start of each monthly update there is a discontinuous jump between the kriged and assimilated value at each location. This jump is due to the fact that we are trying to assimilate the kriged water table depth in the model at a monthly time step and we don't want to use any statistical smoothing technique at this stage, as it will interfere with our study to isolate the effect of this assimilation technique.

CLM4.0 without assimilation performs quite well in regions with shallow water table depth or in seasons when the water table is higher (late winter, spring in this region). This may be because CLM4.0 solves Richard's equation only in the top 10 soil layers, that is, up to 3.8 m depth, while the lower 5 layers just act as thermal slabs. The other reason may be due to the water table depth calculation methodology in CLM4.0 and its parameterization, which is suitable for large-scale (~100km), but not hill-slope or even 1km resolution. The improvements in water table depth modeling by assimilation and these results highlight scale-dependent breakdowns within the CLM4.0 hydrology scheme that needs to be changed for improved high-resolution modeling.

The results discussed above indicate that this assimilation technique can be used in regional-scale high-resolution groundwater modeling for areas with very sparse monitoring stations. This can help improve our groundwater monitoring abilities and studies of the impact of human induced or natural changes on groundwater availability. However, at present, this methodology is somewhat limited due to biases and parameterizations within LSMs, including CLM4.0. This has been

demonstrated here through the calculation of surface runoff, infiltration and fractional saturated area values. Even though calculation of the fractional saturated area is improved with the assimilation scheme, its impact on surface runoff remains incorrect, especially in mountainous regions with very deep groundwater. This parameterization, to calculate surface runoff using groundwater depth values, may well be correct for global climate model scale applications, but will require a more comprehensive scale-dependent type of parameterization for regional applications at higher resolution. These biases also restrict the use of the more accepted Ensemble Kalman Filter based assimilation techniques. Correct representation of surface flow in hyper-resolution hydrologic models will also require much better representation of the subsurface. The importance of subsurface and surface water dynamics for land surface and land atmosphere exchanges has been addressed in various studies [*Bierkens and Van den Hurk*, 2007]. These studies suggest that there exists a strong linkage between the mass, energy, and momentum balances of the subsurface and the land surface, which require integration of two different paradigms.

The assimilation of the interpolated data into CLM4.0 is at a preliminary stage and the application of this technique combined with our use of kriging to interpolate water table observation data in sparsely measured regions requires further evaluation and is the basis of our ongoing research. Nonetheless, these results are very encouraging and will potentially help to usher in a new era and methodology of groundwater assimilation into LSMs.

3.5 Conclusion

The failure of CLM+DART is because of insufficient ensemble spread, which can be changed by making sure that the initial ensemble spread in the initialization files is sufficiently large. In this study we tried to do that by increasing the spin-up time but were unsuccessful. The reason was that for creating ensembles we were perturbing the atmospheric forcing, but the impact due to that on the water table depth variation was minimal. We have developed a new method where we use mathematical functions to create artificial spread in the water table depth at the start of the assimilation process. This initial perturbation of the water table depth ensures a large ensemble spread in water table depth in line with the observation data. Using this approach we believe we can now improve assimilation using the CLM+DART technique.

Application of this kriging-based interpolation and assimilation scheme into CLM4.0 can be expanded to other LSMs and other regions with minimal observation networks. This method may be more useful in areas with significant groundwater dependence and with inadequate recharge. I have been trying to develop this methodology with other observational techniques, including the Gravity Recovery and Climate Experiments (GRACE) and GPS, to further improve global groundwater monitoring capabilities. We can create GRACE water table variation data at 1km

resolution using specific yield data at that resolution for assimilation purpose [*Scanlon et al.*, 2012]. The new CLM4.5 model also has a much better developed specific yield parameterization to calculate water table depth and this should further improve the assimilation results.

As an application of these studies, I'm currently collaborating with a private company interested in groundwater monitoring by implementing the groundwater assimilation methodology to give water table depth information in space and time. The implementation of this technique on ground and the collaboration has been explained in the next chapter of this dissertation and is a manuscript under preparation.

Chapter 4: Application of Groundwater assimilation at high resolution over Paso Robles region, California

4.1 Introduction

Groundwater is a very important resource, especially for more than 2.5 billion people worldwide that completely depend on it for their drinking water needs. In the United States about 23 percent of the freshwater used in 2005 came from groundwater sources. Groundwater accounts for 98% of rural domestic supplies, 35% of public supplies and 42% of irrigation supplies in the United States. It is especially important in those parts of the country that don't have ample surface-water sources, such as the arid Western part of the country, including the California Central Valley and Mid-Western agricultural belt. It often takes more work and cost to access groundwater as opposed to surface water, but where there is little water on the land surface, groundwater can supply the water resource needs of people, animals, and agro-ecosystems. For 2005, most of the U.S. fresh groundwater withdrawals, 68 percent, were for irrigation, while another 19 percent was used for public-supply purposes, mainly to supply drinking water to the population. Groundwater is also crucial for people who supply their own water (domestic use), as over 98 percent of self-supplied domestic water withdrawals are from groundwater. The importance of groundwater increases even more during times of drought when surface water is not abundantly available, thus groundwater acts as a resource bank for water in times of need. Groundwater is being used unsustainably in most parts of the world and globally we are withdrawing 3.5 times more water than sustainable [*Gleeson et al.*, 2012]. The situation is especially alarming in places such as northwestern India [*Rodell et al.*, 2009] and the California Central Valley [*Famiglietti et al.*, 2011; *Steward et al.*, 2013]. The use of groundwater is rapidly growing with increasing population density and the respective increasing pace of consumption.

There is a recognized need to better assess, model and manage groundwater supplies to improve groundwater monitoring and prediction so that this precious resource can be scientifically managed. Presently there are no comprehensive national or global groundwater level networks in existence with uniform coverage of major aquifers, climate zones and land use types [*Hutson*, 2004; *Shah et al.*, 2001]. Even in California, the largest user of groundwater in the country, there is a severe lack of groundwater monitoring resources at sufficient spatial density. The USGS and DWR in California have several groundwater measurement sites but most of these wells do not take measurements continuously nor are they monitored regularly, on top of that the funding for such monitoring stations has actually been decreasing since the early 1990s. Groundwater modeling is also very challenging as even the most sophisticated models struggle with partially known subsurface features, errors in input forcing data and imperfect parameterizations in models. The groundwater models frequently have considerable drift in their outputs, are

poorly integrated with larger land surface processes and are ill suited for modeling groundwater at large continental scales. There is also a huge demand from people to know groundwater within their properties for past, present and future time periods. Groundwater also affects property values for many homeowners, farmers and communities.

There is a recognized need to know groundwater at sufficient spatial and temporal resolution. This can only be achieved through high resolution modeling outputs, but current Land Surface Models (LSMs) lack appropriate groundwater dynamics for high-resolution (1 km and higher) simulations. These problems can be overcome by using data assimilation to guide the high-resolution models towards more correct solutions. The sparse water table measurement values can be interpolated by kriging to obtain a high-resolution spatial map of water table depth in the region at certain times. These calculated spatial values can be inserted/assimilated into the model to improve spatio-temporal predictability. We have developed a methodology for assimilating observed groundwater depth measurements from multiple wells into a high spatial resolution LSM as explained in chapter 3. This assimilation technique has been designed to be useful in places with sparse observation data, with different types of observation data and it can be easily adapted to work with any LSM's that have a functional groundwater component. We have discussed this method in detail and discussed results in chapters 2 and 3 above [*Singh et al.*, 2014a; *Singh et al.*, 2014b]. Our methodology is ripe to help communities in groundwater banking and management in a scientific manner and our aim has been to use the techniques developed to benefit people by bringing it into the market for greater social impact.

4.2 Wellintel collaboration

We believe that the best way to help people with our groundwater modeling methodology would be to collaborate with companies and institutions that are already working on this problem and need such modeling services. Wellintel as a company is dedicated to offer solutions for an average user to monitor and learn about the groundwater levels on their properties. While their aim has been to provide an inexpensive and easy to use method for groundwater monitoring, it fits perfectly into our methodology of using observation data for assimilation to provide high-resolution spatio-temporal groundwater modeling results. Wellintel also saw the need to expand their service to their users and thus we both agreed to collaborate to develop this modeling technique under the UC Berkeley Industrial Alliance Program.

4.2.1 Wellintel

Wellntel is a water technology company that has developed a continuous, cloud-based groundwater level monitoring system designed to provide useful information

to groundwater stakeholders including private well owners like homeowners and farmers, and optionally, to scientists, financiers, insurers, regulators and businesses.

The Wellntel system consists of an inexpensive but accurate permanent sensor that measures and logs static groundwater level, pumping drawdown, recovery rates over time and saves the information securely and privately in the cloud, where users login, see and share their data, set alerts to prevent expensive failures, and can also view their groundwater information combined with other related data, like rainfall or soil moisture. While designed to be sold at a consumer price-point and installable using common tools, the Wellntel sensor uses a sophisticated temperature-compensated, sonar-based technology and a novel algorithm that is not confused by in-well obstruction or noise from pumping or water, and calibrates on-the-fly. It is battery powered, environmentally robust, and communicates wirelessly via a gateway or cell phone to the Internet. By treating virtually any borehole well as a groundwater level testing station and using supplemental geologic and environmental data, Wellntel data supports precise regional groundwater models and statistics that have been previously impossible, due to a lack of continuous measurements. Wellintel aims to make groundwater visible and sustainable with it's water level sensors providing unique water table depth time series data that will expand rapidly to include tens and thousands of location to become the largest such database in the United States.

Figure 4.1: Schematic of a Wellintel data sensor (left panel) and schematic of the data display and well location (right panel)

Wellntel sensor owners are usually private well owners who themselves use the information to lower risk, prevent emergencies, reduce service costs and surprises. They also face the risk of damaging their wells if they pump dry, Wellintel solves

that issue by giving them capacity to constantly monitor water table depth at the well site location. This can save thousands of dollars in well repair or replacement also as budding local groundwater experts; they often share what they know with science or local groundwater experts to strengthen groundwater management methods and policies. Eventually with multiple smart wells in a region, Wellintel will have a high quality, continuous water table depth dataset which can help drive a lot of innovation in groundwater modeling and management.

4.2.2 Synergy between the Hydroclimate Group at UCB and Wellintel

The Hydroclimate Group at UC Berkeley has expertise in regional-to-site scale hydrometeorology and groundwater modeling and data assimilation [*Maxwell and Miller*, 2005; *Miller and Kim*, 1996]. We have developed high-resolution dynamically coupled surface water-groundwater models [*Maxwell and Miller*, 2005; *Pan et al.*, 2008; *Singh et al.*, 2014b] and techniques for assimilating water table depth observations into groundwater models [*Singh et al.*, 2014b] to reduce uncertainty and improve historic and projected predictability. Wellintel has developed an operational groundwater level sensor for real-time monitoring at rural sites and has a business plan to install hundreds of thousands of low cost water table depth sensors, which will provide continuous high quality data. By combining such time-series data and our model-assimilation methods we can run models with observation data assimilation that can give us very high quality land surface data products, which can be used to create high resolution groundwater maps in past, present and future. These data products advance Wellintel's goals by providing their users with unmatched information about groundwater in and around their properties at various times. While we met Wellintel alongside a groundwater conference an important opportunity to collaborate was realized and we partnered with Wellintel to use their water table depth measurements for assimilation into our high-resolution model to provide better assessment of groundwater levels at various spatial resolution for both present and future times to the users.

4.3 Methods

Our goal is to assimilate sensor-based groundwater data into our surface-subsurface terrestrial hydrology model to yield high-resolution spatio-temporal fields. The water table depth can be modeled at high resolution in space and time for historic and projected changes. Here we plan to run the National Center for Atmospheric Research (NCAR) terrestrial hydrology model, the Community Land Model version 4.0 (CLM4.0) that we have advanced for data assimilation application [*Singh et al.*, 2014a; *Singh et al.*, 2014b]. Using the North American Land Data Assimilation System (NLDAS-2) atmospheric data [*Mitchell et al.*, 2004b] as forcing, we generate historic (2003-2006) simulations of groundwater variation with the assimilation of Wellintel flow data. This task will quantify the improvement in groundwater prediction through application of monitoring data. Simulations for this test study will initially be at a 15-minute time step and 1 km length scale. This can be

improved, but there remains a limit due to the length scale of available characterization data.

4.3.1 Model description

The model used in this work for implementing hyper-resolution terrestrial simulations is CLM4.0, the land component of the NCAR Community Earth System Model, which has been discussed in detail previously in this dissertation. The move to the more advanced CLM4.5 under CESM1.2 is under process and we are working on getting the DART assimilation methodology setup with the new CESM model.

4.3.2 Study area

The study area presented here is a 1^{o} x 1^{o} area in central California coast that extends from the western part of the northern California Central Valley (CCV) to the central coast with coordinates -121.00E 35.00N x -120.00E 36.00N and is divided into 100x100 grid cells, each 36 arc-second (~1km), for grid cell based simulations using CLM4.0 (Figure 4.2). This study area contains the Paso Robles groundwater basin that has been experiencing a severe groundwater shortage in recent times as industrial scale agriculture has expanded in the region, which relies on groundwater for their irrigation needs. There have also been a lot of new vineyard developments in the region that have been blamed for much of the decline. Wellintel has a large customer base in this area that are using their groundwater measurement devices to monitor groundwater and are willing to share the data for scientific research purposes. The area has its own challenges with large geographic variations from coastal regions to hilly vineyards to parts of the Central Valley. The study region covers the San Louis Obispo County and contains the Paso Robles groundwater basin.

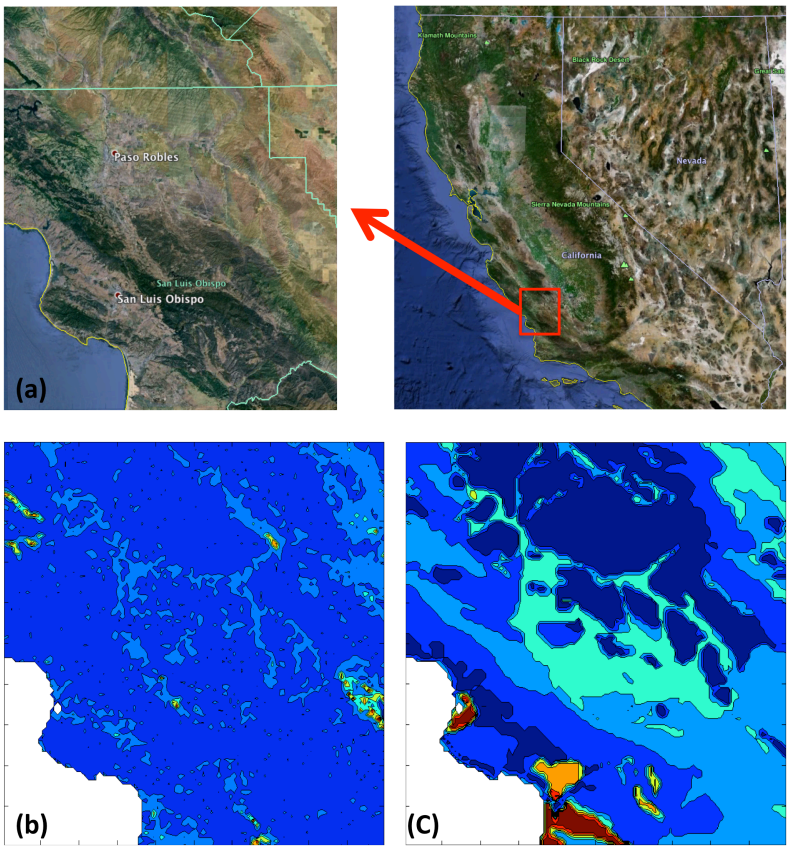

Figure 4.2: (a) The CLM4.0 model test domain, (b) Fraction of saturated area, F_{max}, and (c) Percentage Sand data at 0.01-degree resolution.

4.3.3 Data

Similar to the previously discussed study in chapters 2 and 3, this project required near-surface meteorological forcing data and surface datasets for running the model. We also need groundwater observation data for the kriging-based interpolation leading to assimilation. The high-resolution datasets were custom created by us and the other datasets were obtained from sources explained below. In general, access to datasets is one of the biggest limitations in running a proper high-resolution model and creating such datasets was one of the biggest challenges we faced.

4.3.3.1 Model surface datasets

The input forcing data is very similar to datasets used in the previous chapters though the data had to be created for the Paso Robles region at the resolution at which we wanted to run the model. The topographic data for the model runs in this study was generated at 1-km resolution using 1/3 arc-second (~10-m) resolution data available from the National Elevation Dataset (NED) USGS using methods described earlier in this. The high-resolution soil texture dataset was produced using the STATSGO [*Miller and White*, 1998] dataset available at 1km resolution.

Except for the topographic and soil texture data, which were calculated at 1km resolution by the authors, all other surface and aerosol input data were provided by the NCAR CESM forcing dataset library at 0.23^{o} x 0.31^{o} resolution . We also plan to run the model at much higher reolutions of 100 meters and 30 meters, and we have been creating datasets at these resolutions following the methods described earlier.

4.3.3.2 Model forcing datasets

Simulations at each resolution were forced with the 0.125° x .125° resolution for 01 Jan 1979 – present, with hourly atmospheric forcing data from the North American Land Data Assimilation System (NLDAS-2) atmospheric data, [*Mitchell et al.*, 2004b]. To achieve our goals for near real-time assimilation we are working on incorporating the latest NLDAS-2 forcing data to create forcing for the CESM model. The initialization files were created after spin up of the model from bare soil at each resolution for 20 years to reach thermal and hydrologic equilibrium [*Lo et al.*, 2008].

4.3.3.3 Groundwater data

Groundwater measurement data is obtained from datasets provided by WellIntel through their groundwater sensor network. Additional Wellintel sensors are still being installed and for now we are augmenting these observations with data from the California Department of Water Resources (DWR) managed well sites and the San Luis Obispo County Engineering Department managed well sites. The Wellintel sensors are mainly located in farms and vineyards in the region where users are eager to know the water table depth at their properties and around it.

In total there are more than a dozen Wellintel monitored well sites and over 300 well sites monitored by DWR and San Luis Obispo, but most of DWR and San Luis Obispo County observations are decades old and very few wells have the data for the months required for this study. We need at least a certain number of well observations to apply an ordinary kriging methodology to interpolate water table depth over the whole region and the results improve with more well observations in the region. We also perform cross validation tests to make sure kriging is giving us consistent results and compare it with observations on ground from multiple sources as explained in earlier chapters. The observed data is used to perform kriging for the start of every month, groundwater is temporally and spatially slowly varying with time and this helps to justify this assumption [*V Kumar and Remadevi*, 2006]. A proper data quality check is maintained so that any biased data is not used in the kriging process, first by filtering data on the basis of DWR listed codes for different disruption like leakages or pumping and secondly by removing outlier measurements. The well owners mostly use the US customary units and thus we have done our calculation in feet.

4.3.4 Data Assimilation

Data assimilation is a well-established technique used to improve model prediction, and is widely used in weather forecasting. The groundwater observation data from Wellintel and other sources is used for assimilation into the higher solution CLM4.0 model that should improve the model outputs as we have already established in previous chapters that assimilation of water table depth data into a high-resolution model significantly improves model predicted water table depth calculations. We use the same methods for data assimilation as established in chapter 3 and follow a very similar approach.

The sparse groundwater observation data are kriged using the methods described in chapter 3 and is used to produce interpolated water table depth values at each and every node (Figure 4.3) at the resolution at which we are running the model; 1 km for now but will change as we move forward with model data. As Wellintel is expanding rapidly in the region with aims of hundreds of sensors in a few months the kriged results should improve considerably in the future. The kriged data is assimilated into the model using the direct insertion method for now to check for consistency while we work on combining the latest version of CESM (with CLM 4.5) with the NCAR Data Assimilation Research Testbed (DART) Ensemble Kalman Filter system. Our goal is to set up an automated data assimilation system through which Wellintel water level sensor data can be assimilated at near real-time to provide continuous updating of water table depth variations over the domain.

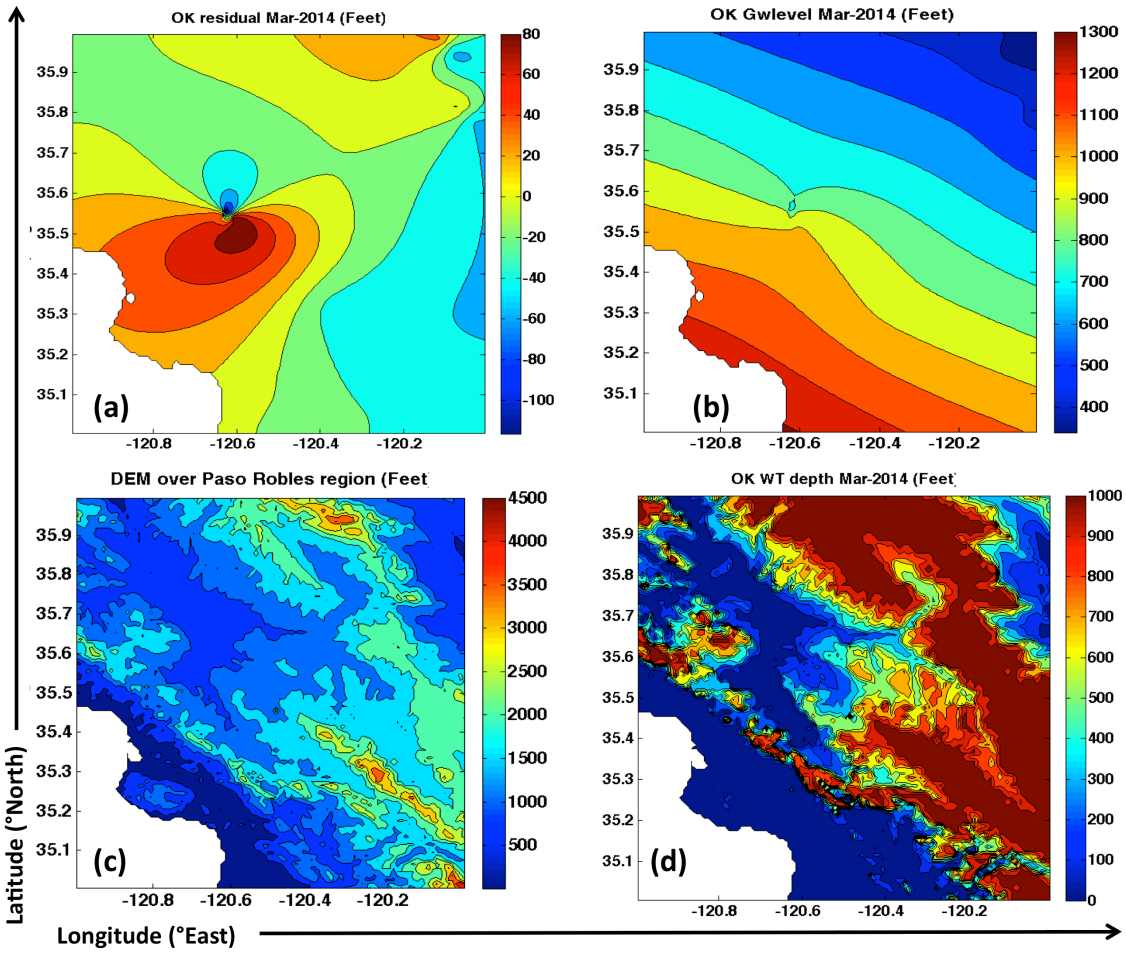

Figure 4.3: Kriging plots for March 20014 (a) Kriged residuals calculated over the site (b) Kriged groundwater level calculated over the site. (c) The DEM over the test region in feet. (d) Water table depth as calculated by the Kriging method and DEM dataset.

4.4 Implementation and traction

We have already established that the high resolution model with groundwater assimilation improves groundwater modeling (Chapter 3) [*Singh et al.*, 2014a], the challenge in Paso Robles has been to implement the methodology on ground and refine it to work with Wellintel data at the resolution and time scale that is needed.

The first step has been to run the model over the area of interest and characterize the water table depth calculated by the model. As we can see in Figure 4.4 the model gives a shallow water table depth throughout the region as expected. This is mainly due to the way groundwater is characterized in CLM4.0 and has been well explained in chapter 2 and chapter 3. The mean water table depth at 1 km resolution (Figure 4.3) does capture some of the spatial heterogeneity and follows the topography to

some extent but on the whole it is extremely shallow.

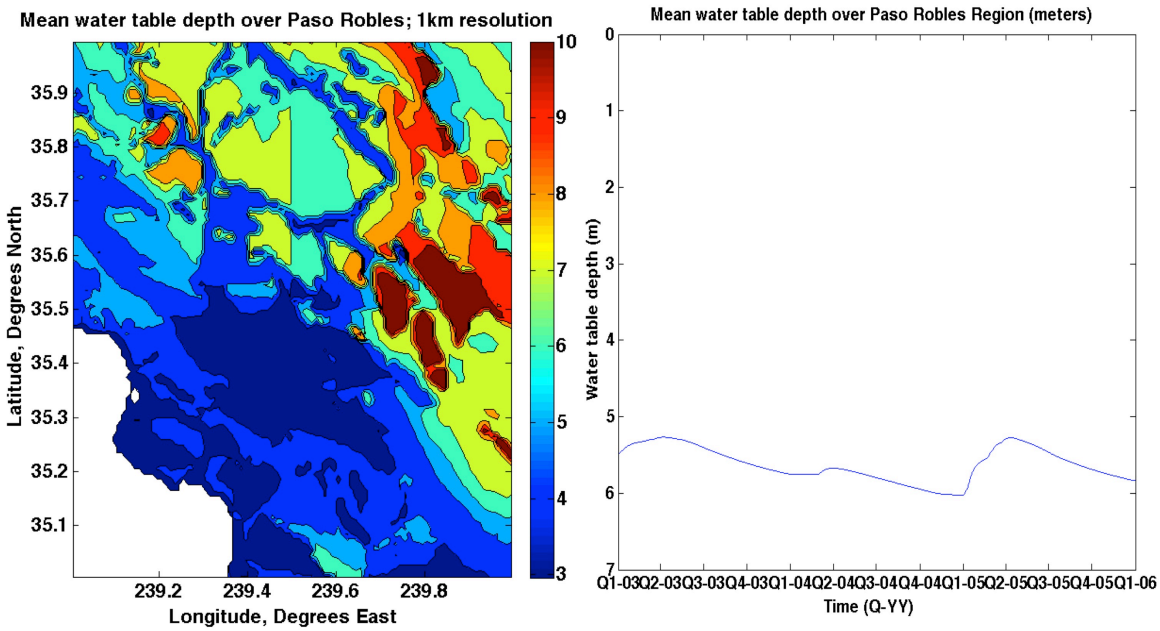

Figure 4.4: Mean water table depth (meters) over the Paso Robles region in time and space from 2003 to 2006

The second step was to do kriging of the observation water table data in the region as we can see in Figure 4.3. This kriged water table shows a much deeper water table profile throughout the region, which is more consistent with the observations. The Kriging output improves with the number of well locations and at present state we are waiting to obtain data from more wells to get good observation-interpolated data for assimilation.

These kriged observation values are then assimilated into the model to give us better water table depth in the region across various temporal and spatial scales. We want to use our improved CLM+DART assimilation methodology that will allow automation in assimilation. The CLM+DART assimilation process is still under construction and we are working with Wellintel to get this into production as soon as possible.

The collaboration is at a stage where we have proven that our methods work at a different location and initial implementation at client site locations. It is set to continue and we aim to provide first information to the users in a couple of months, and then provide an automated updating and data file transport procedure.

4.5 Discussion

Our experience in land surface modeling in the region shows that the high-resolution model with data assimilation will be very useful for the users. We are still awaiting more datasets from the region to get better-interpolated observation data that can be assimilated in CLM4.0. Our models are currently setup to predict water table depth and the collaboration is going strong with a lot of learning on both sides. Our group is very enthusiastic about our research being implemented on the ground as a real-world application that directly benefits users at a large scale. We are working to solve the issues of making the production of useful data quick, which can then benefit the company and its users immensely.

There is a lot of work that needs to be done to make this research more beneficial to the users community. We are planning to initiate monthly-to-seasonal groundwater forecasting with our assimilated models in the region. Water table depth forecasting will utilize the tested system with uncertainty analysis based on the historic simulations. By conditioning model simulations with weather input we will forecast water table depth. Our long experience in historic and projected regional atmospheric and land surface-subsurface modeling using reanalysis data, observations, and general circulation model data as initial and boundary conditions will prove invaluable to this task [*Kueppers et al.*, 2008; *Kyriakidis et al.*, 2001; *Miller and Kim*, 1996; *Miller et al.*, 1999; *Tseng et al.*, 2012]. Water table depth can be predicted for the region for various climate change scenarios thus helping users assess value and risk associated with their property. The models can be forced with climate change scenario projections and the outputs obtained can help to probabilistically inform users of future impacts on the groundwater availability in and around their property. Meaningful statistics of groundwater change as a function of weather variations and water withdrawal will be at the monthly-to-seasonal time scale and will give users knowledge on projections of water table changes that are coupled to weather variation and expected withdrawal rates. Projected outputs of seasonal forecasts for water table depth will inform users enabling them to better prepare for the incoming season.

We are also working to develop an easily accessible Graphical User Interface design based on Geographic Information System (GIS) spatial maps of data generated by the assimilation of Wellintel water level sensor data into CLM4.0. Maps will be overlaid on a GIS platform and be made accessible to users via smart phones and the web. The data in the servers used by the frontend viewing platform can be updated continuously with the model outputs. The model outputs at a chosen timescale can be overlaid over a map of the region using the GIS interface and users can view the spatio-temporal water table depth variations via the web and smart phones. The interface will be made intuitive such that changes in water table depth in and around their property are easily viewed. We will incorporate statistical graphs and charts to provide important information tailored to user needs.

4.6 Conclusion

Our collaboration with Wellintel is still developing and additional needs to be done and are learning about critical issues that users on the ground care about and which datasets are most useful to them. We are working with Wellintel to develop specific datasets in the time frame that is deemed most useful.

The collaboration has a promising future with our team gearing up to run the model at even higher resolution with improved parameterizations and real-time assimilation using the CLM+DART setup. We are also working towards a set of static characterizations of the stratigraphy, soil, and urban infrastructure, as well as dynamic characterizations for vegetation and any other time-varying variables. Wellintel is all set to expand rapidly this year, thus providing the best water table depth measurement datasets in the country at near real-time.

Chapter 5: Conclusion and Recommendation

Modeling groundwater is challenging: groundwater is not a readily visible quantity and is difficult to measure, with limited sets of observations available. Although groundwater models can reproduce water table and head variations, partially known geologic structure, errors in the input forcing fields, and imperfect LSM parameterizations can lead to considerable drift in modeled land surface states. As a result, these models frequently produce biased results that are very different from observations. While many hydrologic groups are grappling with developing better models, it is also possible to instead make the existing models more robust through data assimilation of observed groundwater data. The goal of this research was to develop a methodology for high-resolution LSM runs and to improve hydrologic modeling through observation data assimilation, and then to apply the improved model for groundwater monitoring and banking.

The high resolution CLM4.0 simulations in this dissertation show that the model physics performs well at these resolutions, and leads to better modeling of water and energy budget terms. The assimilation methodology answers the critical question of how to improve groundwater modeling in LSMs that lack sufficient sub-surface parameterizations, and also how to run these LSMs globally at hyper-resolution scales.

5.1 Summary and Conclusion

My doctoral dissertation had three distinct research goals. The first was to run a commonly used LSM at hyper resolution, showing that doing so did not break the model and in fact improved modeling results. The second was to develop an observation data assimilation methodology that improves the high-resolution model. The third goal was to show the applicability of this approach as part of a real-world need for a well-defined user community.

The need for improved accuracy drives the development of such *hyper-resolution* LSMs, which can be implemented at continental scales with resolutions of 1 km or finer. In Chapter 2 we reported our research incorporating fine-scale grid resolutions and surface data into NCAR Community Land Model version 4.0 (CLM v4.0) for simulations at 1 km, 25 km, and 100 km resolutions using 1 km soil and topographic information. Multi-year model runs were performed over the southwestern United States, including the entire state of California and the Colorado River basin, with results demonstrating changes in the total amount of CLM-modeled water storage and in the spatial and temporal distributions of water in snow and soil reservoirs, as well as in surface fluxes and energy balance. We have also demonstrated the critical scales at which important hydrological processes—such as snow water equivalent, soil moisture content, and runoff—begin to more accurately capture the magnitude of the land water balance for the entire domain.

This proves that grid resolution is itself a critical component of accurate model simulations, and of hydrologic budget closure.

We also compared simulation outputs to station and gridded observations of model fields. Although the higher grid resolution model is not driven by high-resolution forcing, grid resolution changes alone nonetheless yield a significant reduction in differences between the model's output and direct observations; the RMSE decreases by more than 35%, 36%, 34%, and 12% for soil moisture, terrestrial water storage anomaly, sensible heat, and snow water equivalent, respectively. Additionally, we performed a 100 m resolution simulation over a spatial sub-domain, the results of which indicate that parameters such as drainage, runoff, and infiltration are significantly impacted when hillslope scales of ~100 m or finer are considered. We also show how limitations of the current model physics, including the absence of lateral flow between grid cells, can affect model simulation accuracy.

The results presented in Chapter 2 are encouraging, but also point to limitations in improving an LSM only by increasing spatial resolution and surface datasets. As shown in the water table depth analysis, increased model resolution alone cannot compensate for errors in parameterization and for the lack of sub-surface information in CLM4.0. This problem can be solved by providing additional information to the model in the form of water table depth data assimilation.

In Chapter 3, I provided the development and verification of a methodology for assimilating observed groundwater depth measurements from multiple wells into the high spatial resolution LSM. A kriging-based interpolation technique overcomes the problem of spatially and temporally sparse observations, and the interpolated data is assimilated into CLM4.0 at 1 km resolution in a test region in northern California. Direct insertion and EAKF-based techniques are used for assimilation, with direct insertion producing better results and showing major improvement in the simulation of water table depth. The Linear Pearson correlation coefficient between the observed well data and the assimilated model is 0.810, but with the non-assimilated model it is only 0.107. This improvement is most significant when water table depth is greater than 5 m. Assimilation also improves soil moisture content, especially in the dry season when the water table is at its lowest. Other variables, including sensible heat flux, ground evaporation, infiltration, and runoff are analyzed to quantify the effects of this new assimilation methodology. Though the changes in these variables seem small, they can nonetheless be critical in coupled models, and the improved simulation of groundwater in the assimilated model can explain the change in results.

In chapter 4 I describe the industrial collaboration to implement our assimilation technique for a domain along the central California coast. This assimilation technique has been designed to be useful in regions with sparse and varied observation data, and it can be easily adapted to work in other LSMs that have a functional groundwater component. Using observation data from Wellintel Inc. we have implemented this methodology and are bringing this forward as an operational

procedure by setting up our model over a large region centered near Paso Robles where the company has been running a pilot for its water table depth measuring devices. Wellintel is set to expand rapidly this year, providing the best water table depth measurement datasets in the country at near real-time. The aim of this collaboration is to provide users with actionable water table depth data in and around their properties for the recent past, present, and for future projections. This methodology combined with Wellintel data is being developed into a groundwater-management and groundwater-banking tool.

This is the first time that groundwater assimilation has been attempted in a high-resolution LSM, and hence this project provides a unique methodology for applying the benefits of a global LSM that can be run at hyper resolution with data assimilation to improve its groundwater modeling. The whole setup offers a powerful tool to researchers for modeling land surface parameters better than ever before, as it is readily transferable to most any LSM.

5.2 Recommendations

Model outputs do not show much improvement as resolution increases from ~100 km to ~25 km because there is little difference in topographic information between the resolutions; these scales are still too coarse. Most model parameters begin to improve at 1 km resolution, but others, such as drainage, infiltration, and runoff improve only as we reach 100 m or hill slope scale resolution. There is a recognized need to increase resolution to at least hill slope scale—100 m resolution or finer—with better representation of the sub-surface soil texture and stratigraphy. The next step is to develop parameterizations at the required resolution and to improve how variables such as water table depth are calculated in the CLM4.0 physics. Parameter calibration efforts will likely be fruitless (or at least much less effective) if an appropriate model resolution is not achieved first.

To achieve the maximum impact in our work, we need to strengthen our collaboration with partners such as Wellintel Inc. This collaboration has a promising future, with our team gearing up to run the model at even higher resolution with improved parameterizations and real-time assimilation using the CLM+DART setup. We are also working toward a set of static characterizations of the stratigraphy, soil, and urban infrastructure, as well as dynamic characterizations for vegetation and other time-varying variables. The CLM+DART methodology needs to be improved, as that might be the best method for data assimilation.

References

Anderson, J. L. (2001), An Ensemble Adjustment Kalman Filter for Data Assimilation, *Monthly Weather Review*, *129*(12), 2884-2903.

Baldocchi, D. D. (2011), FLuxnet Data Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC), edited.

Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology *Hydrological Sciences Bulletin*, *24*(1), 43-69.

Bierkens, M. F., and B. J. Van den Hurk (2007), Groundwater convergence as a possible mechanism for multi‐year persistence in rainfall, *Geophysical research letters*, *34*(2).

Branstetter, M. L. (2001), Development of a parallel river transport algorithm and applications to climate studies, University of Texas at Austin.

Center, N. O. H. R. S. (2004), Snow Data Assimilation System (SNODAS) Data Products at NSIDC, edited, Boulder, Colorado USA: National Snow and Ice Data Center.

Chen, F., W. T. Crow, P. J. Starks, and D. N. Moriasi (2011), Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture, *Advances in Water Resources*, *34*(4), 526-536.

Christensen, O. B., J. H. Christensen, B. Machenhauer, and M. Botzet (1998), Very High-Resolution Regional Climate Simulations over Scandinavia‚ÄîPresent Climate, *Journal of Climate*, *11*(12), 3204-3229.

Chung, J.-w., and J. D. Rogers (2011), Interpolations of Groundwater Table Elevation in Dissected Uplands, *Ground Water*, no-no.

Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resour. Res.*, *14*(4), 601-604.

Cosby, B. J., G. M. Hornberger, R. B. Clapp, and T. R. Ginn (1984), A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils, *Water Resour. Res.*, *20*(6), 682-690.

Decker, M. R., and X. Zeng (2009), Impact of Modified Richards Equation on Global Soil Moisture Simulation in the Community Land Model (CLM3.5), *Journal of Advances in Modeling Earth Systems*, *1*(Art. #5), 22 pp.

Effort, C. o. F. O. f. M. i. t. N. s. S. R., W. Science, T. Board, D. o. Earth, L. Studies, and N. R. Council (2012), *Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites*, The National Academies Press.

Entekhabi, D., and P. S. Eagleson (1989), Land Surface Hydrology Parameterization for Atmospheric General Circulation models Including Subgrid Scale Spatial Variability, *Journal of Climate*, *2*(8), 816-831.

Famiglietti, J. S., and E. F. Wood (1994), Multiscale modeling of spatially variable water and energy balance processes, *Water Resources Research*, *30*(11), 3061-3078.

Famiglietti, J. S., J. W. Rudnicki, and M. Rodell (1998), Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas, *Journal of Hydrology*, *210*(1,Äì4), 259-281.

Famiglietti, J. S., L. Murdoch, V. Lakshmi, and R. P. Hooper (2009), Towards a framework for community modeling in hydrologic science: Blueprint for a community hydrologic modeling platform, paper presented at 2nd Workshop on a Community Hydrologic Modeling Platform, *Univ. of Memphis, Memphis, Tenn., 31 March to 1 April.*

Famiglietti, J. S., M. Lo, S. L. Ho, J. Bethune, K. J. Anderson, T. H. Syed, S. C. Swenson, C. R. de Linage, and M. Rodell (2011), Satellites measure recent rates of groundwater depletion in California's Central Valley, *Geophys. Res. Lett.*, *38*(3), L03403.

Fan, Y., G. Miguez‐Macho, C. P. Weaver, R. Walko, and A. Robock (2007), Incorporating water table dynamics in climate modeling: 1. Water table observations and the equilibrium water table, *J. Geophys. Res.*

Flanner, M. G., and C. S. Zender (2005), Snowpack radiative heating: Influence on Tibetan Plateau climate, *Geophysical Research Letters*, *32*(6), n/a-n/a.

Flanner, M. G., C. S. Zender, J. T. Randerson, and P. J. Rasch (2007), Present-day climate forcing and response from black carbon in snow, *Journal of Geophysical Research: Atmospheres*, *112*(D11), n/a-n/a.

Gent, P. R., S. G. Yeager, R. B. Neale, S. Levis, and D. A. Bailey (2010), Improvements in a half degree atmosphere/land version of the CCSM, *Climate dynamics*, *34*(6), 819-833.

Giorgi, F. (1990), Simulation of Regional Climate Using a Limited Area Model Nested in a General Circulation Model, *Journal of Climate*, *3*(9), 941-963.

Gleeson, T., Y. Wada, M. F. Bierkens, and L. P. van Beek (2012), Water balance of global aquifers revealed by groundwater footprint, *Nature*, *488*(7410), 197-200.

Gochis, D. J., E. R. Vivoni, and C. J. Watts (2010), The impact of soil depth on land surface energy and water fluxes in the North American Monsoon region, *Journal of Arid Environments*, *74*(5), 564-571.

Godin, J. (2012), South Florida Environmental Report*Rep.*

Goovaerts, P. (1997), *Geostatistics for natural resources evaluation*, Oxford university press.

Haddeland, I., B. V. Matheussen, and D. P. Lettenmaier (2002), Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model, *Water Resources Research*, *38*(7), 29-21-29-10.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich (1998), Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resources Research*, *34*(12), 3405-3420.

Hutson, S. S., N. L. Barber, J. F. Kenny, K. S. Linsey, D. S. Lumia, and M. A. Maupin (2004), Estimated use of water in the United States in 2000, *USGS Circular 1268*.

Isaaks, E., and M. Srivastava (1990), *An Introduction to Applied Geostatistics*, Oxford University Press, USA.

Ivanov, V. Y., E. R. Vivoni, R. L. Bras, and D. Entekhabi (2004), Catchment hydrologic response with a fully distributed triangulated irregular network model, *Water Resources Research*, *40*(11), W11102.

Jackson, T. J., M. H. Cosh, R. Bindlish, P. J. Starks, D. D. Bosch, M. Seyfried, D. C. Goodrich, M. S. Moran, and D. Jinyang (2010), Validation of Advanced Microwave Scanning Radiometer Soil Moisture Products, *Geoscience and Remote Sensing, IEEE Transactions on*, *48*(12), 4256-4272.

Jana, R. B., and B. P. Mohanty (2012a), On topographic controls of soil hydraulic parameter scaling at hillslope scales, *Water Resources Research*, *48*(2), n/a-n/a.

Jana, R. B., and B. P. Mohanty (2012b), A topography-based scaling algorithm for soil hydraulic parameters at hillslope scales: Field testing, *Water Resources Research*, *48*(2).

Jin, J., and N. L. Miller (2007), Analysis of the Impact of Snow on Daily Weather Variability in Mountainous Regions Using MM5, *Journal of Hydrometeorology*, *8*(2), 245-258.

Jin, J., N. L. Miller, and N. Schlegel (2010), Sensitivity Study of Four Land Surface Schemes in the WRF Model, *Advances in Meteorology*, *vol. 2010*, 11 pages.

Jones, R. G., J. M. Murphy, and M. Noguer (1995), Simulation of climate change over europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries, *Quarterly Journal of the Royal Meteorological Society*, *121*(526), 1413-1449.

Kluzek, E. (2012), CESM Research Tools: CLM4 in CESM1.0.4 User's Guide Documentation, edited.

Kollet, S. J., and R. M. Maxwell (2008), Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model, *Water Resour. Res.*, *44*(2), W02402.

Kollet, S. J., R. M. Maxwell, C. S. Woodward, S. Smith, J. Vanderborght, H. Vereecken, and C. Simmer (2010), Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources, *Water Resour. Res.*, *46*(4), W04201.

Kueppers, L. M., et al. (2008), Seasonal temperature responses to land-use change in the western United States, *Global and Planetary Change*, *60*(3,Äì4), 250-264.

Kumar, A., C. Peterslidard, Y. Tian, P. Houser, J. Geiger, S. Olden, L. Lighty, J. Eastman, B. Doty, and P. Dirmeyer (2006), Land information system: An interoperable framework for high resolution land surface modeling, *Environmental Modelling & Software*, *21*(10), 1402-1415.

Kumar, V., and Remadevi (2006), Kriging of Groundwater Levels – A Case Study, *Journal of Spatial Hydrology*, *6*.

Kuo, W.-L., T. S. Steenhuis, C. E. McCulloch, C. L. Mohler, D. A. Weinstein, S. D. DeGloria, and D. P. Swaney (1999), Effect of grid size on runoff and soil moisture for a variable-source-area hydrology model, *Water Resources Research*, *35*(11), 3419-3428.

Kyriakidis, P. C., N. L. Miller, and J. Kim (2001), Uncertainty Propagation of Regional Climate Model Precipitation Forecasts to Hydrologic Impact Assessment, *Journal of Hydrometeorology*, *2*(2).

Lawrence, D., et al. (2011), Parameterization Improvements and Functional and Structural Advances in Version 4 of the Community Land Model, *Journal of Advances in Modeling Earth Systems*, *3*(M03001), 27 pp.

Leuangthong, O., KHAN, K. D., and DEUTSCH, C. V (2008), *Solved Problem in Geostatistics* Wiley.

Levis, S., G. B. Bonan, E. Kluzek, P. E. Thornton, A. Jones, W. J. Sacks, and C. J. Kucharik (2012), Interactive Crop Management in the Community Earth System Model (CESM1): Seasonal Influences on Land,ÄìAtmosphere Fluxes, *Journal of Climate*, *25*(14), 4839-4859.

Liang, X., Z. Xie, and M. Huang (2003), A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity (VIC) land surface model, *J. Geophys. Res.*, *108*(D16), 8613.

Lo, M.-H., and J. S. Famiglietti (2010), Effect of water table dynamics on land surface hydrologic memory, *J. Geophys. Res.*, *115*(D22), D22118.

Lo, M.-H., and J. S. Famiglietti (2011), Precipitation response to land subsurface hydrologic processes in atmospheric general circulation model simulations, *J. Geophys. Res.*, *116*(D5), D05107.

Lo, M.-H., P. J. F. Yeh, and J. S. Famiglietti (2008), Constraining water table depth simulations in a land surface model using estimated baseflow, *Advances in Water Resources*, *31*(12), 1552-1564.

Lo, M.-H., J. S. Famiglietti, P. J. F. Yeh, and T. H. Syed (2010), Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data, *Water Resour. Res.*, *46*(5), W05517.

Maxwell, R. M., and N. L. Miller (2005), Development of a Coupled Land Surface and Groundwater Model, *Journal of Hydrometeorology*, *6*(3), 233-247.

Meierdiercks, K. L., J. A. Smith, M. L. Baeck, and A. J. Miller (2010), Analyses of Urban Drainage Network Structure and its Impact on Hydrologic Response1, *JAWRA Journal of the American Water Resources Association*, *46*(5), 932-943.

Meissner, C., and S. Gerd (2009), High-resolution sensitivity studies with the regional climate model COSMO-CLM, *Meteorologische Zeitschrift*, *18*, 543-557.

Miguez-Macho, G., Y. Fan, C. P. Weaver, R. Walko, and A. Robock (2007), Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation, *J. Geophys. Res.*, *112*(D13), D13108.

Miller, and J. Kim (1996), Numerical prediction of precipitation and river flow over the Russian River watershed during the January 1995 California storms, *Bulletin of the American Meteorological Society*, *77*(1), 101-105.

Miller, and R. A. White (1998), A Conterminous United States Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling, *Earth Interactions*, *2*(2), 1-26.

Miller, J. Kim, R. K. Hartman, and J. Farrara (1999), Downscaled Climate and Streamflow Study of the south west united states, *JAWRA Journal of the American Water Resources Association*, *35*(6), 1525-1537.

Mitchell, K. E., et al. (2004a), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *Journal of Geophysical Research: Atmospheres*, *109*(D7), D07S90.

Mitchell, K. E., et al. (2004b), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a

continental distributed hydrological modeling system, *Journal of Geophysical Research: Atmospheres*, *109*(D7).

Moradkhani, H. (2008), Hydrologic Remote Sensing and Land Surface Data Assimilation, *Sensors*, *8*(5), 2986-3004.

Niu, G.-Y., and Z.-L. Yang (2006), Effects of Frozen Soil on Snowmelt Runoff and Soil Water Storage at a Continental Scale, *Journal of Hydrometeorology*, *7*(5), 937-952.

Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, and L. E. Gulden (2005), A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models, *J. Geophys. Res.*, *110*(D21), D21106.

Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su (2007), Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *J. Geophys. Res.*, *112*(D7), D07103.

NRC (2000), Investigating groundwater systems on regional and national scales, edited by N. A. Press, p. 143, Washington, D. C.

Oleson, G.B. Bonan, and J. Feddema (2010a), Technical Description of an Urban Parameterization for the Community Land Model (CLMU), *NCAR Technical Note NCAR*.

Oleson, D. M. Lawrence, G. B. Bonan, M. G. Flanner, E. Kluzek, P. J. Lawrence, S. Levis, S. C. Swenson, and A. D. P.E. Thornton, M. Decker, R. Dickinson, J. Feddema, C.L. Heald, F. Hoffman, J.-F. Lamarque, N. Mahowald, G.-Y. Niu, T. Qian, J. Randerson, S. Running, K. Sakaguchi, A. Slater, R. Stockli, A. Wang, Z.-L. Yang, Xi. Zeng, and Xu. Zeng, (2010b), Technical Description of version 4.0 of the Community Land Model (CLM), *NCAR TECHNICAL NOTE*.

Oleson, et al. (2008), Improvements to the Community Land Model and their impact on the hydrological cycle, *J. Geophys. Res.*, *113*(G1), G01021.

Pan, M., and E. F. Wood (2009), A Multiscale Ensemble Filtering System for Hydrologic Data Assimilation. Part II: Application to Land Surface Modeling with Satellite Rainfall Forcing, *Journal of Hydrometeorology*, *10*(6), 1493-1506.

Qian, T. T., A. Dai, K. E. Trenberth, and K. W. Oleson (2006), Simulation of global land surface conditions from 1948 to 2004. Part I: Forcing data and evaluations, *Journal of Hydrometeorology*, *7*(5), 953-975.

Quinn, P. F., K. J. Beven, and R. Lamb (1995), The in(a/tan/ β) index: How to calculate it and how to use it within the topmodel framework, *Hydrological Processes*, *9*(2), 161-182.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic Data Assimilation with the Ensemble Kalman Filter, *Monthly Weather Review*, *130*(1), 103-114.

Reichle, R. H., M. G. Bosilovich, W. T. Crow, R. D. Koster, S. V. Kumar, S. P. P. Mahanama, B. F. Zaitchik, S. K. Park, and L. Xu (2009), Recent Advances in Land Data Assimilation at the NASA Global Modeling and Assimilation Office
Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, edited, pp. 407-428, Springer, Berlin Heidelberg.

Rigon, R., G. Bertoldi, and T. M. Over (2006), GEOtop: A Distributed Hydrological Model with Coupled Water and Energy Budgets, *Journal of Hydrometeorology*, *7*(3), 371-388.

Rodell, M., I. Velicogna, and J. S. Famiglietti (2009), Satellite-based estimates of groundwater depletion in India, *Nature*, *460*(7258), 999-1002.

Scanlon, B. R., C. C. Faunt, L. Longuevergne, R. C. Reedy, W. M. Alley, V. L. McGuire, and P. B. McMahon (2012), Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, *Proceedings of the national academy of sciences*, *109*(24), 9320-9325.

Shah, T., D. Molden, R. Sakthivadivel, and D. Seckler (2001), Global Groundwater Situation: Opportunities and Challenges, *Economic and Political Weekly*, *36*(43), 4142-4150.

Singh, R., J. T. Reager, N. L. Miller, and Y. Rubin (2014a), Assimilation of Groundwater measurement data at high resolution into the Community Land Model 4.0: Results from Sierra foothill region in Northern California, *Water Resources Research*.

Singh, R., J. T. Reager, N. L. Miller, and J. S. Famiglietti (2014b), Towards hyper-resolution land surface modeling: The effects of fine-scale model grid resolution on CLM4.0 simulations in the Southwestern US, *Water Resources Research*, *In review*.

Skamarock, W., J. Klemp, J. Dudhia, D. Gill, D. Barker, W. Wang, and J. Powers (2005), A Description of the Advanced Research WRF Version 2, *NCAR Tech Notes-468+ STR.*

Steward, D. R., P. J. Bruss, X. Yang, S. A. Staggenborg, S. M. Welch, and M. D. Apley (2013), Tapping unsustainable groundwater stores for agricultural production in the High Plains Aquifer of Kansas, projections to 2110, *Proceedings of the National Academy of Sciences*.

Tian, X. (2008), A land surface soil moisture data assimilation system based on the dual-UKF method and the Community Land Model (DOI 10.1029/2007JD009650), *Journal of Geophysical Research*, *113*(14).

Tseng, Y.-H., S.-H. Chien, J. Jin, and N. L. Miller (2012), Modeling Air-Land-Sea Interactions Using the Integrated Regional Model System in Monterey Bay, California, *Monthly Weather Review*, *140*(4).

VanderKwaak, J. E., and K. Loague (2001), Hydrologic-Response simulations for the R-5 catchment with a comprehensive physics-based model, *Water Resources Research*, *37*(4), 999-1013.

Vivoni, E. R., V. Y. Ivanov, R. L. Bras, and D. Entekhabi (2005), On the effects of triangulated terrain resolution on distributed hydrologic model response, *Hydrological Processes*, *19*(11), 2101-2122.

Wackernagel, H. (1999), *Multivariate Geostatistics: An Introduction With Applications*, {Springer-Verlag Telos}.

Wahr, J., S. Swenson, V. Zlotnicki, and I. Velicogna (2004), Time-variable gravity from GRACE: First results, *Geophys. Res. Lett.*, *31*(11), L11501.

Walker, and P. R. Houser (2001), A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, *Journal of Geophysical Research: Atmospheres*, *106*(D11), 11761-11774.

Walker, and P. R. Houser (2005), *Hydrologic Data Assimilation*.

Wang, A., and X. Zeng (2009), Improving the treatment of the vertical snow burial fraction over short vegetation in the NCAR CLM3, *Adv. Atmos. Sci.*, *26*(5), 877-886.

Wolock, and C. V. Price (1994), Effects of Digital Elevation Model Map Scale and Data Resolution on a Topography-Based Watershed Model, *Water Resources Research*, *30*(11), 3041-3052.

Wolock, and G. J. McCabe (2000), Differences in topographic characteristics computed from 100- and 1000-m resolution digital elevation model data, *Hydrological Processes*, *14*(6), 987-1002.

Wood, E. F., et al. (2011), Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, *47*(5), W05301.

Xie, Z., F. Yuan, Q. Duan, J. Zheng, M. Liang, and F. Chen (2007), Regional Parameter Estimation of the VIC Land Surface Model: Methodology and Application to River Basins in China, *Journal of Hydrometeorology*, *8*(3), 447-468.

Yeh, P. J.-F., and E. A. B. Eltahir (2005a), Representation of Water Table Dynamics in a Land Surface Scheme. Part II: Subgrid Variability, *Journal of Climate*, *18*(12), 1881-1901.

Yeh, P. J.-F., and E. A. B. Eltahir (2005b), Representation of Water Table Dynamics in a Land Surface Scheme. Part I: Model Development, *Journal of Climate*, *18*(12), 1861-1880.

Zaitchik, B. F., M. Rodell, and R. H. Reichle (2008), Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin, *Journal of Hydrometeorology*, *9*(3), 535-548.

Zeng, X., and M. Decker (2009), Improving the Numerical Solution of Soil Moisture,ÄìBased Richards Equation for Land Models with a Deep or Shallow Water Table, *Journal of Hydrometeorology*, *10*(1), 308-319.

Zhang, W., and D. R. Montgomery (1994), Digital elevation model grid size, landscape representation, and hydrologic simulations, *Water Resources Research*, *30*(4), 1019-1028.

Appendix A: Calculation of Topographic Index

Topographic/Wetness index (TI) is a very important parameter in CLM that helps on calculating the maximum saturated fraction (f_{max}). f_{max} and its effect on calculating the various land surface parameters is well explains in Chapter 1. TI is the main way that the topographic information is relayed to the model in CLM. Topographic Index used in this study has been calculated explicitly for each grid cell at the resolution the model is run using the USGS 1/3 arc second (~10m) National Elevation Dataset (NED, Data available from USGS) using the process described in [*Quinn et al.*, 1995; *Wolock and McCabe*, 2000]. TI is defined as:

$$TI = \ln (A/\tan\beta) \qquad (B.1)$$

where A is the upstream or contributing area per unit contour length, and β is the grid cell topographic slope angle [*Beven and Kirkby, 1979*]. TI is a function of both the slope and the upstream contributing area per unit width orthogonal to the flow direction. It is actually the inverse of the stream-power law and therefore relates to fluid flow and deposition within the landscape. TI is less for steeper slopes and more for flat regions.

TI at 10m-resolution was not available even though DEM at that resolution has been available for some time. To use high-resolution information I calculated Topographic Index at 10m-resolution using the method described in [*Beven and Kirkby*, 1979; *Quinn et al.*, 1995]. I used ArcGIS™ to create a model for calculating TI from DEM data obtained from NED (Figure B1). Steps involved in calculation of topographic index as in the model are following.

1. Filling DEM

   The DEM data is analyzed and sink terms are filled.

2. Calculating slope

   Slope of the terrain is calculated and converted into radians. B

3. Calculating flow direction

   Flow direction is encoded as an angle in radians counter-clockwise from east as a continuous (floating point) quantity between 0 and 2 pi. The flow direction angle is determined as the direction of the steepest downward slope on the eight triangular facets formed in a 3 x 3 grid cell window centered on the grid cell of interest. (Figure B1)

4. Calculating flow accumulation

   Flow accumulation is computed in terms of the number of grid cells draining into the grid cell of interest. The contribution at each grid cell is taken initially as one. The contributing area of each grid cell is then taken as its own contribution plus the contribution from upslope neighbors that have some fraction draining to it. The flow from each cell either all drains to one neighbor, if the angle falls along a cardinal (0, p/2, p, 3p/2) or diagonal (p/4, 3p/4, 5p/4, 7p/4) direction, or is on an angle falling between the direct angle to two adjacent neighbor. 'A' is calculated

5. Calculating Topographic Index = ln (A/tanβ)

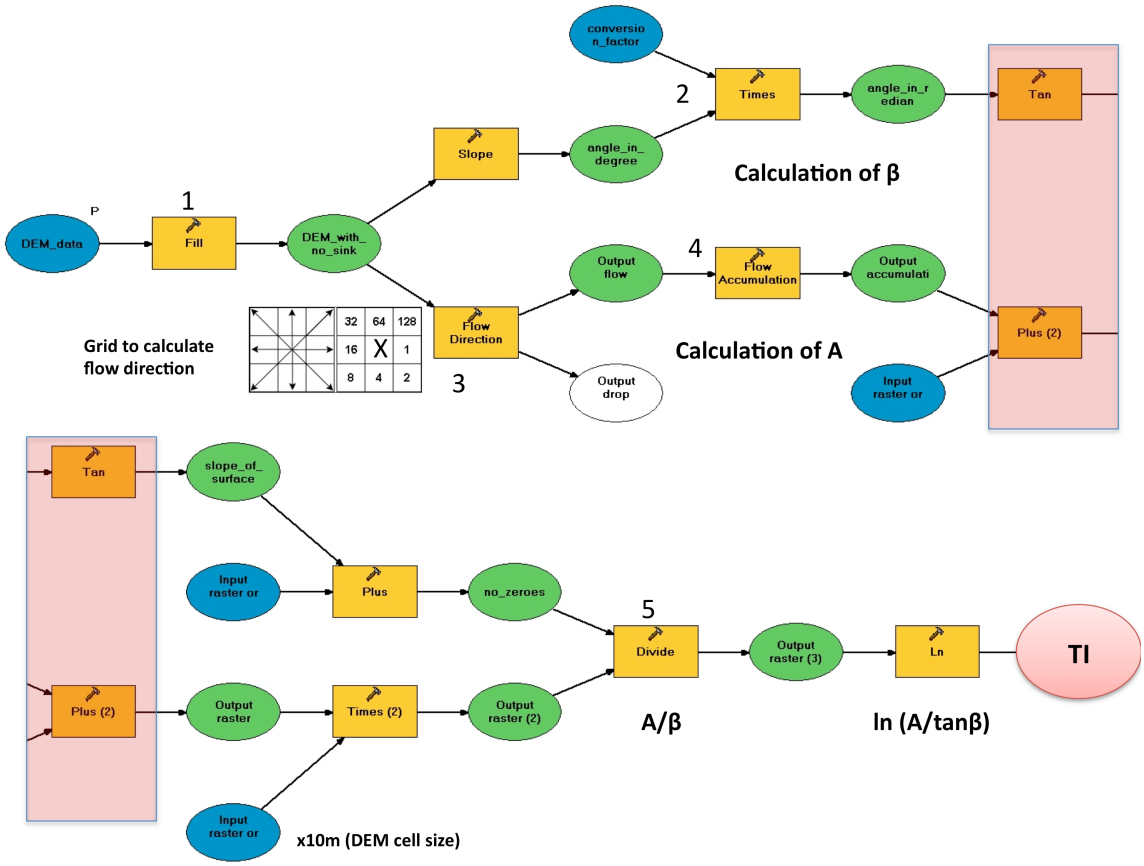

Figure B1: ArcGIS model for calculating Topographic Index from DEM obtained from USGS-NED

The biggest challenge in calculating the TI was the size of the dataset involved in calculations. The DEM dataset over all of SWUS at 10m-resolution was in many gigabytes. Calculation with such large data in ArcGIS took many days to complete.

The following figures show TI calculated at 10m-resolution over the area of interest in this dissertation.

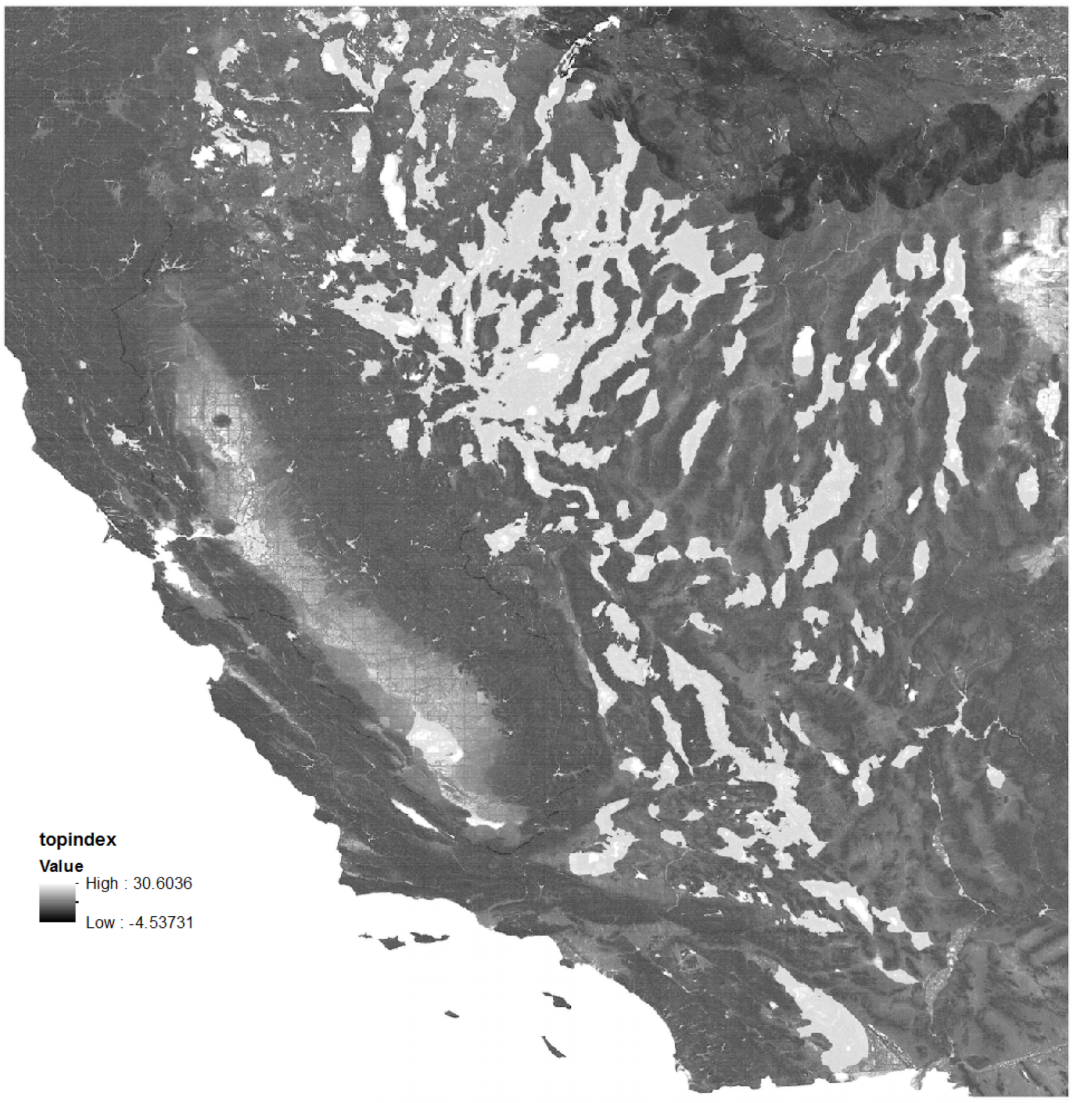

Figure B2: Topographic Index at 10m resolution over the SWUS over which the model was run at 1km resolution as discussed in Chapter 2

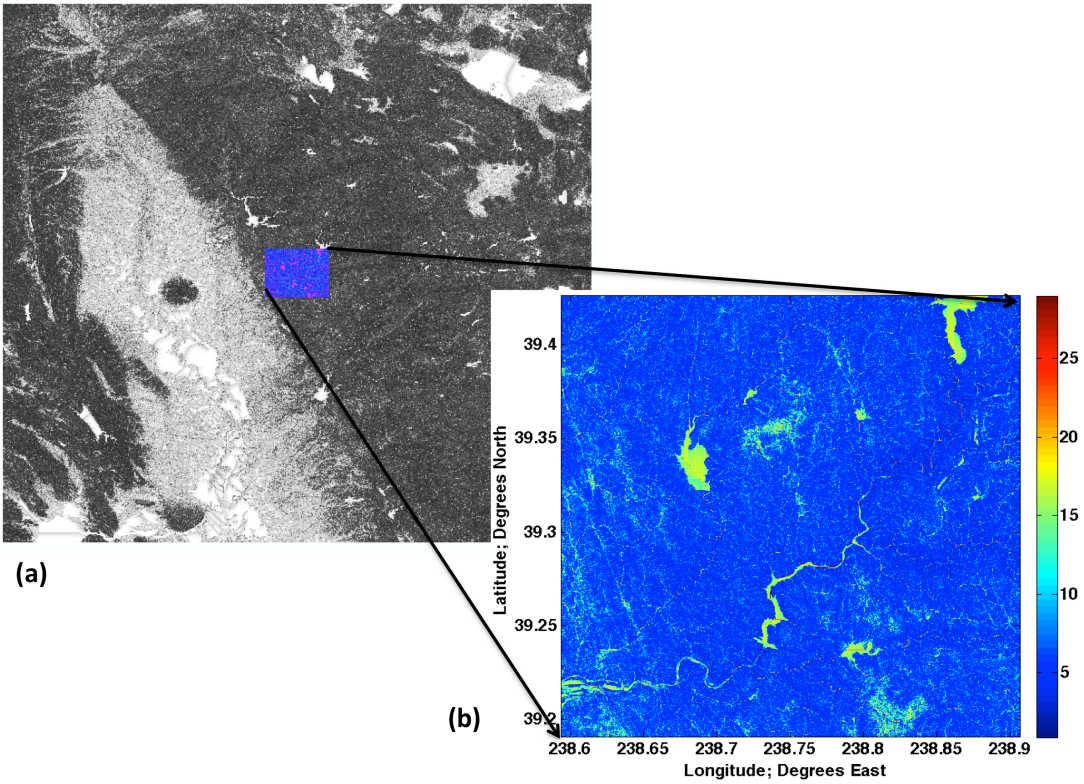

Figure B3: Topographic Index calculated at 10m resolution (a) Topographic Index over Northern California Central Valley, (b) Topographic Index at the region over which the 100m resolution run was done in Chapter 2

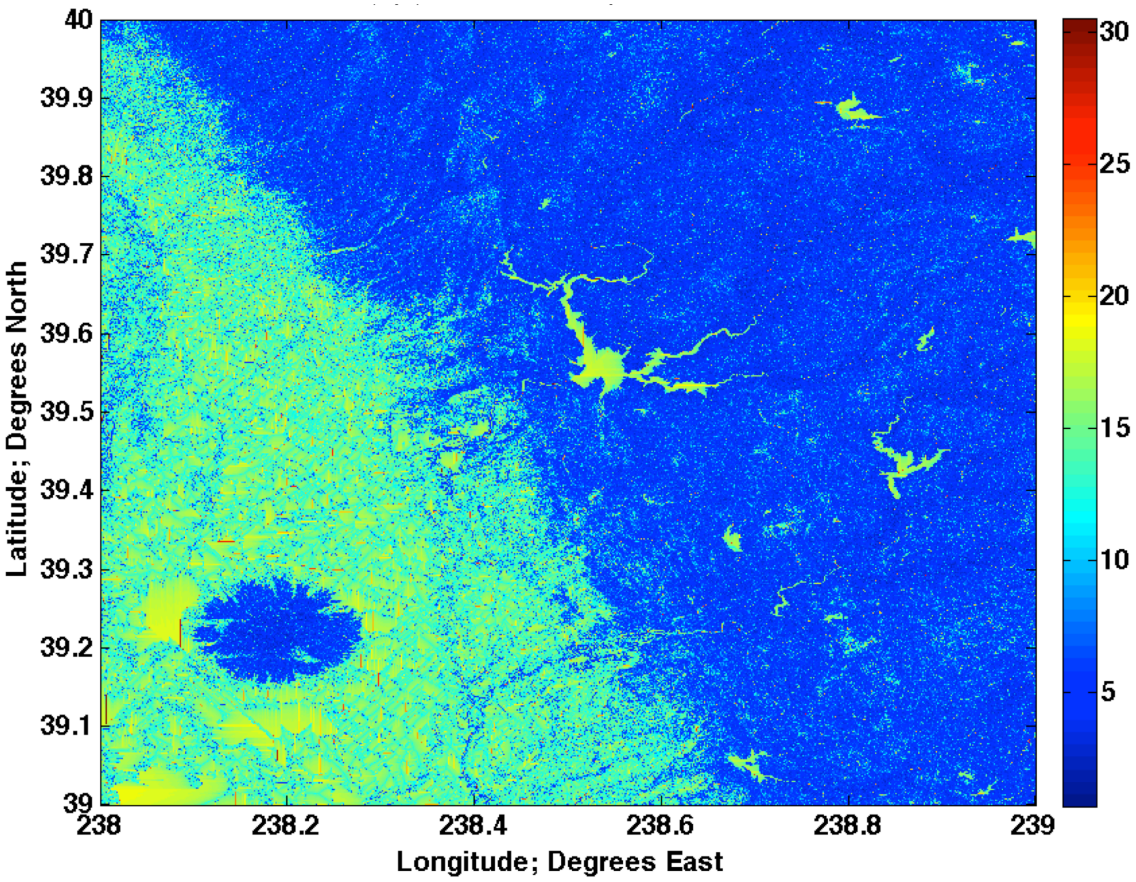

Figure B3: Topographic Index calculated at 10m resolution (a) Topographic Index over Northern California Central Valley, (b) Topographic Index at the region over which the 100m resolution run was done in Chapter 2

Appendix B: Running CLM at 30m resolution

As we showed in Chapter 2, processes like infiltration and runoff are modeled much better at 100 meters or sub 100m resolution. The goal of hyper resolution model is to reach hillslope scale that might be <100m at certain regions. We tested our model setup to run at 30m resolution by creating input files at that resolution. The aim of this study was also to check if the model still performed similarly and the feasibility of this resolution. The TI was calculated at 10m thus it provided f_{max} values at 30m resolution, the soil texture data at this resolution is not available and thus we used the 1km resolution data used earlier. All other input variables were at same resolution as in Chapter 2 and 3.

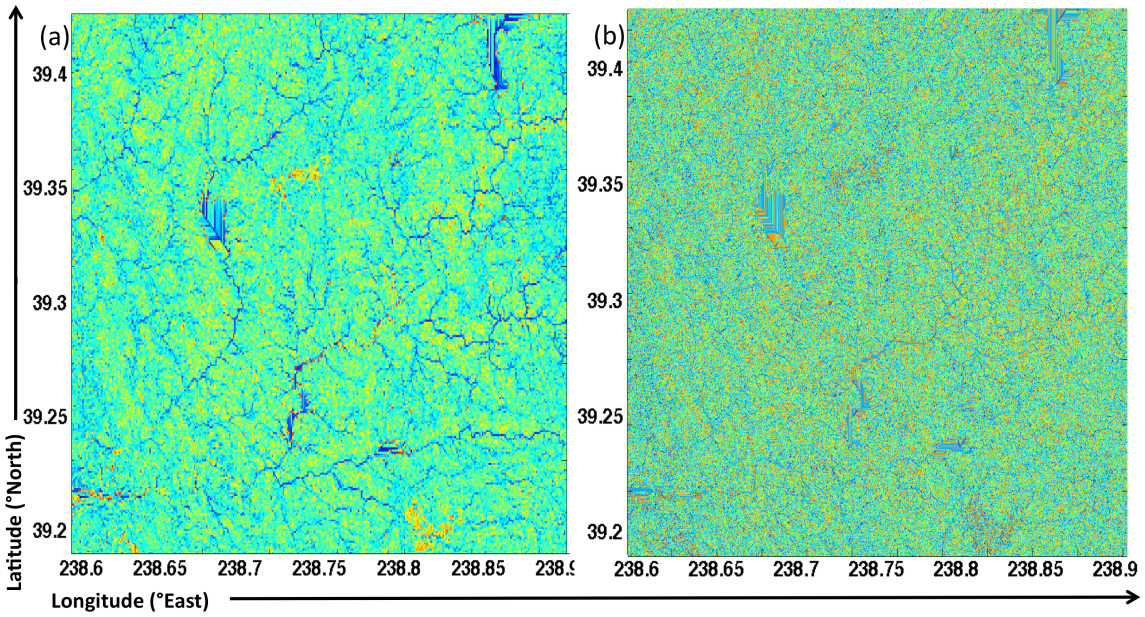

Figure B1: Maximum saturated fraction (f_{max}) data at the region over which the 100m resolution run was done in Chapter 2 (a) at 100m resolution (b) at 30m resolution

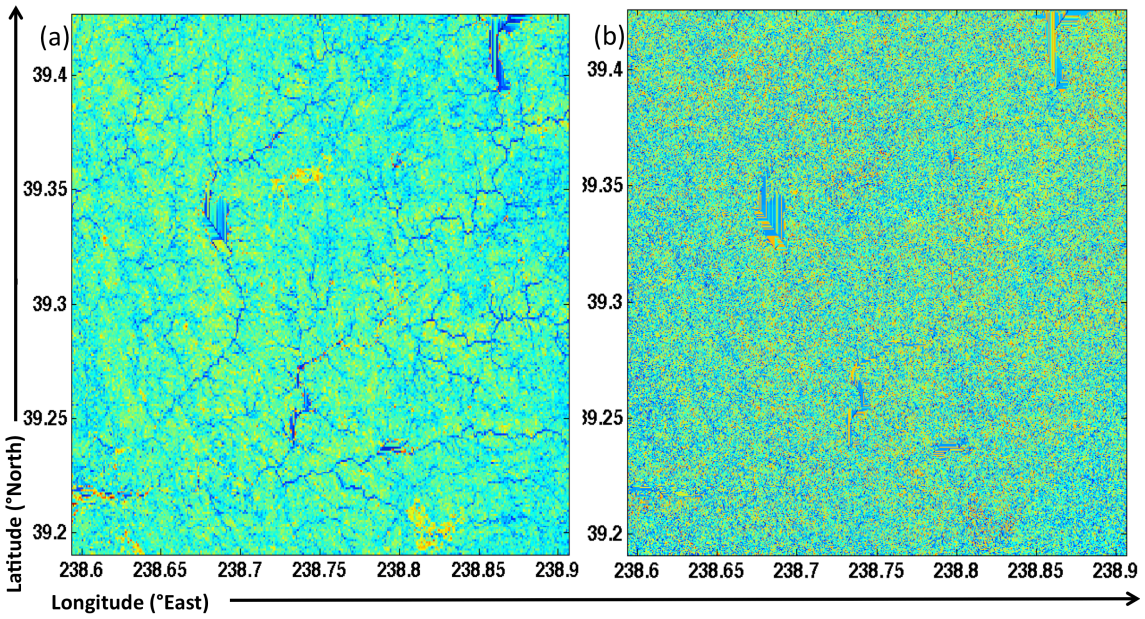

Figure B2: Fractional saturated/impermeable area (f_{sat}) data at the region over which the 100m resolution run was done in Chapter 2 (a) at 100m resolution (b) at 30m resolution