MERRA Analytic Services

John Schnase, Dan Duffy, Glenn Tamkin,

Savannah Strong, Jian Li, Mark Carroll, and Roger Gill

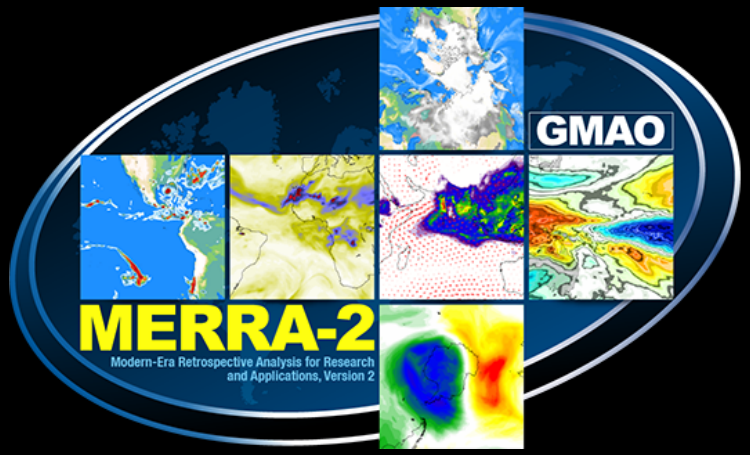
Office of Computational and Information Sciences and Technology

NASA Goddard Space Flight Center

MERRA Analytic Services

The MERRA/AS Project is combining high-performance computing with analytic techniques to better serve MERRA customers by focusing on data reduction and the "Big Data" challenge of upfront data assembly …

- Data Reduction – Correcting, ordering, and simplifying data in support of analytic objectives
- Data Preparation – Preparing heterogeneous climate model outputs so that they can be jointly analyzed
- Data Analysis and Visualization – Applying techniques to discover patterns, correlations, and other higher-order insights of value to a growing community of users

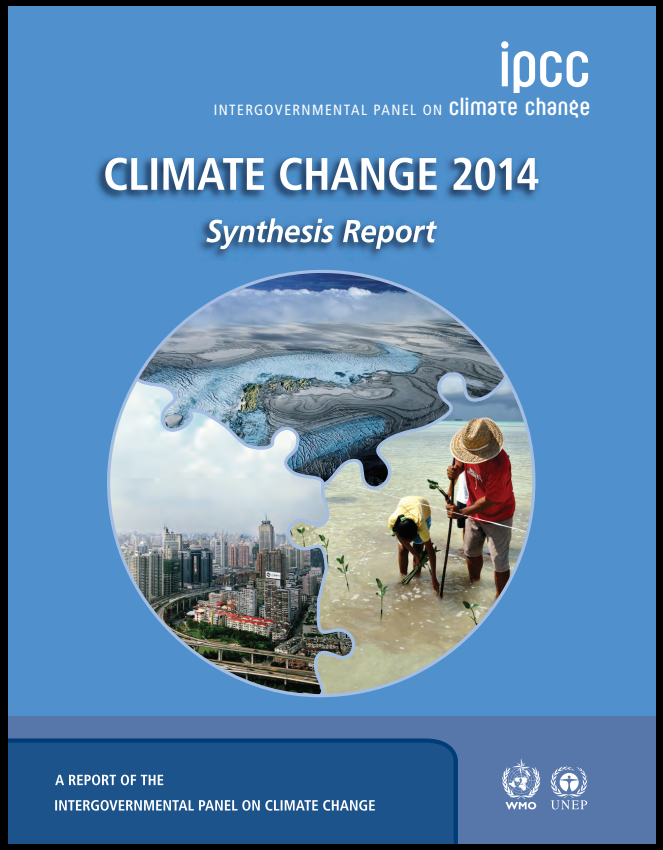


It is estimated that climate scientists spend 50–80% of their time gathering and preparing data for further study (Lohr, 2014)

IPCC-Inspired Design

The Intergovernmental Panel on Climate Change (IPCC) is the leading international body for the assessment of climate change

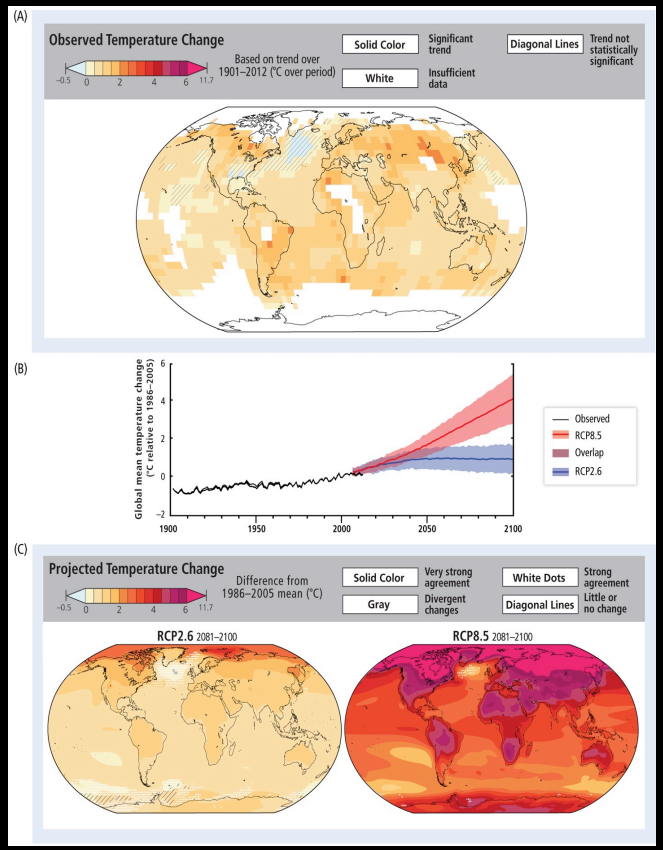
- Open to all members of the United Nations and World Meteorological Organization
- Currently 195 member countries
- Provides world with a clear view of the current state of scientific knowledge through its Assessment Reports
- 2000+ scientists contributed to IPCC's Fifth Assessment Report (AR5)
- IPCC Assessment Reports provide the basis for environmental policy-making throughout the world …



IPCC-Inspired Design

A small collection of data attributes provides the basis for a remarkable amount of intellectual work in the discipline …

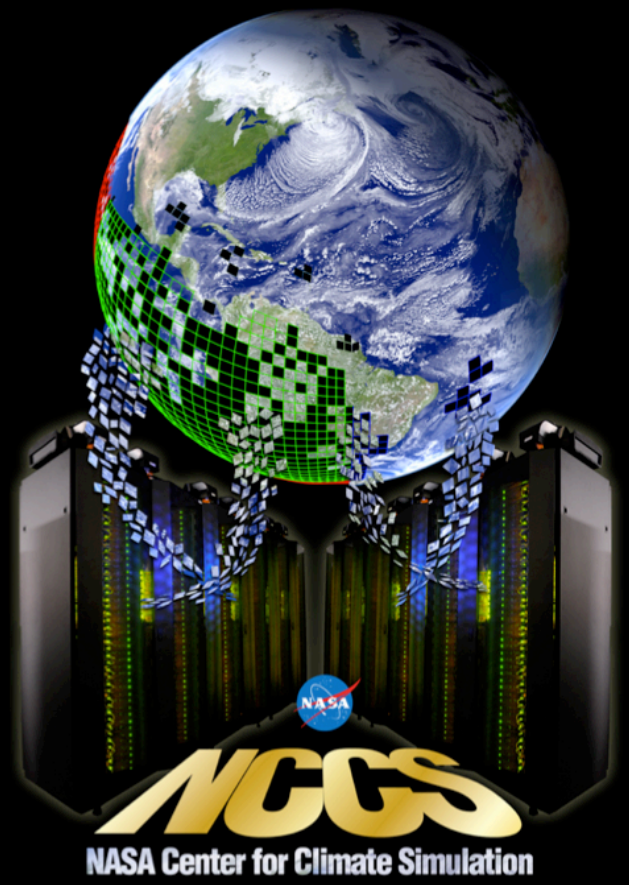
- AR5's published work:
  - Contains 4 million words (50,000 unique)
  - Eight words collectively account for 12,000 direct references to climate model data: *maximum, minimum, average, variance, difference, climatology, anomaly, and trend*
  - Over halve of those references are "trend" …



MERRA Analytic Services

So, how do we use analytic technologies to better serve the MERRA research and applications communities?

- MERRA/AS creates real and virtual collections of the most commonly used products
  => maximum, minimum, average, variance, difference, climatology, anomaly, and trend
- MERRA/AS delivers those products in a highly personalized and tailored form
  > input=variable name, spatial extent, temporal extent
- MERRA/AS computes those products as fast as possible
  => high-performance compute-storage, MapReduce
- MERRA/AS delivers those products through a Web service interface and Python library



Single Reanalysis Estimation of the Contribution of Irrigation to Precipitation

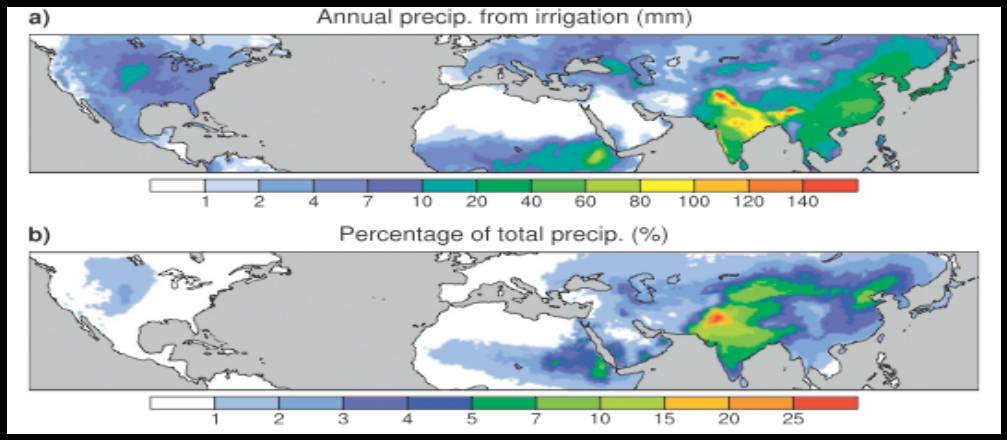
J. Wei tailored climatologies use case …

Study Areas
- Nile Valley
- North China
- California Central Valley
- Northern India/Pakistan

Other Requirements
- 1979 – 2002
- 6-hr time steps
- 18 atmospheric levels

Variables Needed
- Humidity
- Wind speed
- Temperature

Data Wrangled:
- 23 x 365 x 4 x 4 x 18 x 3
  = 7,253,280 "layers" …



Where Does the Irrigation Water Go? An Estimate of the Contribution of Irrigation to Precipitation Using MERRA

JIANGFENG WEI*
Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland
PAUL A. DIRMEYER
Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, Virginia, and Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland
DOMINIK WISSER
Department of Physical Geography, Utrecht University, Utrecht, Netherlands
MICHAEL G. BOSILOVICH
Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland
DAVID M. MOCKO
SAIC and Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 24 May 2012, in final form 21 September 2012)

ABSTRACT

Irrigation is an important human activity that may impact local and regional climate, but current climate model simulations and data assimilation systems generally do not explicitly include it. The European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) shows more irrigation signal in surface evapotranspiration (ET) than the Modern-Era Retrospective Analysis for Research and Applications (MERRA) because ERA-Interim adjusts soil moisture according to the observed surface temperature and humidity while MERRA has no explicit consideration of irrigation at the surface. But, when compared with the results from a hydrological model with detailed considerations of agriculture, the ET from both reanalyses show large deficiencies in capturing the impact of irrigation. Here, a back-trajectory method is used to estimate the contribution of irrigation to precipitation over local and surrounding regions, using MERRA with observation-based corrections and added irrigation-caused ET increase from the hydrological model. Results show substantial contributions of irrigation to precipitation over heavily irrigated regions in Asia, but the precipitation increase is much less than the ET increase over most areas, indicating that irrigation could lead to water deficits over these regions. For the same increase in ET, precipitation increases are larger over wetter areas where convection is more easily triggered, but the percentage increase in precipitation is similar for different areas. There are substantial regional differences in the patterns of irrigation impact, but, for all the studied regions, the highest percentage contribution to precipitation is over local land.

* Current affiliation: Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas.

Corresponding author address: Jiangfeng Wei, Jackson School of Geosciences, The University of Texas at Austin, 2275 Speedway C9000, Austin, TX 78712.
E-mail: jwei@utexas.edu

DOI: 10.1175/JHM-D-12-079.1

1. Introduction

Irrigation is an important human activity that has the potential to impact local and regional climate through the hydrological cycle and surface energy balance (e.g., Chase et al. 1999; Pielke et al. 2011). About two-thirds of the global freshwater withdrawals from surface and underground are used for agriculture (Shiklomanov 2000),

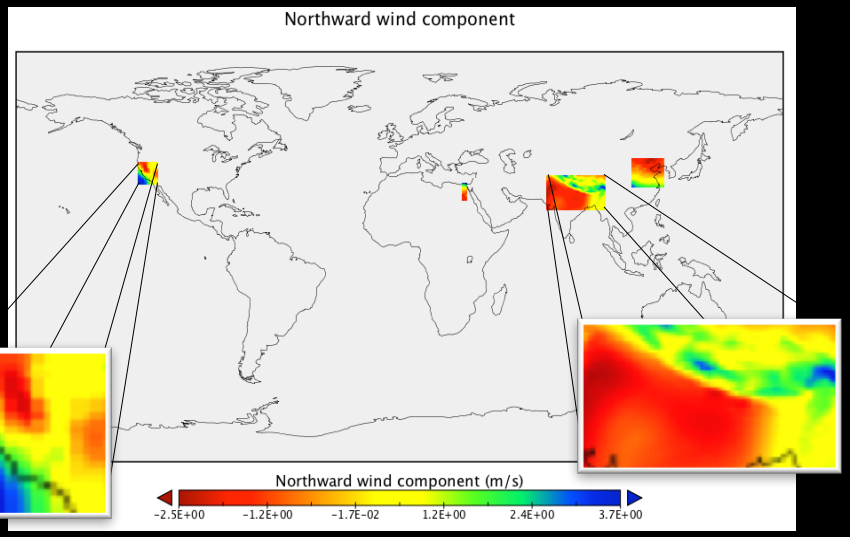
Single Reanalysis Estimation of the Contribution of Irrigation to Precipitation

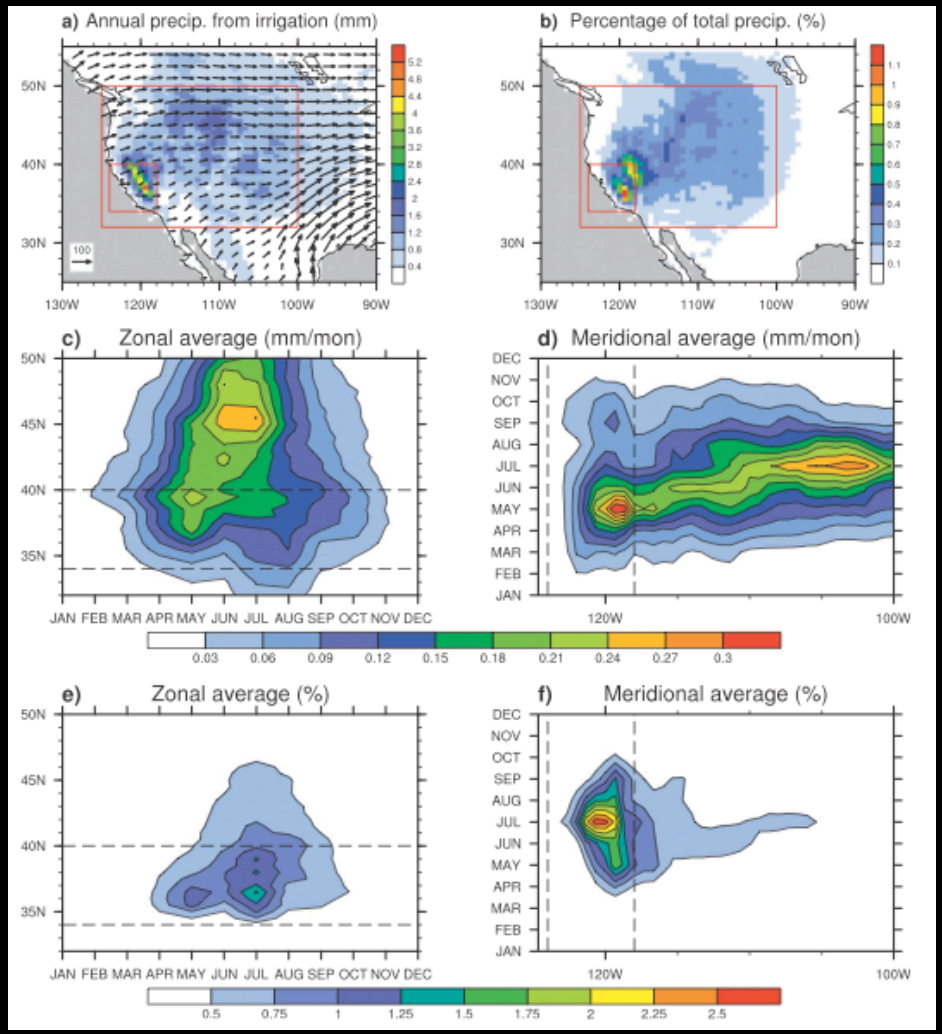
Traditional Approach (hours, days)

- Step 1 – 8.4 TB moved from archive
- Step 2 – Clipping / averaging

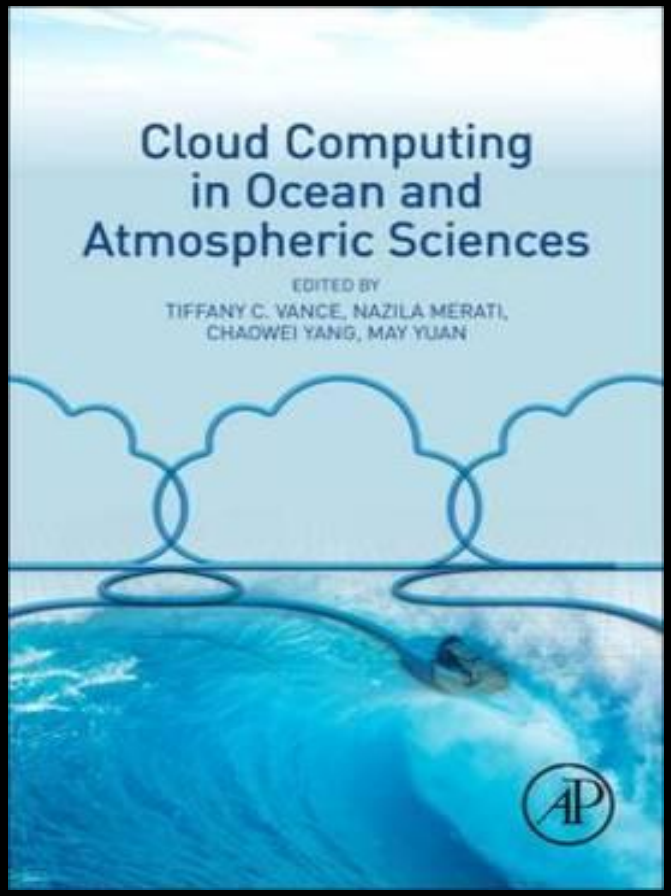
With MERRA/AS and CDSlib (2.5 minutes)

- Step 1 – Clipping / averaging
- Step 2 – 500 MB of final product moved to local workstation in minutes …





For more information …



Schnase, J.L., 2016. Climate Analytics as a Service. In: Vance, T.C., Merati, N., Yang, C., Yuan, M. (Eds.), *Cloud Computing in Ocean and Atmospheric Sciences*. Academic Press, pp. 187-219. ISBN: 9780128031926, http://dx.doi.org/10.1016/B978-0-12-803192-6.00011-6

CHAPTER 11

Climate Analytics as a Service

J.L. Schnase
NASA Goddard Space Flight Center, Greenbelt, MD, USA

INTRODUCTION

Cloud technologies provide an unprecedented opportunity to expand the power and influence of computing in daily life. So far, those opportunities have unfolded in largely ad hoc ways, resulting in a creative chaos that at times can be confusing. Over the years, we have become comfortable with a classic von Neumann perspective on what constitutes a computer. We share mental models and patterns of thinking about how computers are built and how they behave—what in broad terms computing technologies can do, how they do it, and what they cannot do. But when it comes to cloud computing, those patterns have not been established—save one: the concept of *service*. We have developed a shared notion that cloud technologies in an essential way provide the basis for services. By definition, cloud capabilities reside there, not here—the action of helping is conveyed to the user: the user is served. Hence, terms such as Software-as-a-Service and Platform-as-a-Service have become common parlance in the world of cloud computing.

In our efforts to deal with the big data problems of climate science, we are trying to take a deeper dive into our understanding of cloud-computing services. To begin, we ask the fundamental question: What is it that needs to be served? Our answer is *analytics*. But analytics served in a particular way. For now at least, we believe it would be productive to focus on the basics—do simple things well and very fast. We need to garner the agile high-performance computing and storage resources of cloud computing to address climate science's big data problems in a new way—in a way that melds knowledge creation with data creation, curation, discovery, and workflow orchestration. This chapter is an effort to advance the cause.

Our story begins with the observation that big data challenges are generally approached in one of two ways. Sometimes they are viewed as a problem of large-scale data management, in which solutions are offered through an array of traditional storage and database theories and

Cloud Computing in Ocean and Atmospheric Sciences
ISBN 978-0-12-803192-6
http://dx.doi.org/10.1016/B978-0-12-803192-6.00011-6

Cloud Computing in Ocean and Atmospheric Sciences, First Edition, 2016, 187-219