# Application for LLNL/CASC/AIMS Computer Scientist Position

**Research Experience, Background, and Interests and Pertinence to ESGF and UV-CDAT Projects**

**Mark Yashar**
**http://www.linkedin.com/in/markyashar**

**April 8, 2015**

# Outline

- Broad overview of research interests and relevant scientific/ technical skills, including academic coursework
- Examples of research experience and background and **pertinence to ESGF & UV-CDAT projects at LLNL/AIMS in each case**
  - Meteorological and $CO_2$ Regional Modeling
  - Square Kilometer Array (SKA) research & development
  - Dark Energy Research
  - SLAC data handling and archive maintenance
- Summary

# Overview of Research Interests and Scientific/Technical skills

Interested in utilization and development of modeling, data analysis, statistical, mining, reduction, and processing algorithms, software, methods, and techniques in a wide range of possible physical science (e.g.,earth sciences, physics), scientific computing, and engineering disciplines. Examples:

- Monte Carlo methods and techniques (e.g., for parameter and error estimation)
  - Markov Chain Monte Carlo (MCMC) (including Bayesian analysis) and Metropolis Hastings algorithms
- Data and image handling, management, reduction, processing and analysis
- Data visualization and associated software tools

# Overview of Research Interests and Scientific/Technical Skills

- Statistics and error analysis (e.g., propagation of errors, Gaussian distributions, covariance, and correlation)
- Modeling, simulation and model fitting (i.e., **optimization**)
  - Chi-square fitting
  - Fitting data to a straight line
  - General linear least squares
  - Nonlinear models
  - Confidence limits on estimated model parameters
- Scientific/technical writing

# Examples of Research and Technical Experience and Interests

**Meteorological & $CO_2$ Regional Modeling (Supv.: I. Fung, UCB)**

- Carried out research focused on mesoscale and regional atmospheric transport modeling and analysis of anthropogenic and biogenic $CO_2$ emissions for northern California for mutli-scale estimation and quantification of atmospheric $CO_2$ concentrations

# Examples of Research Experience and Interests

**Meteorological & $CO_2$ Regional Modeling (Supv.: I. Fung,UCB)**

- Made extensive use of **WRF** (written mostly in **FORTRAN 90**), **WRF-Chem** coupled weather-air quality model for atmospheric transport simulations, and **WRF-VPRM** biospheric model to simulate $CO_2$ biosphere fluxes and atmospheric $CO_2$ concentrations.
  - Involved use of **R statistical scripting language**, **NCL**, MATLAB, **Python (numpy, netCDF4)**, **Ferret**, and **NCO tools** for additional pre- and post-processing, modification, visualization and analysis of **netCDF** files and output **diagnostic variables** of simulation results (e.g., pressure, temperature, relative humidity)

# Examples of Research Experience and Interests

**Meteorological & $CO_2$ Regional Modeling (Supv.: I. Fung, UCB)**

- Troubleshooted and debugged **WRF**, **WRF-Chem**, and **VPRM** simulation runs and results to increase data efficiency and to decrease time to solve problems
- Installed, compiled, built, configured, and ran WRF, WRF-Chem, and VPRM on **NERSC multi-core supercomputing system** ("Hopper"; Cray XT CLE/Linux x86_64, PGI compiler w/ gcc, distributed memory parallel mode, **MPI**) and submitted batch job scripts to this system to run code in parallel on compute nodes via **MPI** tasks
- Assisted students and post-docs in installing, configuring, and running WRF and WRF-Chem software on "Hopper"

# Examples of Research Experience and Interests

**Meteorological & $CO_2$ Regional Modeling: WRF preprocessing, post-processing, analysis, and visualization**

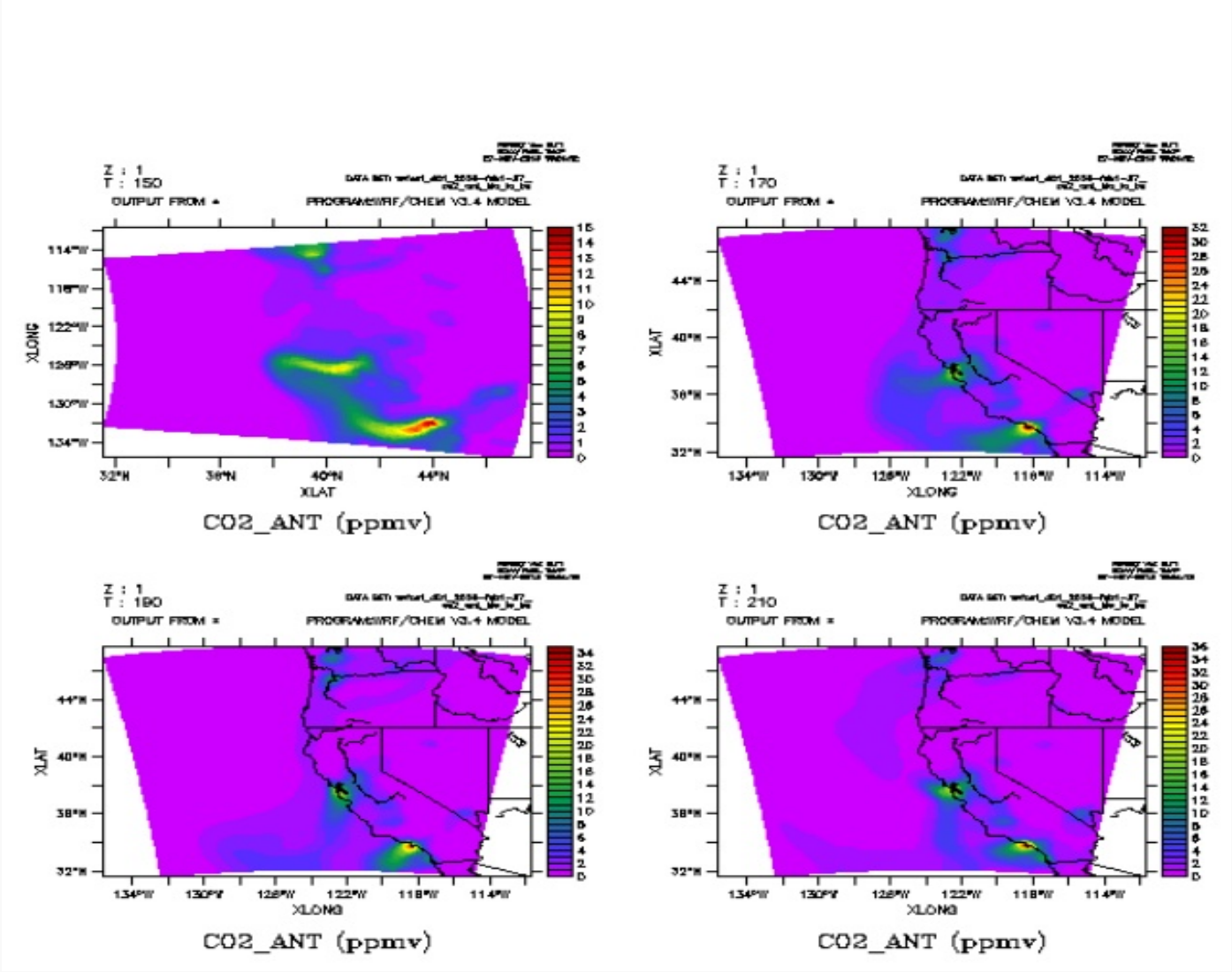
- WRF Preprocessing
  - Modified **FORTRAN 90** code to convert CO2 emissions inventory data files between **HDF5**, binary, and **NetCDF-4** formats to enable and expedite further processing
  - Modified and utilized R scripts to edit **NetCDF-4** files to be in correct format and containing necessary CO2 emission fields and data to be input to WRF simulation runs

# Examples of Research and Technical Experience and Interests

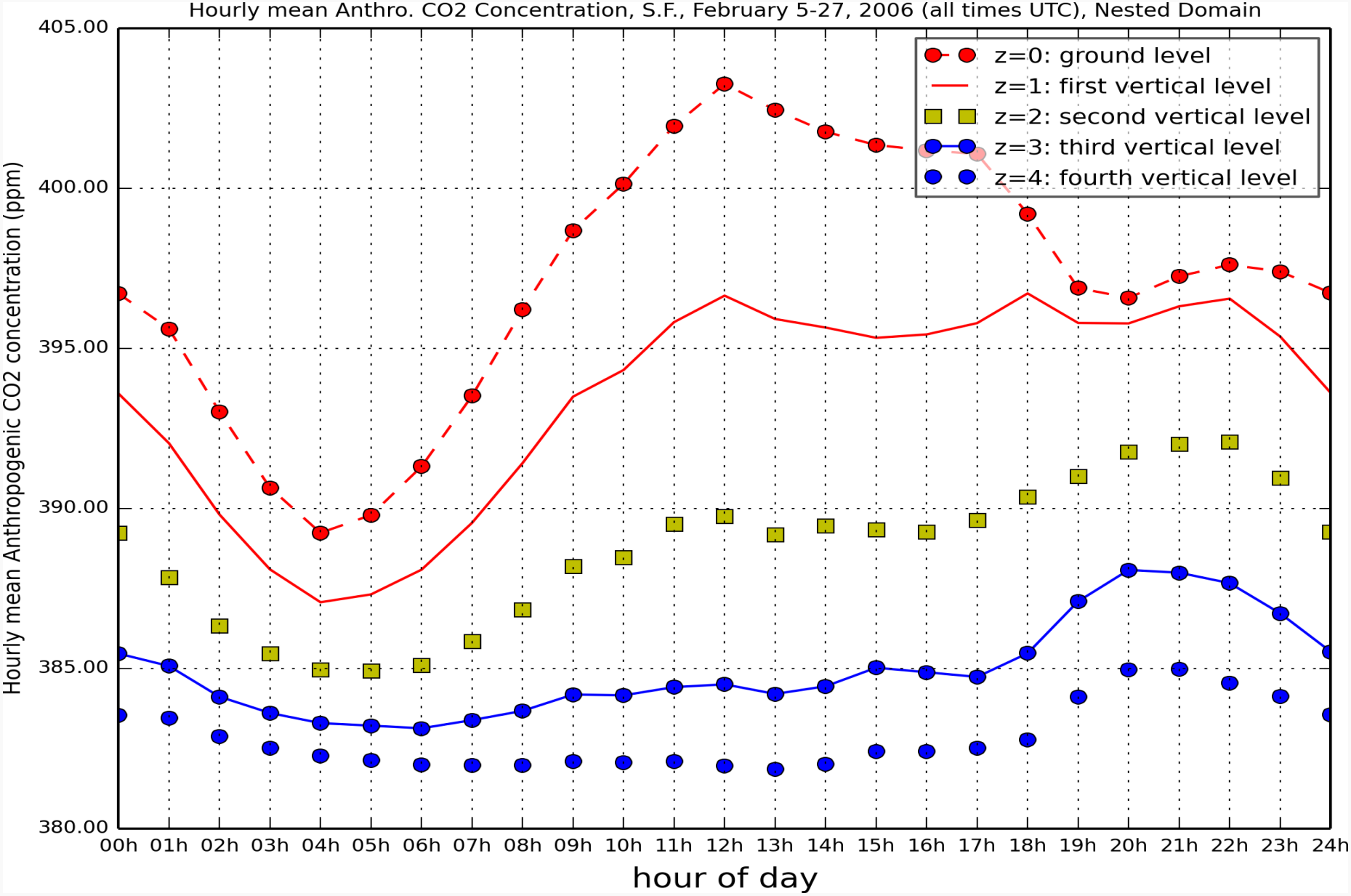
**Meteorological & $CO_2$ Regional Modeling: WRF preprocessing, post-processing, analysis, and visualization**

- WRF visualization and analysis
  - Wrote and utilized **NCL, Ferret, and Python (numpy)** scripts and NCO package (e.g., 'ncdump', 'ncview') to visualize and analyze WRF model output **diagnostic variables** (e.g., CO2 concentrations, pressure, temperature, wind vectors) in form of contour maps, time series plots, and animations
  - Used **NCL, NCO, Ferret, and Python** to compute statistics of model variables (e.g., means)

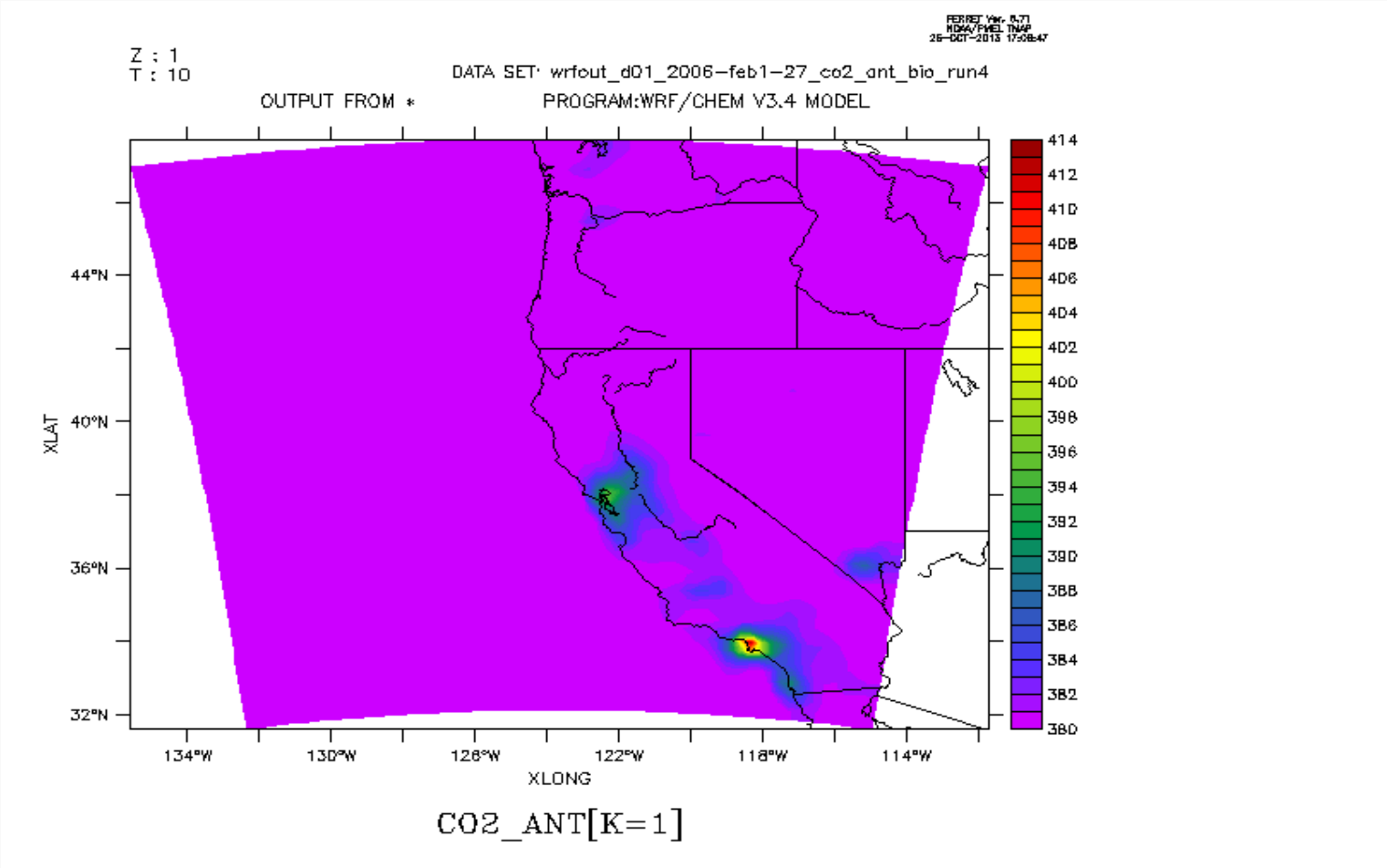
# WRF Simulation Results: Contour Maps



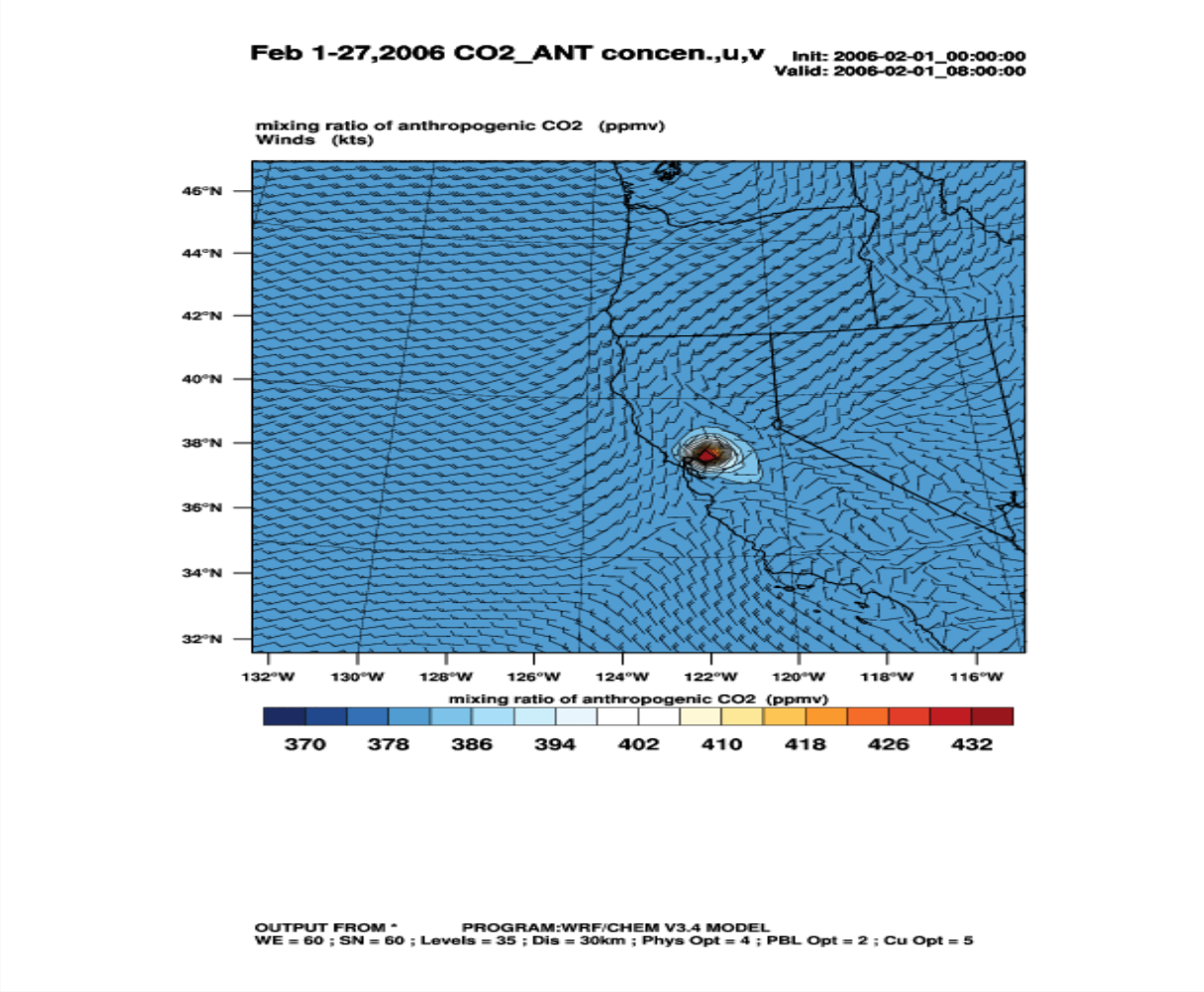
# WRF Simulation Results: Time Series Plots



# WRF Simulation Results: Animations (Ferret)



# WRF Simulation Results: Animations (NCL)



# Meteorological and $CO_2$ Regional Modeling

**Pertinence to ESGF and UV-CDAT Projects**

- WRF meteorological/atmospheric modeling
  - Helpful in understanding **CMIP5,6** framework and data
- **NCL, Ferret, NCO, R, Python** visualization and analysis of WRF model output variables/**diagnostics**
  - User perspective of visualization and analysis tools
  - Assist and contribute in improving and developing UV-CDAT & ESGF data analysis, diagnostic, and visualization tools (e.g., **ESGF toolbox**)
- **NetCDF** and **HDF5** processing and editing/writing
  - Knowledge and experience with NetCDF file format and structure
  - Contribute to development of **CMOR2**

# Meteorological and $CO_2$ Regional Modeling

**Pertinence to ESGF and UV-CDAT Projects**

- Experience installing, configuring, testing, maintaining and running the WRF model on **HPC** platform
  - Help to incorporate **parallel capabilities** for climatology generation and batch processing
  - Add parallel support in UV-CDAT analysis and visualization frameworks where required
  - Implementing algorithms toward analysis of large volumes of data using **Hadoop MapReduce**

# Examples of Research Experience and Interests

**SKA Research and Development (Supv.: A. Kemball, UIUC)**

- R&D in SKA calibration and processing algorithms and computing with a focus on cost and feasibility studies of radio imaging algorithms and direction-dependent calibration errors
  - Evaluated the computational costs of non-deconvolved images of a number of existing radio interferometry algorithms used to deal with non-coplanar baselines in wide field radio interferometry and co-authored a corresponding internal technical report with A. Kemball

# Examples of Research and Technical Experience and Interests

## SKA Research and Development (Supv.: A. Kemball, UIUC)

- **M. Yashar**, A. Kemball, Computational Costs of Radio Imaging Algorithms Dealing with the Non-coplanar Baselines Effect:I, TDP Calibration and Processing Group Memo #3

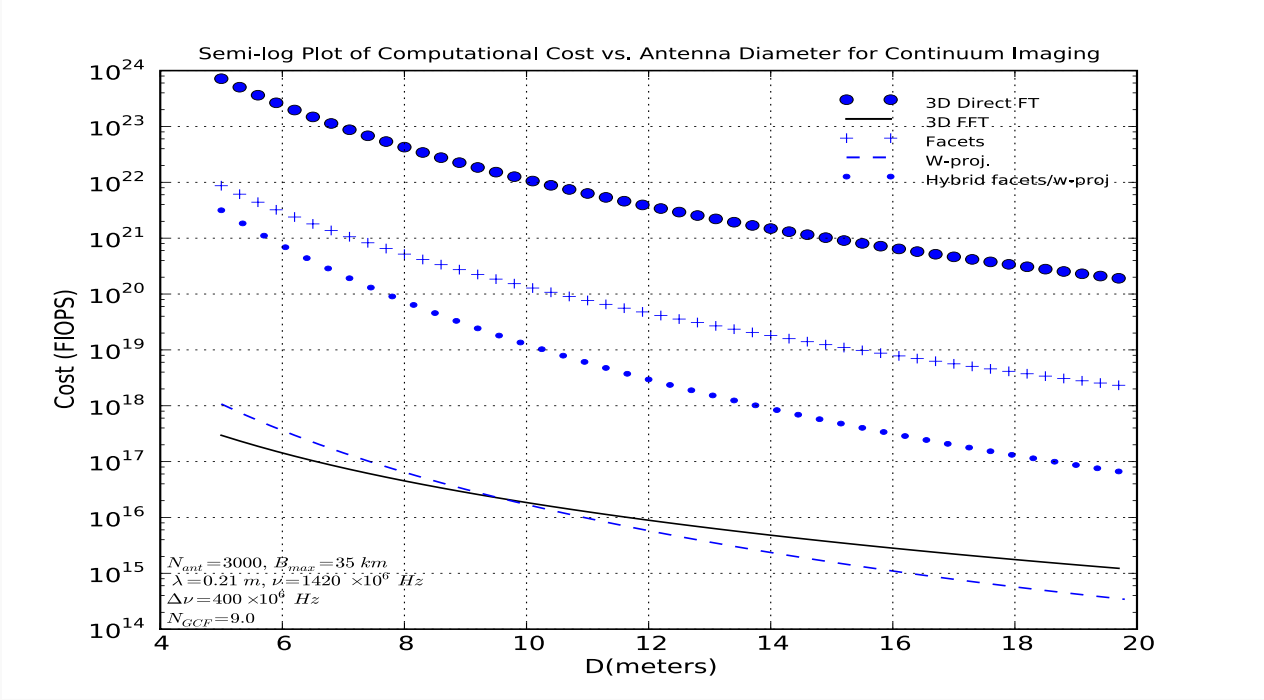



Figure 1: Semi-log y plots of computational costs (without consideration of deconvolution and parallel computing efficiency $\eta$) vs. antenna diameter $D$ for continuum imaging for the 3-D direct FT, 3-D FFT, facets, w-projection, and hybrid facets/w-projection imaging algorithms.

# Examples of Research and Technical Experience and Interests

**SKA Research and Development (Supv.: A. Kemball, UIUC)**

- Implemented numerical and imaging simulations (Meqtrees, CASA – **Python,C++**) and Monte Carlo simulations (**Python--numpy**) to address cost, feasibility, dynamic range, and image fidelity issues related to calibration and processing for SKA and dependence of these issues on key antenna and feed design parameters (e.g., sidelobe level, mount type). Co-authored corresponding technical memo: A. Kemball, T. Cornwell, **M. Yashar**, Calibration and Processing Constraints on Antenna and Feed Designs for SKA: I, TDP Calibration and Processing Group Memo #4
  - Installed, built, compiled, configured **C++ software development (OOP)** environment for CASA (gdb, ddd, Eclipse debuggers, GNU tools) and made modifications to C++ code to carry out simulations and test imaging algorithms

# Examples of Research and Technical Experience and Interests

**SKA Research and Development: Miscellaneous Computing/ Software Projects**

- Installed, configured, and maintained multiple 64-bit Vmware (Workstation 7.1) **virtual machines (VM)** on host Linux CentOS 5.x system using ISO disk images. Various guest OS (e.g., CentOS and Ubuntu) were installed on newly created VMs and VM root partition was expanded.
- Assisted in installing, configuring, and maintaining **Java-based** Liferay Portal software with Apache Tomcat Server Bundle and **MySQL Database** on a Linux machine to facilitate communication between team members, upload and maintain library of internal technical documents, etc.

# SKA Research and Development

**Relevance to ESGF and UV-CDAT Work**

- SKA R&D **Python (numpy, scipy, matplotlib)** data and statistical analysis, visualization (combined with academic background in data analysis, etc.)
  - ➢ Contribute to ESGF/UV-CDAT **data mining** and **machine learning** work (e.g., Python scikit-learn)
- Experience working in **C++ software development** environment and with OOP
  - ➢ Assist in development of ESGF node software stack and UV-CDAT software development (e.g., CDMS development, developing and testing integrated tools)
- Knowledge and experience installing, configuring, utilizing, and maintaining VMs
  - ➢ Contribute to ESGF cloud computing deployment projects

# Examples of Research Experience and Interests

**Dark Energy Research (Supv.: A. Albrecht, UCD)**

- MCMC analysis (involving extensive use of MATLAB code) of dark energy quintessence model (IPL or Ratra-Peebles) that included utilization of DETF data models that simulated current and future data sets from new and proposed observational programs
- Wrote and submitted batch job scripts to run MATLAB MCMC code on Linux computing cluster to expedite running of MCMC simulations and generation of MCMC output
- Troubleshooted and debugged MATLAB code, simulation runs and results

# Examples of Research Experience and Interests

## Dark Energy Research (Supv.: A. Albrecht, UCD)

- Lead author of paper published in Physical Review D on research results: **M. Yashar**, B. Bozek, A. Albrecht, A. Abrahamse, M. Barnard, Exploring Parameter Constraints on Quintessential Dark Energy: The Inverse Power Law Model, Physical Review D, 79, 103004, 2009.
  - From the associated likelihood contours, found that the respective increase in constraining power with higher quality data sets produced by analysis gave results that were broadly consistent with the DETF for the dark energy parameterization that they used. Also found, consistent with other findings, that for a universe containing dark energy described by the IPL potential, a cosmological constant can be excluded by high quality "Stage 4" experiments by well over 3 sigma.

# Examples of Research and Technical Experience and Interests

## Dark Energy Research (Supv.: A. Albrecht, UCD)

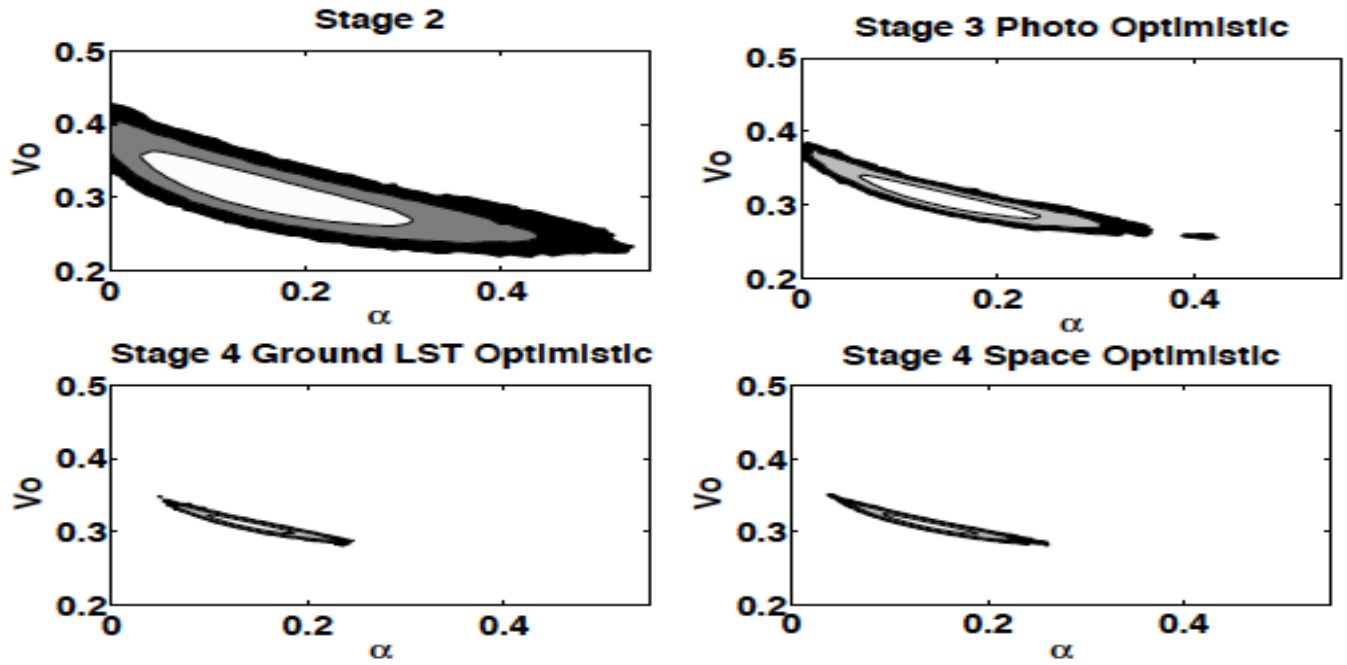



FIG. 9: $V_0 - \alpha$ $1\sigma$ (68.27%), $2\sigma$ (95.44%) and $3\sigma$ (99.73%) likelihood contours for DETF optimistic combined data sets generated from a selected IPL background cosmological model.

# Dark Energy Research

## Pertinence to ESGF and UV-CDAT Projects

- Experience and knowledge using MCMC to solve optimization problems (i.e., parameter fitting), along with academic background with other statistics and data analysis algorithms
  - ➢ Contribute to ESGF and UC-CDAT data analysis, data mining, and machine learning activities
- Research and academic background in physics and cosmology
  - ➢ May enable contribution to development potential future ESFG physics or astronomy platforms or nodes

# Examples of Research and Technical Experience and Interests

**SLAC Data Handling and Archive Maintenance (Supv.: E. Bloom, P. Kunz, SLAC)**

- Handled and processed data and maintained data archive for Unconventional Stellar Aspect X-ray astronomy experiment
  - Downloaded raw data files from NRL and processed them to create FITS files for scientist's use locally
  - Submitted batch jobs to other SLAC computing systems and clusters
  - Assisted in writing Perl and UNIX shell scripts to automate and expedite data processing, handling, and maintenance tasks (including some error checking and debugging)
  - Copied raw data files to computer tape cartridges

# SLAC Data Handling and Archive Maintenance

## Relevance to ESGF and UV-CDAT Projects

- SLAC X-ray astronomy data handling, archive maintenance, error checking and quality control of data and data processing
  - Assist with various ESGF data management, data archive maintenance, data accessibility, data storage, and quality control tasks

# Summary

- Overview of research interests and scientific/technical skills
- Examples of research experience and interests
  - Meteorological and $CO_2$ regional modeling
  - Square Kilometer Array (SKA) research & development
  - Dark Energy research
  - SLAC data analysis and archive maintenance
- Pertinence to ESGF and UV-CDAT projects.

# Thank You

- Look forward to
  - helping you solve problems and make a contribution
  - acquiring new computational and analysis skills as necessary (and come up to speed on things quickly)

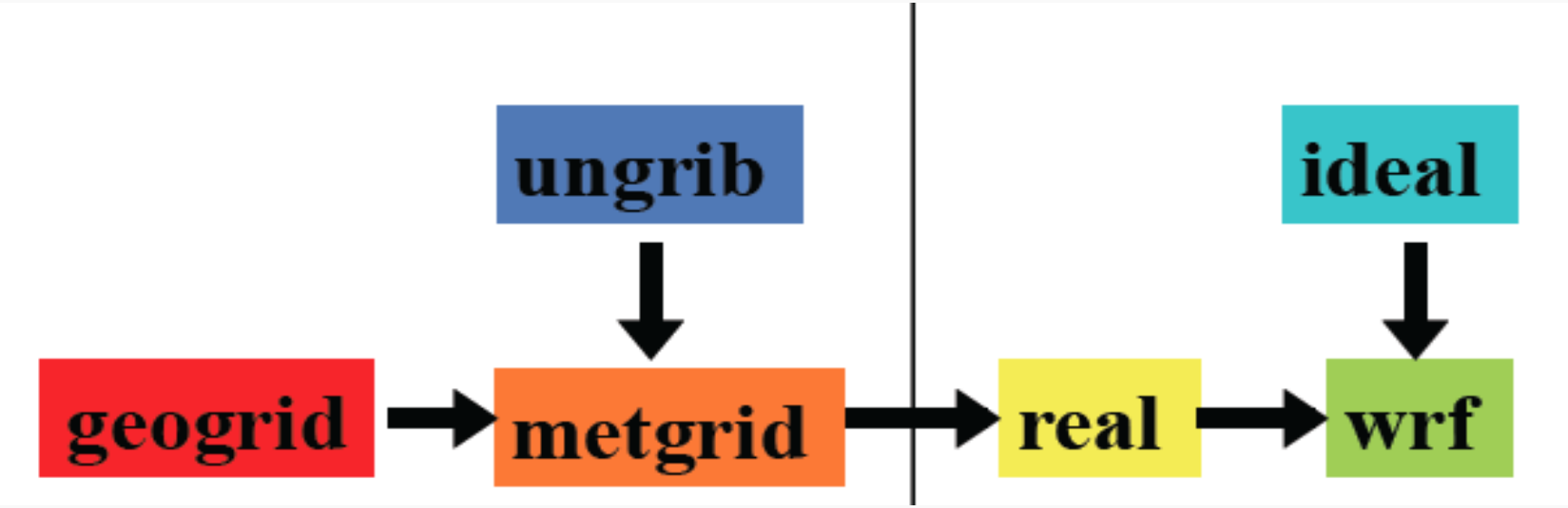
Thank you. Questions?

# Extra Slides

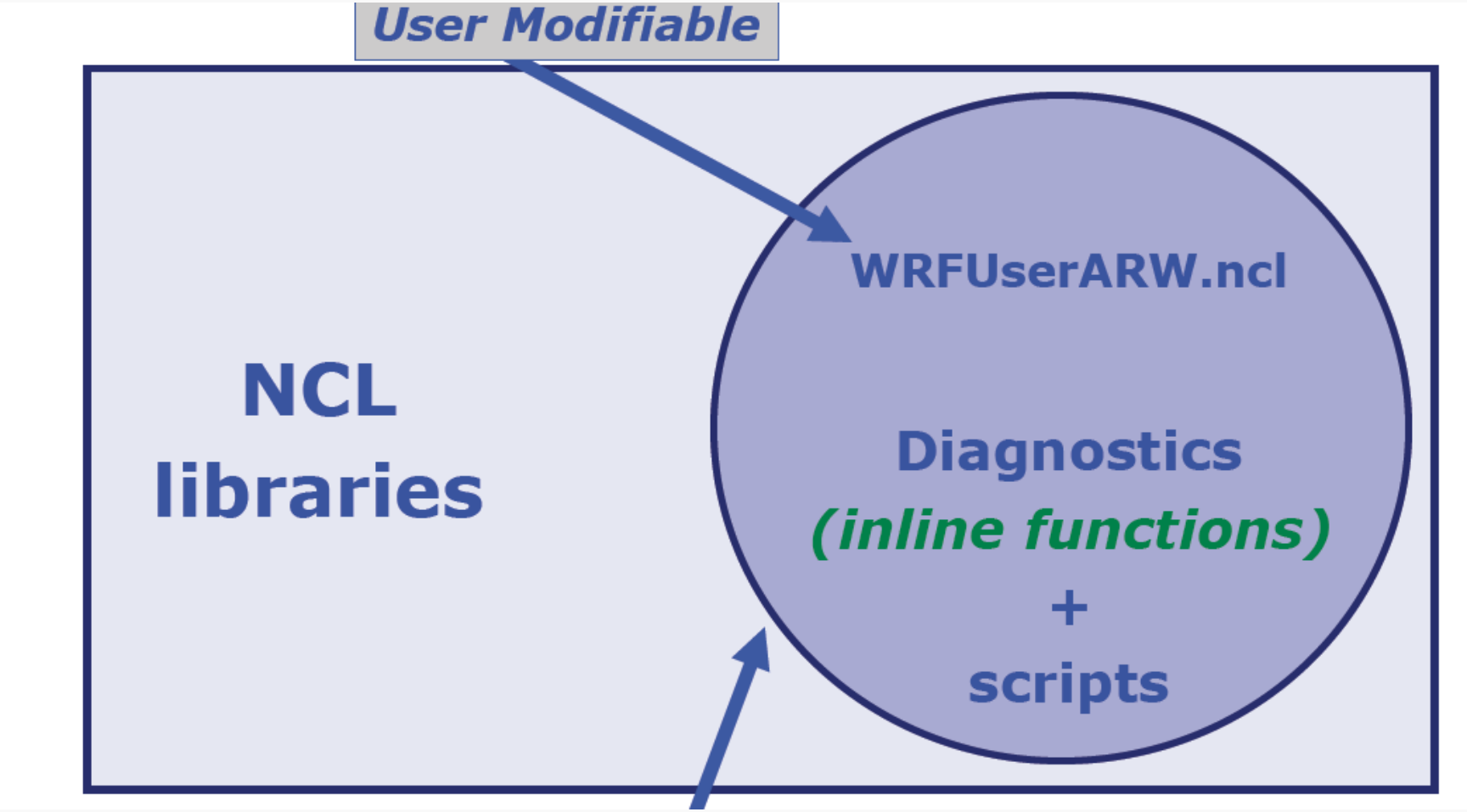
# Markov Chain Monte Carlo

- Technique often applied to solve **optimization problems** (e.g., parameter fitting) in large-dimensional spaces
  - MCMC simulates likelihood surface for set of parameters by sampling from posterior distribution via series of random draws
  - Chain steps semi-stochastically in parameter space via, e.g., **Metropolis-Hastings algorithm**, such that more probable values of parameter space are stepped to more often

# WPS and WRF Program Flow (Dudhia, 2013)



# NCL & WRF (Bruyere, 2013)



# Dark Energy Research (Additional Details)

- Carried out research project with Prof. Albrecht's research group at UCD involving MCMC analysis of dark energy quintessence model (known as Inverse Power Law (IPL) model that included utilization of Dark Energy Task Force (DETF) data models that simulated current and future data sets from new and proposed observational programs (e.g., future SNe, Ia, BAO, weak lensing, CMB observations, etc.)
  - Wrote, modified, and submitted batch job scripts to run MATLAB MCMC code on a Linux computing cluster to expedite running of MCMC simulations and generation of MCMC output

# Dark Energy Research (Additional Details)

- Generated simulated data sets for Lambda-CDM background cosmology as well as case where dark energy was provided by specific IPL model.
- Used MCMC algorithm to map likelihood around each fiducial model and moving to succession of random point in space using Metropolis-Hastings stepping algorithm.
- From associated likelihood contours, found that respective increase in constraining power with higher quality data sets produced by analysis gave results that were broadly consistent with DETF parameterization that they used.

# Dark Energy Research (Additional Details)

- Also found, consistent with other findings, that for universe containing dark energy described by IPL model, cosmological constant can be excluded by high quality "Stage 4" experiments by well over 3σ
- Troubleshooted and debugged simulation runs and results
- Lead author of paper published in Physical Review D on research results, using MC, MCMC, Metropolis-Hastings, Bayesian, and various mathematical modeling and uncertainty quantification methods/skills.

# Dark Energy Research (Additional Details)

- Assisted graduate student in generating 3-D $\chi^2$ plots with MATLAB that helped develop intuition into actual physical behavior of Albrecht-Skordis (AS) dark energy quintessence model – better understanding than would have been allowed by running full MCMC on larger parameter space
  - This systematic investigation revealed some numerical problems and issues in student's analysis of AS model.