Regional Climate Model Evaluation System

: fostering inter-disciplinary research on model evaluation against satellite observations with the help of statistical methods

Huikyo Lee

Science Data Understanding group (http://dus.jpl.nasa.gov),

Jet Propulsion Laboratory (JPL), California Institute of Technology

© 2018 California Institute of Technology. Government sponsorship acknowledged.

Acknowledgement

- My special thanks of gratitude to the Regional Climate Model Evaluation System (RCMES) team at Jet Propulsion Laboratory (JPL).

https://rcmes.jpl.nasa.gov/content/rcmes-team

Dr. Duane Waliser

Principal Investigator - Climate Science
Email: duane.e.waliser@jpl.nasa.gov

Dr. Duane Waliser is the Principal Investigator and Climate Science lead for RCMES. His principle research interests lie in climate dynamics and in global atmosphere-ocean modeling, with emphasis on prediction and predictability, and model evaluation and improvement. He joined JPL in 2004 with interests in utilizing new and emerging satellite data sets to study weather and climate as well as advance our model simulation and forecast capabilities, particularly for long-range weather and short-term climate applications. Dr. Waliser is also the Chief Scientist for the Earth Science and Technology Directorate at JPL, a Visiting Associate in the Geological and Planetary Sciences Division at Caltech and an Adjunct Professor in the Atmospheric and Oceanic Sciences Department at UCLA.

Dr. Kyo Lee

Co-Investigator - Climate and Data Science
Email: Huikyo.Lee@jpl.nasa.gov

Dr. Huikyo Lee a data scientist. Dr. Lee's skills in big data management, climate science, and statistics have allowed him to update RCMES and develop research collaborations internationally. Dr. Lee has led several RCMES workshops in international conferences (Lund, Sweden in 2014; Stockholm, Sweden in 2016; San Jose, Costa Rica in 2016), at California State University in 2016, and at the Asia-Pacific Economic Cooperation Climate Center (APCC) in 2014. Dr. Lee received JPL's voyager award in recognition of his heroic efforts to advance RCMES. Dr. Lee has a Ph. D. in Atmospheric Sciences from University of Illinois at Urbana-Champaign.

Alex Goodman

RCMES Data Scientist
Email: alexander.goodman@jpl.nasa.gov

Alex Goodman is a data scientist. With a strong background in both software development and climate science, his roles in the development of RCMES are wide and varied, but he has particularly strong interests in data visualization, distributed computing, and software package management. He received an M.S. in Atmospheric Science from Colorado State University in 2016.

Dr. Jinwon Kim

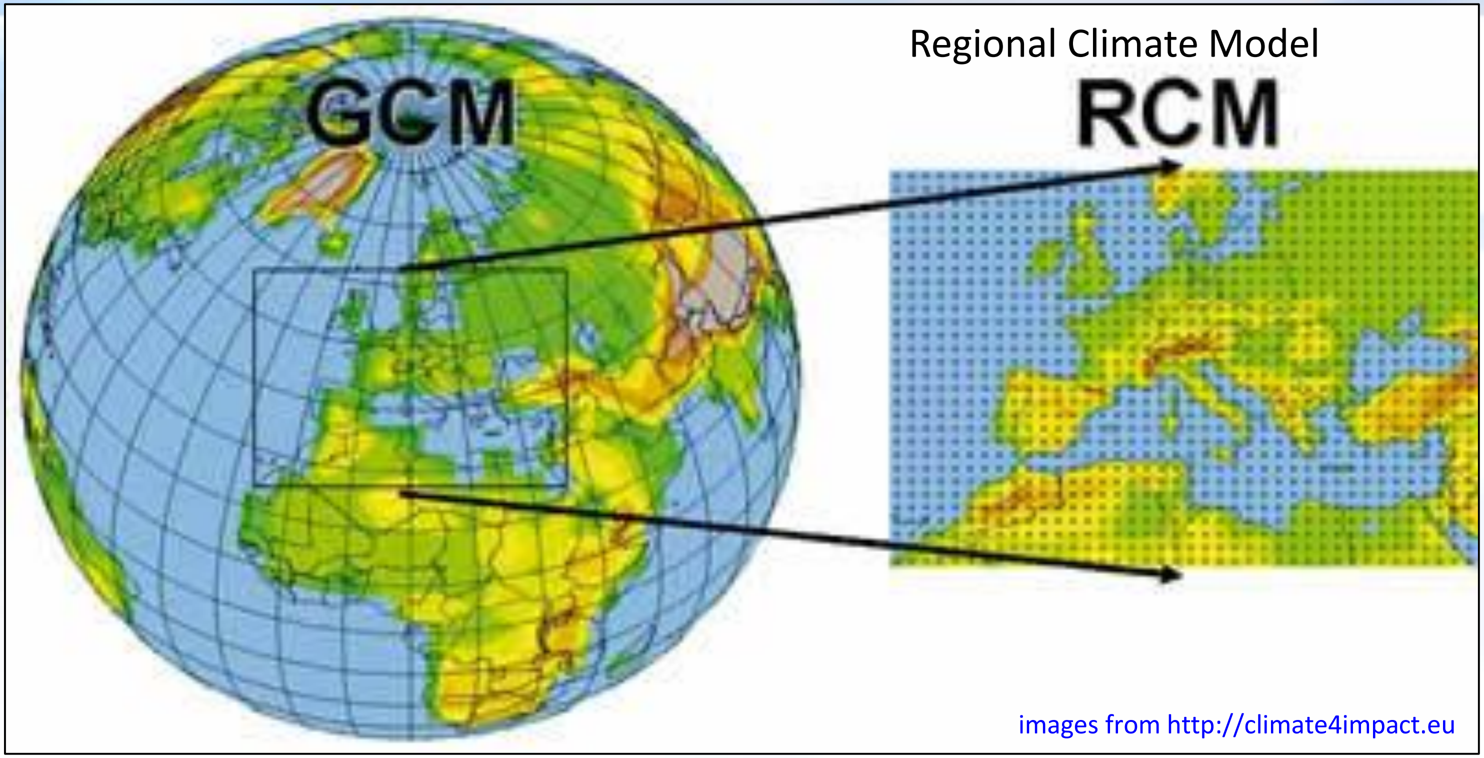
Regional Climate Scientist and Modeler
Email: jkim@atmos.ucla.edu

Dr. Jinwon Kim is a Researcher in the Joint Institute for Regional Earth System Science and Engineering (JIFRESSE) at UCLA. Dr. Kim is an expert in the area of climate modeling, including work in multi-model ensembles and regional climate comparisons.

Outline

- Introduction
  - NASA's downscaling assessment
- Systematic evaluation of CORDEX RCMs using RCMES
- Use cases of RCMES
  - Evaluation of hourly rainfall characteristics in RCM simulations
- Toward the model evaluation of the future
  - Combining multiple metrics into one: multi-objective optimization
  - Multi-resolution investigation of climate models

Dynamical Downscaling



Use a RCM **much higher spatial resolution** and possibly **improved physical process** representation over the area desired for impacts assessment with the boundary values from the GCM

courtesy of Duane Waliser and Jonathan Case

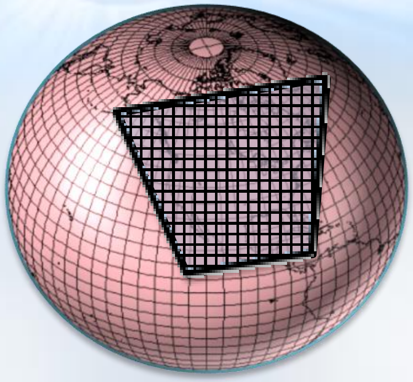
images from http://www.cordex.org/

Downscaling Assessment by Multi-NASA Centers (JPL, GSFC, MSFC, AMES)

- High resolution is thought to be crucial for the accurate representation of some processes/phenomena.
- Downscaling assessment questions:
  - Under ideal forcing conditions (e.g. high-quality re-analyses), how good are RCMs at replicating important weather and climate processes/phenomena?
  - Under what conditions does downscaling (RCMs driven by GCMs) give valid results?
  - Do high resolution RCMs (5km or finer) offer anything that can't be obtained via coarser resolution GCMs (25km or coarser)?

courtesy of Duane Waliser and Jonathan Case

Types of Simulations



I. NASA Unified-WRF (NU-WRF) with MERRA-2 six-hourly reanalysis boundary conditions

- CONUS domain, resolutions: 24 km, 12 km, 4 km
- Characterize fidelity of fine-grid RCM for important processes/phenomena.

A) Evaluate fine-grid: Score = RCM

II. NASA GEOS-5 Replay mode constrained by MERRA-2

- Global domain, resolution: 12.5 km

B) Evaluate fine-grid GCM: Score = fGCM

C) Evaluate coarse-grid GCM: Score = cGCM

D) Compare A), B) and C):

Compare RCM, fGCM and cGCM

images from http://rcmes.jpl.nasa.gov

courtesy of Duane Waliser and Jonathan Case

Assessing the Value of Dynamical Downscaling

Used together, Scores RCM, fGCM and cGCM quantify the value of dynamic downscaling for the given process/phenomenon.

Dynamic downscaling of little/no value.

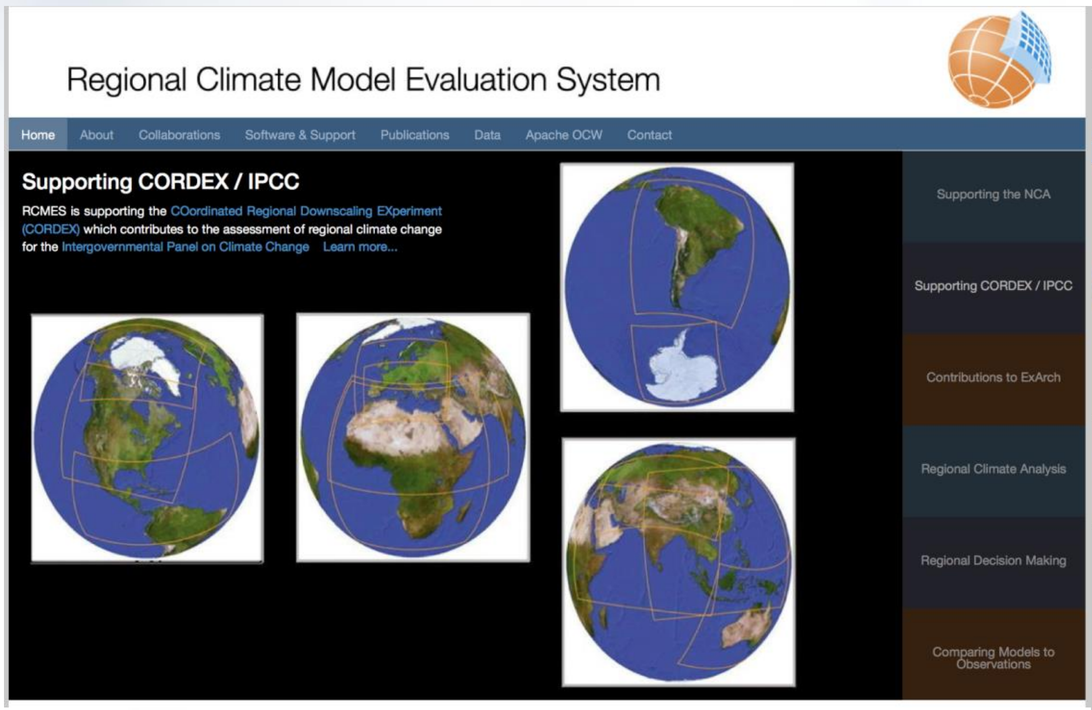
Dynamic downscaling of high value.

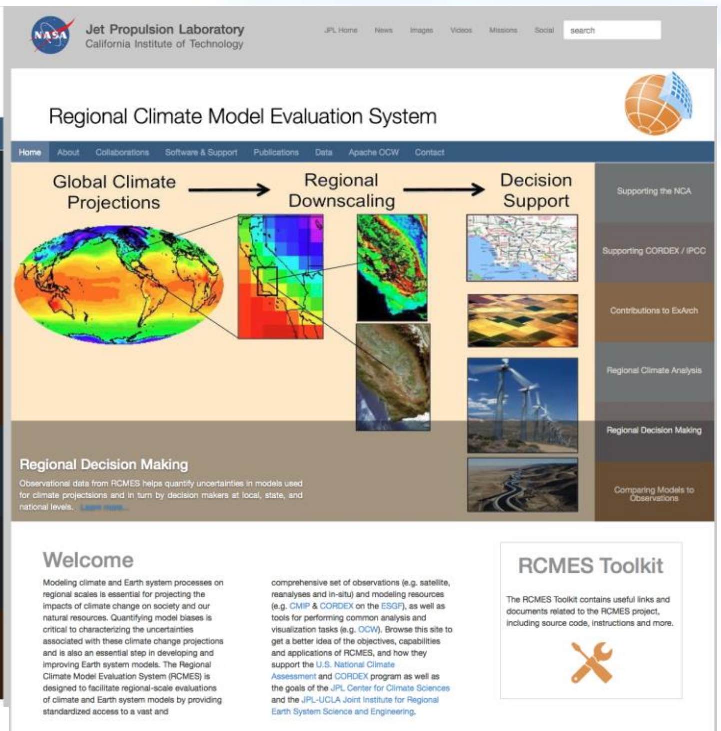
Outline

- Introduction
- Systematic evaluation of CORDEX RCMs using the Regional Climate Model Evaluation System
  - The following slides are provided by Alexander Goodman at JPL
  - *Lee et al.* (2018), Regional Climate Model Evaluation System (RCMES) powered by Apache Open Climate Workbench (OCW) v1.3.0, in preparation.
- Use cases of RCMES
  - Evaluation of hourly rainfall characteristics in RCM simulations
- Toward the model evaluation of the future
  - Combining multiple metrics into one: multi-objective optimization
  - Multi-resolution investigation of climate models

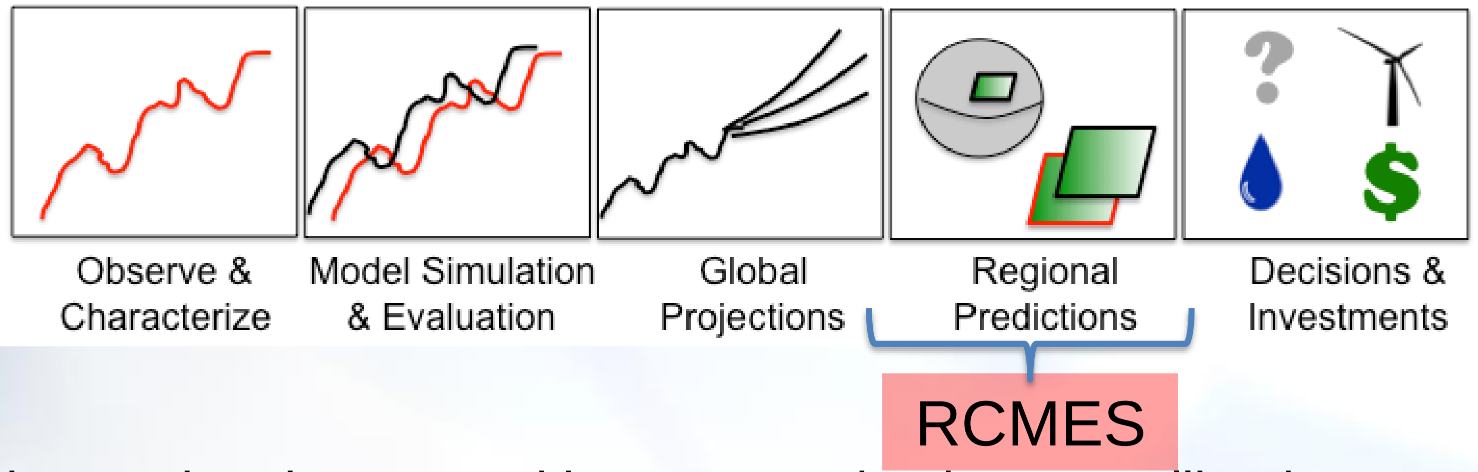
The Regional Climate Model Evaluation System (RCMES, http://rcmes.jpl.nasa.gov)

- Joint collaboration: JPL/NASA, UCLA
- Python-based open source software powered by the Apache Open Climate Workbench library





Regional Climate Model Evaluation System : Motivation & Goals



GOALS

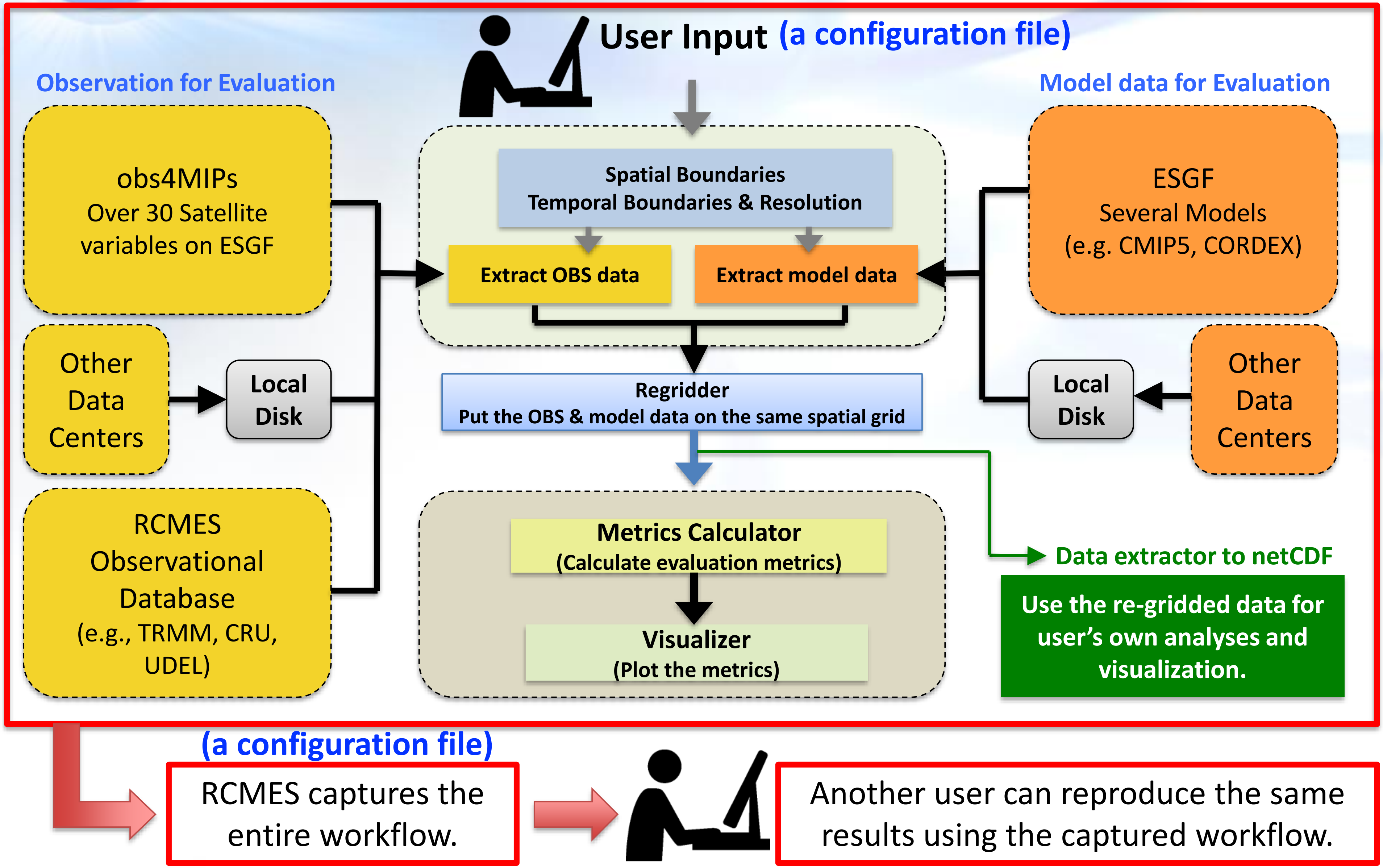
- Make observation datasets, with some emphasis on satellite data, more accessible to the regional climate modeling community.
- Make the evaluation process for regional climate models simpler, quicker and physically more comprehensive.
- Provide researchers more time to spend on analysing results and less time coding and worrying about file formats, data transfers, etc.

BENEFITS

- Quantify model strengths/weaknesses for development/improvement efforts.
- Improved understanding of uncertainties in predictions.

Regional Climate Model Evaluation System

High-Level Architecture



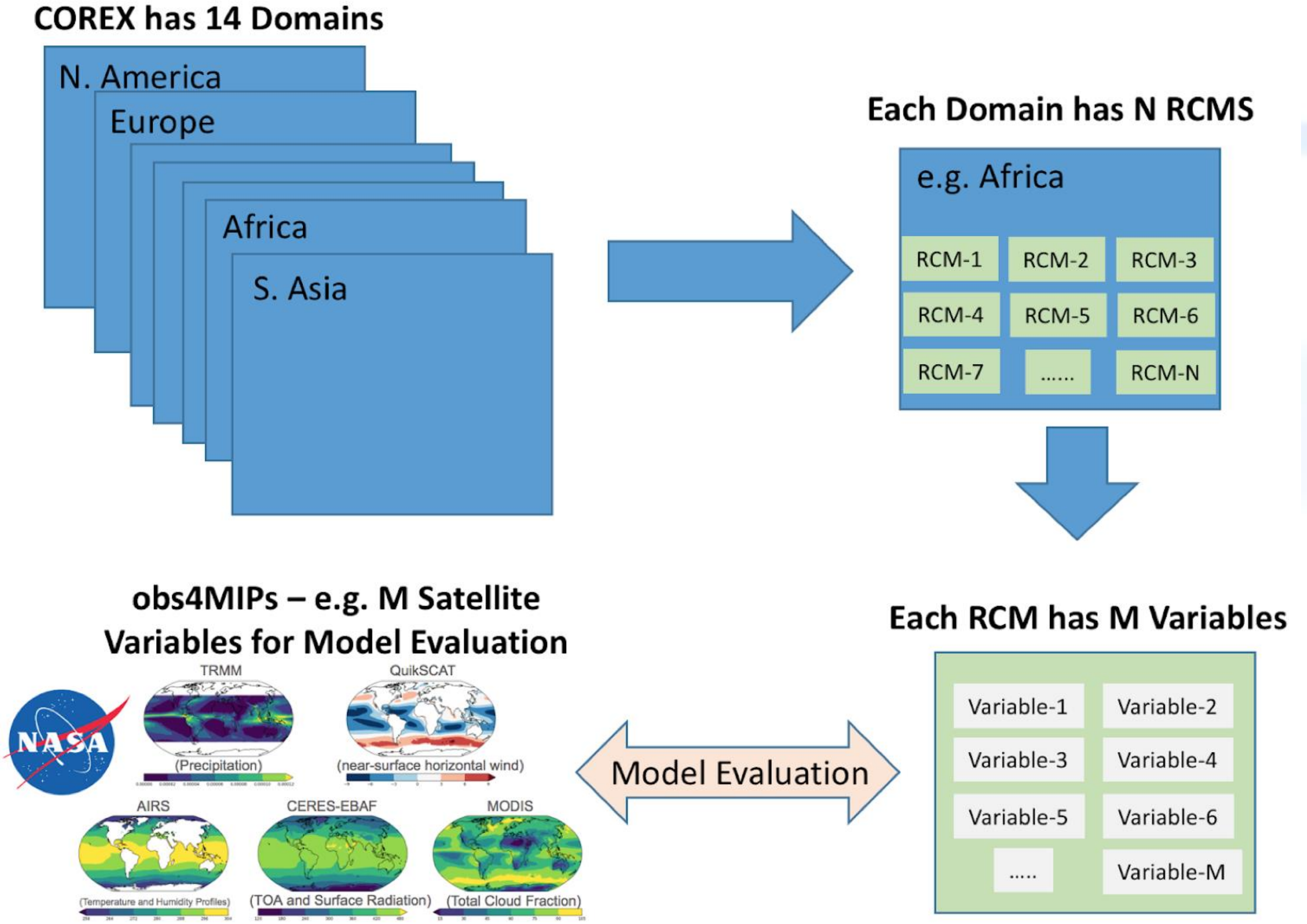
A Systematic Evaluation Of CORDEX Simulations Using Obs4MIPs

Leverage NASA-sponsored RCMES and NASA- co-sponsored obs4MIPs to benefit WCRP/CORDEX

GOALS

- Make it easy to perform systematic evaluations with CORDEX domains, models and variables
  - Utilize the Regional Climate Model Evaluation System
- Leverage NASA satellite observational datasets in evaluations
  - Utilize the Observations for Model Intercomparisons Project (Obs4MIPs)
  - Evaluate the simulate variables other than temperature and precipitation
- Increase communication and collaboration between CORDEX domains

The multi-model, multi-variable evaluation against satellite observations from obs4MIPs over the 14 CORDEX domains



Why do we need “Systematic Evaluation”?


```yaml
workdir: /home/goodman/data_processing/CORDEX/analysis/NAM-44/CERES-EBAF/rlus/annual
output_netcdf_filename: rlus_CERES-EBAF_NAM-44_annual.nc

# (RCMES will temporally subset data between month_start and month_end.
# If average_each_year is True (False), seasonal mean in each year is (not) calculated and used for metrics calcul
time:
    maximum_overlap_period: True
    temporal_resolution: monthly
    month_start: 1
    month_end: 12
    average_each_year: True

space:
    boundary_type: CORDEX NAM

regrid:
    regrid_on_reference: True

datasets:
  - loader_name: local_split
    name: CERES-EBAF
    file_path: /proj3/data/obs4mips/rlus_CERES-EBAF_L3B_Ed2-8_*.nc
    variable_name: rlus
  - loader_name: local_split
    name: UQAM-CRCM5
    file_path: /proj3/data/CORDEX/NAM-44/rlus/rlus_NAM-44_ECMWF-ERAINT_evaluation_r1i1p1_UQAM-CRCM5_v1_mon_*.nc
    variable_name: rlus
    lat_name: lat
    lon_name: lon
  - loader_name: local_split
    name: SMHI-RCA4
    file_path: /proj3/data/CORDEX/NAM-44/rlus/rlus_NAM-44_ECMWF-ERAINT_evaluation_r1i1p1_SMHI-RCA4_v1_mon_*.nc
    variable_name: rlus
    lat_name: lat
    lon_name: lon
  - loader_name: local_split
    name: DMI-HIRHAM5
    file_path: /proj3/data/CORDEX/NAM-44/rlus/rlus_NAM-44_ECMWF-ERAINT_evaluation_r1i1p1_DMI-HIRHAM5_v1_mon_*.nc
    variable_name: rlus
    lat_name: lat
    lon_name: lon
  - loader_name: local_split
    name: MOHC-HadRM3P
    file_path: /proj3/data/CORDEX/NAM-44/rlus/rlus_NAM-44_ECMWF-ERAINT_evaluation_r1i1p1_MOHC-HadRM3P_v1_mon_*.nc
    variable_name: rlus
    lat_name: lat
    lon_name: lon
```


Season

Domain

Observations

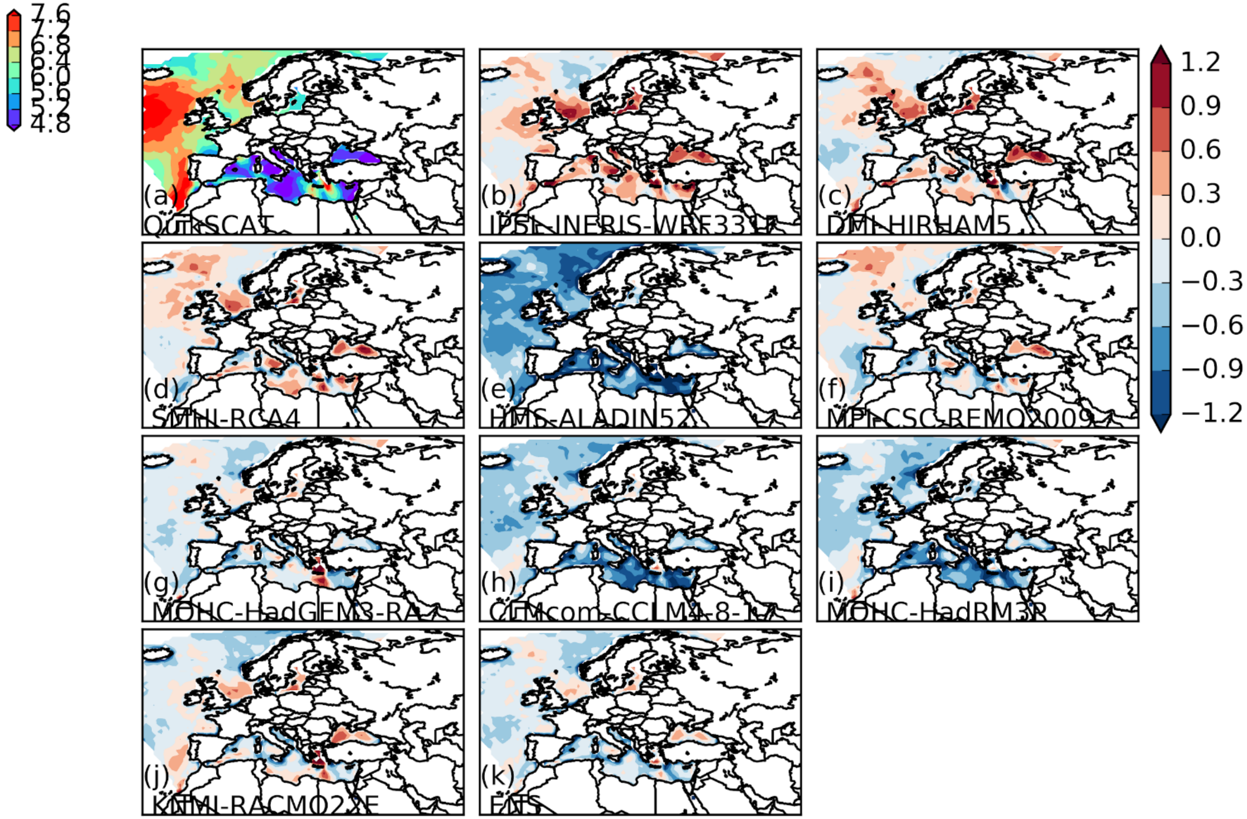
Models

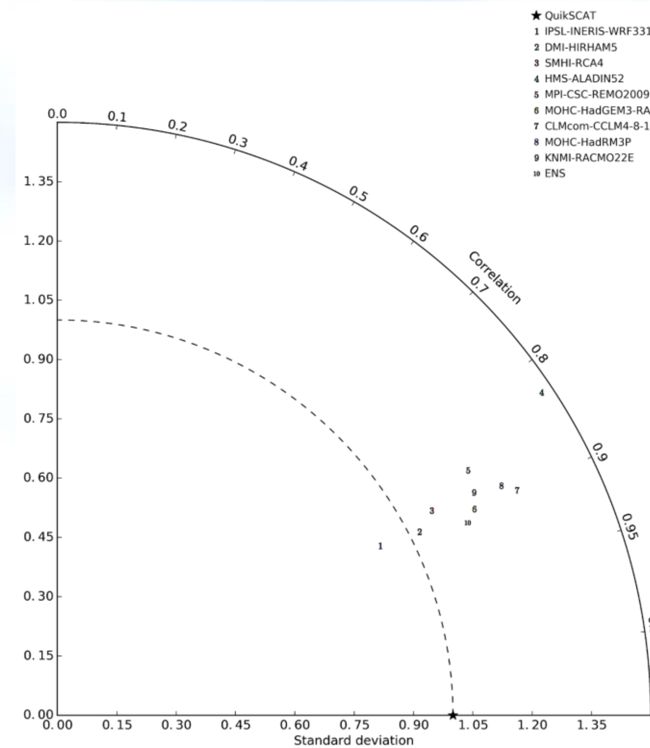
- This CF file (namelist file) is necessary to run each evaluation combination (CORDEX Domain, Season and Variable), forming a large “evaluation matrix”.
- 14 variables x 13 domains x 3 seasons x ~10 models > **5000** evaluations

- Three input parameters
  1. **CORDEX domain name**
  2. **obs. data location**
  3. **model data location**

QuikSCAT Zonal Sea-Surface Wind (m/s) (summer)

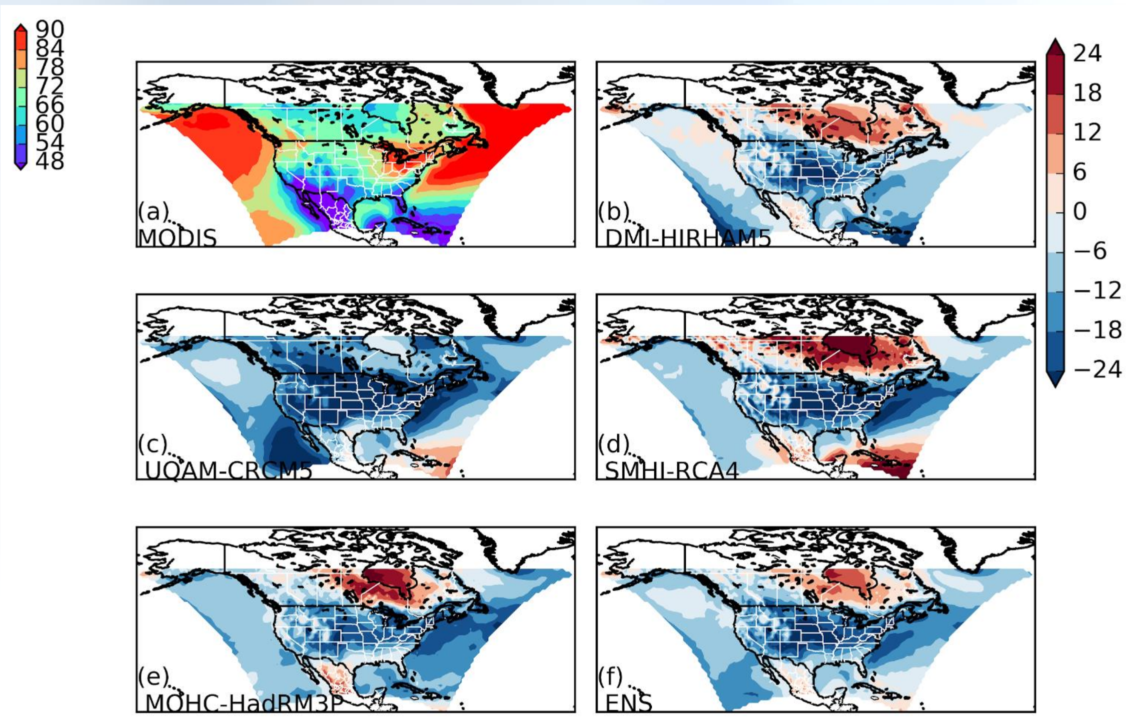
Europe

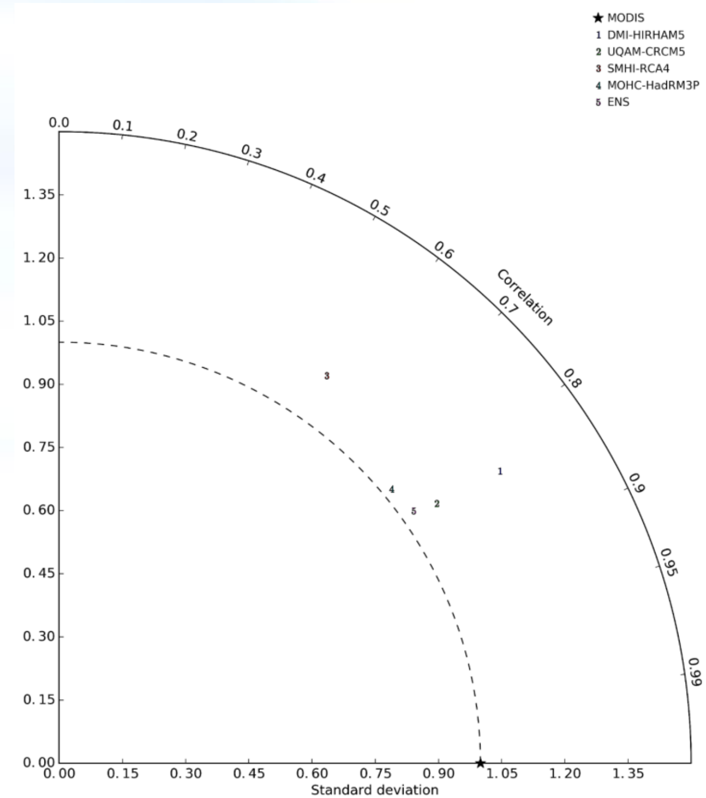




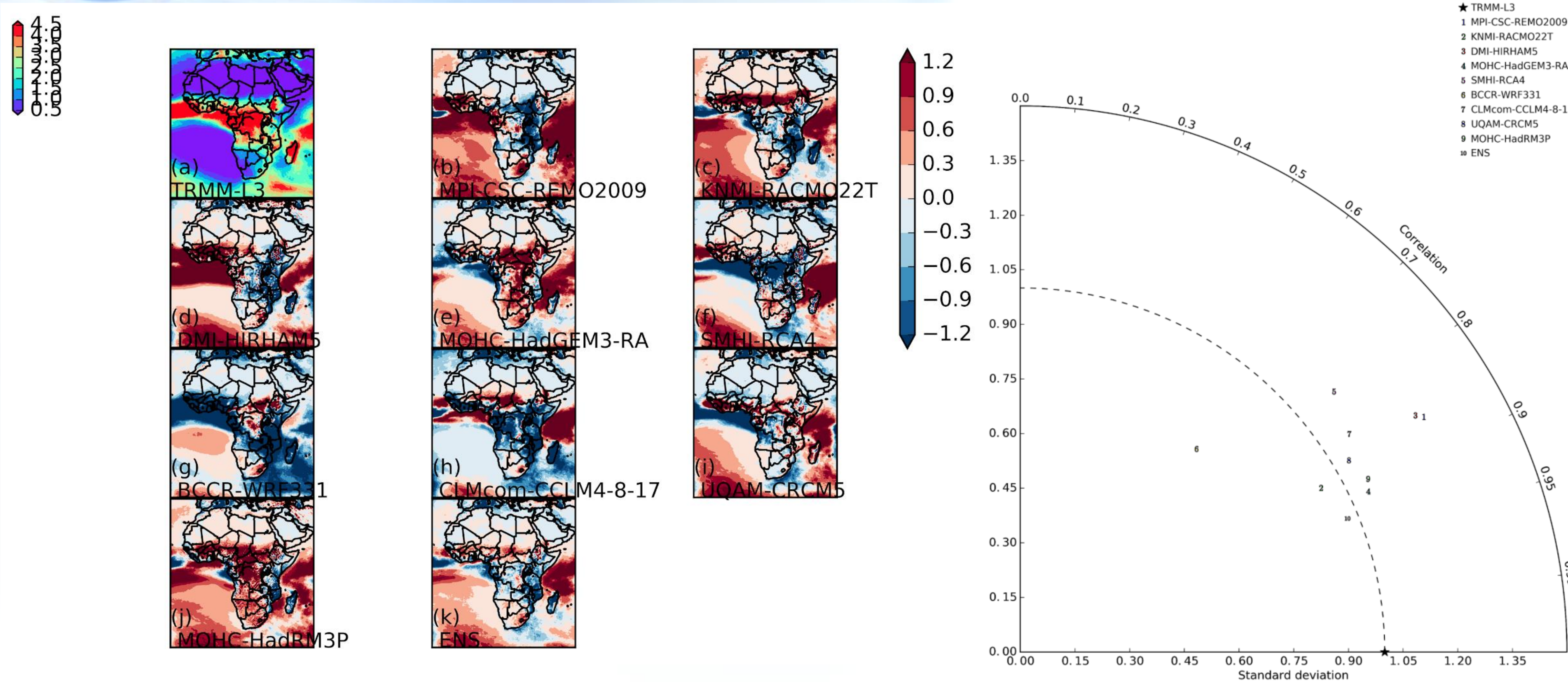
MODIS Total Cloud Fraction (%) (winter)

North America





TRMM Precipitation (mm/month) (annual)
Africa



- Initial Evaluations have focused on the three domains (North America, Europe, and Africa) with the most simulations on the ESGF.
- Analyze and post remaining evaluations with simulations (not on ESGF but) made available to us (e.g. East Asia).
- We will publish these results on the RCMES website.

Outline

- Introduction
- Systematic evaluation of CORDEX RCMs using the Regional Climate Model Evaluation System
- Use cases of RCMES
  - Evaluation of hourly rainfall characteristics in RCM simulations

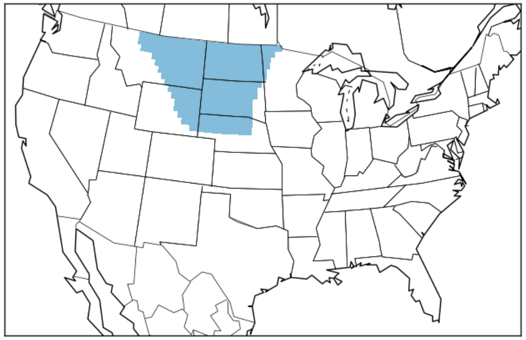
Lee et al. (2017), Evaluating hourly rainfall characteristics over the U.S. Great Plains in dynamically downscaled climate model simulations using NASA-Unified WRF, JGR.

- Toward the model evaluation of the future
  - Combining multiple metrics into one: multi-objective optimization
  - Multi-resolution investigation of climate models

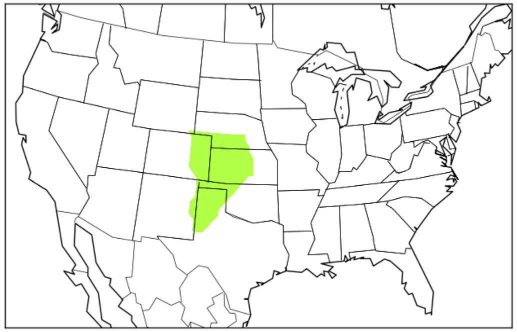
Hourly Precipitation [mm/hr] datasets over the Great Plains in summer (JJA)

	Resolution	Period
Observations		
Stage IV	4 km, hourly	2002-2010
GPM	0.1^{o}, 30 minutes	2014-2015
NU-WRF with spectral nudging		
WRF24	24 km, hourly	2002-2010
WRF12	12 km, hourly	2002-2010
WRF04	4 km, hourly	2000-2004
GEOS replay		
GEOS	12.5 km, hourly	2002-2009

10: Northern Plains



11: Central Plains



12: Southern Plains

from [Bukovsky 2011]

Challenges in evaluating high-resolution model output

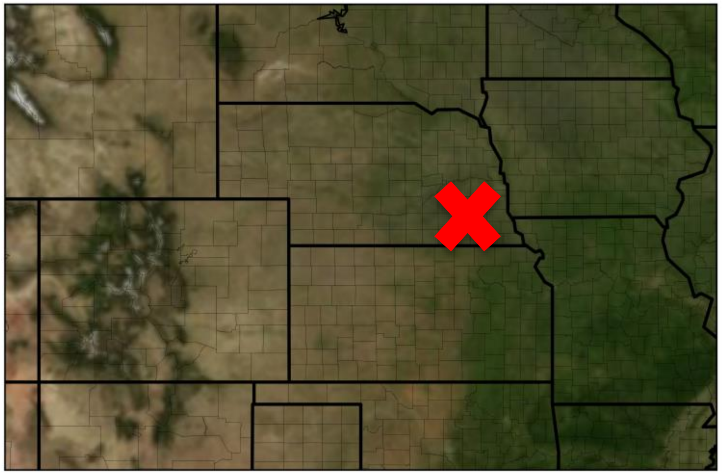
- The high-resolution climate simulation results from GEOS and NU-WRF are massive and stored on the NASA Center for Climate Simulation (NCCS) servers at Goddard.

Model resolution	Data size (two-dimensional variables only)
24 km	560 G
12 km	2.2 T
4 km	(expected to be more than 8 Terabytes)

- Conventional model evaluation techniques require regridding of datasets.
- B4 simulation is only available for five years (2000 - 2004). GPM data became available in 2014.

Objectives

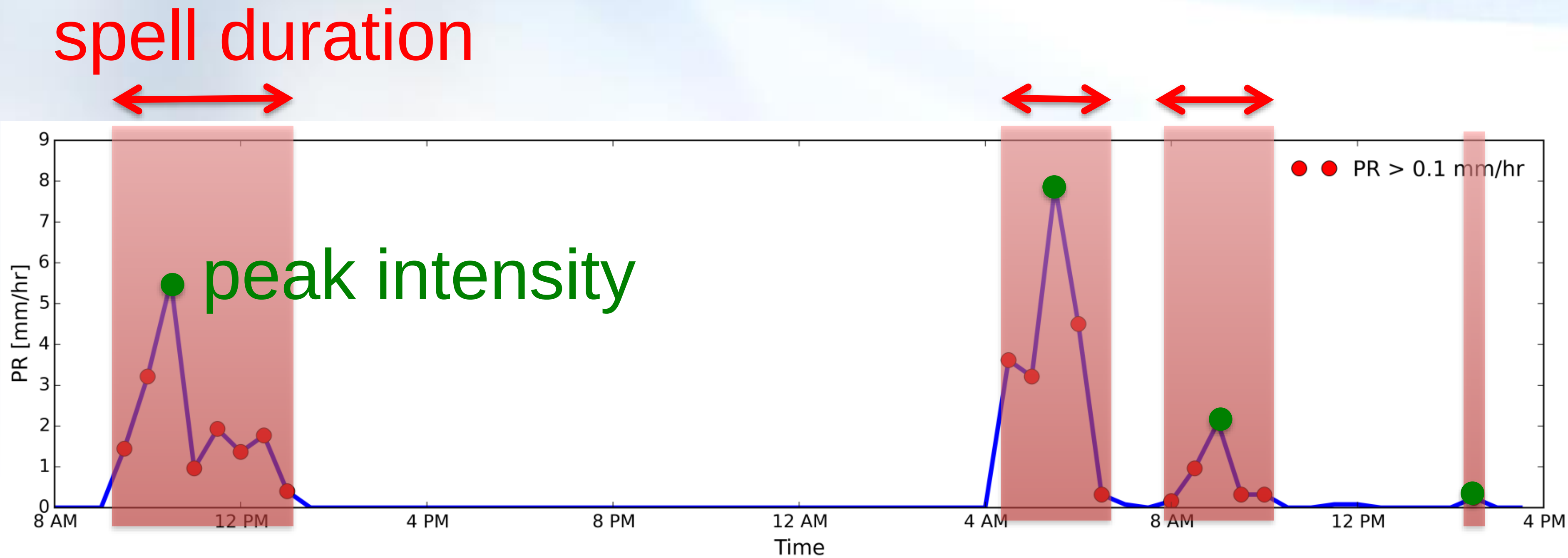
- Evaluate the NU-WRF and GEOS simulations using a metric which
  - does not require transfer of massive datasets,
  - does not require regridding of original output,
  - highlights advantages of high-resolution simulations.



- When it rains at somewhere in the Great Plains, would there be a probability density function to forecast **how (strong/long/much) it will rain**?

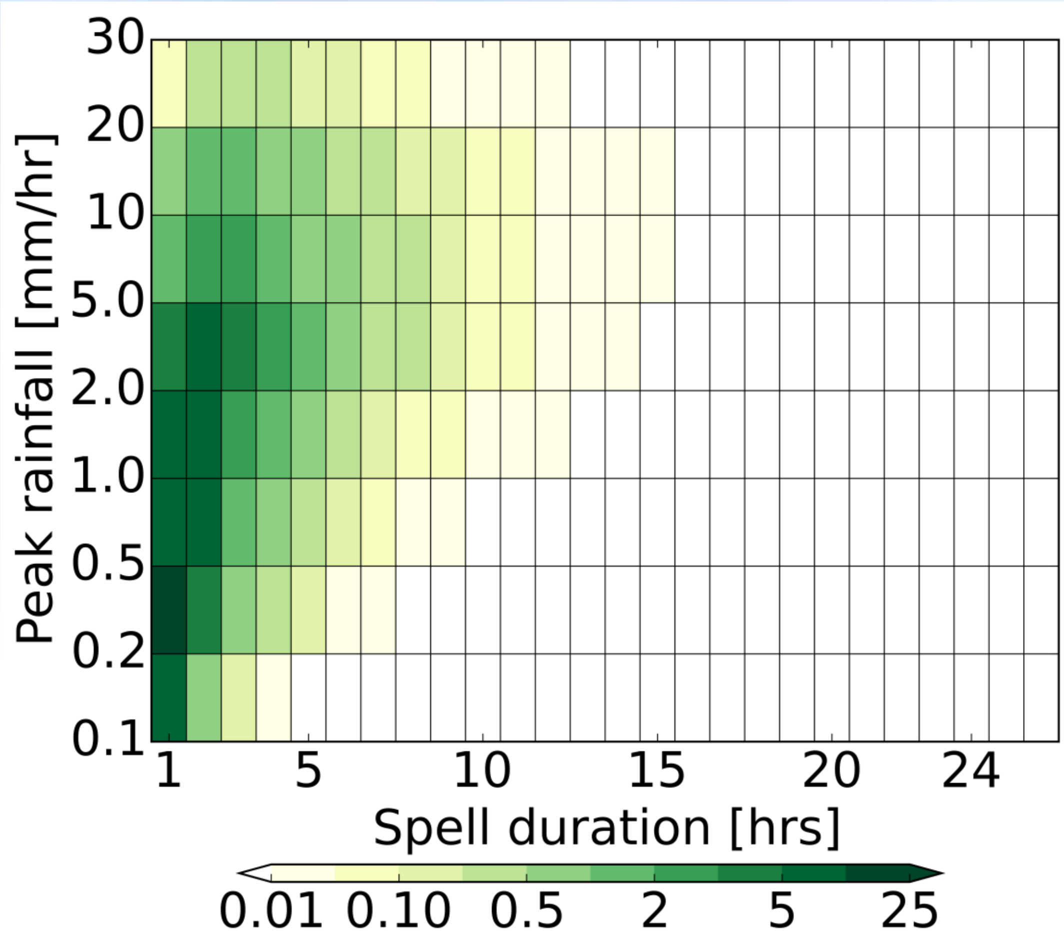
Definition of wet spell duration [hours] and peak rainfall [mm/hr] [*Kendon et al.*, 2014]

half-hourly GPM precipitation
(at Lat.=40.25N & Lon.=98.15W on June 27th and 28th, 2014)



(spell duration, peak intensity) over a season at every grid points in a region of interest → joint distributions

A Joint Probability Distribution Function (JPDF) of Rainfall Duration and Intensity



x-axis: duration

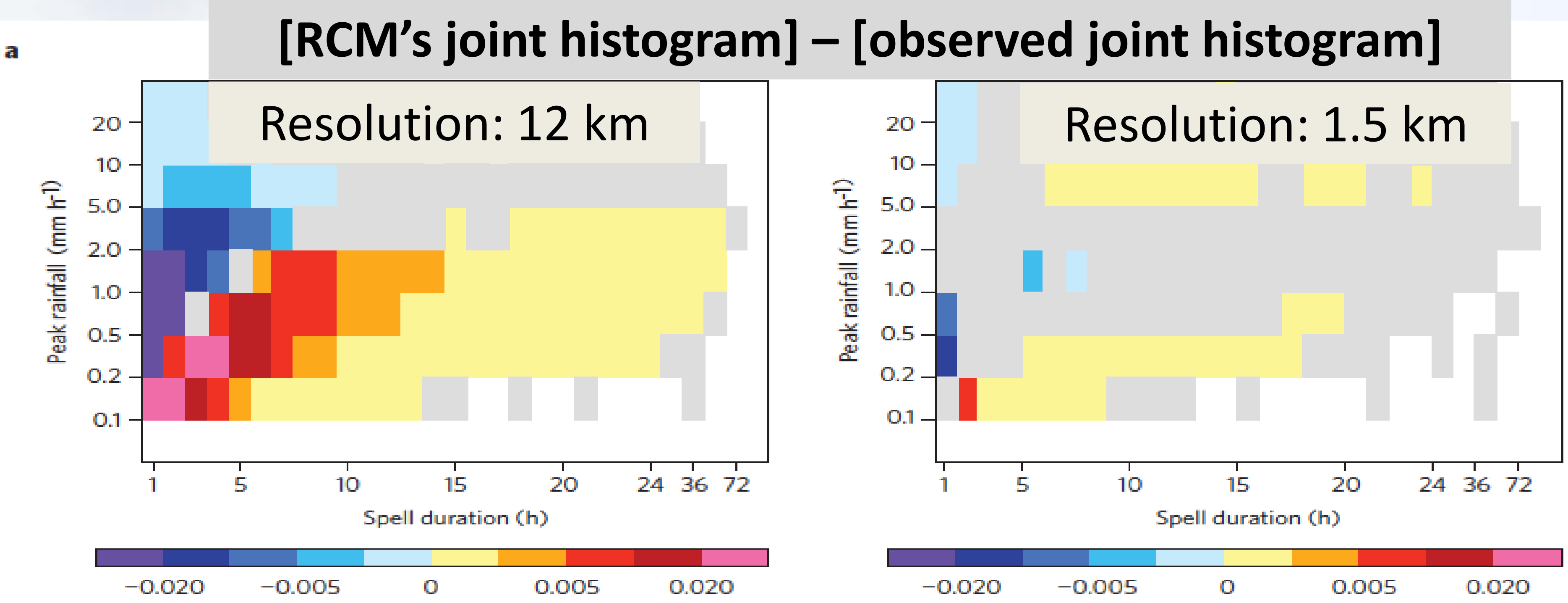
y-axis: peak intensity

color: probability density [%]

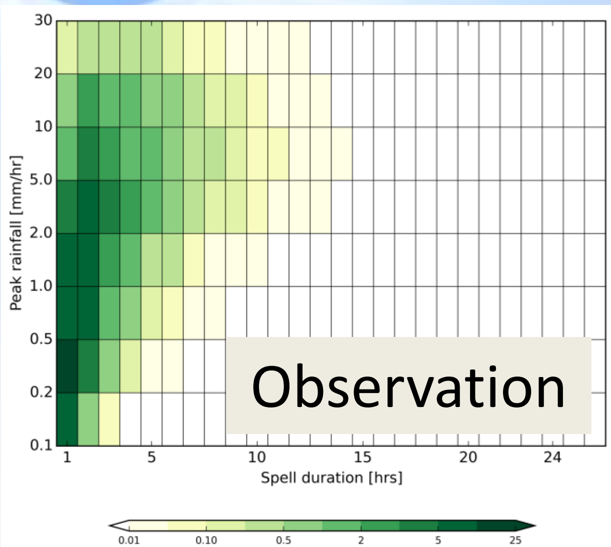
$\Sigma p(\text{duration, intensity}) = 100\ \%$

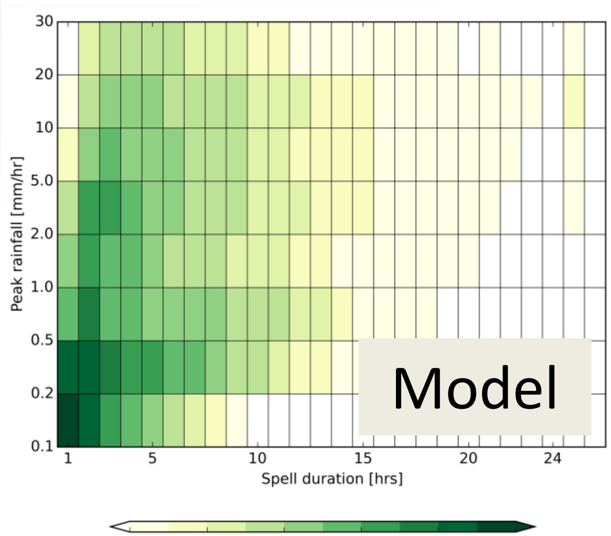
What is the added value of high-resolution RCM simulations?

- from Figure 4 in *Kendon et al.* [2014]: model biases in the joint probability distribution of wet spell duration (x-axis) vs. peak intensity (y-axis).
- The 1.5 km model simulates more realistic hourly rainfall characteristics including extremes.

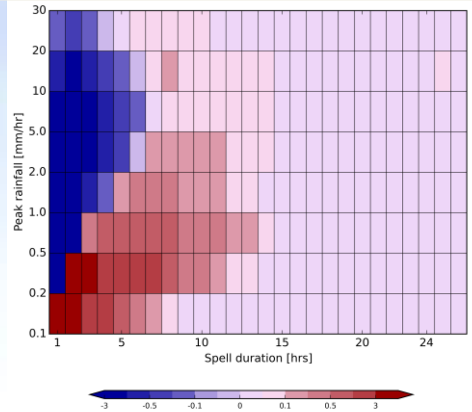


Comparison of two JPDFs

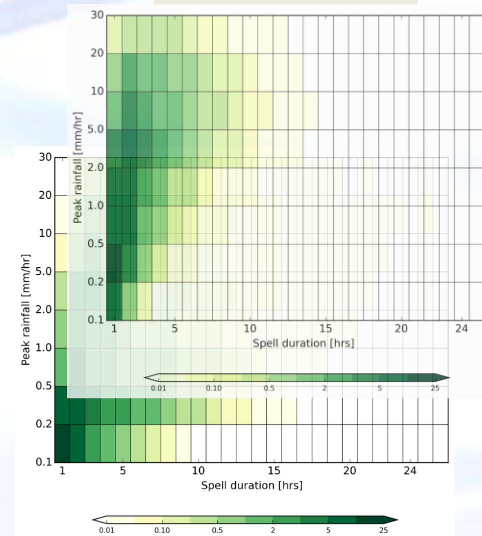




Model - Observation



Overlap

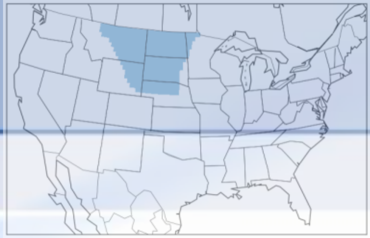


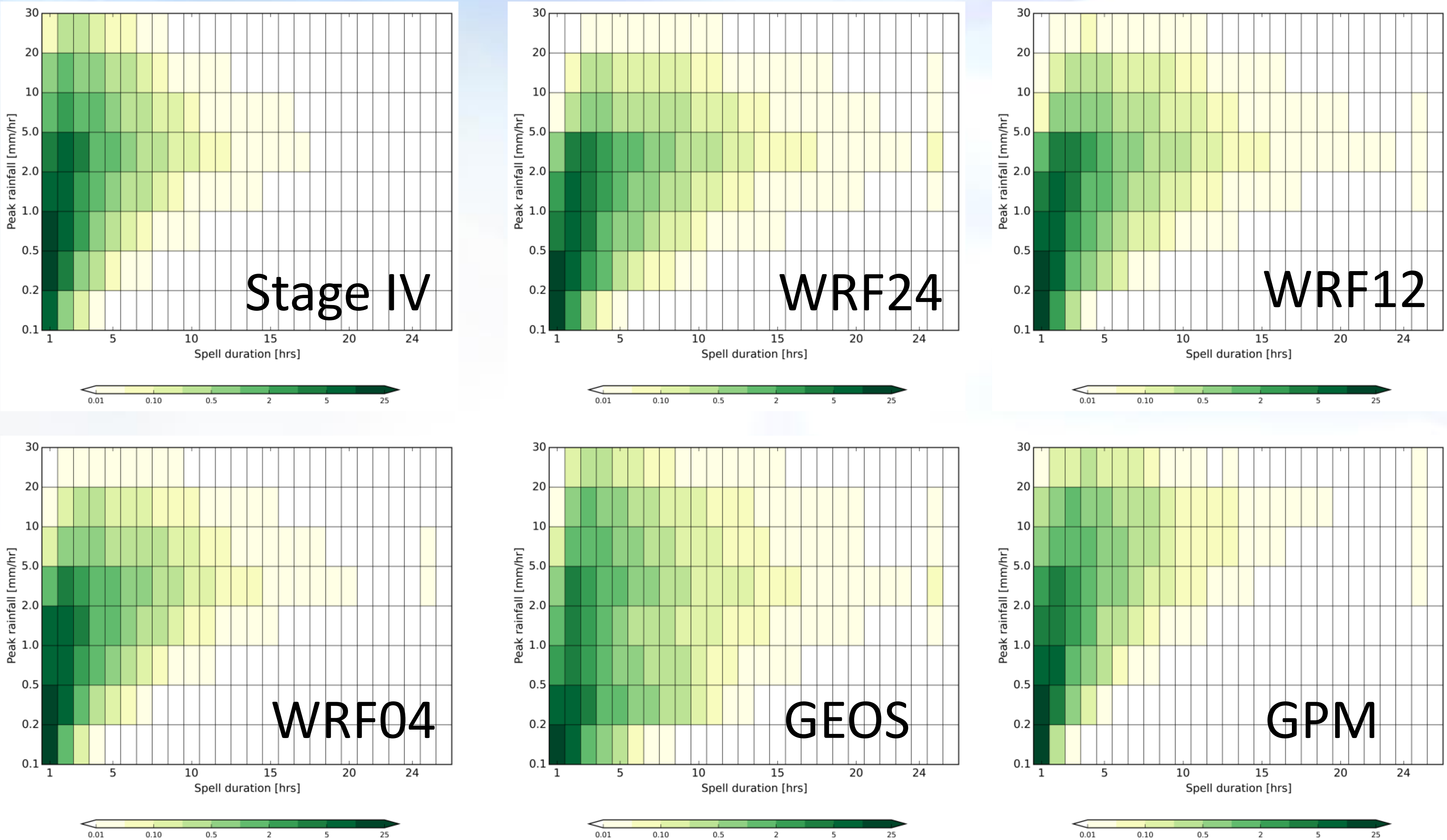
- As a measure of similarity between two joint PDFs, the overlap ratio was calculated (0%: no overlap, 100%: perfect overlap).

$$\text{overlap ratio } [\%] = \sum_{x,y} min[f_1(x,y), f_2(x,y)]$$

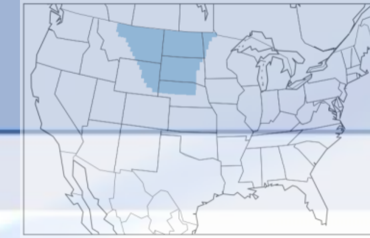
Joint PDF for each dataset (summer: JJA)

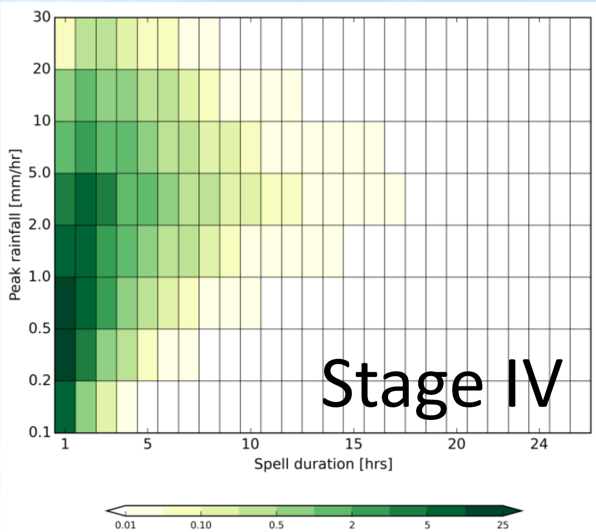
: from precipitation data at all grid points in the Northern Plains

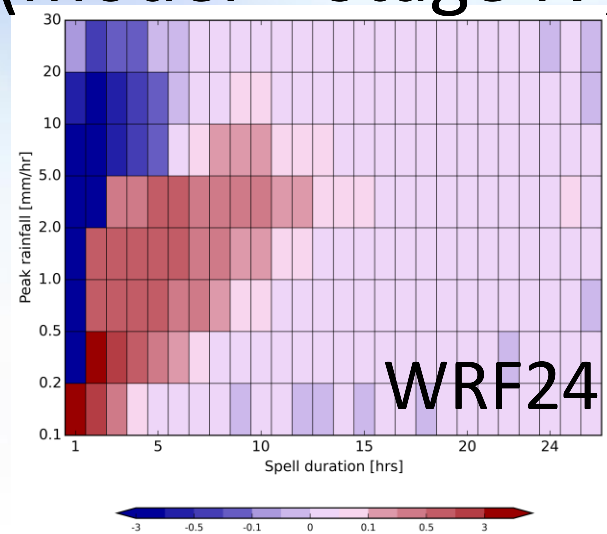


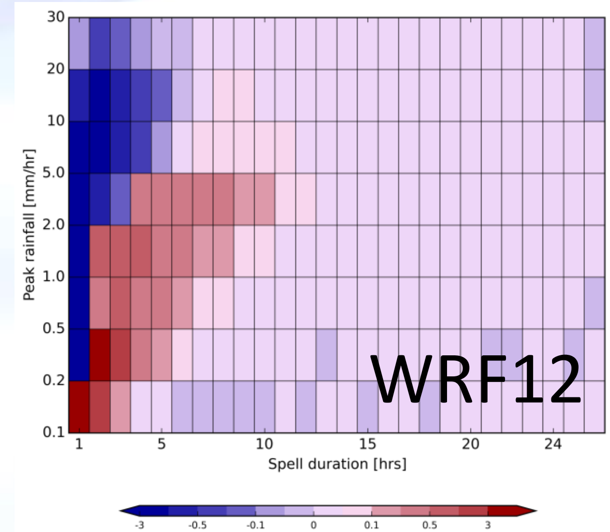


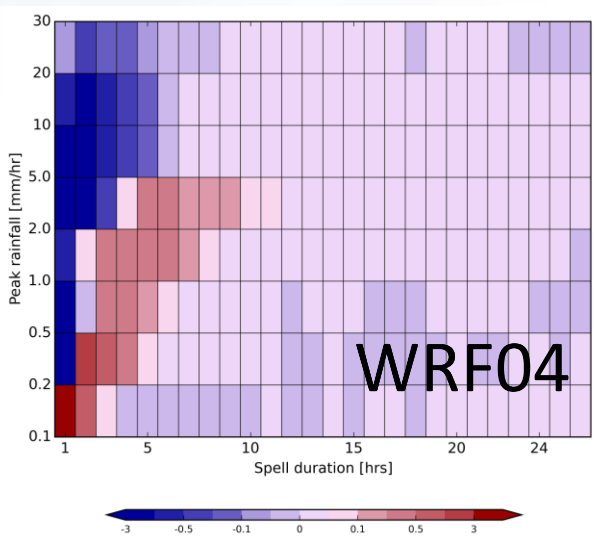
Comparison of JPDFs: Northern Plain, Summer (model – Stage IV)

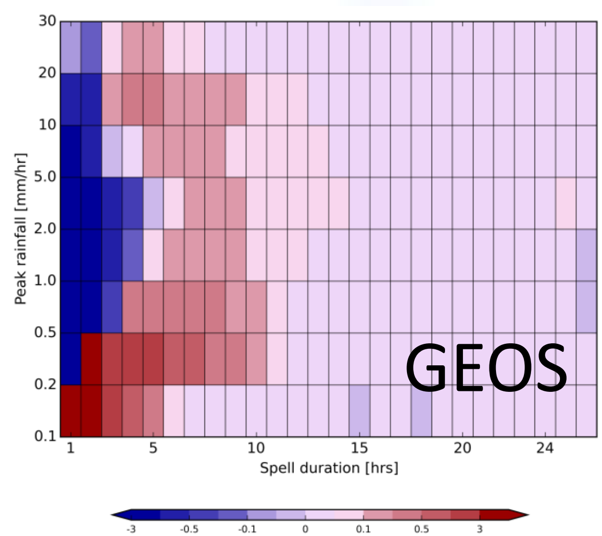






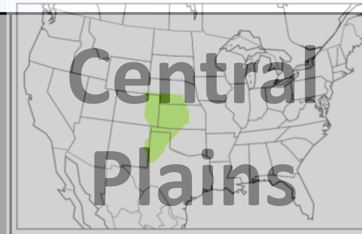




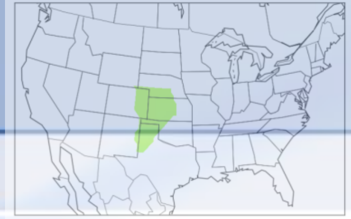


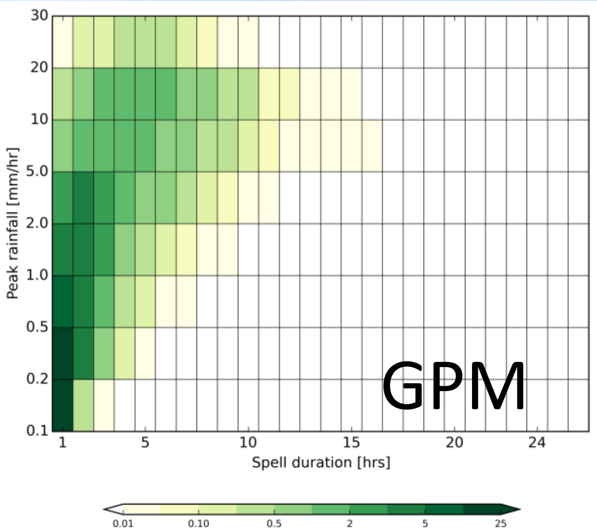
- Less frequent **short-duration downpour events** in both NU-WRF and GEOS regardless of resolutions.

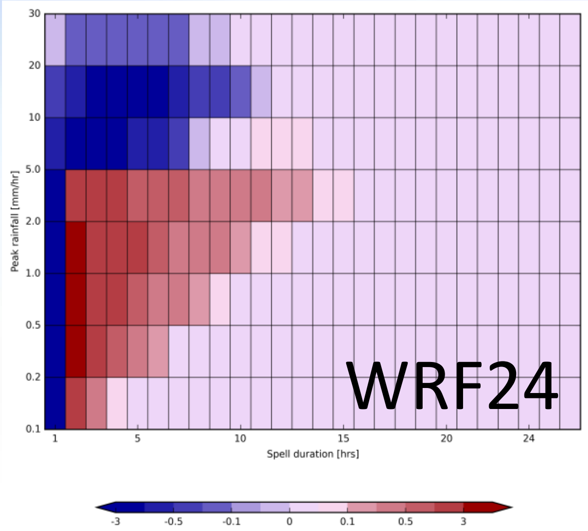
Overlap of the joint probability functions of rainfall duration and peak intensity (observation data: Stage IV)

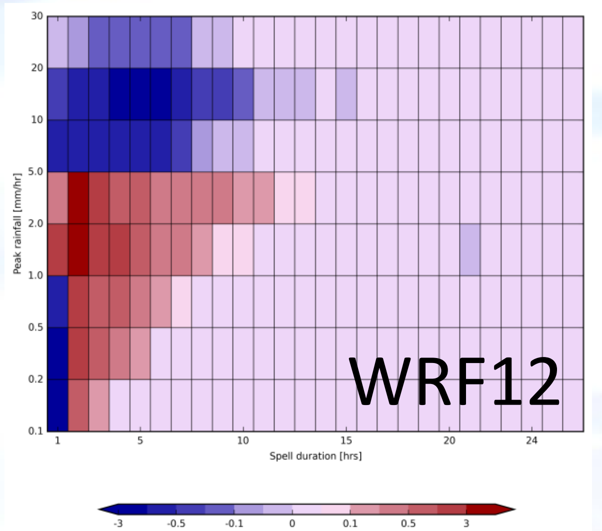
[%]	Northern Plains 	Central Plains 	Southern Plains 
WRF24	70.0	67.4	61.6
WRF12	79.3	78.1	74.4
WRF04 (2000-2004)	**83.9**	**81.2**	**82.3**
GEOS (2002-2009)	63.8	51.4	29.8

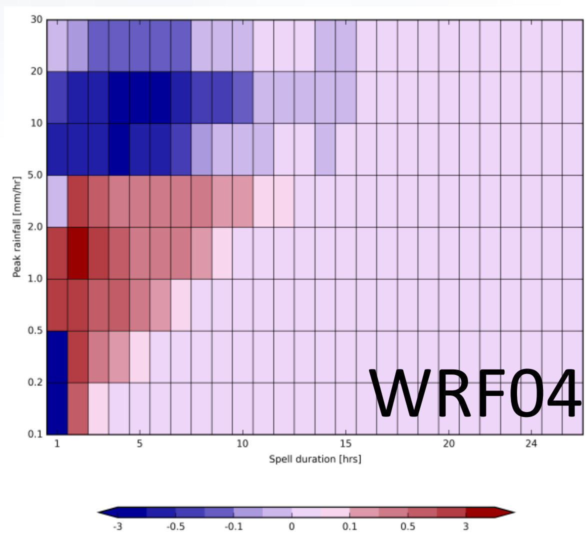
Comparison of JPDFs: Central Plain, Summer (model – GPM)

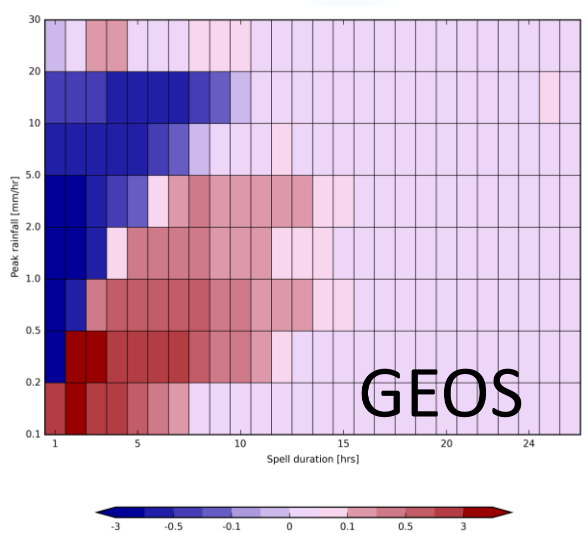




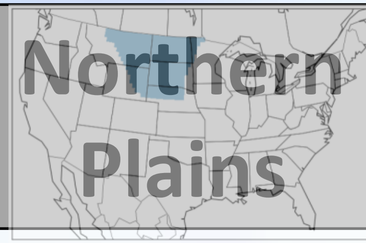
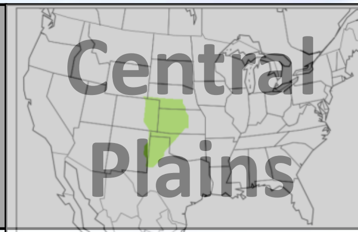
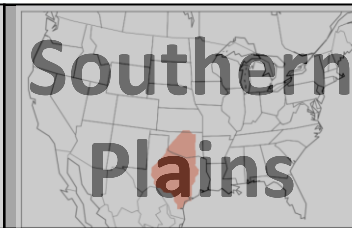




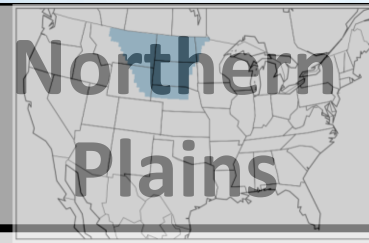
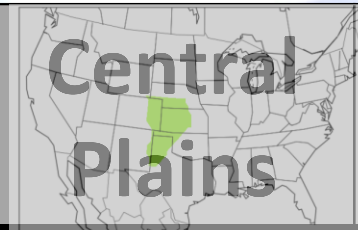




Overlap of the joint probability functions of rainfall duration and peak intensity (observation data: GPM IMERG)

[%]	 Northern Plains	 Central Plains	 Southern Plains
WRF24	67.7	62.3	57.3
WRF12	77.2	72.3	69.6
WRF04 (2000-2004)	**82.7**	**76.8**	**76.5**
GEOS (2002-2009)	72.1	60.6	37.9
MERRA2	38.1	29.4	17.9

Spectral nudging effect
(observation data: Stage IV)

[%]	 Northern Plains	 Central Plains	 Southern Plains
WRF24, 600 km	70.0	67.4	61.6
WRF24, 2000 km	69.6	67.2	62.2
WRF24, no nudging	65.8	67.6	63.8
WRF12, 600 km	79.3	78.1	74.4
WRF12, 2000 km	79.5	78.5	75.3
WRF12, no nudging	74.1	76.6	75.4

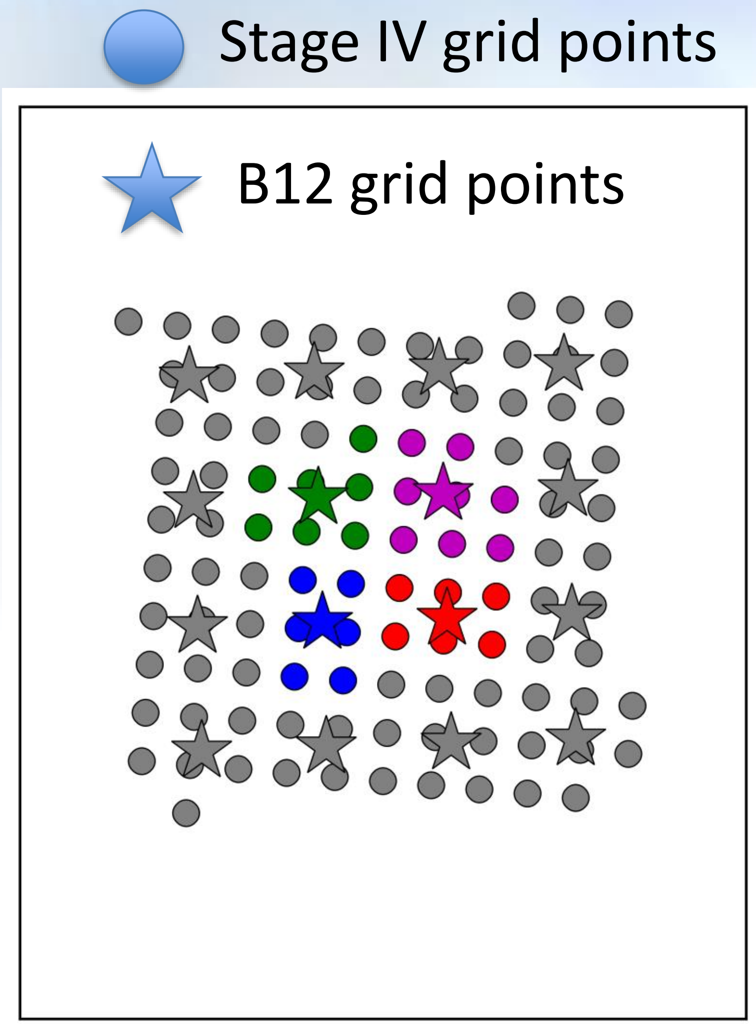
Cumulus parameterization effect (observation data: Stage IV)

[%]	 Northern Plains	 Central Plains	 Southern Plains
WRF24, Grell 3-D ensemble	70.0	67.4	61.6
B24, Betts-Miller-Janjic	57.1	54.4	52.5
B24, New Kain-Fritsch	62.2	57.9	55.1
B24, New simplified Arakawa-Schubert	69.8	72.0	70.3
WRF12, Grell 3-D ensemble	79.3	78.1	74.4
WRF04, Grell 3-D ensemble	**83.9**	**81.2**	**82.3**

Issues

- The difference in number of samples. For the Northern Great Plains,
  - WRF24: 1829 grid points
  - WRF12: 7296 grid points
  - WRF04: 65732 grid points
- The difference in data periods:
  - Stage IV, WRF24 and WRF12: 2002-2010
  - WRF04: 2000-2004
  - GPM: 2014-2015
- The difference between Stage IV and GPM

Remapping Stave IV data onto WRF12 and WRF24 grid points

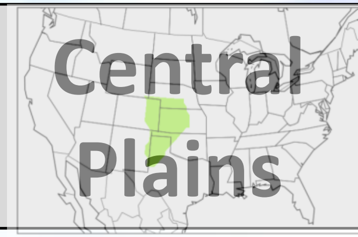


- H_0: At any grid point, B24 (or B12) can simulate the precipitation spatially averaged over nearby Stage IV grid points.

- Aggregate of Stage IV data near each B12 (or B24) grid point and calculate average.

Comparison with averaged Stage IV

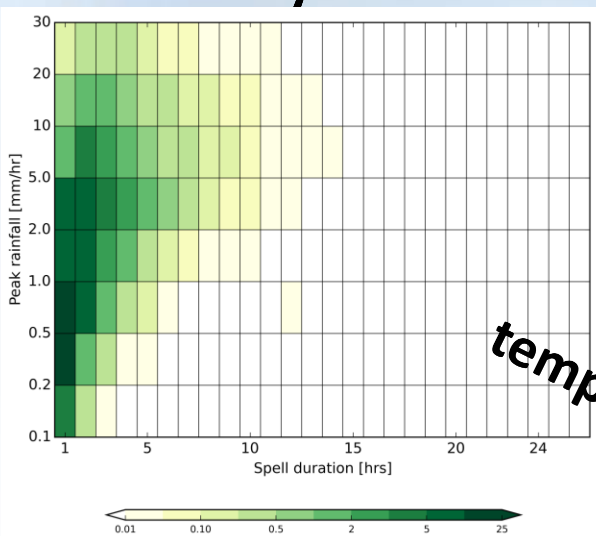
* all NU-WRF results with spectral nudging, 600 km

[%]	 Northern Plains	 Central Plains	 Southern Plains
WRF24	70.0->82.3	67.4->81.6	61.6->76.8
WRF12	79.3->84.6	78.1->82.3	74.4->79.3
WRF04 (2000-2004)	**83.9**	**81.2**	**82.3**

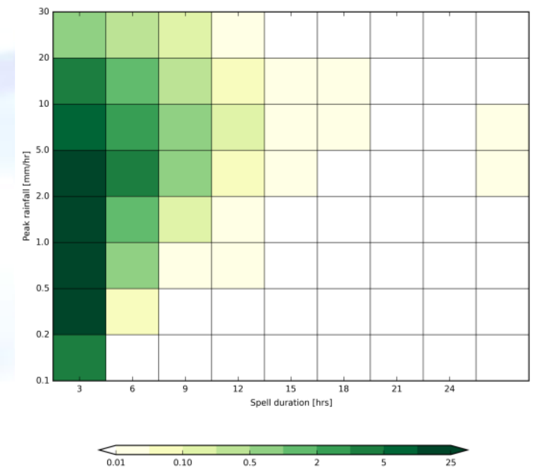
- WRF12 and WRF24 show reasonable agreement with the aggregated Stage IV data at the coarse resolution grid points.

Lost information due to temporal averaging (1) [Stave IV, JJA in 2002, Northern Great Plains]

Original JPDF from hourly data

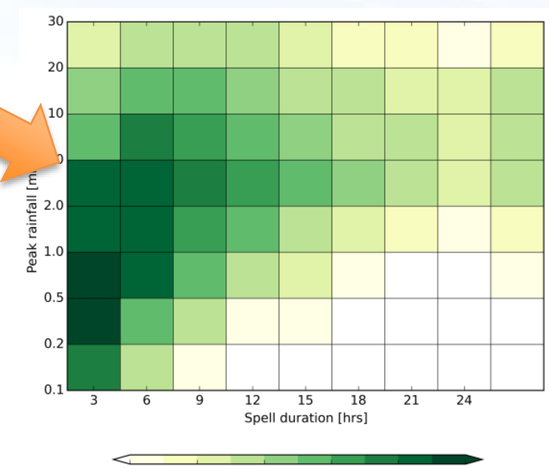


rebinning



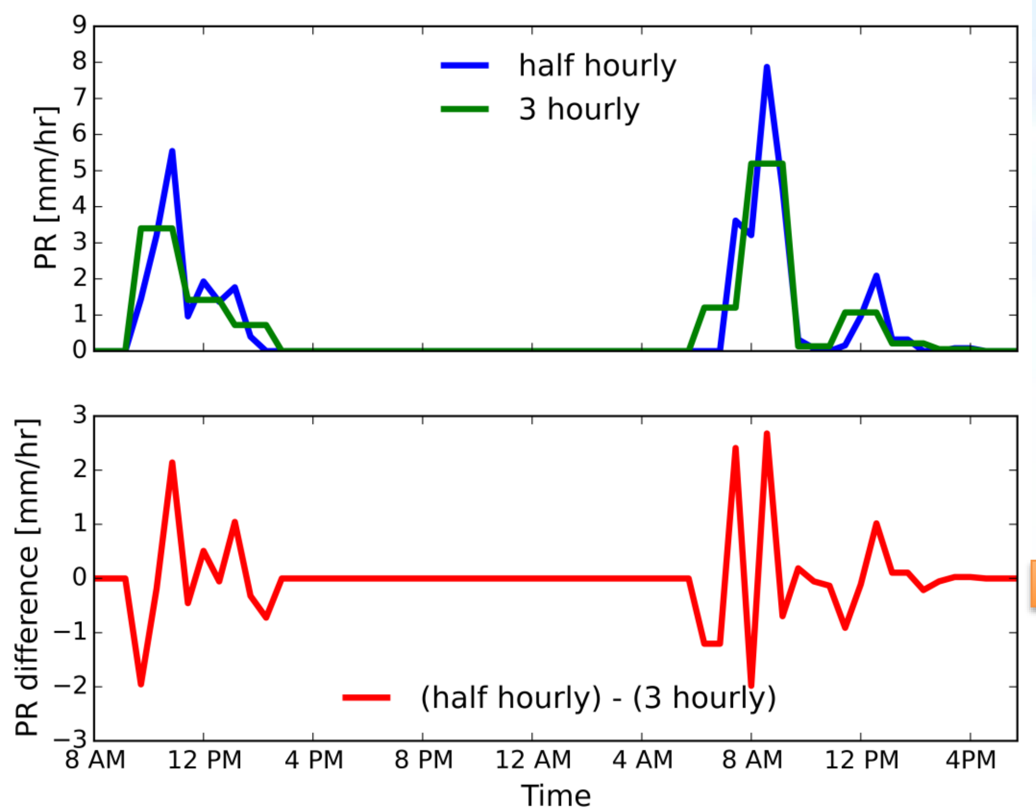
temporal averaging

3 hourly data



- ex) rainfall at 00, 02, 04, 05 UTC : two events for one hour, one event for two hours => one event for six hours
- Even the joint PDFs with the same bin widths look totally different: averaging artifacts.

Lost information due to temporal averaging (2) [half-hourly GPM precipitation]



Can this be added value of high resolution simulations?

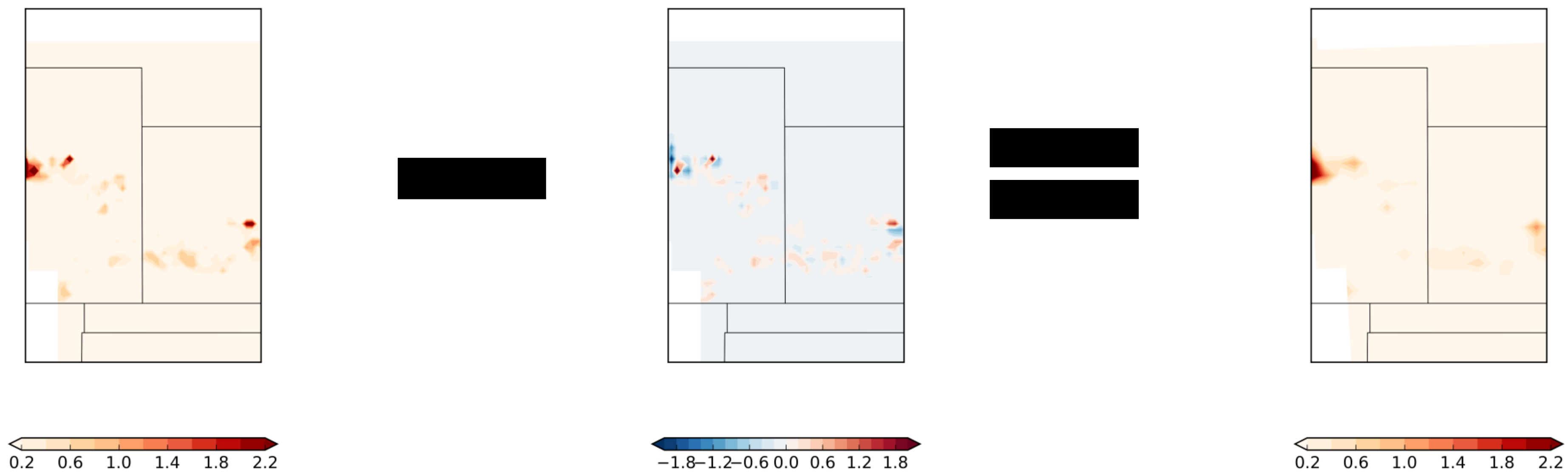
- Higher frequency variability in precipitation than an hour or less have potential to be better reproduced by RCMs with high resolution (WRF04) than low resolution simulations (WRF12 and WRF24).

Lost information due to spatial upscaling of high-resolution datasets [GPM, July 08/2014, 01:30 AM]

A. GPM precipitation (0.1°)

B. fine-scale spatial pattern

A-B. GPM precipitation aggregated into B24 grids



- By upscaling high resolution data into low resolution grid points, we lose fine-scale spatial information that is related to high frequency variability.
- The JPDF of rainfall duration and peak intensity from hourly data without regridding captures valuable information from high frequency variability and fine-scale spatial pattern.
- Satellite observations with high temporal and spatial resolutions such as GPM will highlight added value of high resolution simulations.

Issues

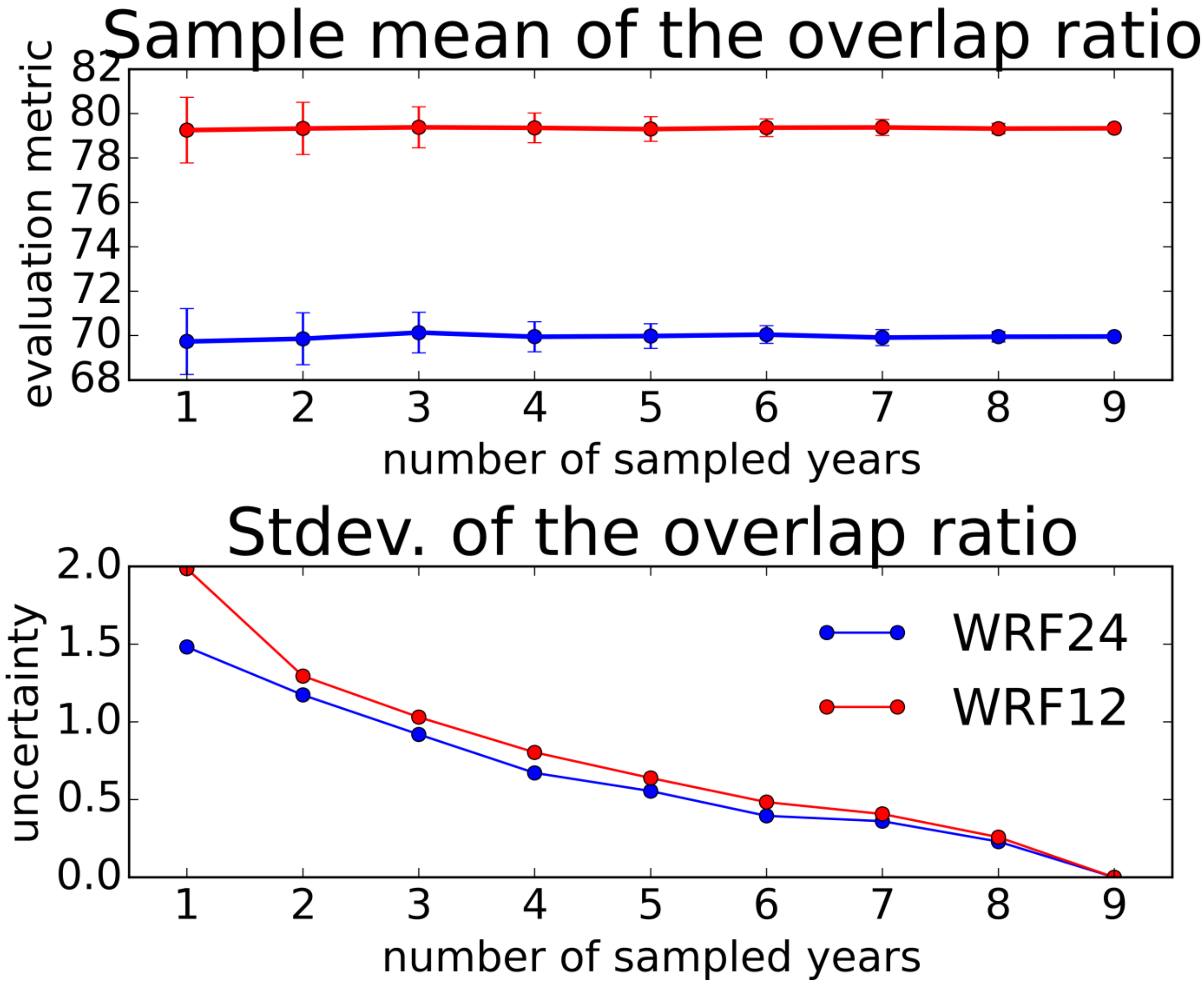
- The difference in number of samples. For the Northern Great Plains,
  - WRF24: 1829 grid points
  - WRF12: 7296 grid points
  - WRF04: 65732 grid points
- The difference in data periods:
  - Stage IV, WRF24 and WRF12: 2002-2010
  - WRF04: 2000-2004
  - GPM: 2014-2015
- The difference between Stage IV and GPM

Uncertainty of the overlap ratio due to temporal subsampling

- H_0: Uncertainty of the overlap ratio between WRF04 and Stage IV can be estimated by comparing subsampled JPDFs from WRF12 and WRF24 with Stage IV JPDF.
- Sampling 2002-2010 without replacement:

```
for sample_size = 1, 8
    for number_sample = 1, 100
        sub_jpdf = CALC_JPDF(          )
        overlap_ratio = COMPARE(stage4_jpdf, sub_jpdf)
```

- ex) (2002, 2003, 2005) (2004, 2007, 2008) (2002, 2009, 2010) ... (2002, 2004, 2010)

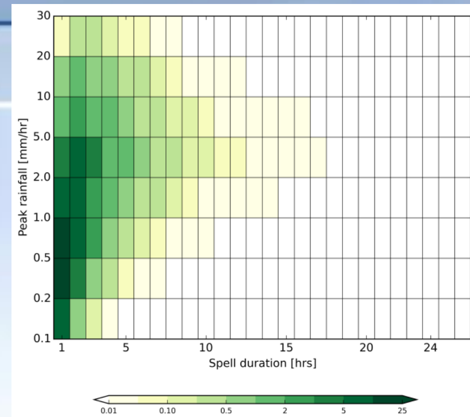


- The uncertainty is defined as standard deviation of the evaluation metric values (overlap ratio) for random subsamples.
- The uncertainty of the overlap ratio decreases by building a JPDF using output for a longer period.

Issues

- The difference in number of samples. For the Northern Great Plains,
  - WRF24: 1829 grid points
  - WRF12: 7296 grid points
  - WRF04: 65732 grid points
- The difference in data periods:
  - Stage IV, WRF24 and WRF12: 2002-2010
  - WRF04: 2000-2004
  - GPM: 2014-2015
- The difference between Stage IV and GPM

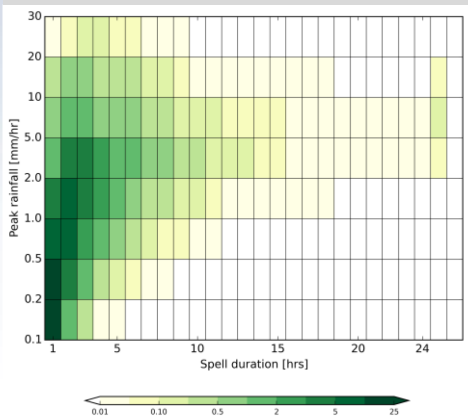
Stage IV [2002-2010]



Northern: 83.8 %
Central : 82.4 %
Southern: 84.7 %

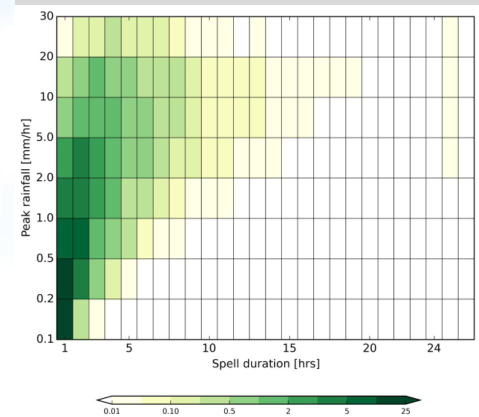
Northern: 82.1 %
Central : 87.4 %
Southern: 80.9 %

Stage IV [2014]



Northern: 81.0 %
Central : 84.5 %
Southern: 84.6 %

GPM [2014]



[%]	Northern Plains	Central Plains	Southern Plains
WRF04 & Stage IV	83.9	81.2	82.3
WRF04 & GPM	82.7	76.8	76.5

Conclusions (1)

- The three NU-WRF simulations (WRF24, WRF12 and WRF04) commonly show less frequent **short-duration downpour events**.
- Spectral nudging effect is not significant in the joint probability distribution function of precipitation peak intensity and duration.
- The effect of cumulus parameterization schemes is smaller than horizontal resolution.

Conclusions (2)

- In the Great Plains, WRF04 can simulate the most similar rainfall characteristics to Stage IV and GPM at its original grid points. The overlap ratios between WRF04 and the observations are similar to those within the observations.
- Added value of high-resolution simulations: lost information when spatially or temporally averaging high-resolution data.
- In other words, WRF04 run's value is allowing us to fully utilize high-resolution observation datasets (Stage IV and GPM) as they are.
- Without the full 11 years (2000-2010) of simulations or observations, we can still show the added value of WRF04 simulation with the uncertainty due to temporal sampling.

Outline

- Introduction
- Systematic evaluation of CORDEX RCMs using the Regional Climate Model Evaluation System
- Use cases of RCMES
  - Evaluation of hourly rainfall characteristics in RCM simulations
- Toward the model evaluation of the future

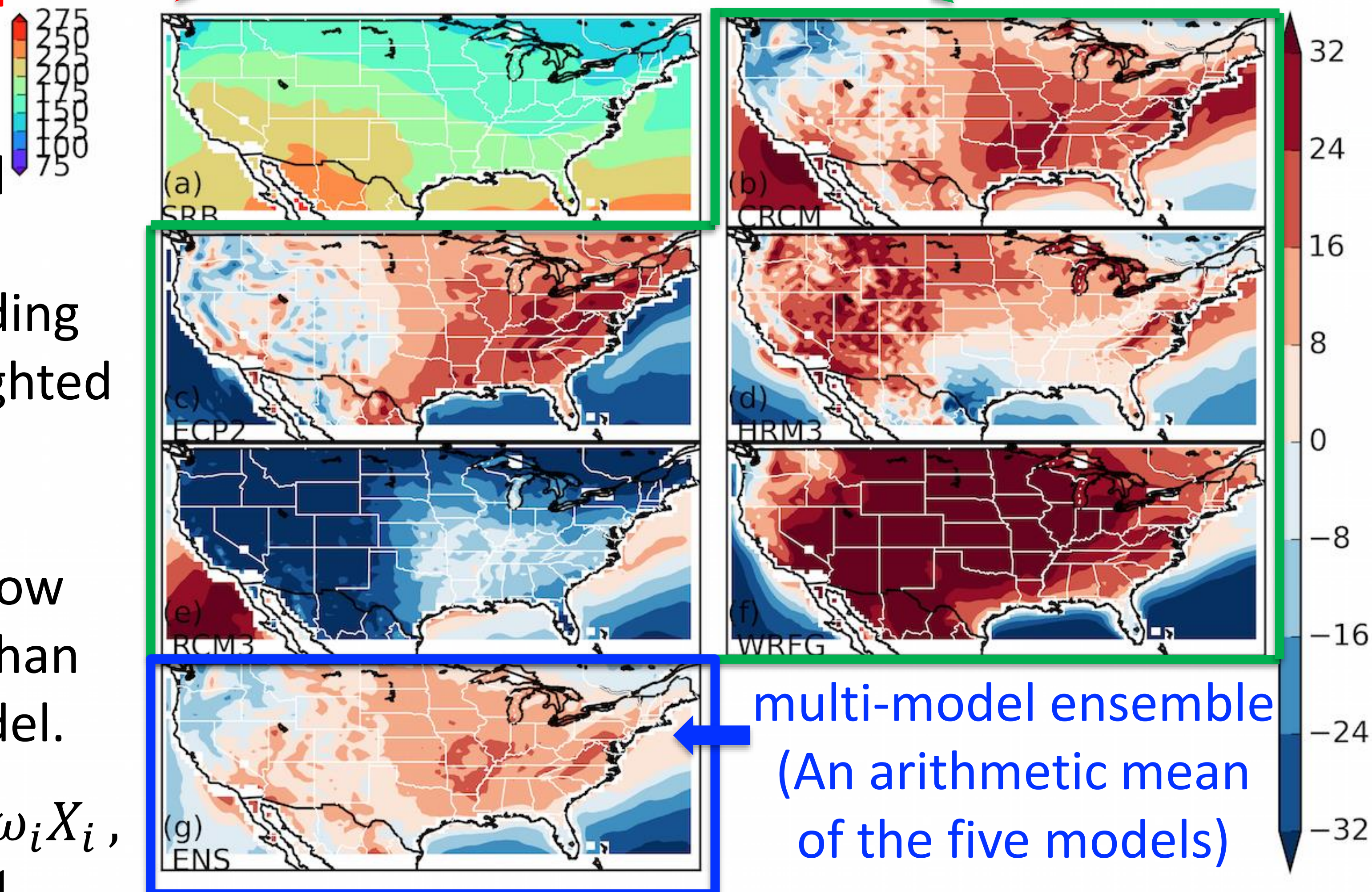
NASA's Computational Modeling Algorithms and Cyberinfrastructure program will support these two projects to further improve RCMES.

  - Combining multiple metrics into one: multi-objective optimization
  - Multi-resolution investigation of climate models

Evaluation of the annual-mean surface insolation using RCMES

satellite-based observation

five models

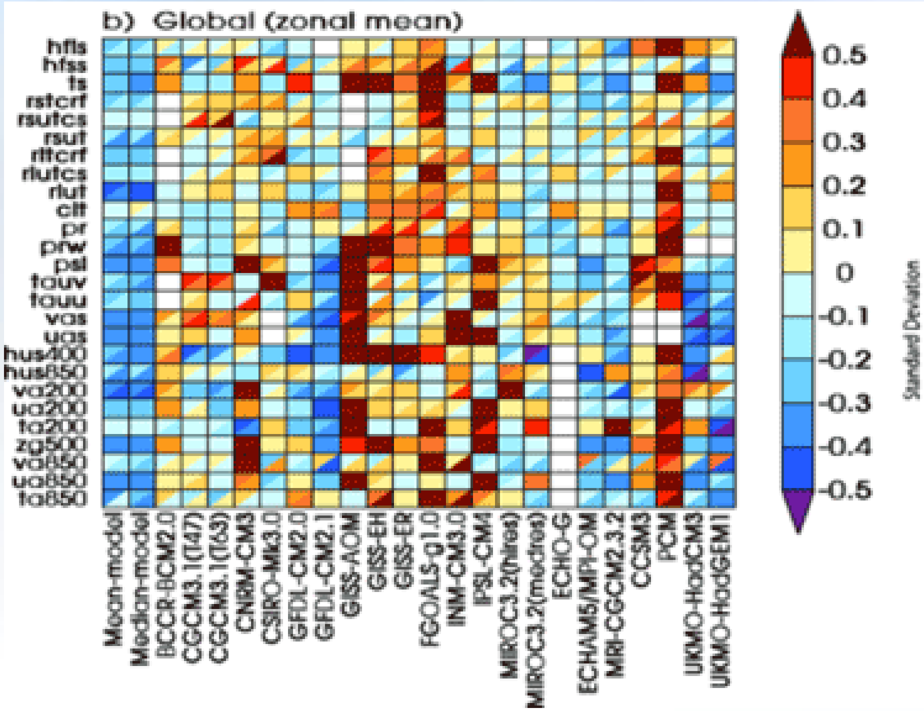


multi-model ensemble (An arithmetic mean of the five models)

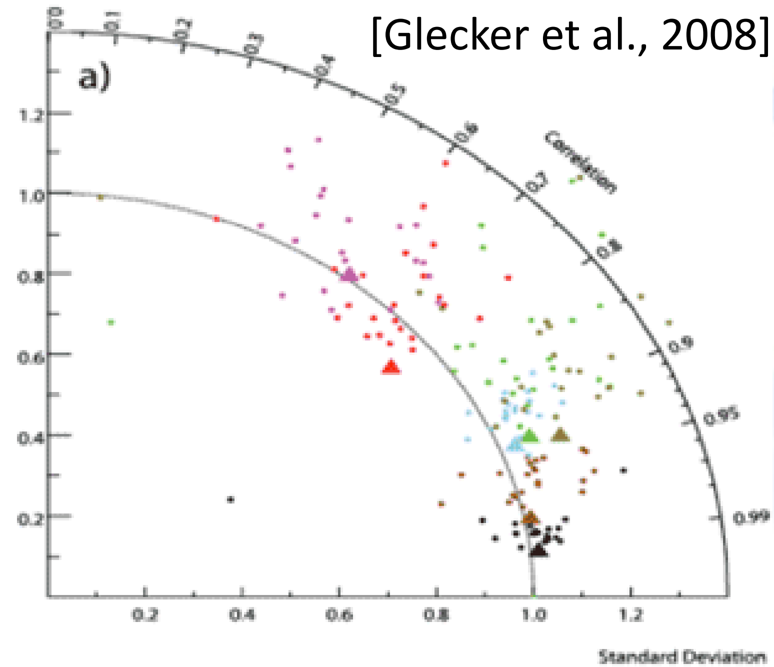
Multi-model ensembles (*ENS*), including equally weighted multi-model multi-model averages, generally show better skill than a single model.

$ENS = \sum_{i=1}^{n} \omega_i X_i$, $\sum_{i=1}^{n} \omega_i = 1$

How to combine multiple evaluation metrics into one?



[Glecker et al., 2008]



- ~~The more metrics, the better.~~
- Trade-offs between metrics
- Representativeness and meaningfulness of the combined metrics is key.

Conventional approach: evaluation => multi-model ensemble

n climate models	Evaluation metric
Model 1: $M_1(x,y,t)$	$Score_1$
Model 2: $M_2(x,y,t)$	$Score_2$
Model 3: $M_3(x,y,t)$	$Score_3$
.	
.	
.	
Model *n*: $M_n(x,y,t)$	$Score_n$

a weighted multi-model ensemble

$$Ens(x,y,t) = \sum_{i=1}^{n} \omega_i M_i(x,y,t)$$

$$\sum_{i=1}^{n} \omega_i = 1$$

This is a process of building a multi-model ensemble based on model evaluation result.

ex) If Model 3 shows better performance against observations than Model 2, $\omega_3 > \omega_2$.

Evaluation of multi-model ensembles

- Let's assume that we evaluate model to find **the best weighted multi-model ensemble**.
- [$\omega_1, \omega_2, \ldots, \omega_n$] are random numbers between 0 and 1.

$$Ens(x,y,t) = \sum_{i=1}^{n} \omega_i M_i(x,y,t) \qquad \sum_{i=1}^{n} \omega_i = 1$$

- By changing [$\omega_1, \omega_2, \ldots, \omega_n$] , we can calculate scores of multiple evaluation metrics for each *Ens(x,y,t)*.

	Metric 1	**Metric 2**	**…**	**Metric k**
$0.8*M_1 + 0.15*M_2 + 0.05*M_3$	0.2	0.3	…	0.6
$0.3*M_1 + 0.60*M_2 + 0.1*M_3$	0.5	0.1	…	0.5
…	…	best scores		…
0.1*M_1 + **0.2***M_2+**0.7***M_3	0.9	0.8	…	0.7

0.1, 0.2 and 0.7 could be interpreted as combined metrics for the three models.

How to Estimate $[\omega_1^*, \omega_2^*, \ldots, \omega_n^*]$ (weighting factors for the best ensemble)? => Multi-Objective optimization

- The objective optimization is a mathematical process of finding the best solution involving objective functions to be optimized simultaneously.

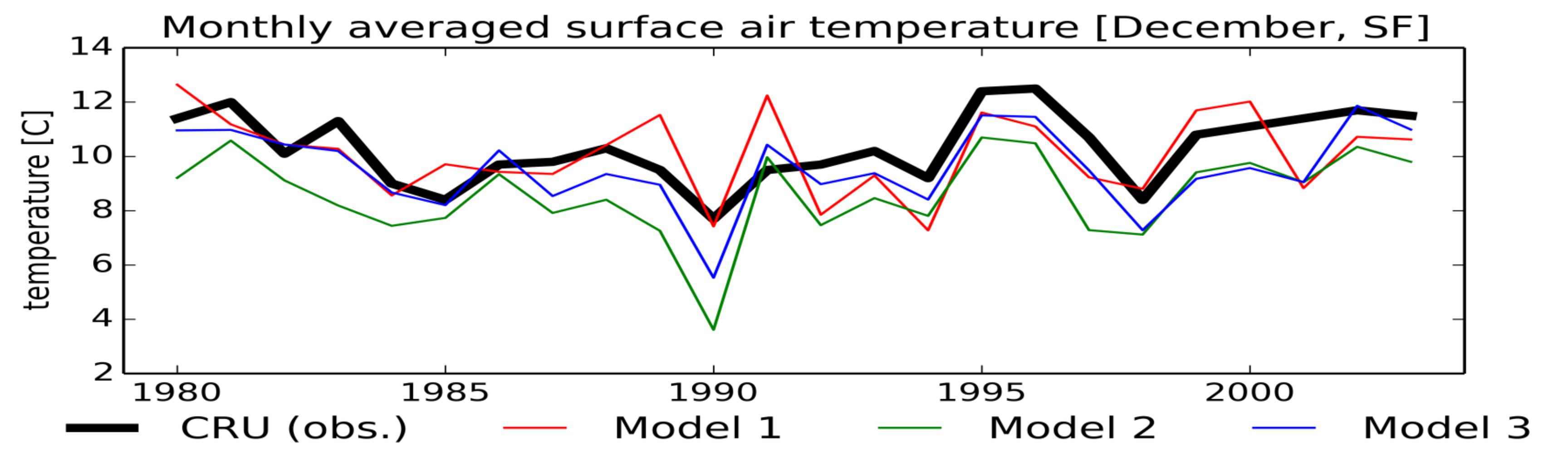
- Multi-objective optimization has been applied in many fields of science, economics and engineering.

- In climate model evaluation,
  - solution: weighting factors
  - objective functions: model evaluation metrics

from http://www.noesissolutions.com/

In most of cases, there does not exist a single solution. In general, there exist a number of Pareto optimal solutions.

Example) near surface air temperature in San Francisco [1980-2003, monthly averaged in December]

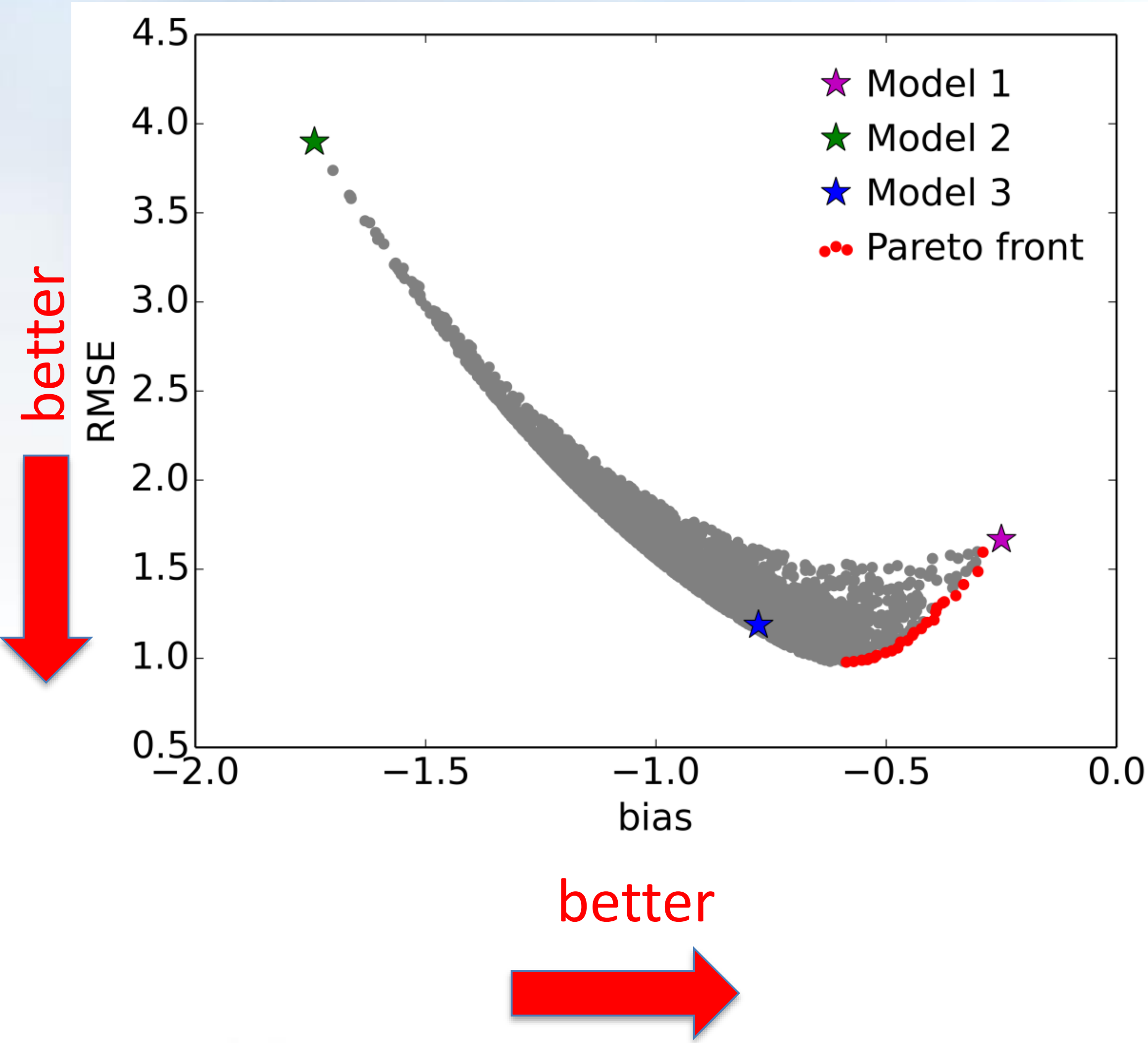


	Metric 1: Bias	Metric 2: RMSE
Model 1 (M1)	**-0.25**	1.67
Model 2 (M2)	-1.74	3.90
Model 3 (M3)	-0.78	**1.19**

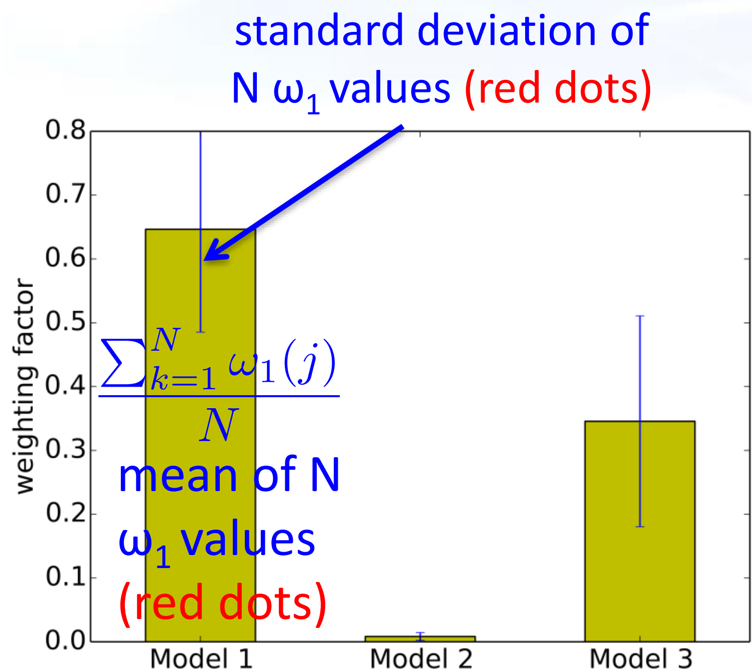
- What about the bias and RMSE of weighted ensemble time series?

$$Ens(t) = \omega_1 \times M1(t) + \omega_2 \times M2(t) + \omega_2 \times M3(t)$$

Optimized (minimized |bias| and RMSE) Weighting Factors [ω_1, ω_2, ω_3]



- Each grey point represents the bias and RMSE of one combination of the three models.



Big Data challenges of the optimization

- The bias in the example is a linear metric.

For example, a bias of a model ensemble ($ENS = \sum_{i=1}^{n} \omega_i M_i$) is same as weighted average of each model (M_i)'s bias.

- However, most of popularly used metrics are not linear. So the metric scores for each simulation, valuable information as they are, can not be used to find the best (optimized) ensemble.

Model	Metric 1	Metric 2	…	Metric k
M_1	0.9	0.3	…	0.4
M_2	0.5	0.5	…	0.3
…	…	…	…	…
M_n	0.95	0.8	…	0.5

Ens.	Metric 1	Metric 2	…	Metric k
Ens 1	0.4	0.2	…	0.1
Ens 2	0.3	0.3	…	0.7
…	…	…	…	…
Ens N	0.6	0.6	…	0.4

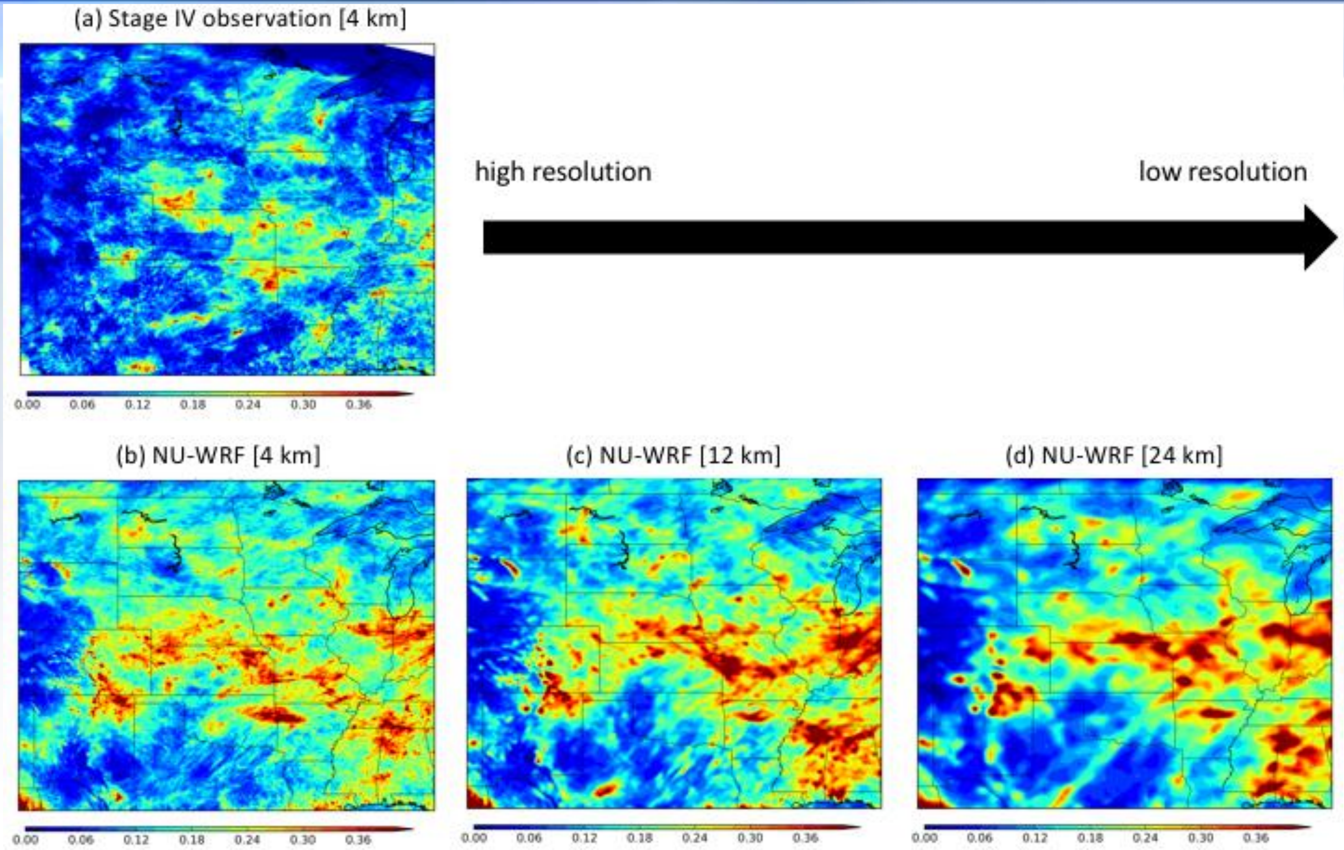
Outline

- Introduction
- Systematic evaluation of CORDEX RCMs using the Regional Climate Model Evaluation System
- Use cases of RCMES
  - Evaluation of hourly rainfall characteristics in RCM simulations
- Toward the model evaluation of the future

NASA's Computational Modeling Algorithms and Cyberinfrastructure program will support these two projects to further improve RCMES.

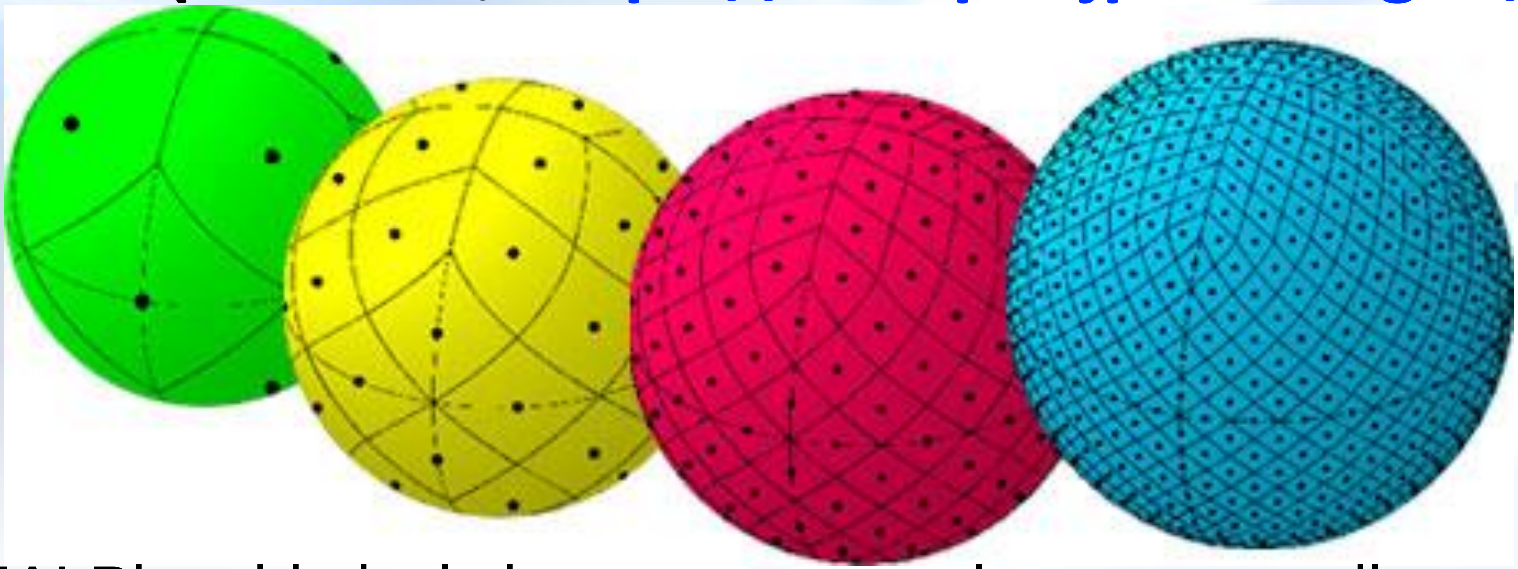
  - Multi-resolution investigation of climate models
  - Combining multiple metrics into one: multi-objective optimization

Motivation



There is a strong need for the development and application of infrastructure for observation-based evaluations of spatial patterns in key climate variables simulated with various spatial resolutions.

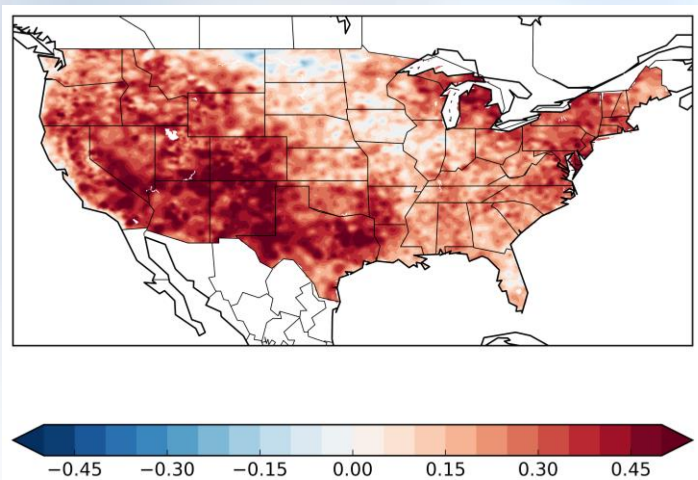
Hierarchical Equal Area isoLatitude Pixelization (HEALPix, https://healpix.jpl.nasa.gov/)



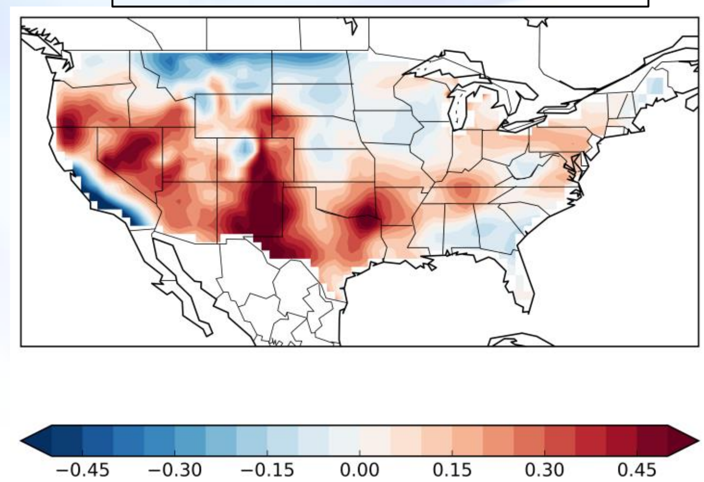
- All HEALPix grid pixels have exact equal areas regardless of their geographical locations and resolutions.
- Geographical coordinates of HEALPix pixels are fixed, and higher resolution pixels are hierarchically nested within lower resolution pixels.
- Generating low-resolution data from high-resolution data on HEALPix grids is a simple and efficient averaging process that utilizes their unique hierarchical data structure.

Case study: fine-scale (< 10 km) variability of long-term temperature trends

observed temperature trend in summer 1979-2012 [K/decade] (resolution 5 km)

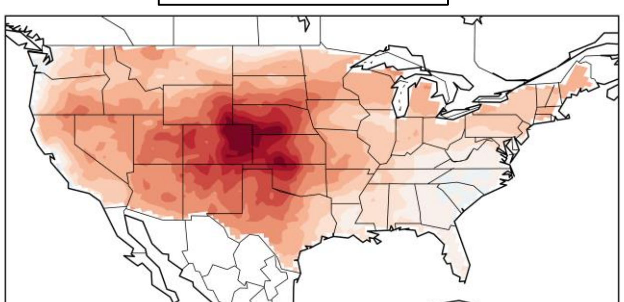


ERA interim (80 km)

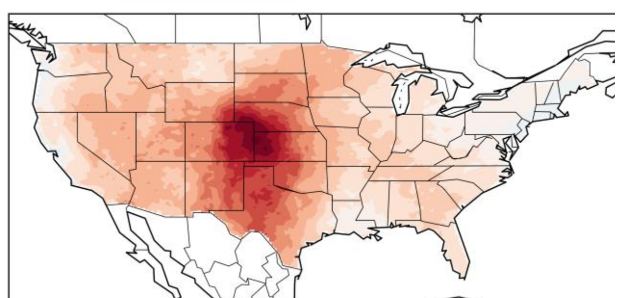


downscaled by CRCM

(44 km)



(22 km)



(11 km)