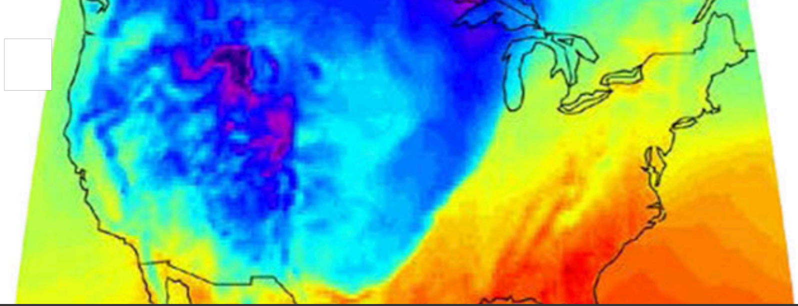
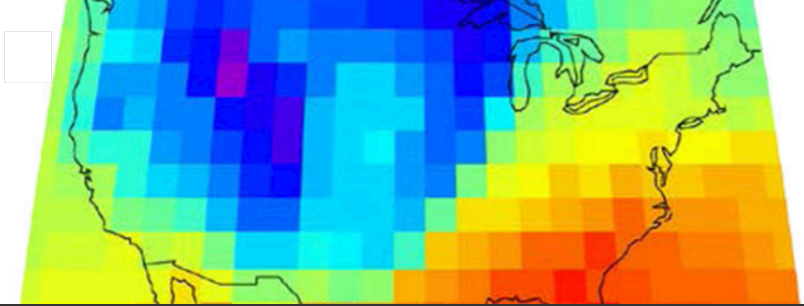



Comparing the performance of multiple statistical downscaling approaches using a perfect model framework

Anne Stoner[1], Katharine Hayhoe[1], Keith Dixon[2], John Lanzante[2], and Ian Scott-Fleming[1]

[1] Climate Science Center, Texas Tech University, Texas, USA

[2]Geophysical Fluid Dynamics Laboratory, New Jersey, USA

The Statistical Downscaling Conundrum

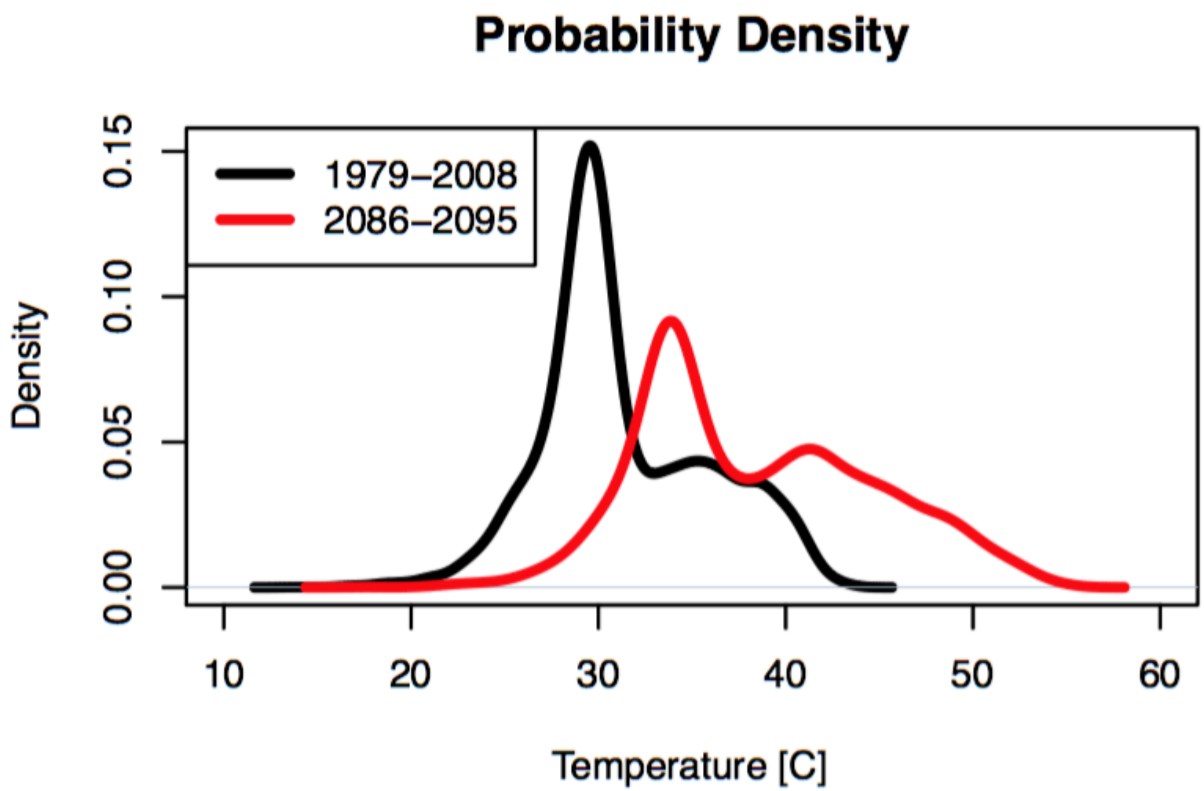
- Most Empirical Statistical Downscaling Methods (ESDMs) are developed and applied after a cursory evaluation focusing on historical periods, to make sure the output is reasonable.
- During development, ESDMs are not usually optimized to meet performance metrics but rather to produce a certain product, given specific inputs
- For decades, the field has lacked a standardized comparison tool to evaluate and compare different SDMs on an equal footing -> both during and after evaluation

Stationarity assumption

One critical assumption implicit to all ESD methods is that of **statistical stationarity**, which presumes the statistical relationships between GCM output and observed climate data remain constant over time.

Stationarity assumption

One critical assumption implicit to all ESD methods is that of **statistical stationarity**, which presumes the statistical relationships between GCM output and observed climate data remain constant over time.



ESDMs are traditionally evaluated on the historical period.

- By dividing the historical record up into training and validation periods, producing downscaled output for a period that can be compared with independent observations
- But how do we know if the model will perform equally well on future data? Will the stationarity assumption hold?

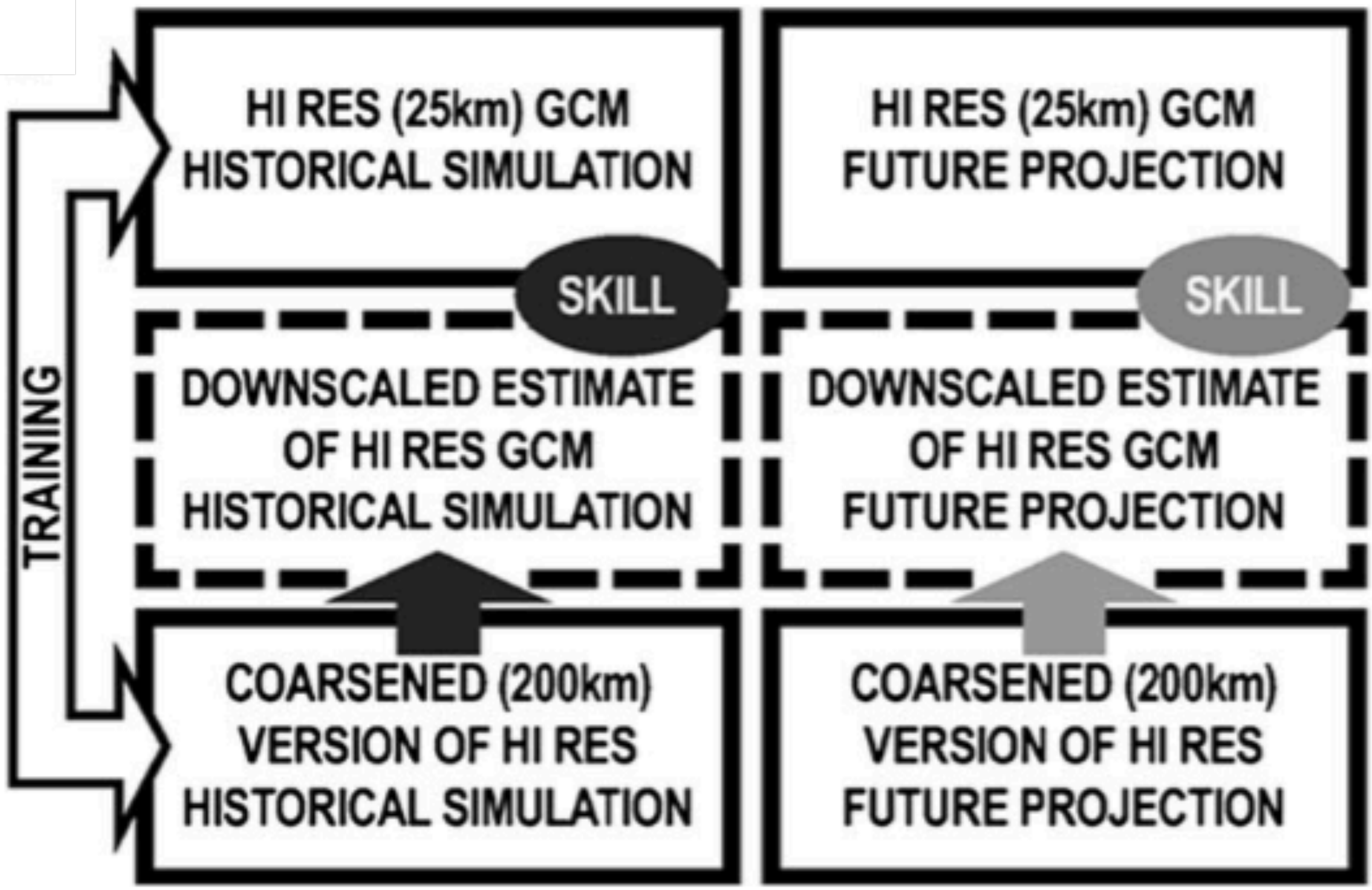
ESDMs are traditionally evaluated on the historical period.

- By dividing the historical record up into training and validation periods, producing downscaled output for a period that can be compared with independent observations
- But how do we know if the model will perform equally well on future data? Will the stationarity assumption hold?

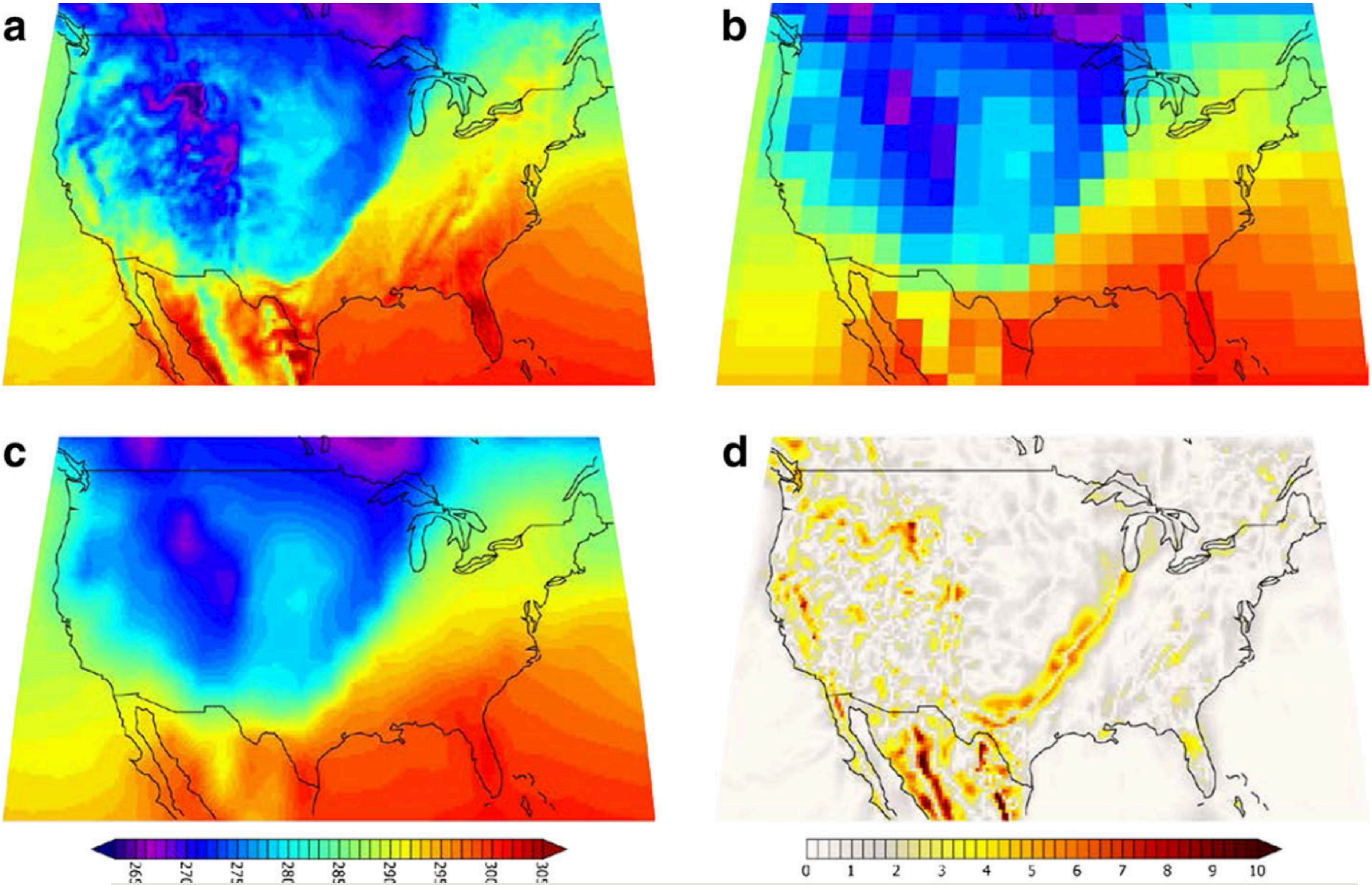
We have developed a new "perfect model" framework to evaluate ESDMs and tests for stationarity *in the future.*

- Using high-resolution GCM output as "observations" for the future, the ESDM is trained on historical high-resolution output and generates future projections that are compared with the dynamically-generated high-resolution projections.

“Perfect model” experimental design



“Perfect model” experimental design



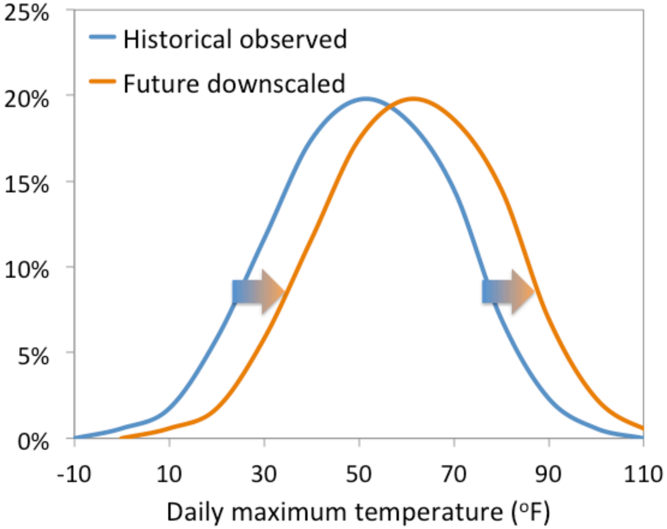
What are we using this framework to do?

FIRST, we coded up 3 frequently-used ESDMs (delta, quantile mapping, and ARRM), and compared their ability to reproduce high-resolution GCM simulations for the future, for a range of temperature and precipitation quantiles and impact-relevant thresholds.

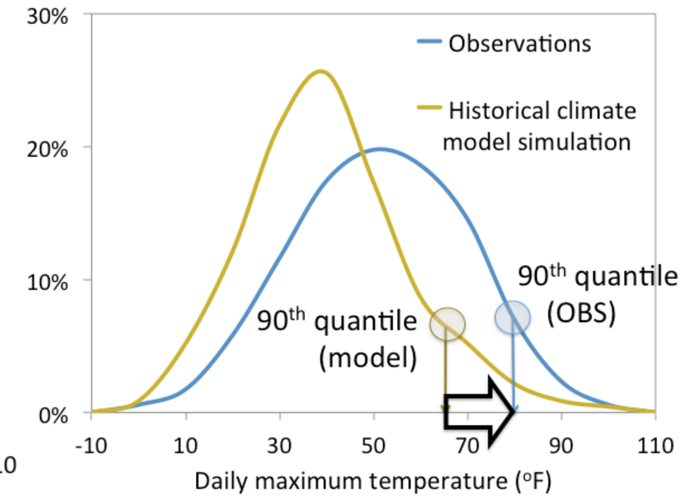
Our comparison shows that simple methods can be adequate for some purposes, but complex methods are necessary for extremes.

Even complex methods can have problems in areas of rapidly varying topography, such as coastlines.

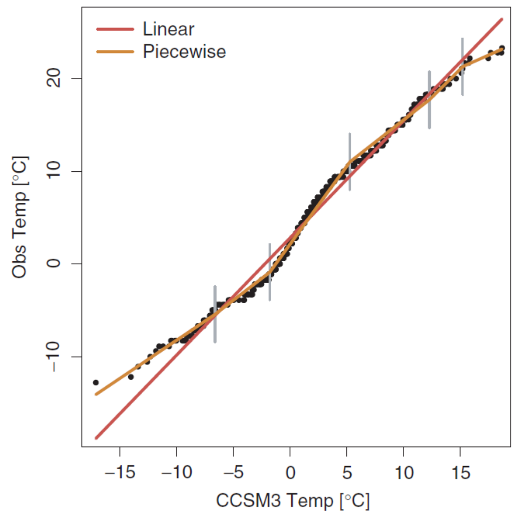
STEP ONE: Comparing 3 Existing ESDMs



ONE. THE DELTA METHOD

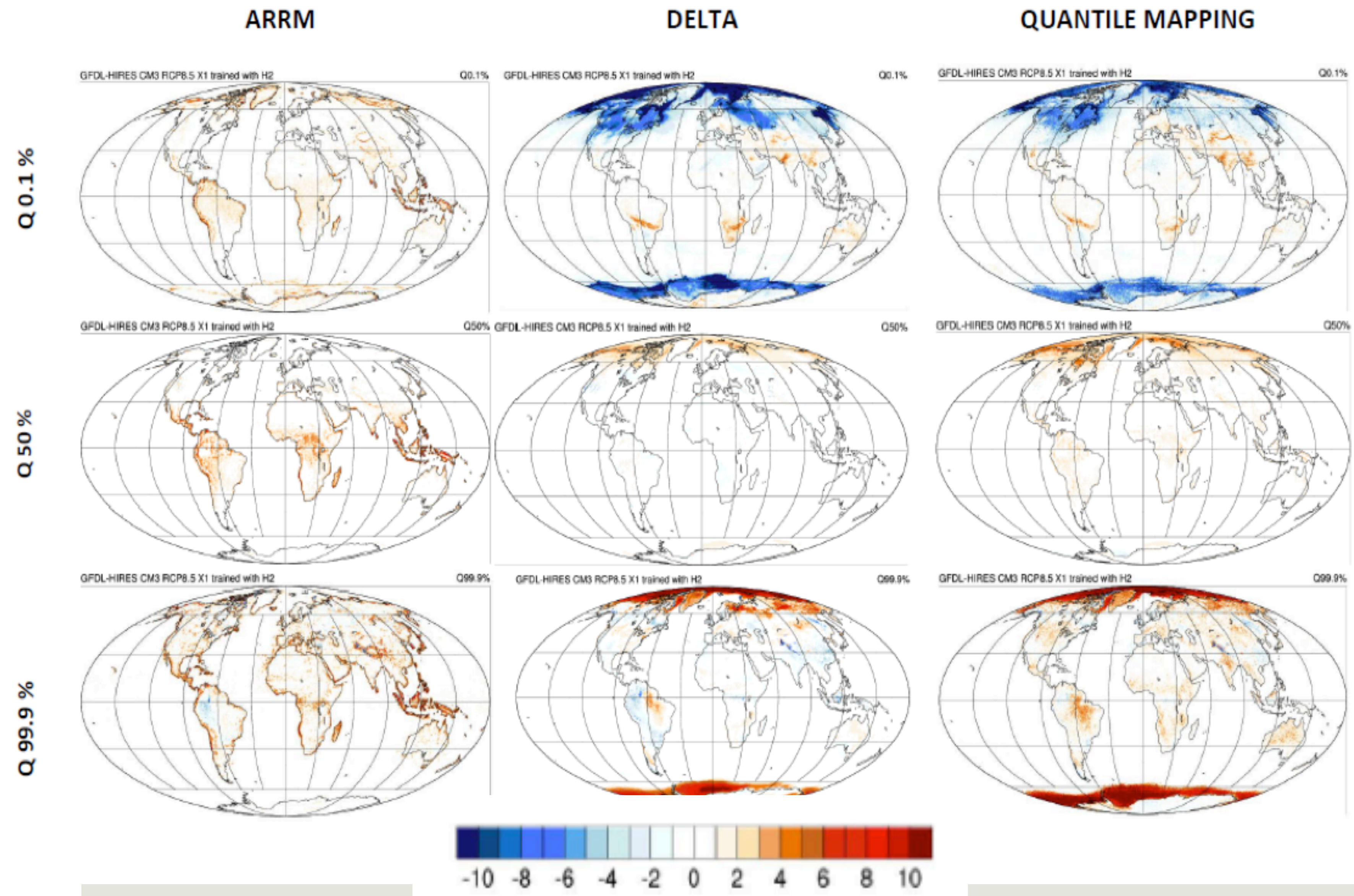


TWO. EMPIRICAL QUANTILE MAPPING (BCSD)

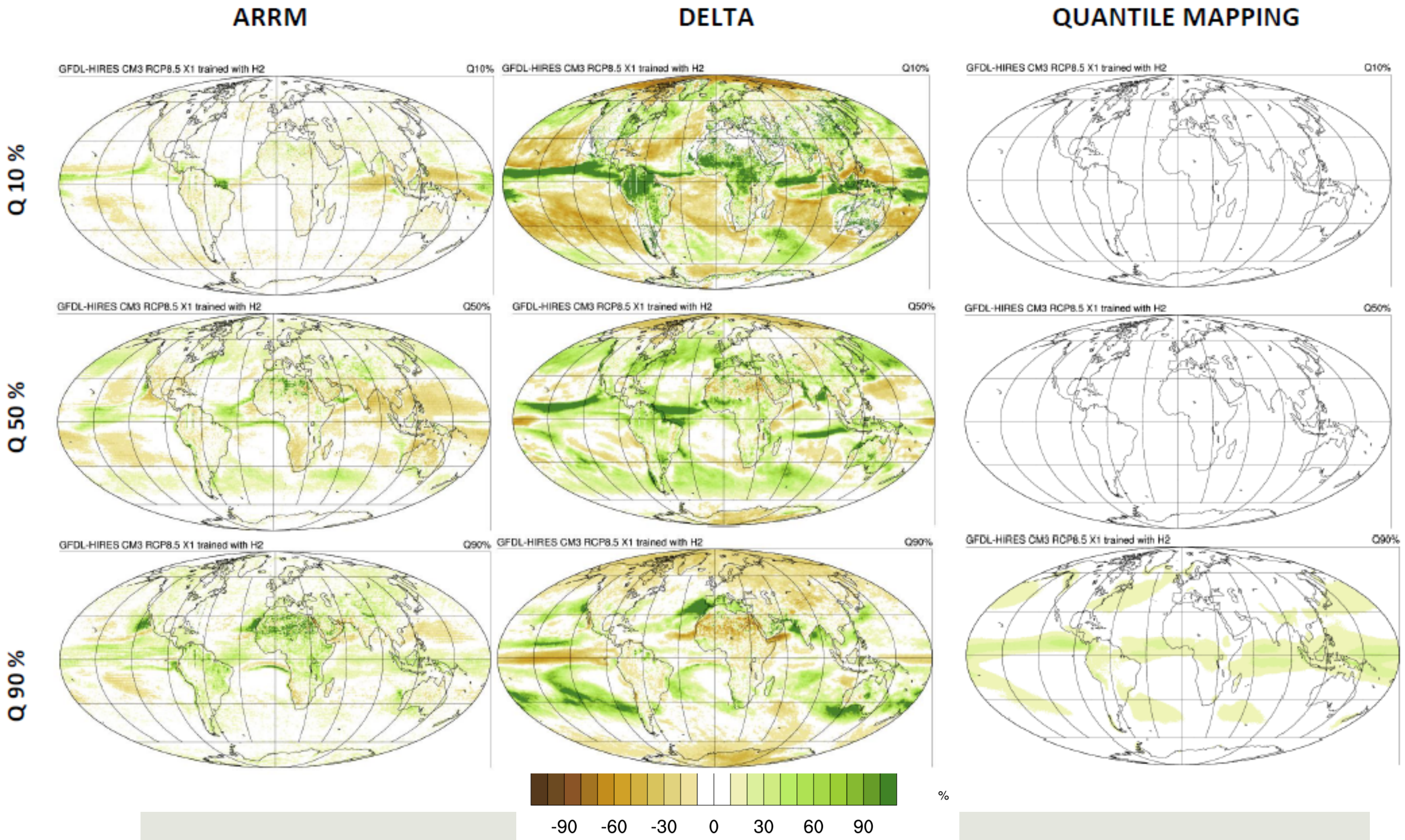


THREE. PARAMETRIC QUANTILE MAPPING (ARRM)

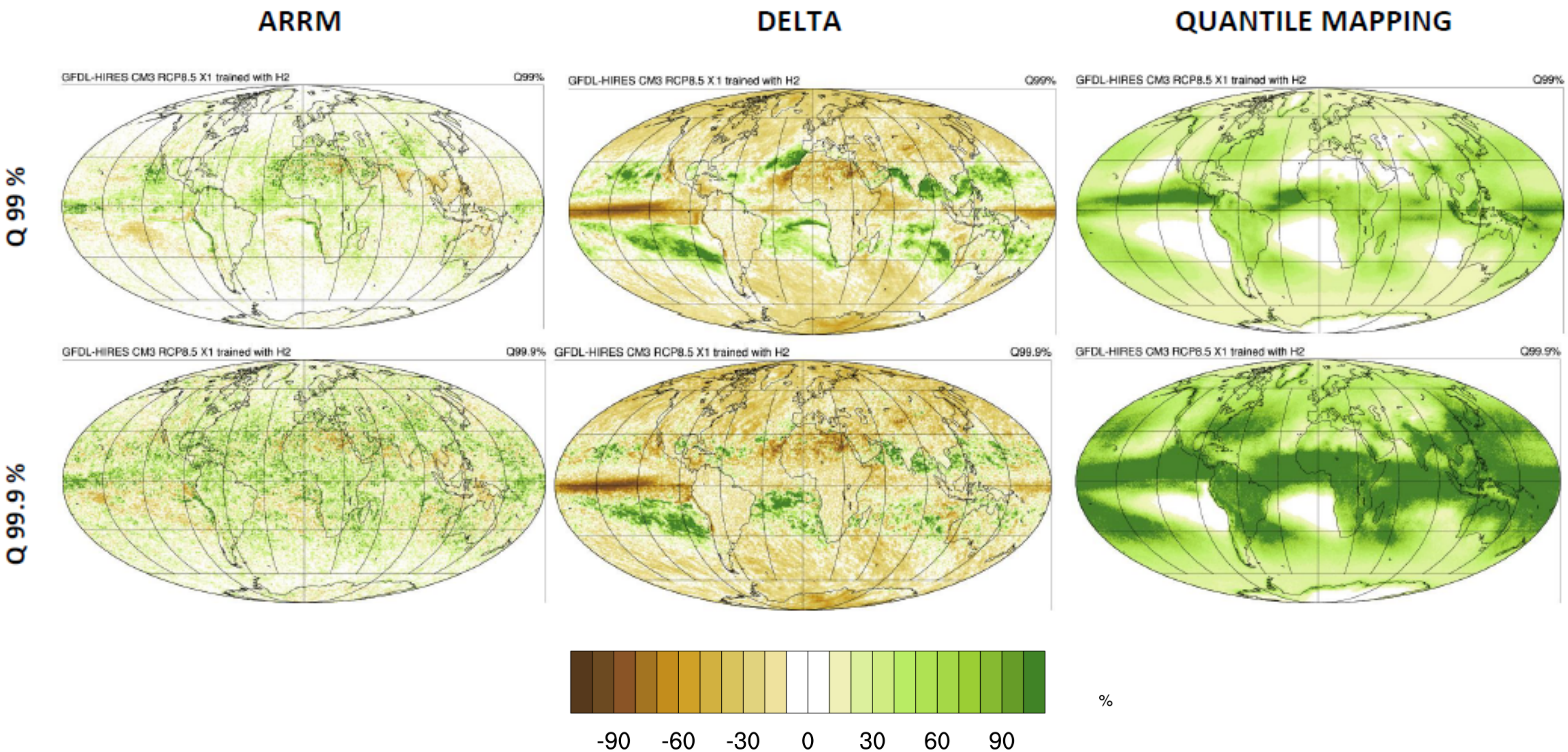
STEP ONE: Comparing 3 Existing ESDMs
Daily Maximum Temperature Bias 2086-2095



RESULTS: Daily Wet-Day Not-So-Extreme Precipitation Bias in 2086-2095



RESULTS: Daily Wet-Day Extreme Precipitation Bias in 2086-2095



What are we using this framework to do?

SECOND, we just finished a brand-new non-parametric ESDM -- using the perfect model framework to evaluate and test the statistical methods and physical assumptions being made at every step along the way.

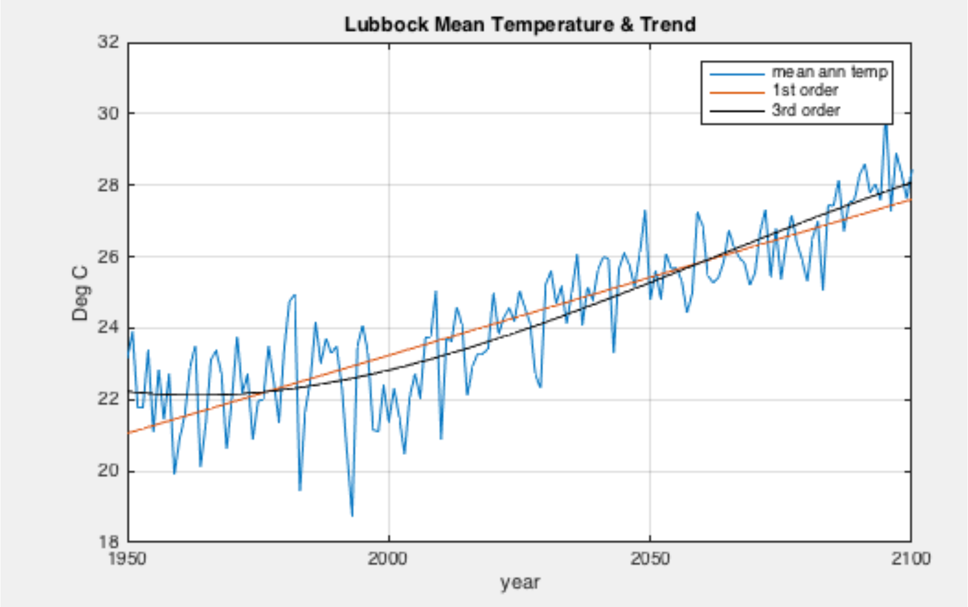
This process greatly increased the complexity of the new ESDM, but also its global relevance, as each aspect of the model was generalized to produce stationarity at any location, from Mt Everest to the Amazon.

The end result is ~100x more computationally efficient, with significantly reduced biases.

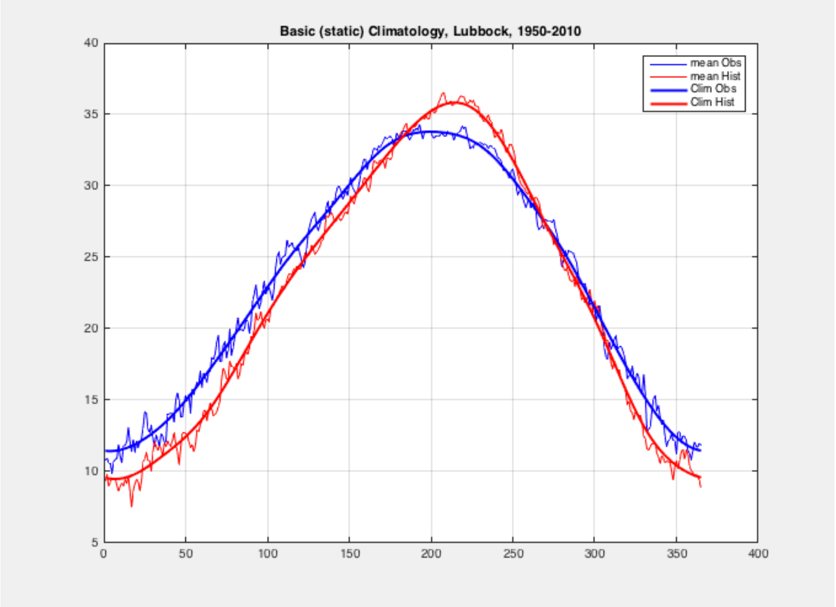
STEP TWO: Developing New ESDM

The new non-parametric kernel density model (ARRM2) separates the signal into 4 separate components, and models each differently:

ONE. Long-term trend



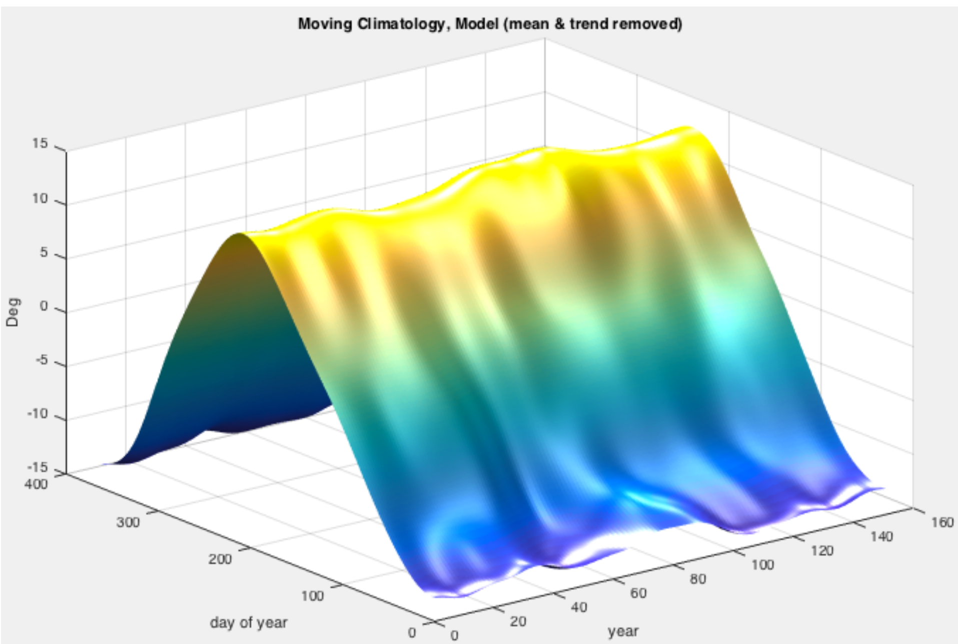
TWO. Static annual climatology



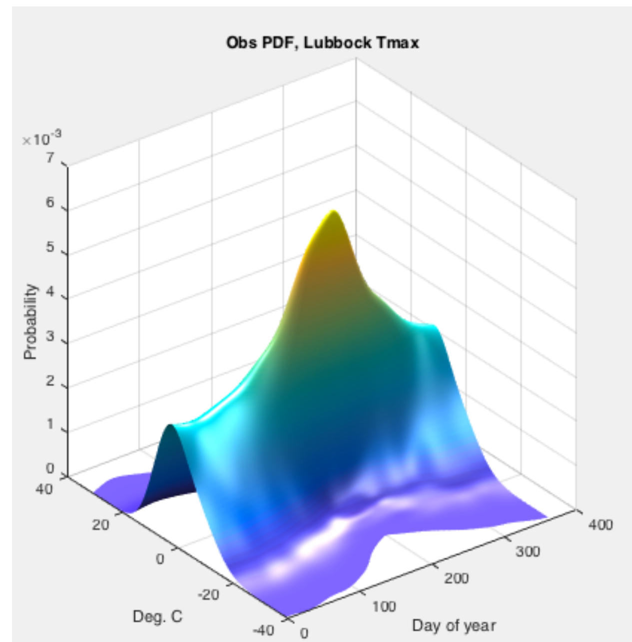
STEP TWO: Developing New ESDM

The new non-parametric kernel density model (ARRM2) separates the signal into 4 separate components, and models each differently:

THREE. Dynamic climatology



FOUR. High-frequency anomalies



STEP TWO: Developing New ESDM

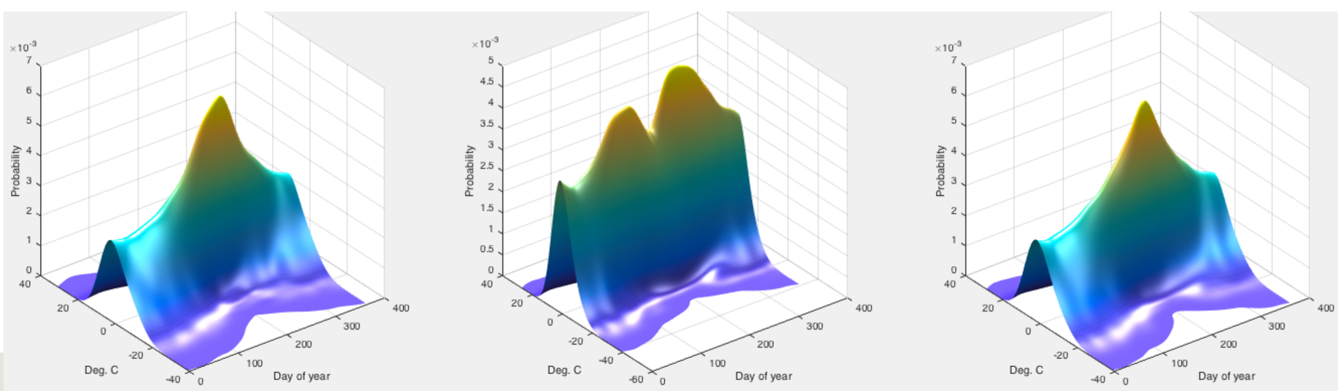
The model builds mappings between observations and GCM for the individual components.

KDE (kernel density estimation) is used to estimate underlying PDFs.

QQ (quantile-quantile) mapping creates a continuous mapping surface between 2 PDF distribution surfaces.

- ARRMv1 mapped on 12 independent monthly PDFs
- ARRMv2 maps on 365 correlated daily PDFs

OBSERVED HISTORICAL **MODEL HISTORICAL** **DOWNSCALED HISTORICAL**

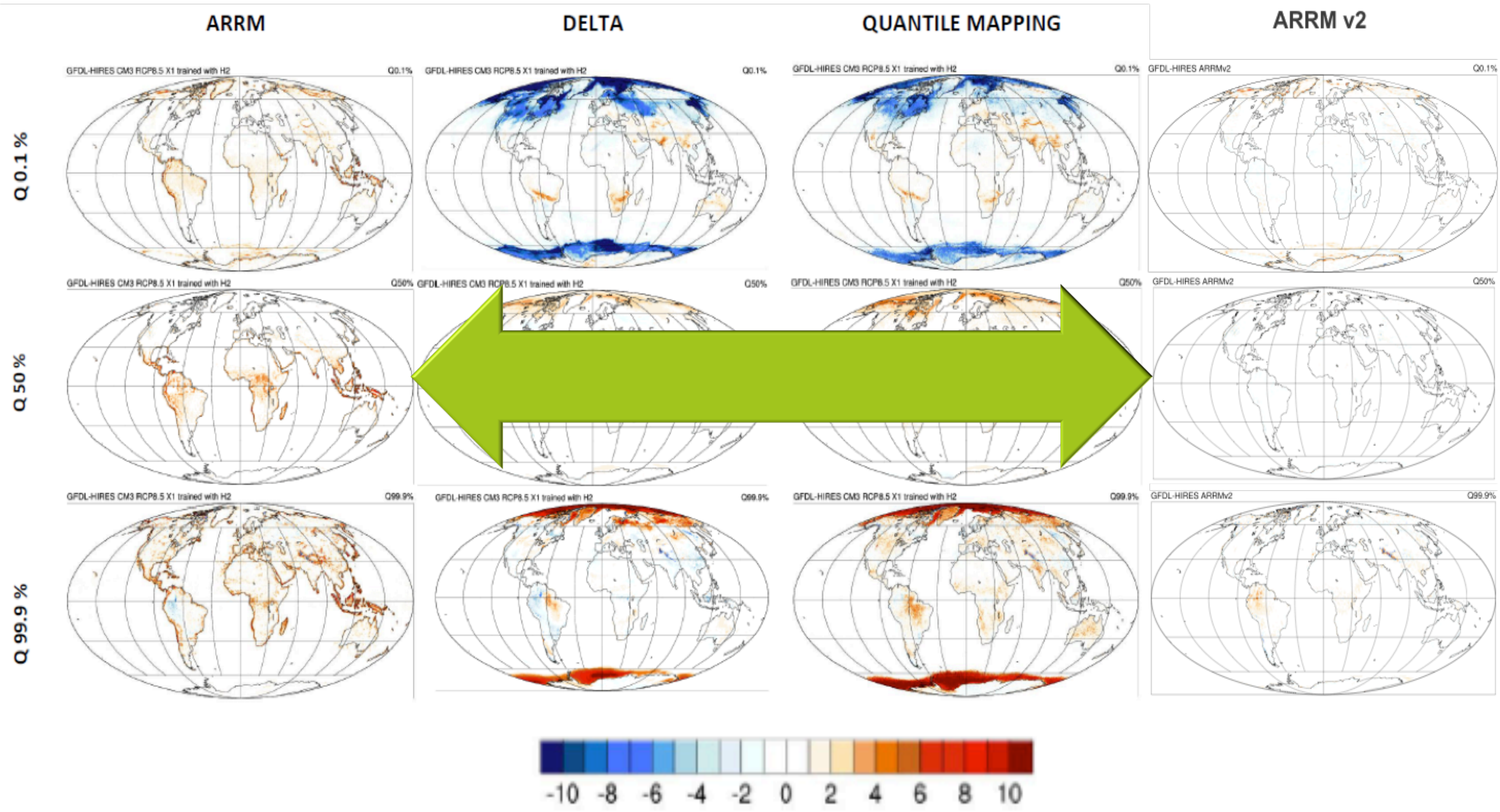


STEP TWO: Developing New ESDM

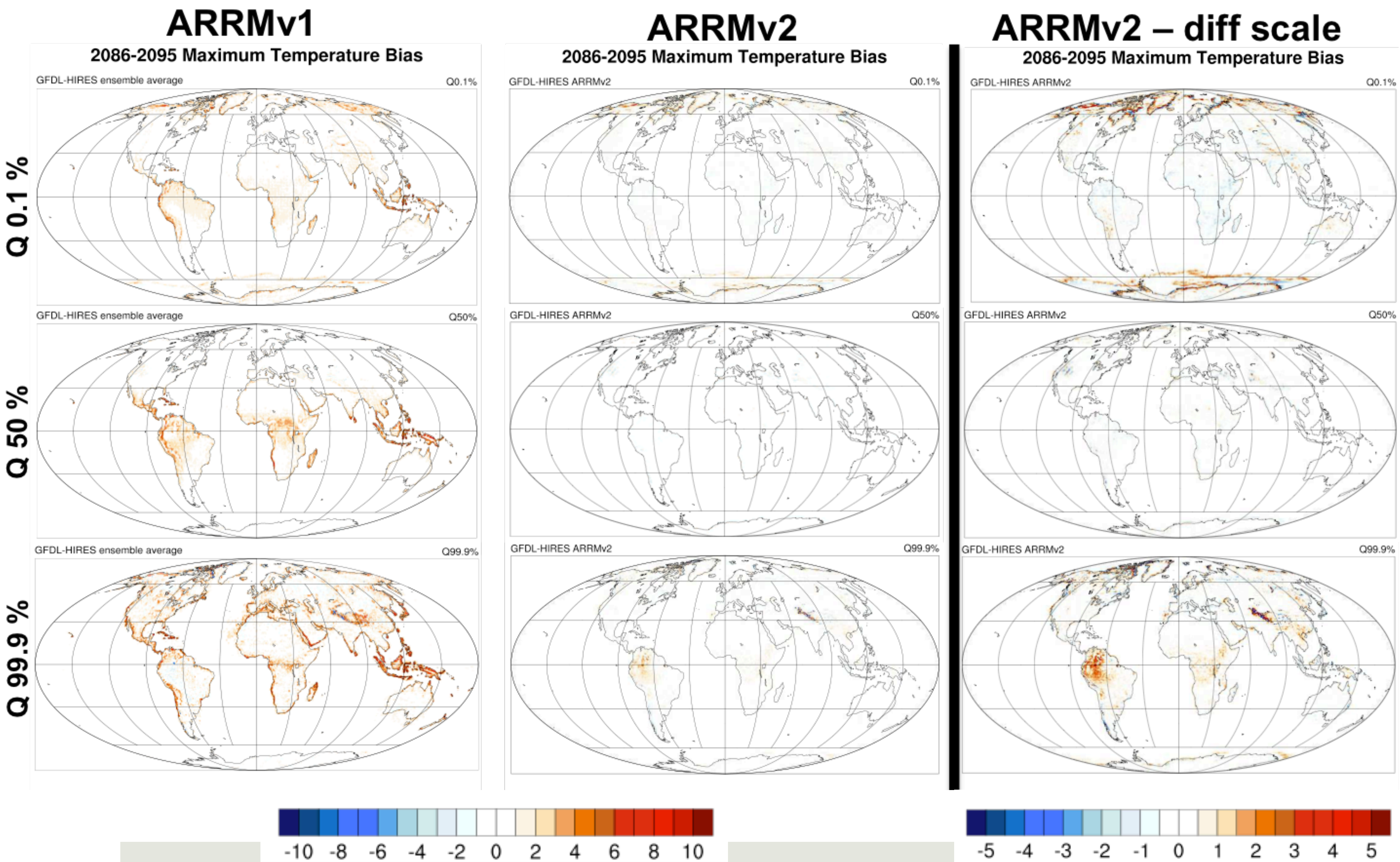
The final step is to recombine the four adjusted components to construct an estimate of the future signal.

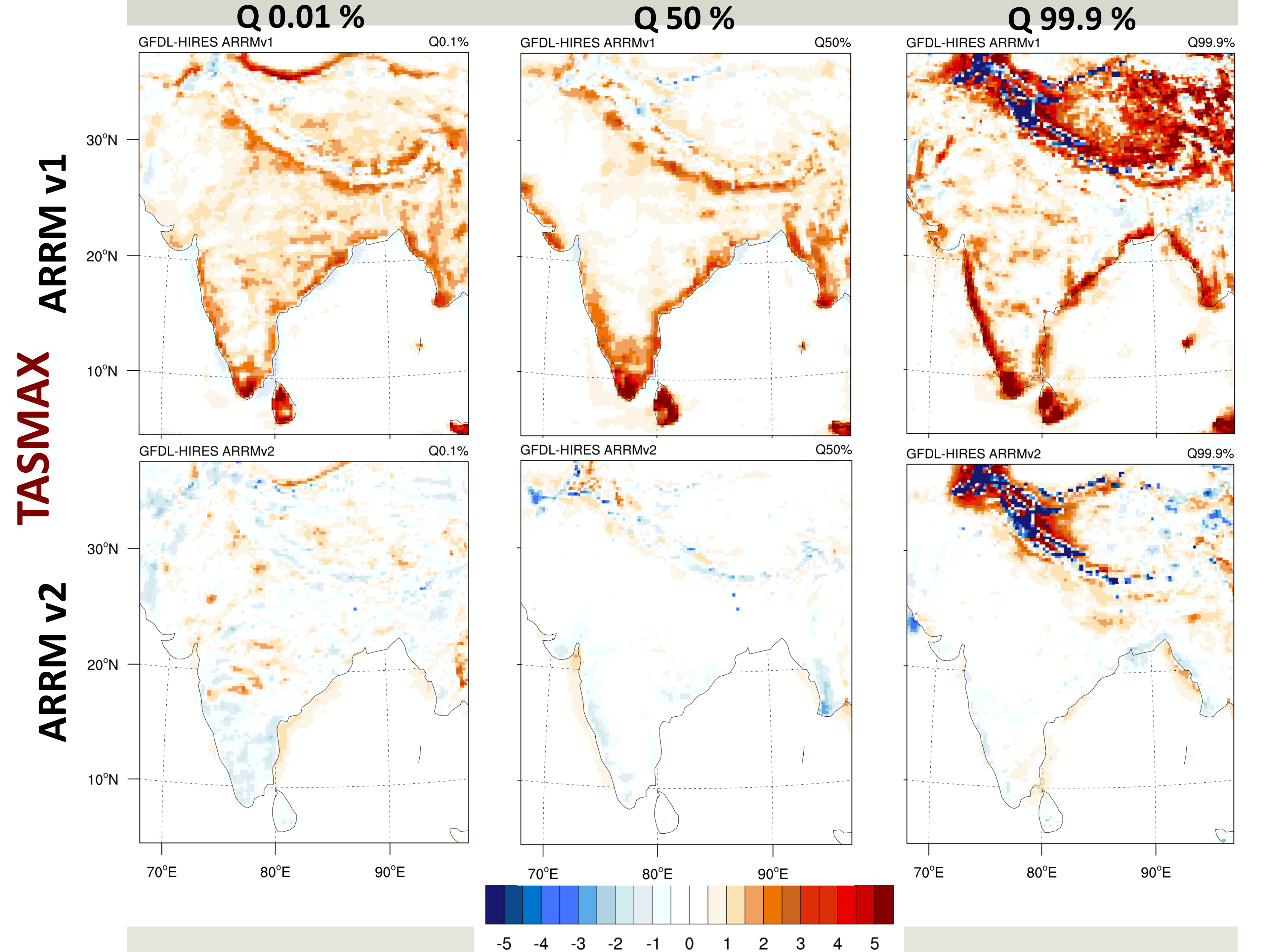
- Adjusted Long-term trend
  - LONG TERM TREND (MODEL) + (MEAN (OBS) – MEAN (HISTORICAL MODEL))
- Adjusted Climatology
  - DYNAMIC CLIMATOLOGY(MODEL) + (STATIC CLIMATOLOGY (OBS) – STATIC CLIMATOLOGY (MODEL))
- Adjusted Anomalies
  - Daily anomalies, with each day remapped so PDF surface of Model anomalies matches the PDF surface of Observations

RESULTS: Daily Maximum Temperature Bias in 2086-2095

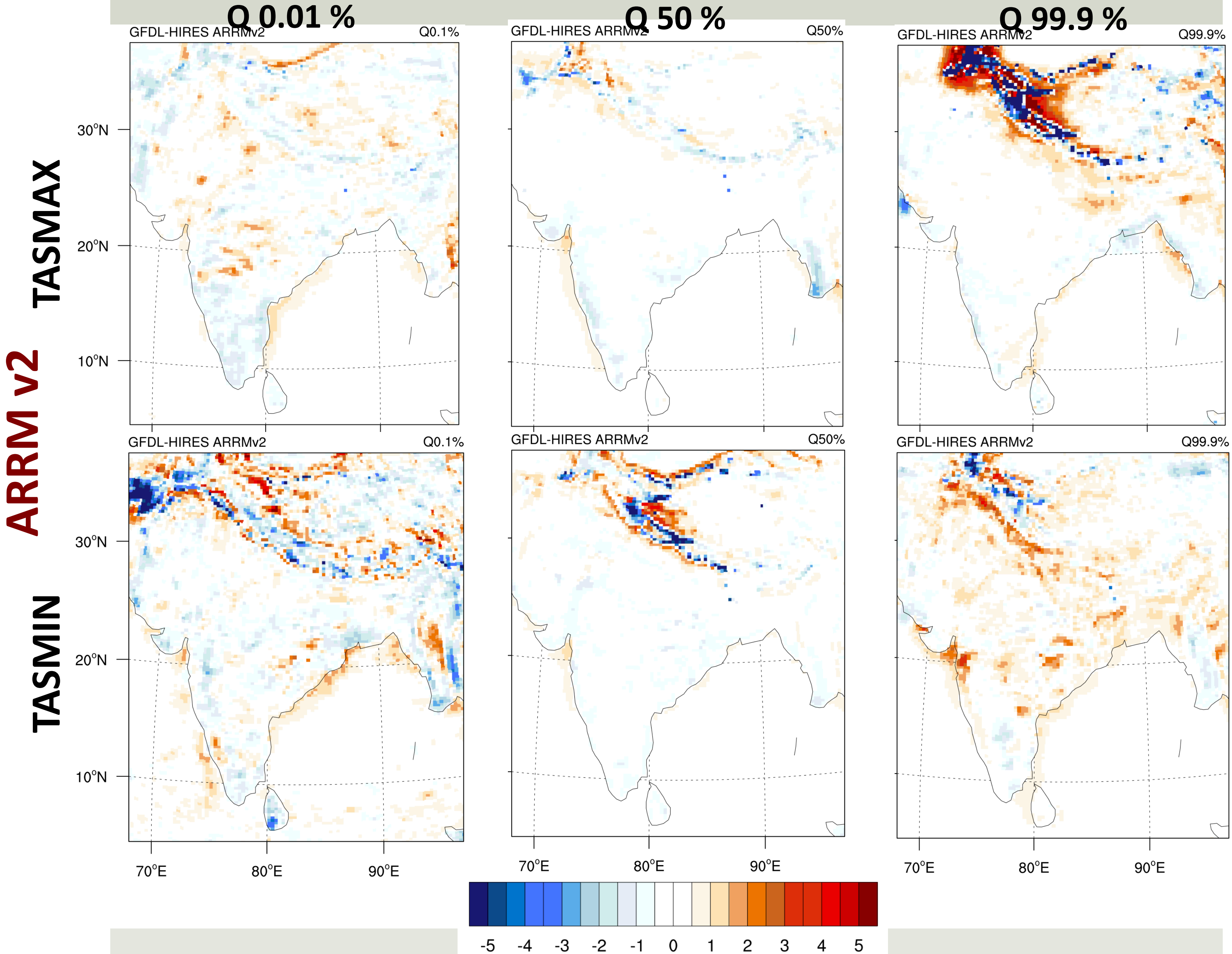


ARRM1 vs. ARRM 2 – does using the perfect model framework in development make a difference?

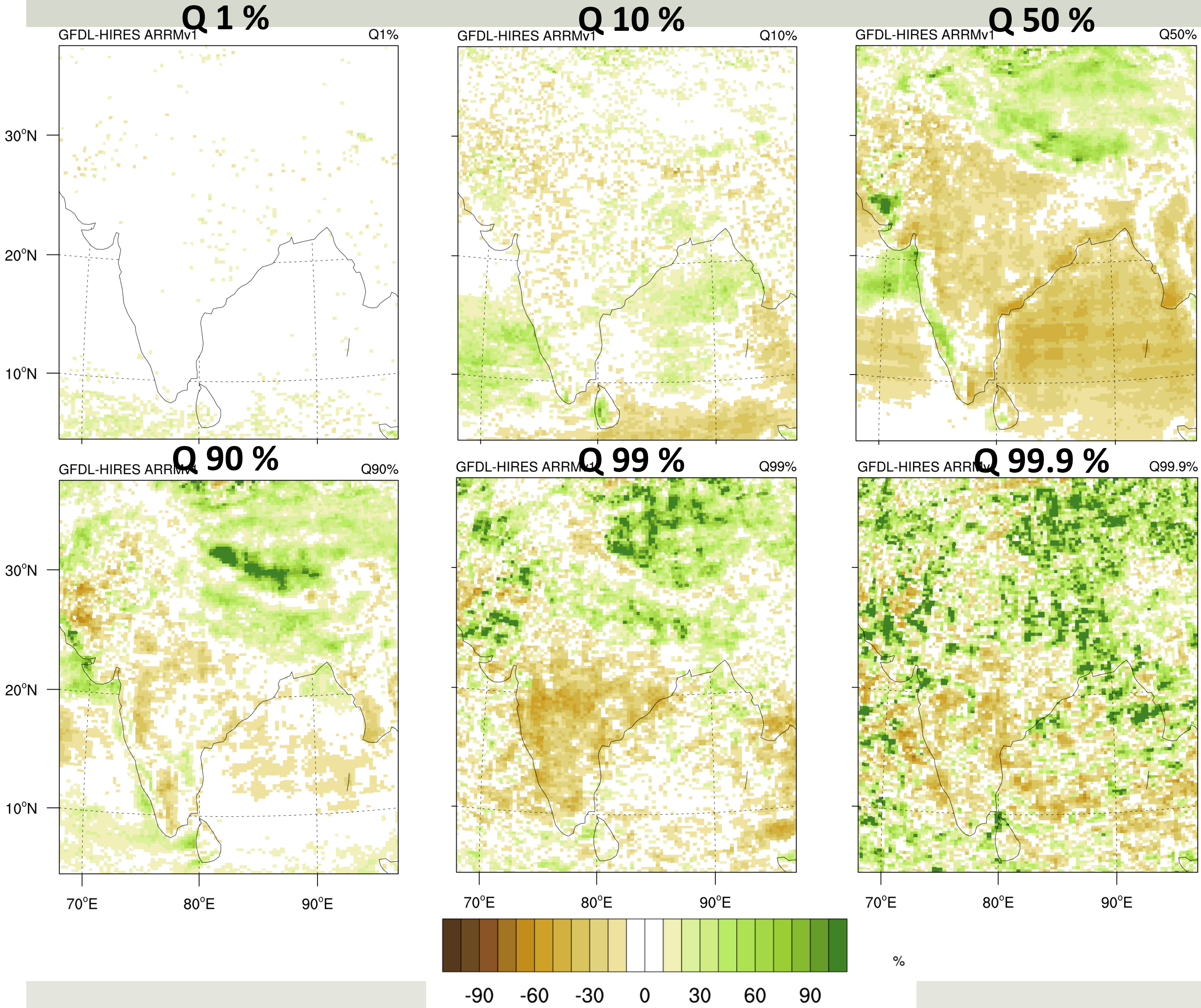




Q 0.01 %
Q 50 %
Q 99.9 %
ARRM v1
ARRM v2
TASMAX
GFDL-HIRES ARRMv1
Q0.1%
Q50%
Q99.9%
GFDL-HIRES ARRMv2
30°N
20°N
10°N
70°E
80°E
90°E
-5
-4
-3
-2
-1
0
1
2
3
4
5


Q 0.01 %
Q 50 %
Q 99.9 %
TASMAX
TASMIN
ARRM v2
GFDL-HIRES ARRMv2
Q0.1%
Q50%
Q99.9%
30°N
20°N
10°N
70°E
80°E
90°E
-5
-4
-3
-2
-1
0
1
2
3
4
5

ARRM v1 PRECIPITATION



Does the stationarity assumption hold?

DELTA

- Temperature: At mid to low latitudes, not at high latitudes
- Precipitation: In very few regions, definitely not in equatorial region

EMPIRICAL QUANTILE MAPPING (BCSD)

- Temperature: More so in mid to low latitudes than high latitudes
- Precipitation: Excellent until 90^{th} percentile, then rapid degradation

Does the stationarity assumption hold?

PARAMETRIC QUANTILE MAPPING (ARRM v1)

- Temperature: Better inland than in coastal regions and winter than summer
- Precipitation: Reasonable until the 99th percentile, less so in equatorial regions

NON-PARAMETRIC MAPPING (ARRM v2)

- Temperature: Shows improvement in many regions, particularly at the tails of the distribution
- Still a bias in very varied topographic regions (e.g. Himalayas)

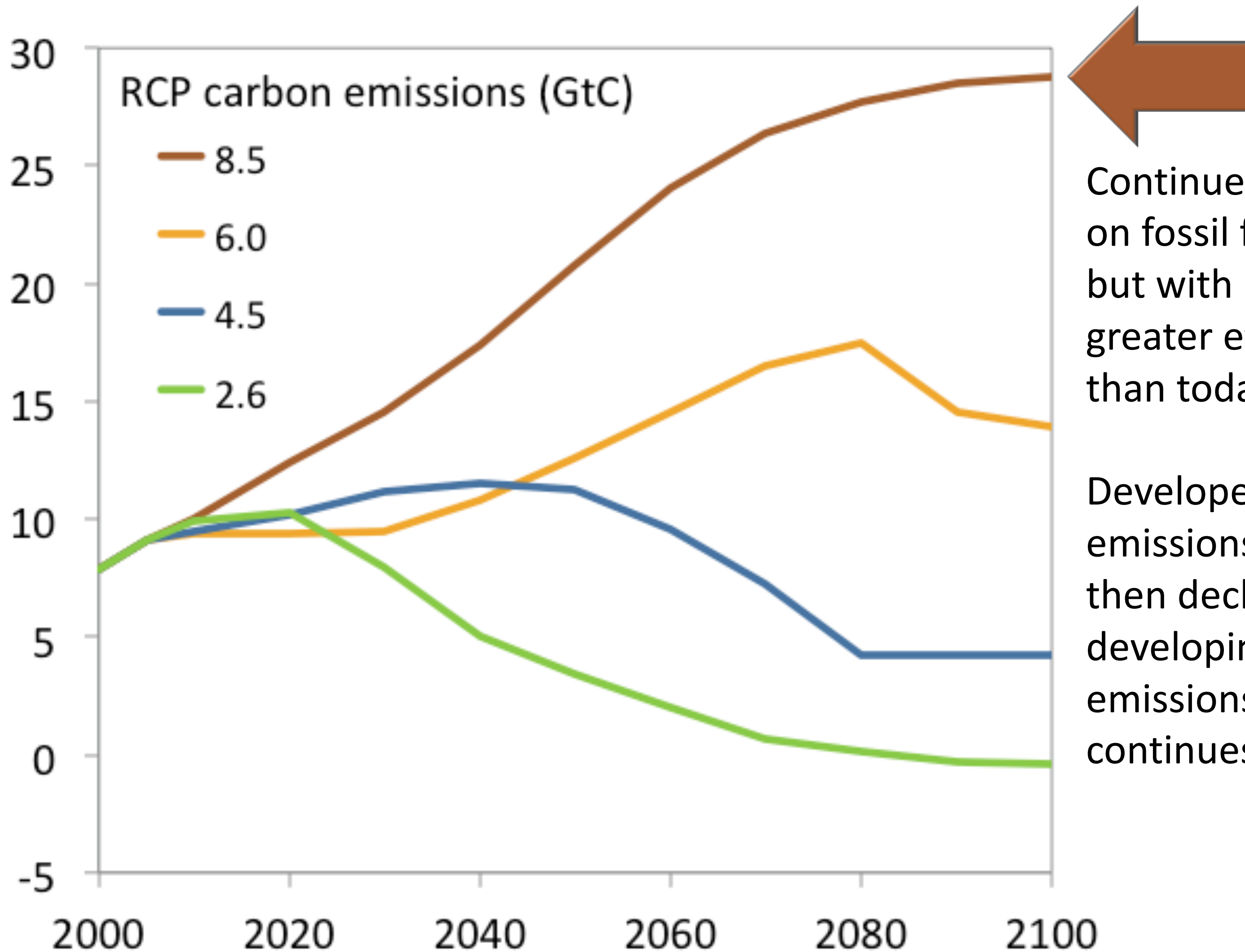
CONCLUSIONS

- The “perfect model” framework for evaluating downscaling methods consistently identifies geographic locations and quantiles at which the stationarity assumption is violated.
- For temperature, all ESDMs show reasonable stationarity in the middle of the distribution in most regions but degrade toward the tails and at high latitudes, especially for simpler methods.
- For precipitation, methods show sharp differences depending on the quantile of the distribution. This has important implications for application of ESDM output to impact assessment.
- Using the “perfect model” framework as a development tool has created an ESDM with biases at least equal to, and generally lower than, its predecessor; upcoming research will test ARRMv2 biases in precipitation and relative humidity.

For the climate data exercise, we are using:

- 6 CMIP5 Global Climate Models (GCMs), selected for their ability to reproduce the Indian Monsoon and their long development history
  - CCSM4
  - GFDL-ESM2G
  - IPSL-CM5A-LR
  - MIROC5
  - MPI-ESM-LR
  - MRI-CGCM3
- 2 future Representative Concentration Pathways (RCPs)
  - The higher RCP 8.5
  - The mid-low RCP 4.5

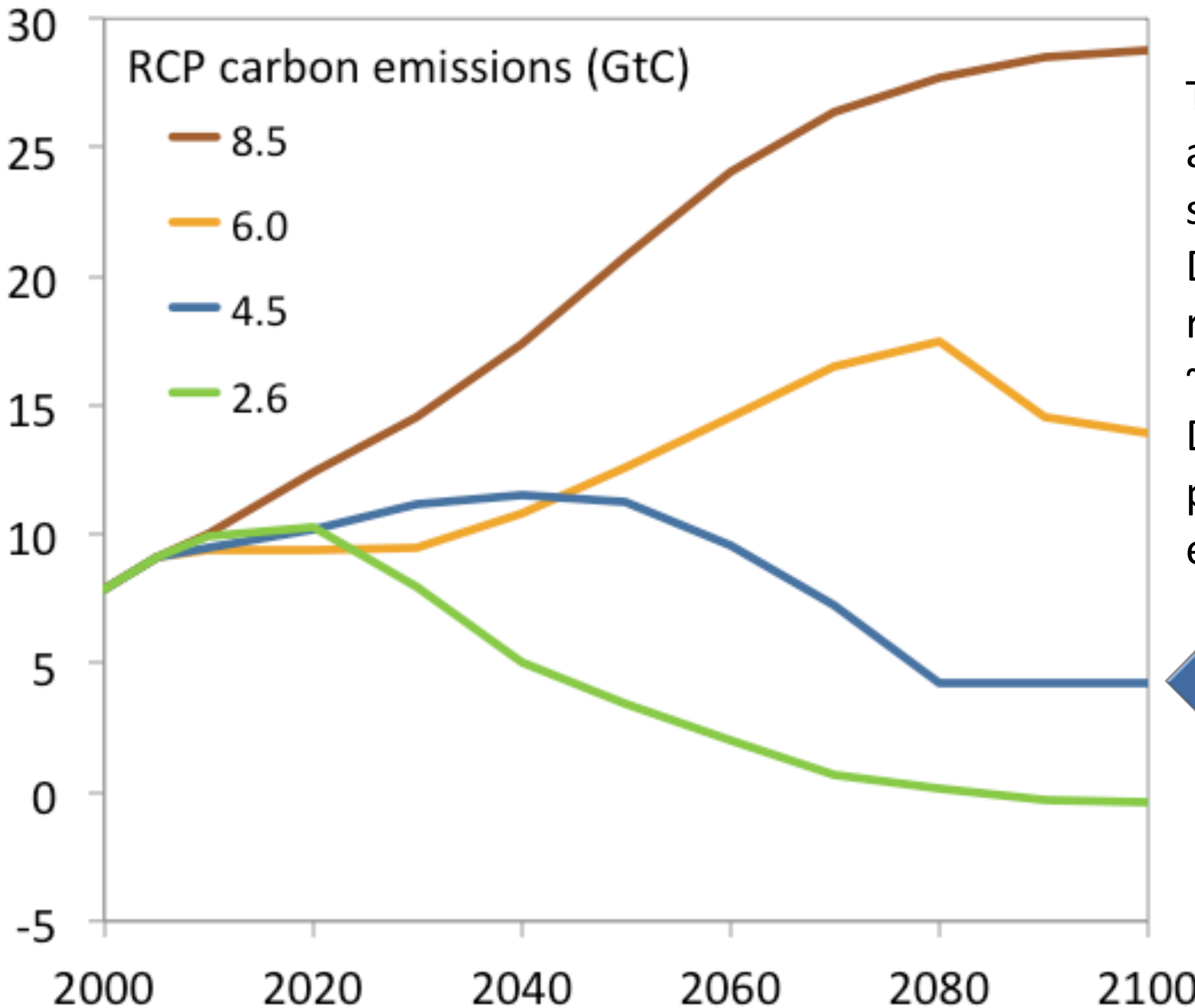
Two Future Scenarios: Higher and Lower



Continued reliance on fossil fuels, but with much greater efficiency than today.

Developed nations' emissions peak, then decline, while developing nations' emissions growth continues.

Two Future Scenarios: Higher and Lower

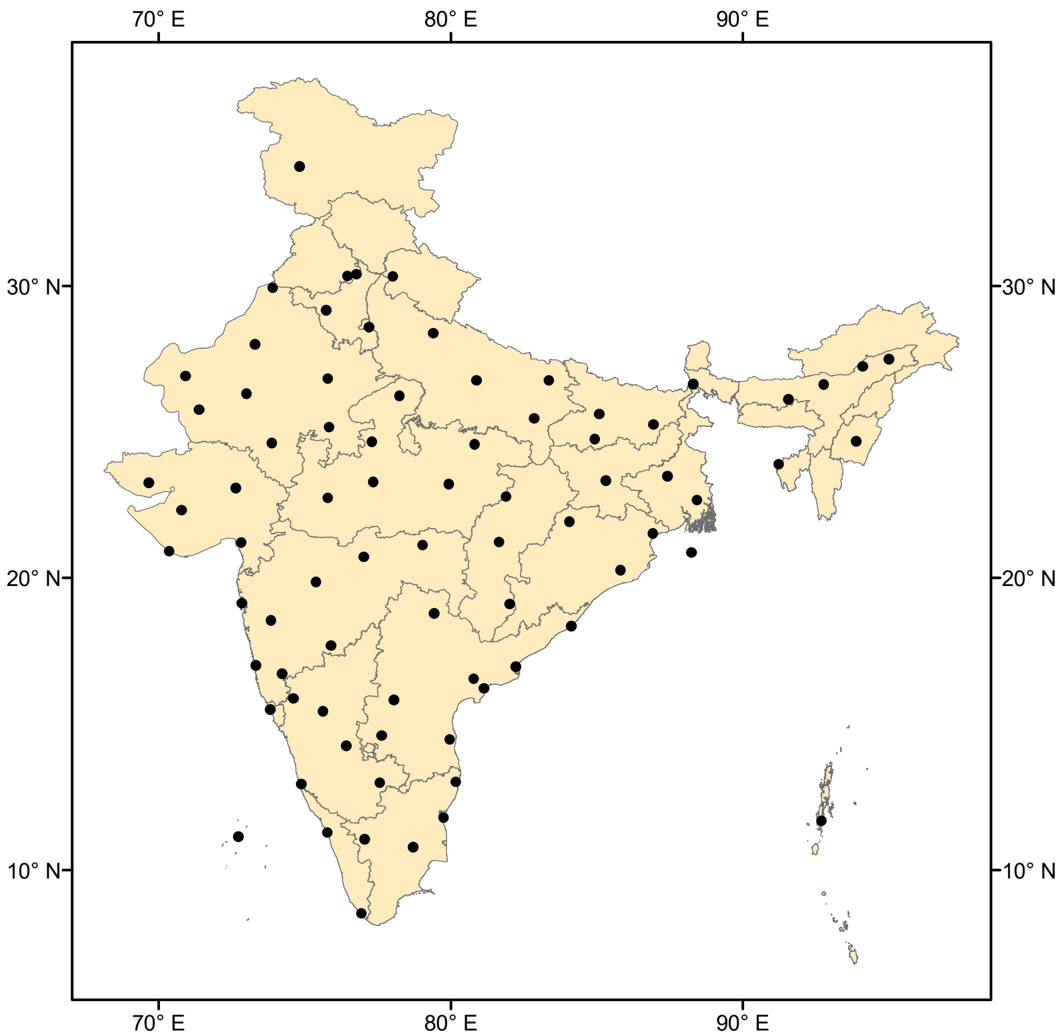


Transition to alternative energy sources.
Developed nations reduce emissions ~80% by 2050
Developing nations participate in emission reductions.

For the climate data exercise, we are using:

- 6 CMIP5 Global Climate Models (GCMs), selected for their ability to reproduce the Indian Monsoon and their long development history
- 2 future Representative Concentration Pathways (RCPs)
- 3 variables: daily maximum and minimum temperature, 24 hour cumulative precipitation

For the climate data exercise, we are using:



64 out of 79 long-term weather stations that have sufficient daily maximum and minimum temperature and 24 hour cumulative precipitation to be downscaled

For the climate data exercise, we are using:

- 6 CMIP5 Global Climate Models (GCMs), selected for their ability to reproduce the Indian Monsoon and their long development history
- 2 future Representative Concentration Pathways (RCPs)
- 3 variables: daily maximum and minimum temperature, 24 hour cumulative precipitation
- 64 out of 79 weather stations
- 2 sets of downscaled projections
  - Projections for individual weather stations, downscaled using ARRMv2 (for temperature) and ARRMv1 (for precipitation)
  - Gridded projections covering all of India, downscaled using NASA NEX

THANK YOU!