Using Multivariate Adaptive Constructed Analogs (MACA) data product for climate projections

Maria Herrmann and Ray Najjar
Chesapeake Hypoxia Analysis and Modeling Program (CHAMP)
Conference Call
2017-04-21

CHAMP project needs

- Downscaled climate projections to serve as inputs to the Chesapeake watershed and estuarine models
- Climate variables:
  - air temperature
  - precipitation
  - humidity
  - downwelling short wave radiation
  - wind velocity (magnitude and direction)
- Spatial resolution: 1/8^{o}
- Temporal resolution: daily

MACA Downscaled Climate Projections Product

- MACA: MULTIVARIATE ADAPTIVE CONSTRUCTED ANALOGS method of statistical downscaling

- Product (Abatzoglou and Brown, 2012): downscaled daily output of 20 CMIP5 GCMs from native resolution to 1/24^{o} (~4 km) or 1/16^{o} (~6 km)
  - developed and hosted at the University of Idaho; John Abatzoglou, PI
  - available at http://maca.northwestknowledge.net/index.php

- Approach in a nutshell: a training dataset (gridded observations) is used to (2) remove historical BIASES and match (2) SPATIAL PATTERNS in GCM output

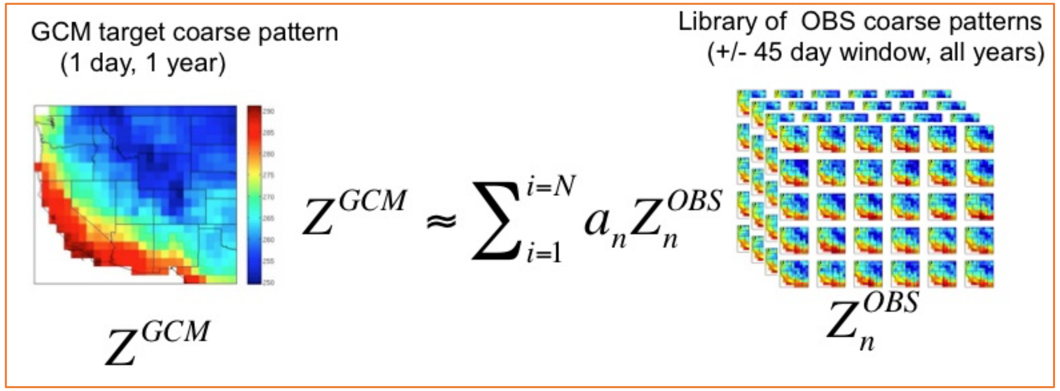
- Downscaled variables: air temperature, humidity, precipitation, downwelling shortwave radiation, wind

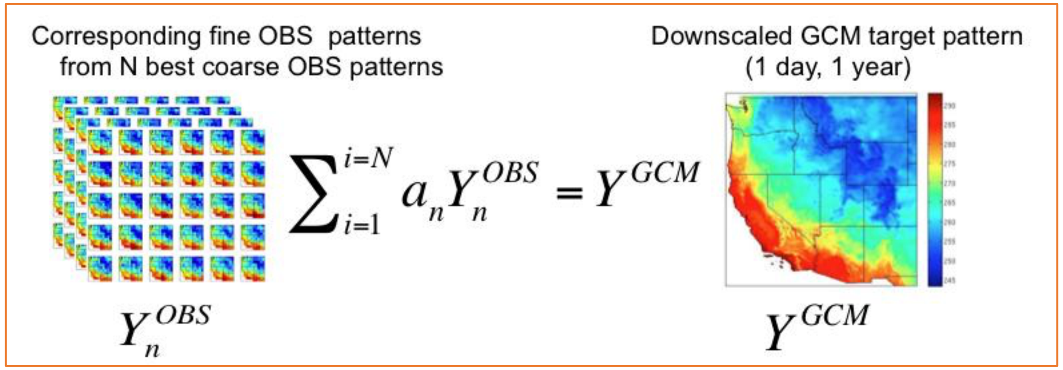
Multivariate Adaptive **Constructed Analogs** method

- GCM data and observations (training dataset) are interpolated to a common coarse grid

- A best fit approach is applied to a daily *target* GCM data by identifying days in the observation record that had similar, or analogous, spatial patterns for a suite of variables. The procedure identifies the top 100 (changed to 10 in v2) patters, or analogs, that occurred no-more than 45 days from the day of year for the *target* GCM day. Analogs are selected based on the lowest root-mean-square error (RMSE) to the *target* pattern. Rather than only using a single "best" analog, the constructed analog approach uses a superposition of the top patterns and through matrix inversion estimates coefficients for each pattern that can then be used to develop a linear model for the target day.

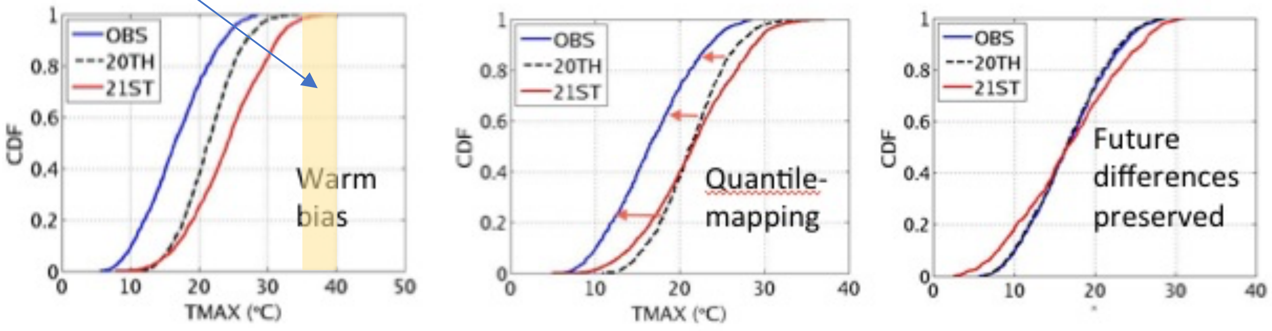
- The coefficients are then applied to superpose the fine-scale observation data and to create the downscaled GCM data.





Multivariate **Adaptive** Constructed Analogs method

1. In addition to bias correction, epoch adjustment is used to adaptively account for *disappearing/novel analogs*. The seasonal and yearly trends at each grid point are calculated using a 21-day, 31-year running mean of the data and removed. These trends are replaced as the final step in MACA. This step helps to avoid the lack of suitable weather analogs in a changing climate.
2. At each grid point, the cumulative distribution function (CDF) of a 15-day window and all years of data(historical or future) is considered for both the GCM data and the observation data. The bias correction step maps the CDF of the historical GCM data and the future GCM data to the CDF of the observation data using a non-parametric quantile-mapping method. Quantile differences between the historical and future GCM data are preserved after the bias correction step. This bias correction is applied first to coarse GCM data and secondly to the fine GCM data after the constructed analogs step.



Raw GCM Data

After Epoch Adjustment

After Bias Correction

MACA Data Products Overview

- Daily output from 20 CMIP5 GCMs downscaled to $1/24^{o}$ (~4 km) or $1/16^{o}$ (~6 km) resolution
    - 20 models were selected from over 40 CMIP5 models based on whether daily output was available for all variables of interest
    - Only one ensemble run from each model was used
- Two time periods
    - 1950-2005 (historical)
    - 2006-2100 (future)
- Two Representative Concentration Pathways (RCPs) for the future:
    - RCP4.5 (an additional 4.5 W/m^2 is added by 2100 compared to preindustrial conditions; moderate climate action)
    - RCP8.5 (an additional 8.5 W/m^2 is added by 2100; no climate action)
- Two downscaling methods:
    - MACAv1
    - MACAv2
- Two training datasets:
    - METDATA (Abatzoglou, 2013)
    - LIVNEH (Livneh et al., 2013)

Differences between MACAv1 and MACAv2 downscaling

Epoch adjustment: variables
In v2, the trend in all variables is removed at the start.
In v1, the trend for only tasmax/tasmin,uas/vas was removed at the start.

Epoch adjustment: periods
In v2, the trend is removed for both the historical and future periods.
In v1, the trend was only removed in the future period.

Bias correction at coarse scale: preserving trends
In v2, after the coarse bias correction of the multiplicative variables (i.e. pr, huss,wind speed), the trend is removed again (as the coarse bias correction can modify this trend).

Constructed Analogs: number of patterns
In v2, we use 10 patterns.
In v1, we use 100 patterns(there may be some overfitting here).

Constructed Analogs: the error
In v2, we interpolate the error in the constructed coarse pattern and add it to the constructed downscaling at the fine level.
In v1, we throw away the coarse error from the constructed analogs method.

Bias Correction at Fine Scale: independently or jointly
In v2, we bias correct pr, huss, rsds independently, but bias correct tasmax and tasmin jointly with the corrected precipitation.
In v1, we bias correct all variables independently.

Comparison of METDATA and LIVNEH training datasets

METDATA

- Spatial resolution of 1/24-degree (~4-km)
- Daily timescales
- Variables: near-surface minimum/maximum temperature, minimum/maximum relative humidity, precipitation, downward solar radiation, **wind components**, and specific humidity.

This dataset was created by bias-correcting daily and sub-daily mesoscale reanalysis and assimilated precipitation from the NASA North American Land Data Assimilation System (NLDAS-2, Mitchell et al., 2004) using monthly temperature, precipitation and humidity from Parameter-elevation Regressions on Independent Slopes Model (PRISM, Daly et al., 2008). The data were validated against an extensive network of weather stations including RAWS, AgriMet, AgWeatherNet, and USHCN-2 showing skill in correlation and RMSE comparable to that derived from interpolation using station observations.

LIVNEH

- Spatial resolution of 1/16 deg (~ 6-km)
- Daily timescales
- Variables: near-surface minimum/maximum temperature, precipitation and **wind speed** from the years 1950-2011.

This dataset was created by incorporating day observations of maximum and minimum temperature as well as accumulated precipitation from National Weather Service Cooperative Observer stations across the country. Climatological estimates of monthly precipitation from the Parameter-elevation Regressions on Independent Slopes Model (PRISM, Daly et al., 2008) are further used to account for spatial variability in precipitation. Temperatures are adjusted across elevated terrain using a standard 6.5C/km lapse rate.

The training dataset for MACAv2-LIVNEH also includes the additional variables of downward solar radiation and specific humidity. These variables have been estimated from the temperature and precipitation data in the Livneh dataset using the MT-CLIM (Glassy et al., 1994) algorithm. The MT-CLIM algorithm utilizes the latitude, elevation, slope, and aspect of the location. Radiation estimates are based on the observation that the diurnal temperature range (the difference between maximum and minimum temperature each day) is closely related to the daily average atmospheric transmittance.

Comparison of the three available MACA data products

	MACAv1-METDATA	MACAv2-LIVNEH	MACAv2-METDATA
Status	Available	Available	Available
Training Dataset	Abatzoglou et. al, 2012 (Years 1979-2010)	Livneh et. al, 2013 (Years 1950-2011)	Abatzoglou et. al, 2012 (Years 1979-2012)
Resolution	4-km (1/24-deg)	~6-km (1/16-deg)	4-km (1/24-deg)
Spatial Extent	WUSA (31.02-49.1N, -124.77 to -103.02E)	Contiguous USA (CONUS) and Columbia Basin into Canada	Contiguous USA (CONUS)
Spatial Projection	WGS 1984 datum	WGS 1984 datum	WGS 1984 datum
Temporal Resolution	daily(on METDATA grid) monthly aggregations (adjusted to 4-km PRISM grid)	daily	daily
Temporal Extent	1950-2100	1950-2099	1950-2099
Downscaled Variables	tasmax,tasmin,rhsmax, rhsmin,pr,rsds, uas,vas,huss	tasmax,tasmin,pr, rsds,was,huss	tasmax,tasmin,rhsmax, rhsmin,pr,rsds, uas,vas,huss
Leap Days	No leap days	Yes, even if model excludes leap days	Yes, even if model excludes leap days
Size of individual netcdf files	3GB-daily, 100MB-monthly	2-4GB compressed-daily,0.1GB compressed-monthly	2.5GB-daily
Size of total dataset	13.7TB-daily, 0.3TB-monthly	5TB-daily,0.5TB-monthly	20TB-daily

List of CMIP5 GCMs used in MACA product

Model Name	Model Country	Model Agency	Atmosphere Resolution(Lon x Lat)	Ensemble Used
bcc-csm1-1	China	Beijing Climate Center, China Meteorological Administration	2.8 deg x 2.8 deg	r1i1p1
bcc-csm1-1-m	China	Beijing Climate Center, China Meteorological Administration	1.12 deg x 1.12 deg	r1i1p1
BNU-ESM	China	College of Global Change and Earth System Science, Beijing Normal University, China	2.8 deg x 2.8 deg	r1i1p1
CanESM2	Canada	Canadian Centre for Climate Modeling and Analysis	2.8 deg x 2.8 deg	r1i1p1
CCSM4	USA	National Center of Atmospheric Research, USA	1.25 deg x 0.94 deg	r6i1p1
CNRM-CM5	France	National Centre of Meteorological Research, France	1.4 deg x 1.4 deg	r1i1p1
CSIRO-Mk3-6-0	Australia	Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Centre of Excellence, Australia	1.8 deg x 1.8 deg	r1i1p1
GFDL-ESM2M	USA	NOAA Geophysical Fluid Dynamics Laboratory, USA	2.5 deg x 2.0 deg	r1i1p1
GFDL-ESM2G	USA	NOAA Geophysical Fluid Dynamics Laboratory, USA	2.5 deg x 2.0 deg	r1i1p1
HadGEM2-ES	United Kingdom	Met Office Hadley Center, UK	1.88 deg x 1.25 deg	r1i1p1
HadGEM2-CC	United Kingdom	Met Office Hadley Center, UK	1.88 deg x 1.25 deg	r1i1p1
inmcm4	Russia	Institute for Numerical Mathematics, Russia	2.0 deg x 1.5 deg	r1i1p1
IPSL-CM5A-LR	France	Institut Pierre Simon Laplace, France	3.75 deg x 1.8 deg	r1i1p1
IPSL-CM5A-MR	France	Institut Pierre Simon Laplace, France	2.5 deg x 1.25 deg	r1i1p1
IPSL-CM5B-LR	France	Institut Pierre Simon Laplace, France	2.75 deg x 1.8 deg	r1i1p1
MIROC5	Japan	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies,and Japan Agency for Marine-Earth Science and Technology	1.4 deg x 1.4 deg	r1i1p1
MIROC-ESM	Japan	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	2.8 deg x 2.8 deg	r1i1p1
MIROC-ESM-CHEM	Japan	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	2.8 deg x 2.8 deg	r1i1p1
MRI-CGCM3	Japan	Meteorological Research Institute, Japan	1.1 deg x 1.1 deg	r1i1p1
NorESM1-M	Norway	Norwegian Climate Center, Norway	2.5 deg x 1.9 deg	r1i1p1

Guidelines from MACA website

Uncertainty in climate projections comes from:

- anthropogenic forcing: changes in greenhouse gases, aerosols, and land use
  uncertainty is relatively low before mid-century, but increases substantially by late 21st century
- the sensitivity of the climate system to changes in forcing
  uncertainty is significant and model-dependent
- natural unforced variability
  uncertainty can be particularly important for near-term projections and for variables with a low signal-to-noise ratio (e.g., precipitation)

Model evaluation over given domain:

- studies have shown that climate projections from a random set of models yield results similar to those from the best models; therefore it may not be absolutely critical to cull or weight GCMs

How many models to consider:

- using all available models will not encompass the full range of potential futures
- but incorporating a small number of projections is not suggested as it will provide a limited sample
- suggested number: at least 10 models

Historical baseline:

- the full 56 year historical period (1950-2005) from each GCM should be used; the MACA process maps GCM values to values from the training observation dataset, so that only the full 56 years will give the proper statistics held in the training dataset; using a subset of the historical period will not guarantee that the extremes typical of the data will be seen or that the resulting statistics will be typical of the historical period.

GCM selection process

Metrics

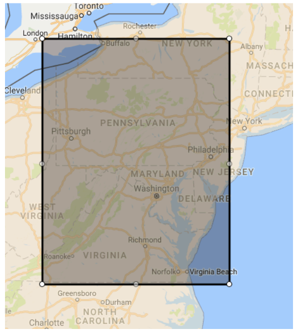
- mean temperature
- precipitation
- humidity
- SWR
- wind speed

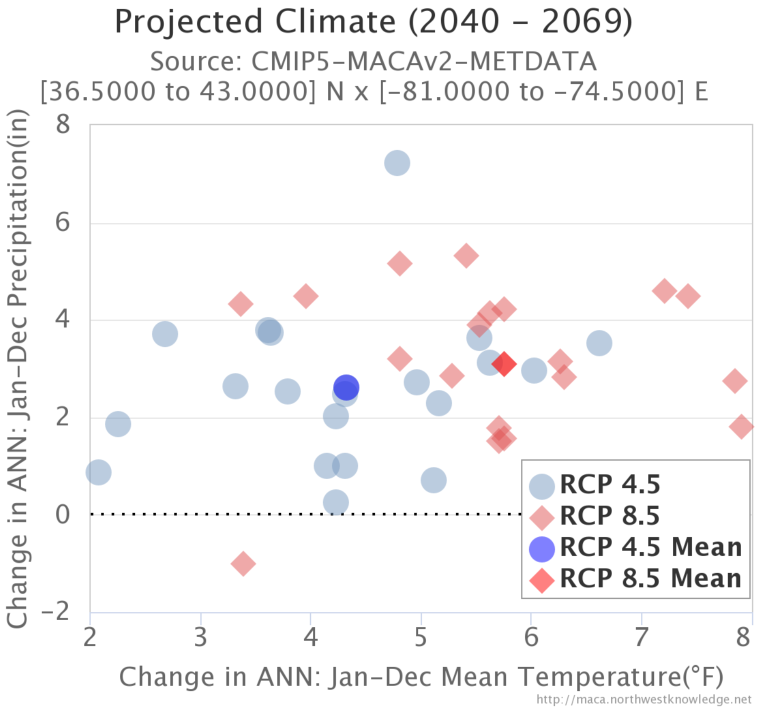
Time periods: 1950-2005 (historical) vs. 2040-2069 (mid-century)

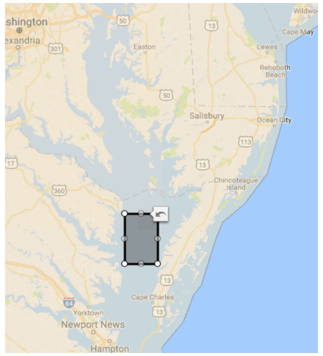
Spatial domains: watershed vs. estuary

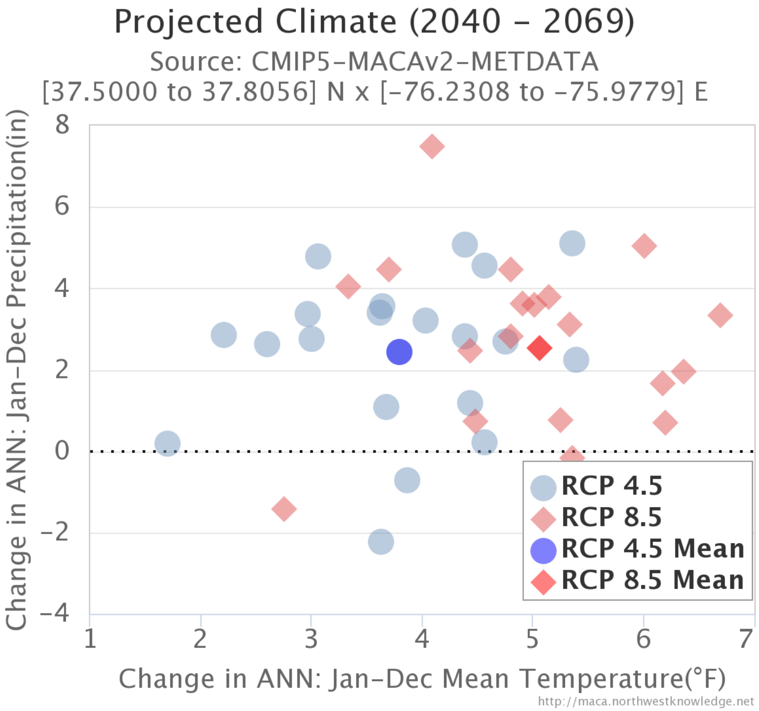
Calculation steps:

average over time (years, months, seasons?)
average in space
calculate change for each metric for each model between the time periods
calculate statistics of the changes over the set of models









Summary of proposed strategy

- MACAv2_METDATA
  - improved downscaling method
  - wind direction in addition to wind sped
- Historical period: 1950-2005
- Mid-century period: 2040-2069
- Maximize number of GCMs rather than RCPs
  - Before mid-century, uncertainty in climate projections is driven more by the uncertainty in climate sensitivity, rather that the uncertainty in anthropogenic forcing; the latter becomes relatively more important by late 21st century

Questions to the group

- Realistically, what is the max number of projection scenarios we could use?
- Do we need min/max temperature and humidity or just means?
- Do we agree on daily, $1/8^{o}$?
- Do we agree on two domains (watershed vs. estuary) and can I get the masks?

References

http://maca.northwestknowledge.net/index.php (main source of info for this presentation)

Abatzoglou, J.T., 2013. Development of gridded surface meteorological data for ecological applications and modelling. International Journal of Climatology 33, 121-131. (METDATA training dataset)

Abatzoglou, J.T., Brown, T.J., 2012. A comparison of statistical downscaling methods suited for wildfire applications. International Journal of Climatology 32, 772-780. (MACA method)

Livneh, B., Rosenberg, E.A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K.M., Maurer, E.P., Lettenmaier, D.P., 2013. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. Journal of Climate 26, 9384. (LIVNEH training dataset)