Creation of Pseudo Worlds

Kate Willett

Summary:
I have created 8 worlds of station data, each with a different background climate and error model combination.

I have a document listing the applied errors for each station for each world

Homogenisation algorithms can be run on these worlds.
HOW WELL DOES THE ALGORITHM DO AT BRINGING THE ERROR MODELS BACK TO GROUND TRUTH?
I can compare the % of discontinuities correctly detected/missed/incorrectly detected for each world.
I can compare the 'truth' trends/mean/variance with that of the homogenised pseudo-worlds to see how similar they are

Creating the clean 'truth' pseudo-worlds:
I have taken HadCM3 surface Tmax and Tmin daily model data for three different climate types: 1950-2000 for CONTROL, 2000-2050 for SO2_HIST (natural forcings only) and A1B (all forcings) for the globe - timestamps are all saved to 1950-2000 for convenience.

I have downscaled these grids to match HadGHCND stations over three regions:

AIP=Australia including Pacific Islands 110E-180E, 50S-10S
TSA=Tropical Southeast Asia including southern Japanese Islands, Philipines, Thailand etc. Not India 90E-150E, 10S-30N
NWE=Northwest Europe to East Germany border 10W-15E, 40N-70N

Each pseudo-station has its climatological mean and variance (by month) nudged to match that of its related real-world station. White noise is added to reduce the autocorrelation in pseudo-stations that are derived from the same model grid-box.

Each pseudo-station has data removed to match the temporal coverage of its related real-world station.

Creating the 'broken' pseudo-worlds:
I have created three error types of pseudo worlds - many small breaks (MSB), few large breaks (FLB) and hotch-potch breaks (HPB) (a mix of both). All have both abrupt and gradual discontinuities and some are common to geo-political regions. Applied to the three climate types this constitutes 8 pseudo-worlds:

CONTROL no breaks 1950-2000
CONTROL many small breaks 1950-2000
CONTROL few large breaks 1950-2000
A1B (inc. background warming trend) many small breaks 2000-2050
A1B (inc. background warming trend) few large breaks 2000-2050
A1B (inc. background warming trend) hotch potch of all 2000-2050
SO2_HIST (natural forcing) many small breaks 2000-2050
; SO2_HIST (natural forcing) few large breaks 2000-2050

BREAKPOINT STATISTICS:
Minimum gap = 6 months for many small and hotch potch and 12 months for few large

few large breaks = 3-5 breaks per station
many small breaks = 5-10 breaks per station
hotchpotch breaks = 3-10 breaks per station

gradual = one systematic slow increase/decrease in the mean & variance per station - NEVER GEOGRAPHICALLY CONSISTENT

(Urbanisation will affect diurnal temp range - warmer at night as well as warmer generally - this is not accurately characterised yet)

All changes to Tmax and Tmin are applied identically but may result in slightly different actual changes.

delta mean only = ~30% of breaks in station
delta variance only = ~30% of breaks in station
delta mean and variance only = ~40% of breaks in station

~40% of breaks per station will be geographically/politically consistent to within +/- 1 year
~30% of stations to have a 'gradual inhomogeneity' - is this representative of the real world? Possibly not but don't want to cloud the ability to detect breaks by confusing things too much

Some changes to the mean are applied seasonally - so they are relative to the variance for each month during the annual cycle
All changes to the variance are seasonal to some extent – higher variance months will increase more than lower variance months

Large Breaks:
MEAN = increase/decrease by ~1 deg C (0.5-1.5 deg C)
VARIANCE = increase/decrease by 0.2-4 StDev (0.6-0.8 or 1.2 to 1.4 * StDev)

Small Breaks:
MEAN = increase/decrease by 0.25 deg C (-0.5 to 0.5 deg C)
VARIANCE = increase/decrease by 0.01-0.2 StDev (0.8-1.0 to 1.0-1.2 * StDev)

Gradual Inhomogeneities:
MEAN = increase/decrease by ~1 deg C (0.5-1.5 deg C)
VARIANCE = increase/decrease by 0.1-0.3 StDev (0.7-0.9 or 1.1 to 1.3 * StDev)

delta variance and delta mean applied to all data points after break until next break

all in both Tmin and Tmax identically

gradual changes applied as a block change to all data before break and then gradually between the break locations - all 'many small breaks', 'few large breaks' and hotchpotch worlds will be identical

abrupt changes to all worlds within category (i.e. MSB, FLB, HPB) will be identical

Method of Adding Breaks:

DELTA VARIANCE - go through data in each break period
; - group into months
; - make month clims over break period - subtract from values to get anoms
; - DO WE NEED TO REMOVE TREND?
; - YES - IF YOU DON'T DETREND THEN YOU CHANGE THE MEAN AND ACTUAL TREND OF THE DATA - tested
; - CONCLUSION - ANOMALISE, DETREND (USING mid-point AS CONSTANT)
; - multiply by change in variance
; - add Trend and period clims back

; now apply gradual change over a period in the same way
; - first change variance for all data up to beginning of gradual break location
; - then apply slope in variance from there up to end of break location such that the end is equal to the variance of the remaining data

DELTA MEAN
; now apply change to the mean with no seasonal variance
; - as these are additive it doesn't matter about converting to anomalies - although we should think about how the bias changes over the diurnal cycle when dealing with sub-daily data

; now apply change to the mean with seasonal variance
; get the max and min month climatological st dev for the annual cycle
; min month st dev has 0.75*delta mean applied, max st dev has full delta mean applied, all others are somewhere in between depending on the relative magnitude of their st dev
; Apply this month specific delta mean

; now apply gradual change over a period in the same way
; - first change in mean for all data up to beginning of gradual break location
; - then apply slope in mean from there up to end break loc such that the end is equal to the mean of the remaining data

IMPROVEMENTS FOR LATER?
Addition of more white noise errors?
Tmax/Tmin physically consistent changes - i.e. urban vs rural stations will see different changes in their daytime and nighttime (Tmax/Tmin) obs due to radiation and wind speed differences
Tie each applied error to a specific cause with suitable physical characteristics i.e. station move, instrument renewal, urbanisation etc.
Provide metadata of some known discontinuties