Supporting Information to

Temperature variability projections remain uncertain after constraining them to best performing SMILEs

By N. Maher & L. Suarez-Gutierrez

Model Performance Evaluation of Temperature Variability

This section encloses the temperature variability rank-frequency evaluation framework results for all 11 SMILEs considered in this study globally, against GISTEMPv4 and ERSSTv5 observations over land and ocean regions respectively, both for detrended and non-detrended temperature anomalies. For the purposes of this evaluation, monthly mean temperature anomalies are relative to the period 1961–1990, and model output data are regridded to match the observational grid.

Enclosed figures include:
Maps for the rank-frequency evaluation for December to February (DJF; top) and June to August (JJA; bottom) monthly mean temperature anomalies, presented first for each model, followed by time series and rank-histograms DJF (first) and JJA (second) for spatially aggregated land and ocean regions, presented second, for the following 11 SMILEs, in order of appearance: ACCESS-ESM1.5, CanESM2, CanESM5, CESM1-LE, CESM2-LE, CSIRO-MK3.6, GFDL-ESM2M, GFDL-SPEAR-MED, MIROC6, MPI-GE5 and MPI-GE6.

These results are separated into the following subsections:

- Detrended Land Surface Air Temperatures
- Non-Detrended Land Surface Air Temperatures
- Detrended Ocean Surface Temperatures
- Non-Detrended Ocean Surface Temperatures

Detrended Land Surface Air Temperatures

Rank-frequency variability evaluation framework for detrended 2m air temperature (TAS) anomalies over land grid cells.

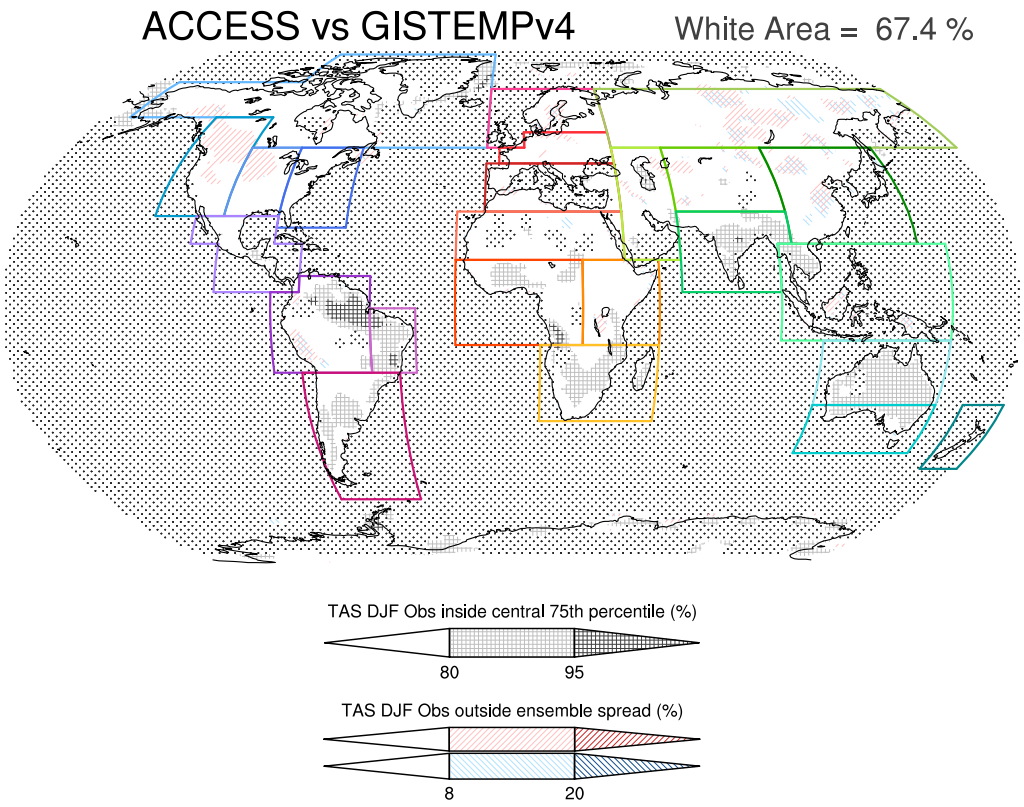
Maps show grid-cell evaluation of the simulated DJF and JJA monthly mean temperature anomalies for the 11 SMILEs included in this study against GISTEMPv4 observations globally. Gray hatching represents where observations cluster within the 75th percentile bounds of the ensemble (12.5th to 87.5th percentiles) for more than 80% of months (light grey) or for more than 95% of months (dark grey). Red and blue shading represents where observations are larger than the ensemble maximum (red) or smaller than the ensemble minimum (blue), respectively, for more than 8% of the months (light red and blue) or for more than 20% of the months (dark red and blue). Dotted areas represent ocean areas or grid cells where observations are missing and are therefore excluded from this analysis. Colored boxes demark the boundaries of each land region assessed. The percentage of assessed grid-cells that present none of these biases in given at the top (white area).

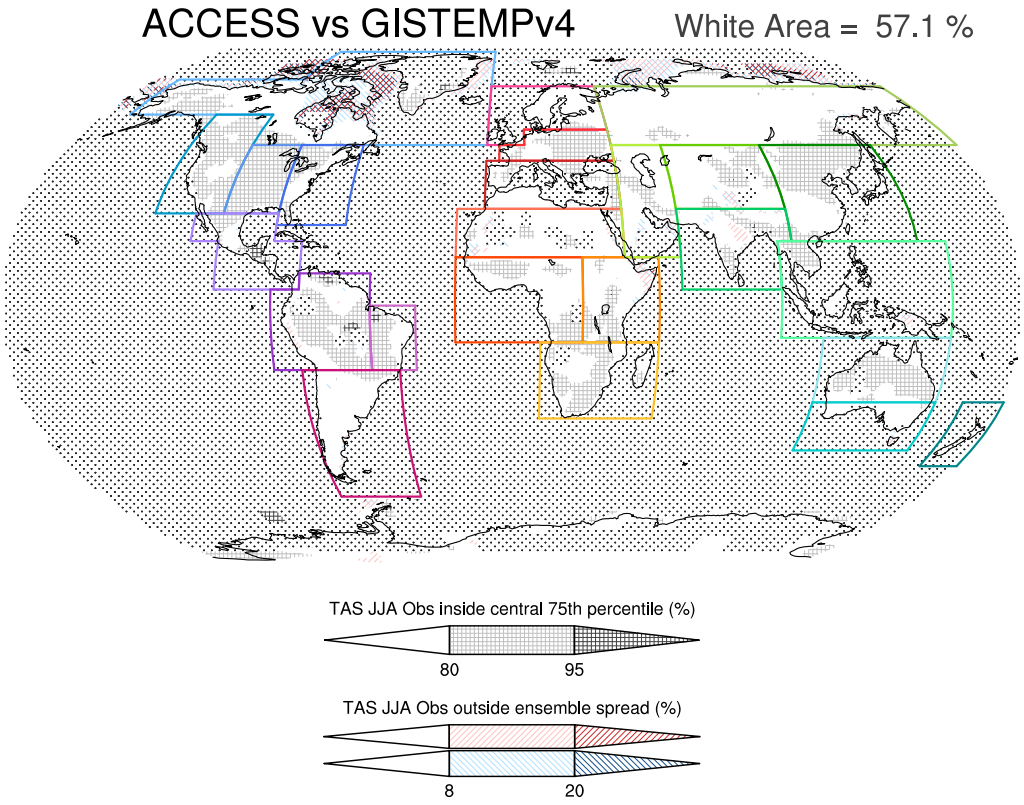
Time series and rank frequency histograms show spatially aggregated DJF and JJA TAS for each land regions for all 11 SMILEs. Time series show the ensemble maximum and minimum (coloured lines) and central 75th percentile ensemble spread (shading) are shown against observations (black dots).

Rank histograms represent the frequency of each place that observations would take in a list of ensemble members ordered by ascending temperature anomaly values. Rank 0 indicates observations are below the minimum ensemble value, and rank n, with n the number of ensemble members, indicates that observations exceed the maximum ensemble value for that particular month. For a model that perfectly represents observations over an infinitely-long observational record, all ranks should occur with similar frequency and this histogram should be roughly flat.

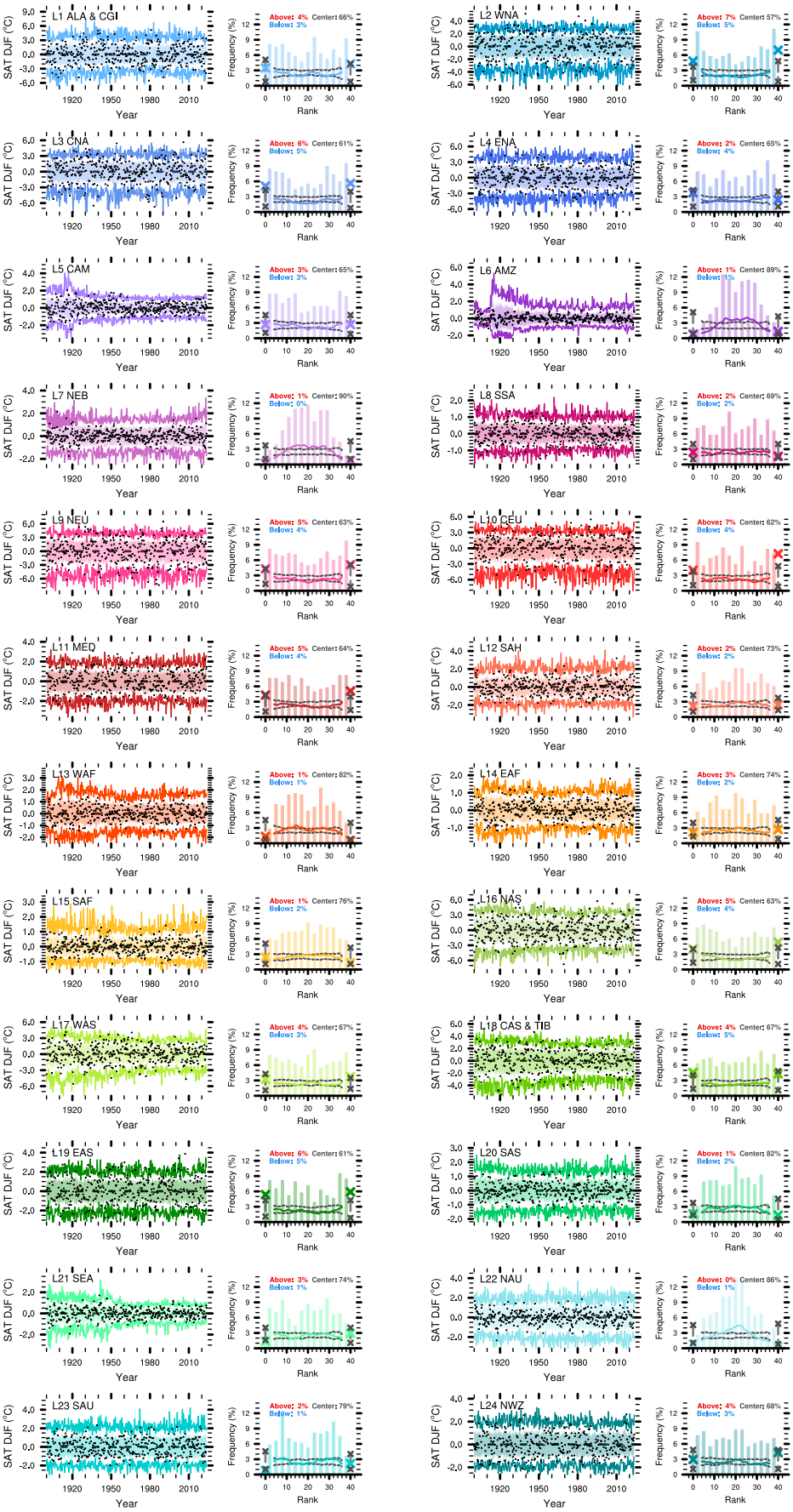
To illustrate how internal variability may affect rank frequencies given the non-infinite record length considered, we also include a perfect-model comparison, which shows the range of rank frequencies that each ensemble member would take if it were observations. If the rank exhibited by observations (colors) is within this perfect-model range (grey), the rank frequency evaluation shows an adequate model performance, and any deviations from a perfectly frank rank histogram can be assumed to be within the range of deviations that could be caused by internal variability. Lines in the rank histogram illustrate the rank histogram's slope, as the mean rank frequency over a centered 6-bin window for observations (solid colored lines), and the 5-95th perc. perfect model range (gray dashed lines). Crosses represent the frequency of minimum (0) and maximum (number of members) ranks for observations (colors), and for the 5-95th perc. perfect model range (gray).

Percentages at the top of the rank histograms show how often monthly anomalies fall above or below ensemble limits (red and blue, respectively) and within the 75th perc. Range (grey), analogous to the criteria chosen for the map-based evaluation but for spatially aggregated values.

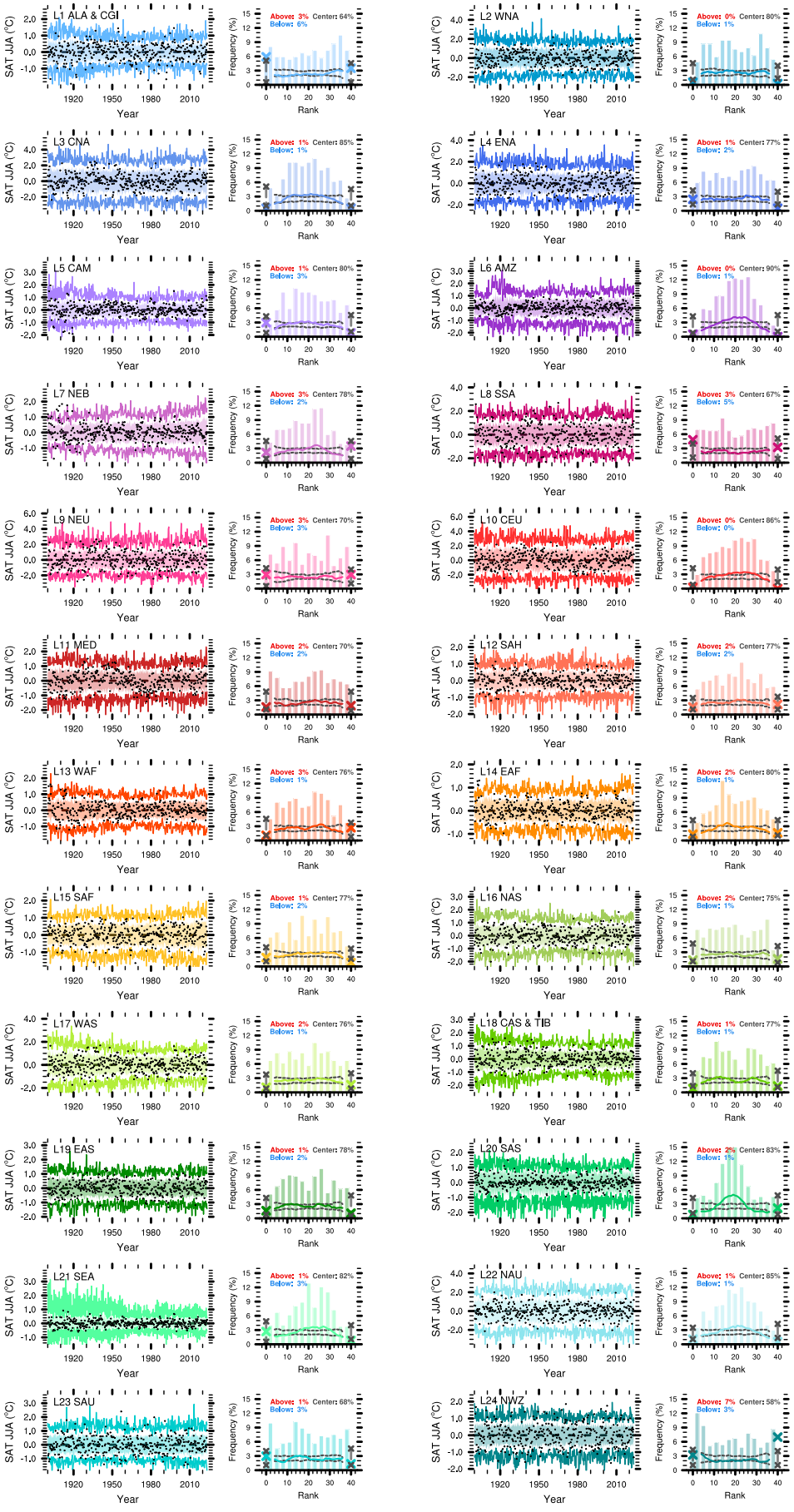

ACCESS vs GISTEMPv4
White Area = 67.4 %
TAS DJF Obs inside central 75th percentile (%)
80
95
TAS DJF Obs outside ensemble spread (%)
8
20


ACCESS vs GISTEMPv4
White Area = 57.1 %
TAS JJA Obs inside central 75th percentile (%)
80
95
TAS JJA Obs outside ensemble spread (%)
8
20

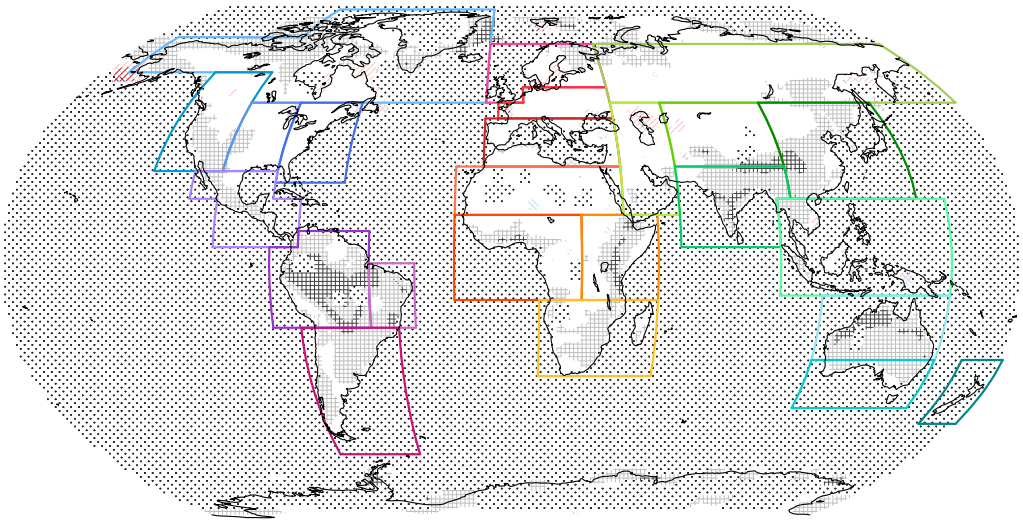
ACCESS vs GISTEMPv4 TAS DJF



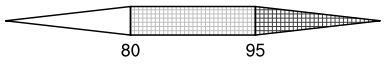
ACCESS vs GISTEMPv4 TAS JJA



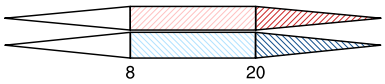
CanESM2 vs GISTEMPv4 White Area = 53.9 %



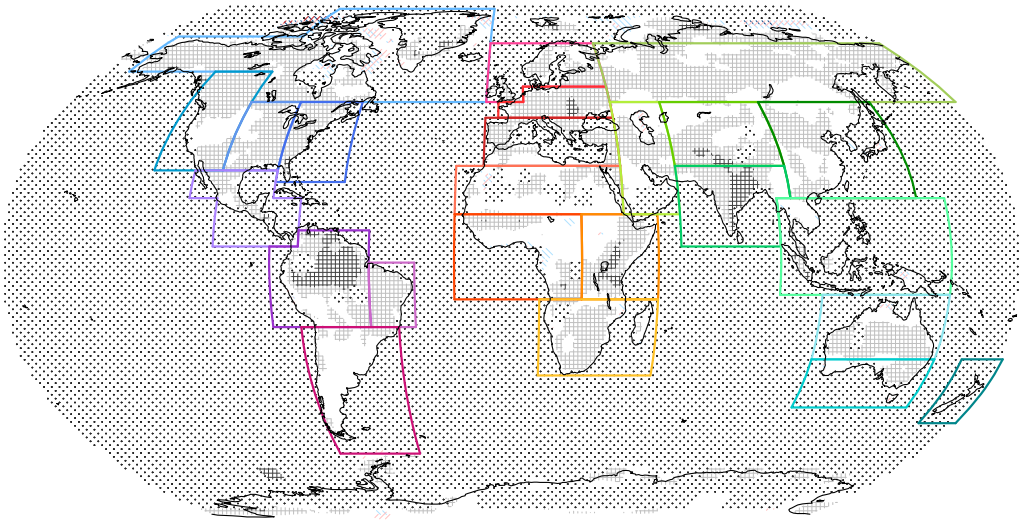
TAS DJF Obs inside central 75th percentile (%)



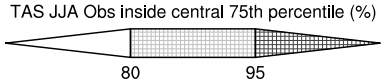
TAS DJF Obs outside ensemble spread (%)



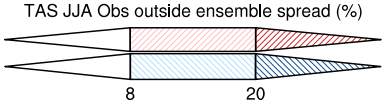
CanESM2 vs GISTEMPv4 White Area = 40.4 %



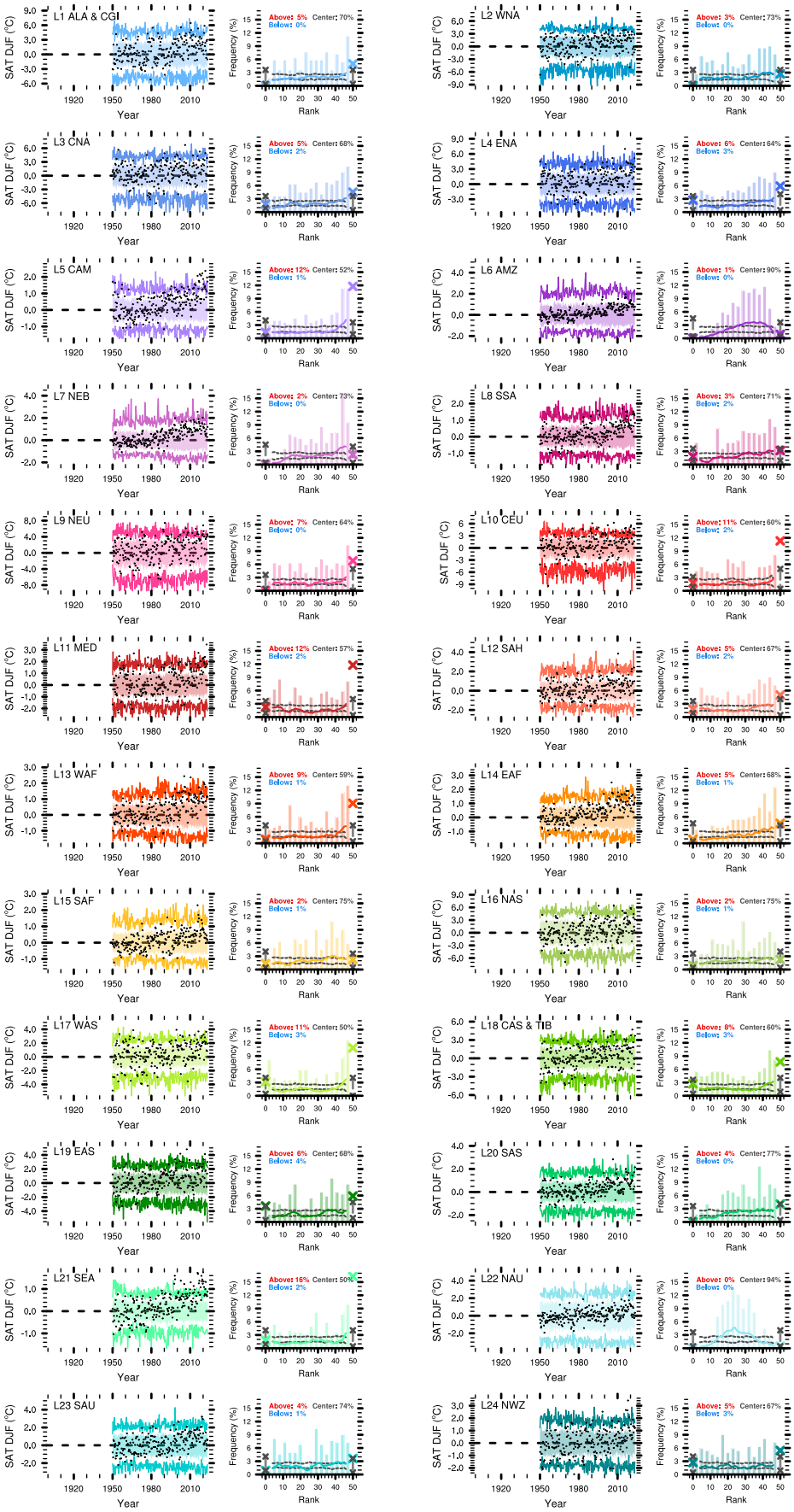
TAS JJA Obs inside central 75th percentile (%)



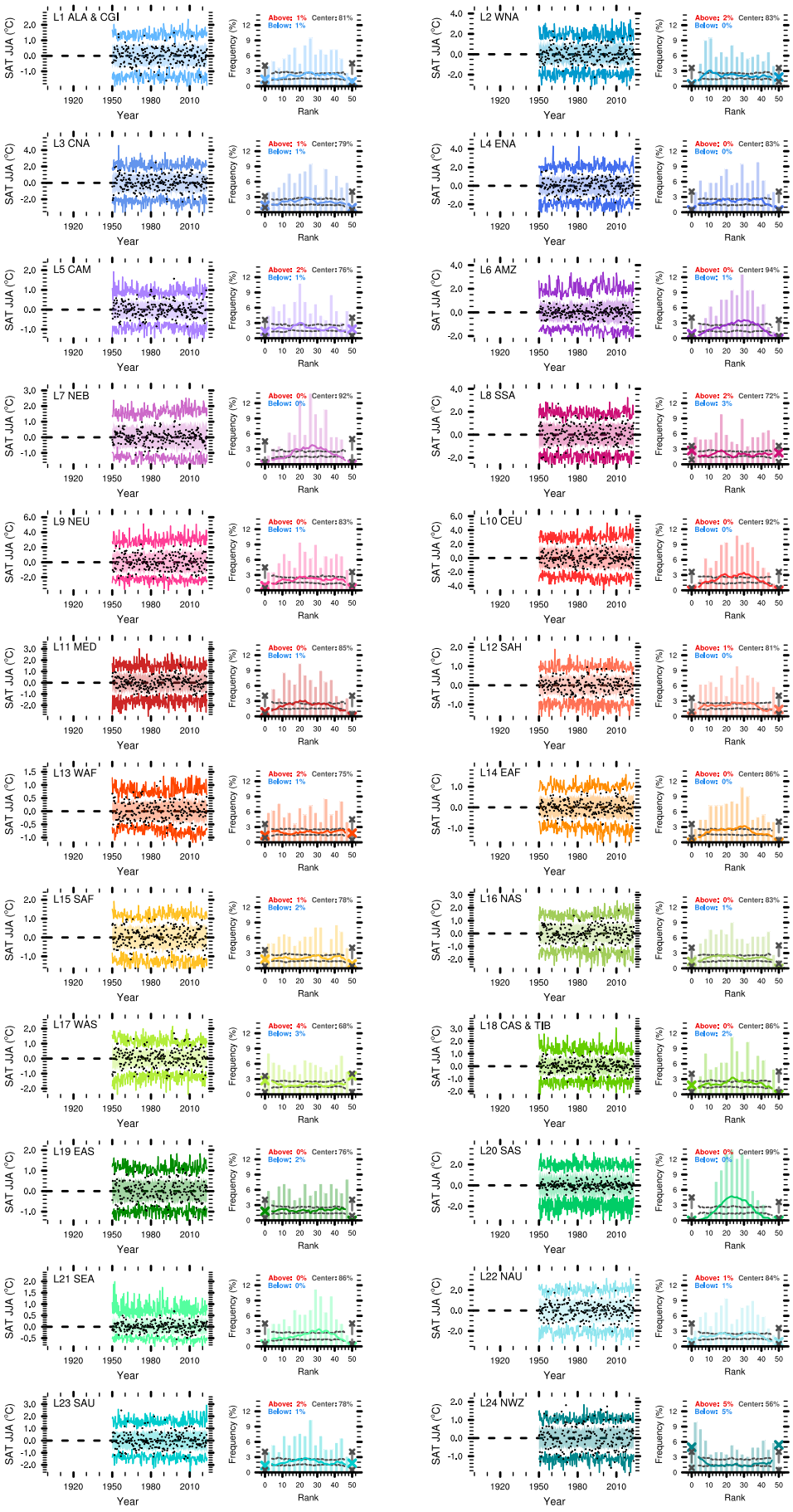
TAS JJA Obs outside ensemble spread (%)

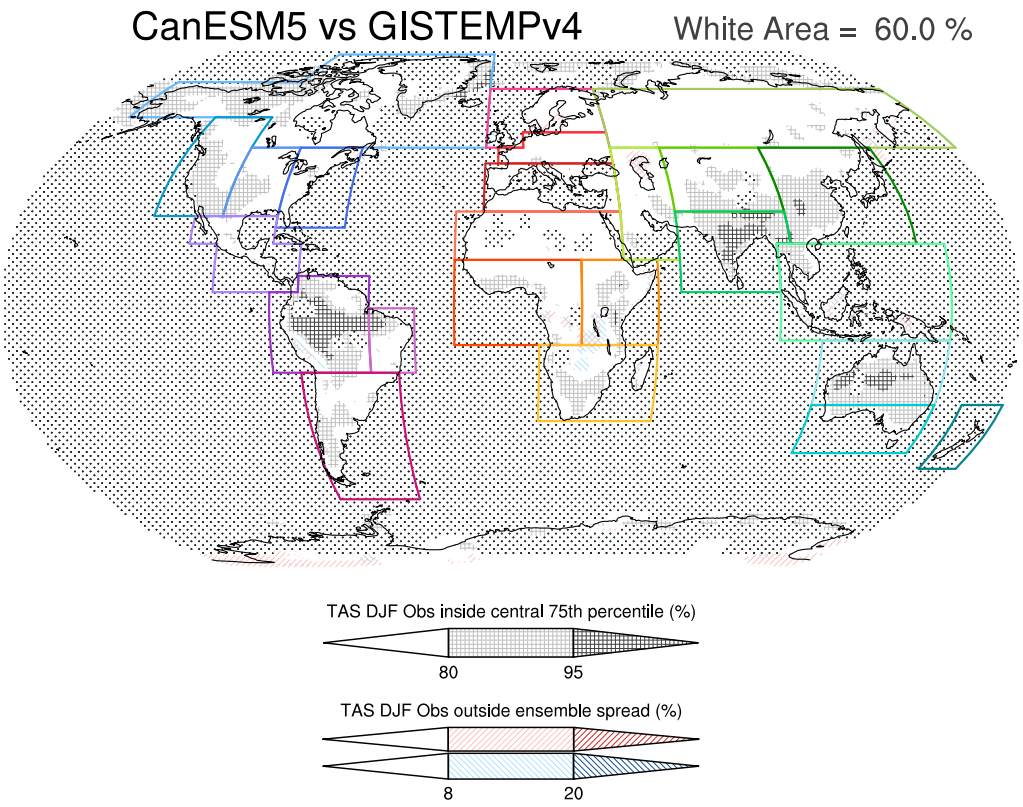


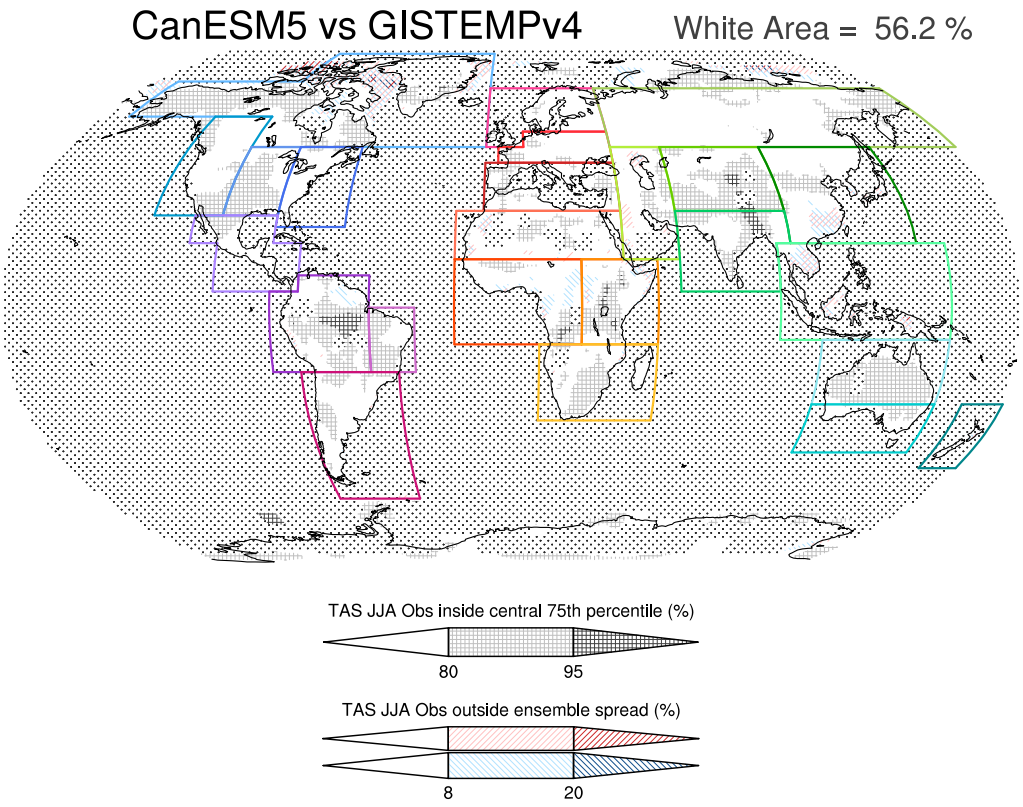
CanESM2 vs GISTEMPv4 TAS DJF



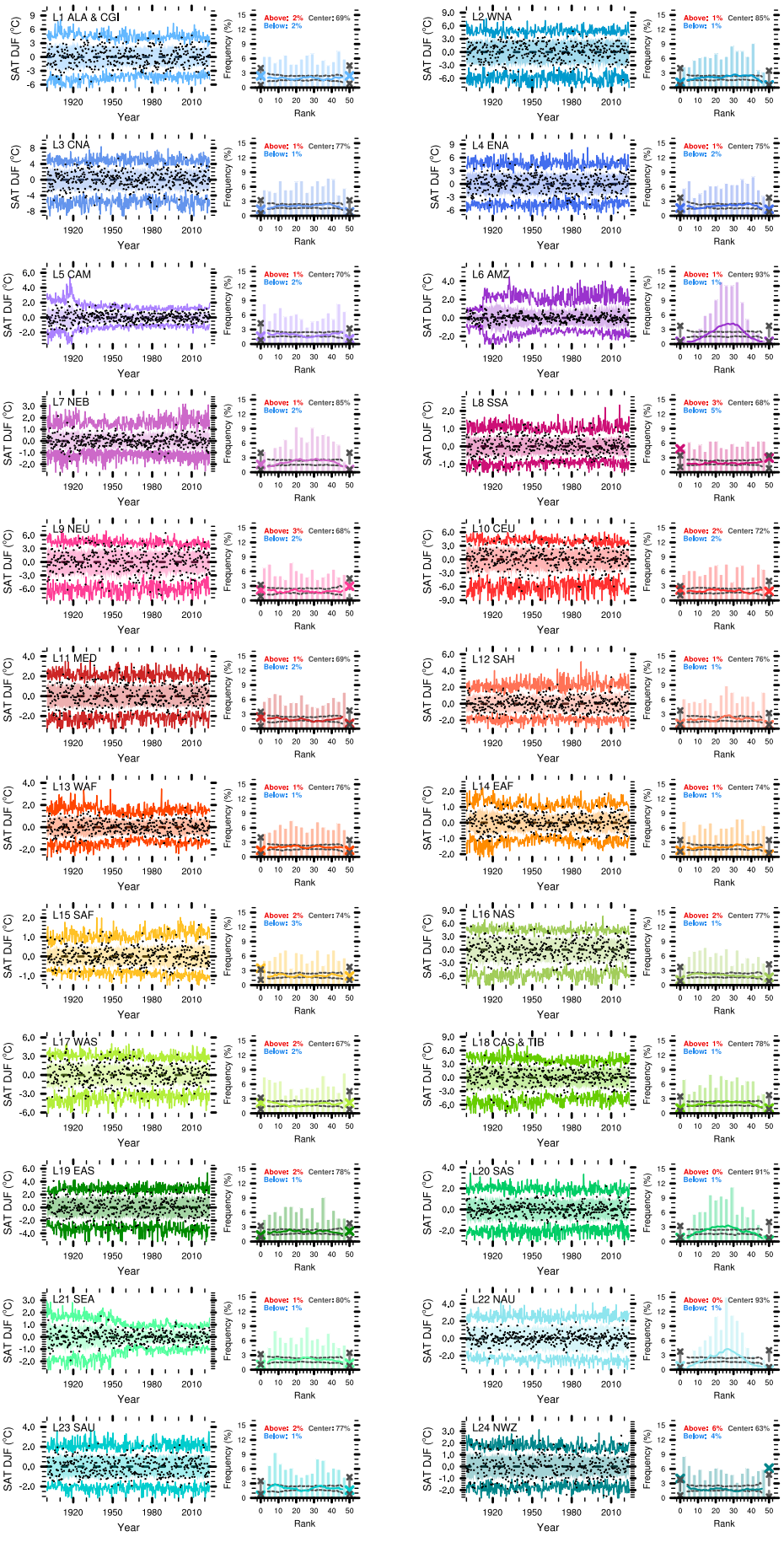
CanESM2 vs GISTEMPv4 TAS JJA



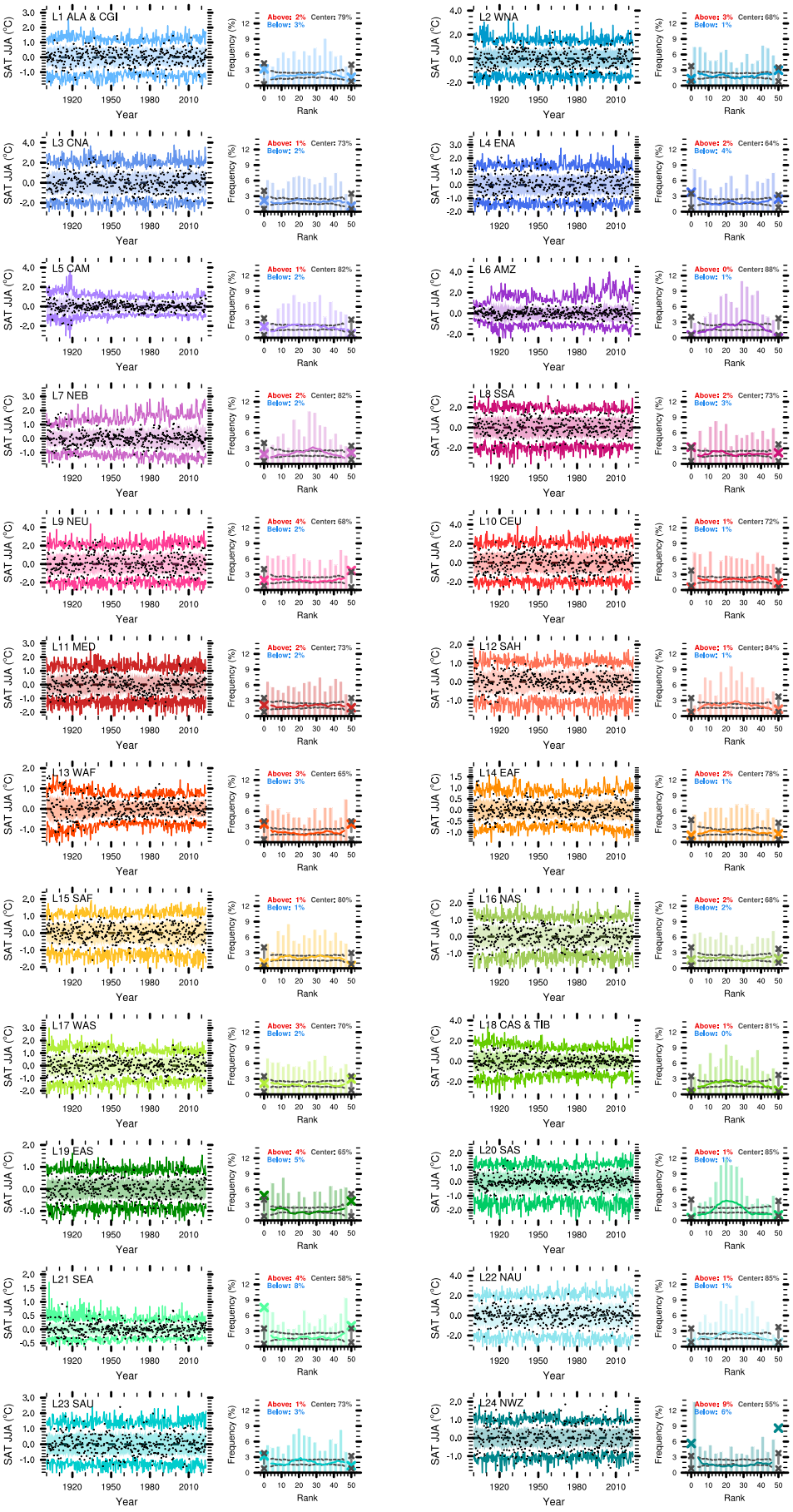

CanESM5 vs GISTEMPv4
White Area = 60.0 %
TAS DJF Obs inside central 75th percentile (%)
80
95
TAS DJF Obs outside ensemble spread (%)
8
20


CanESM5 vs GISTEMPv4
White Area = 56.2 %
TAS JJA Obs inside central 75th percentile (%)
80
95
TAS JJA Obs outside ensemble spread (%)
8
20

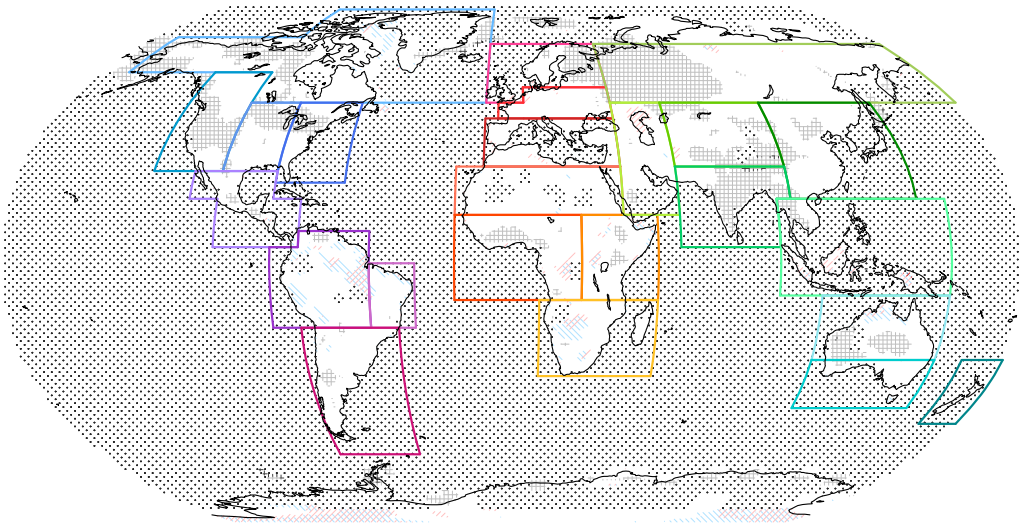
CanESM5 vs GISTEMPv4 TAS DJF



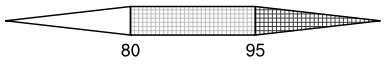
CanESM5 vs GISTEMPv4 TAS JJA



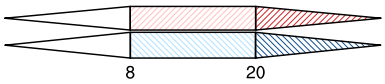
CESM-LE vs GISTEMPv4 White Area = 67.6 %



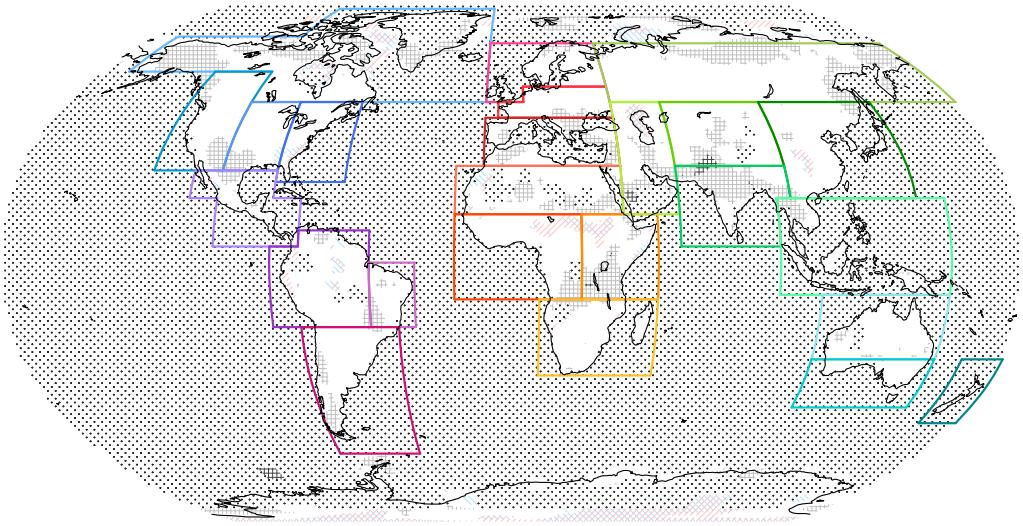
TAS DJF Obs inside central 75th percentile (%)



TAS DJF Obs outside ensemble spread (%)



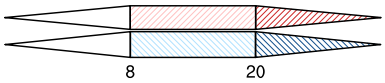
CESM-LE vs GISTEMPv4 White Area = 71.9 %



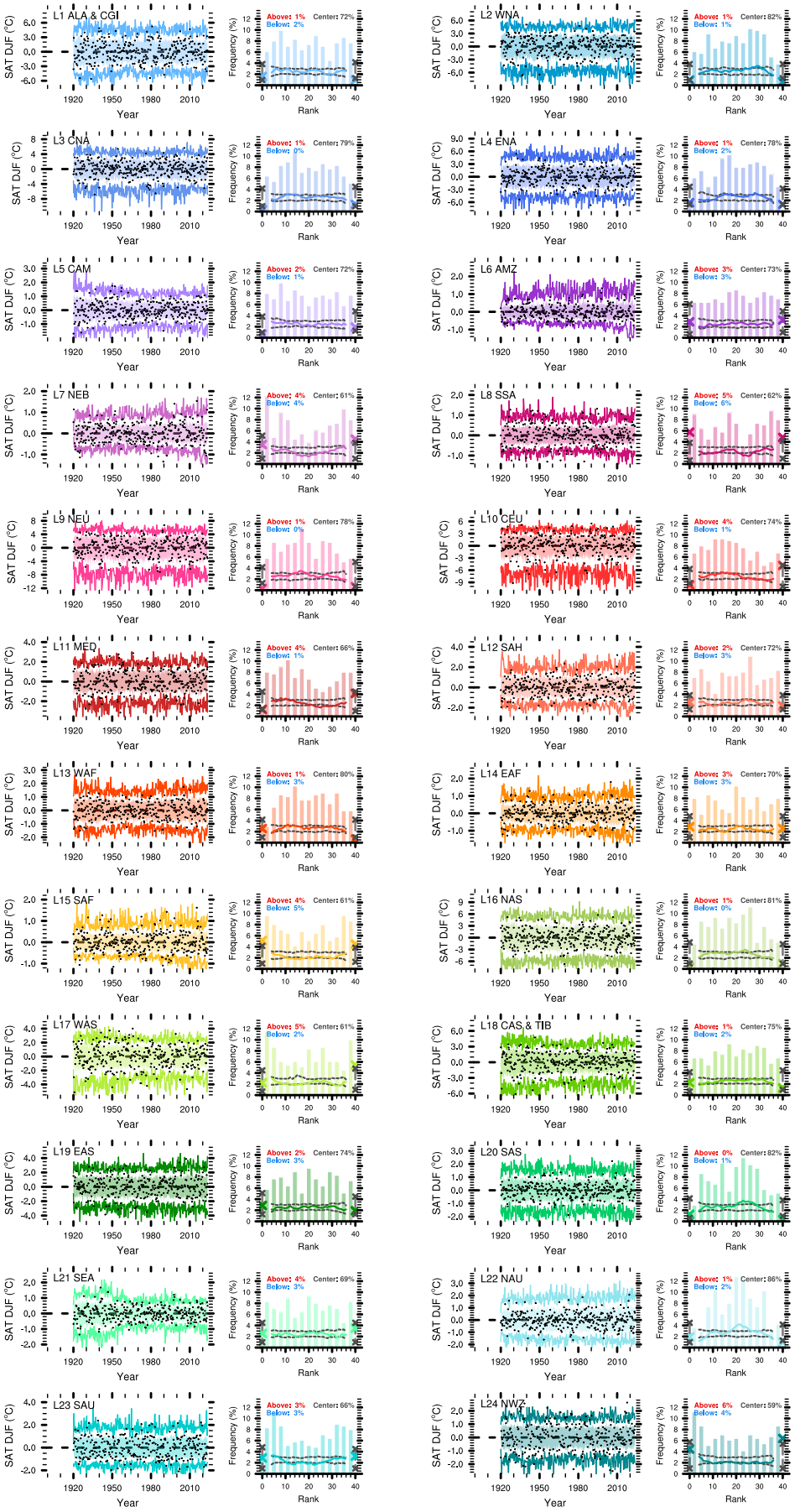
TAS JJA Obs inside central 75th percentile (%)



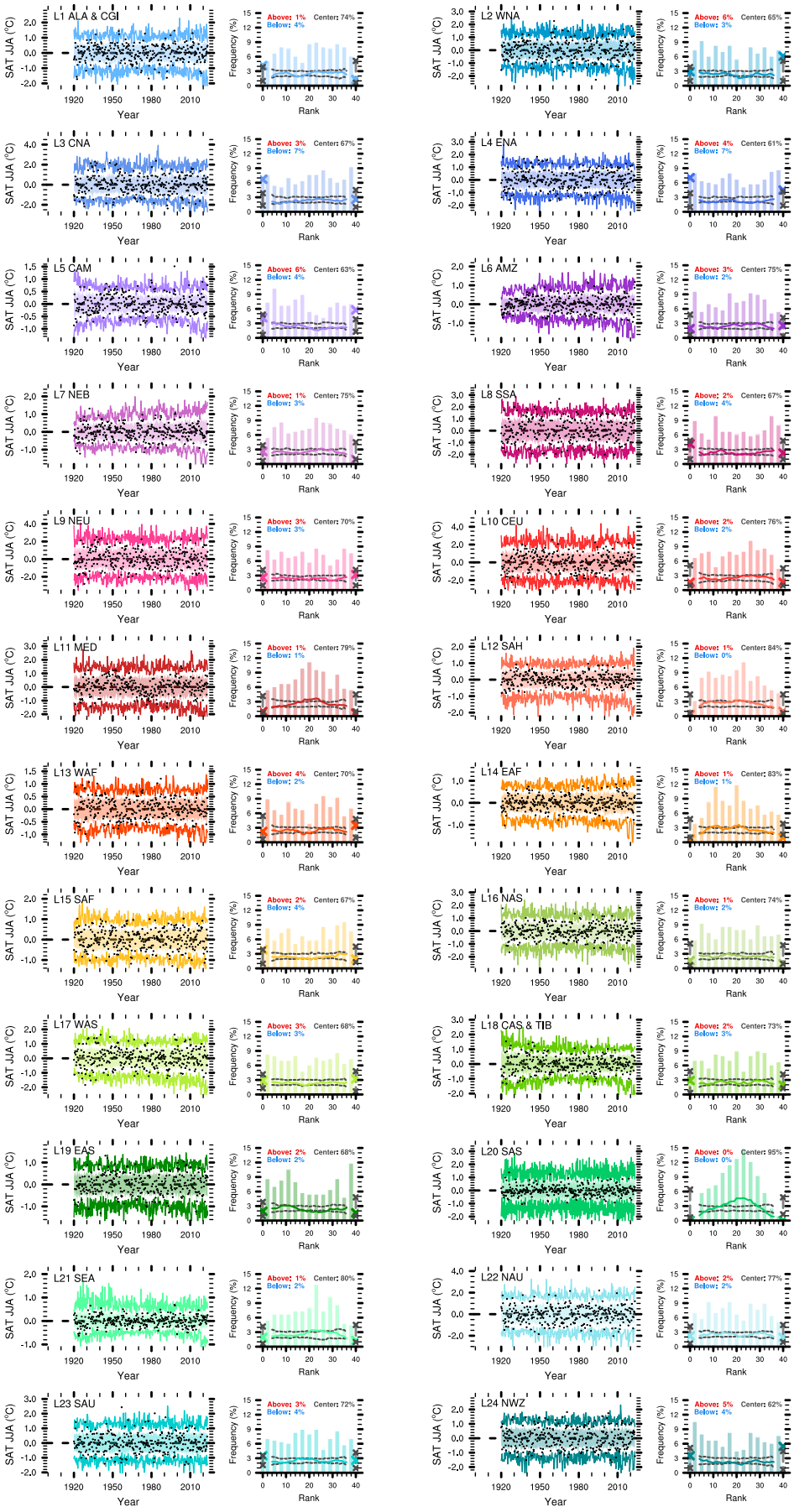
TAS JJA Obs outside ensemble spread (%)



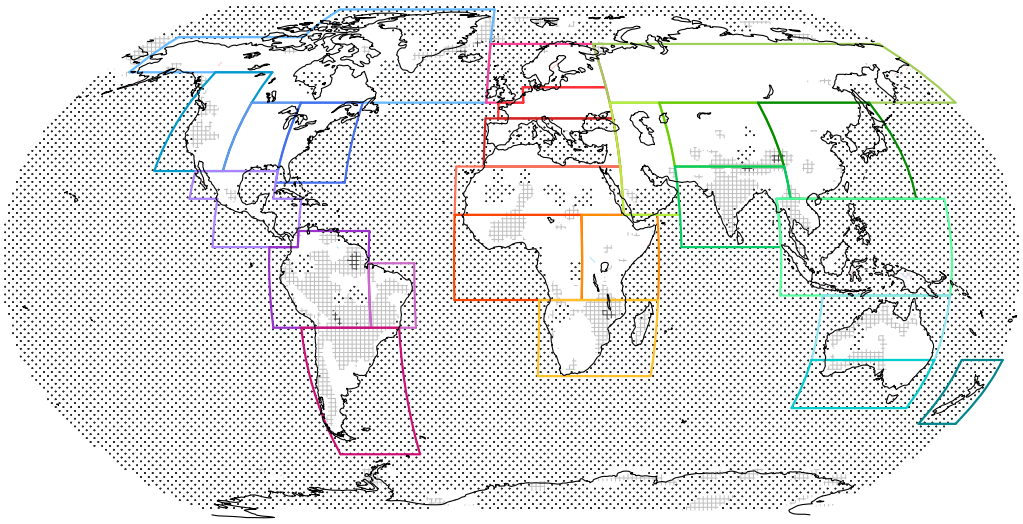
CESM-LE vs GISTEMPv4 TAS DJF



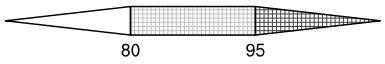
CESM-LE vs GISTEMPv4 TAS JJA



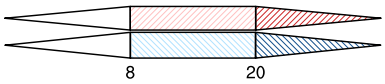
CESM2-LE vs GISTEMPv4 White Area = 79.0 %



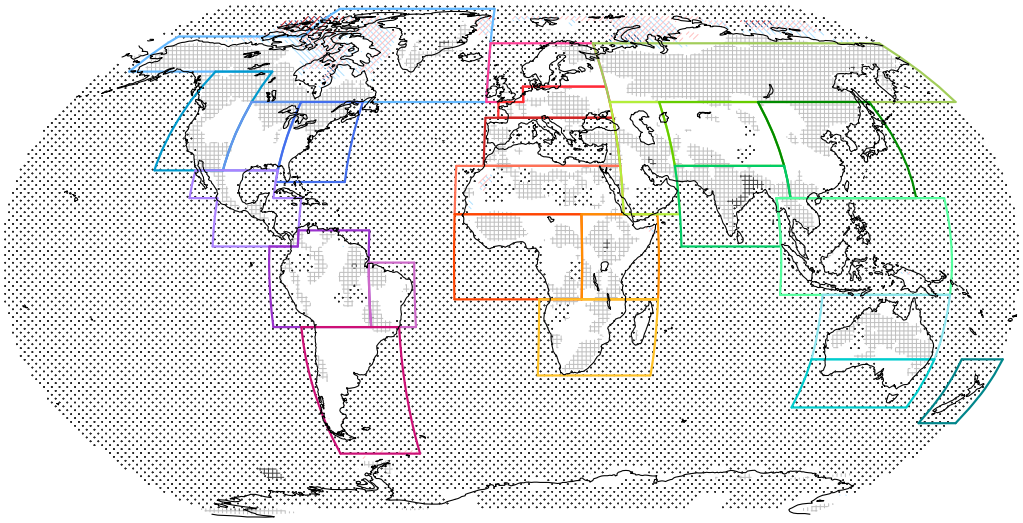
TAS DJF Obs inside central 75th percentile (%)



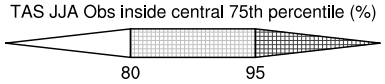
TAS DJF Obs outside ensemble spread (%)



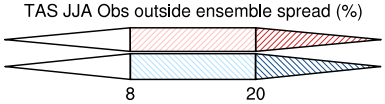
CESM2-LE vs GISTEMPv4 White Area = 54.8 %



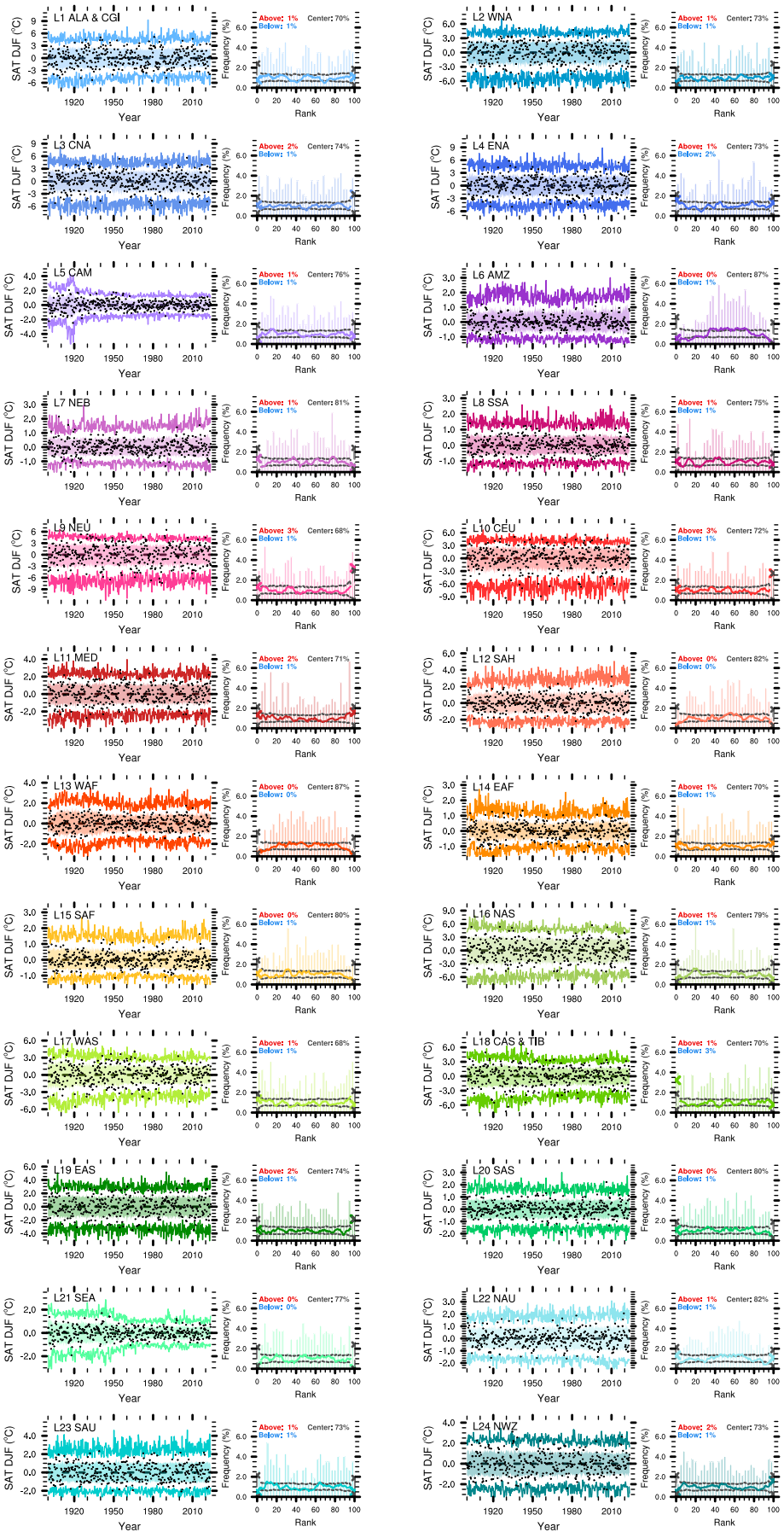
TAS JJA Obs inside central 75th percentile (%)



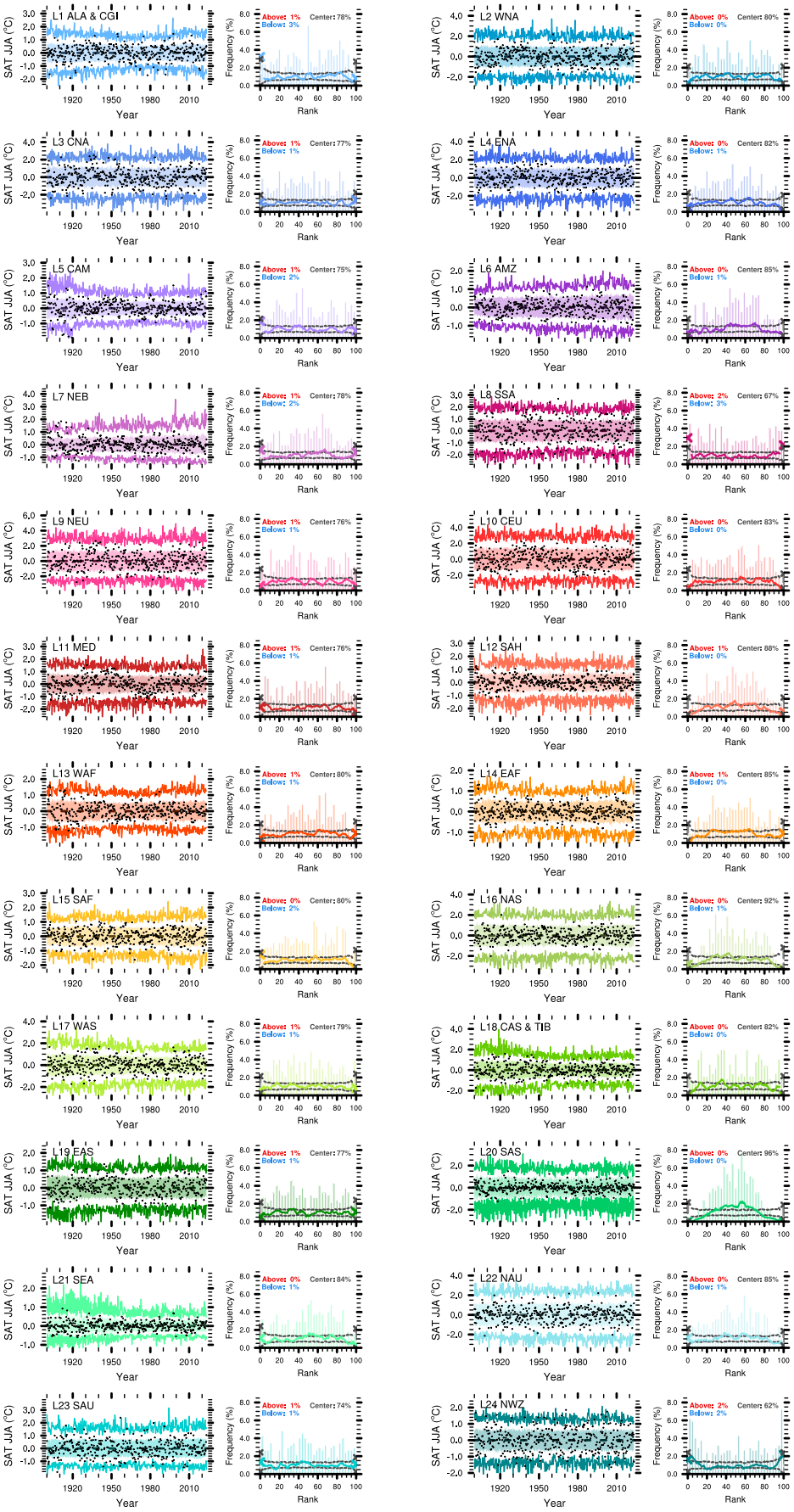
TAS JJA Obs outside ensemble spread (%)



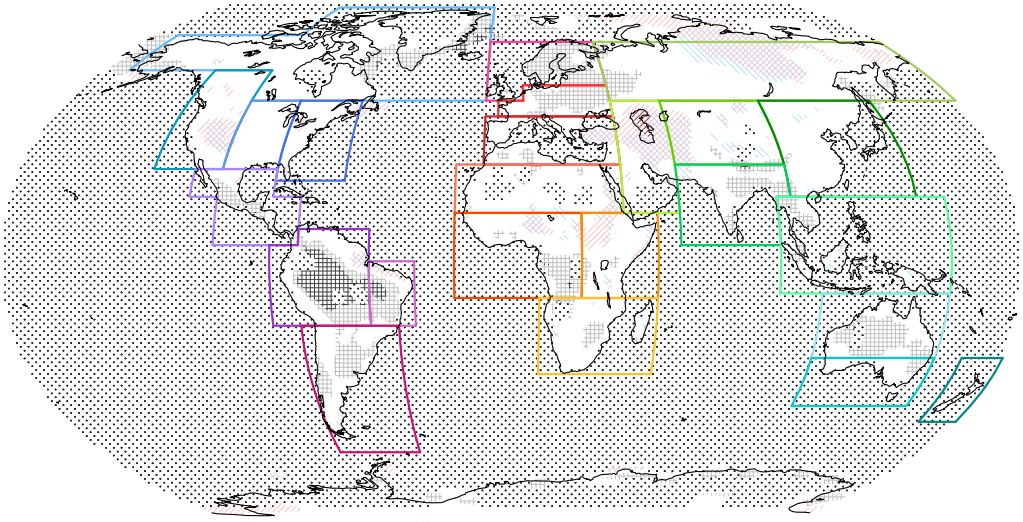
CESM2-LE vs GISTEMPv4 TAS DJF



CESM2-LE vs GISTEMPv4 TAS JJA



CSIRO-Mk360 vs GISTEMPv4 White Area = 62.0 %



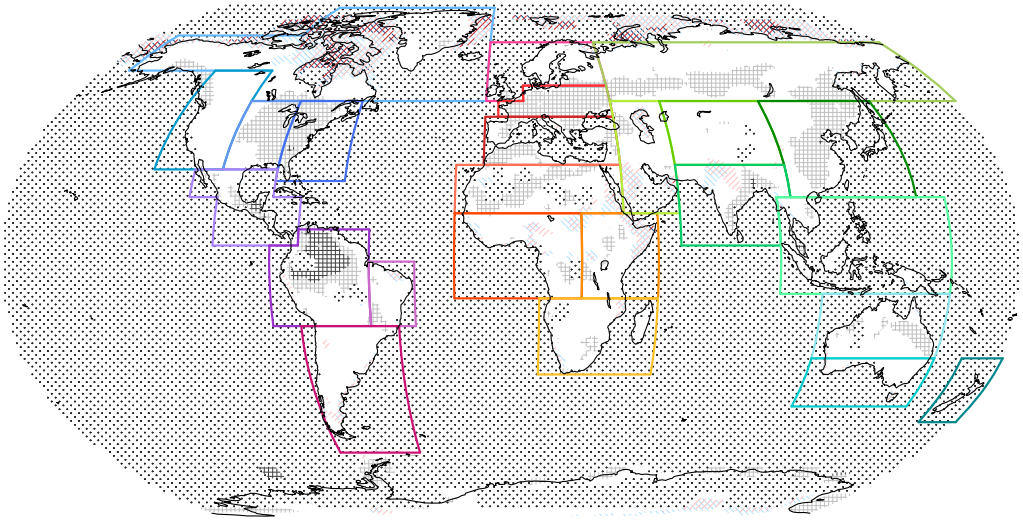
TAS DJF Obs inside central 75th percentile (%)

80 95

TAS DJF Obs outside ensemble spread (%)

8 20

CSIRO-Mk360 vs GISTEMPv4 White Area = 60.2 %



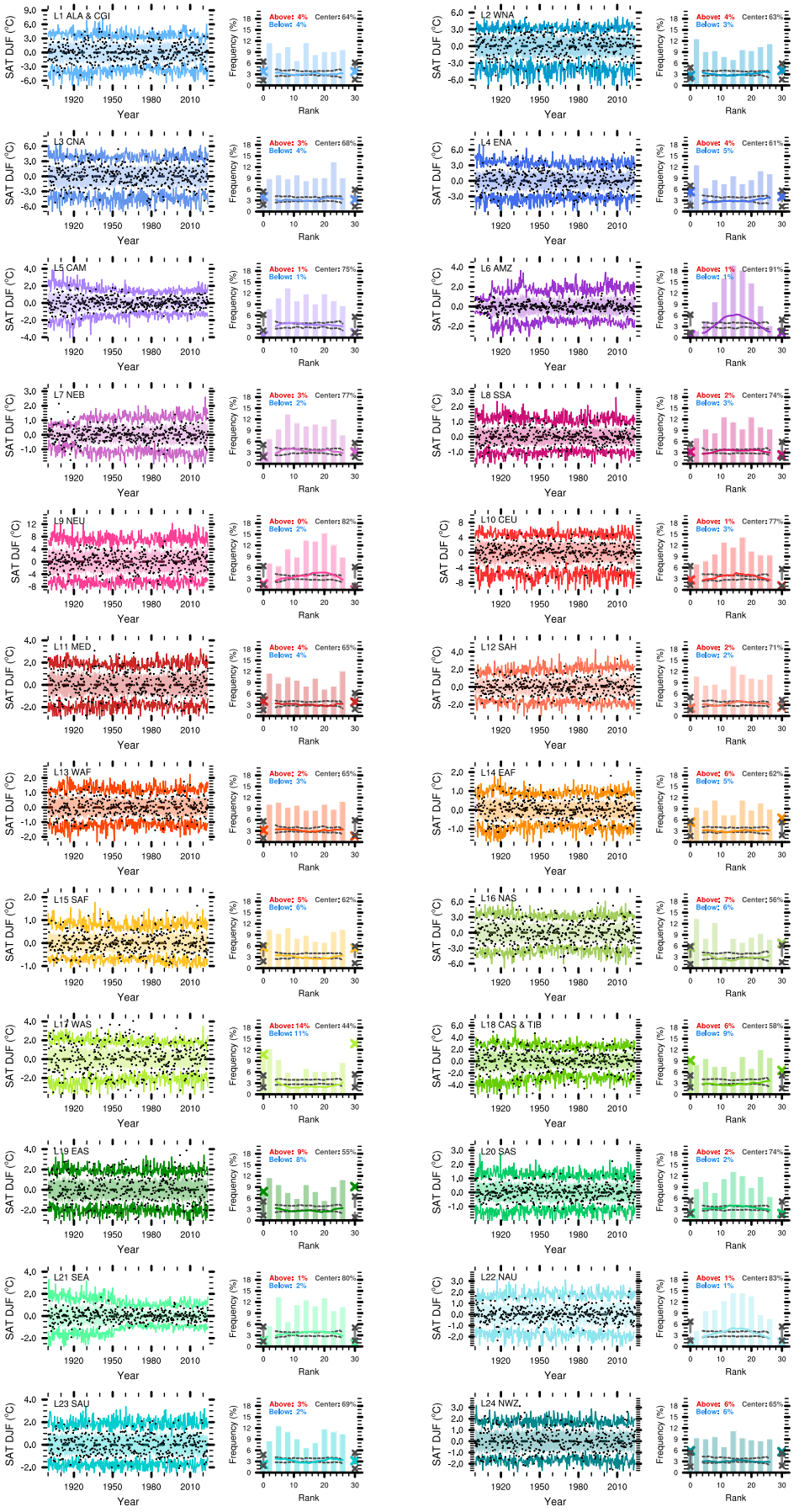
TAS JJA Obs inside central 75th percentile (%)

80 95

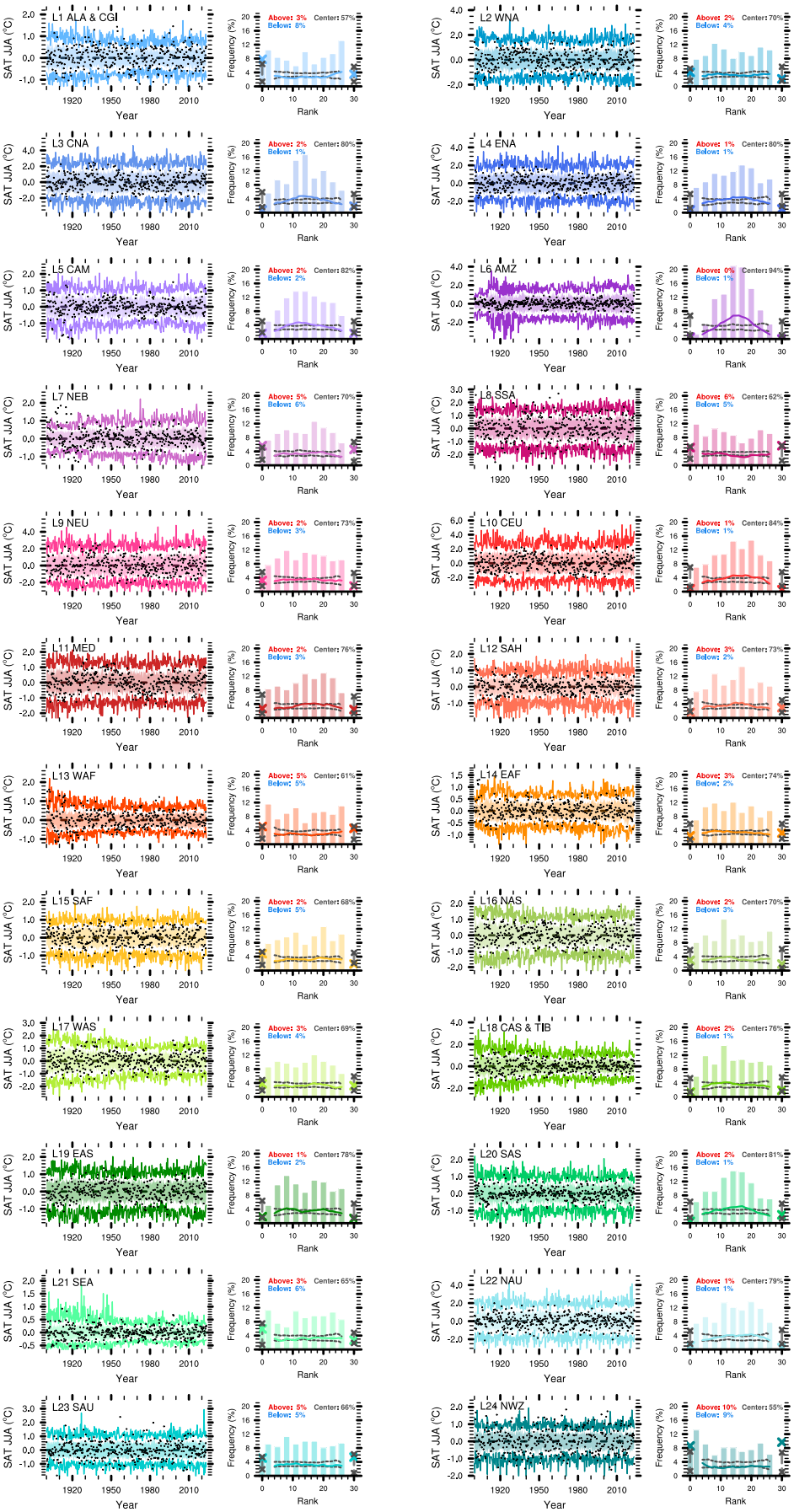
TAS JJA Obs outside ensemble spread (%)

8 20

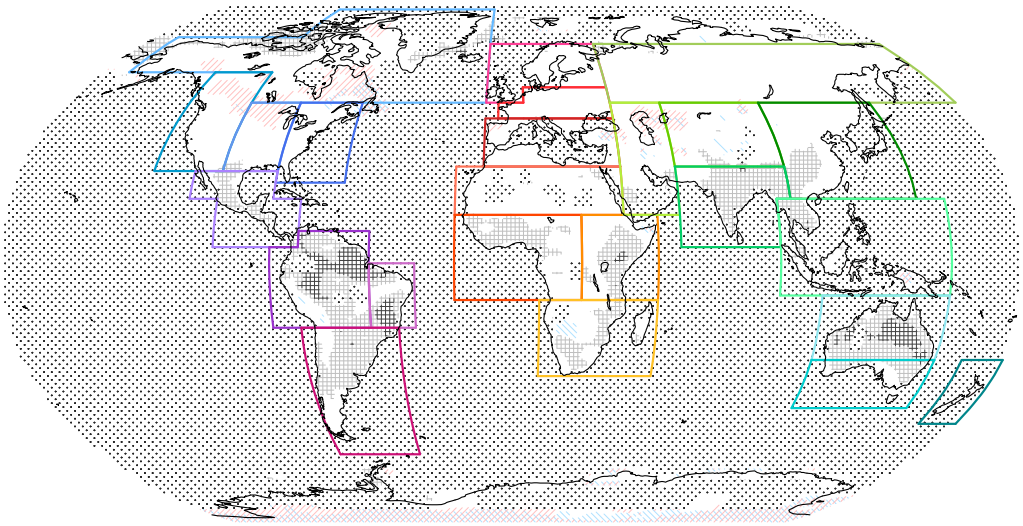
CSIRO-Mk360 vs GISTEMPv4 TAS DJF



CSIRO-Mk360 vs GISTEMPv4 TAS JJA



GFDL-ESM2M vs GISTEMPv4 White Area = 63.5 %



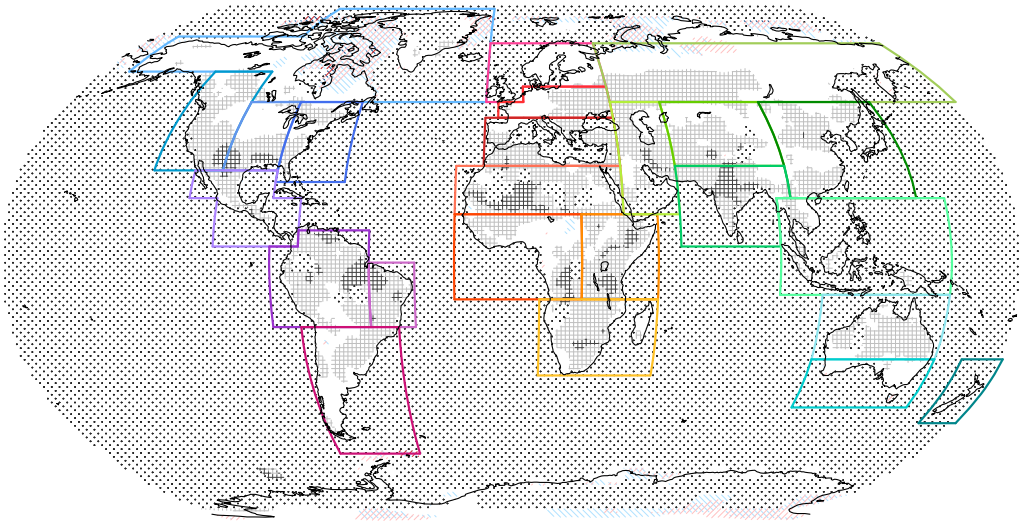
TAS DJF Obs inside central 75th percentile (%)

80 95

TAS DJF Obs outside ensemble spread (%)

8 20

GFDL-ESM2M vs GISTEMPv4 White Area = 42.0 %



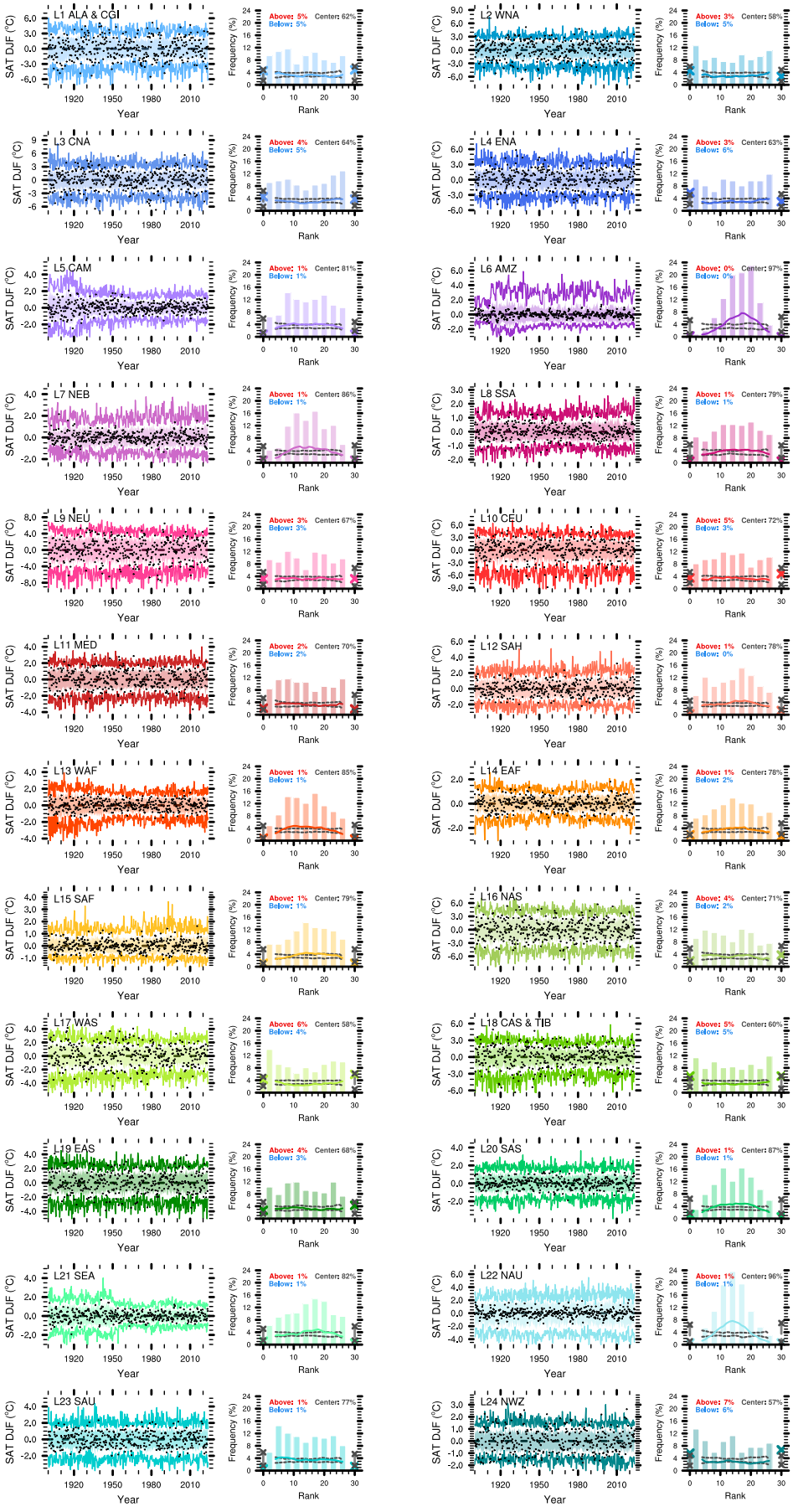
TAS JJA Obs inside central 75th percentile (%)

80 95

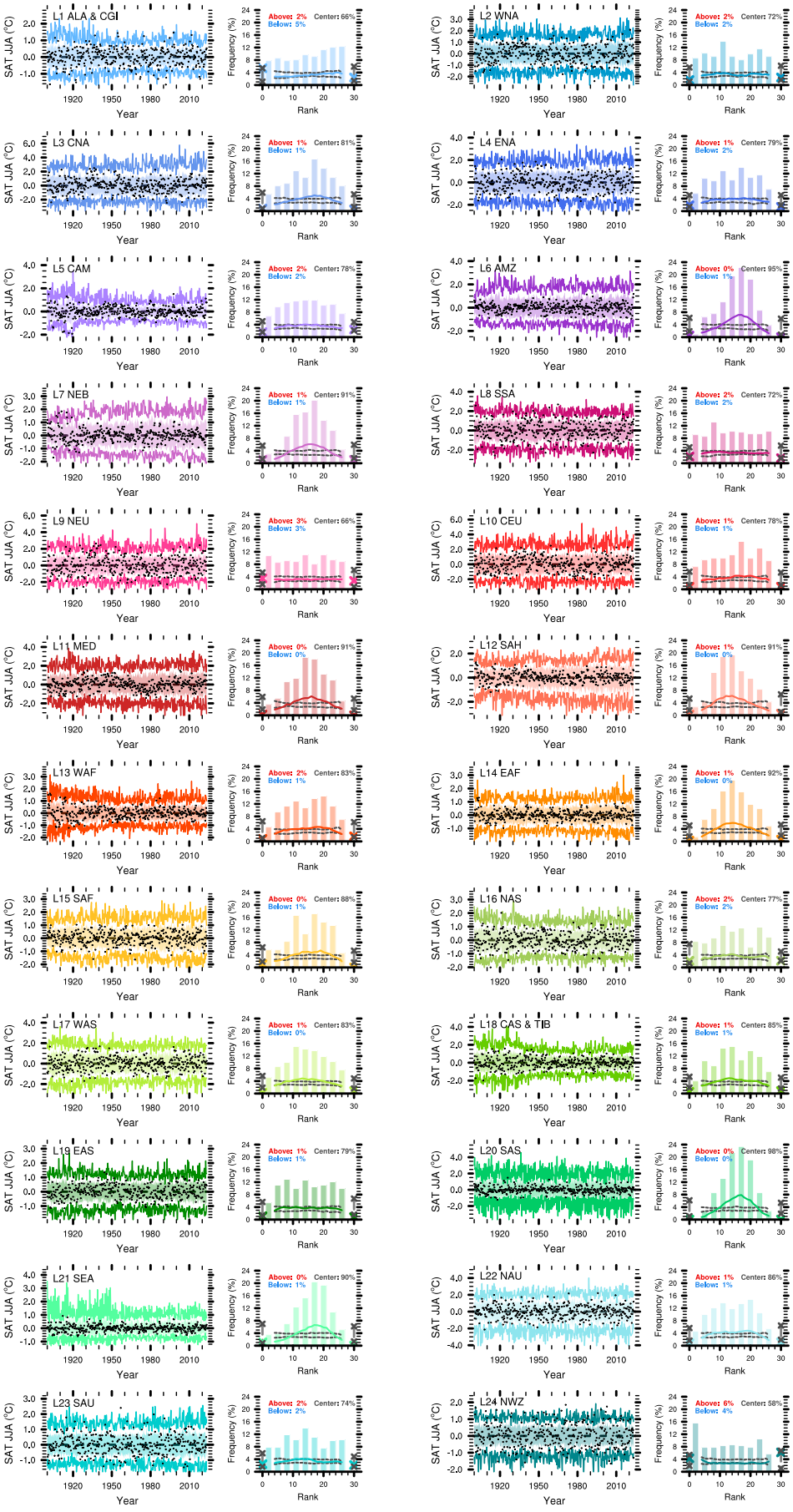
TAS JJA Obs outside ensemble spread (%)

8 20

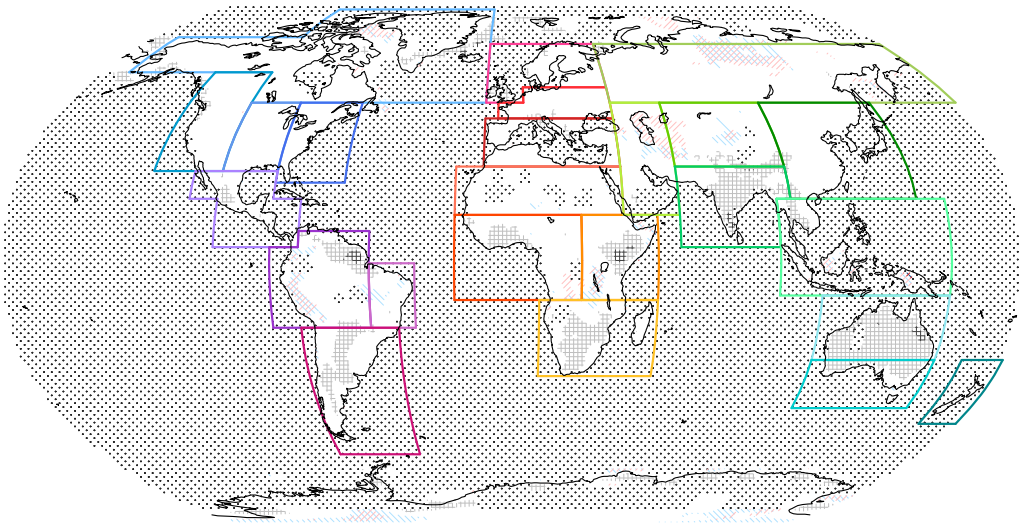
GFDL-ESM2M vs GISTEMPv4 TAS DJF



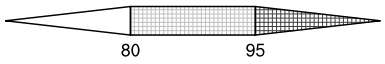
GFDL-ESM2M vs GISTEMPv4 TAS JJA



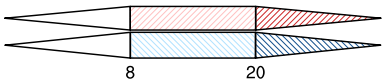
GFDL-SPEAR vs GISTEMPv4 White Area = 74.3 %



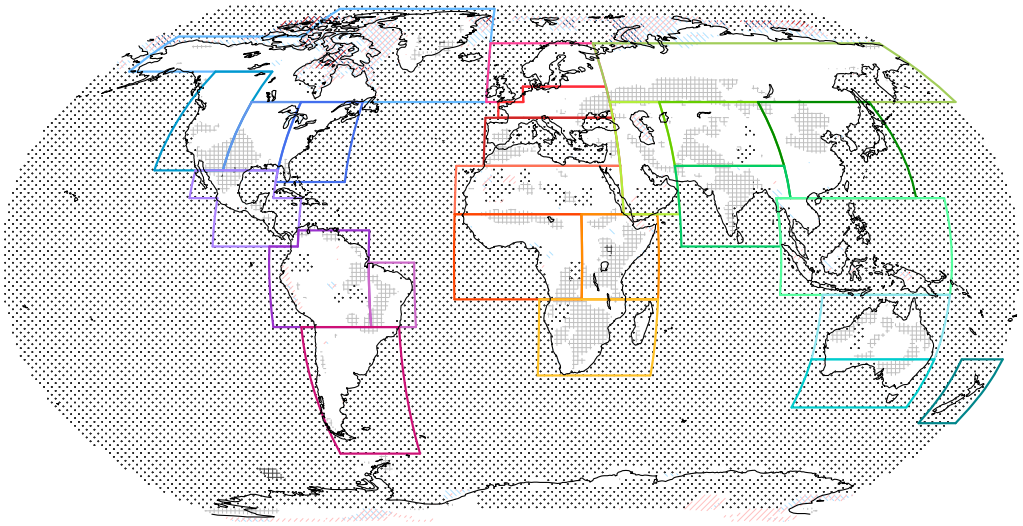
TAS DJF Obs inside central 75th percentile (%)



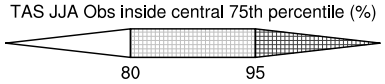
TAS DJF Obs outside ensemble spread (%)



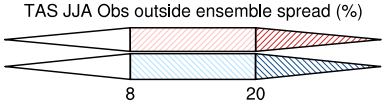
GFDL-SPEAR vs GISTEMPv4 White Area = 68.1 %



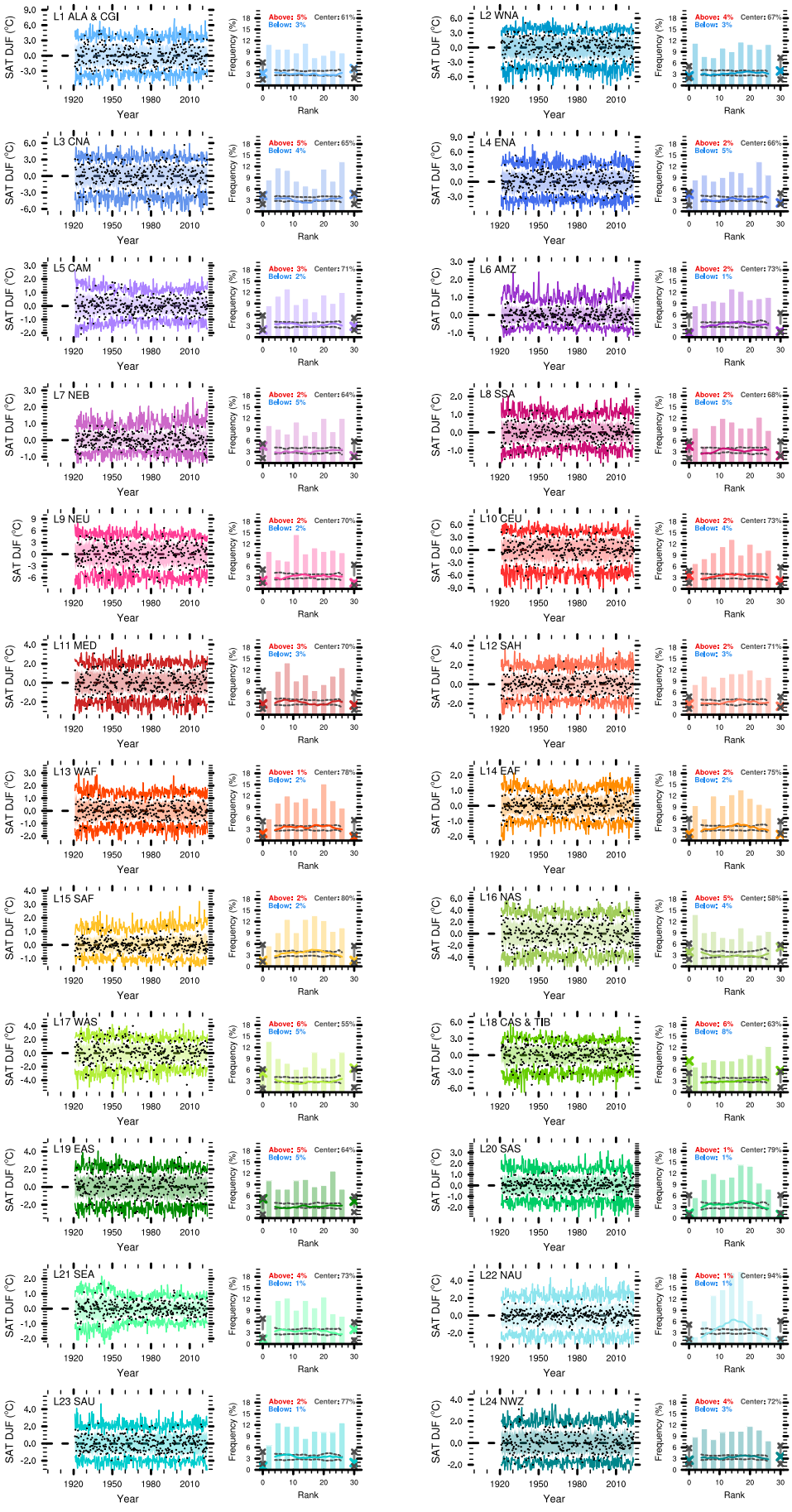
TAS JJA Obs inside central 75th percentile (%)



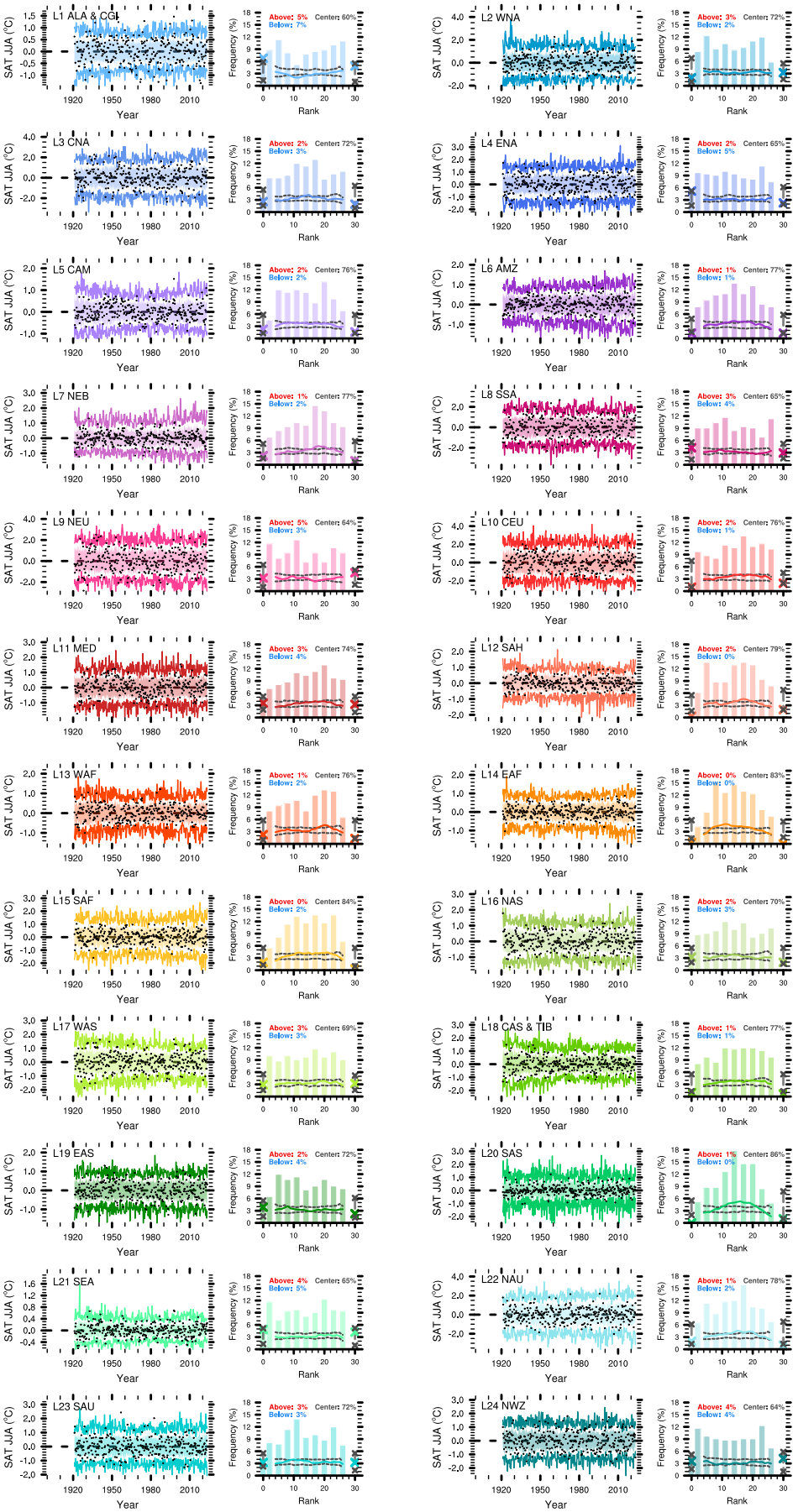
TAS JJA Obs outside ensemble spread (%)

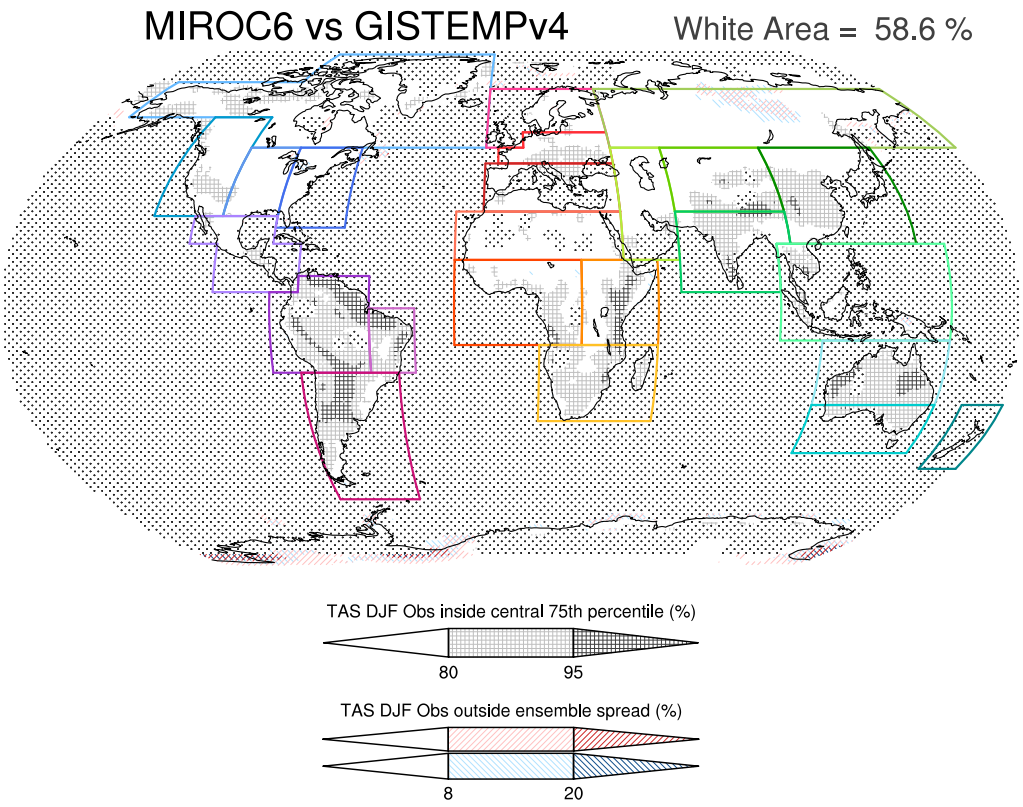


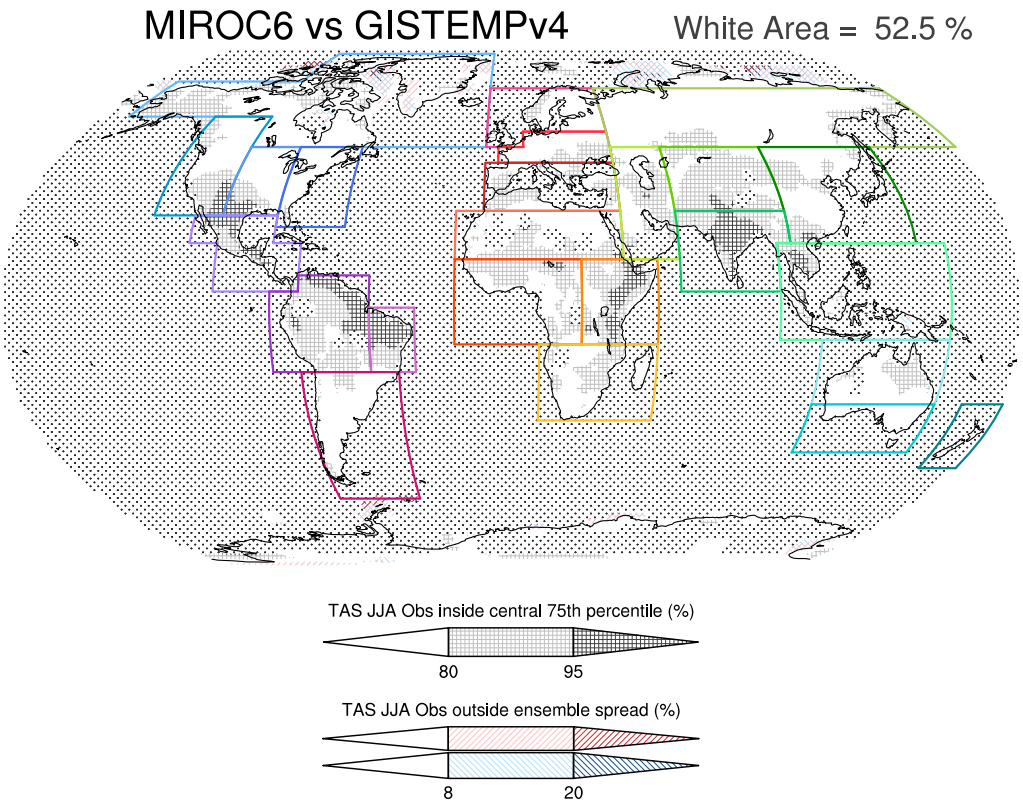
GFDL-SPEAR vs GISTEMPv4 TAS DJF



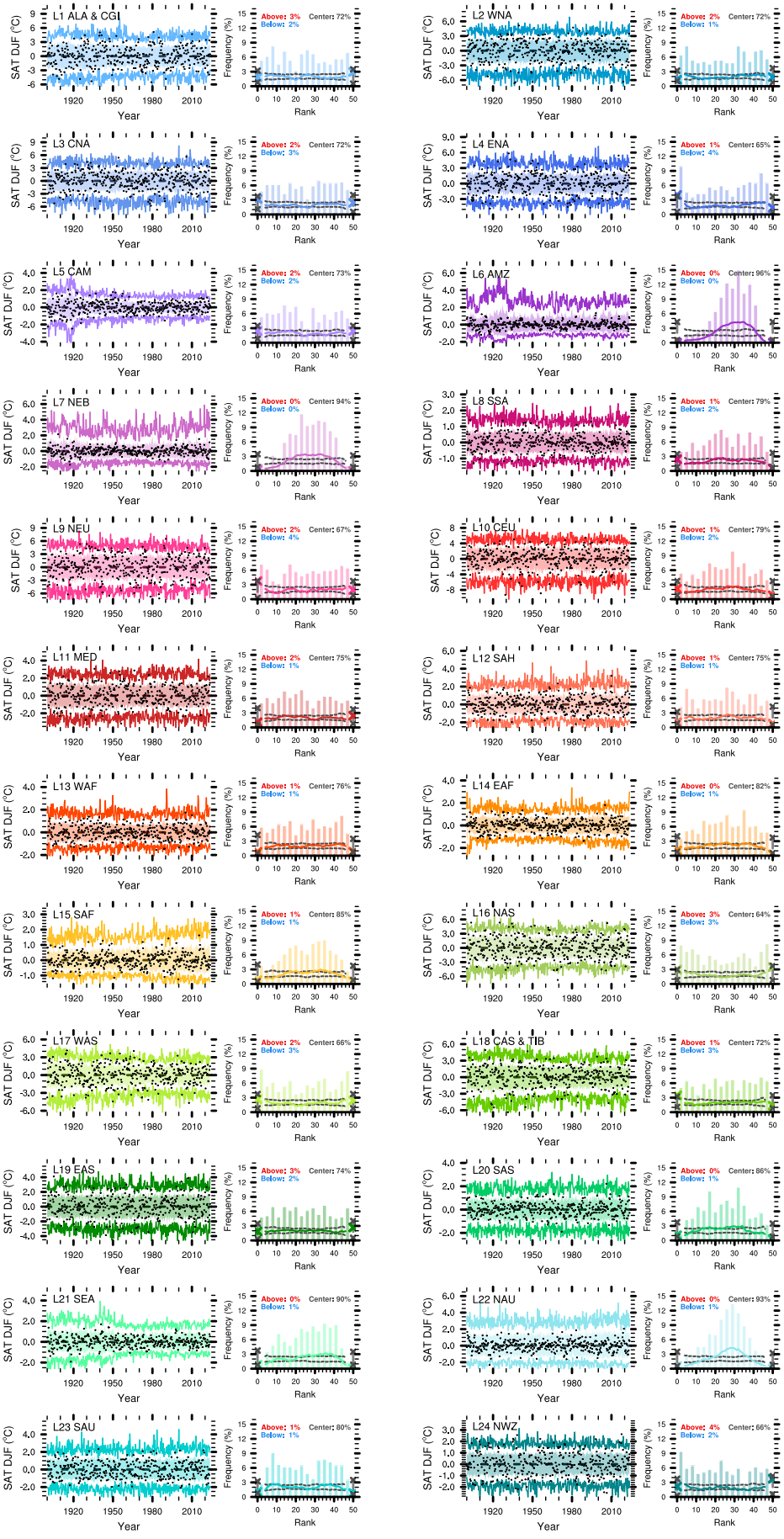
GFDL-SPEAR vs GISTEMPv4 TAS JJA



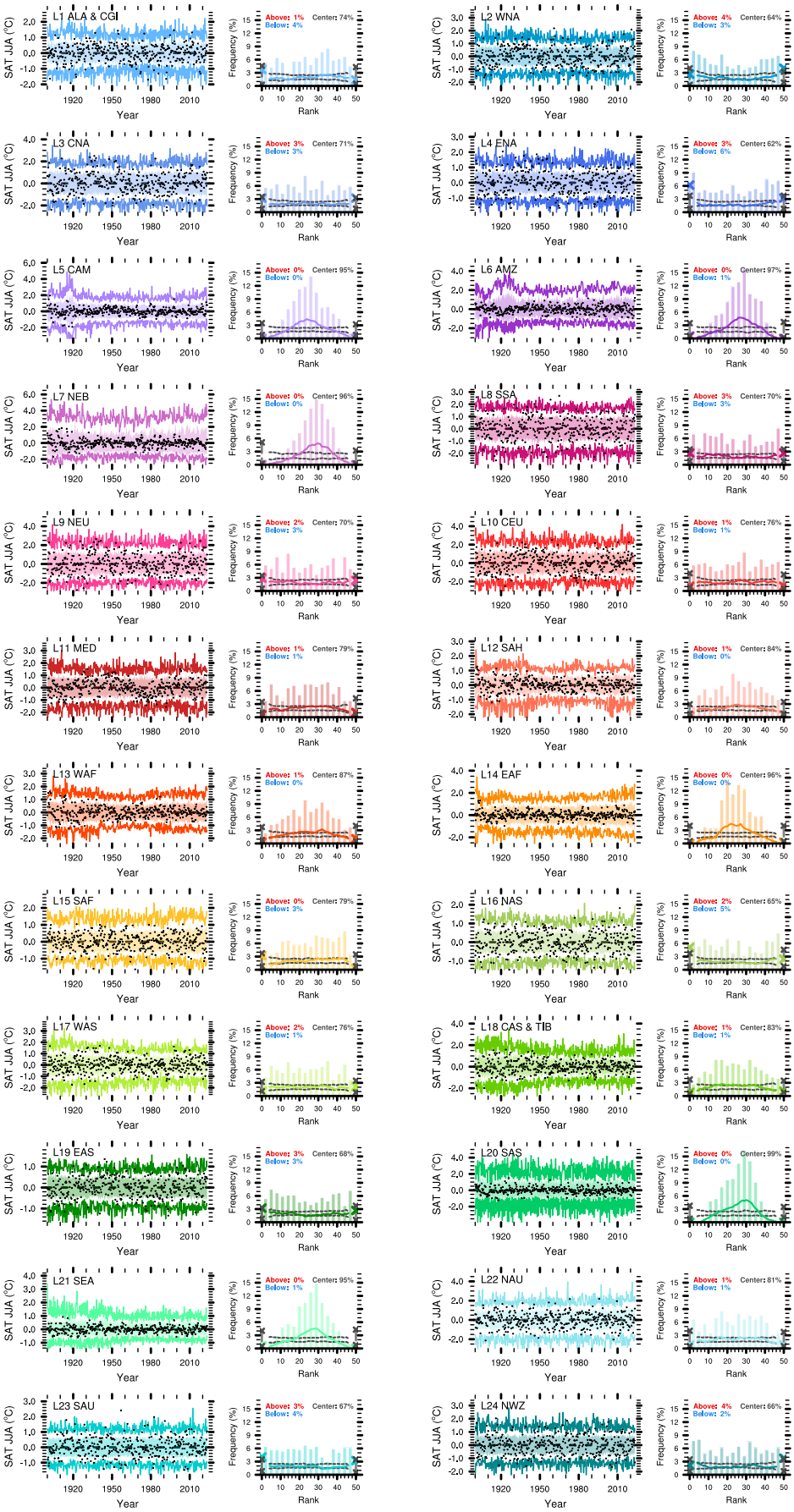

MIROC6 vs GISTEMPv4
White Area = 58.6 %
TAS DJF Obs inside central 75th percentile (%)
80
95
TAS DJF Obs outside ensemble spread (%)
8
20


MIROC6 vs GISTEMPv4
White Area = 52.5 %
TAS JJA Obs inside central 75th percentile (%)
80
95
TAS JJA Obs outside ensemble spread (%)
8
20

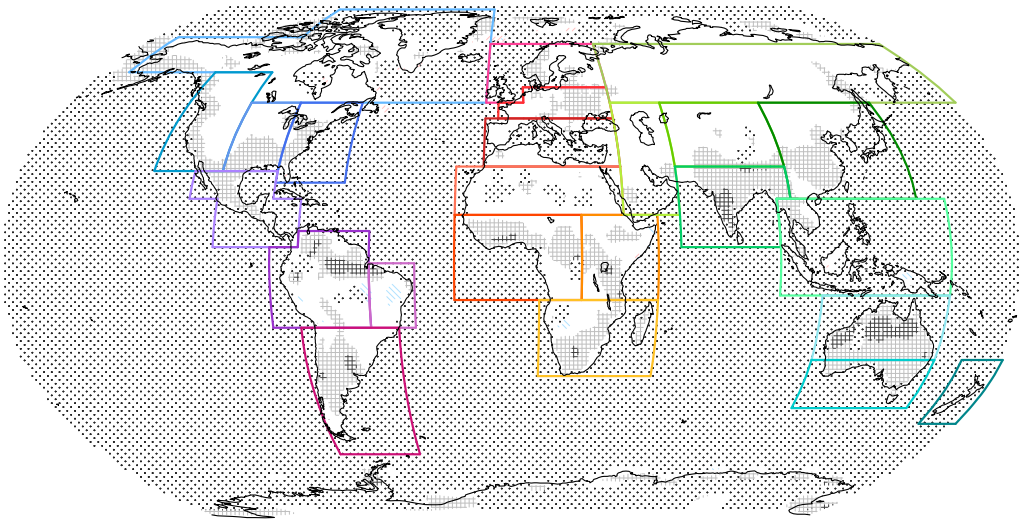
MIROC6 vs GISTEMPv4 TAS DJF



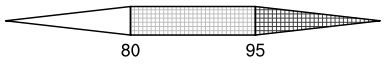
MIROC6 vs GISTEMPv4 TAS JJA



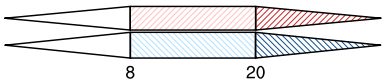
MPI-GE5 vs GISTEMPv4 White Area = 65.6 %



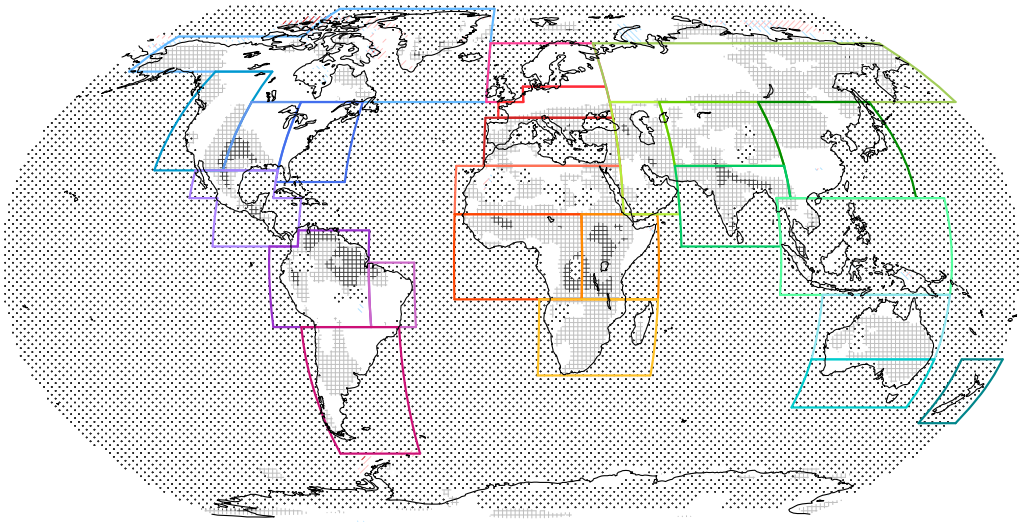
TAS DJF Obs inside central 75th percentile (%)



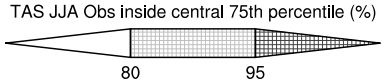
TAS DJF Obs outside ensemble spread (%)



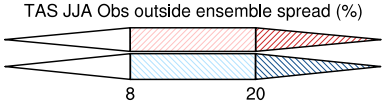
MPI-GE5 vs GISTEMPv4 White Area = 51.0 %



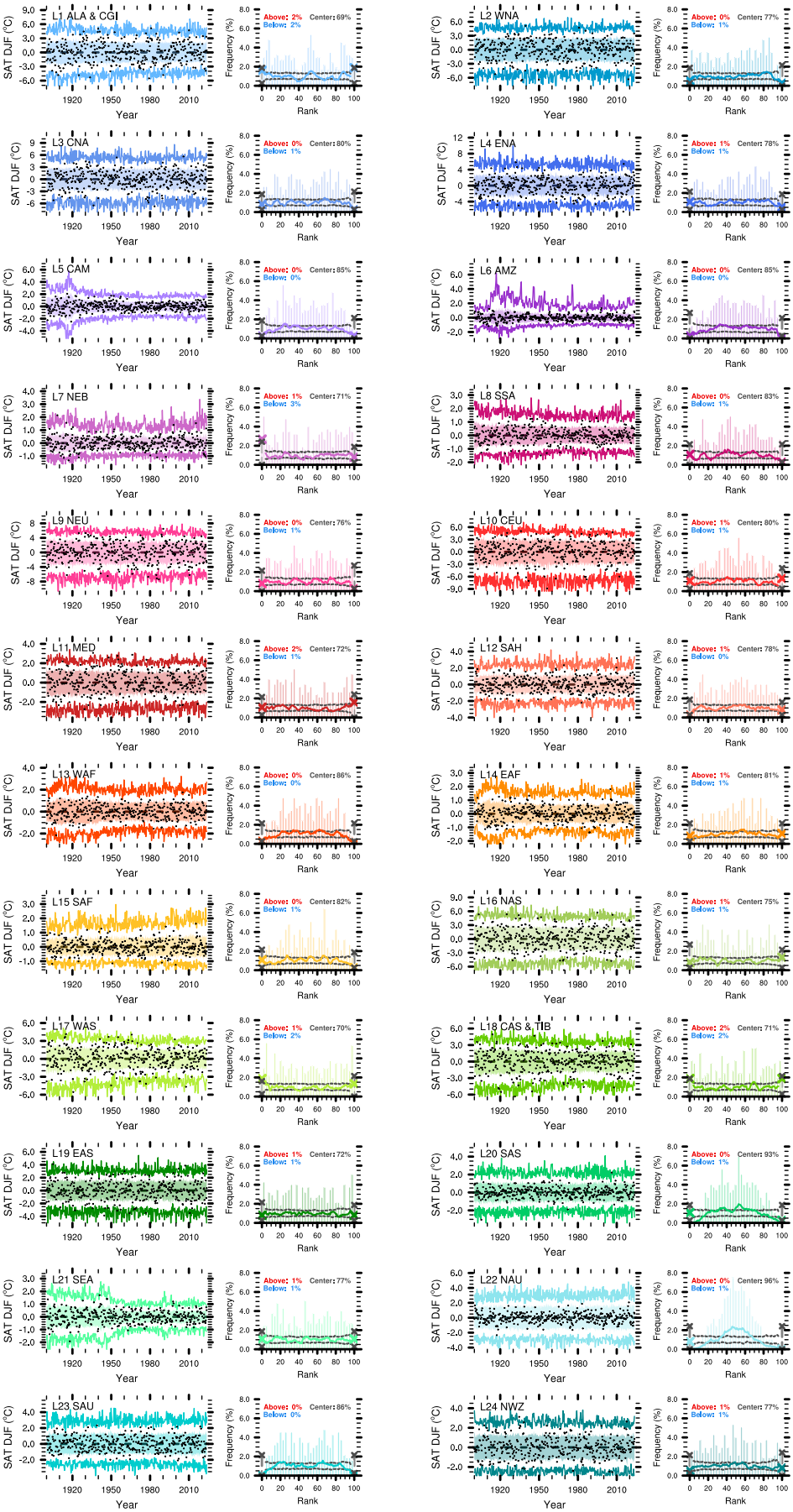
TAS JJA Obs inside central 75th percentile (%)



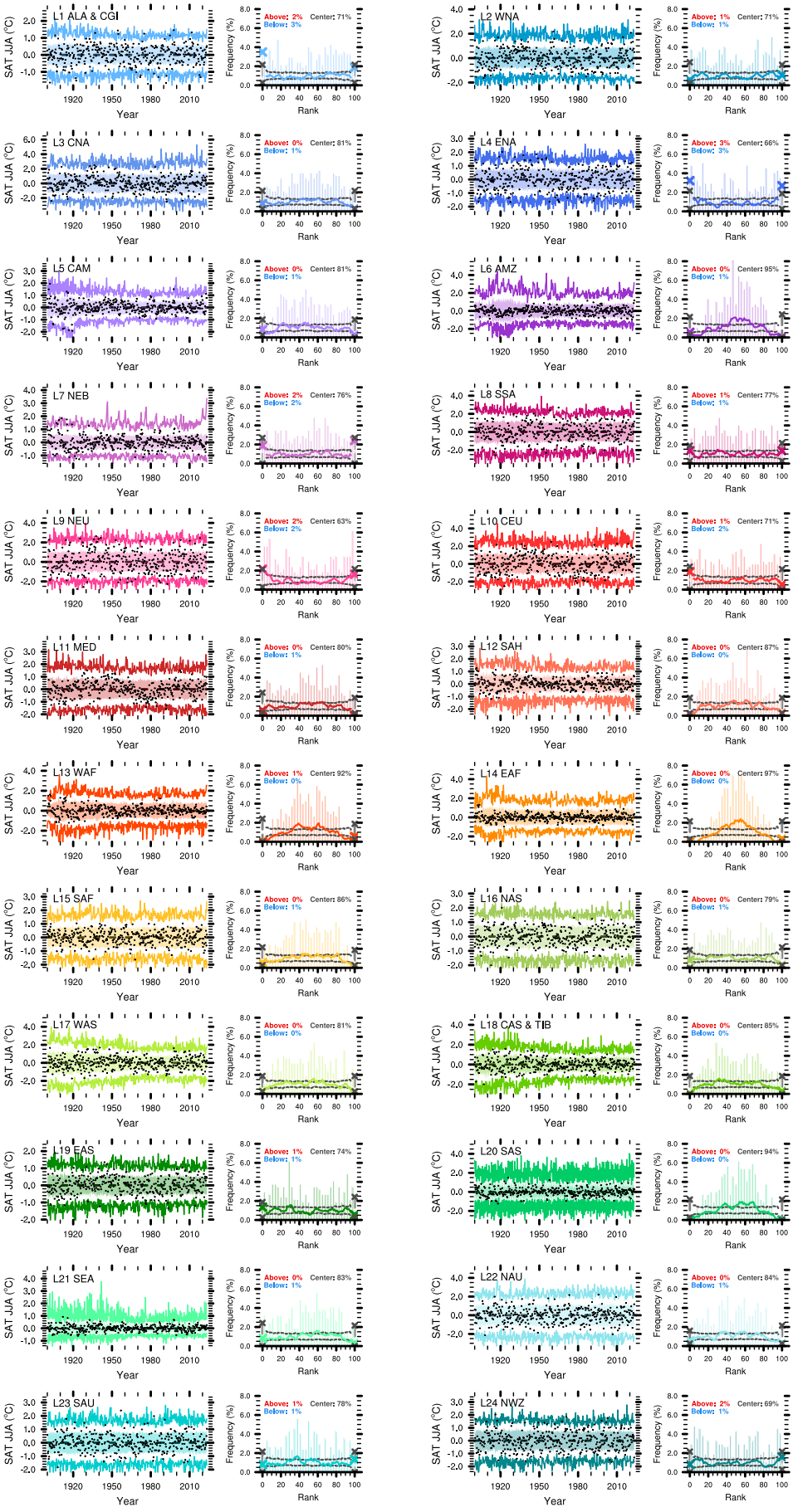
TAS JJA Obs outside ensemble spread (%)



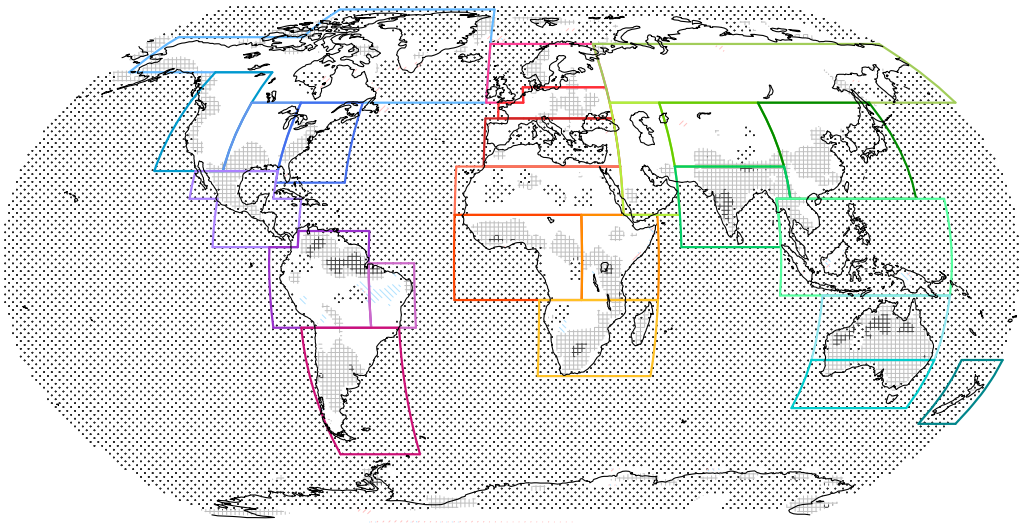
MPI-GE5 vs GISTEMPv4 TAS DJF



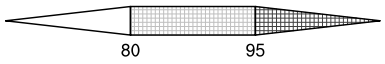
MPI-GE5 vs GISTEMPv4 TAS JJA



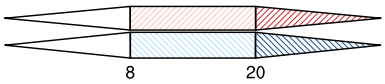
MPI-GE6 vs GISTEMPv4 White Area = 68.0 %



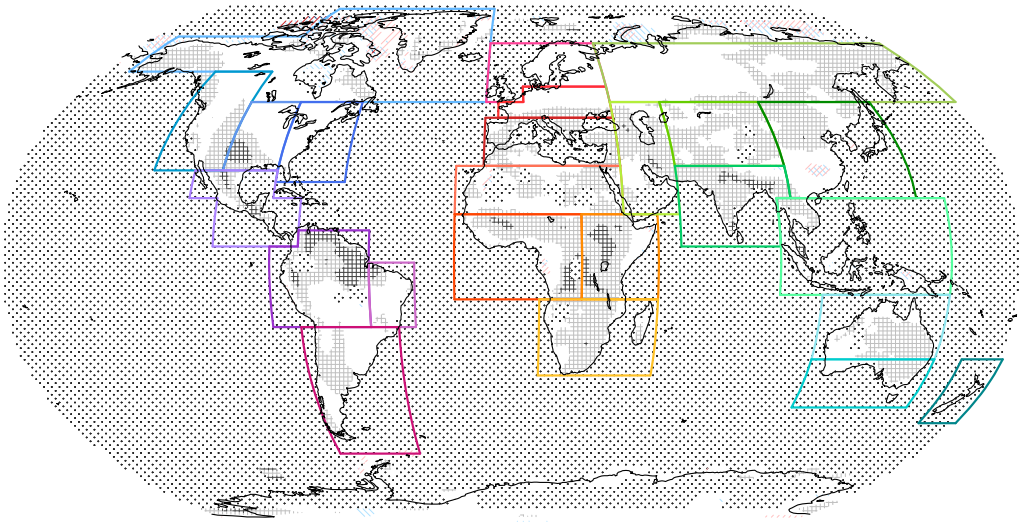
TAS DJF Obs inside central 75th percentile (%)



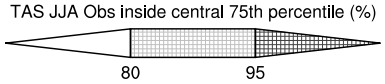
TAS DJF Obs outside ensemble spread (%)



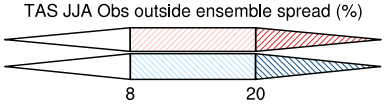
MPI-GE6 vs GISTEMPv4 White Area = 50.7 %



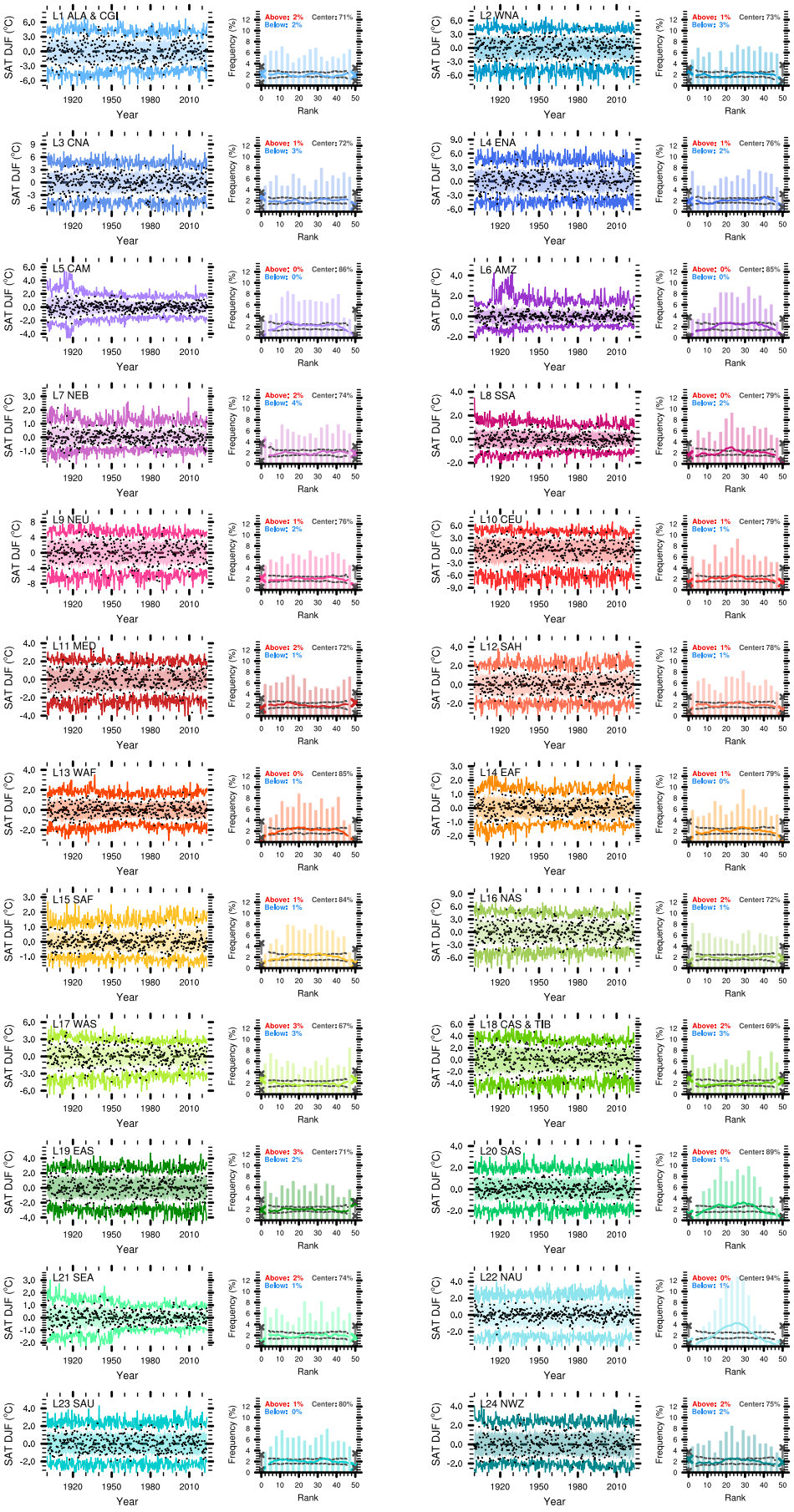
TAS JJA Obs inside central 75th percentile (%)



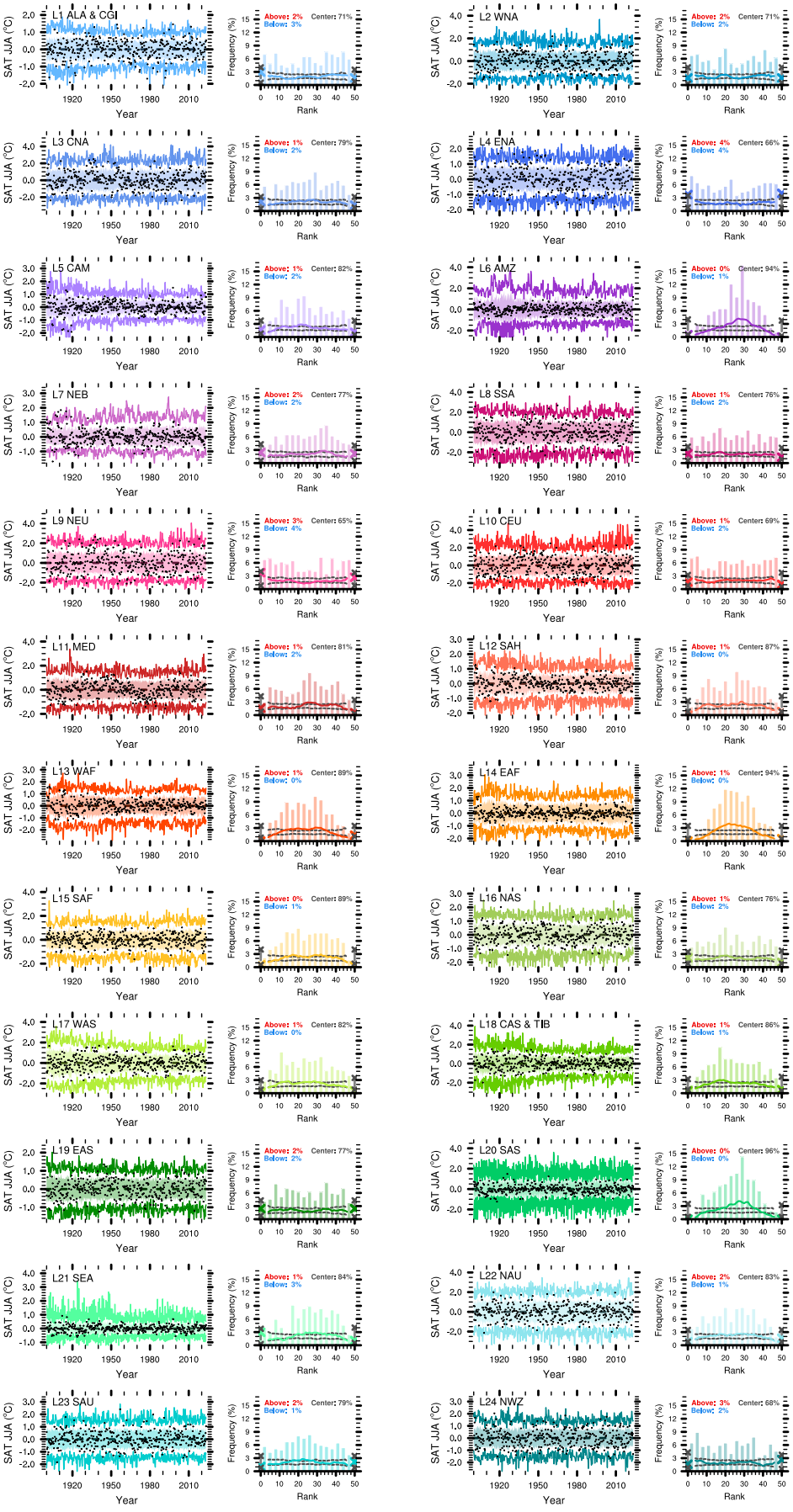
TAS JJA Obs outside ensemble spread (%)



MPI-GE6 vs GISTEMPv4 TAS DJF



MPI-GE6 vs GISTEMPv4 TAS JJA



Non-Detrended Land Surface Air Temperatures

Rank-frequency variability evaluation framework for non-detrended 2m air temperature (TAS) anomalies over land grid cells.

Maps show grid-cell evaluation of the simulated DJF and JJA monthly mean temperature anomalies for the 11 SMILEs included in this study against GISTEMPv4 observations globally. Gray hatching represents where observations cluster within the 75th percentile bounds of the ensemble (12.5th to 87.5th percentiles) for more than 80% of months (light grey) or for more than 95% of months (dark grey). Red and blue shading represents where observations are larger than the ensemble maximum (red) or smaller than the ensemble minimum (blue), respectively, for more than 8% of the months (light red and blue) or for more than 20% of the months (dark red and blue). Dotted areas represent ocean areas or grid cells where observations are missing and are therefore excluded from this analysis. Colored boxes demark the boundaries of each land region assessed. The percentage of assessed grid-cells that present none of these biases in given at the top (white area).

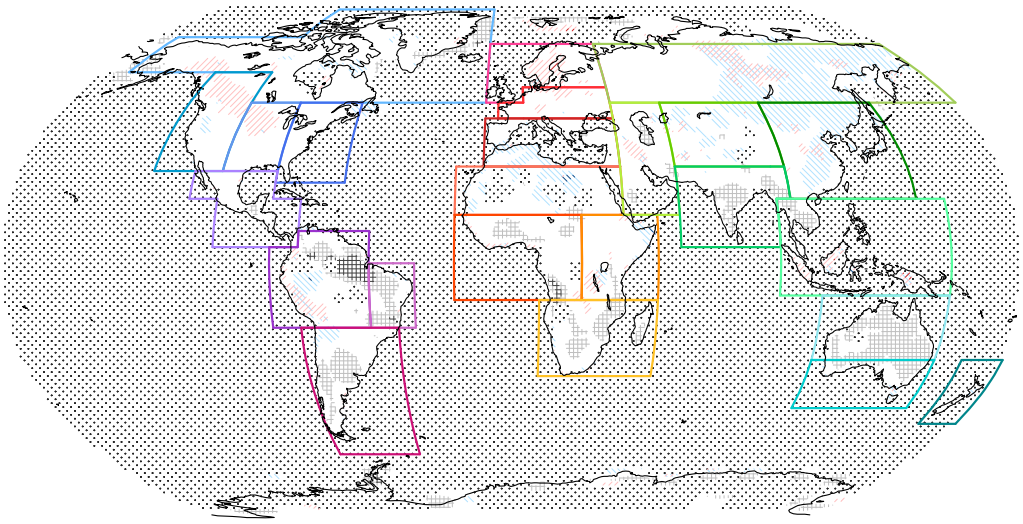
Time series and rank frequency histograms show spatially aggregated DJF and JJA TAS for each land regions for all 11 SMILEs. Time series show the ensemble maximum and minimum (coloured lines) and central 75th percentile ensemble spread (shading) are shown against observations (black dots).

Rank histograms represent the frequency of each place that observations would take in a list of ensemble members ordered by ascending temperature anomaly values. Rank 0 indicates observations are below the minimum ensemble value, and rank n, with n the number of ensemble members, indicates that observations exceed the maximum ensemble value for that particular month. For a model that perfectly represents observations over an infinitely-long observational record, all ranks should occur with similar frequency and this histogram should be roughly flat.

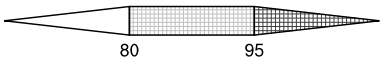
To illustrate how internal variability may affect rank frequencies given the non-infinite record length considered, we also include a perfect-model comparison, which shows the range of rank frequencies that each ensemble member would take if it were observations. If the rank exhibited by observations (colors) is within this perfect-model range (grey), the rank frequency evaluation shows an adequate model performance, and any deviations from a perfectly frank rank histogram can be assumed to be within deviations that could be caused by internal variability. Lines in the rank histogram illustrate the rank histogram's slope, as the mean rank frequency over a centered 6-bin window for observations (solid colored lines), and the 5-95th perc. perfect model range (gray dashed lines). Crosses represent the frequency of minimum (0) and maximum (number of members) ranks for observations (colors), and for the 5-95th perc. perfect model range (gray).

Percentages at the top of the rank histograms show how often monthly anomalies fall above or below ensemble limits (red and blue, respectively) and within the 75th perc. Range (grey), analogous to the criteria chosen for the map-based evaluation but for spatially aggregated values.

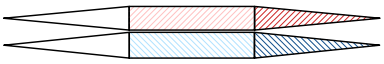
ACCESS vs GISTEMPv4 White Area = 61.3 %



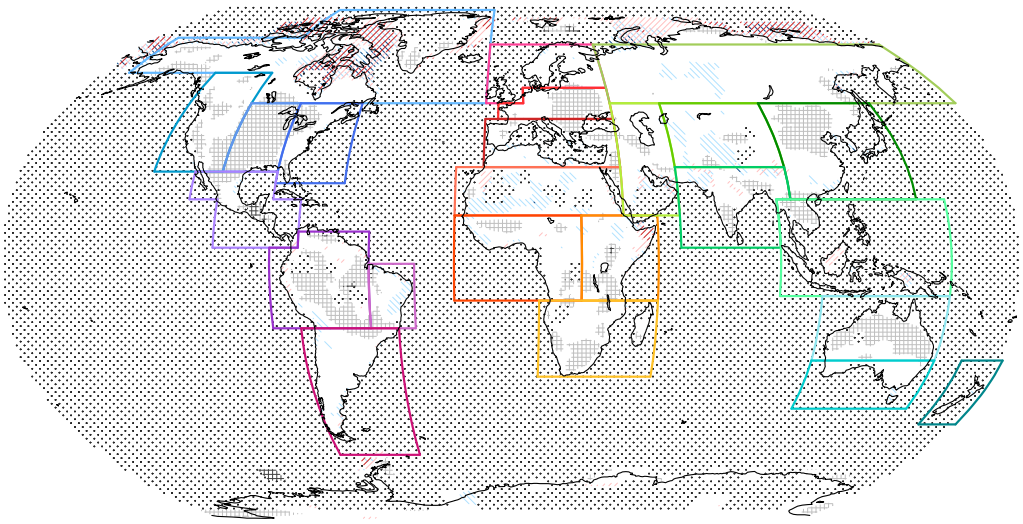
TAS DJF Obs inside central 75th percentile (%)



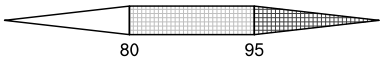
TAS DJF Obs outside ensemble spread (%)



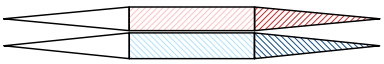
ACCESS vs GISTEMPv4 White Area = 59.0 %



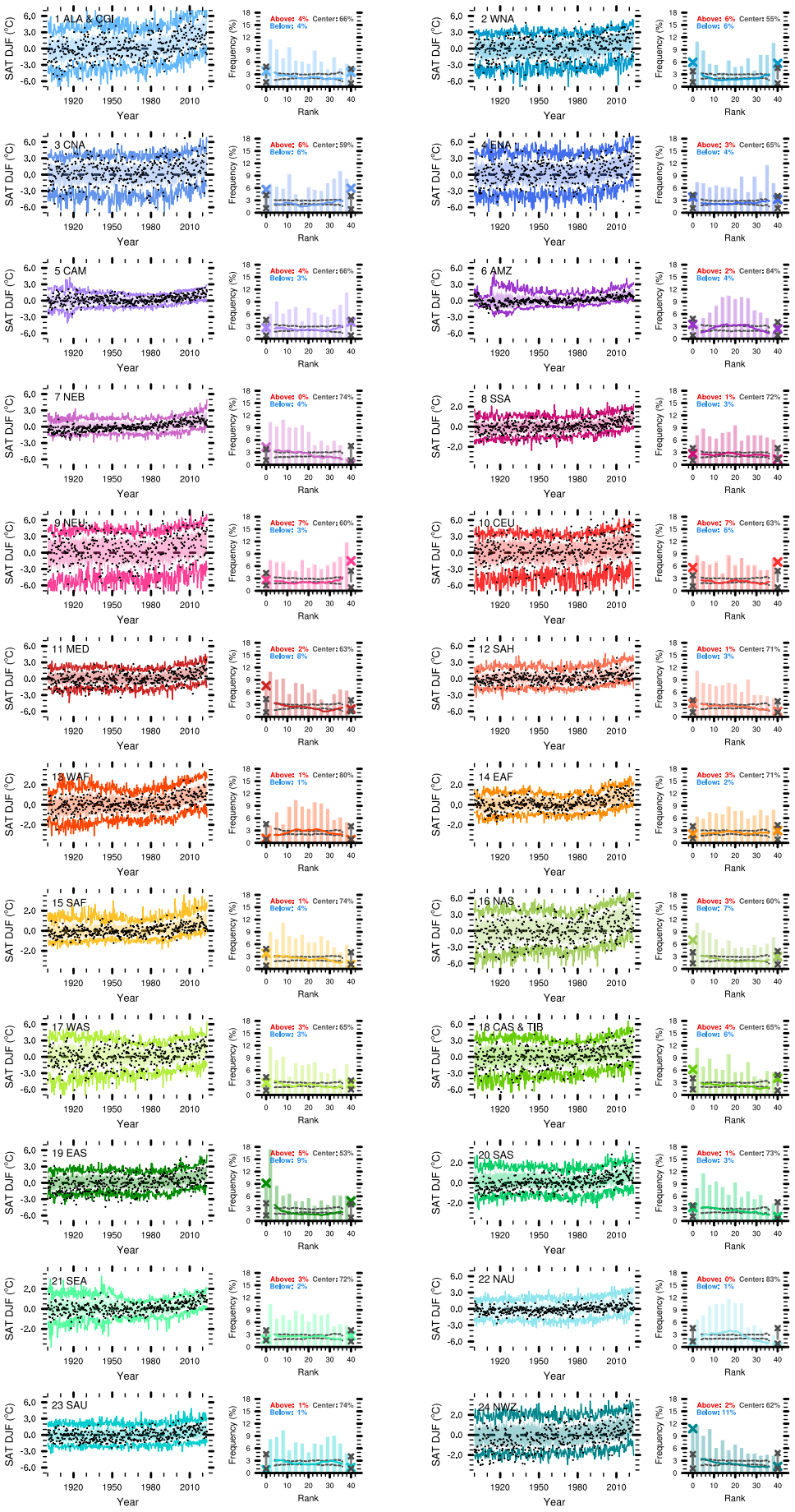
TAS JJA Obs inside central 75th percentile (%)



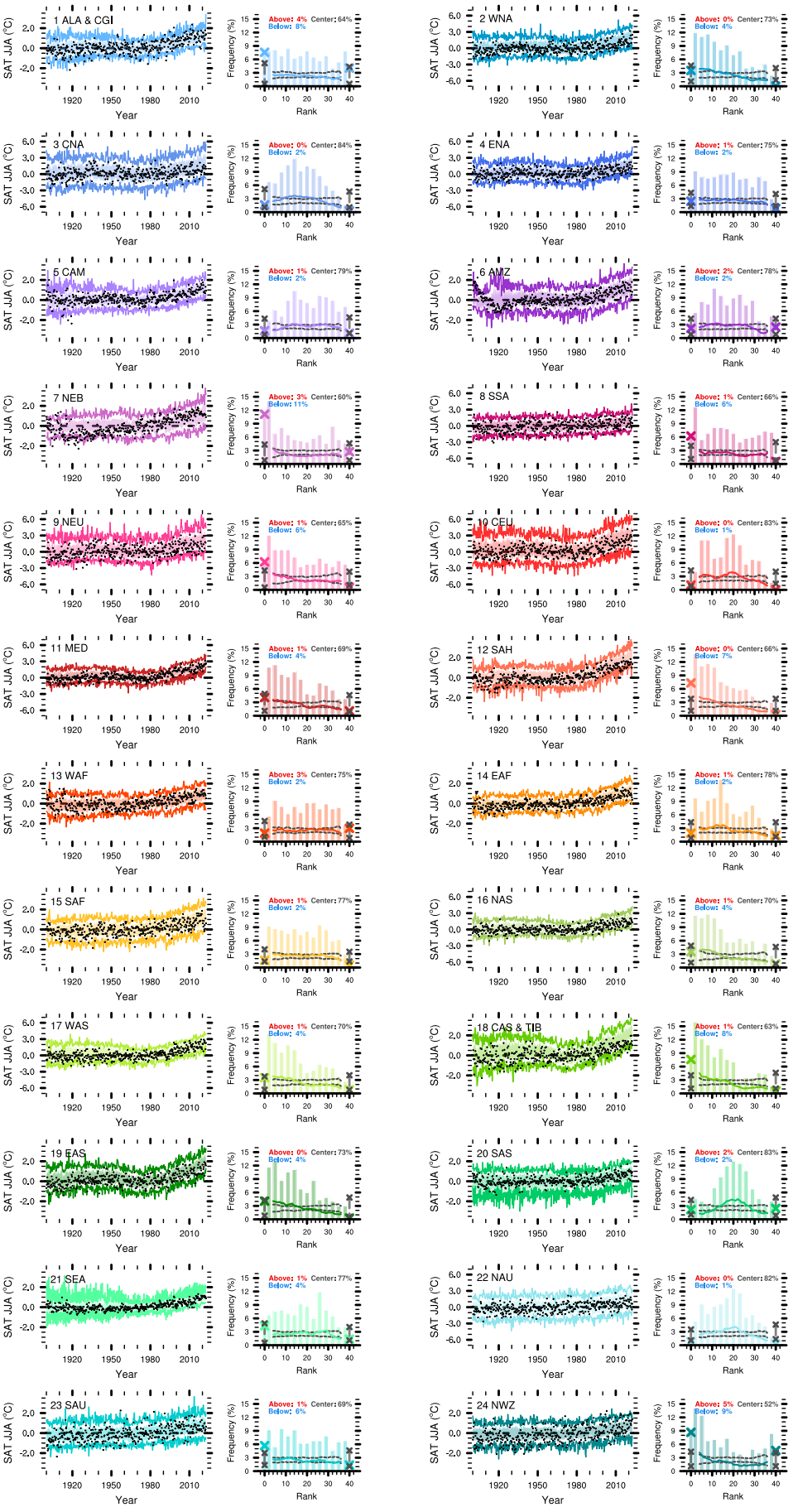
TAS JJA Obs outside ensemble spread (%)



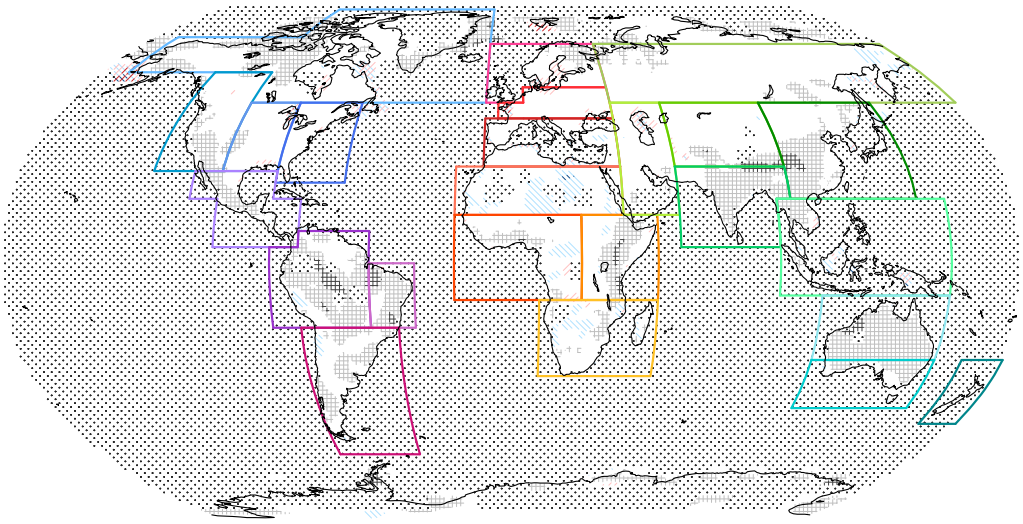
ACCESS vs GISTEMPv4 TAS DJF



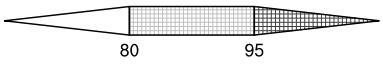
ACCESS vs GISTEMPv4 TAS JJA



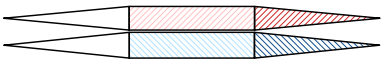
CanESM2 vs GISTEMPv4 White Area = 57.7 %



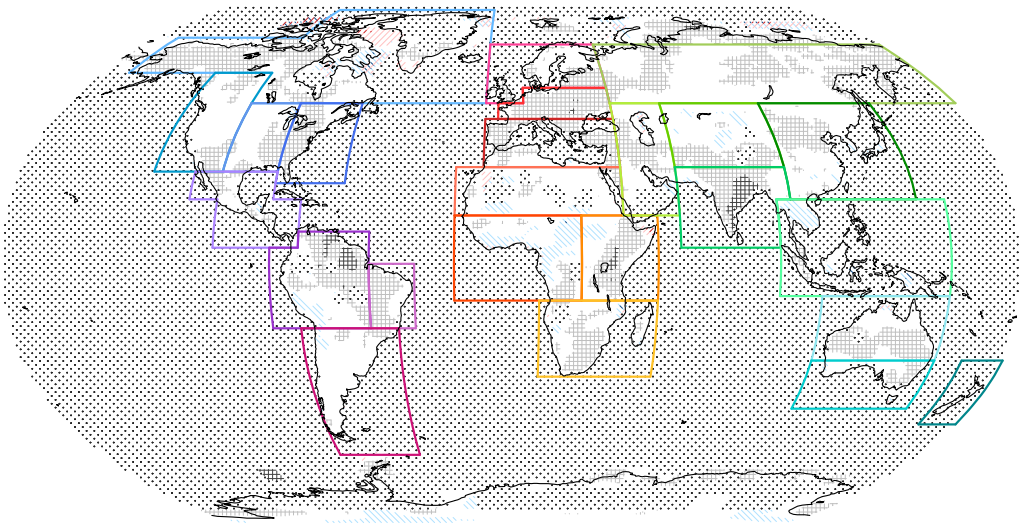
TAS DJF Obs inside central 75th percentile (%)



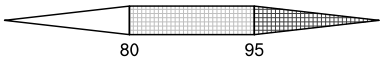
TAS DJF Obs outside ensemble spread (%)



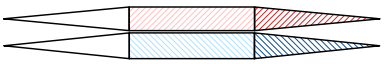
CanESM2 vs GISTEMPv4 White Area = 50.9 %



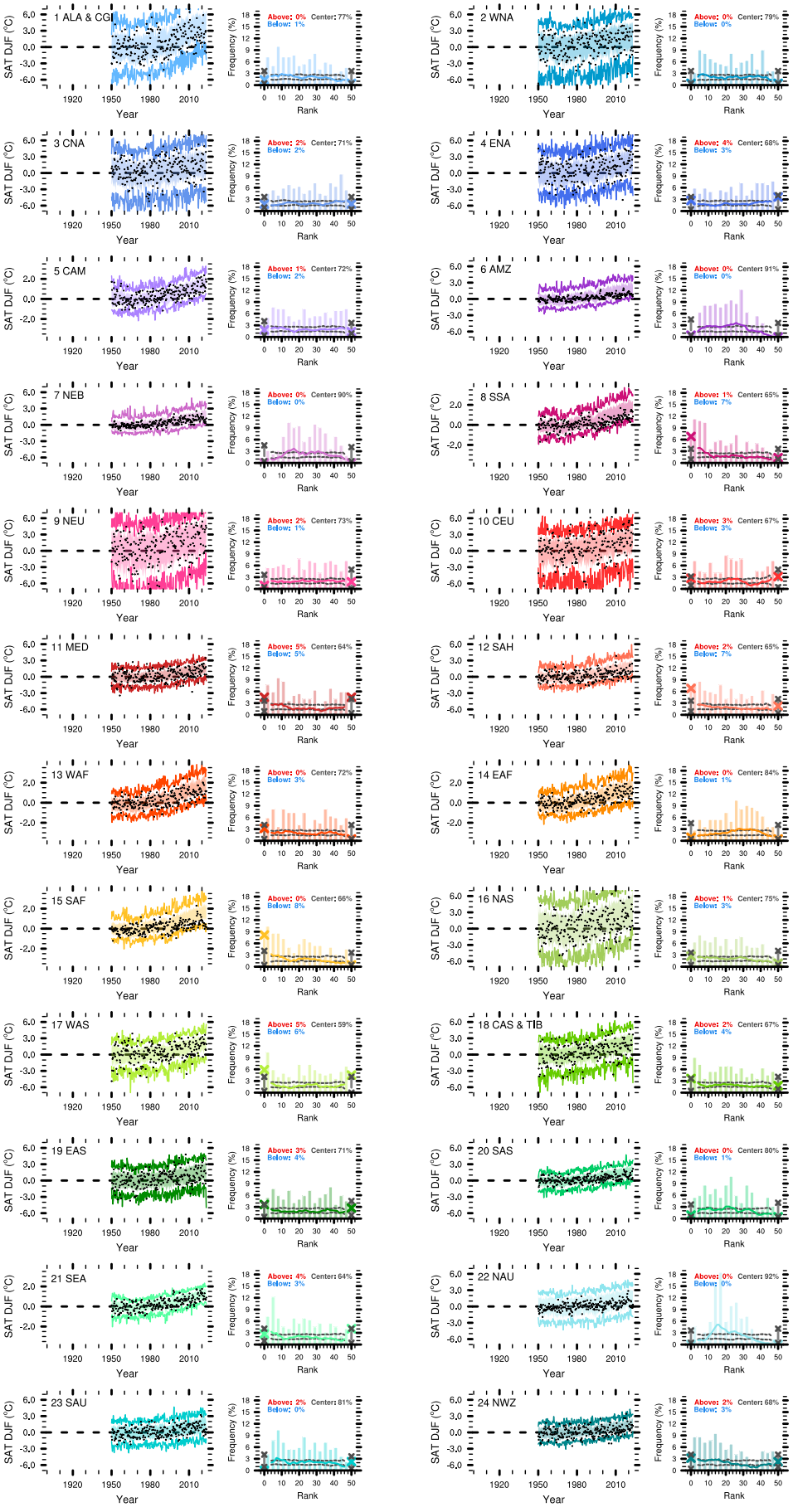
TAS JJA Obs inside central 75th percentile (%)



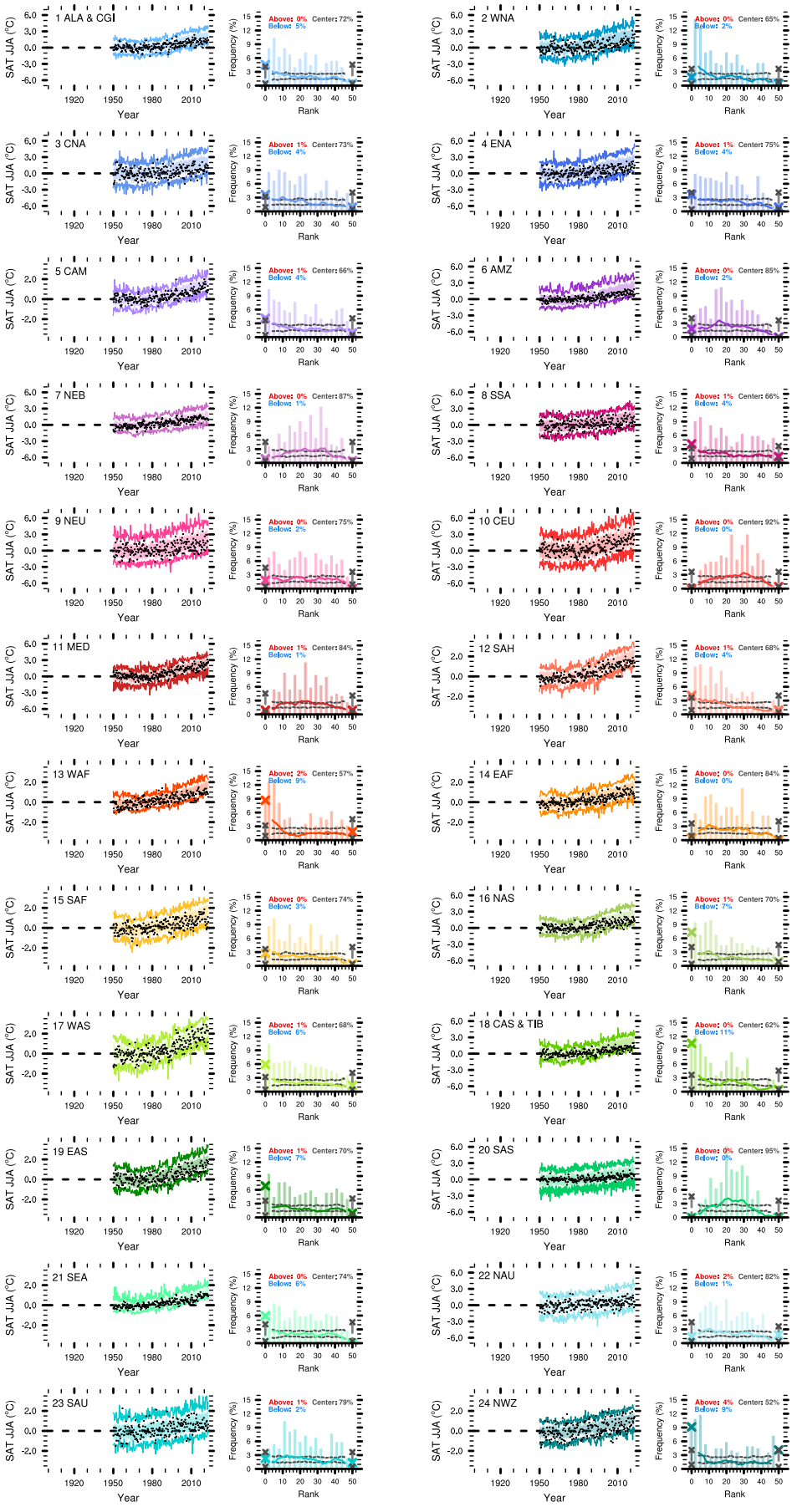
TAS JJA Obs outside ensemble spread (%)



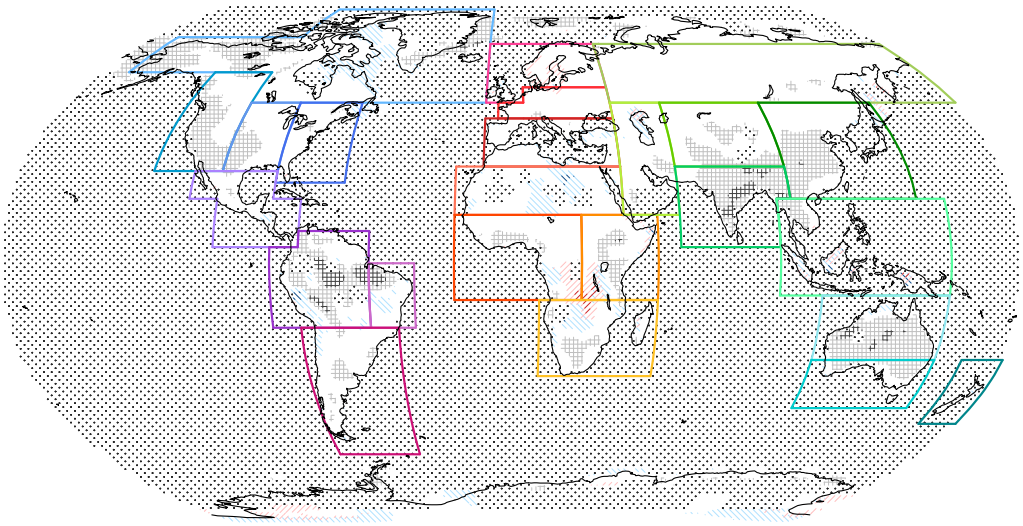
CanESM2 vs GISTEMPv4 TAS DJF



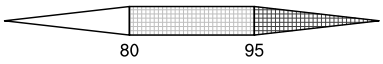
CanESM2 vs GISTEMPv4 TAS JJA



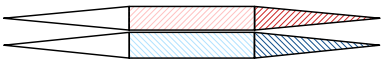
CanESM5 vs GISTEMPv4 White Area = 60.2 %



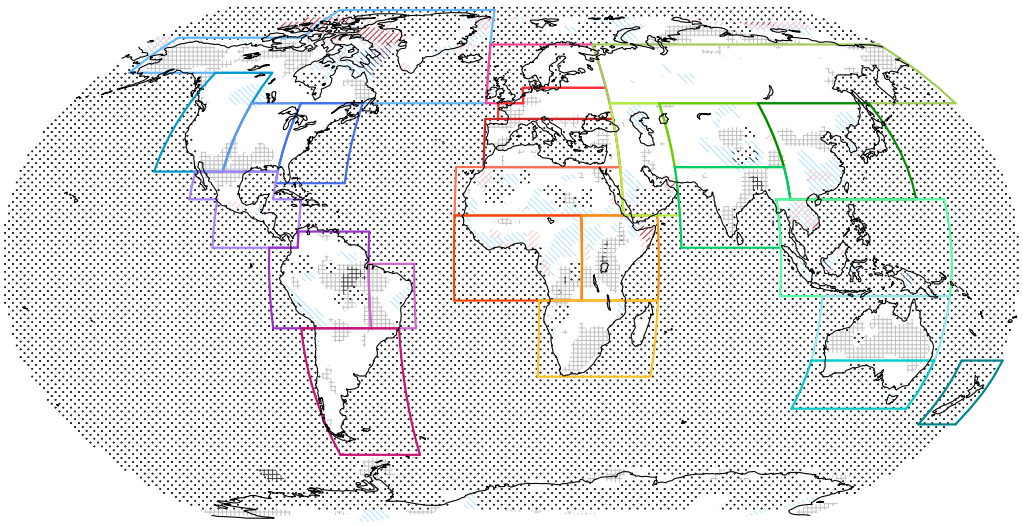
TAS DJF Obs inside central 75th percentile (%)



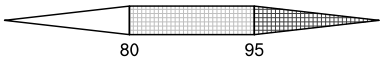
TAS DJF Obs outside ensemble spread (%)



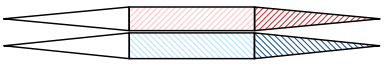
CanESM5 vs GISTEMPv4 White Area = 61.0 %



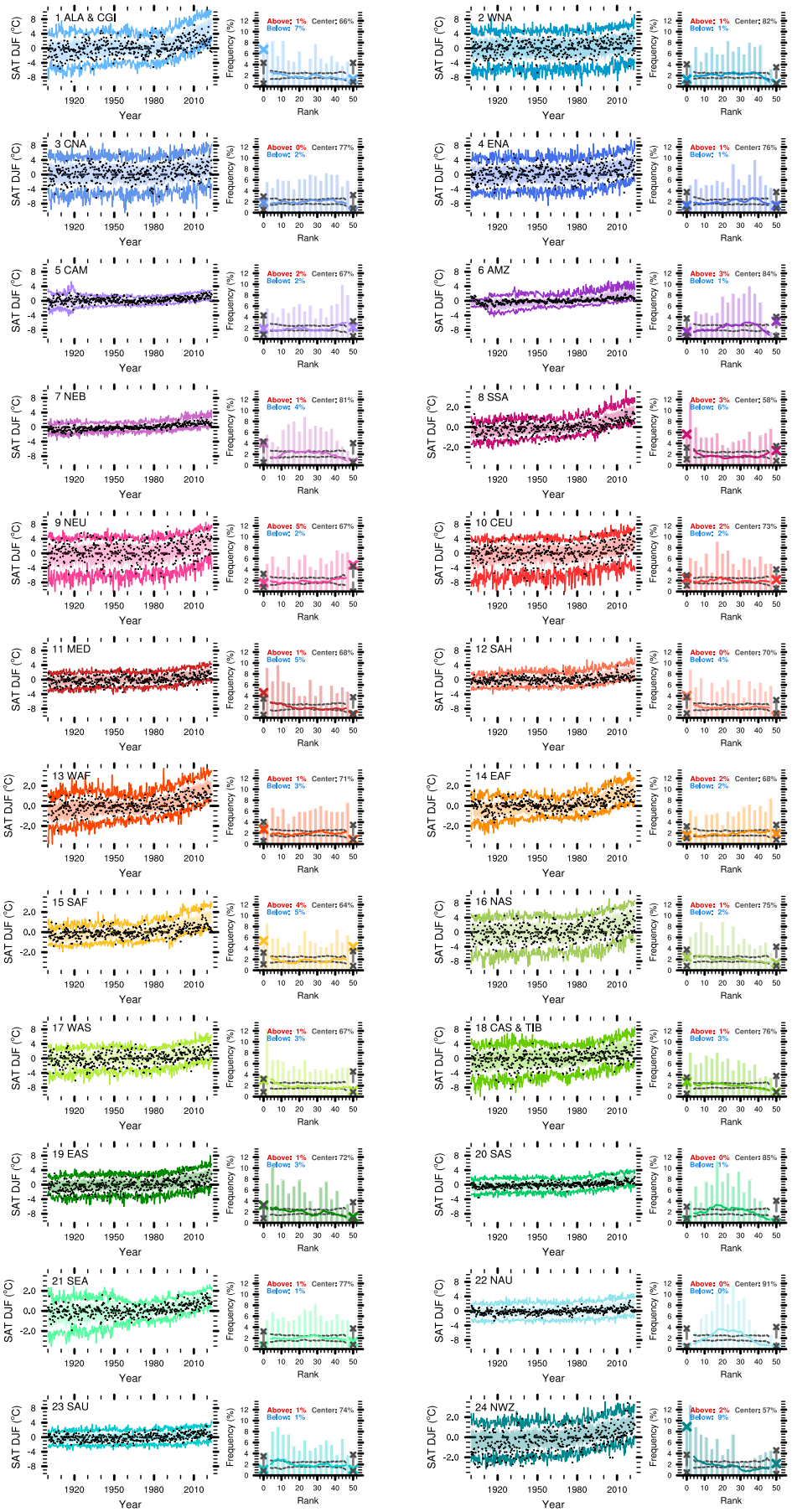
TAS JJA Obs inside central 75th percentile (%)



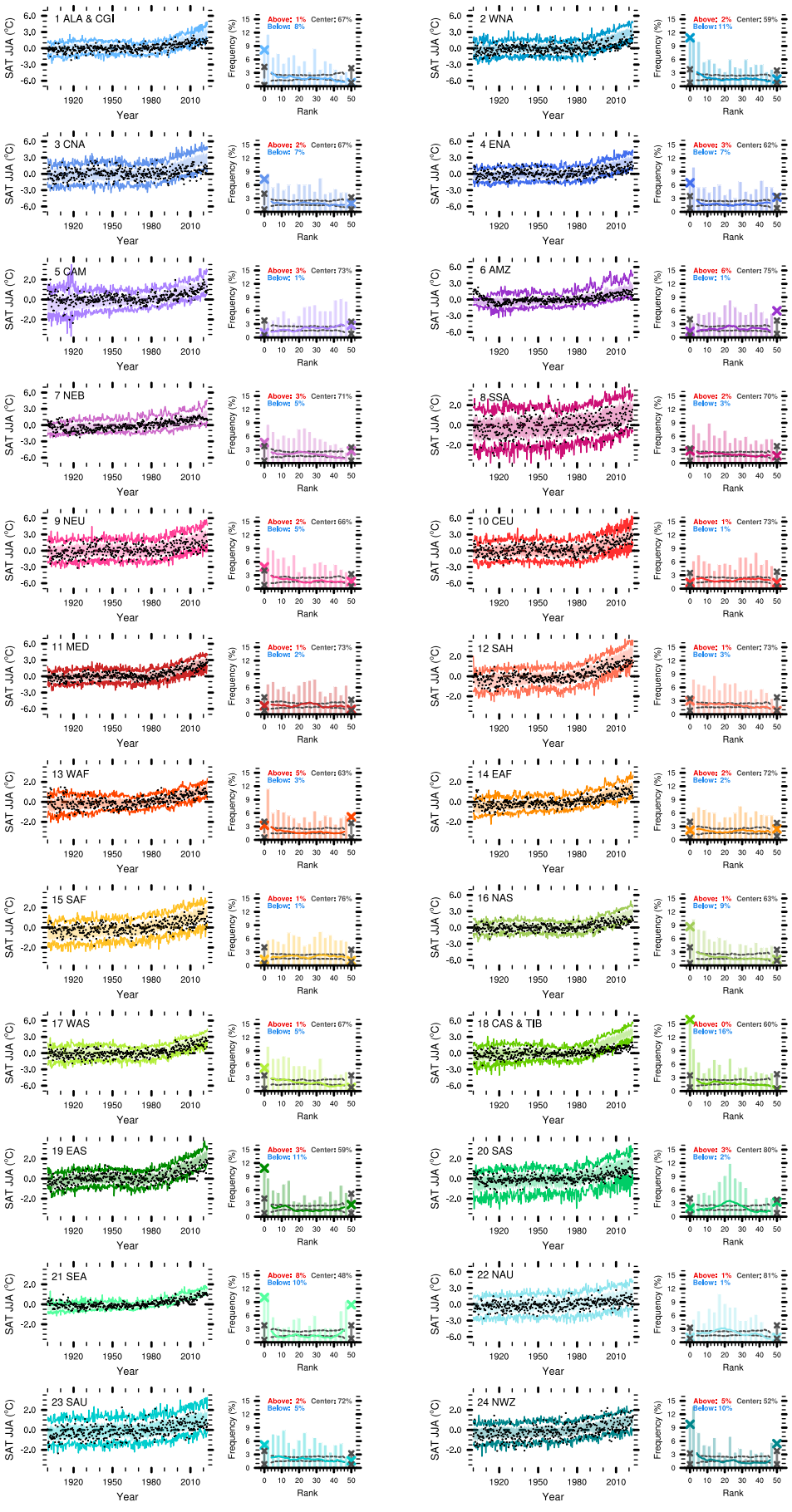
TAS JJA Obs outside ensemble spread (%)



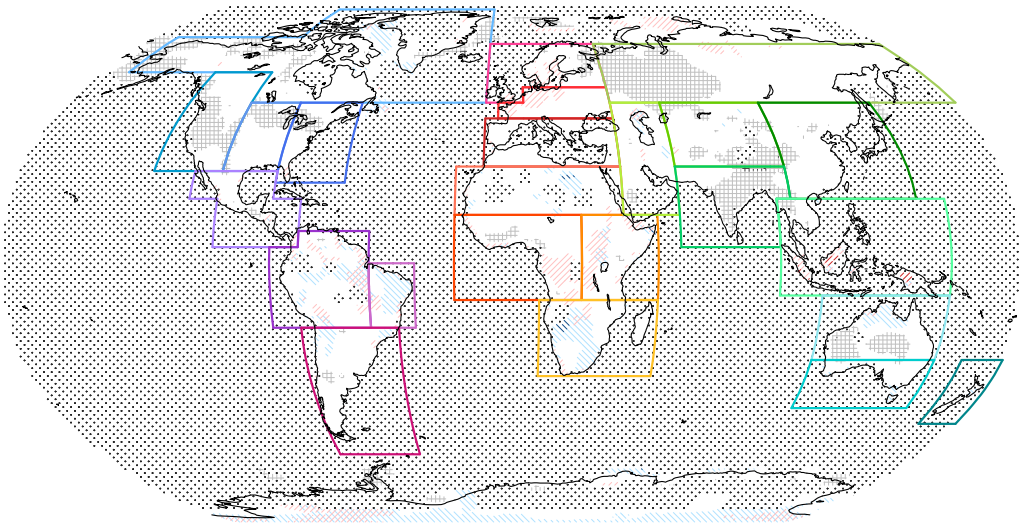
CanESM5 vs GISTEMPv4 TAS DJF



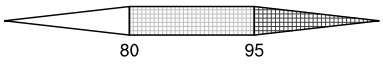
CanESM5 vs GISTEMPv4 TAS JJA



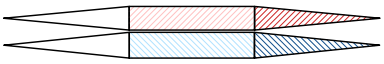
CESM-LE vs GISTEMPv4 White Area = 64.1 %



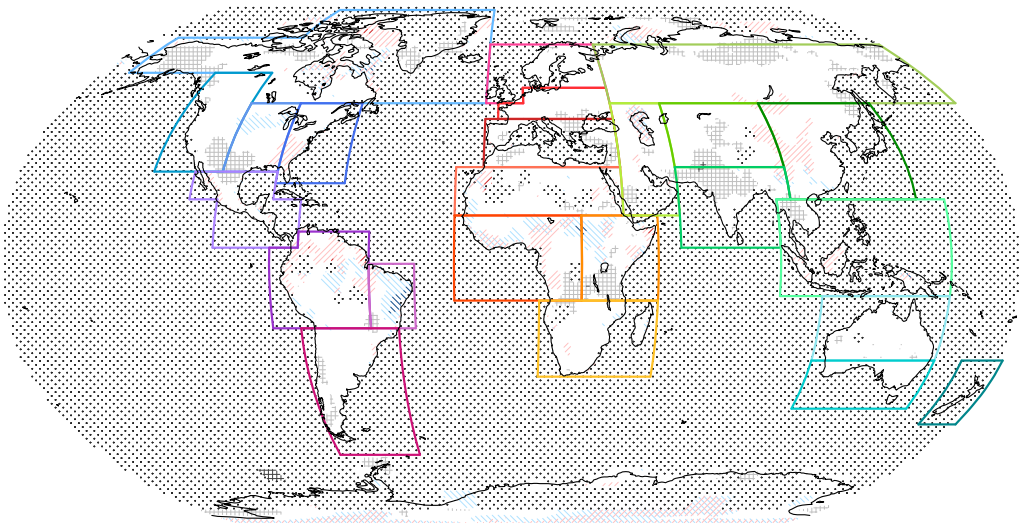
TAS DJF Obs inside central 75th percentile (%)



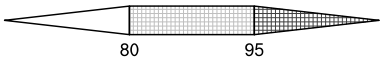
TAS DJF Obs outside ensemble spread (%)



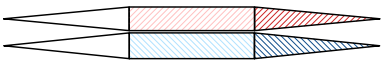
CESM-LE vs GISTEMPv4 White Area = 65.9 %



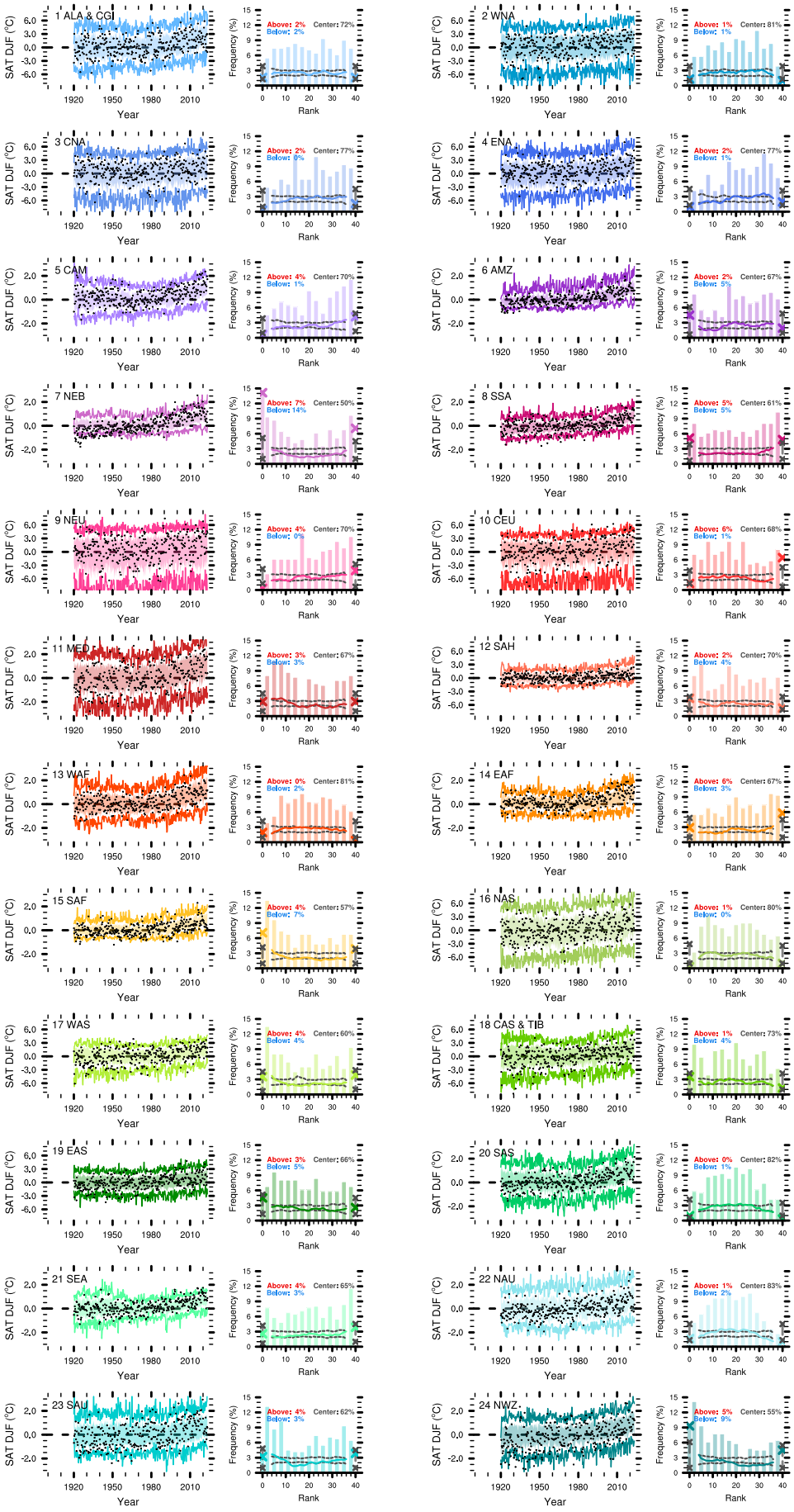
TAS JJA Obs inside central 75th percentile (%)



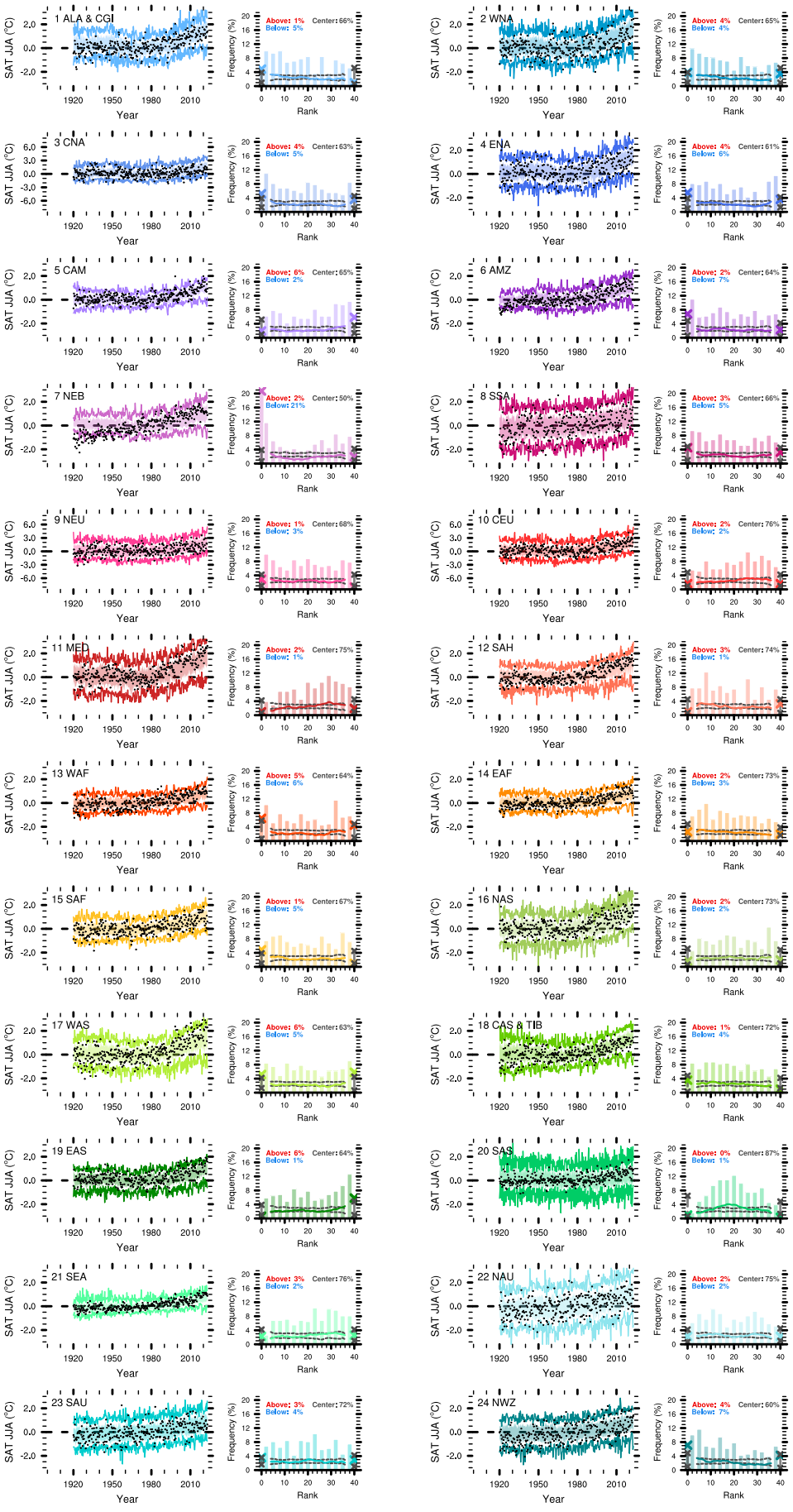
TAS JJA Obs outside ensemble spread (%)



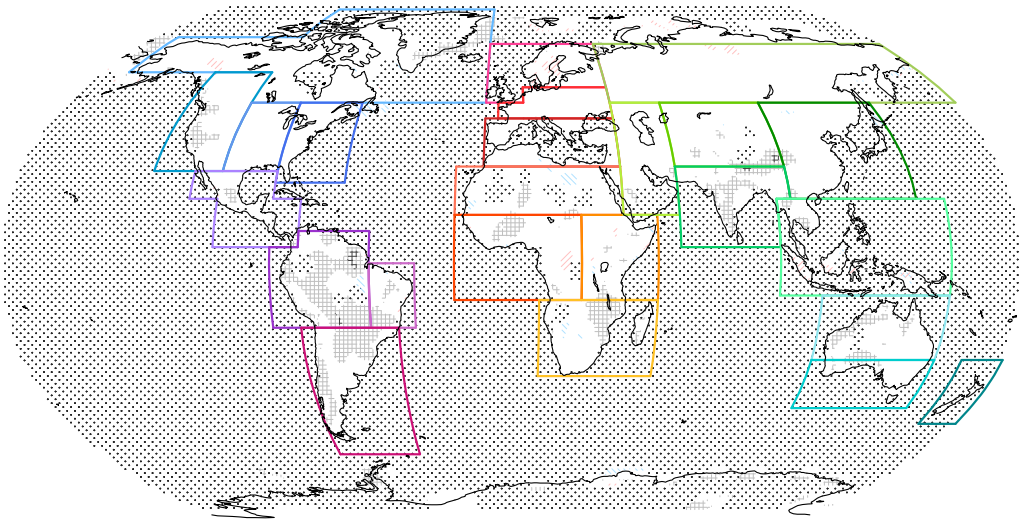
CESM-LE vs GISTEMPv4 TAS DJF



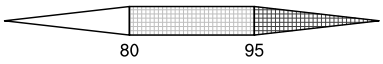
CESM-LE vs GISTEMPv4 TAS JJA



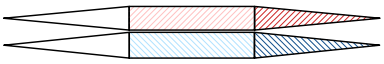
CESM2-LE vs GISTEMPv4 White Area = 81.4 %



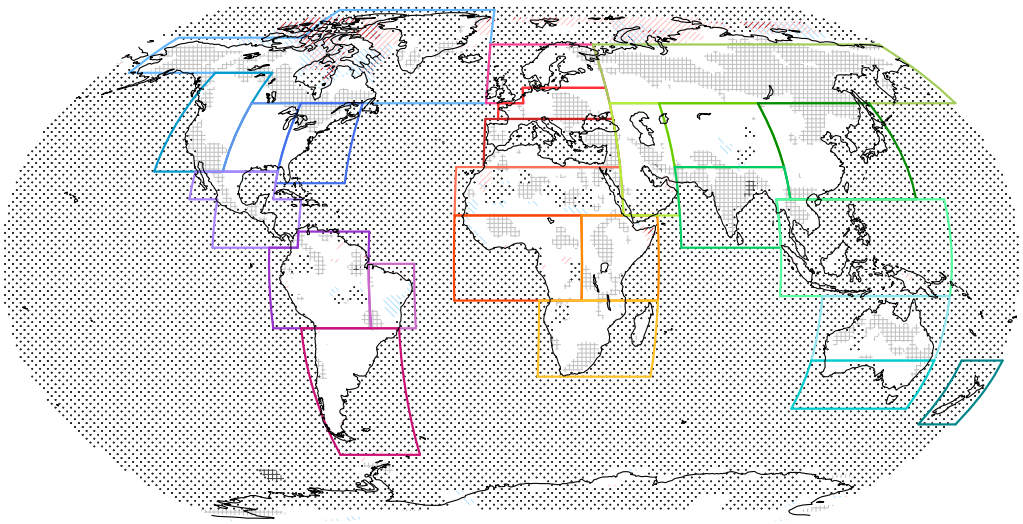
TAS DJF Obs inside central 75th percentile (%)



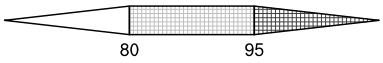
TAS DJF Obs outside ensemble spread (%)



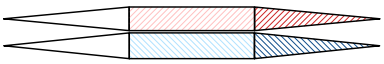
CESM2-LE vs GISTEMPv4 White Area = 63.5 %



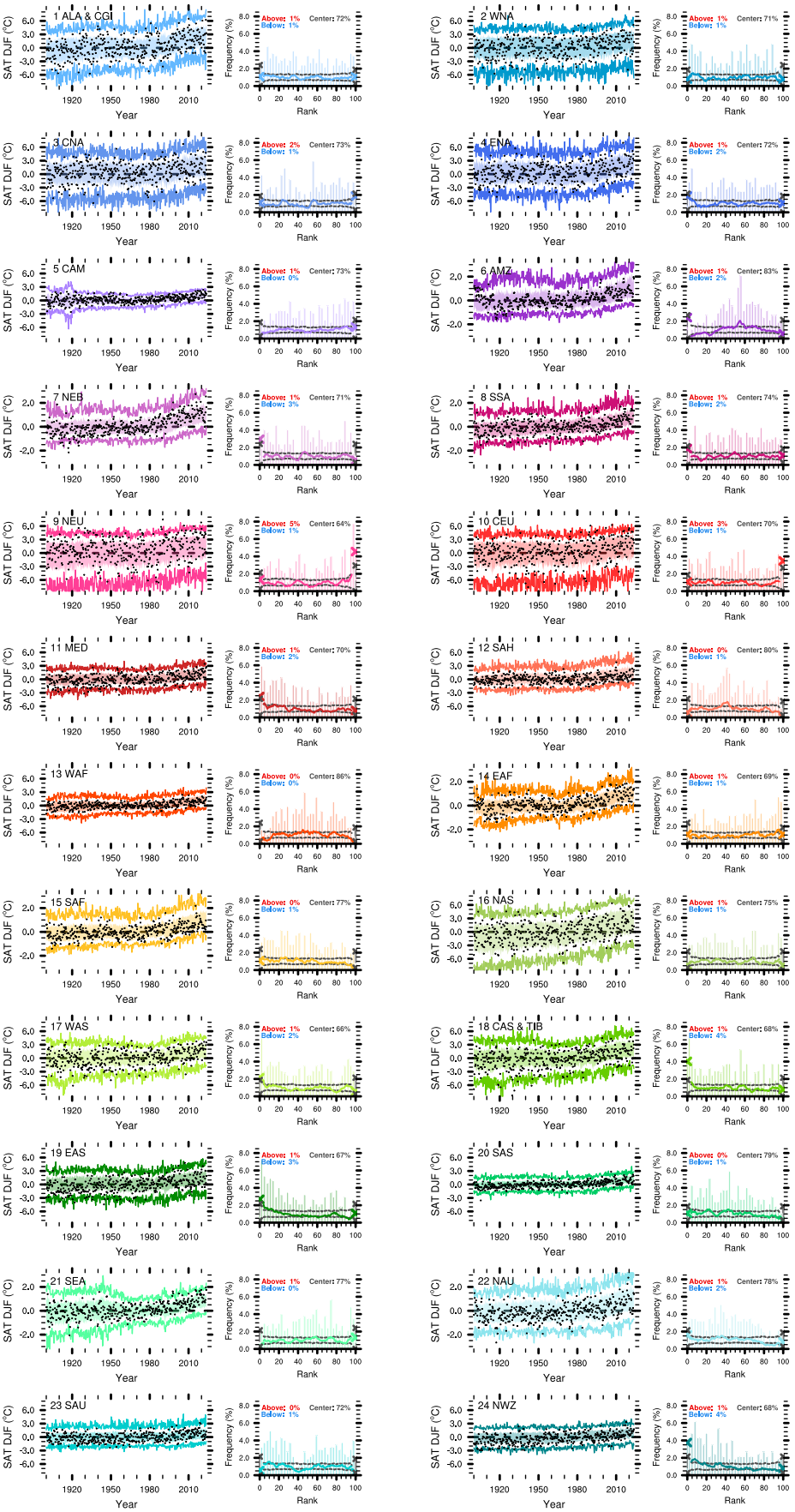
TAS JJA Obs inside central 75th percentile (%)



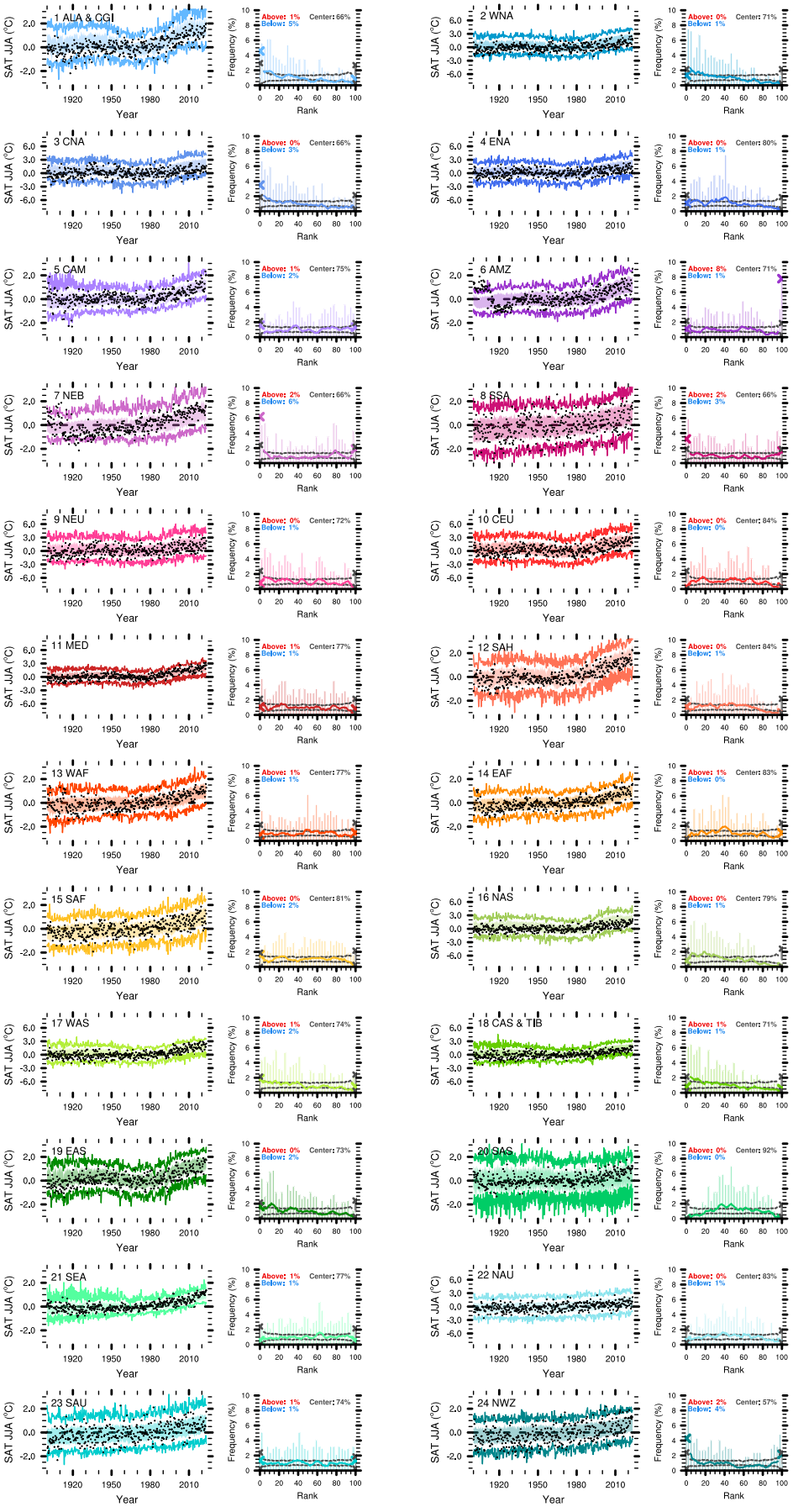
TAS JJA Obs outside ensemble spread (%)



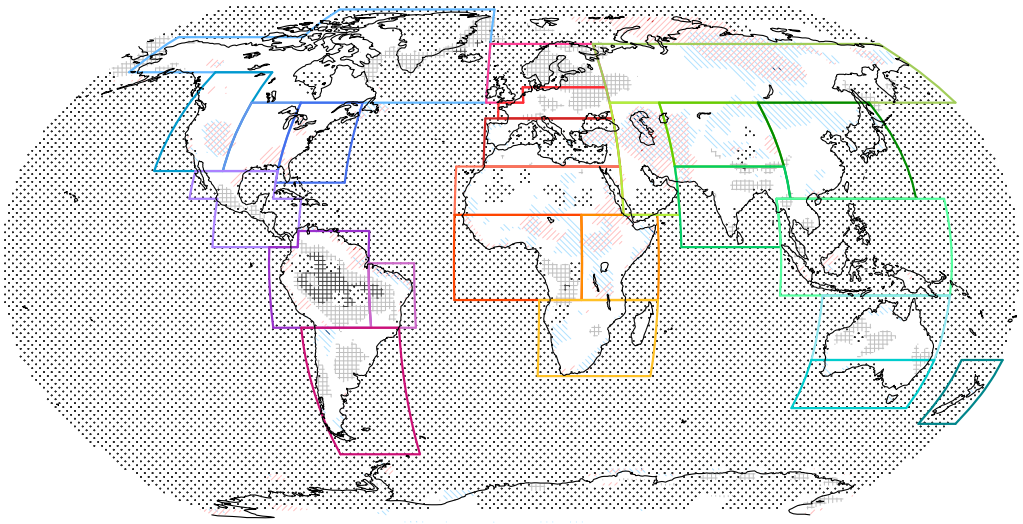
CESM2-LE vs GISTEMPv4 TAS DJF



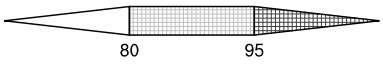
CESM2-LE vs GISTEMPv4 TAS JJA



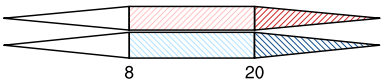
CSIRO-Mk360 vs GISTEMPv4 White Area = 57.4 %



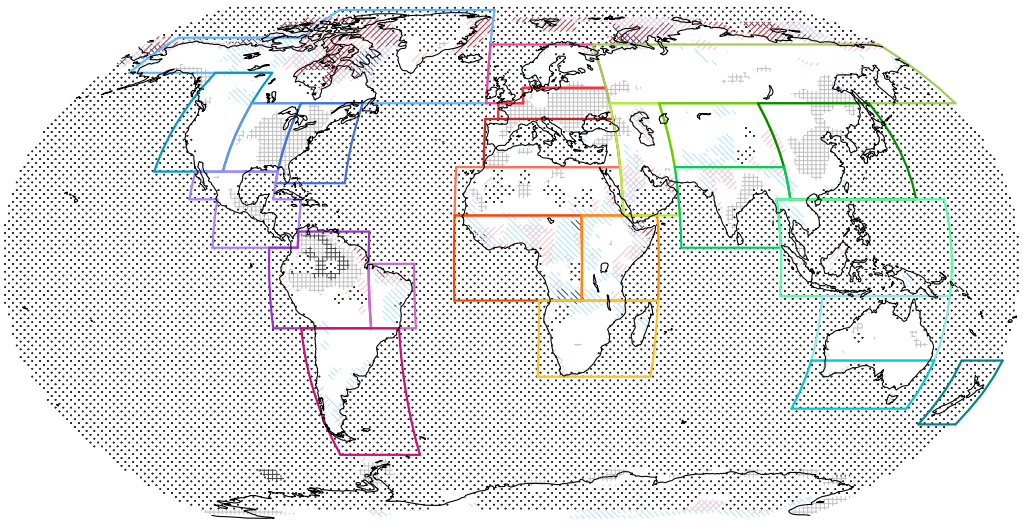
TAS DJF Obs inside central 75th percentile (%)



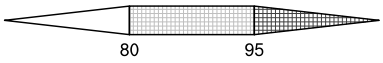
TAS DJF Obs outside ensemble spread (%)



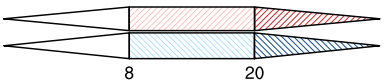
CSIRO-Mk360 vs GISTEMPv4 White Area = 54.9 %



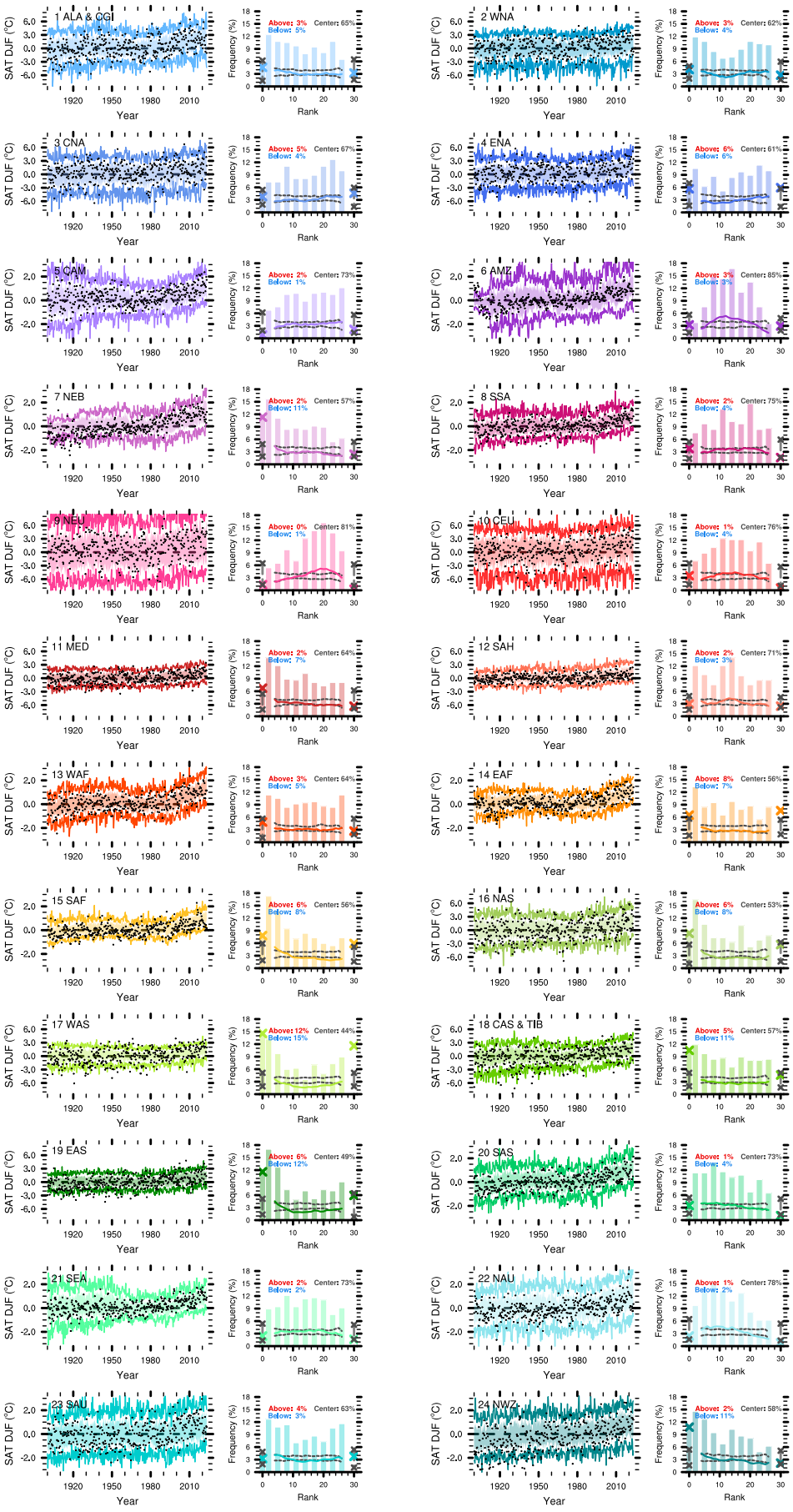
TAS JJA Obs inside central 75th percentile (%)



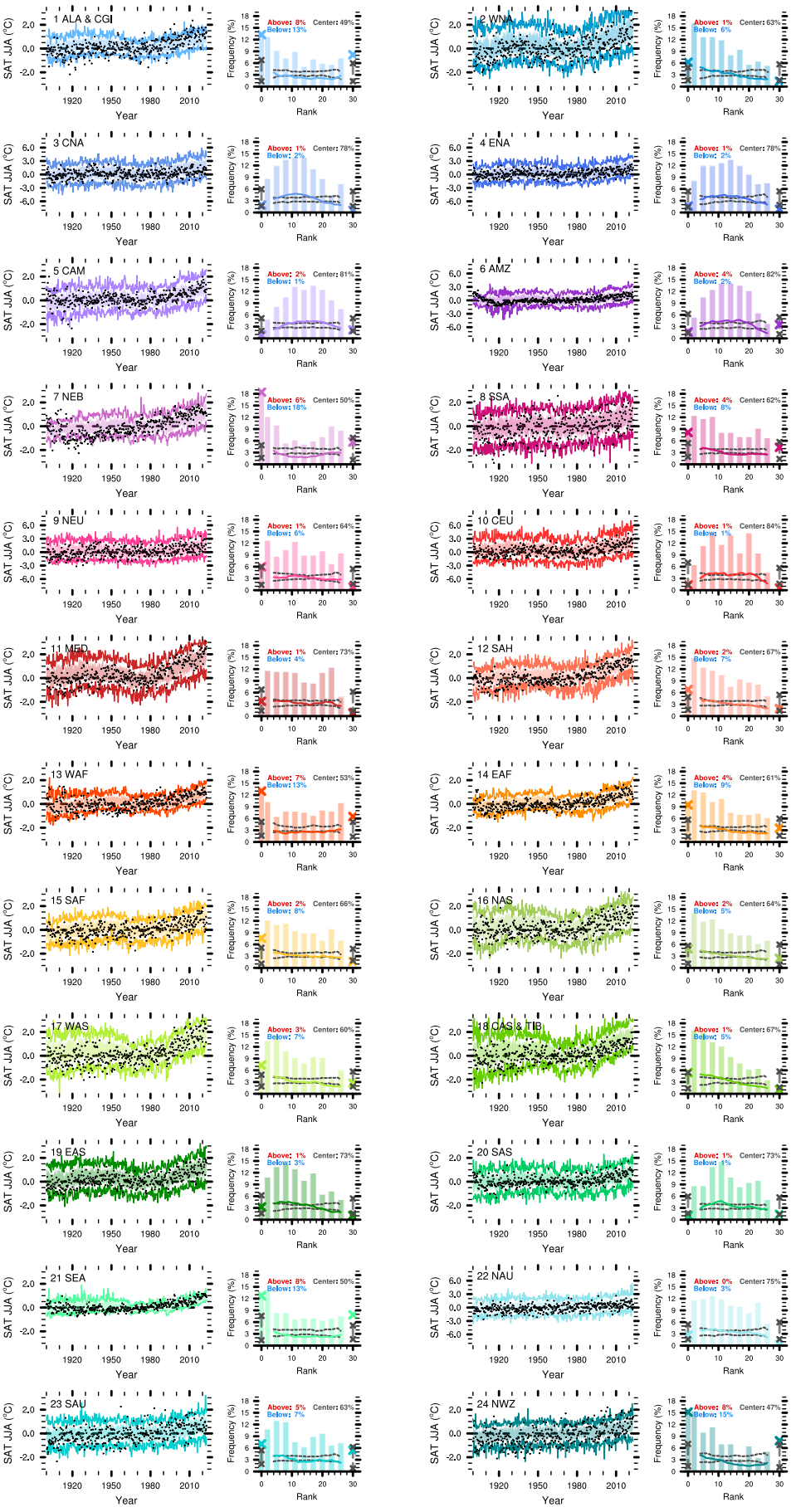
TAS JJA Obs outside ensemble spread (%)



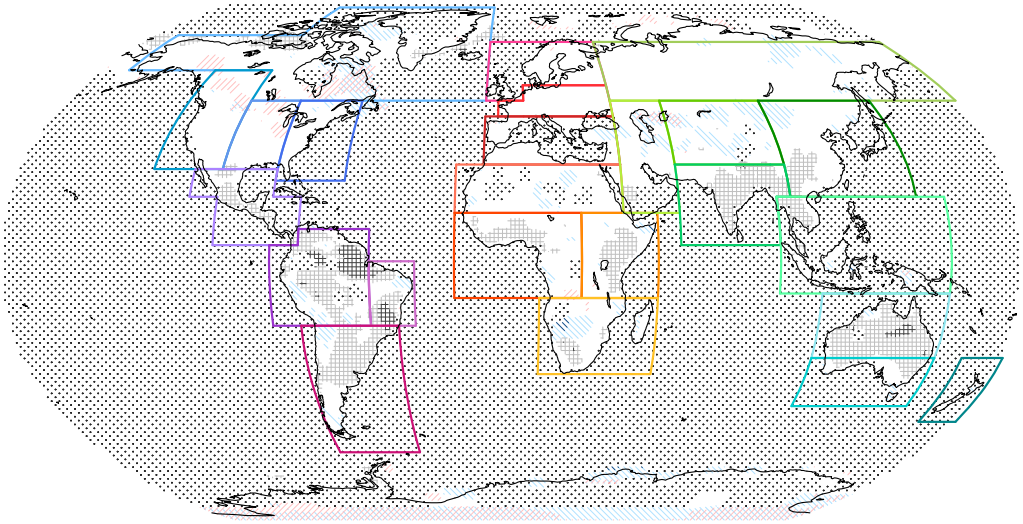
CSIRO-Mk360 vs GISTEMPv4 TAS DJF



CSIRO-Mk360 vs GISTEMPv4 TAS JJA



GFDL-ESM2M vs GISTEMPv4 White Area = 62.6 %



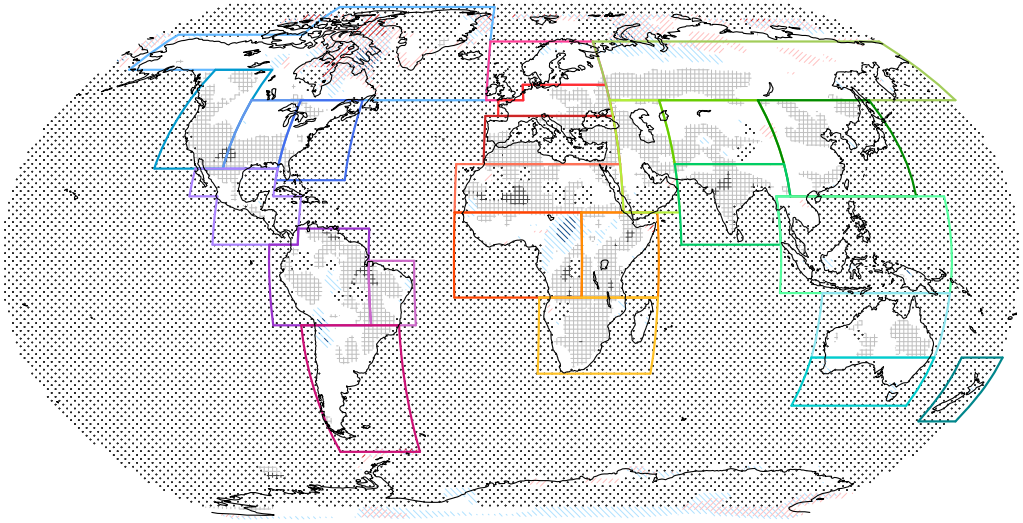
TAS DJF Obs inside central 75th percentile (%)

80 95

TAS DJF Obs outside ensemble spread (%)

8 20

GFDL-ESM2M vs GISTEMPv4 White Area = 50.5 %



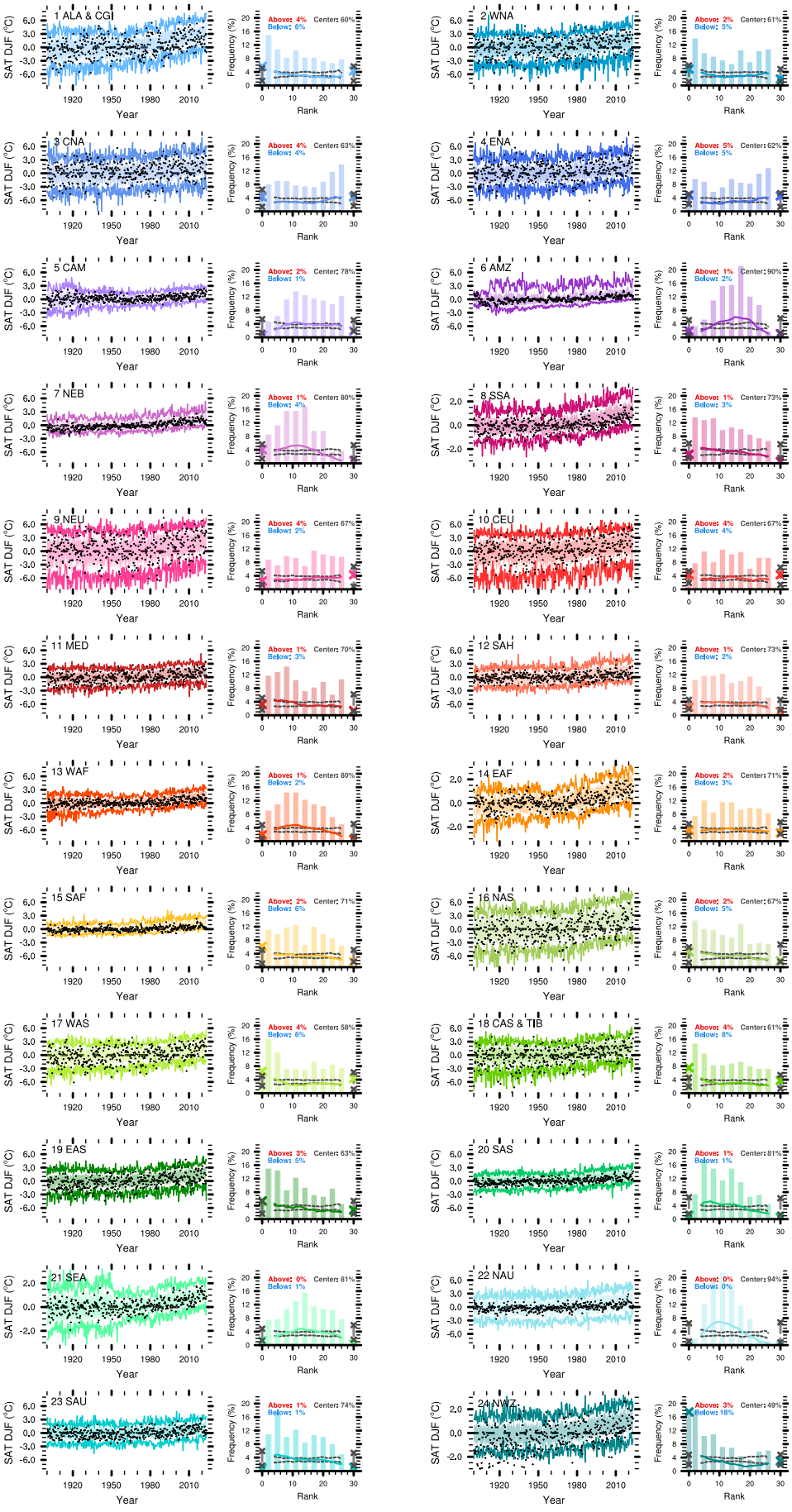
TAS JJA Obs inside central 75th percentile (%)

80 95

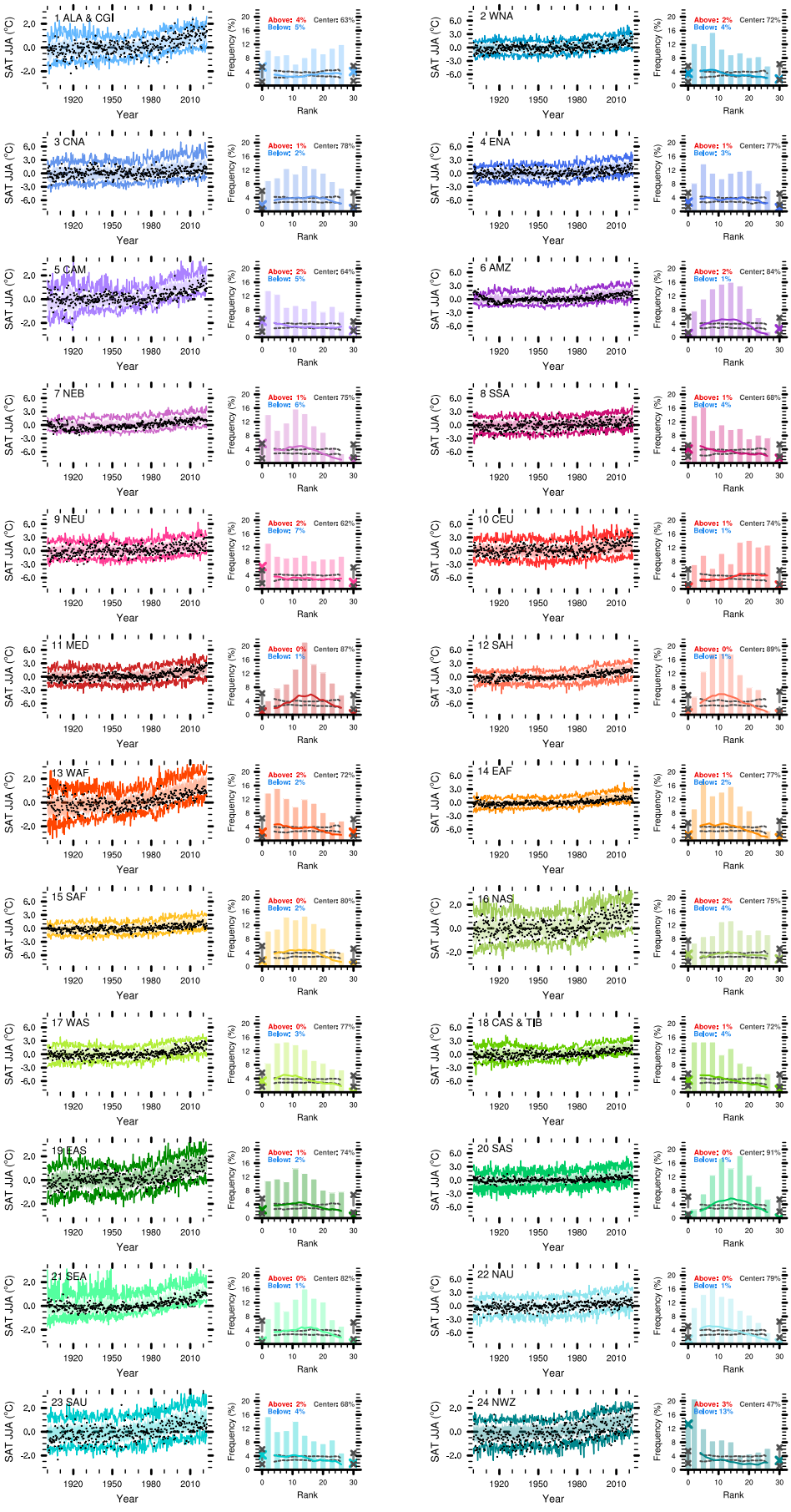
TAS JJA Obs outside ensemble spread (%)

8 20

GFDL-ESM2M vs GISTEMPv4 TAS DJF

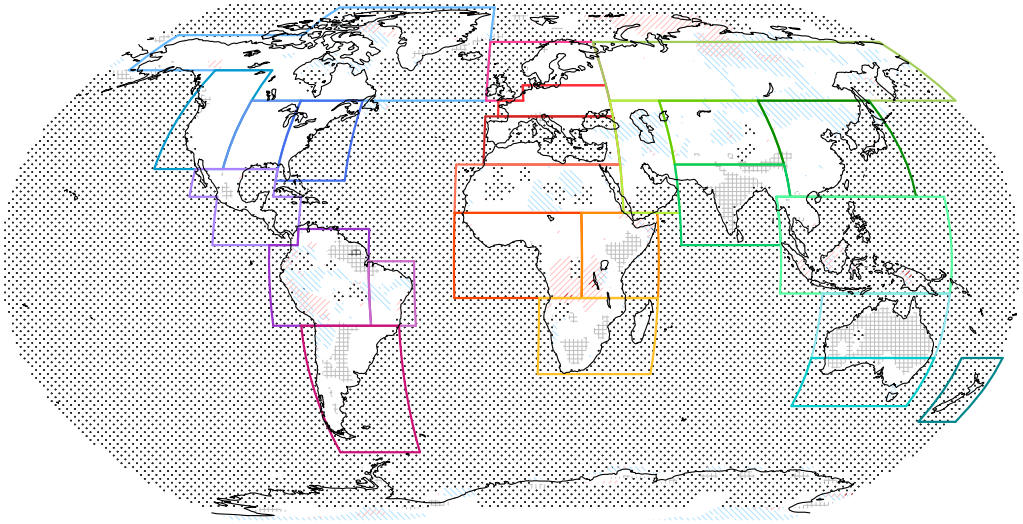


GFDL-ESM2M vs GISTEMPv4 TAS JJA



GFDL-SPEAR vs GISTEMPv4

White Area = 67.9 %



TAS DJF Obs inside central 75th percentile (%)

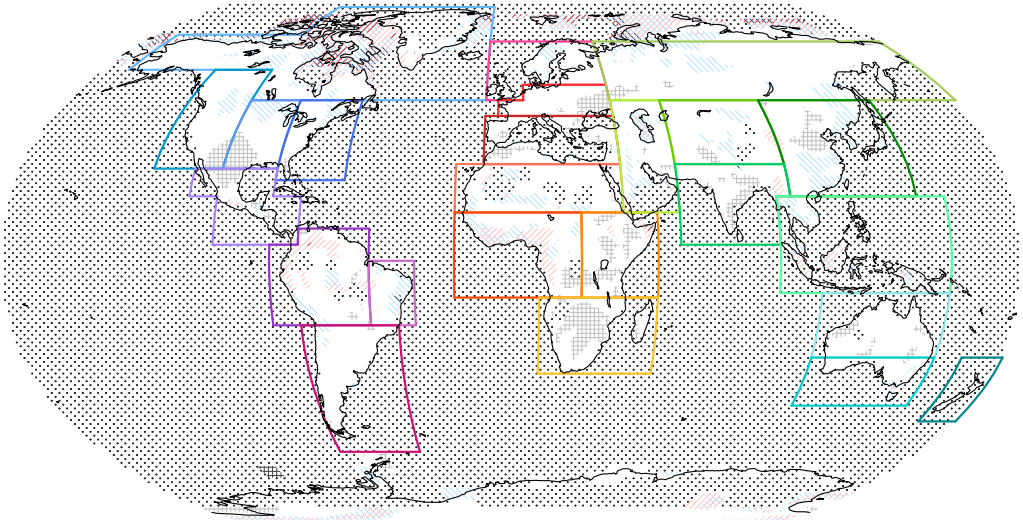
80 95

TAS DJF Obs outside ensemble spread (%)

8 20

GFDL-SPEAR vs GISTEMPv4

White Area = 64.0 %



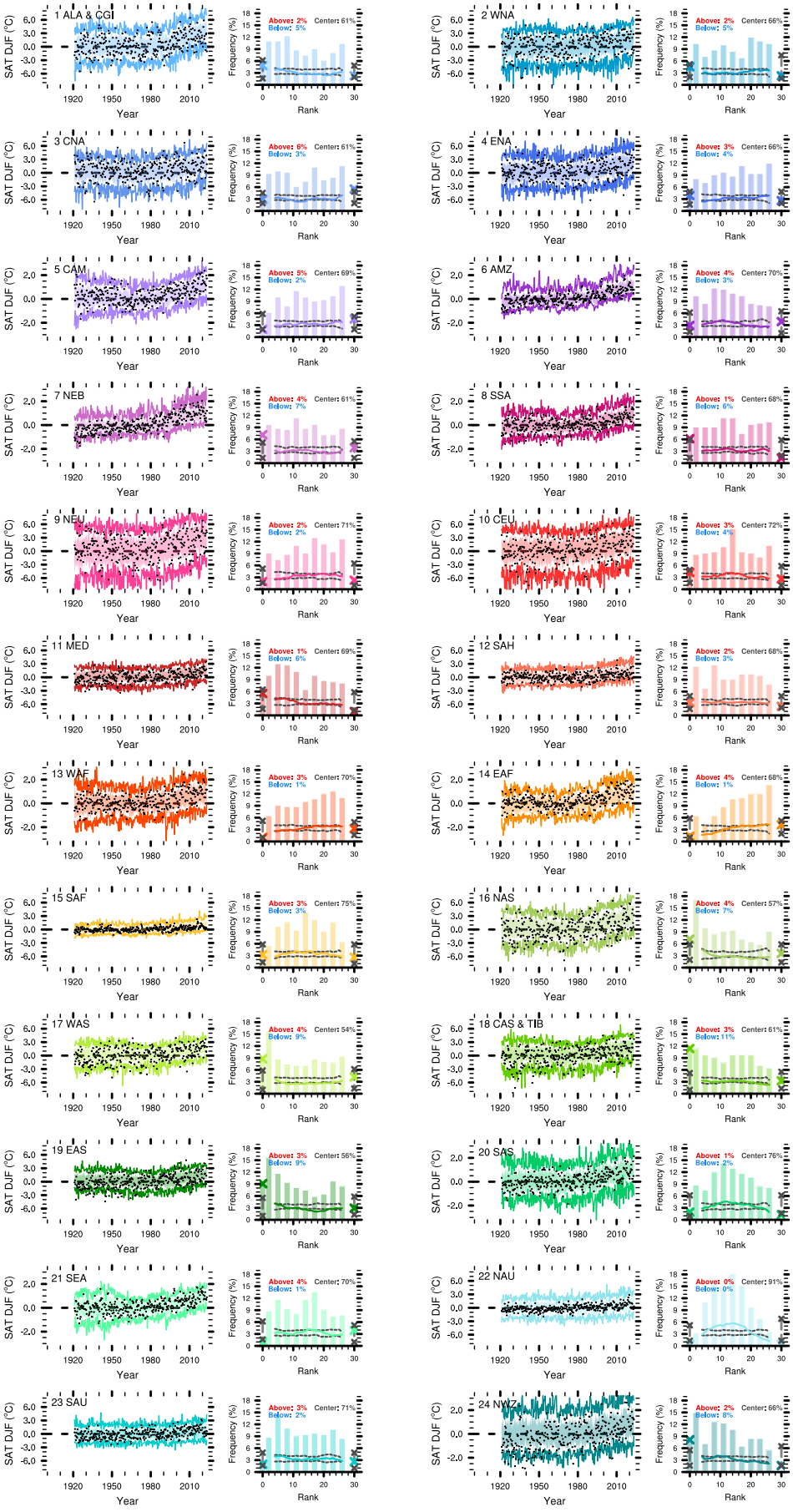
TAS JJA Obs inside central 75th percentile (%)

80 95

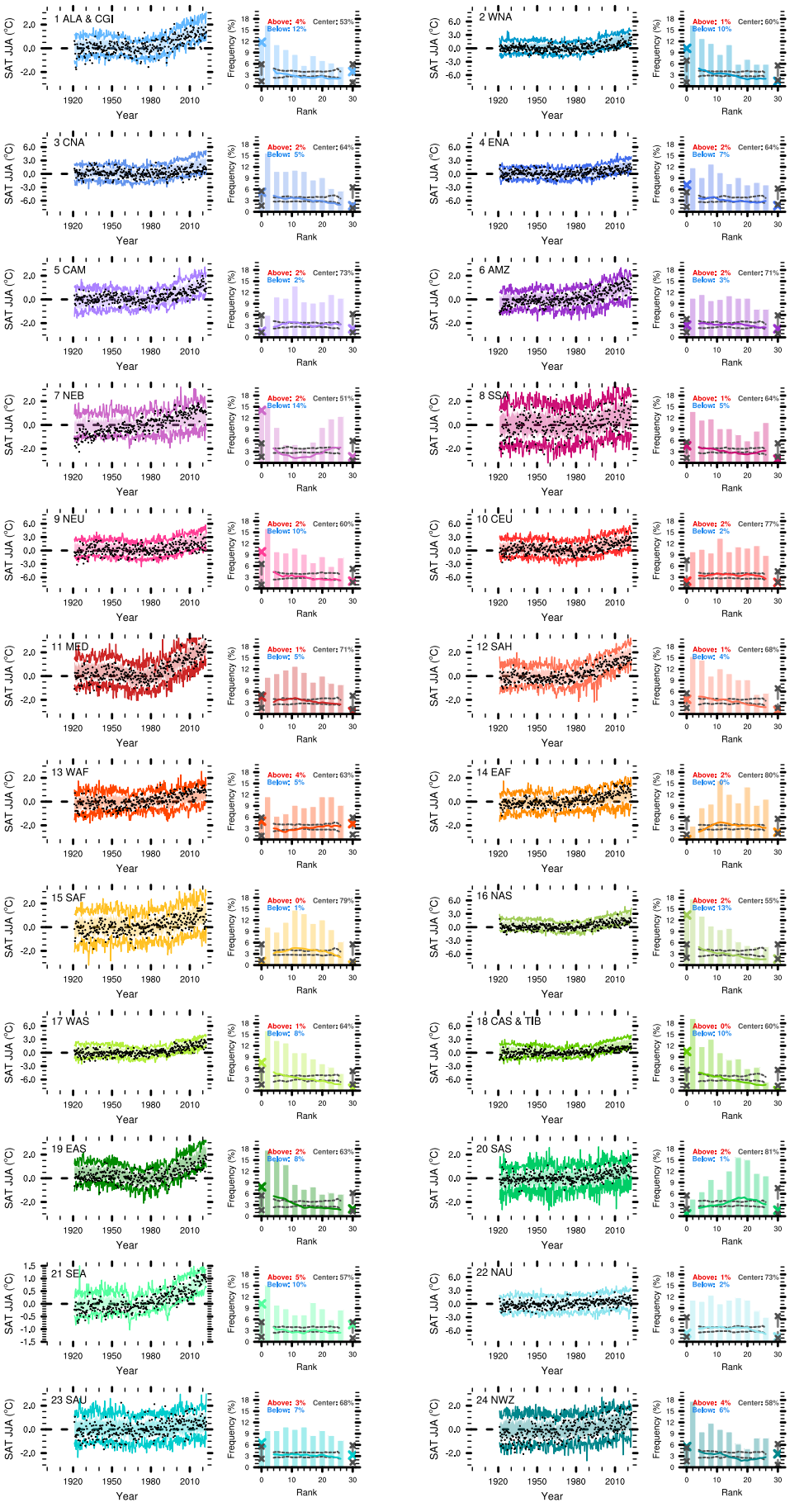
TAS JJA Obs outside ensemble spread (%)

8 20

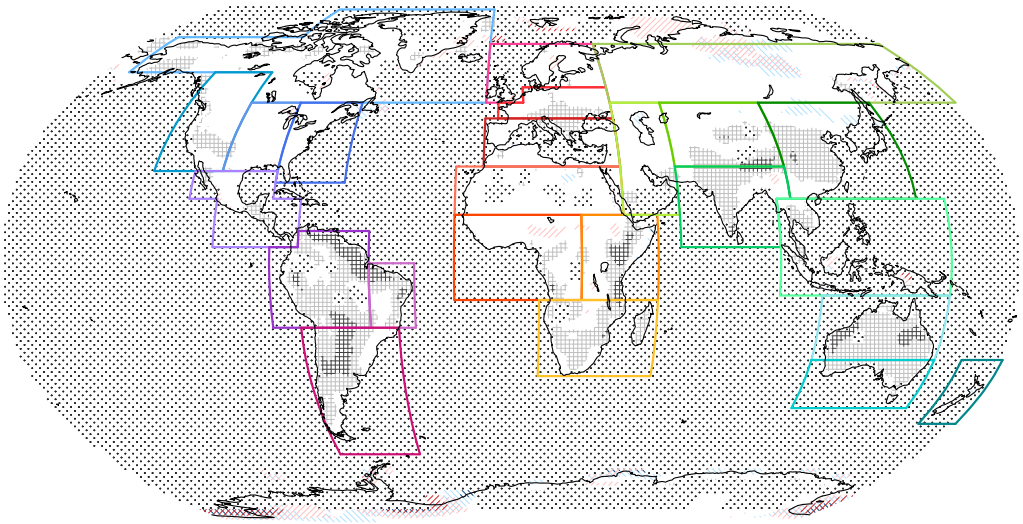
GFDL-SPEAR vs GISTEMPv4 TAS DJF



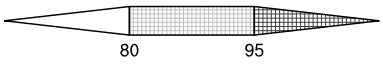
GFDL-SPEAR vs GISTEMPv4 TAS JJA



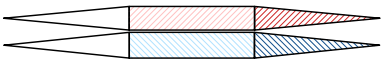
MIROC6 vs GISTEMPv4 White Area = 60.3 %



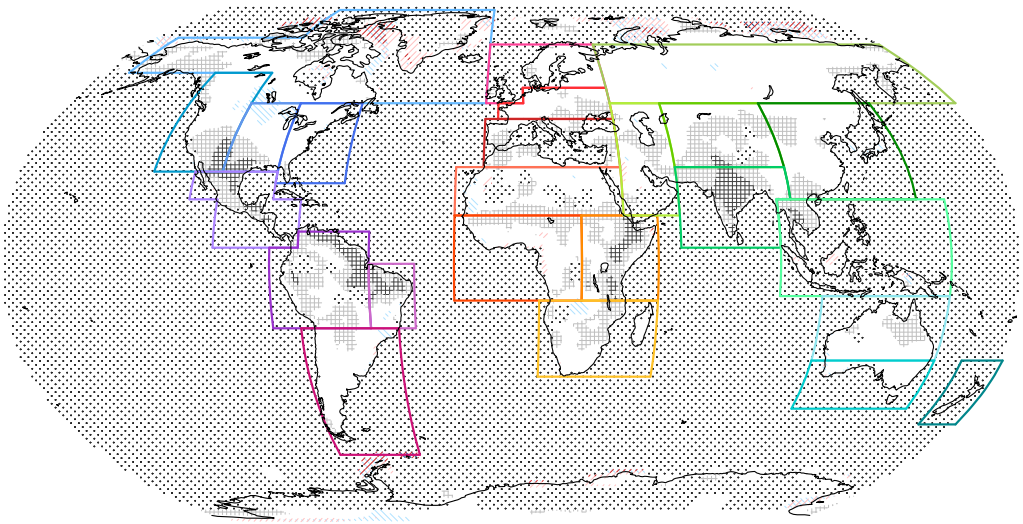
TAS DJF Obs inside central 75th percentile (%)



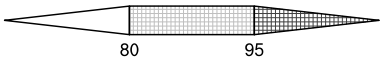
TAS DJF Obs outside ensemble spread (%)



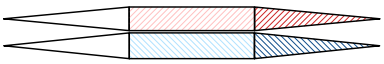
MIROC6 vs GISTEMPv4 White Area = 59.5 %



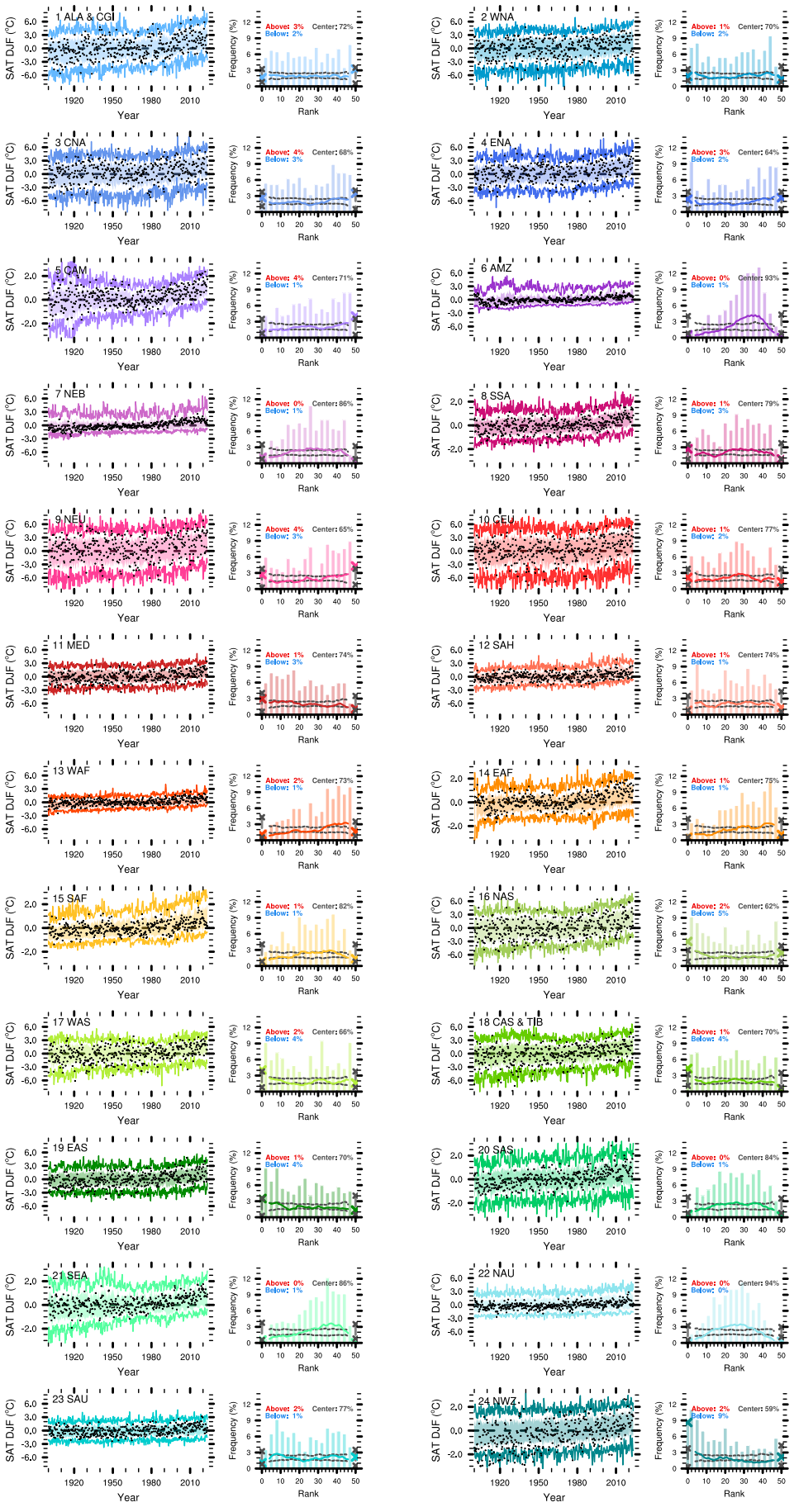
TAS JJA Obs inside central 75th percentile (%)



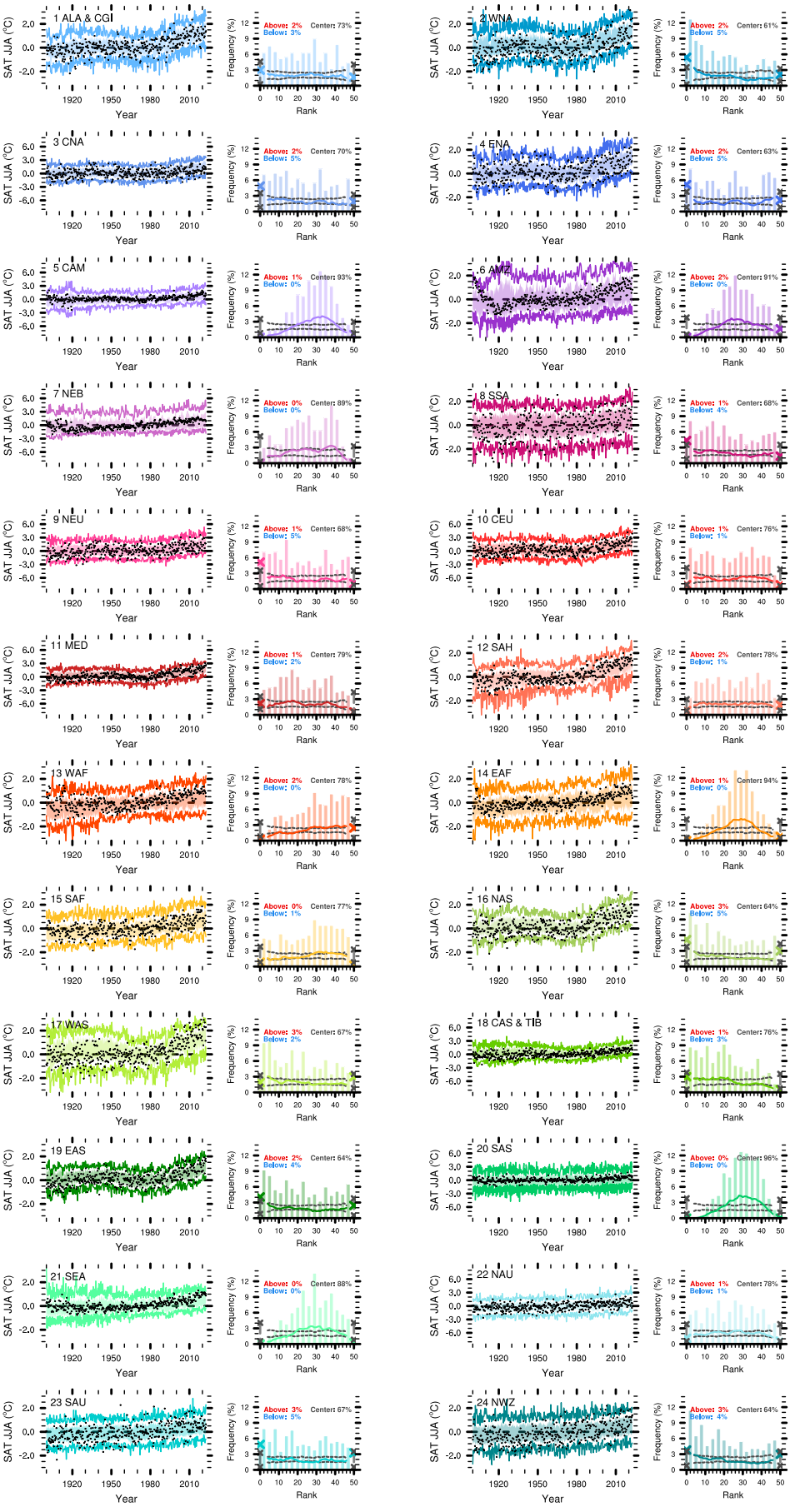
TAS JJA Obs outside ensemble spread (%)



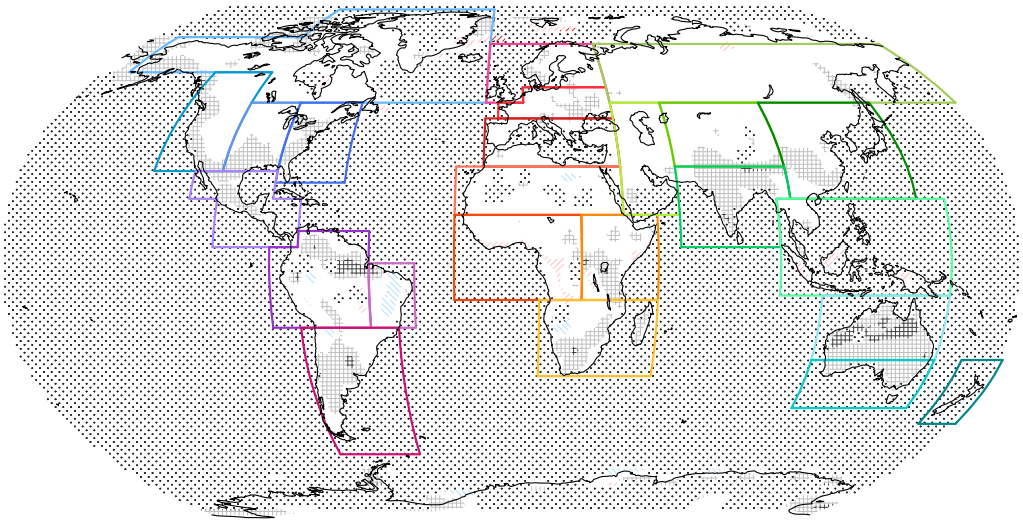
MIROC6 vs GISTEMPv4 TAS DJF



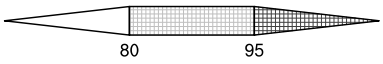
MIROC6 vs GISTEMPv4 TAS JJA



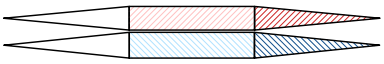
MPI-GE5 vs GISTEMPv4 White Area = 69.8 %



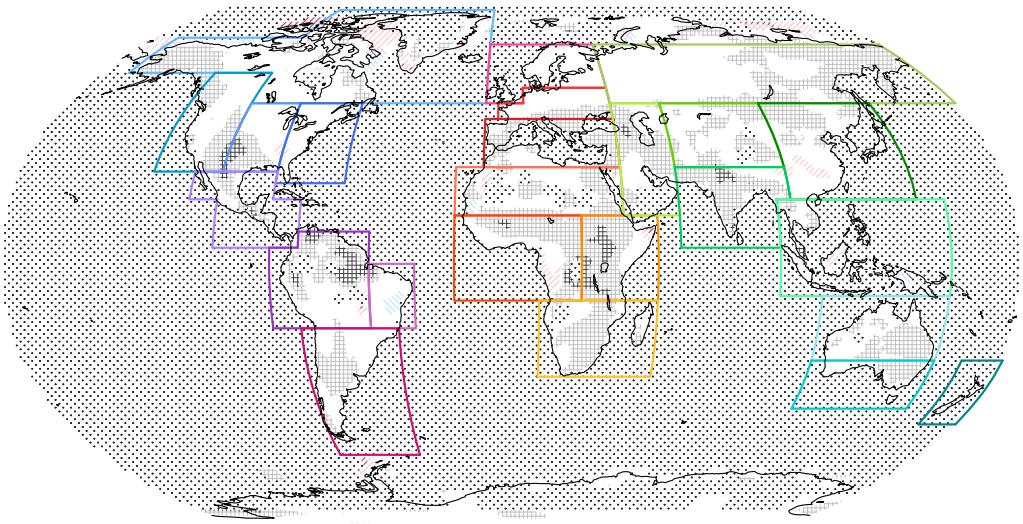
TAS DJF Obs inside central 75th percentile (%)



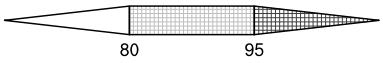
TAS DJF Obs outside ensemble spread (%)



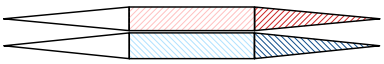
MPI-GE5 vs GISTEMPv4 White Area = 55.8 %



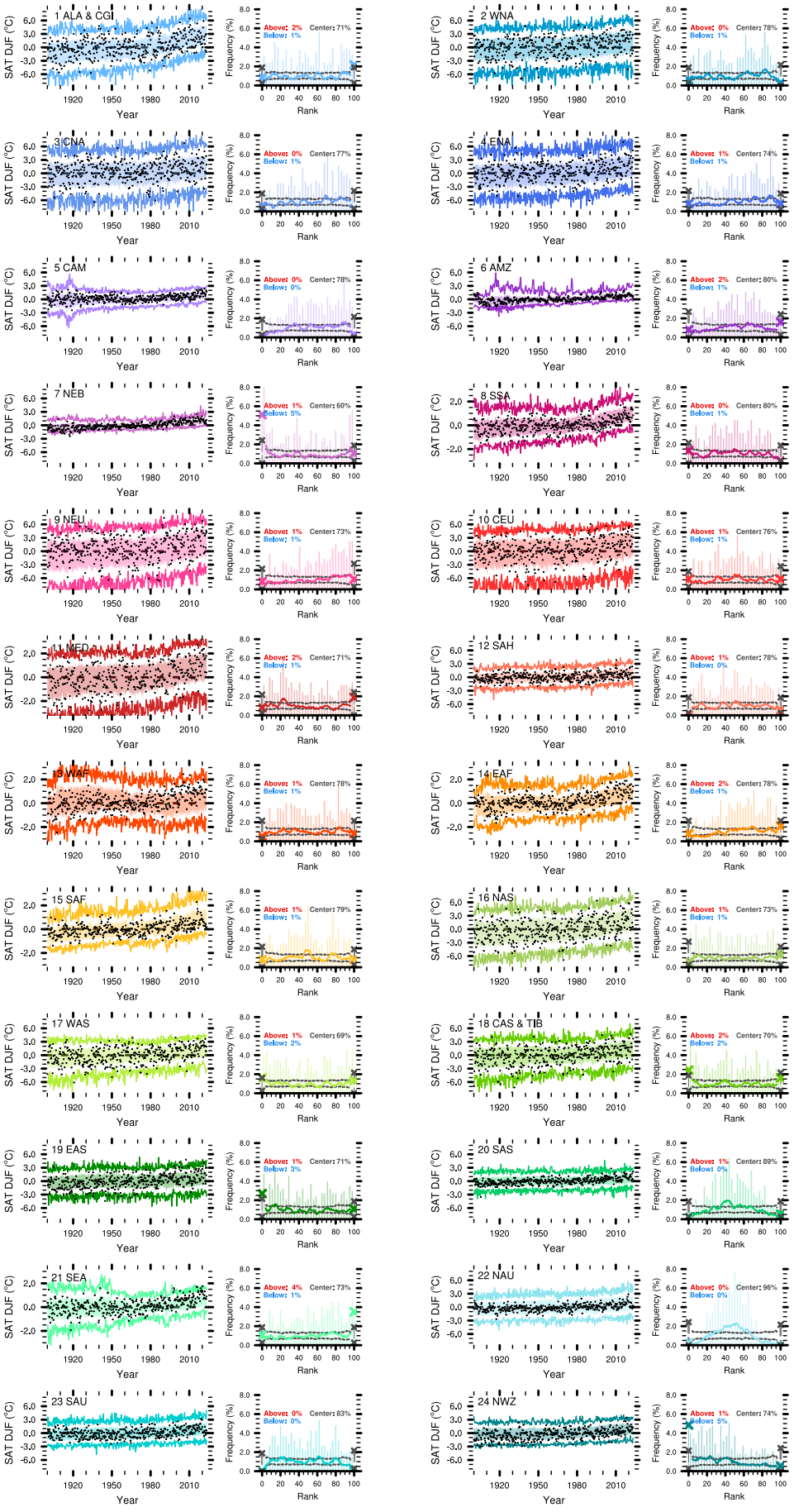
TAS JJA Obs inside central 75th percentile (%)



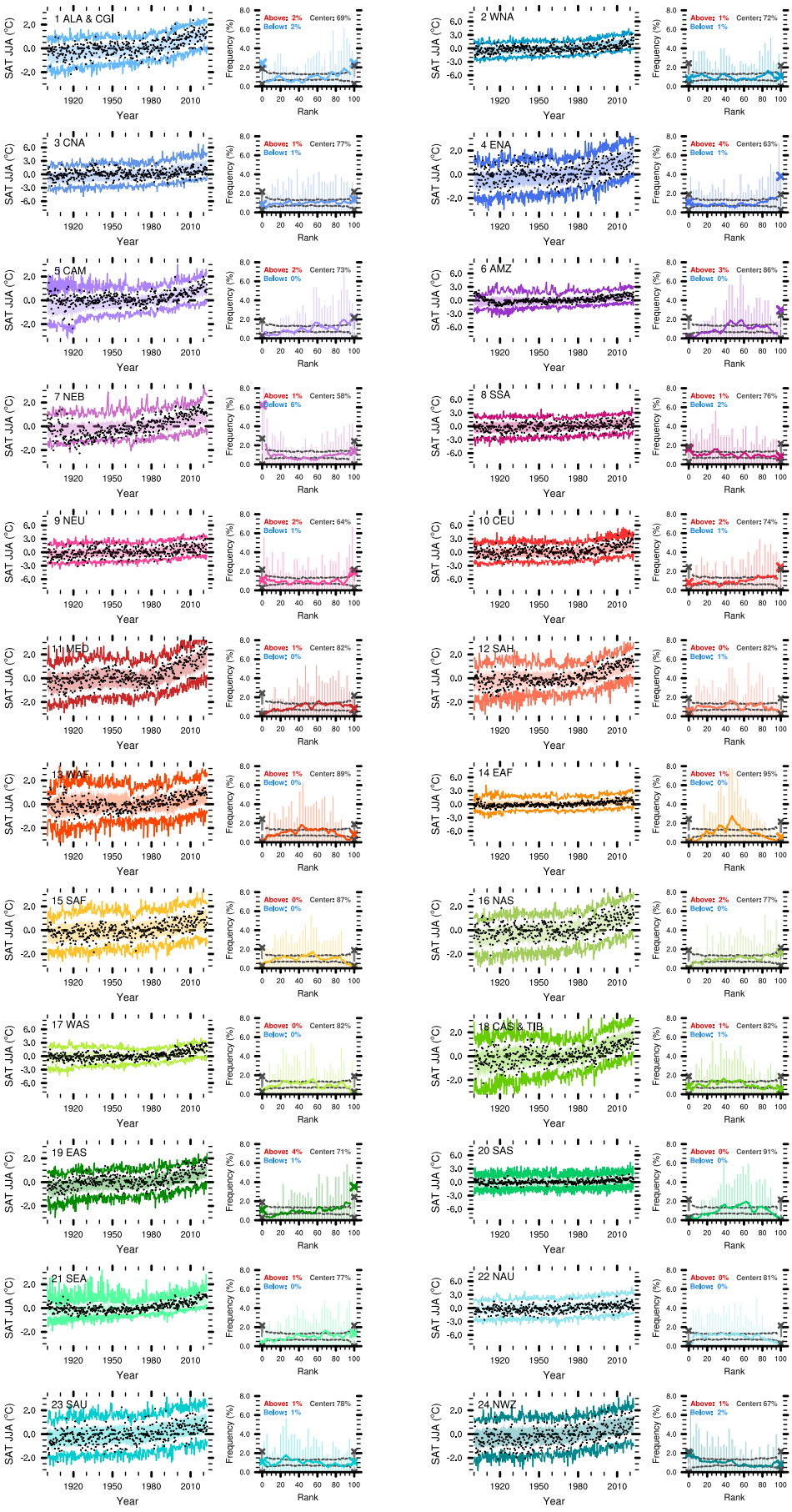
TAS JJA Obs outside ensemble spread (%)



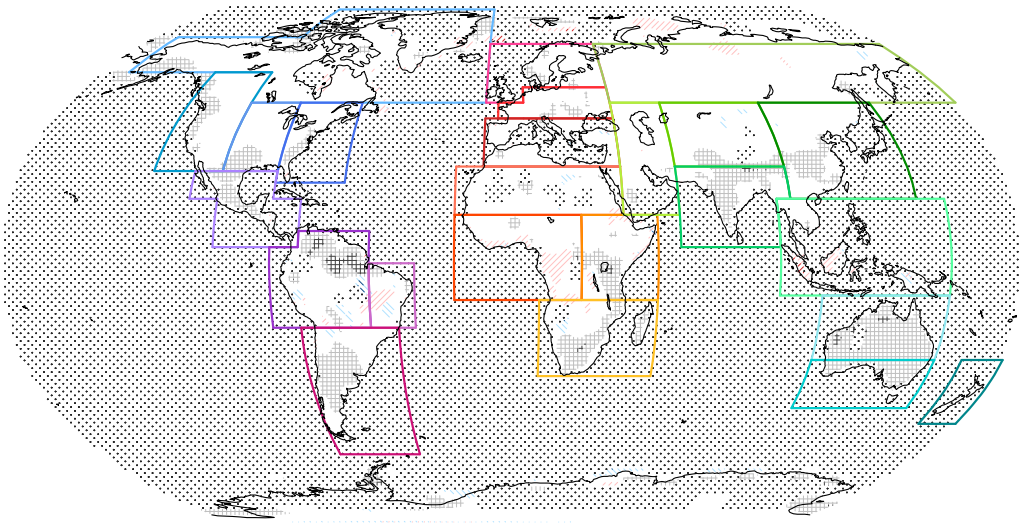
MPI-GE5 vs GISTEMPv4 TAS DJF



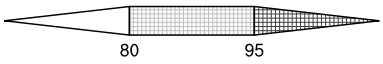
MPI-GE5 vs GISTEMPv4 TAS JJA



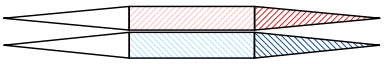
MPI-GE6 vs GISTEMPv4 White Area = 70.0 %



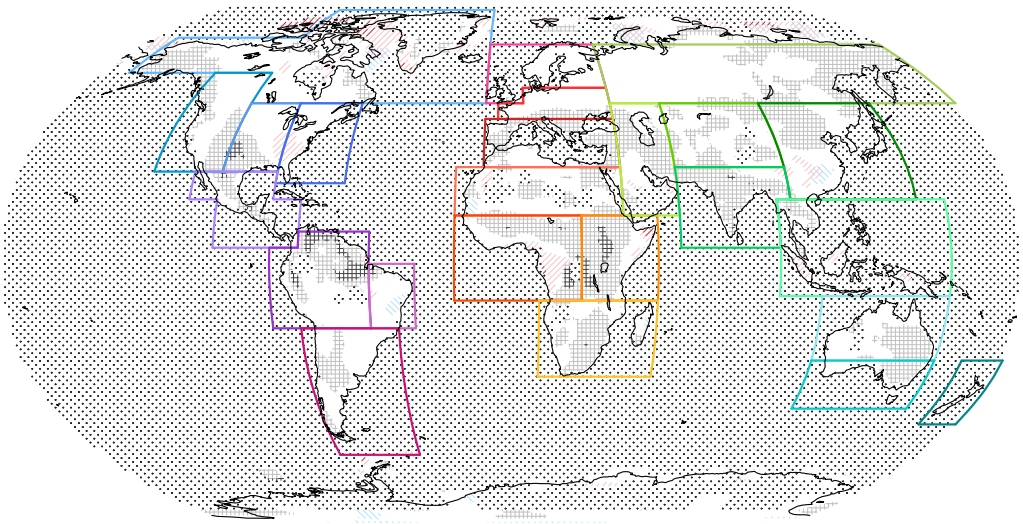
TAS DJF Obs inside central 75th percentile (%)



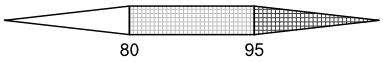
TAS DJF Obs outside ensemble spread (%)



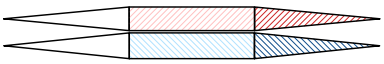
MPI-GE6 vs GISTEMPv4 White Area = 54.1 %



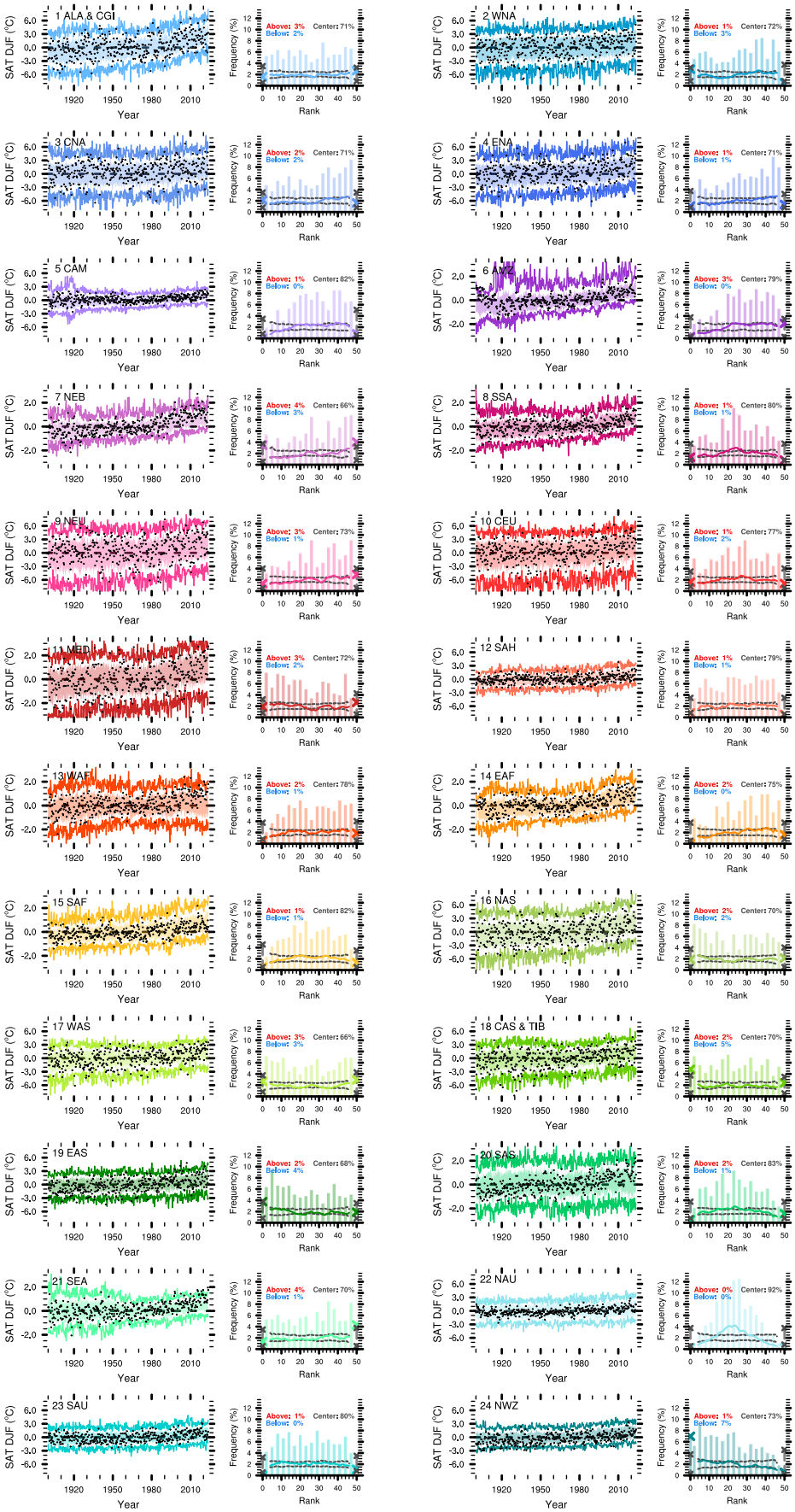
TAS JJA Obs inside central 75th percentile (%)



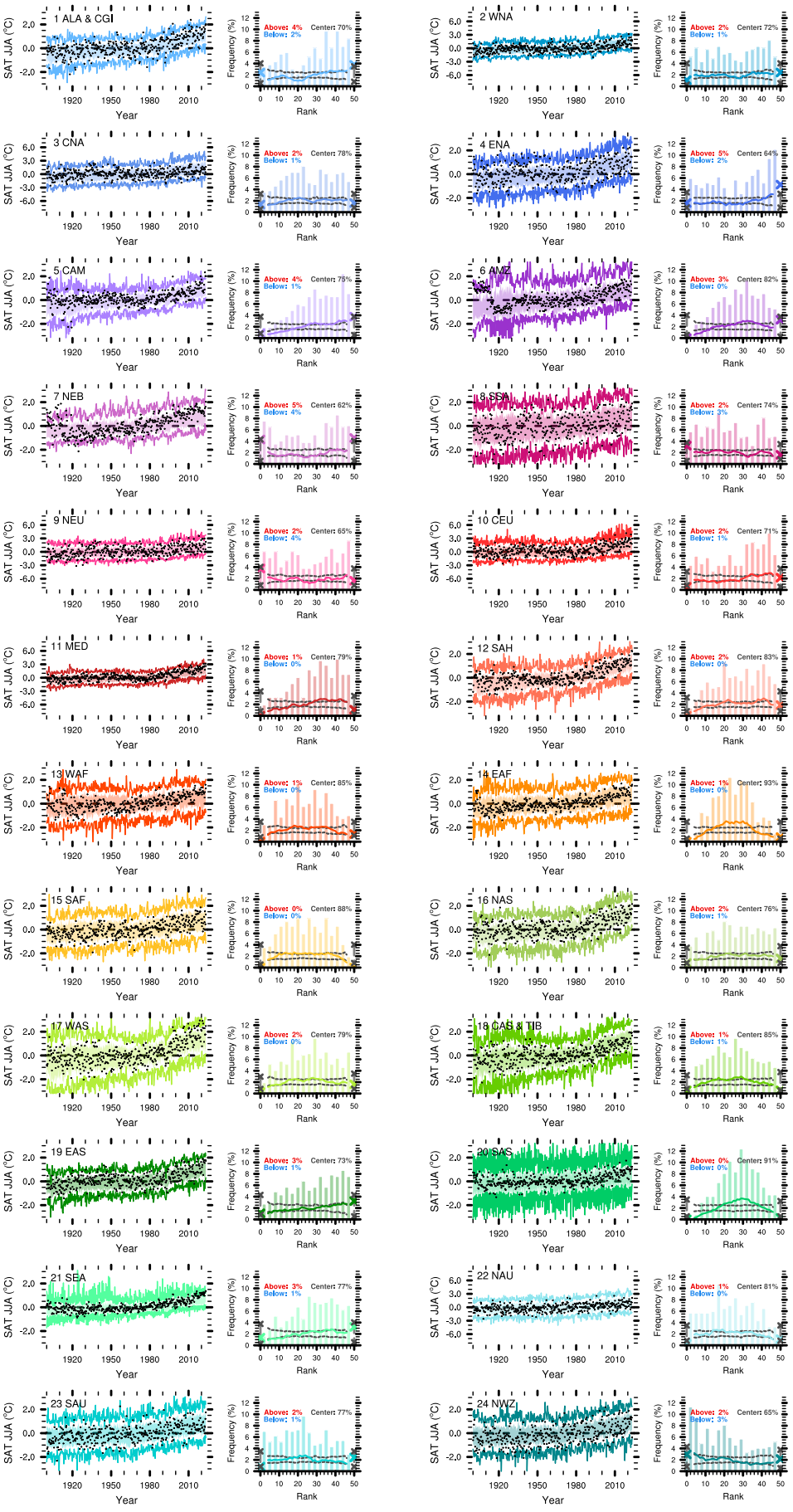
TAS JJA Obs outside ensemble spread (%)



MPI-GE6 vs GISTEMPv4 TAS DJF



MPI-GE6 vs GISTEMPv4 TAS JJA



Detrended Ocean Surface Temperatures

Rank-frequency variability evaluation framework for detrended sea surface temperature (SST) anomalies over ocean grid cells.

Maps show grid-cell evaluation of the simulated DJF and JJA monthly mean temperature anomalies for the 11 SMILEs included in this study against ERSSTv5 observations globally. Gray hatching represents where observations cluster within the 75th percentile bounds of the ensemble (12.5th to 87.5th percentiles) for more than 80% of months (light grey) or for more than 95% of months (dark grey). Red and blue shading represents where observations are larger than the ensemble maximum (red) or smaller than the ensemble minimum (blue), respectively, for more than 8% of the months (light red and blue) or for more than 20% of the months (dark red and blue). Dotted areas represent ocean areas or grid cells where observations are missing and are therefore excluded from this analysis. Colored boxes demark the boundaries of each ocean region assessed. The percentage of assessed grid-cells that present none of these biases in given at the top (white area).

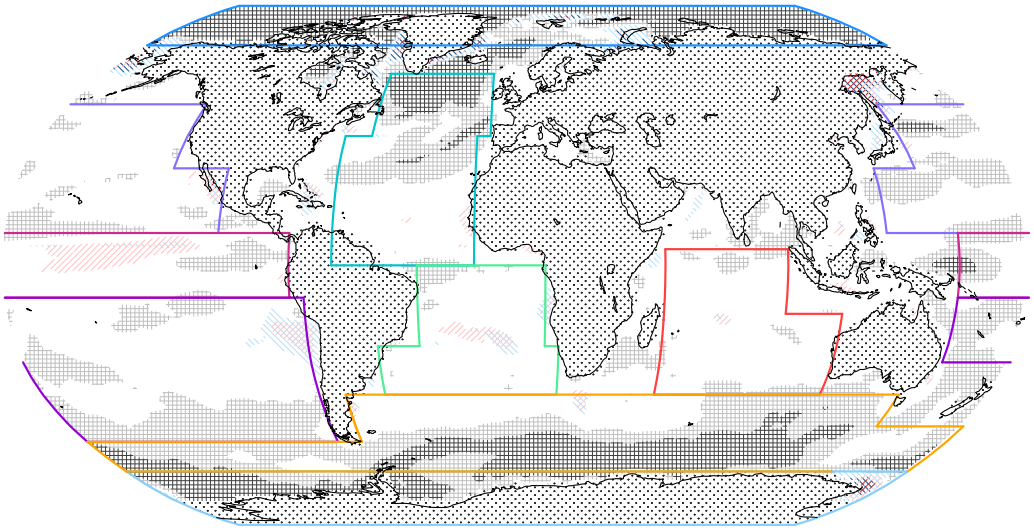
Time series and rank frequency histograms show spatially aggregated DJF and JJA SSTs for each ocean regions for all 11 SMILEs. Time series show the ensemble maximum and minimum (coloured lines) and central 75th percentile ensemble spread (shading) are shown against observations (black dots).

Rank histograms represent the frequency of each place that observations would take in a list of ensemble members ordered by ascending temperature anomaly values. Rank 0 indicates observations are below the minimum ensemble value, and rank n, with n the number of ensemble members, indicates that observations exceed the maximum ensemble value for that particular month. For a model that perfectly represents observations over an infinitely-long observational record, all ranks should occur with similar frequency and this histogram should be roughly flat.

To illustrate how internal variability may affect rank frequencies given the non-infinite record length considered, we also include a perfect-model comparison, which shows the range of rank frequencies that each ensemble member would take if it were observations. If the rank exhibited by observations (colors) is within this perfect-model range (grey), the rank frequency evaluation shows an adequate model performance, and any deviations from a perfectly frank rank histogram can be assumed to be within deviations that could be caused by internal variability. Lines in the rank histogram illustrate the rank histogram's slope, as the mean rank frequency over a centered 6-bin window for observations (solid colored lines), and the 5-95th perc. perfect model range (gray dashed lines). Crosses represent the frequency of minimum (0) and maximum (number of members) ranks for observations (colors), and for the 5-95th perc. perfect model range (gray).

Percentages at the top of the rank histograms show how often monthly anomalies fall above or below ensemble limits (red and blue, respectively) and within the 75th perc. Range (grey), analogous to the criteria chosen for the map-based evaluation but for spatially aggregated values.

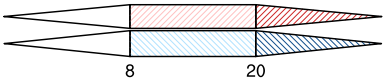
ACCESS vs ERSSTv5 White Area = 59.0 %



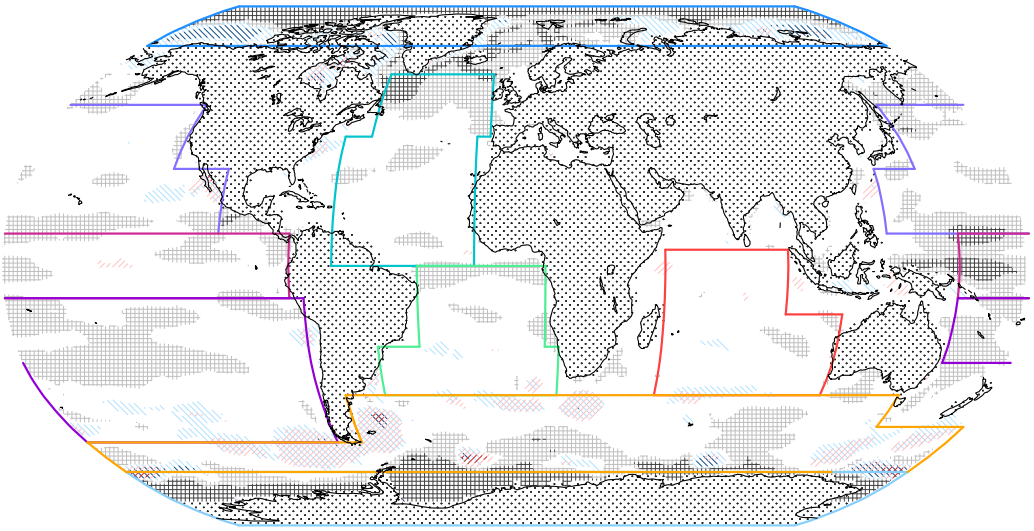
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



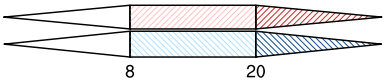
ACCESS vs ERSSTv5 White Area = 62.7 %

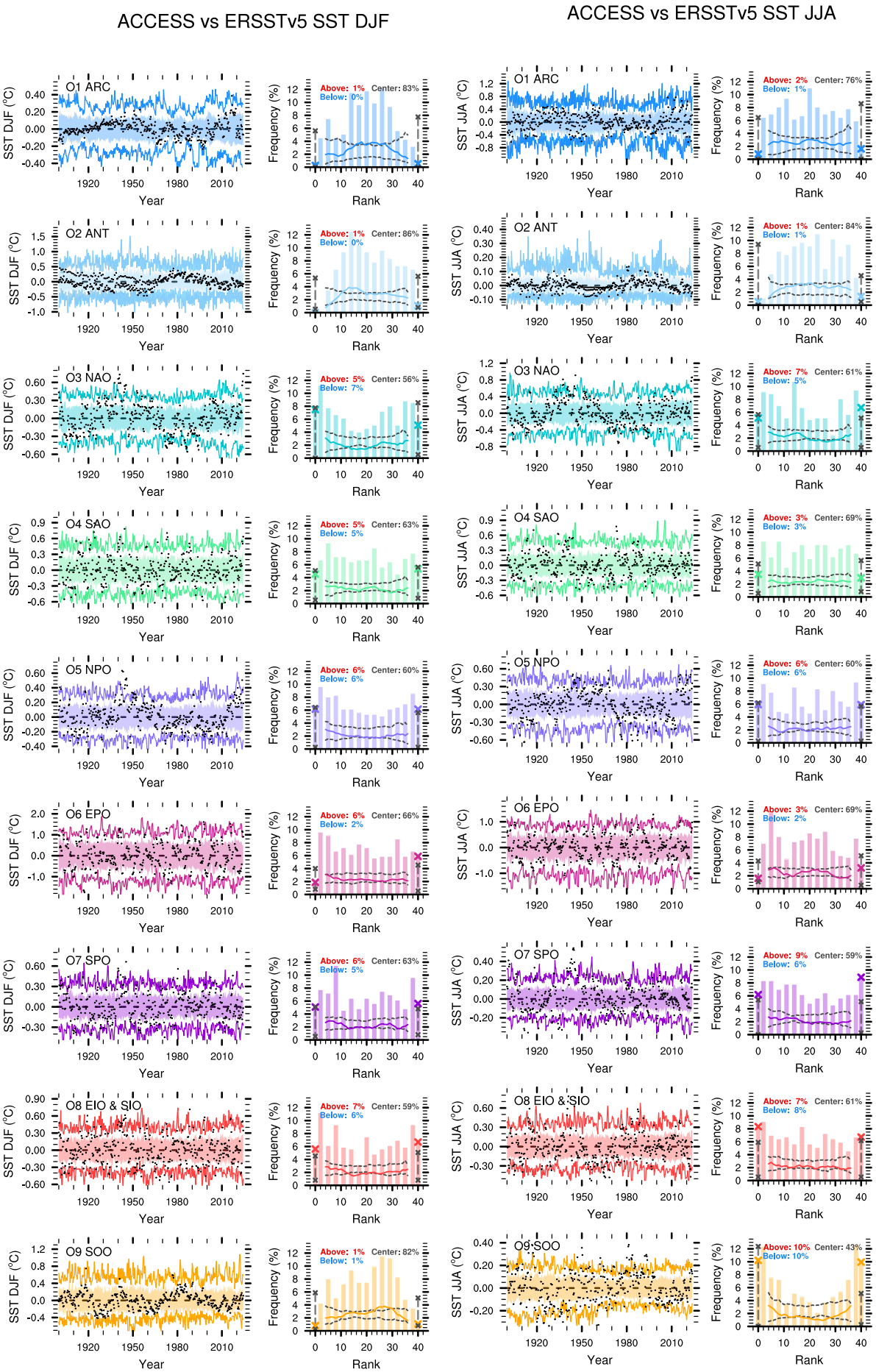


SST JJA Obs inside central 75th percentile (%)

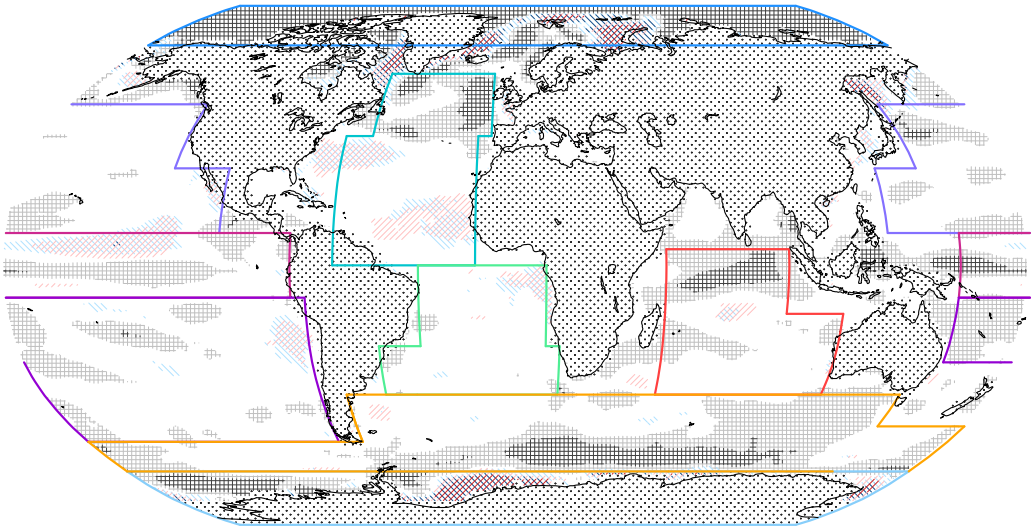


SST JJA Obs outside ensemble spread (%)




ACCESS vs ERSSTv5 SST DJF
ACCESS vs ERSSTv5 SST JJA
O1 ARC
O2 ANT
O3 NAO
O4 SAO
O5 NPO
O6 EPO
O7 SPO
O8 EIO & SIO
O9 SOO
SST DJF (°C)
SST JJA (°C)
Year
Frequency (%)
Rank
Above: 1% Center: 83% Below: 0%
Above: 2% Center: 76% Below: 1%
Above: 1% Center: 86% Below: 0%
Above: 1% Center: 84% Below: 1%
Above: 5% Center: 56% Below: 7%
Above: 7% Center: 61% Below: 5%
Above: 5% Center: 63% Below: 5%
Above: 3% Center: 69% Below: 3%
Above: 6% Center: 60% Below: 6%
Above: 6% Center: 60% Below: 6%
Above: 6% Center: 66% Below: 2%
Above: 3% Center: 69% Below: 2%
Above: 6% Center: 63% Below: 5%
Above: 9% Center: 59% Below: 6%
Above: 7% Center: 59% Below: 6%
Above: 7% Center: 61% Below: 8%
Above: 1% Center: 82% Below: 1%
Above: 10% Center: 43% Below: 10%

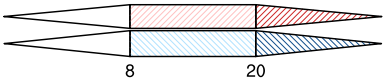
CanESM2 vs ERSSTv5 White Area = 57.0 %



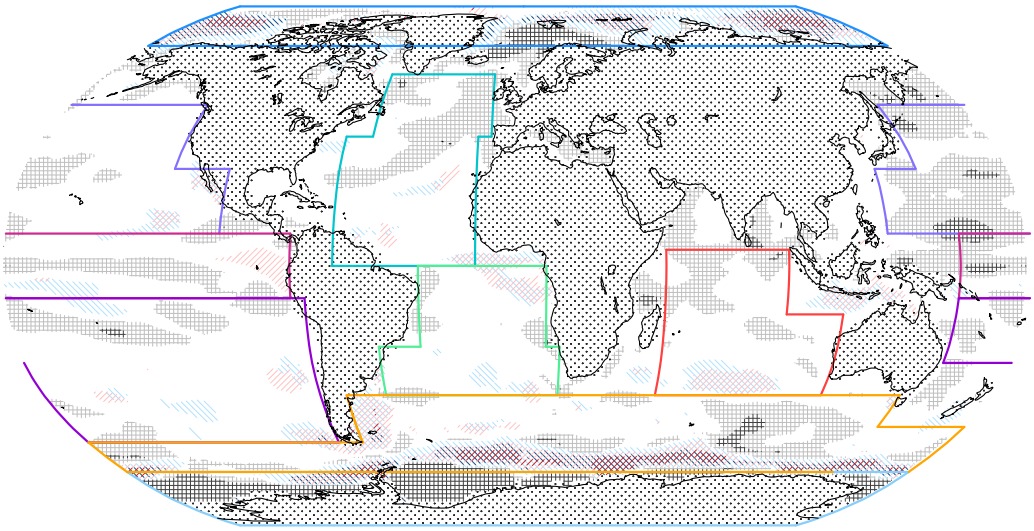
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



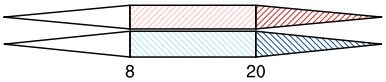
CanESM2 vs ERSSTv5 White Area = 56.2 %



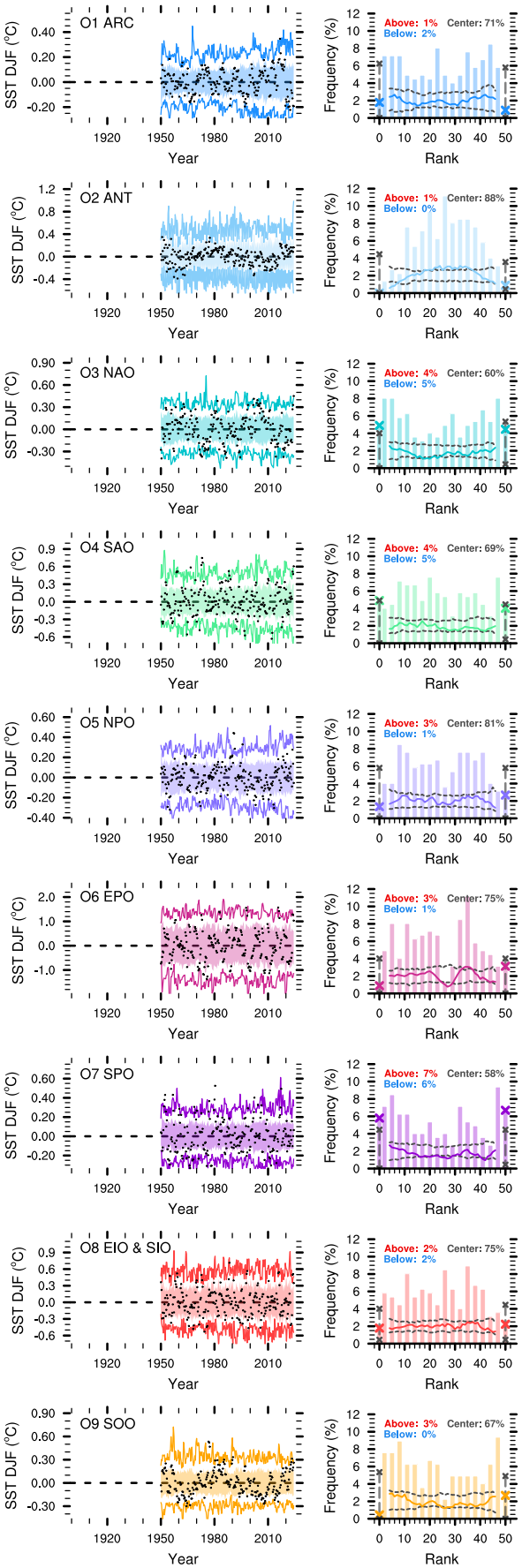
SST JJA Obs inside central 75th percentile (%)



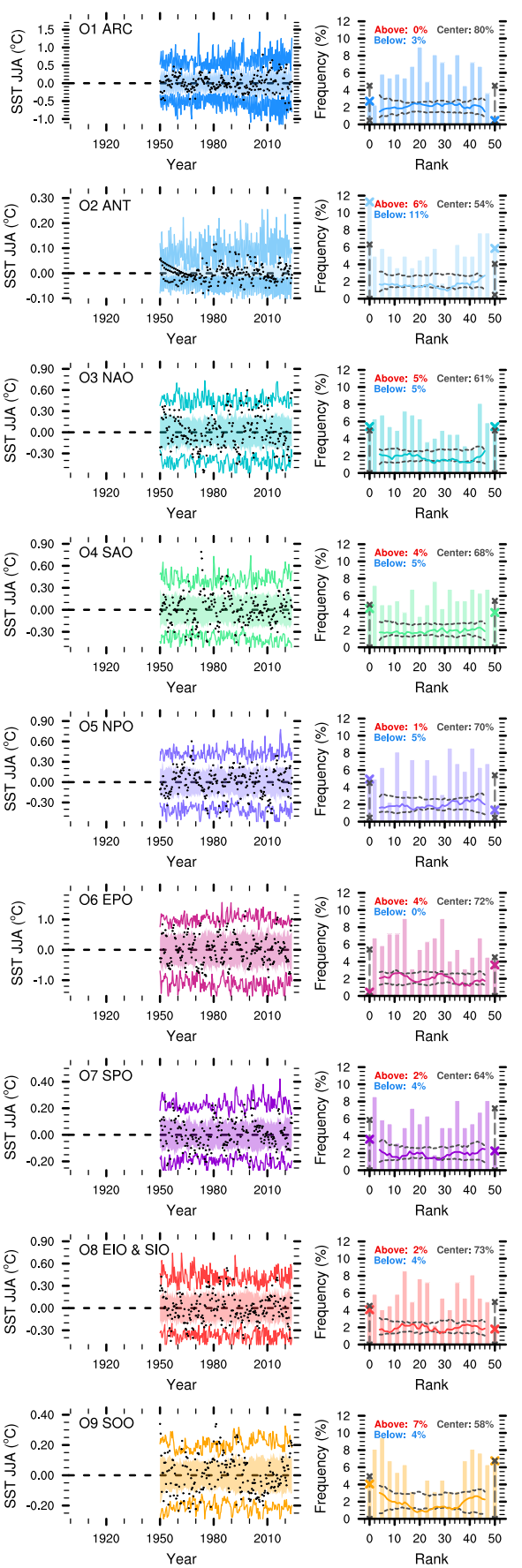
SST JJA Obs outside ensemble spread (%)



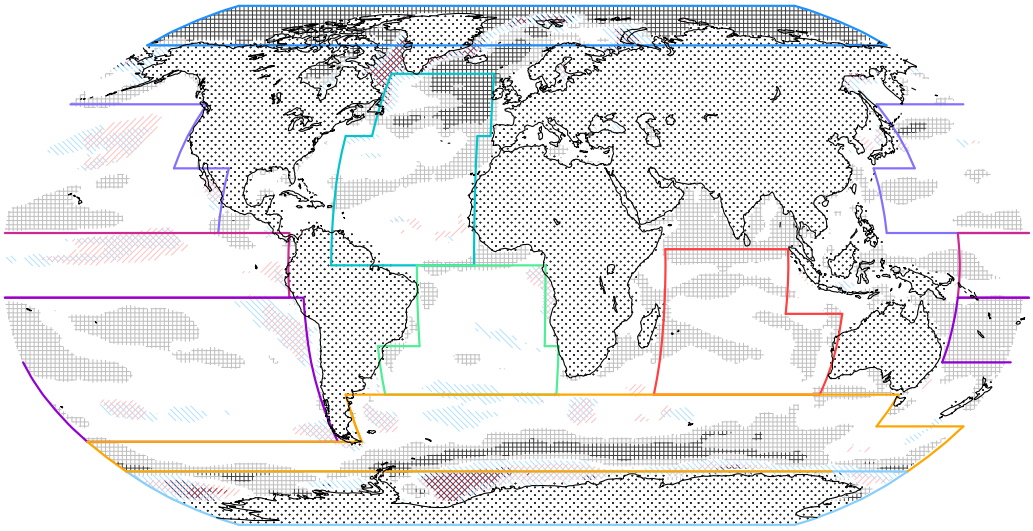
CanESM2 vs ERSSTv5 SST DJF



CanESM2 vs ERSSTv5 SST JJA



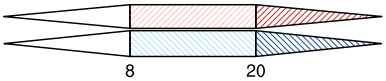
CanESM5 vs ERSSTv5 White Area = 58.4 %



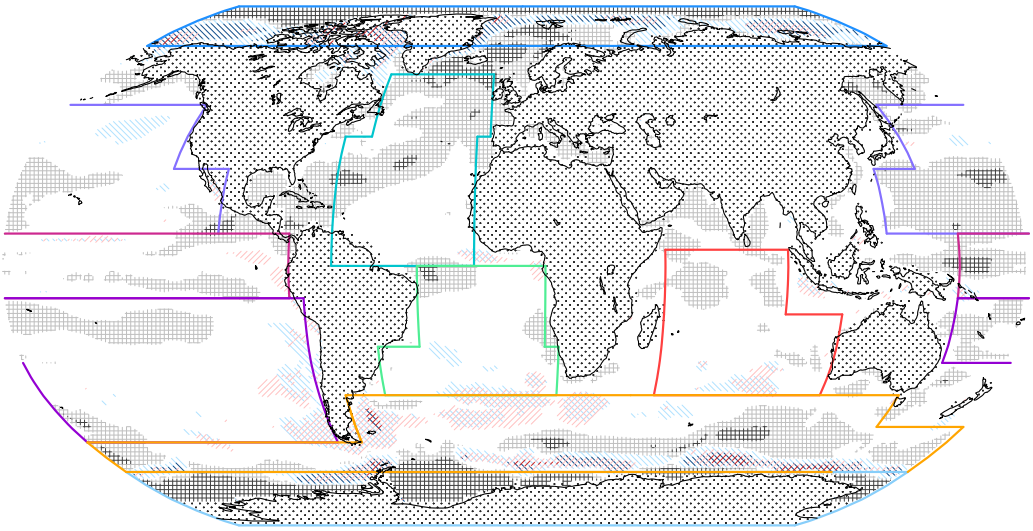
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



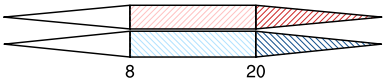
CanESM5 vs ERSSTv5 White Area = 56.2 %



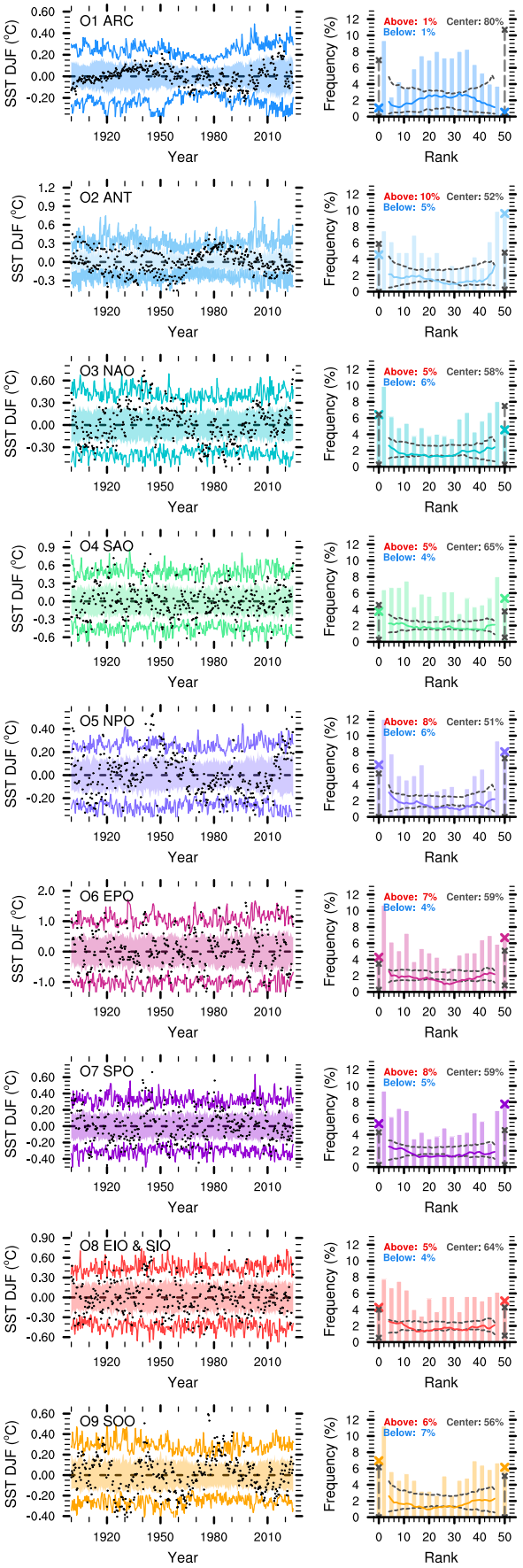
SST JJA Obs inside central 75th percentile (%)



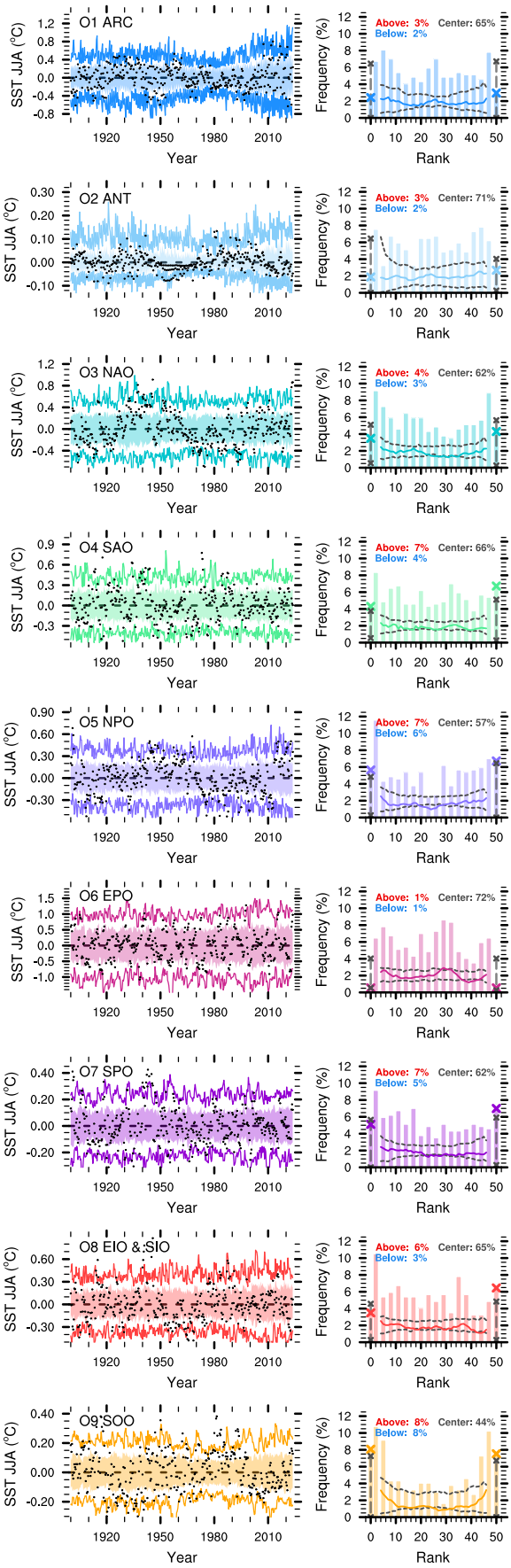
SST JJA Obs outside ensemble spread (%)



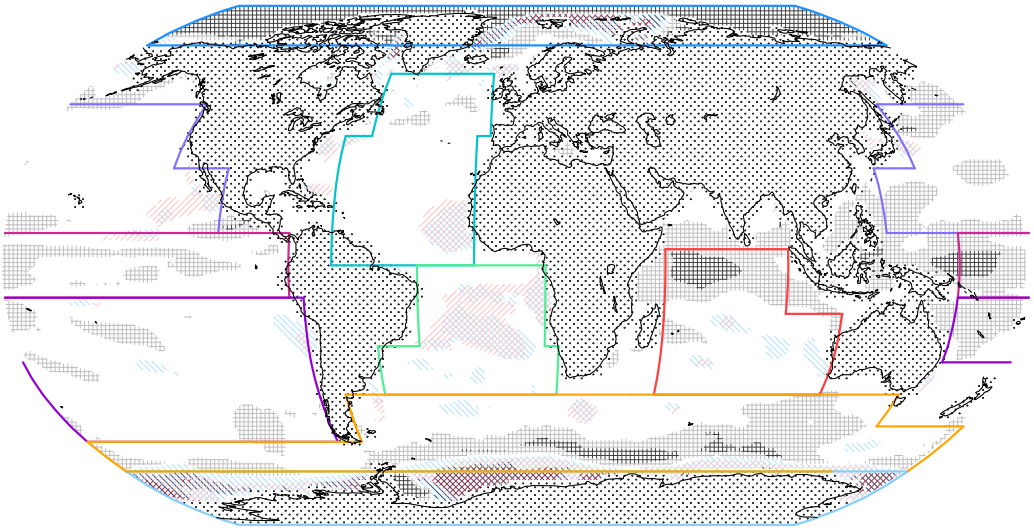
CanESM5 vs ERSSTv5 SST DJF



CanESM5 vs ERSSTv5 SST JJA



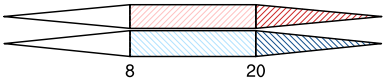
CESM-LE vs ERSSTv5 White Area = 62.8 %



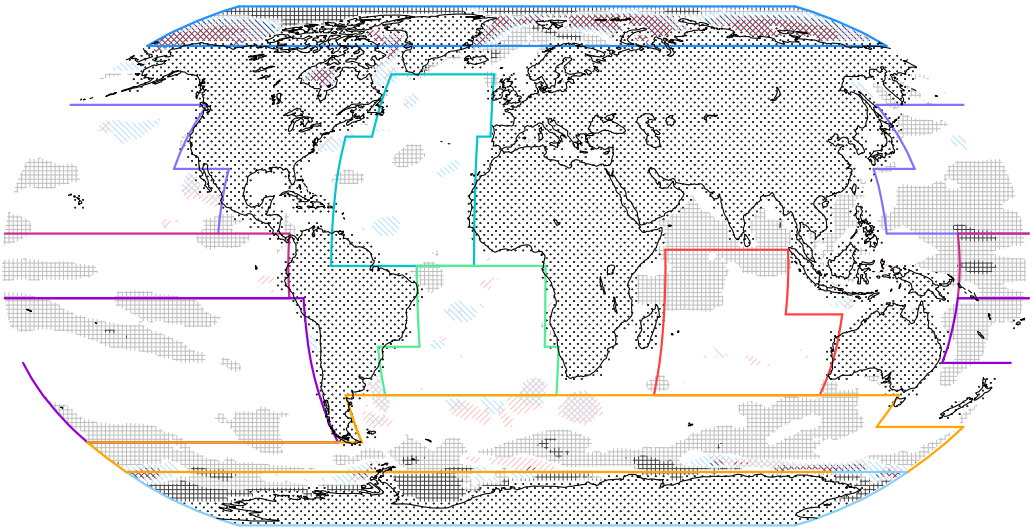
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



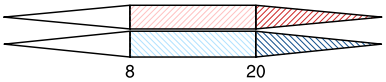
CESM-LE vs ERSSTv5 White Area = 65.9 %



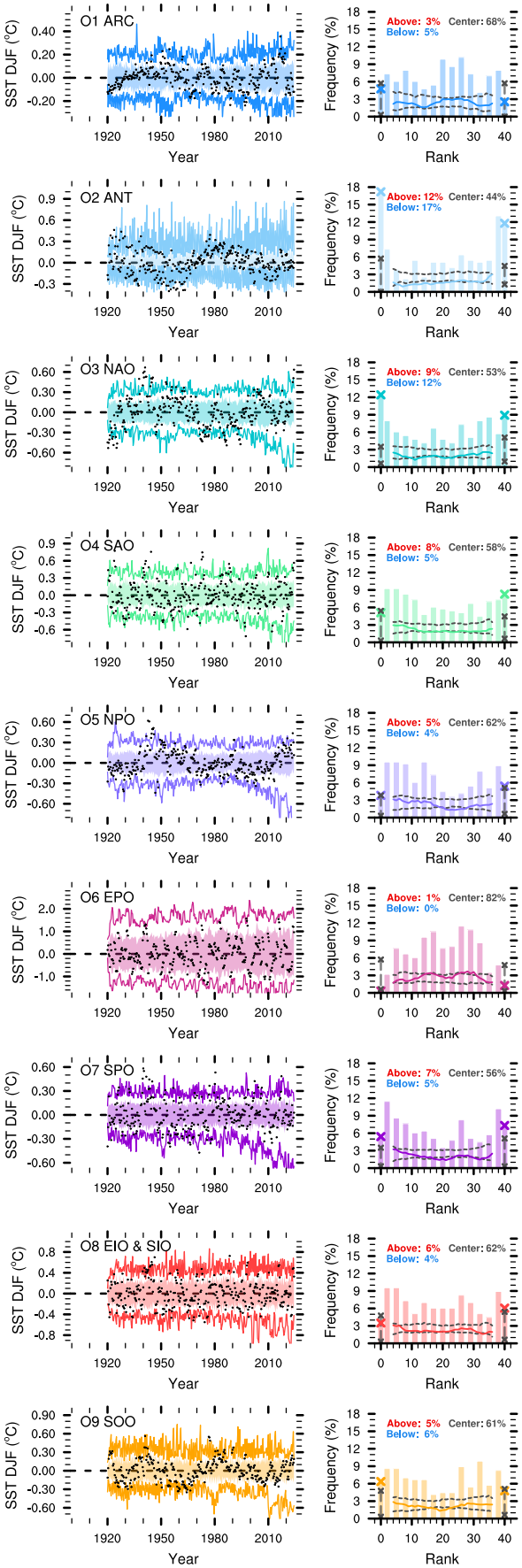
SST JJA Obs inside central 75th percentile (%)



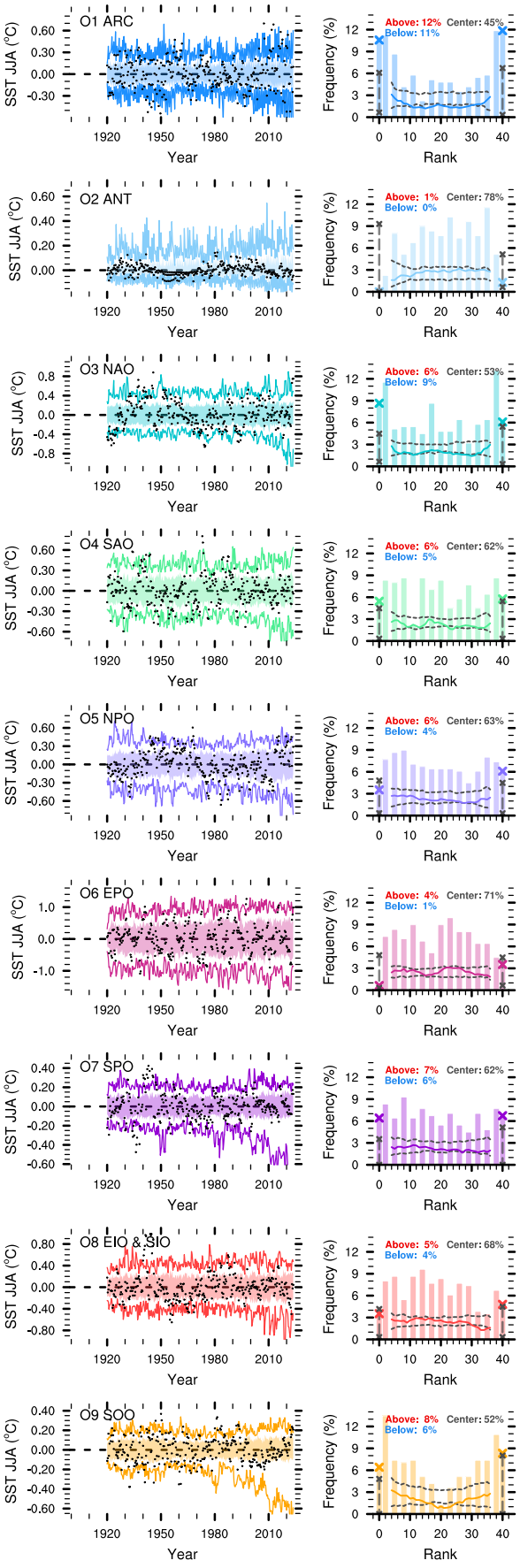
SST JJA Obs outside ensemble spread (%)



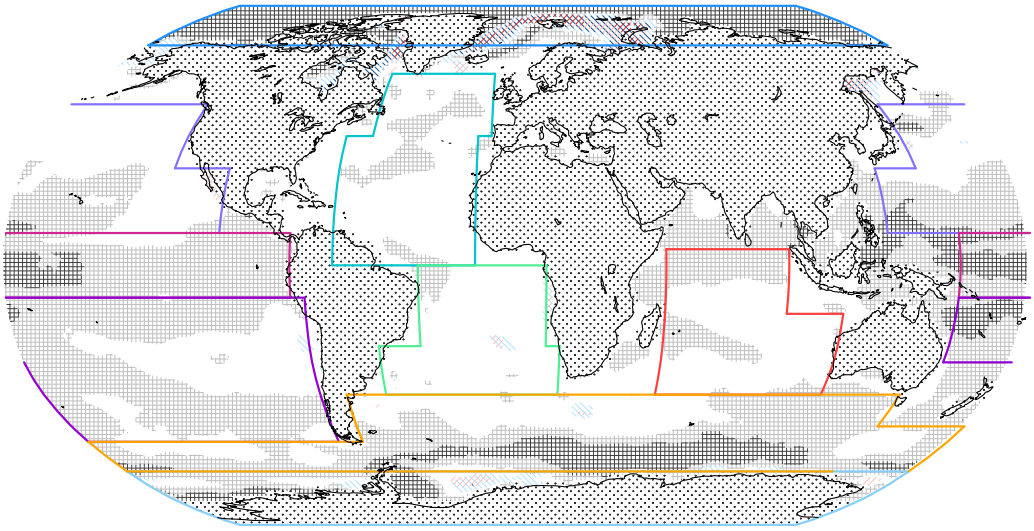
CESM-LE vs ERSSTv5 SST DJF



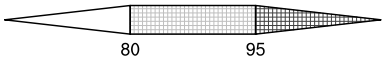
CESM-LE vs ERSSTv5 SST JJA



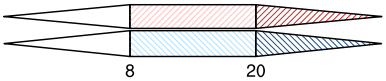
CESM2-LE vs ERSSTv5 White Area = 47.0 %



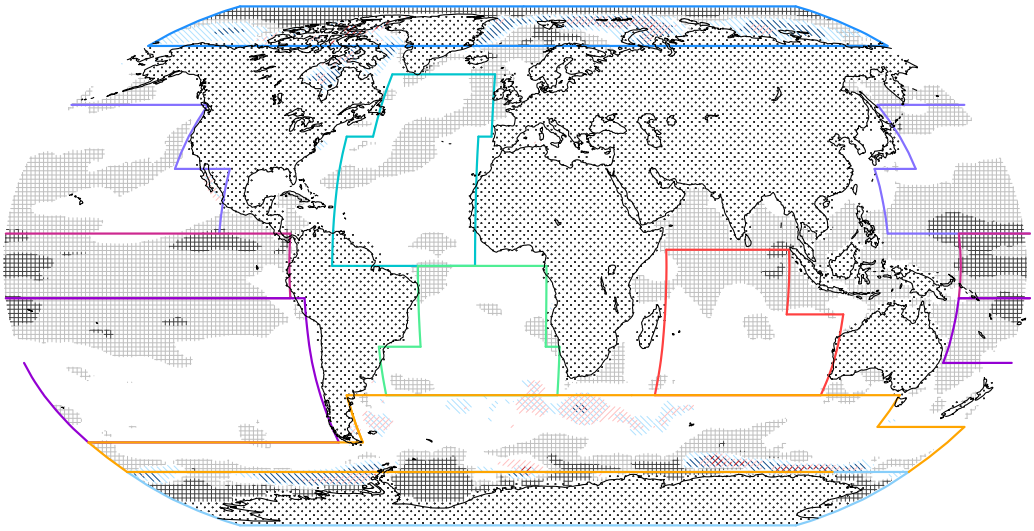
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



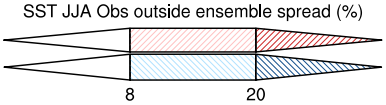
CESM2-LE vs ERSSTv5 White Area = 53.2 %



SST JJA Obs inside central 75th percentile (%)

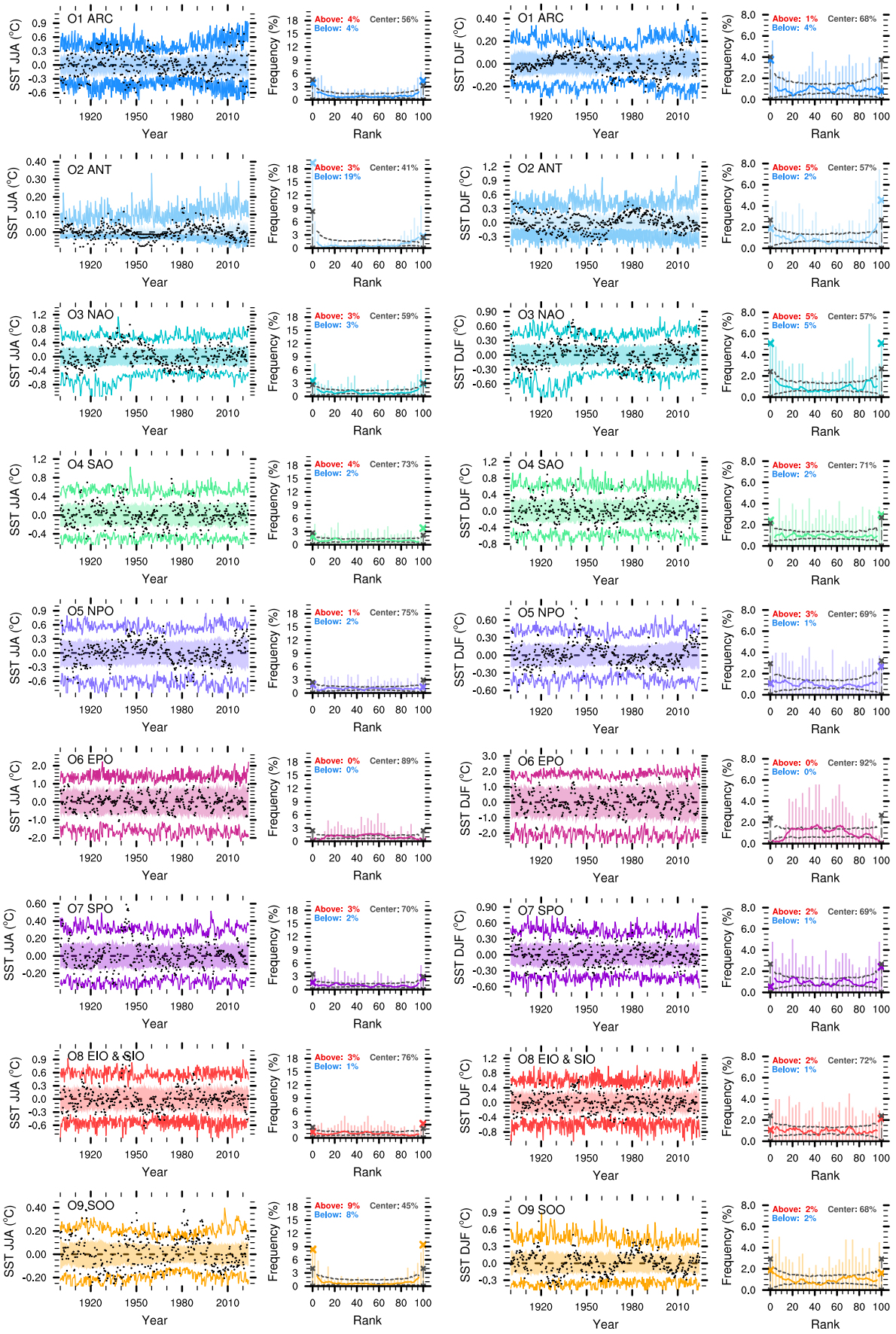


SST JJA Obs outside ensemble spread (%)

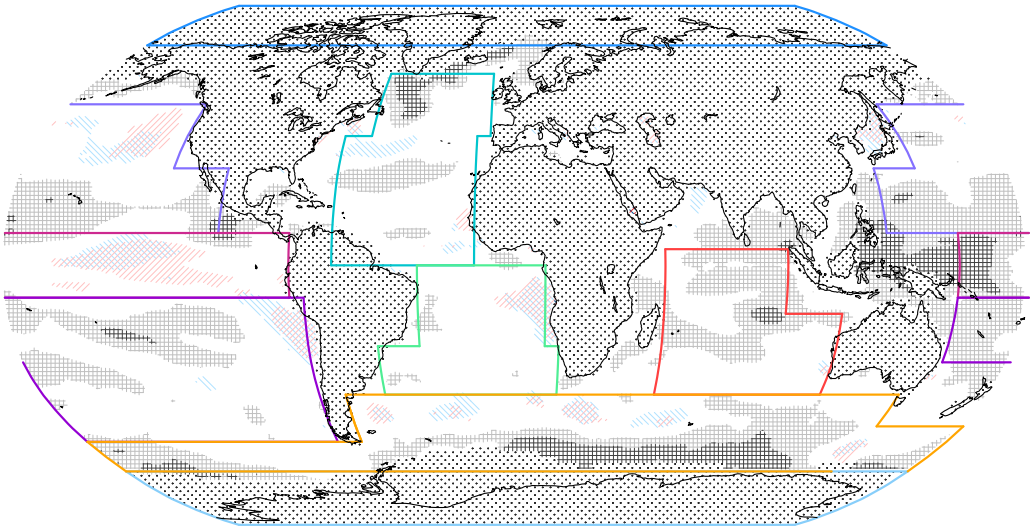


CESM2-LE vs ERSSTv5 SST JJA

CESM2-LE vs ERSSTv5 SST DJF



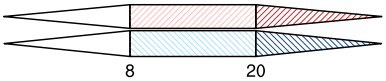
CSIRO-Mk360 vs ERSSTv5 White Area = 57.8 %



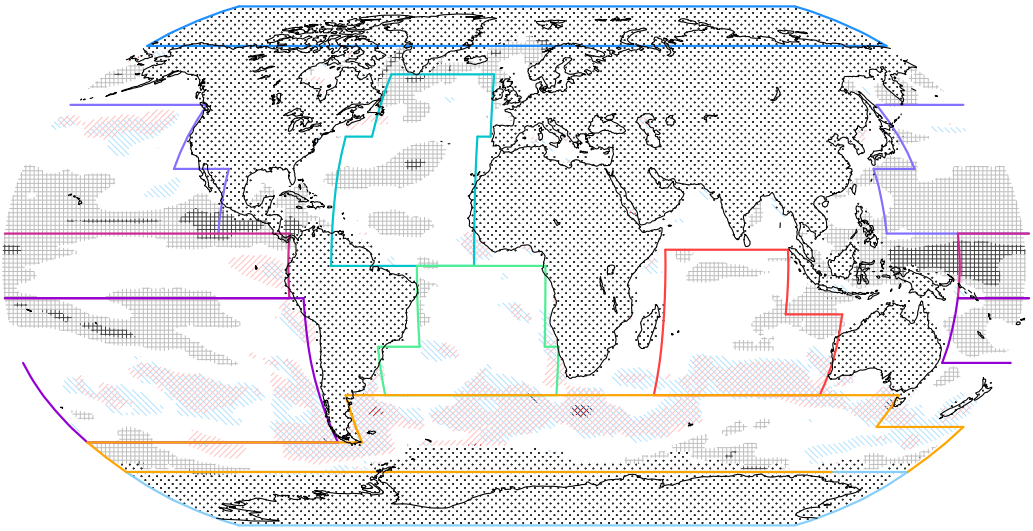
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



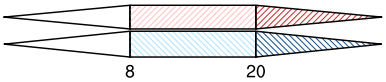
CSIRO-Mk360 vs ERSSTv5 White Area = 55.7 %



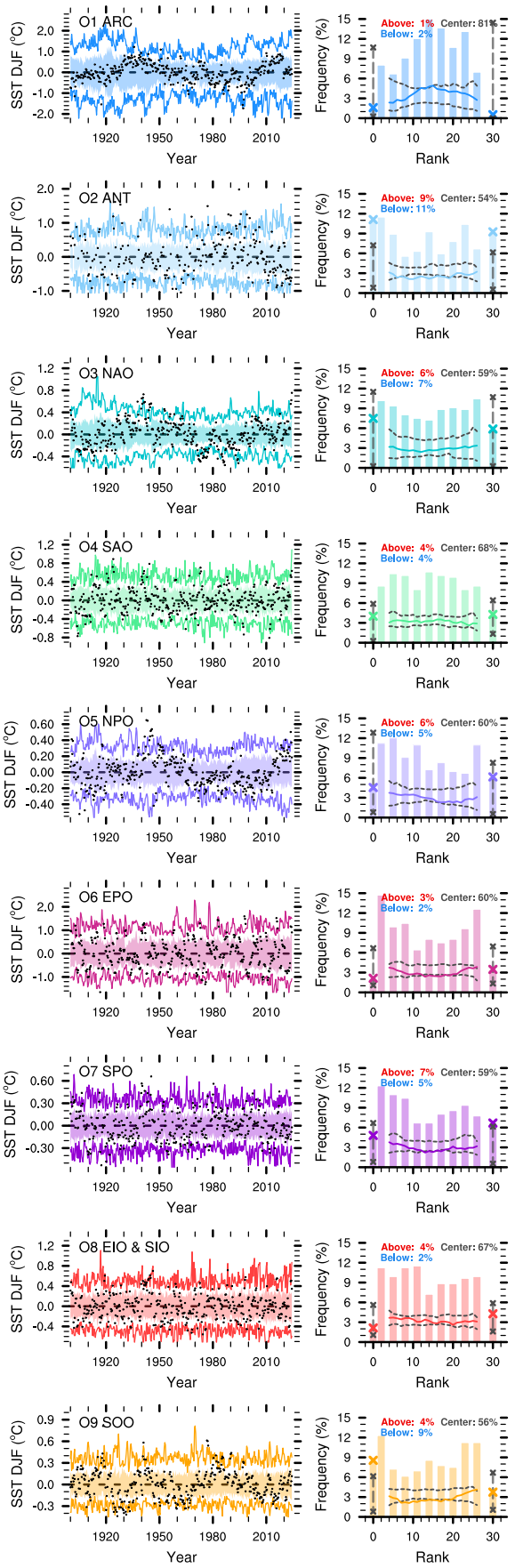
SST JJA Obs inside central 75th percentile (%)



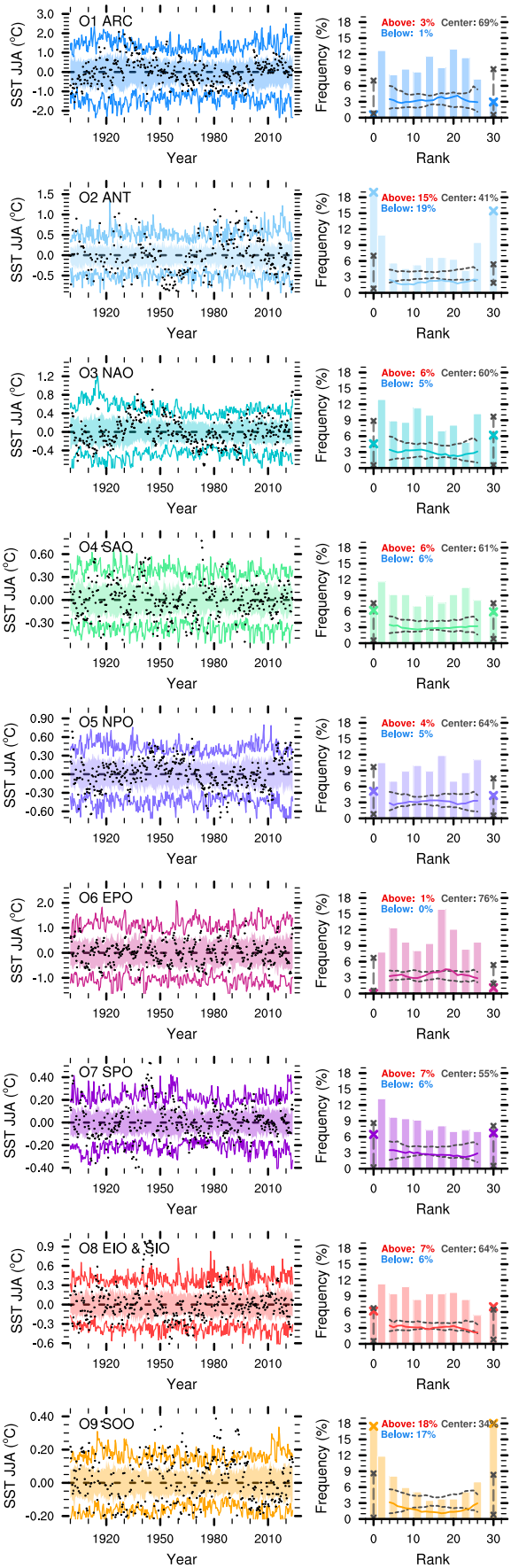
SST JJA Obs outside ensemble spread (%)



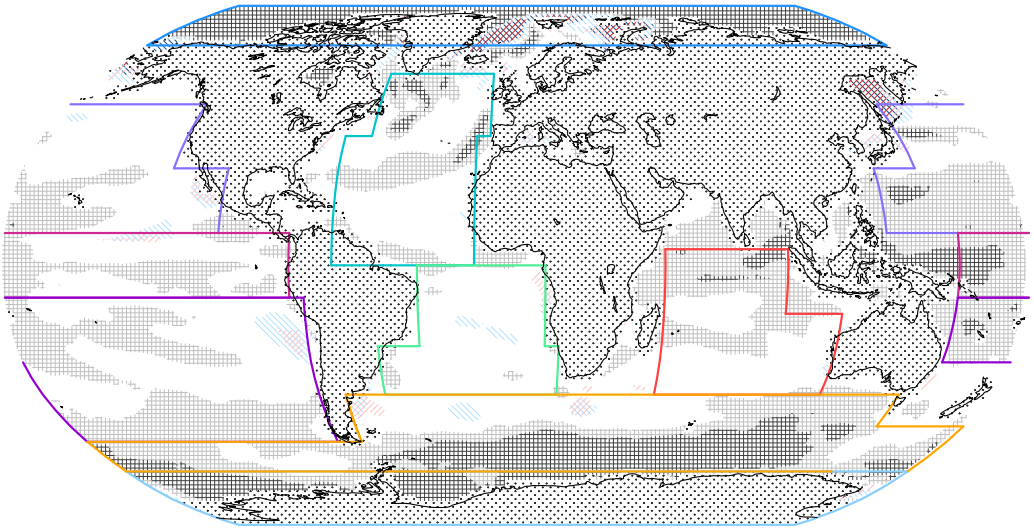
CSIRO-Mk360 vs ERSSTv5 SST DJF



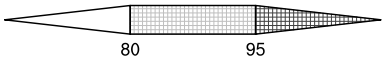
CSIRO-Mk360 vs ERSSTv5 SST JJA



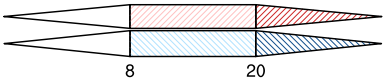
GFDL-ESM2M vs ERSSTv5 White Area = 49.3 %



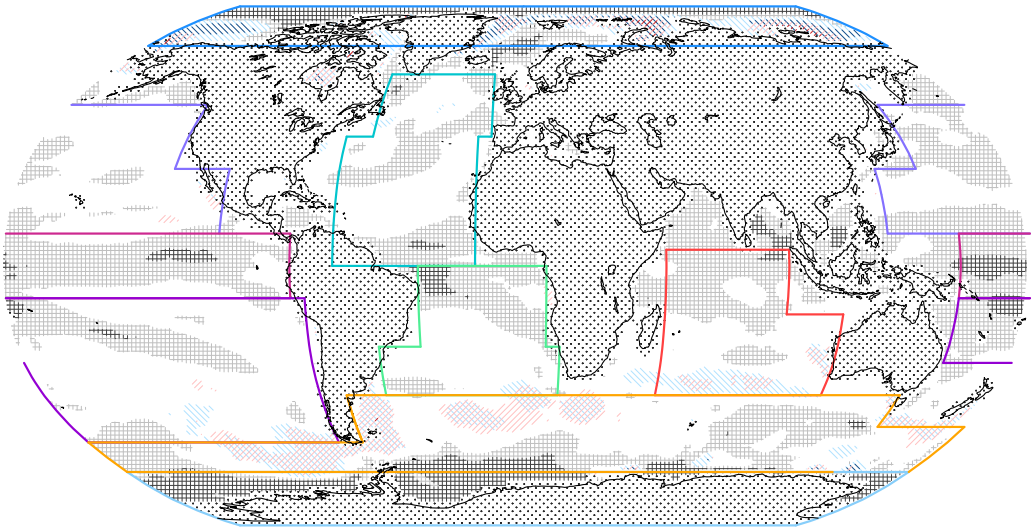
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



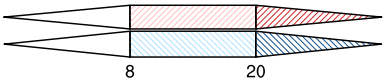
GFDL-ESM2M vs ERSSTv5 White Area = 49.6 %



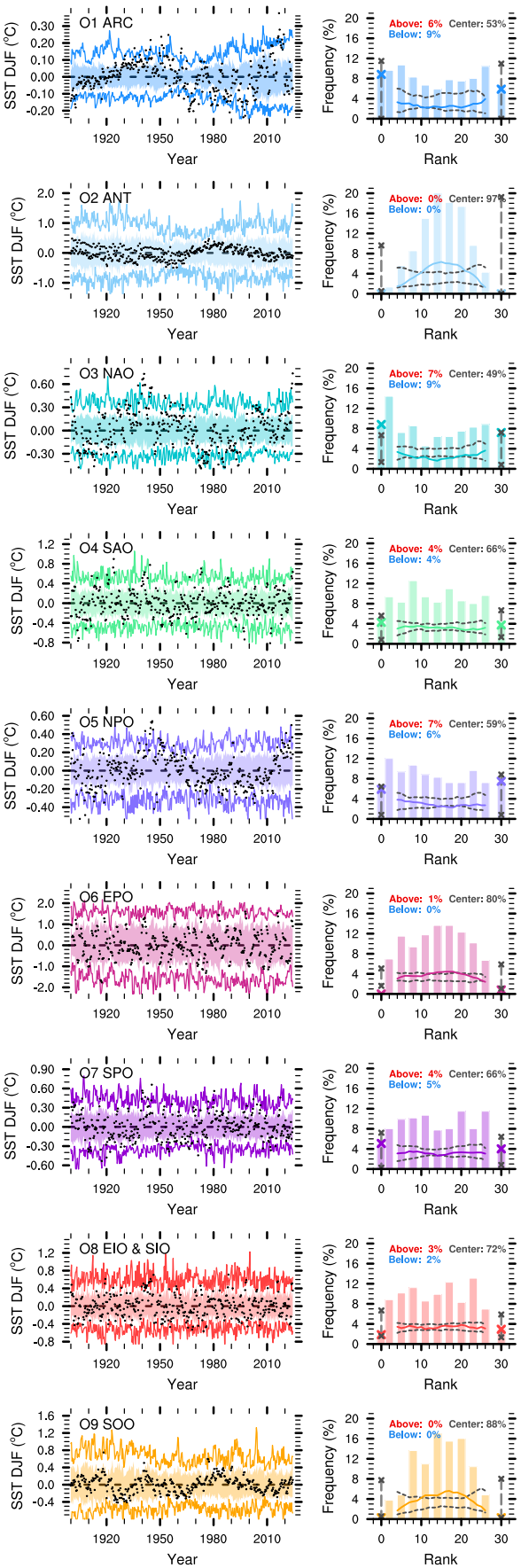
SST JJA Obs inside central 75th percentile (%)



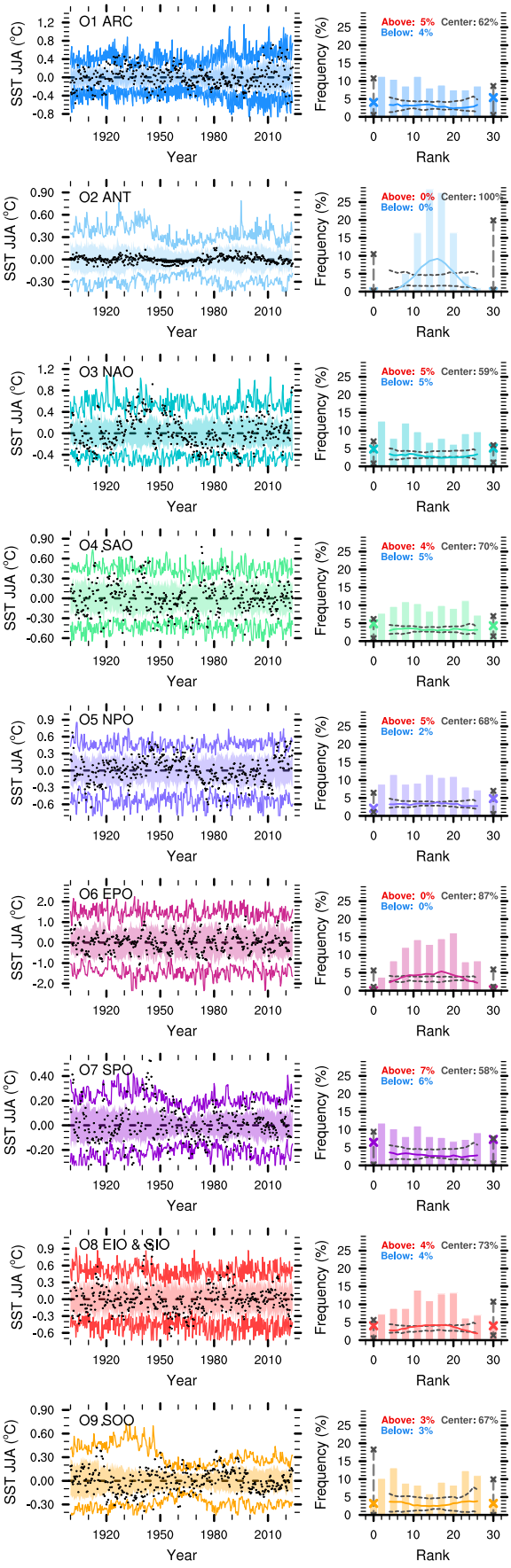
SST JJA Obs outside ensemble spread (%)



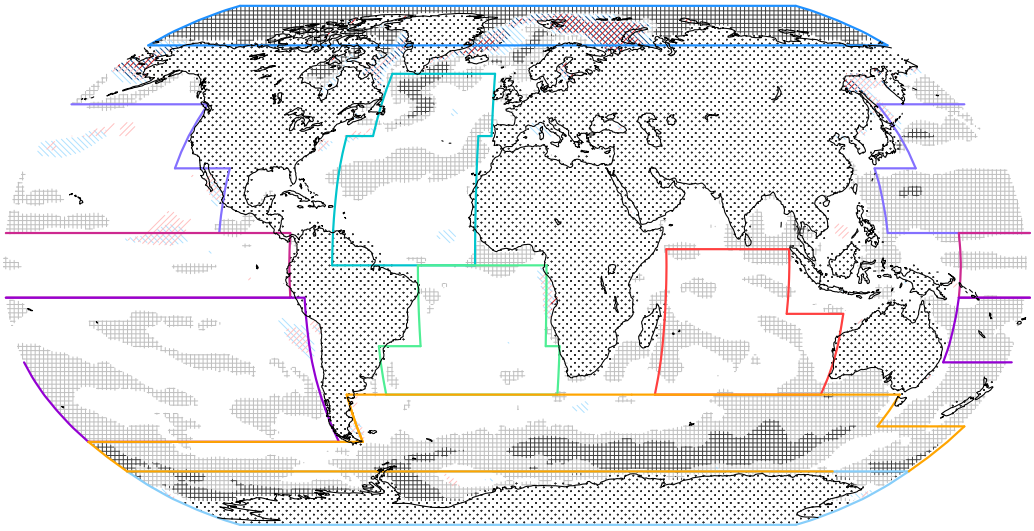
GFDL-ESM2M vs ERSSTv5 SST DJF



GFDL-ESM2M vs ERSSTv5 SST JJA



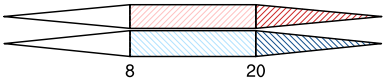
GFDL-SPEAR vs ERSSTv5 White Area = 58.9 %



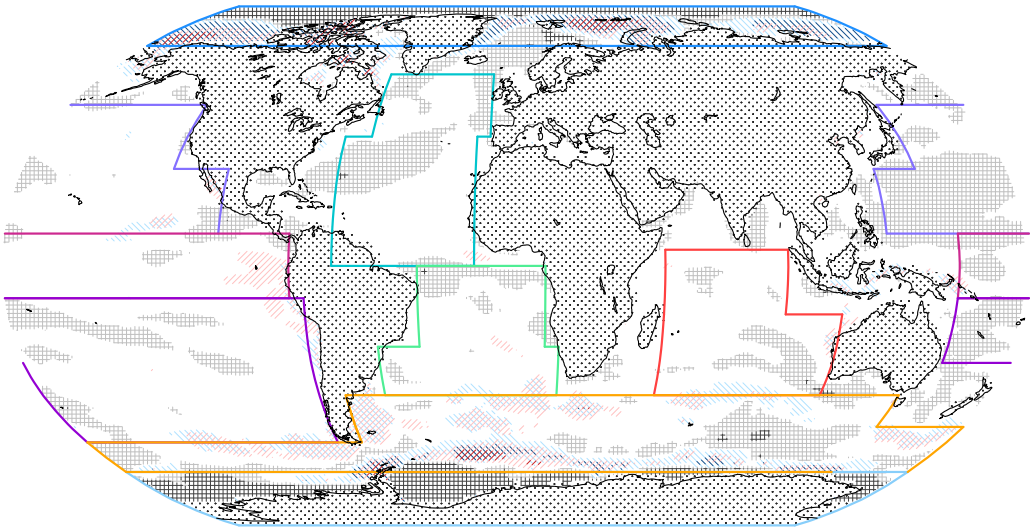
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



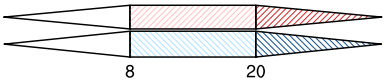
GFDL-SPEAR vs ERSSTv5 White Area = 62.0 %



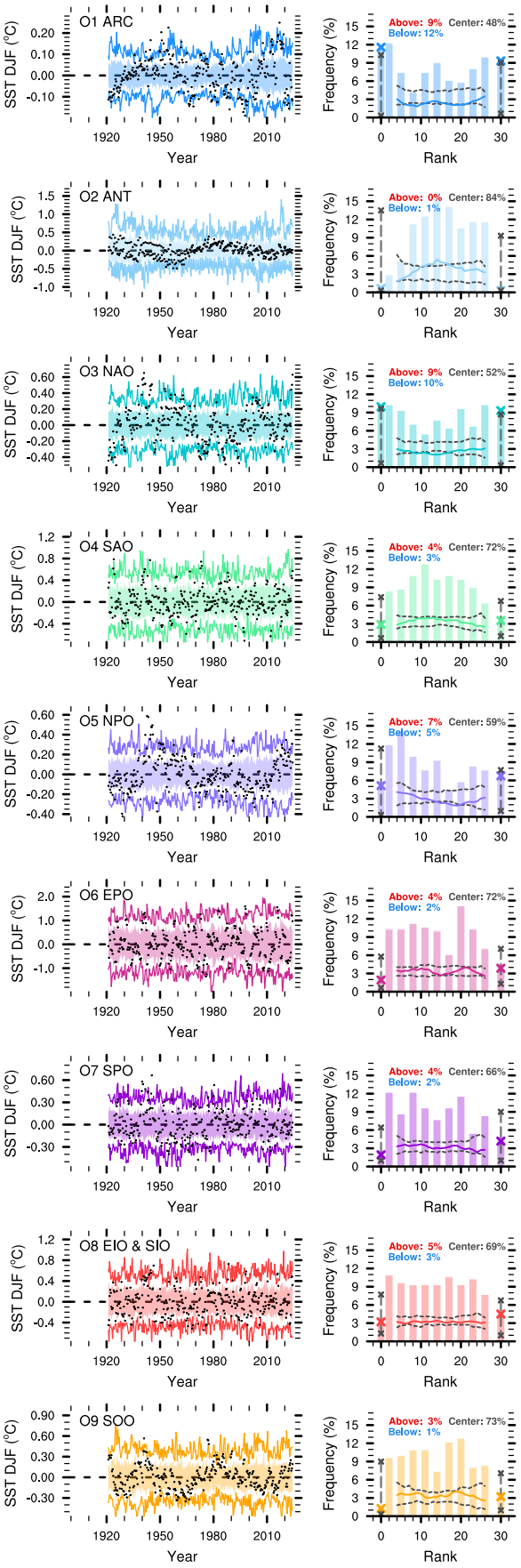
SST JJA Obs inside central 75th percentile (%)



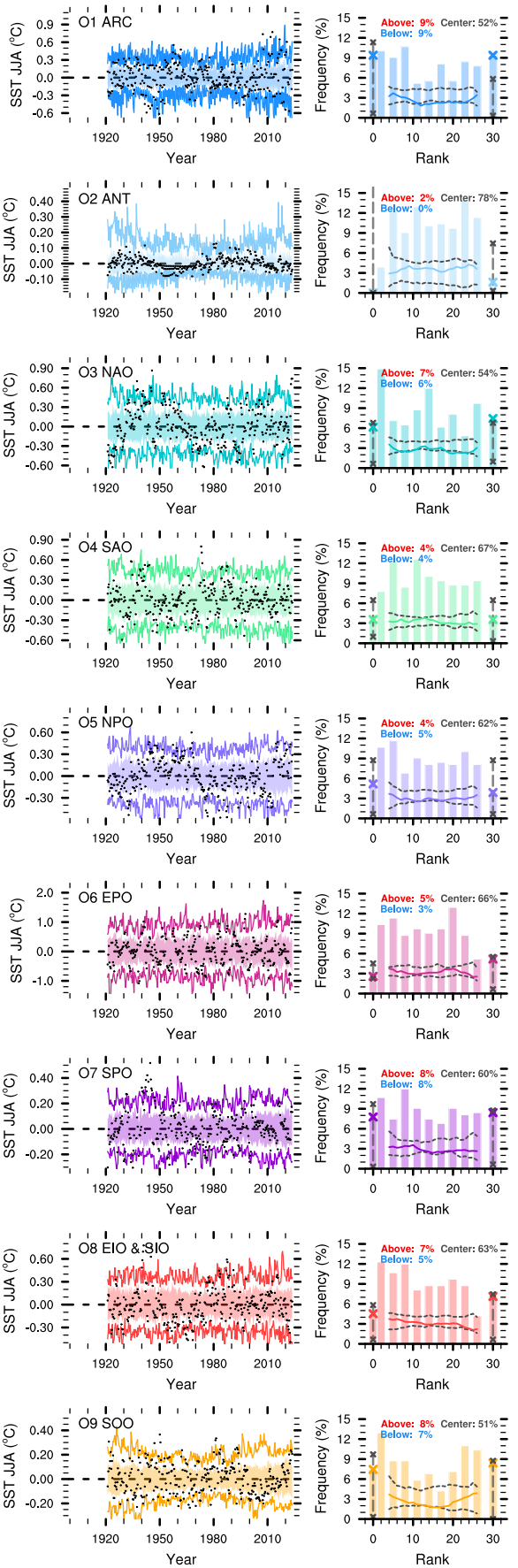
SST JJA Obs outside ensemble spread (%)



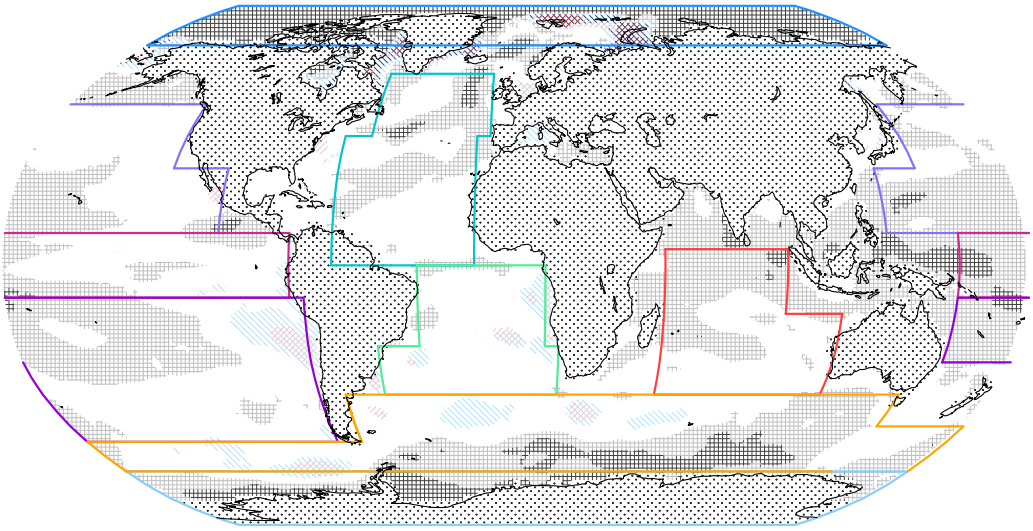
GFDL-SPEAR vs ERSSTv5 SST DJF



GFDL-SPEAR vs ERSSTv5 SST JJA



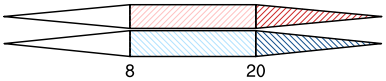
MIROC6 vs ERSSTv5 White Area = 49.7 %



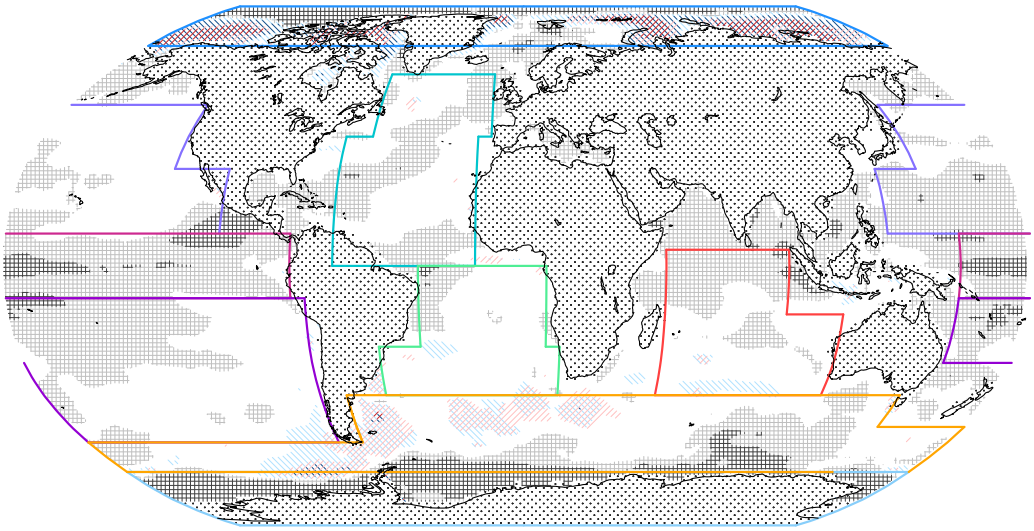
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



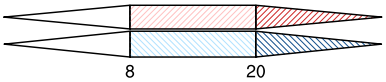
MIROC6 vs ERSSTv5 White Area = 46.4 %



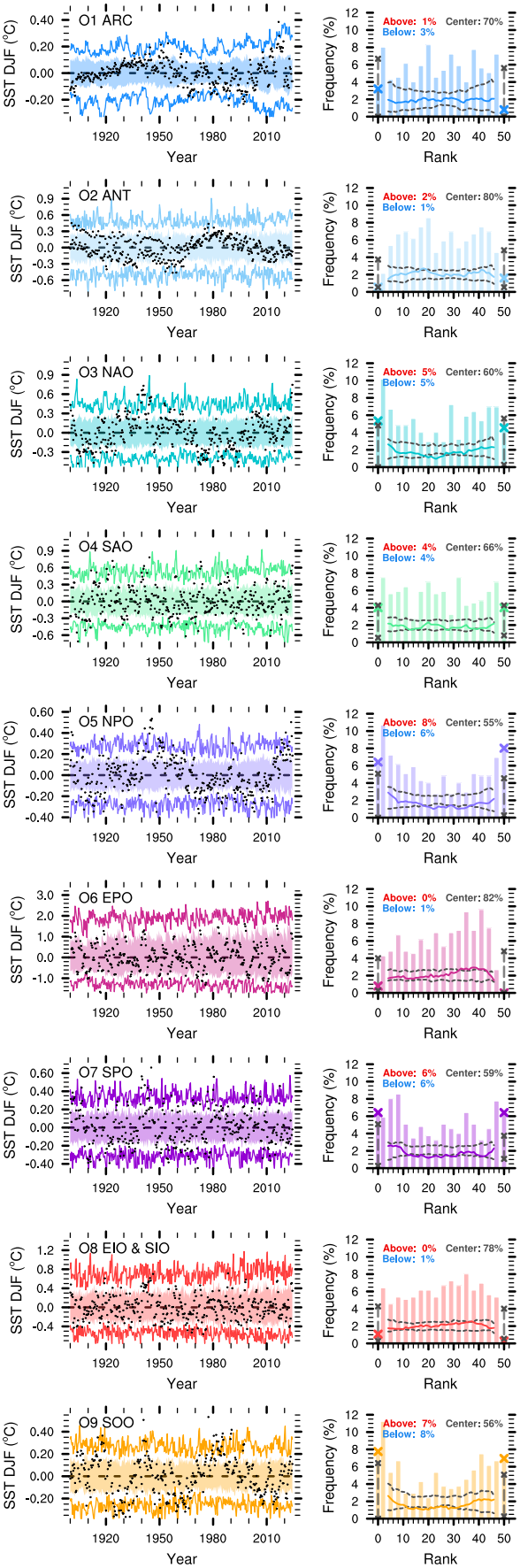
SST JJA Obs inside central 75th percentile (%)



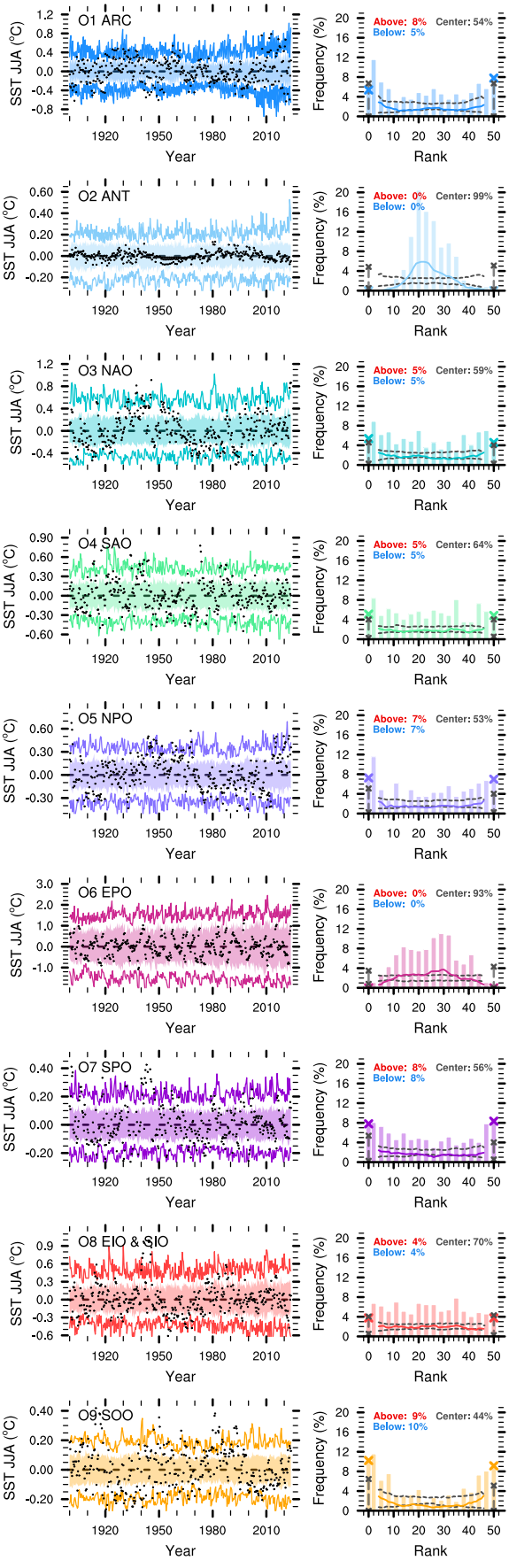
SST JJA Obs outside ensemble spread (%)



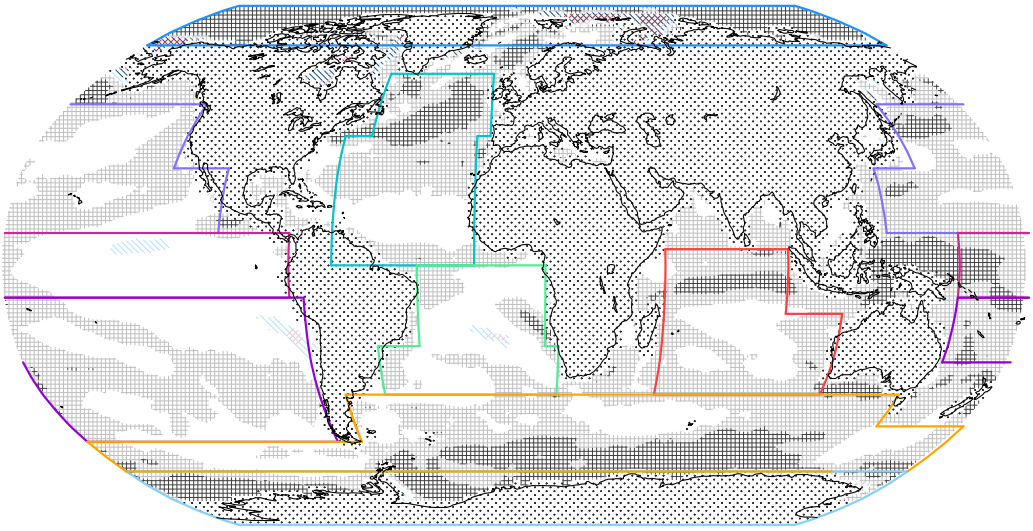
MIROC6 vs ERSSTv5 SST DJF



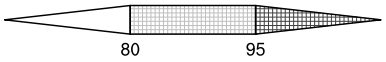
MIROC6 vs ERSSTv5 SST JJA



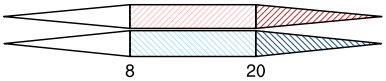
MPI-GE5 vs ERSSTv5 White Area = 37.6 %



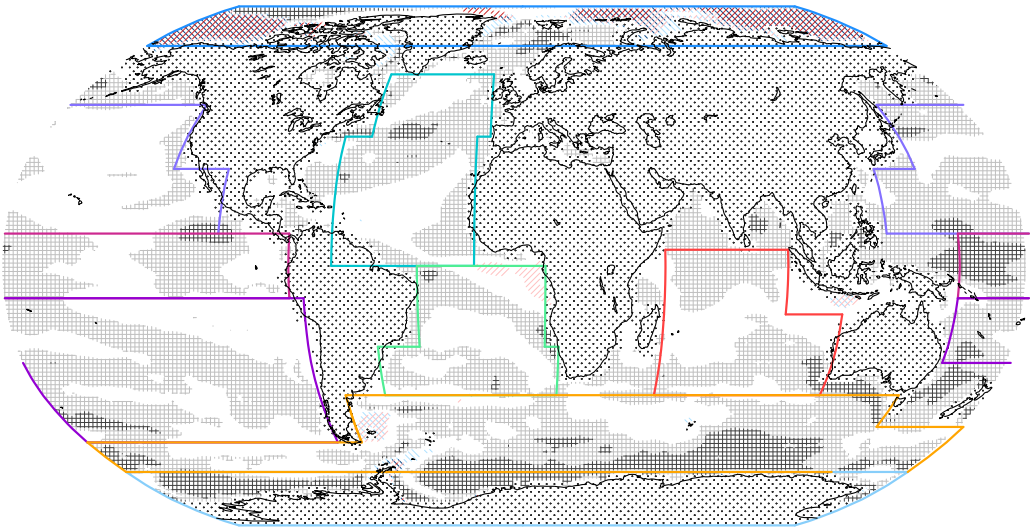
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



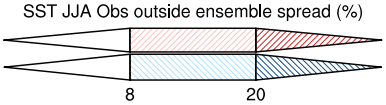
MPI-GE5 vs ERSSTv5 White Area = 40.2 %



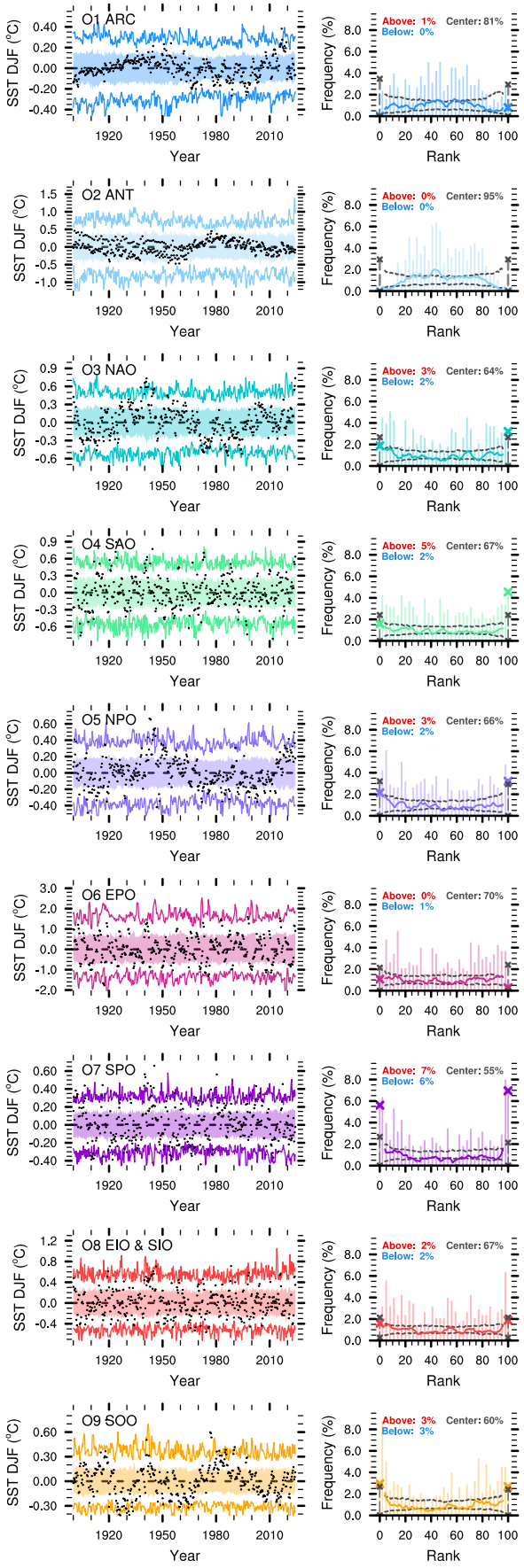
SST JJA Obs inside central 75th percentile (%)



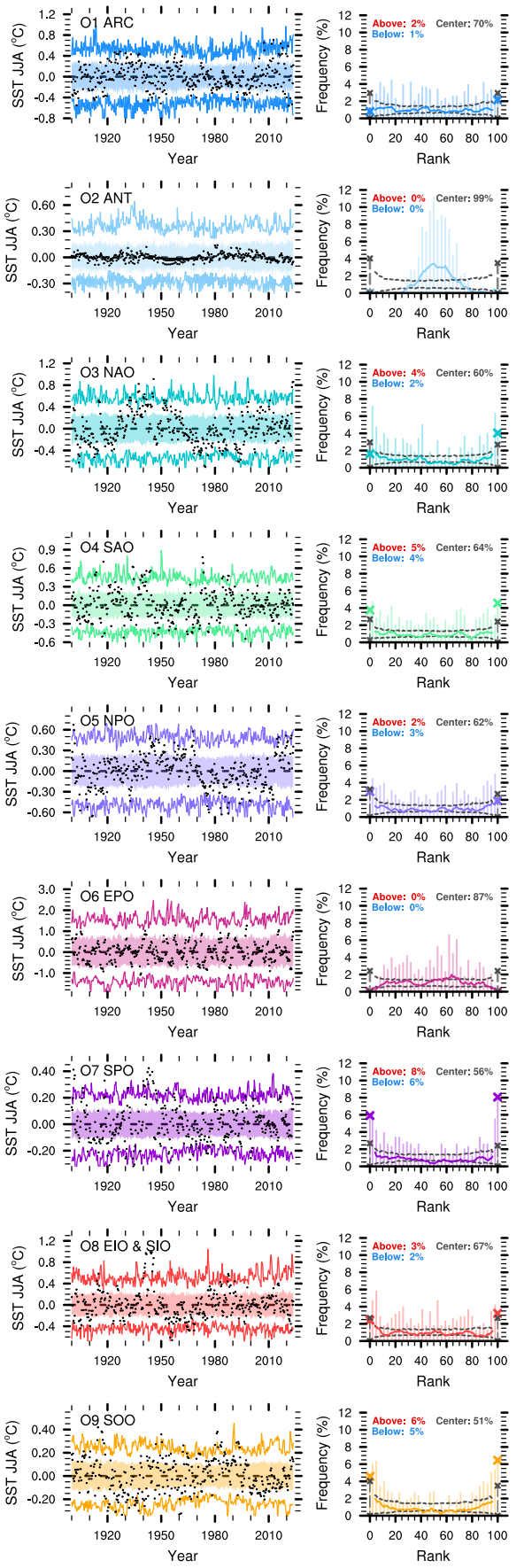
SST JJA Obs outside ensemble spread (%)



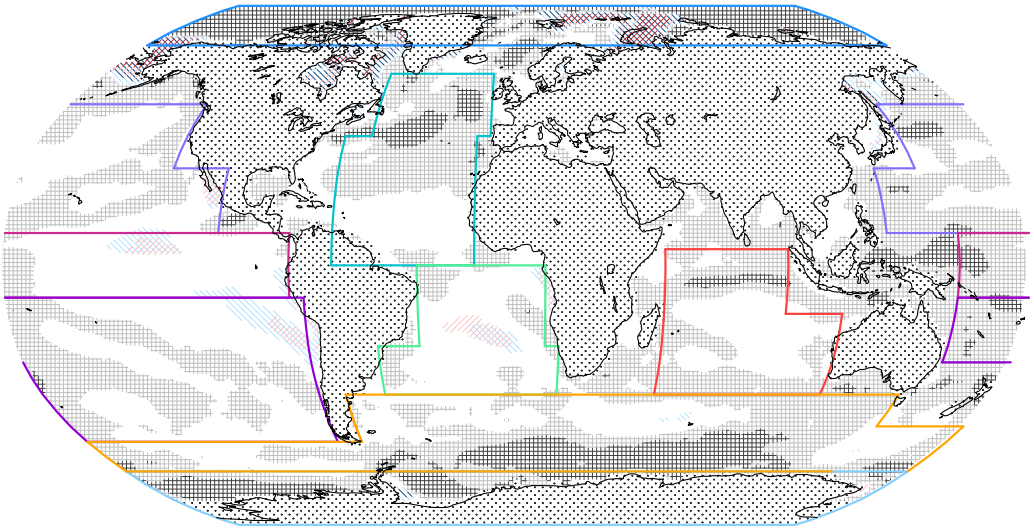
MPI-GE5 vs ERSSTv5 SST DJF



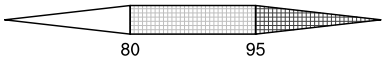
MPI-GE5 vs ERSSTv5 SST JJA



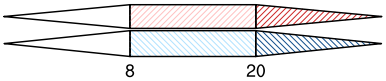
MPI-GE6 vs ERSSTv5 White Area = 41.9 %



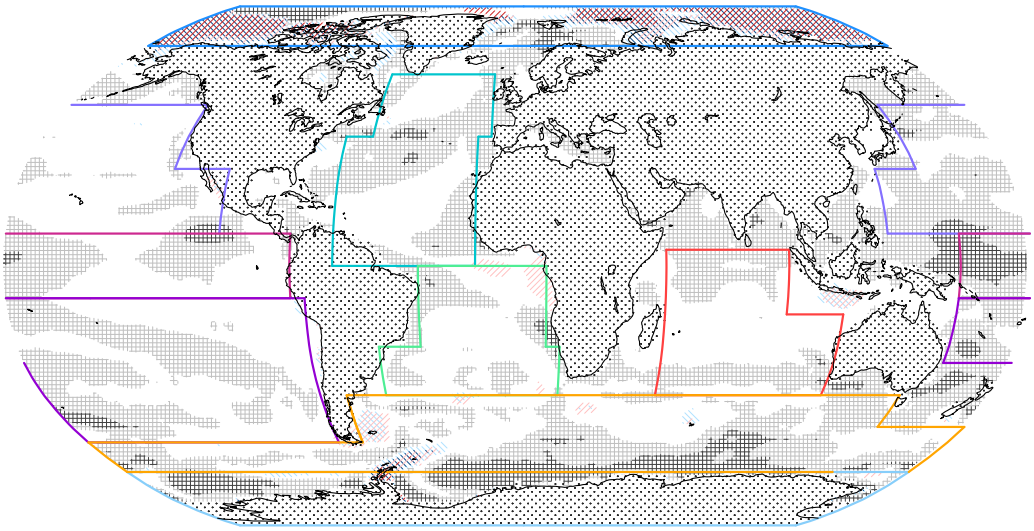
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



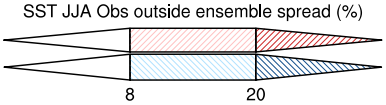
MPI-GE6 vs ERSSTv5 White Area = 45.9 %



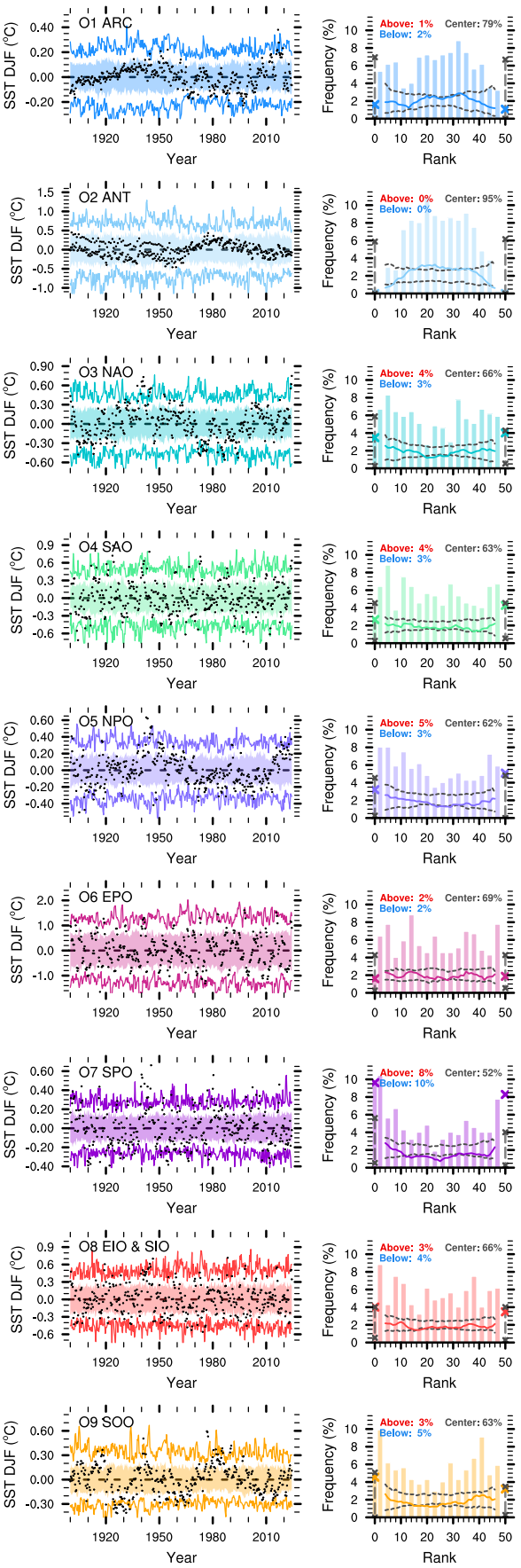
SST JJA Obs inside central 75th percentile (%)



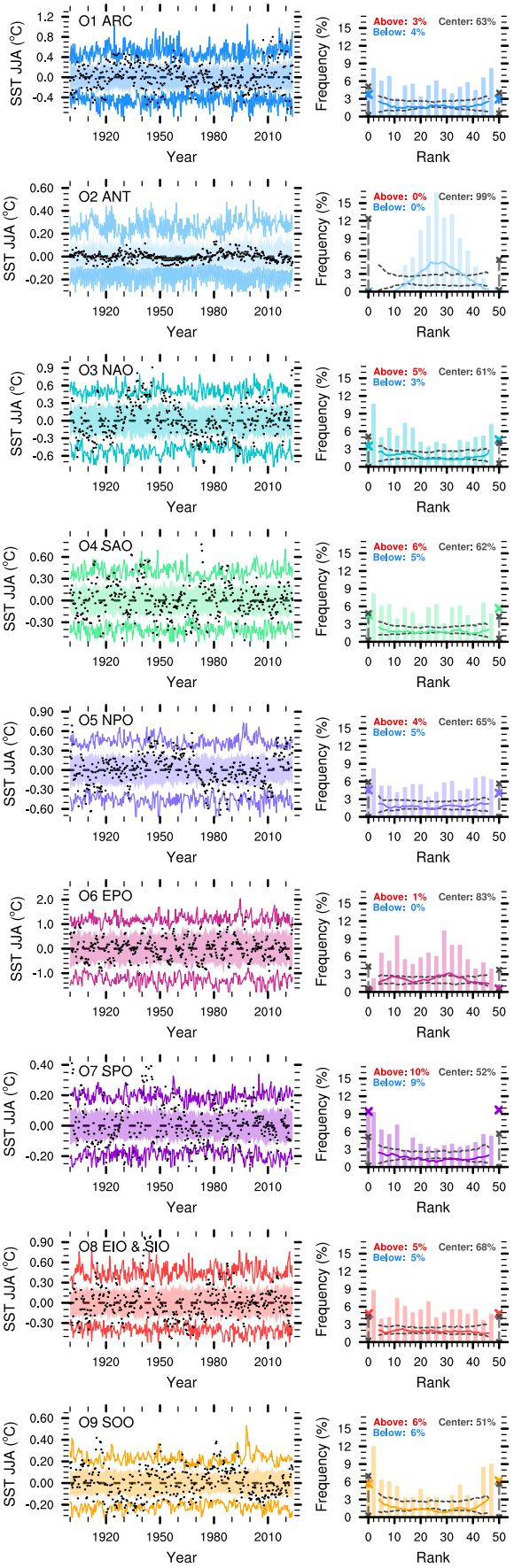
SST JJA Obs outside ensemble spread (%)



MPI-GE6 vs ERSSTv5 SST DJF



MPI-GE6 vs ERSSTv5 SST JJA



Non-Detrended Ocean Surface Temperatures

Rank-frequency variability evaluation framework for non-detrended sea surface temperature (SST) anomalies over ocean grid cells.

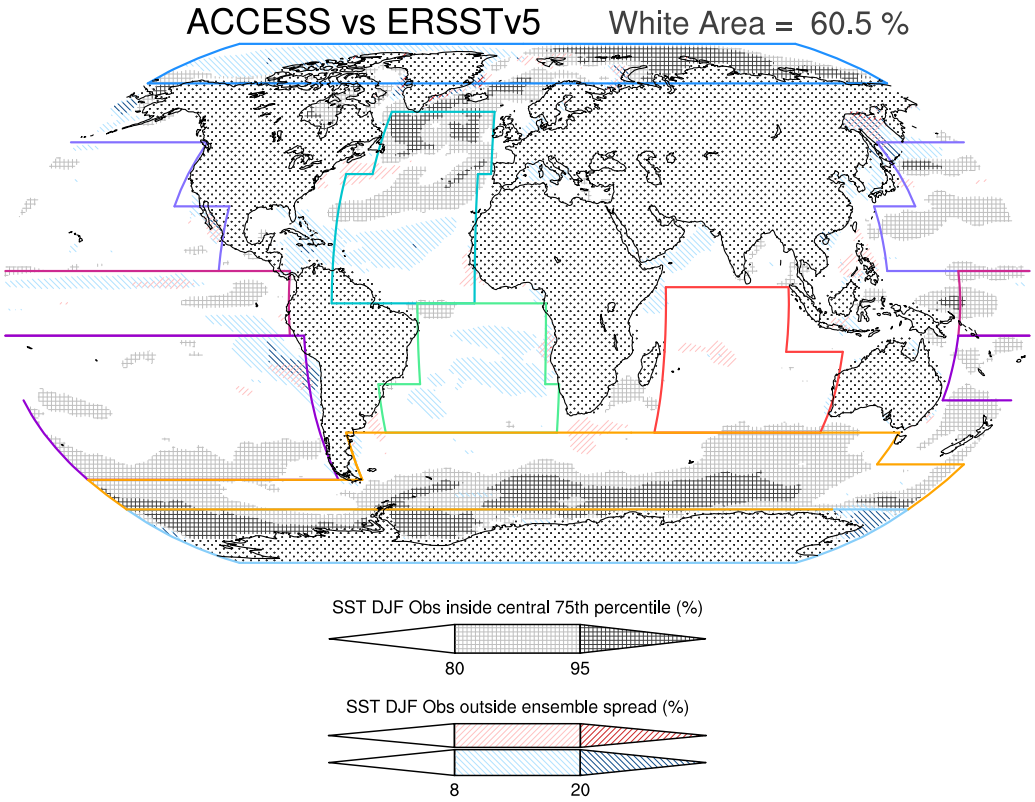
Maps show grid-cell evaluation of the simulated DJF and JJA monthly mean temperature anomalies for the 11 SMILEs included in this study against ERSSTv5 observations globally. Gray hatching represents where observations cluster within the 75th percentile bounds of the ensemble (12.5th to 87.5th percentiles) for more than 80% of months (light grey) or for more than 95% of months (dark grey). Red and blue shading represents where observations are larger than the ensemble maximum (red) or smaller than the ensemble minimum (blue), respectively, for more than 8% of the months (light red and blue) or for more than 20% of the months (dark red and blue). Dotted areas represent ocean areas or grid cells where observations are missing and are therefore excluded from this analysis. Colored boxes demark the boundaries of each ocean region assessed. The percentage of assessed grid-cells that present none of these biases in given at the top (white area).

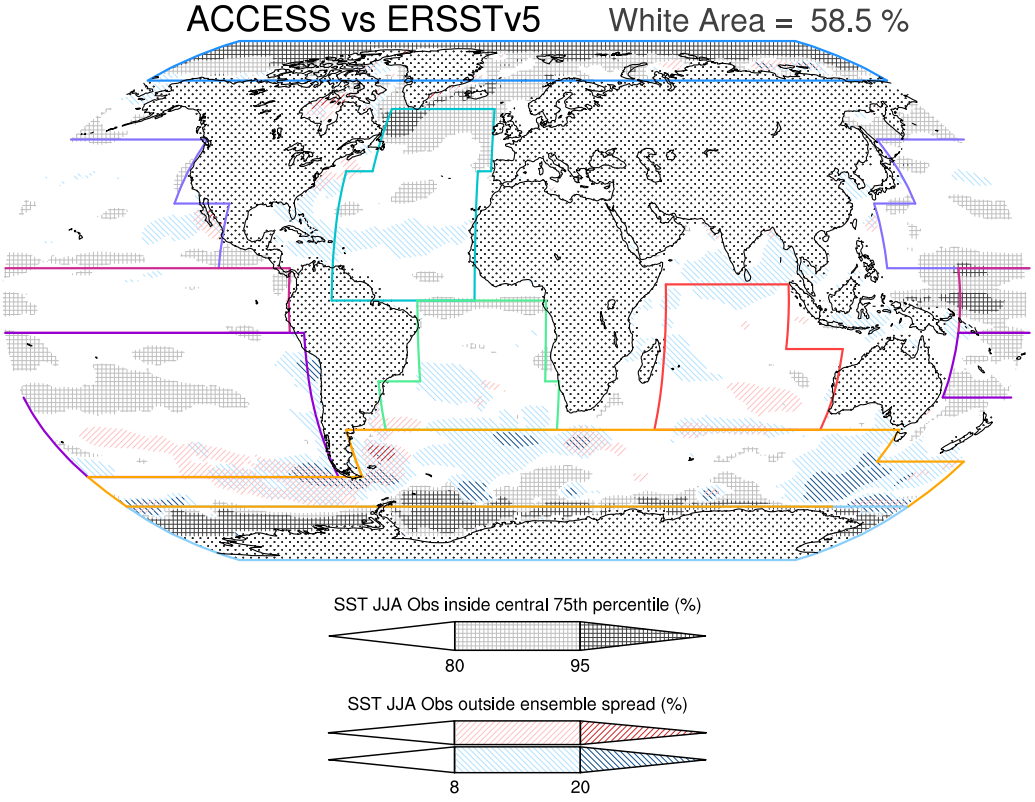
Time series and rank frequency histograms show spatially aggregated DJF and JJA SSTs for each ocean regions for all 11 SMILEs. Time series show the ensemble maximum and minimum (coloured lines) and central 75th percentile ensemble spread (shading) are shown against observations (black dots).

Rank histograms represent the frequency of each place that observations would take in a list of ensemble members ordered by ascending temperature anomaly values. Rank 0 indicates observations are below the minimum ensemble value, and rank n, with n the number of ensemble members, indicates that observations exceed the maximum ensemble value for that particular month. For a model that perfectly represents observations over an infinitely-long observational record, all ranks should occur with similar frequency and this histogram should be roughly flat.

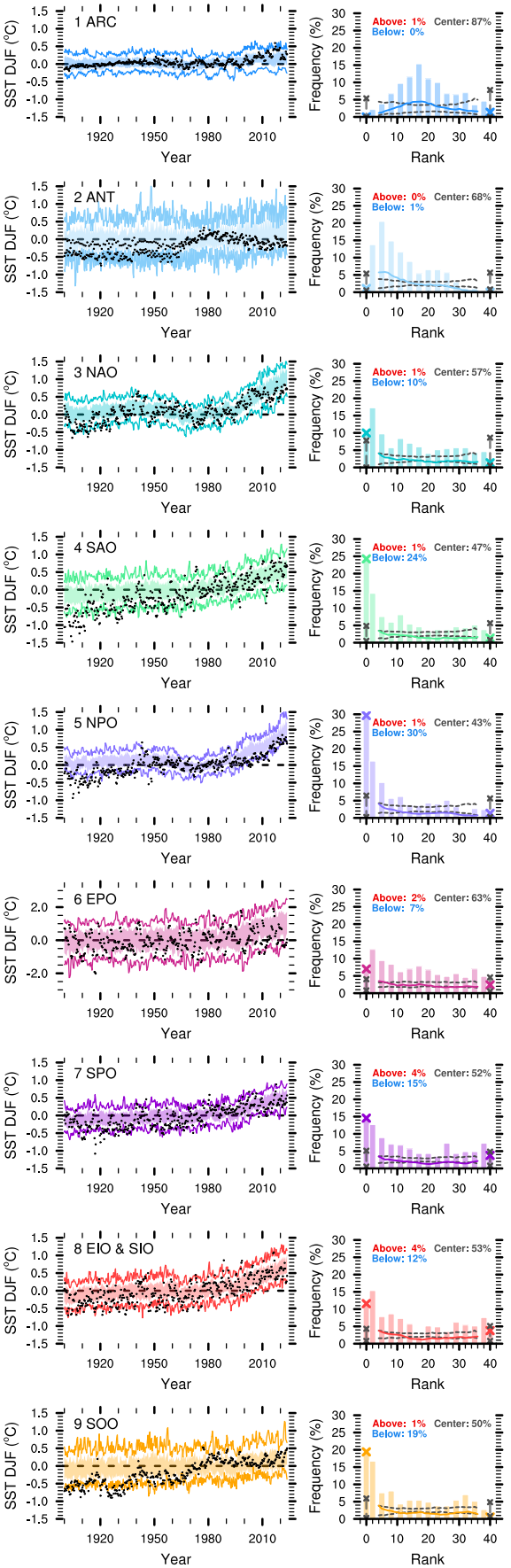
To illustrate how internal variability may affect rank frequencies given the non-infinite record length considered, we also include a perfect-model comparison, which shows the range of rank frequencies that each ensemble member would take if it were observations. If the rank exhibited by observations (colors) is within this perfect-model range (grey), the rank frequency evaluation shows an adequate model performance, and any deviations from a perfectly frank rank histogram can be assumed to be within deviations that could be caused by internal variability. Lines in the rank histogram illustrate the rank histogram's slope, as the mean rank frequency over a centered 6-bin window for observations (solid colored lines), and the 5-95th perc. perfect model range (gray dashed lines). Crosses represent the frequency of minimum (0) and maximum (number of members) ranks for observations (colors), and for the 5-95th perc. perfect model range (gray).

Percentages at the top of the rank histograms show how often monthly anomalies fall above or below ensemble limits (red and blue, respectively) and within the 75th perc. Range (grey), analogous to the criteria chosen for the map-based evaluation but for spatially aggregated values.

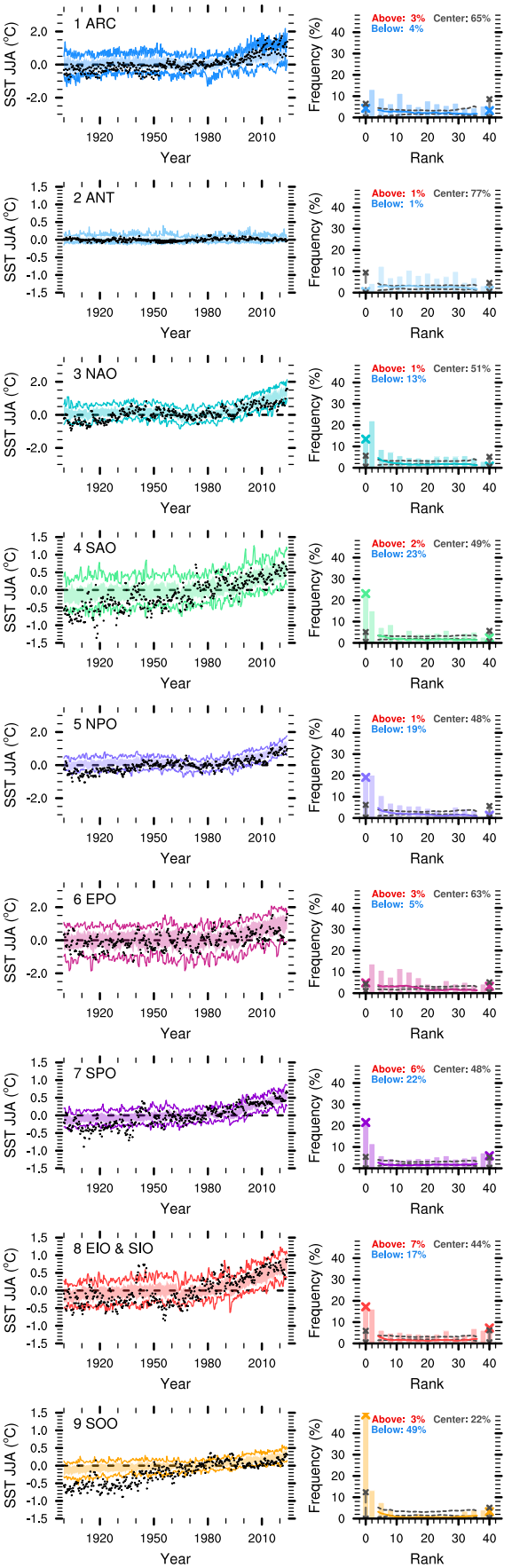

ACCESS vs ERSSTv5
White Area = 60.5 %
SST DJF Obs inside central 75th percentile (%)
80
95
SST DJF Obs outside ensemble spread (%)
8
20


ACCESS vs ERSSTv5
White Area = 58.5 %
SST JJA Obs inside central 75th percentile (%)
80
95
SST JJA Obs outside ensemble spread (%)
8
20

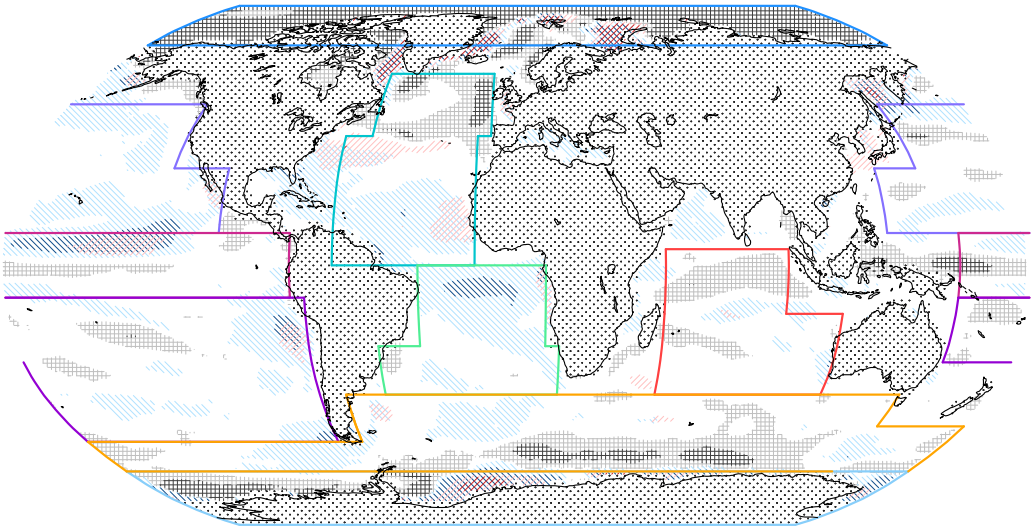
ACCESS vs ERSSTv5 SST DJF



ACCESS vs ERSSTv5 SST JJA



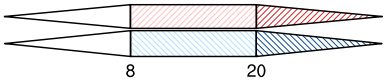
CanESM2 vs ERSSTv5 White Area = 51.8 %



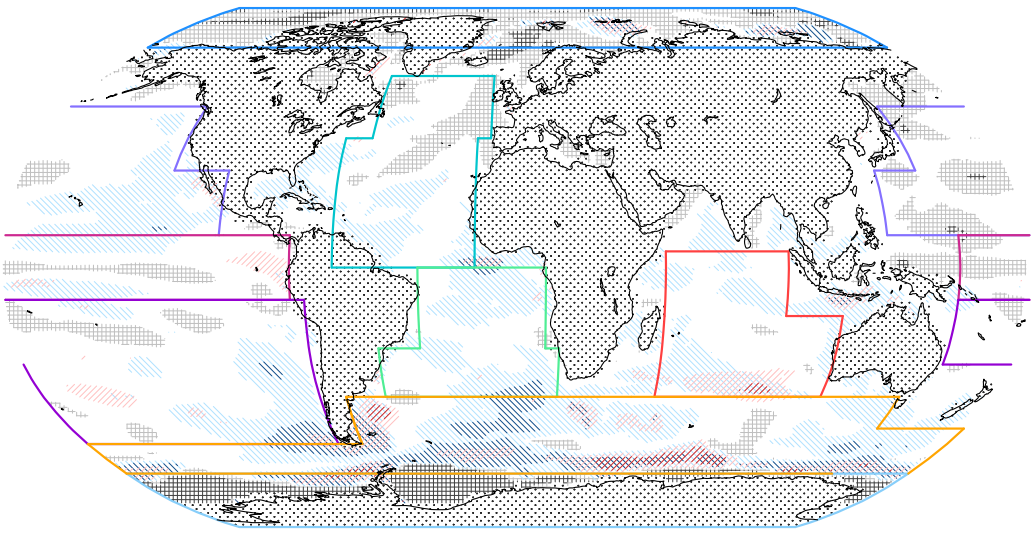
SST DJF Obs inside central 75th percentile (%)



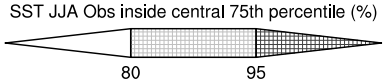
SST DJF Obs outside ensemble spread (%)



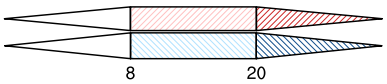
CanESM2 vs ERSSTv5 White Area = 48.5 %



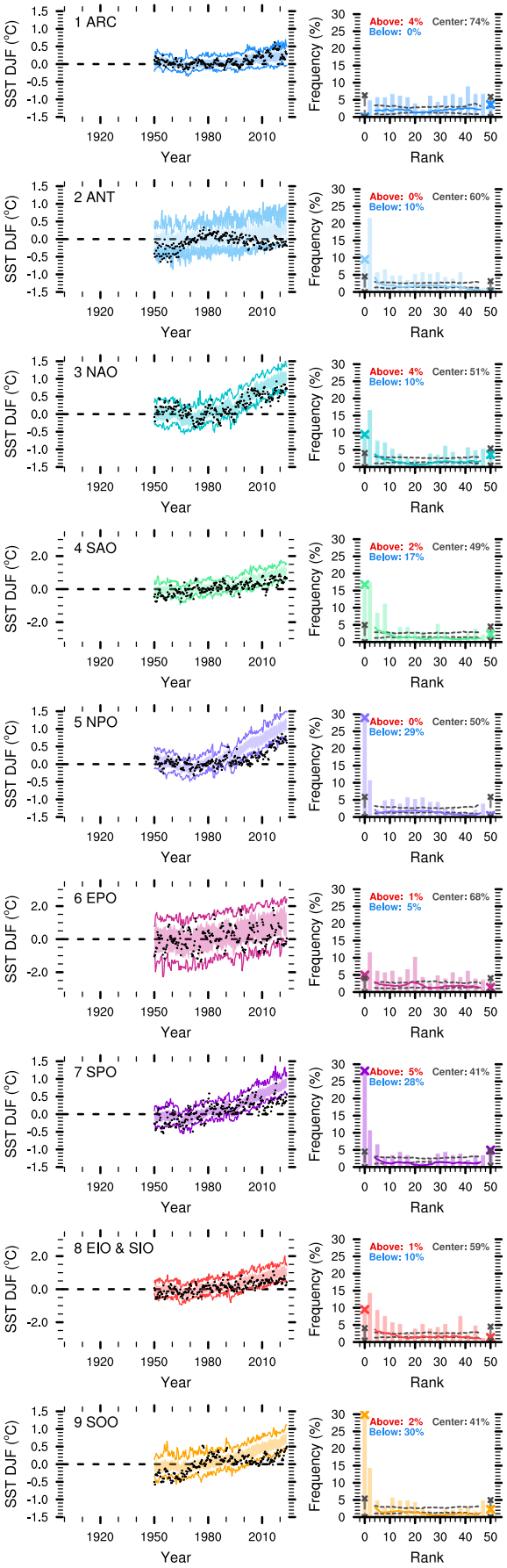
SST JJA Obs inside central 75th percentile (%)



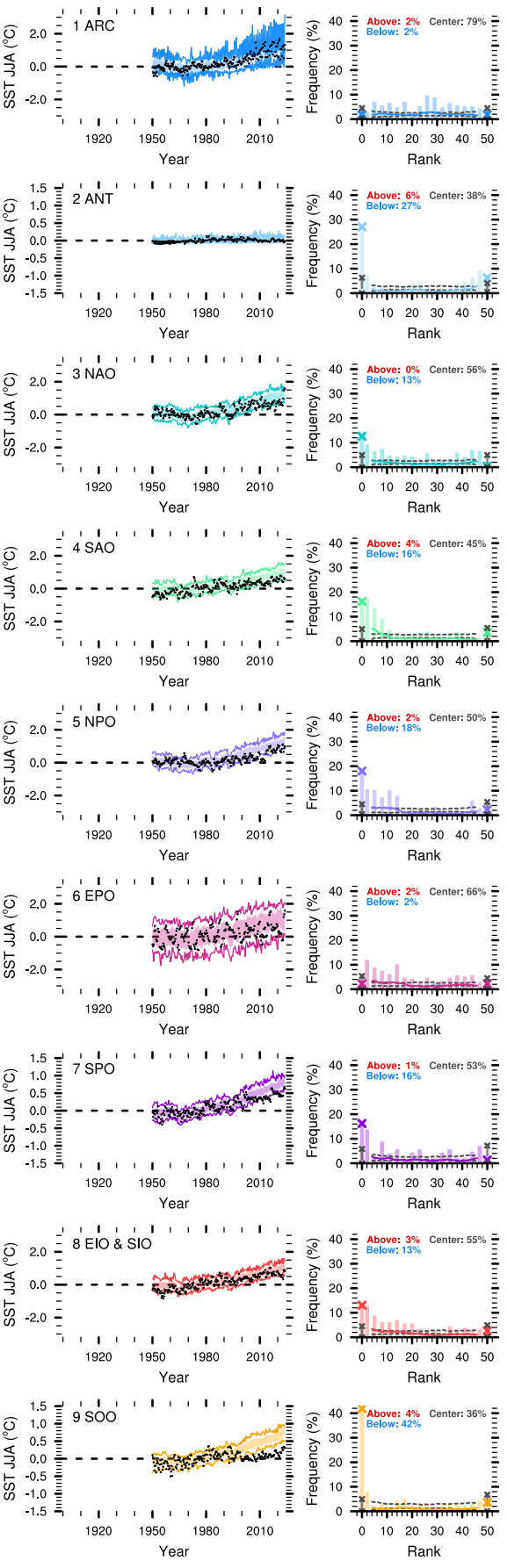
SST JJA Obs outside ensemble spread (%)



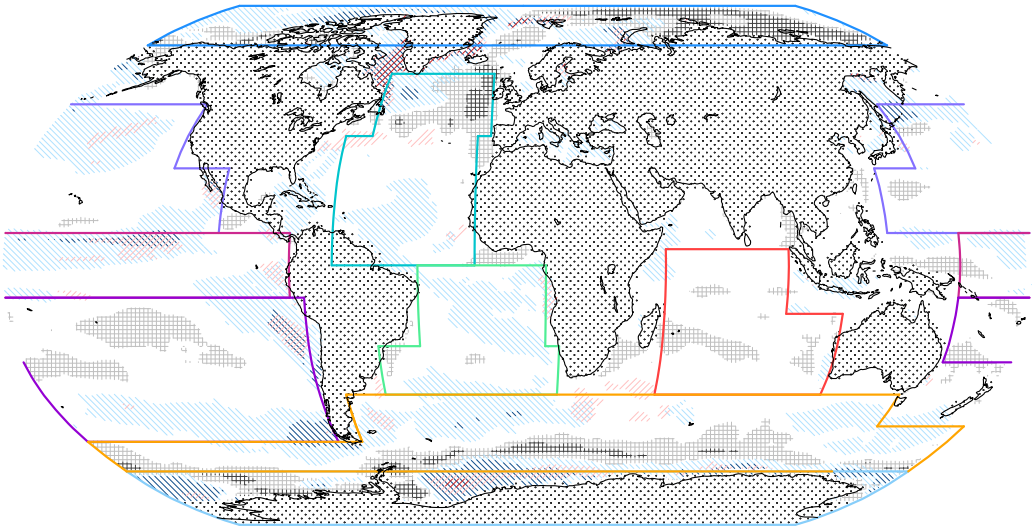
CanESM2 vs ERSSTv5 SST DJF



CanESM2 vs ERSSTv5 SST JJA



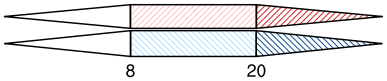
CanESM5 vs ERSSTv5 White Area = 55.0 %



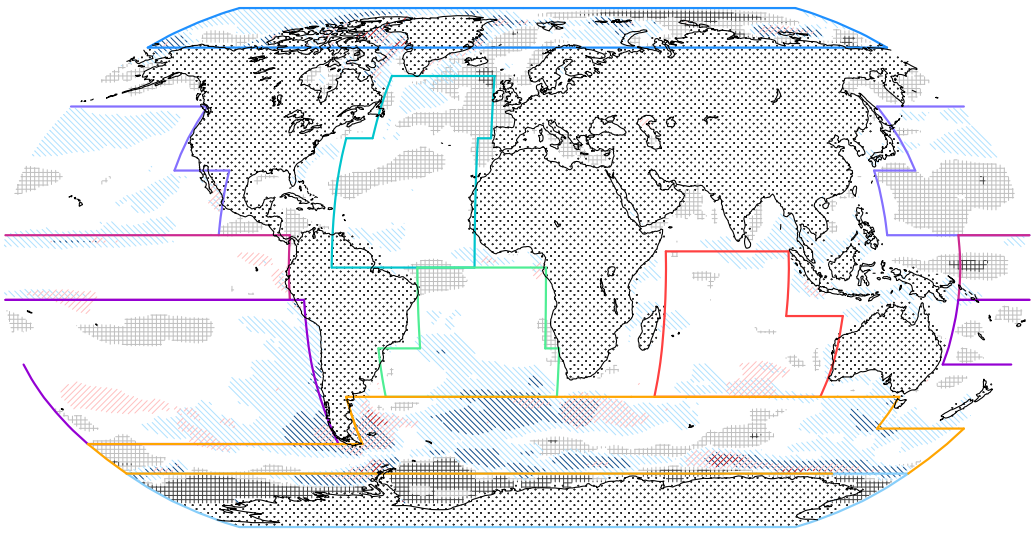
SST DJF Obs inside central 75th percentile (%)



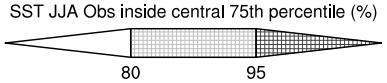
SST DJF Obs outside ensemble spread (%)



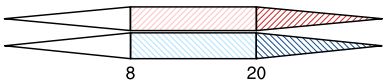
CanESM5 vs ERSSTv5 White Area = 55.9 %



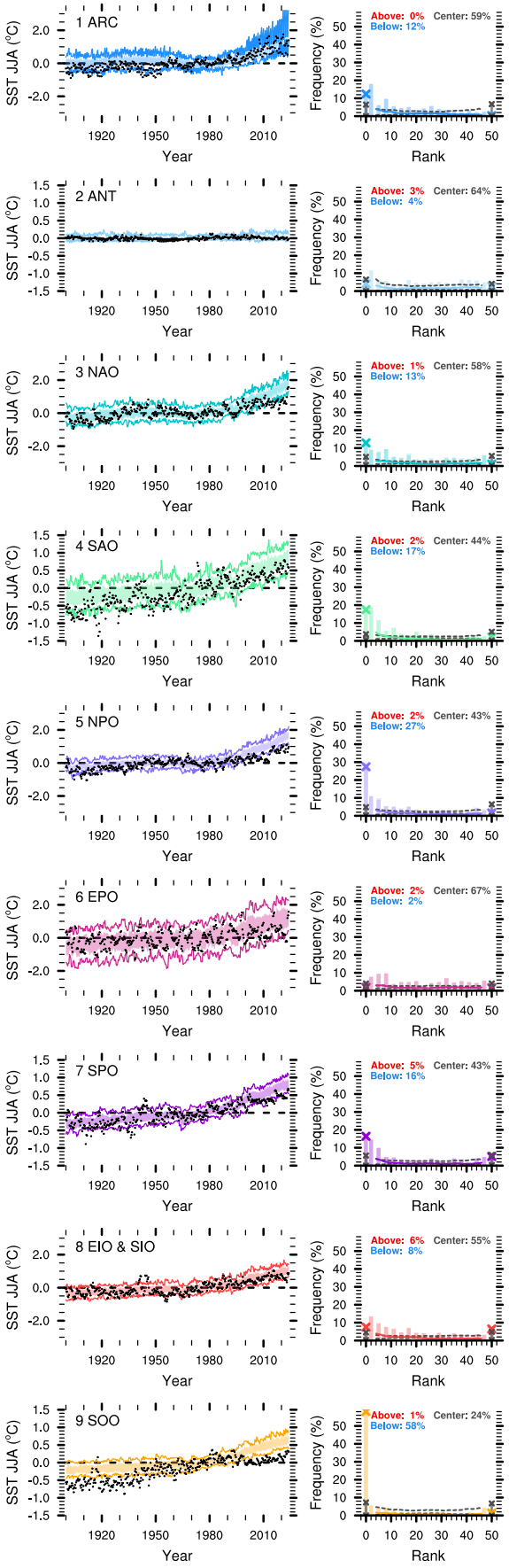
SST JJA Obs inside central 75th percentile (%)



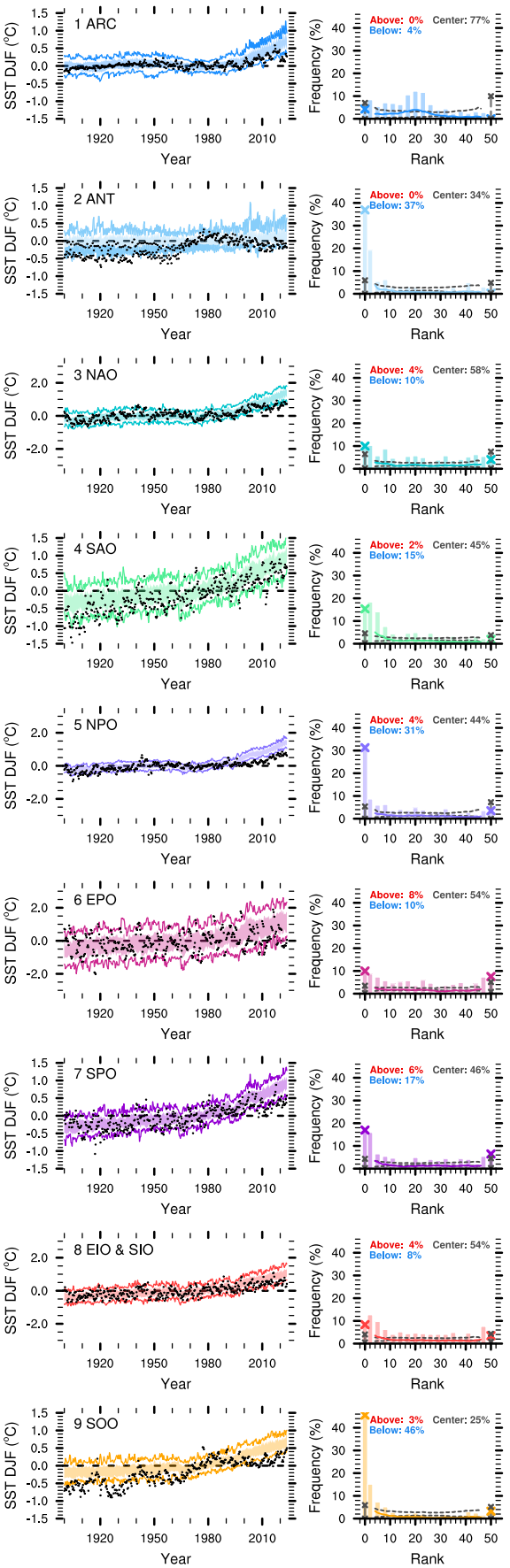
SST JJA Obs outside ensemble spread (%)



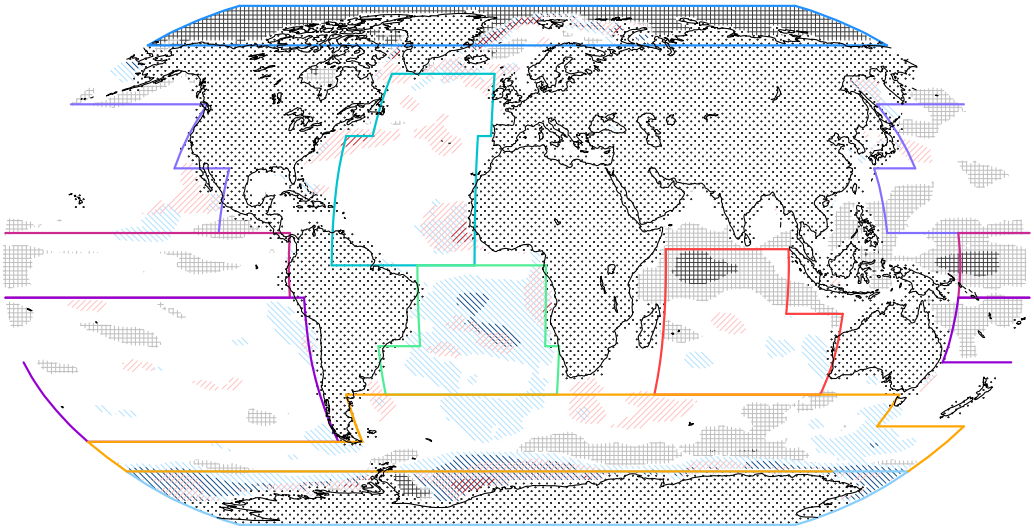
CanESM5 vs ERSSTv5 SST JJA



CanESM5 vs ERSSTv5 SST DJF



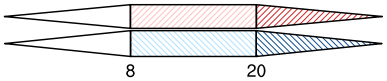
CESM-LE vs ERSSTv5 White Area = 61.8 %



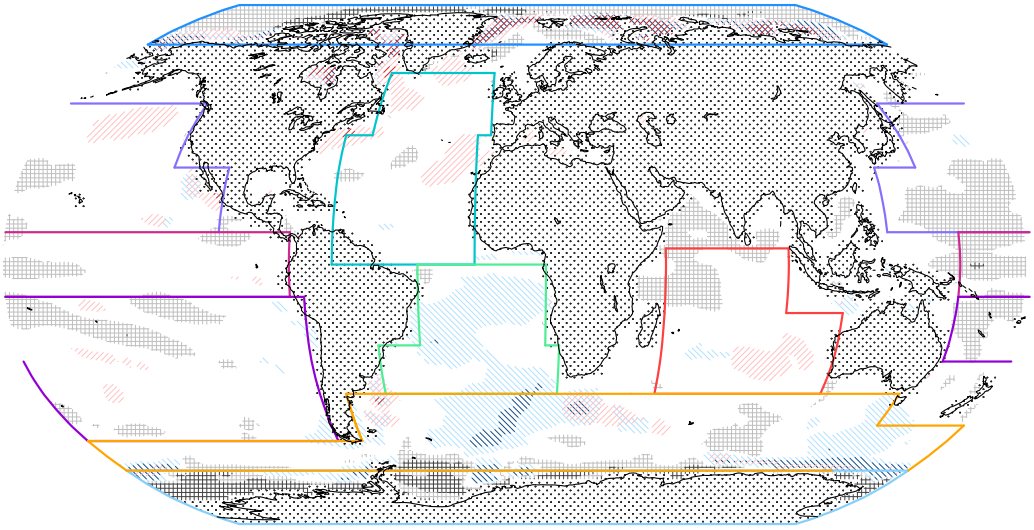
SST DJF Obs inside central 75th percentile (%)



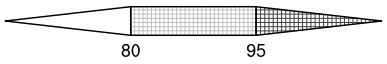
SST DJF Obs outside ensemble spread (%)



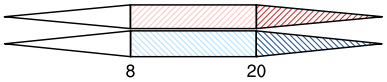
CESM-LE vs ERSSTv5 White Area = 64.7 %



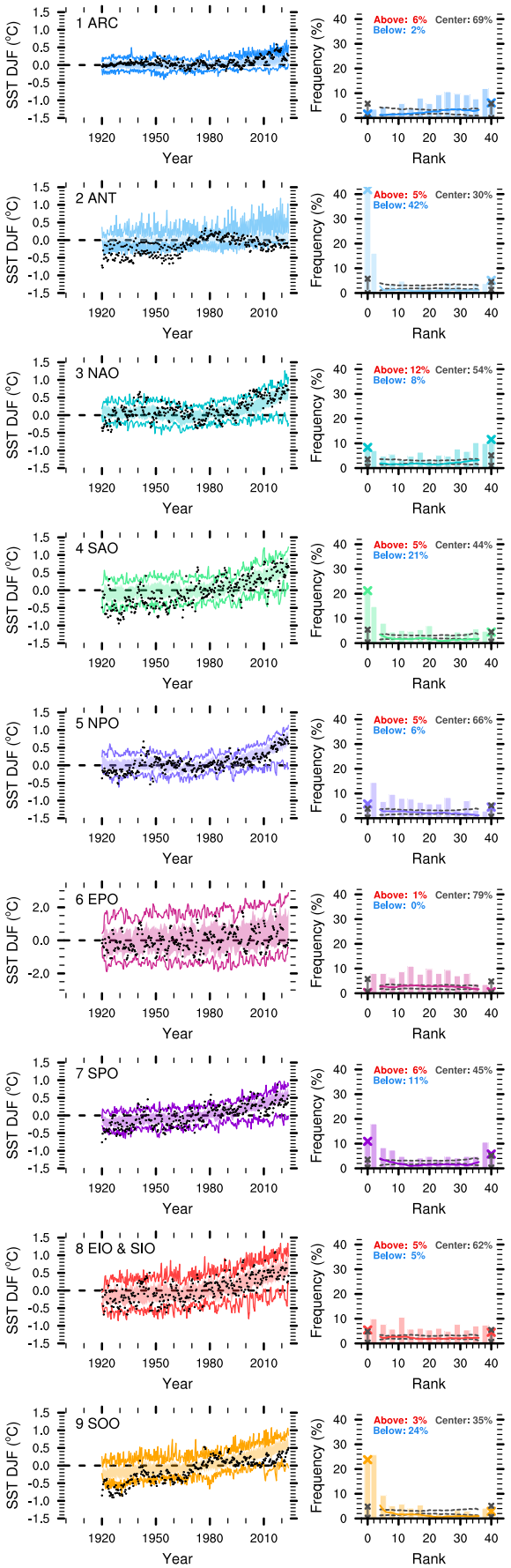
SST JJA Obs inside central 75th percentile (%)



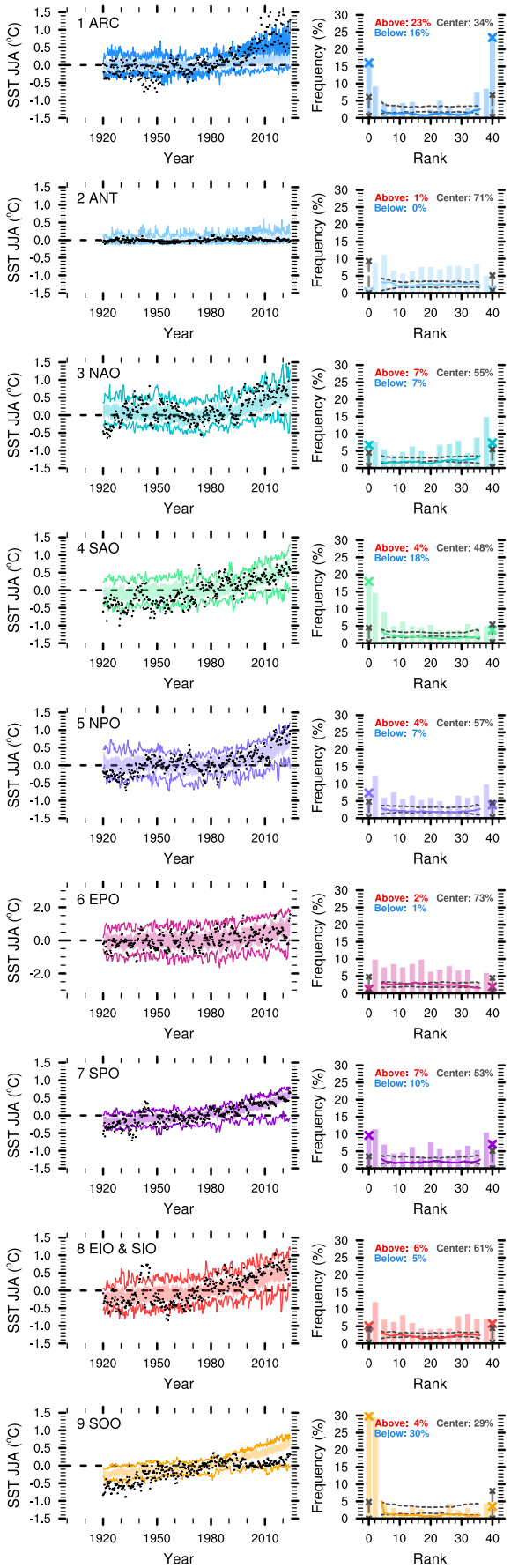
SST JJA Obs outside ensemble spread (%)



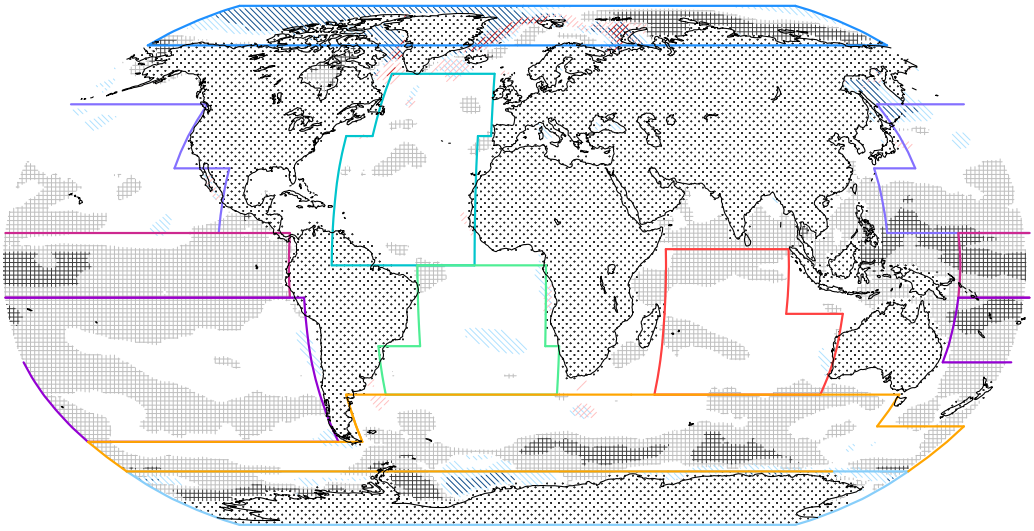
CESM-LE vs ERSSTv5 SST DJF



CESM-LE vs ERSSTv5 SST JJA



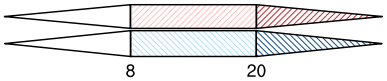
CESM2-LE vs ERSSTv5 White Area = 52.8 %



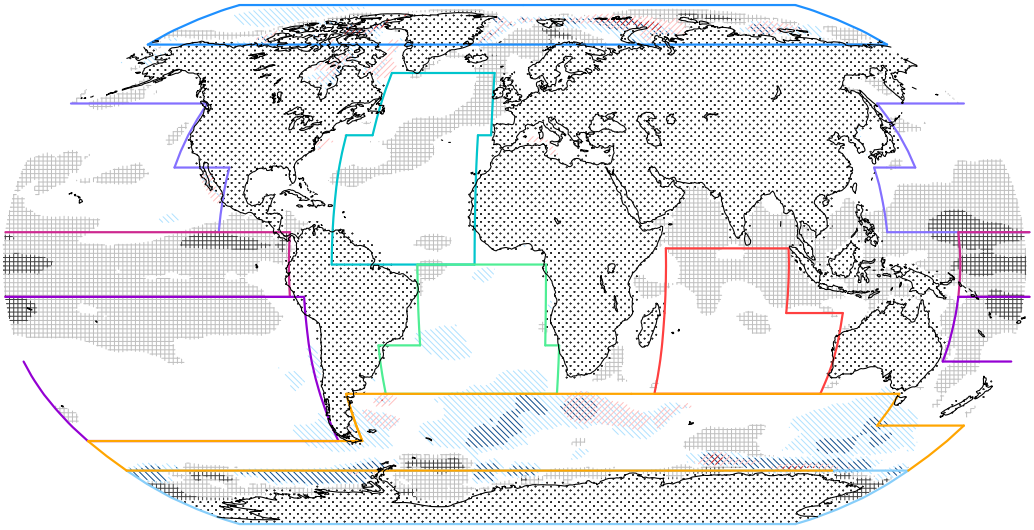
SST DJF Obs inside central 75th percentile (%)



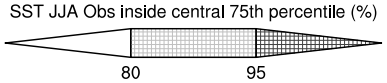
SST DJF Obs outside ensemble spread (%)



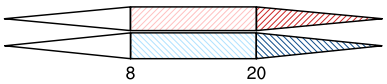
CESM2-LE vs ERSSTv5 White Area = 56.3 %



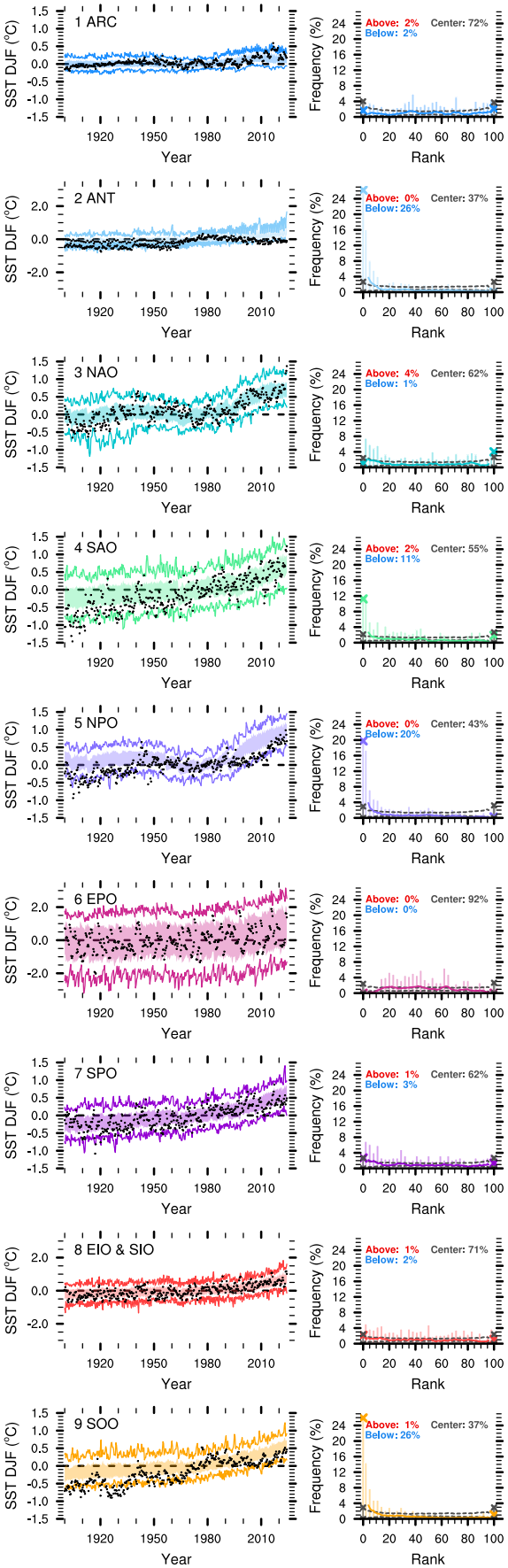
SST JJA Obs inside central 75th percentile (%)



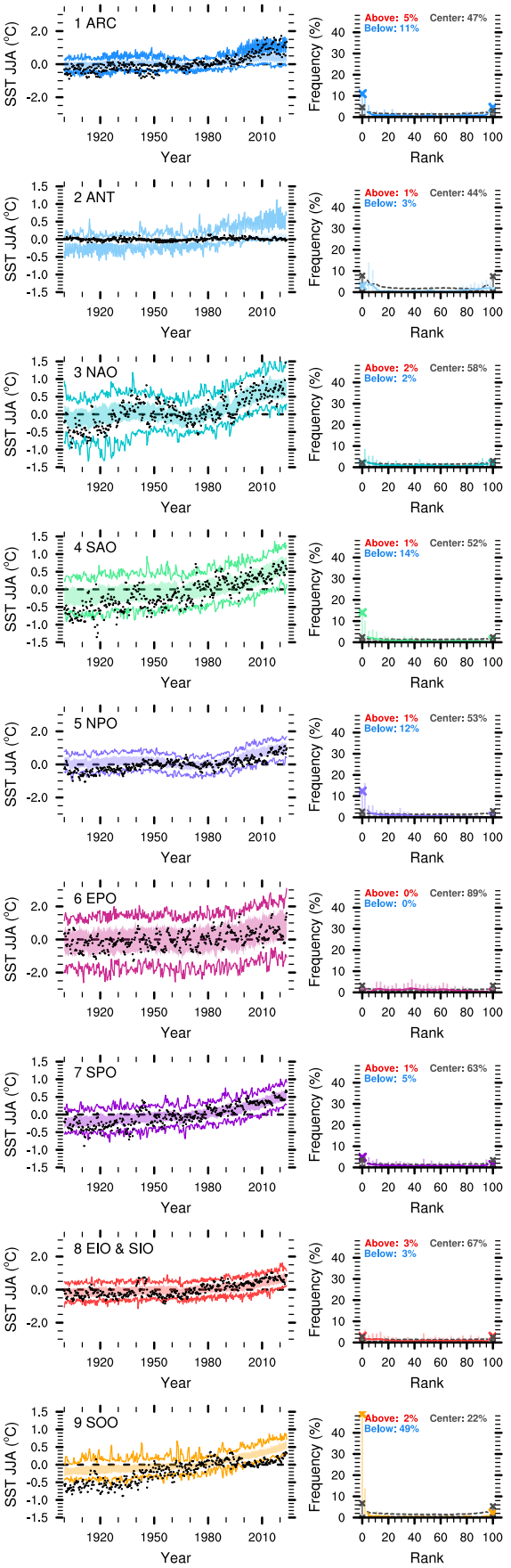
SST JJA Obs outside ensemble spread (%)



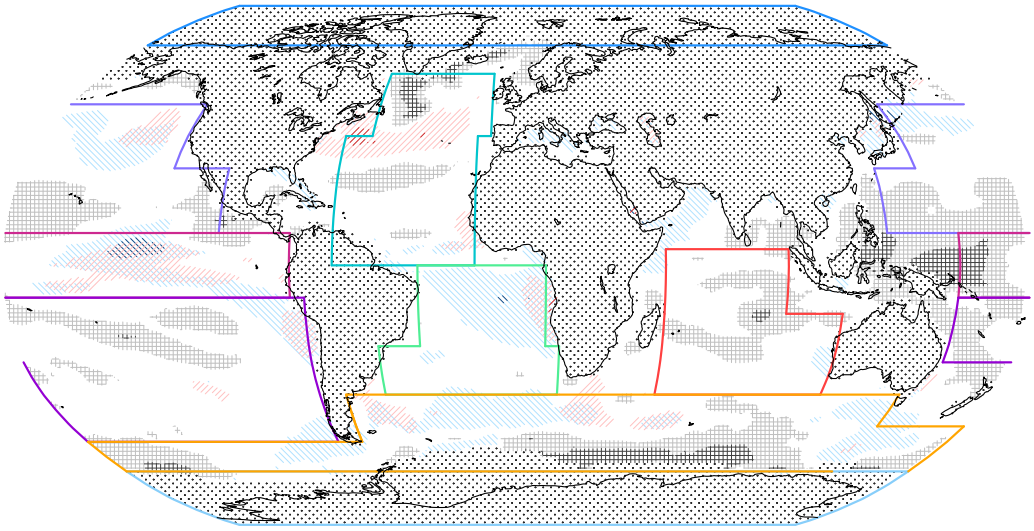
CESM2-LE vs ERSSTv5 SST DJF



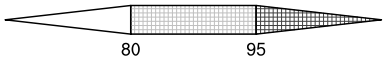
CESM2-LE vs ERSSTv5 SST JJA



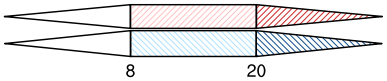
CSIRO-Mk360 vs ERSSTv5 White Area = 54.9 %



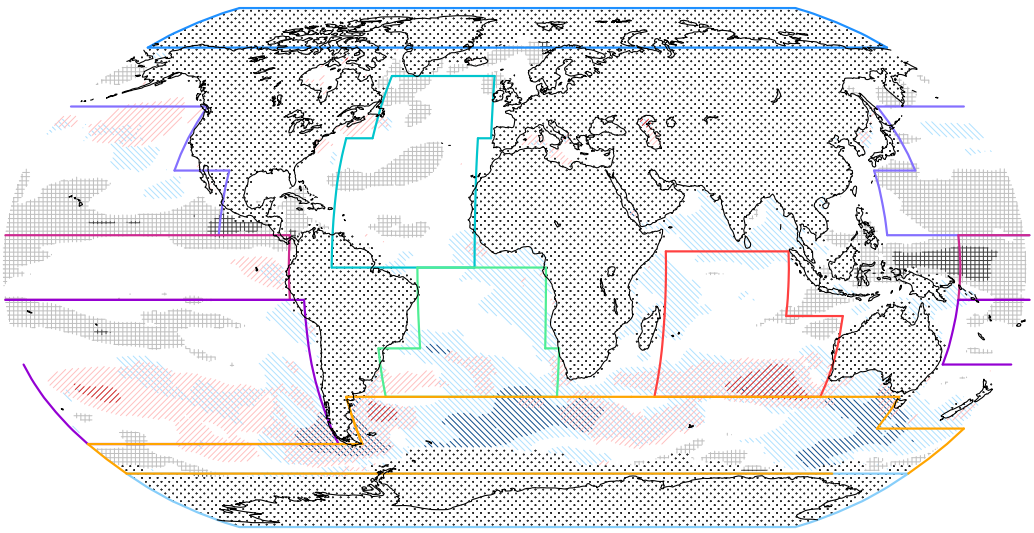
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



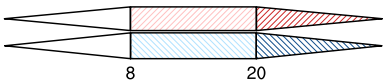
CSIRO-Mk360 vs ERSSTv5 White Area = 47.3 %



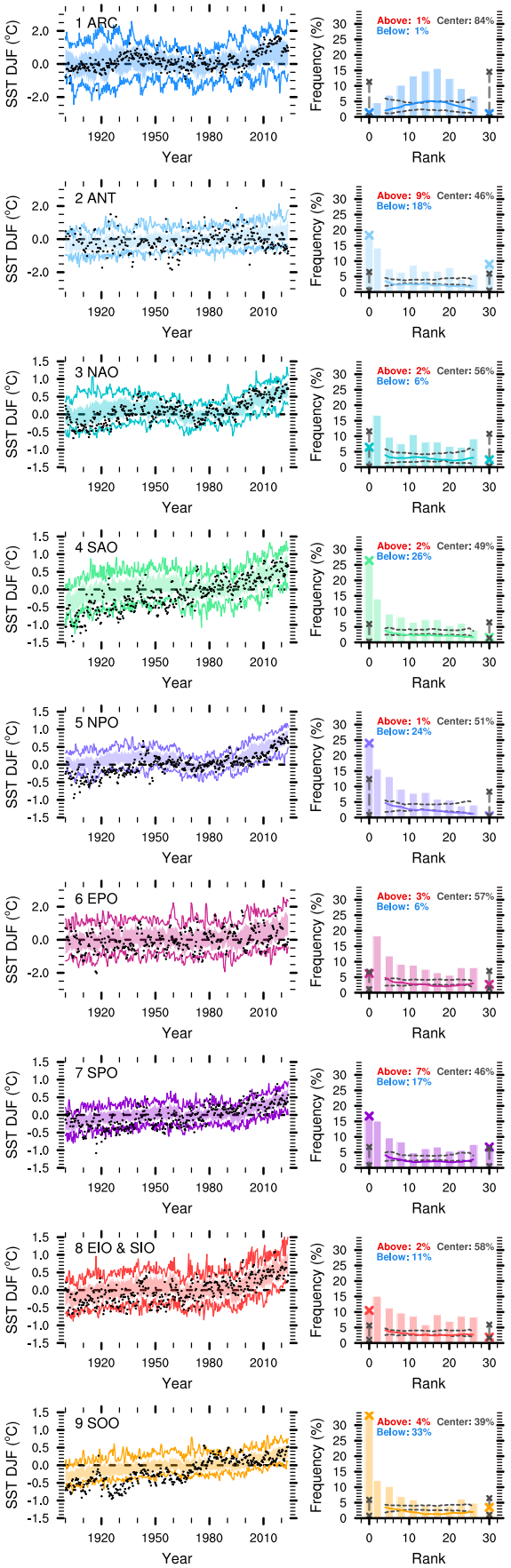
SST JJA Obs inside central 75th percentile (%)



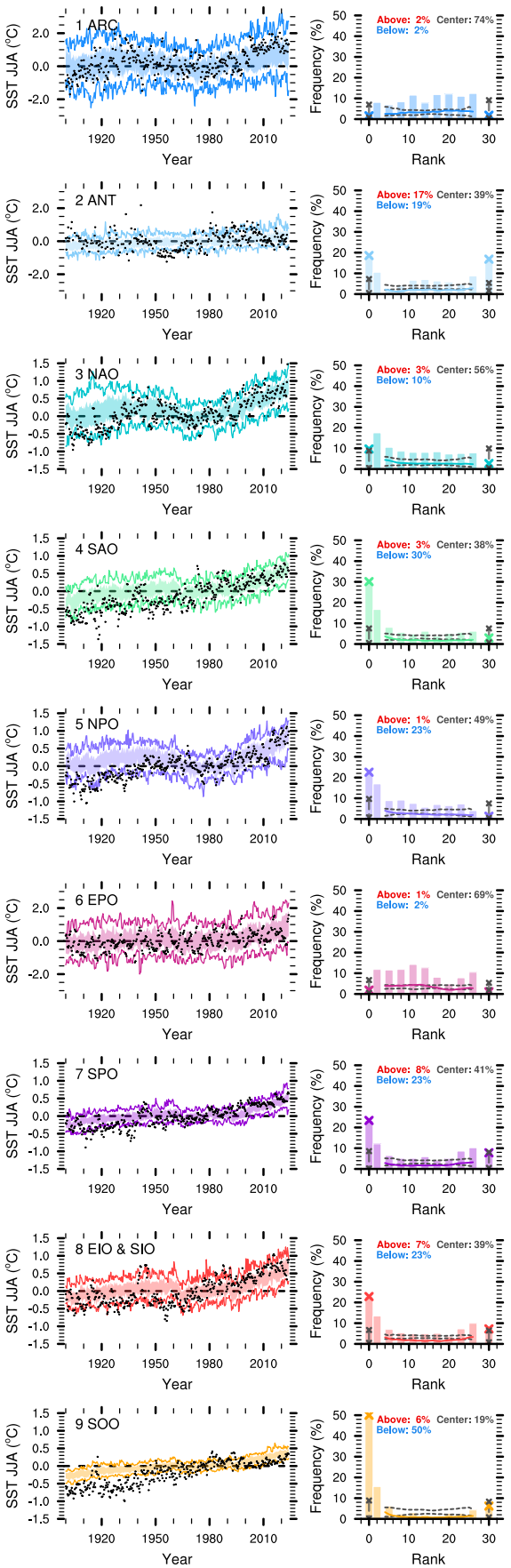
SST JJA Obs outside ensemble spread (%)



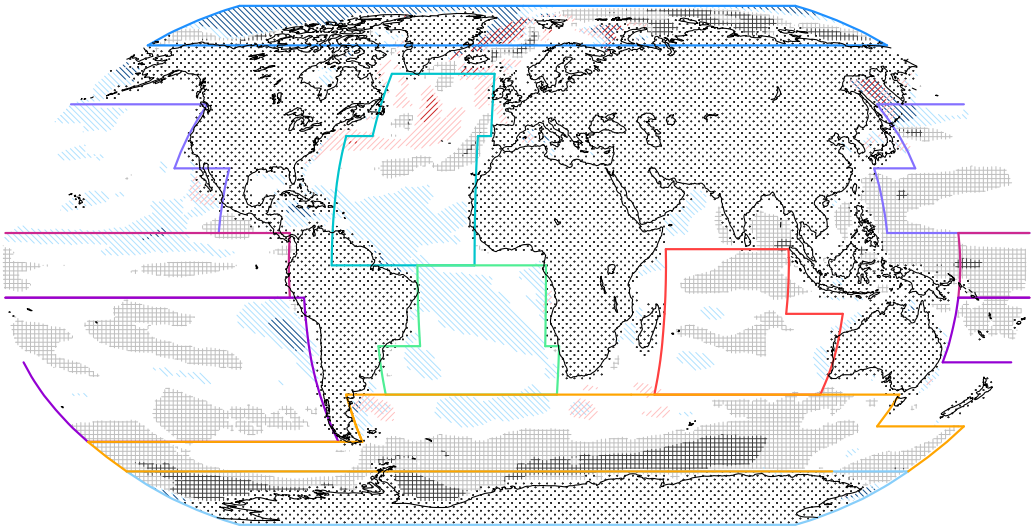
CSIRO-Mk360 vs ERSSTv5 SST DJF



CSIRO-Mk360 vs ERSSTv5 SST JJA



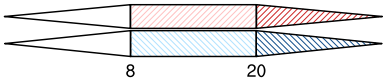
GFDL-ESM2M vs ERSSTv5 White Area = 53.6 %



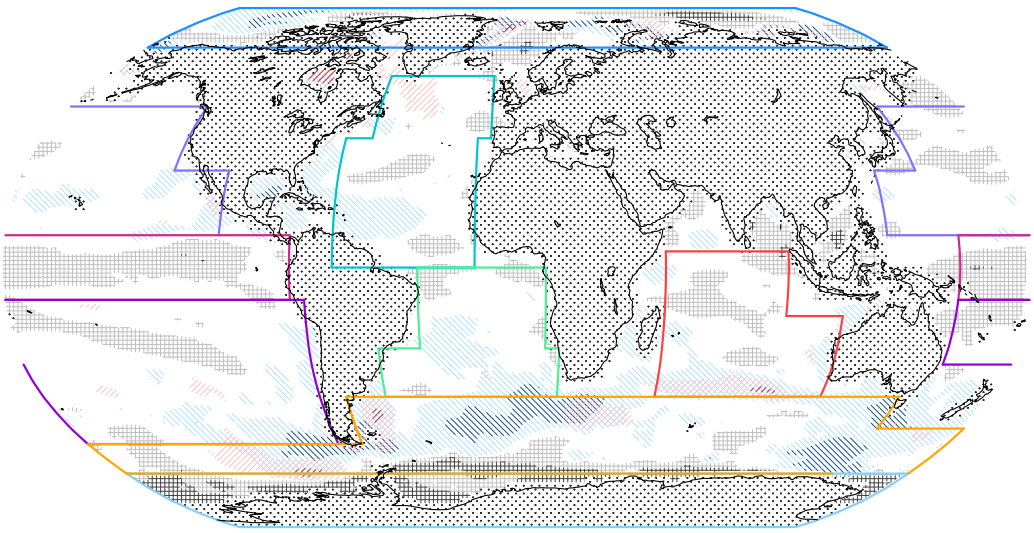
SST DJF Obs inside central 75th percentile (%)



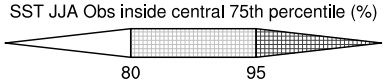
SST DJF Obs outside ensemble spread (%)



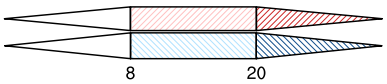
GFDL-ESM2M vs ERSSTv5 White Area = 54.0 %



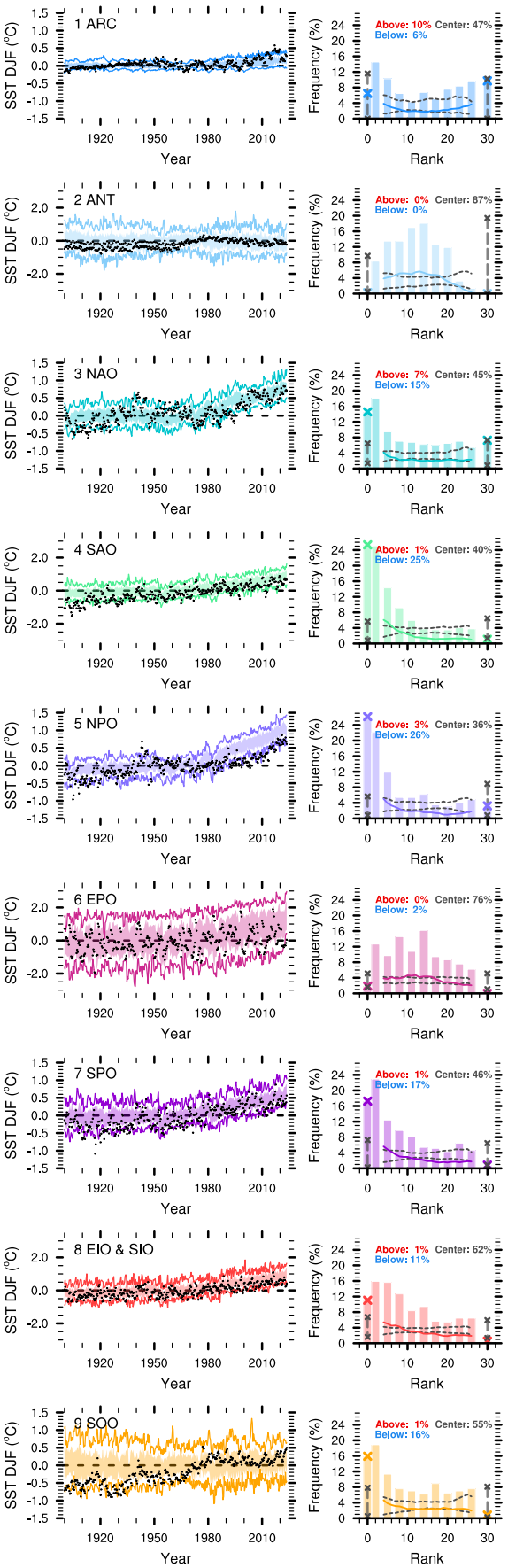
SST JJA Obs inside central 75th percentile (%)



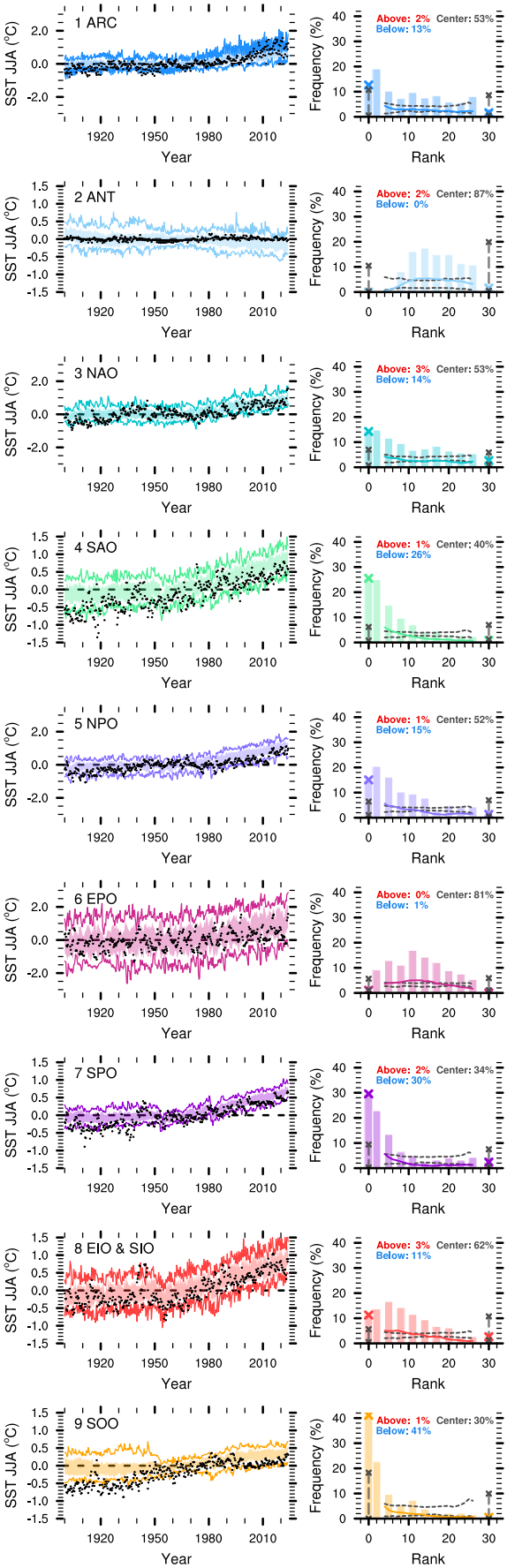
SST JJA Obs outside ensemble spread (%)



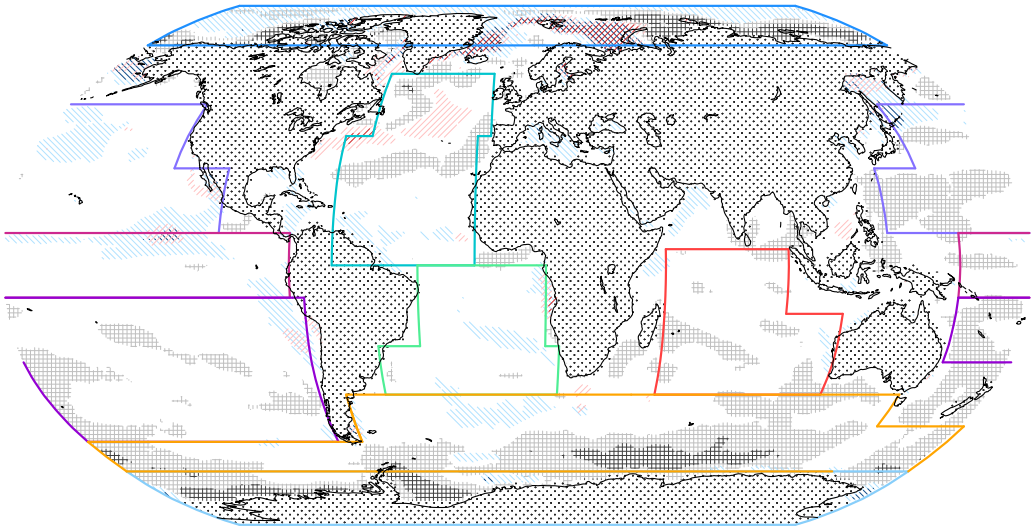
GFDL-ESM2M vs ERSSTv5 SST DJF



GFDL-ESM2M vs ERSSTv5 SST JJA



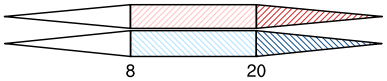
GFDL-SPEAR vs ERSSTv5 White Area = 62.7 %



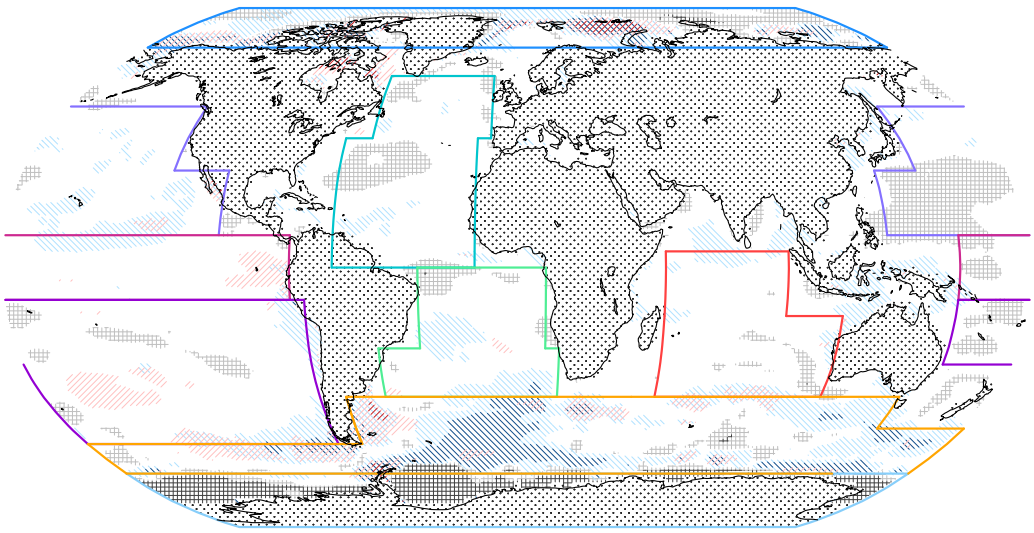
SST DJF Obs inside central 75th percentile (%)



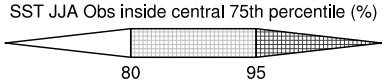
SST DJF Obs outside ensemble spread (%)



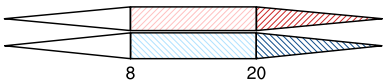
GFDL-SPEAR vs ERSSTv5 White Area = 62.5 %



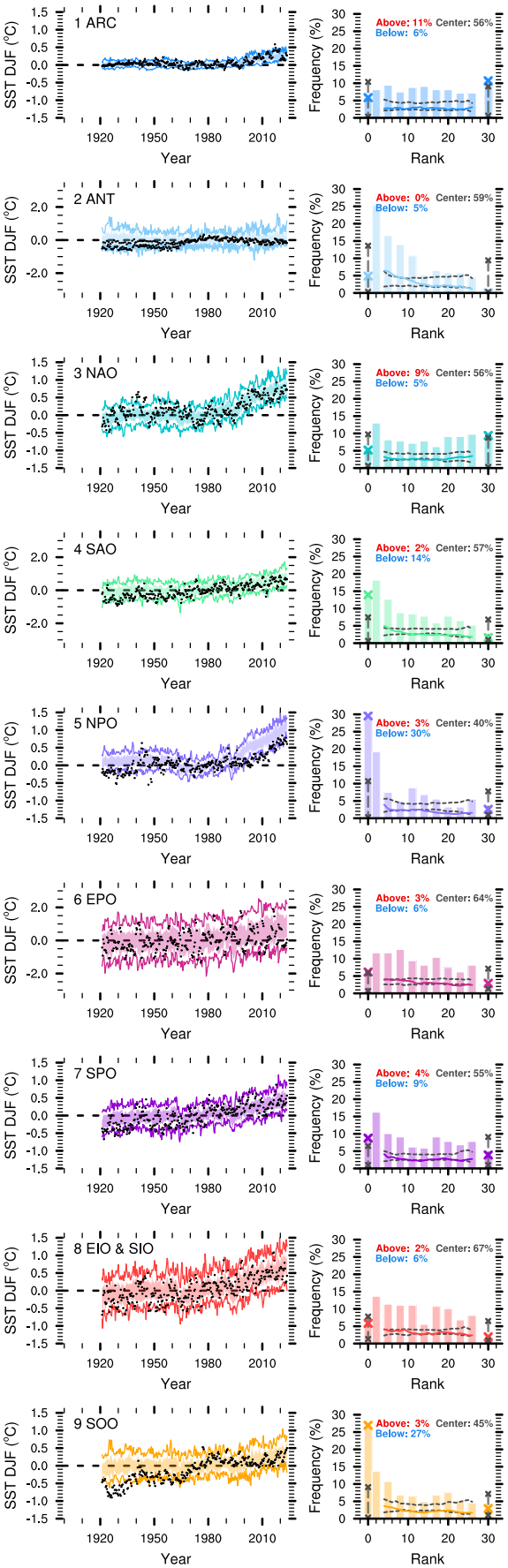
SST JJA Obs inside central 75th percentile (%)



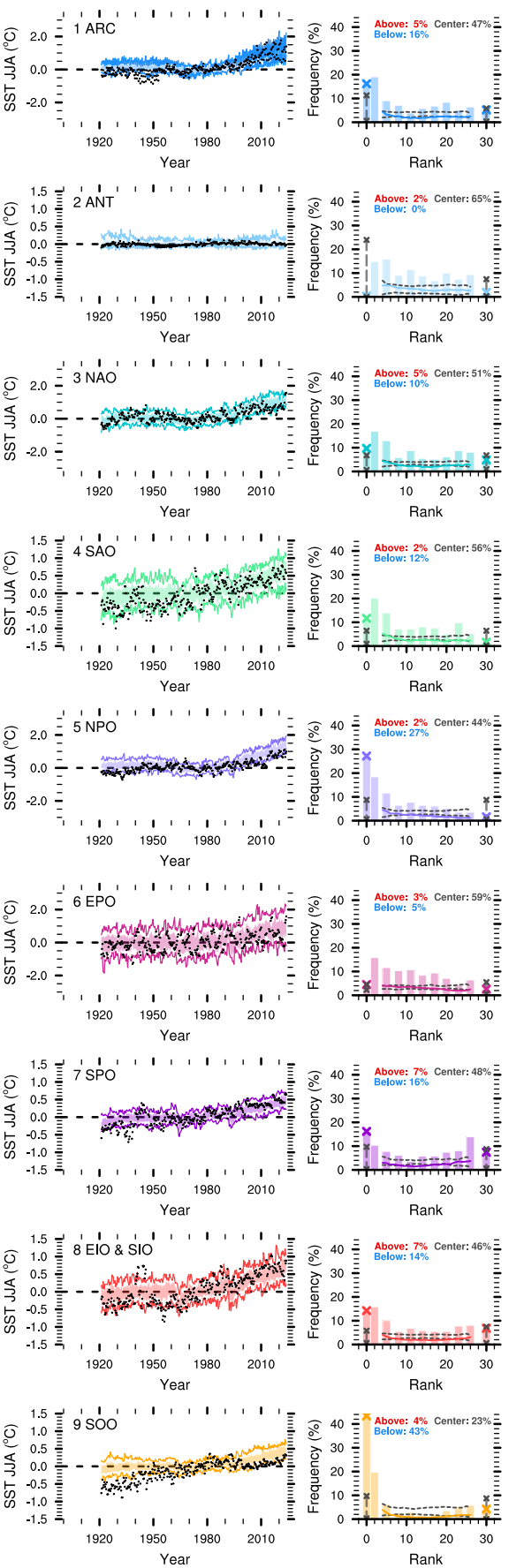
SST JJA Obs outside ensemble spread (%)



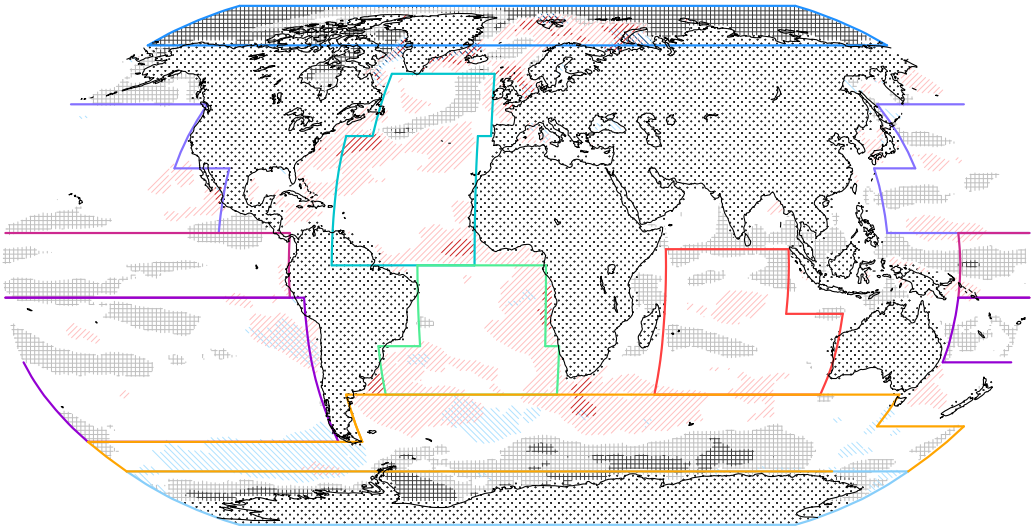
GFDL-SPEAR vs ERSSTv5 SST DJF



GFDL-SPEAR vs ERSSTv5 SST JJA



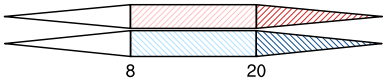
MIROC6 vs ERSSTv5 White Area = 57.4 %



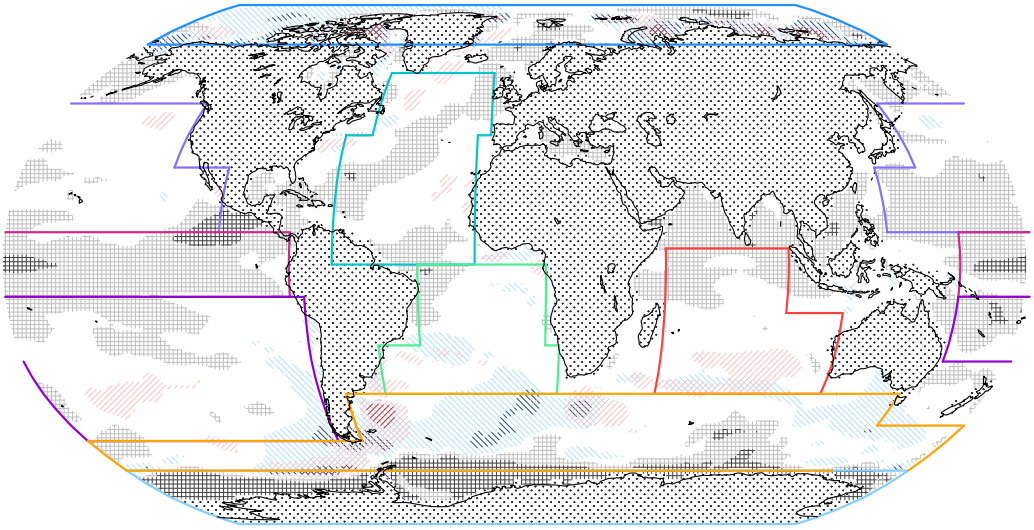
SST DJF Obs inside central 75th percentile (%)



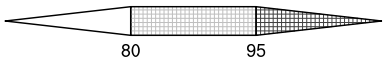
SST DJF Obs outside ensemble spread (%)



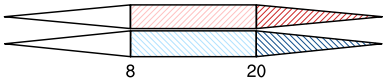
MIROC6 vs ERSSTv5 White Area = 48.6 %



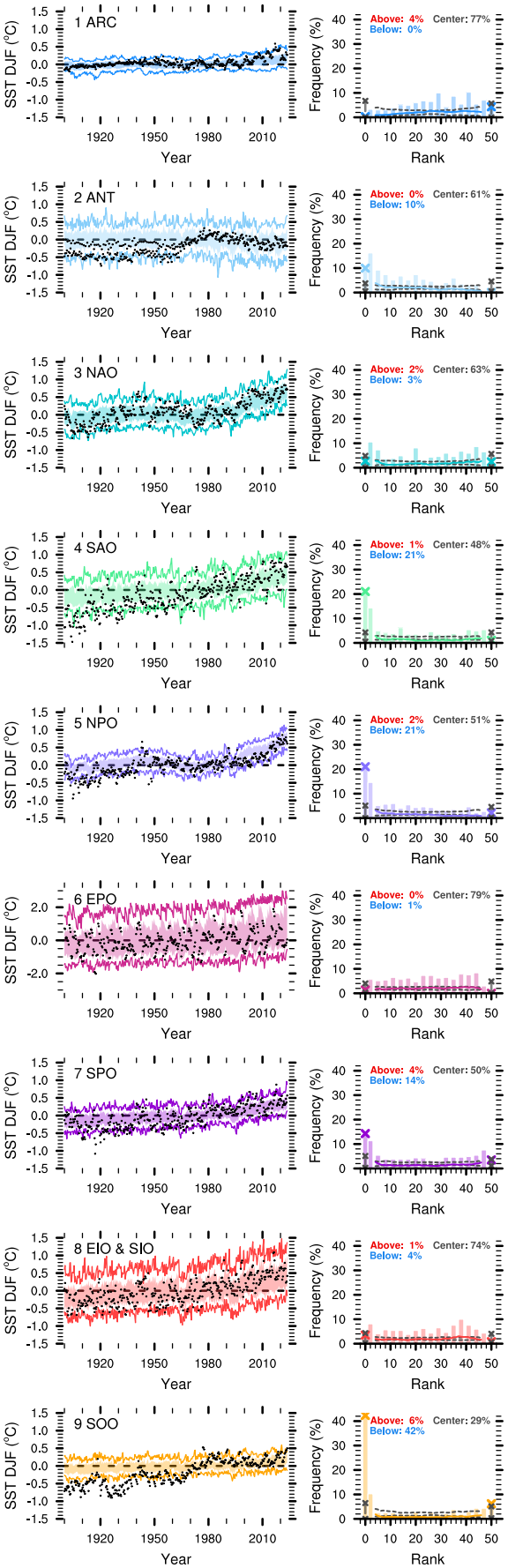
SST JJA Obs inside central 75th percentile (%)



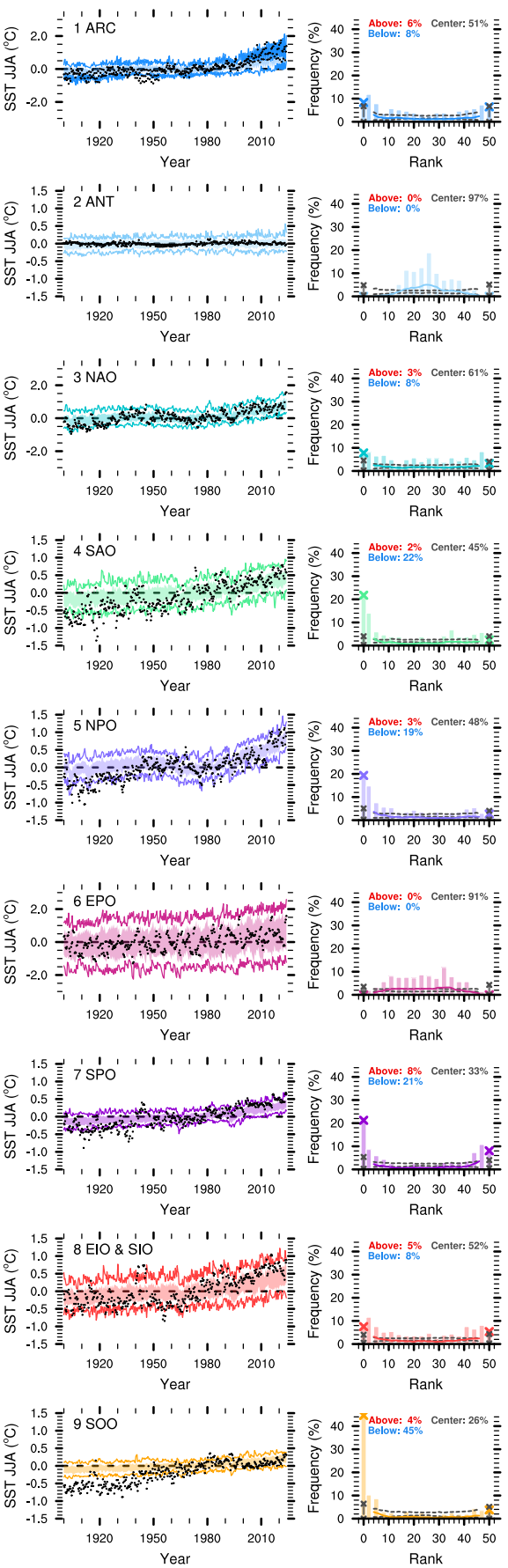
SST JJA Obs outside ensemble spread (%)



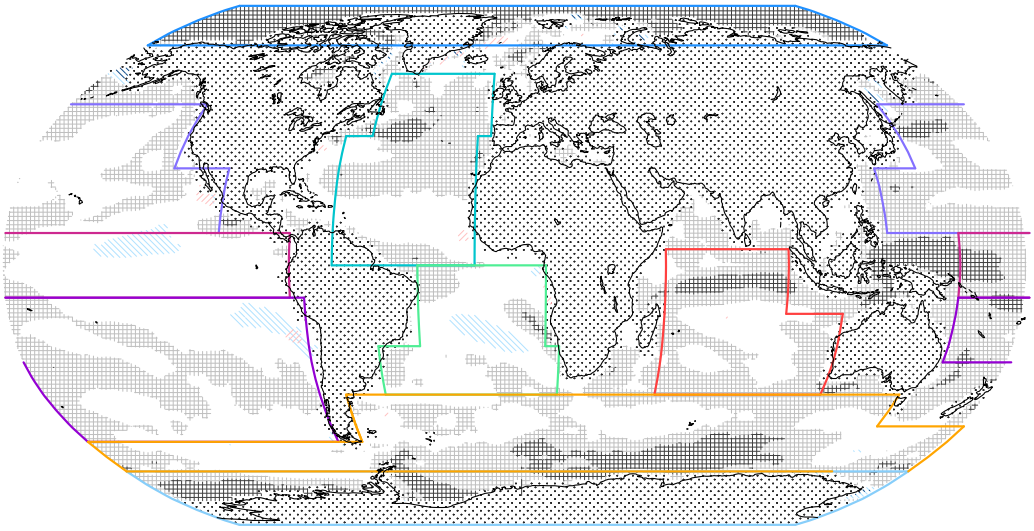
MIROC6 vs ERSSTv5 SST DJF



MIROC6 vs ERSSTv5 SST JJA



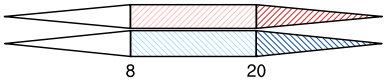
MPI-GE5 vs ERSSTv5 White Area = 43.9 %



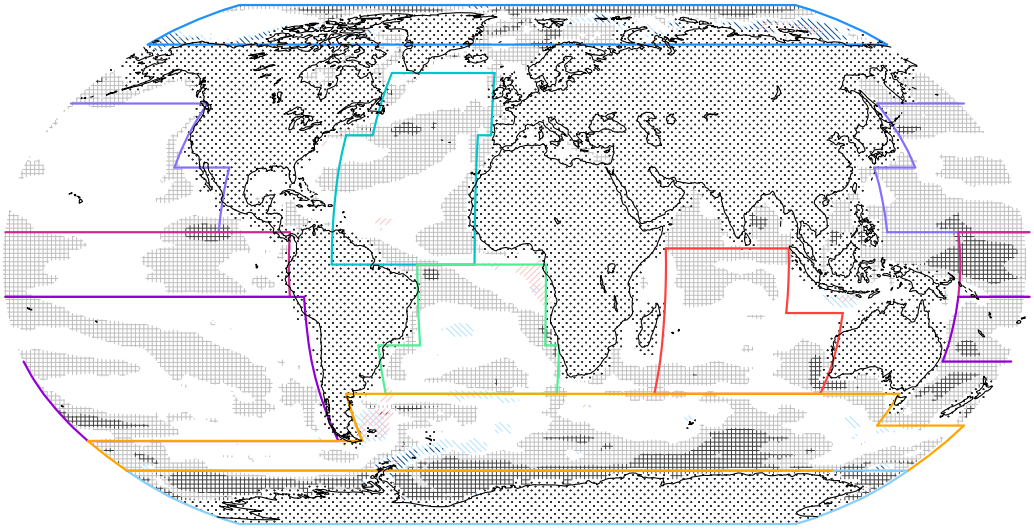
SST DJF Obs inside central 75th percentile (%)



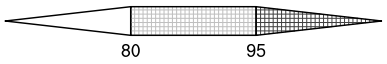
SST DJF Obs outside ensemble spread (%)



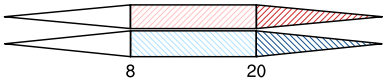
MPI-GE5 vs ERSSTv5 White Area = 51.0 %



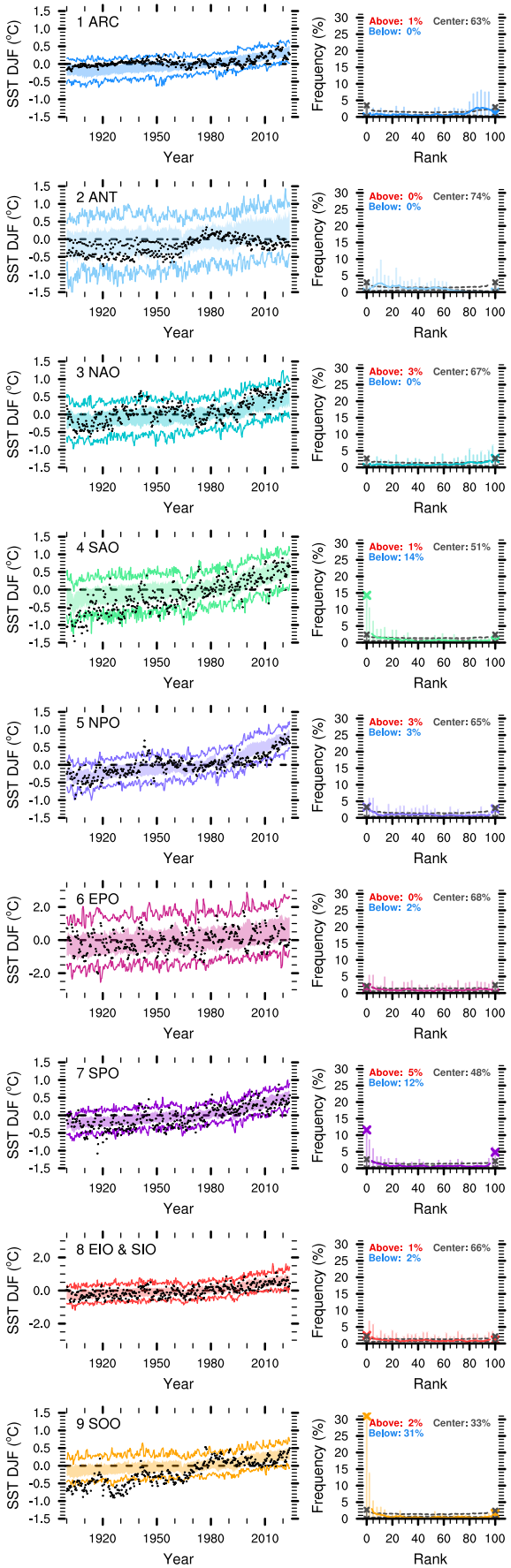
SST JJA Obs inside central 75th percentile (%)



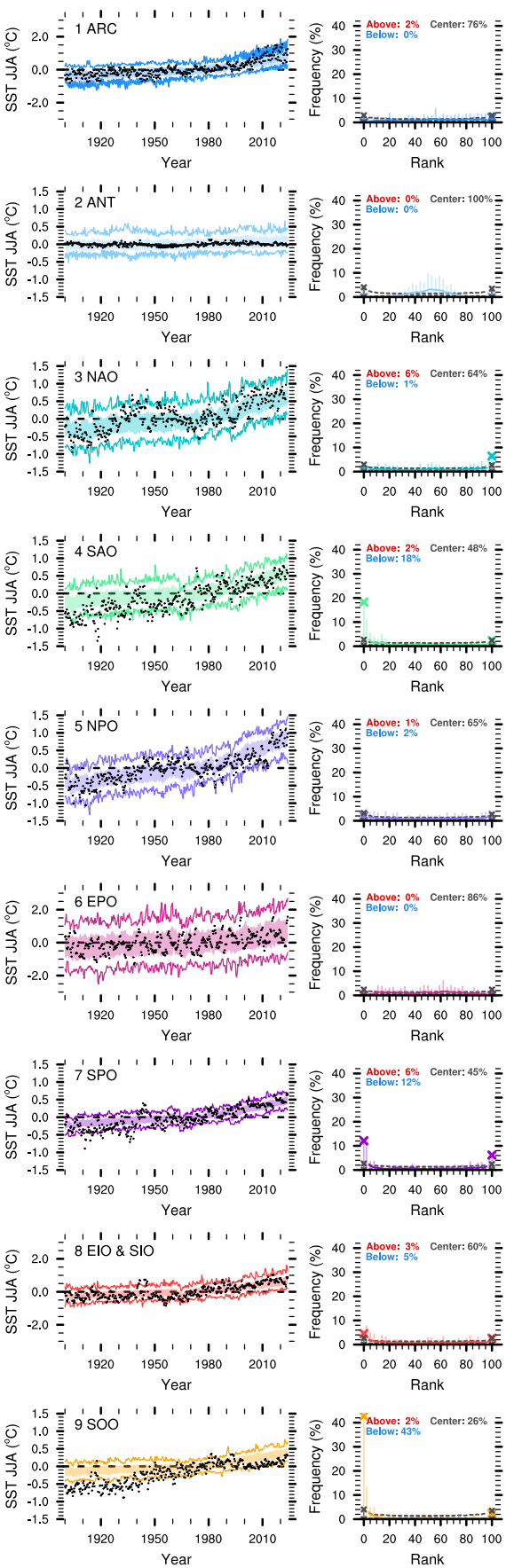
SST JJA Obs outside ensemble spread (%)



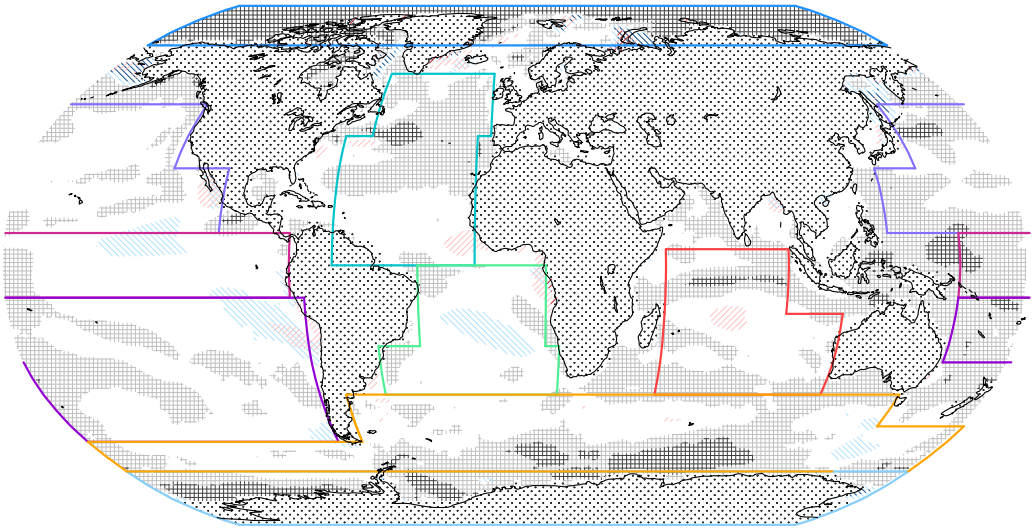
MPI-GE5 vs ERSSTv5 SST DJF



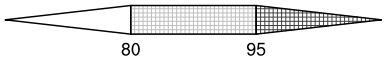
MPI-GE5 vs ERSSTv5 SST JJA



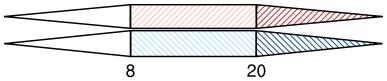
MPI-GE6 vs ERSSTv5 White Area = 49.3 %



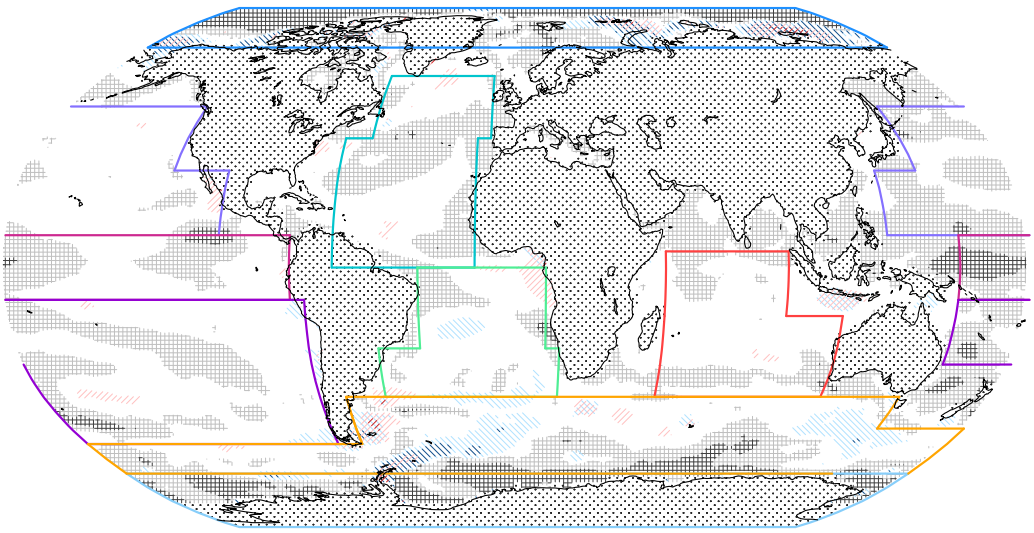
SST DJF Obs inside central 75th percentile (%)



SST DJF Obs outside ensemble spread (%)



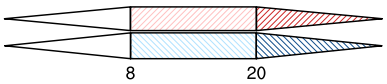
MPI-GE6 vs ERSSTv5 White Area = 53.0 %



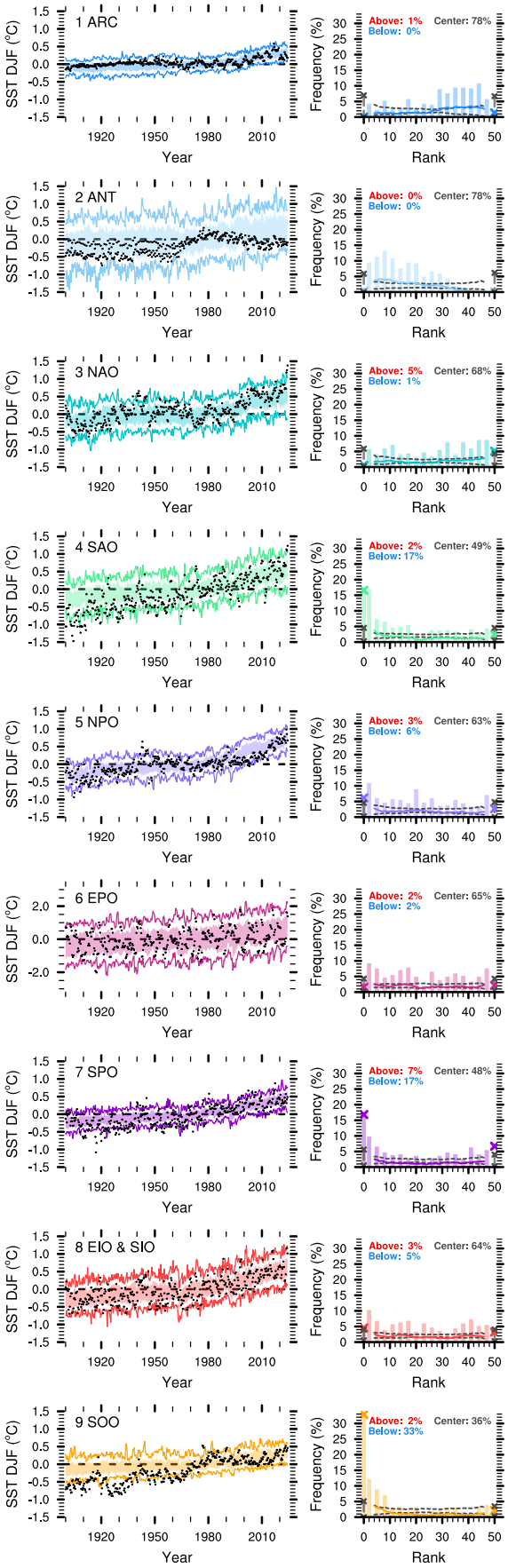
SST JJA Obs inside central 75th percentile (%)



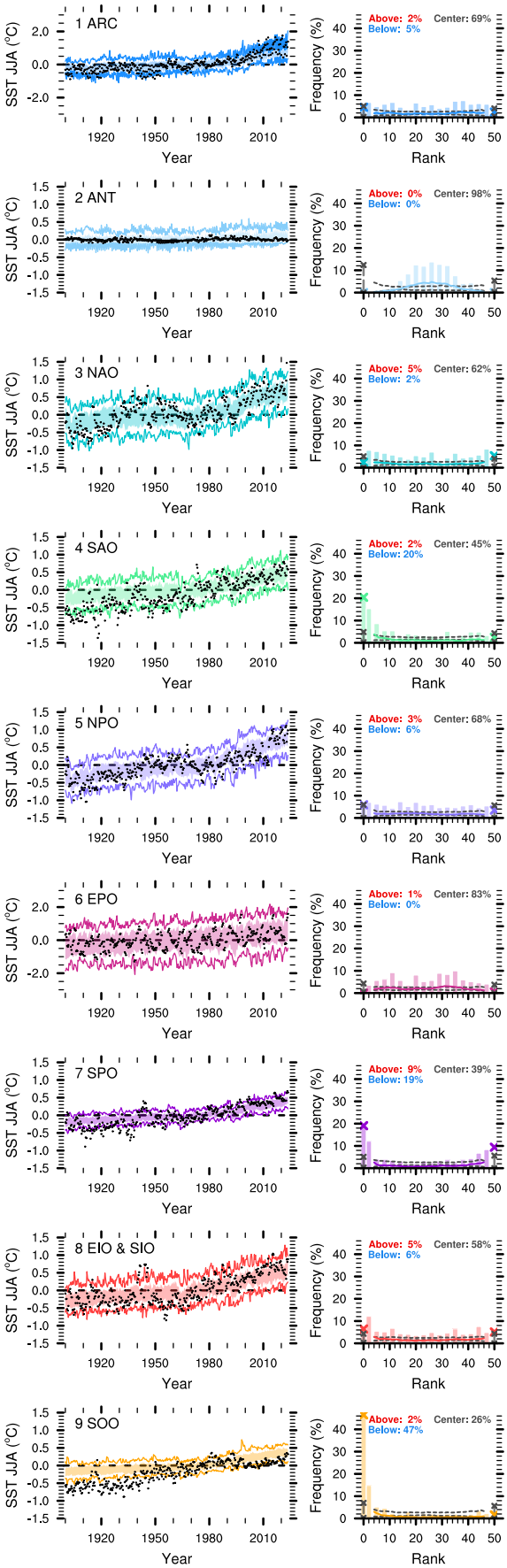
SST JJA Obs outside ensemble spread (%)



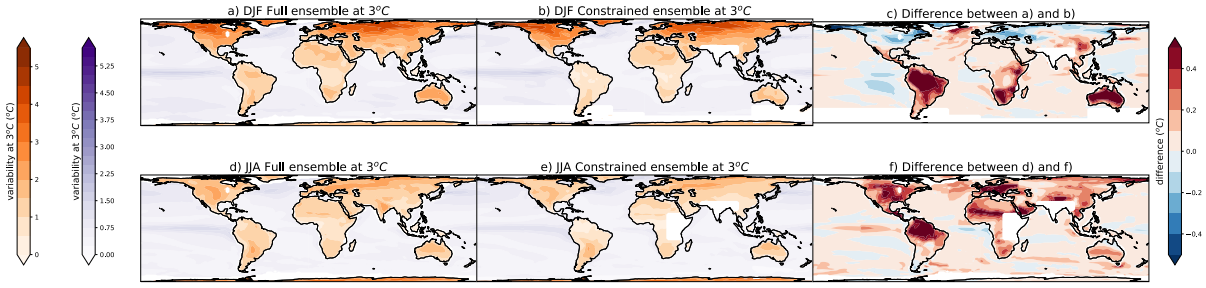
MPI-GE6 vs ERSSTv5 SST DJF



MPI-GE6 vs ERSSTv5 SST JJA



Constrained versus Unconstrained variability projections at 3˚C of global mean warming



Supplementary Figure 2.1: Difference between the full and constrained ensembles at 3˚C of warming. *a) Temperature variability at 3˚C averaged across the full ensemble for DJF, b) temperature variability at 3˚C averaged across the constrained ensemble for DJF, c) difference between a and b. d) Temperature variability at 3˚C averaged across the full ensemble for JJA, e) temperature variability at 3˚C averaged across the constrained ensemble for JJA, c) difference between d) and e).*