# Temporal validation

**Radan HUTH**

Faculty of Science, Charles University,
Prague, CZ

Institute of Atmospheric Physics, Prague, CZ

# What is it?

- validation in the temporal domain
- validation of temporal behaviour
- 2 different issues fall here
  - short-term (day-to-day) variability
  - long-term variations (trends)

# Why is it important?

- short-term variability
  - many impact sectors (models) are sensitive to it
    - agriculture
    - hydrology
- long-term variations (trends)
  - key property in relation to climate change

# Short-term variability

- various aspects
  - temperature (and some other variables)
    - persistence (temporal autocorrelations)
    - day-to-day changes (variations) – empirical distributions
    - extended extreme events (heat waves, cold spells)
  - precipitation
    - separate evaluation of
      - precipitation occurrence / non-occurrence (binary variable)
      - precipitation amounts (continuous variable)
    - wet / dry periods
    - transition probabilities (wet→wet, dry→wet)
    - “binary persistence” – quantifiable e.g. by Heidke “skill” score
    - not much sense in examining temporal properties of precipitation amounts – perhaps only in very wet climates

# Short-term variability

- issue that must be considered: grid box vs. stations
- gridbox (gridpoint) representation (whether in RCM or gridded observations) may not truly represent station characteristics of temporal behaviour and extremes
- (smoothing effect)
- must be kept in mind when interpreting results
- e.g. Osborn & Hulme: Development of a relationship between station and grid-box rainday frequencies for climate model evaluation, *J. Climate* 1997

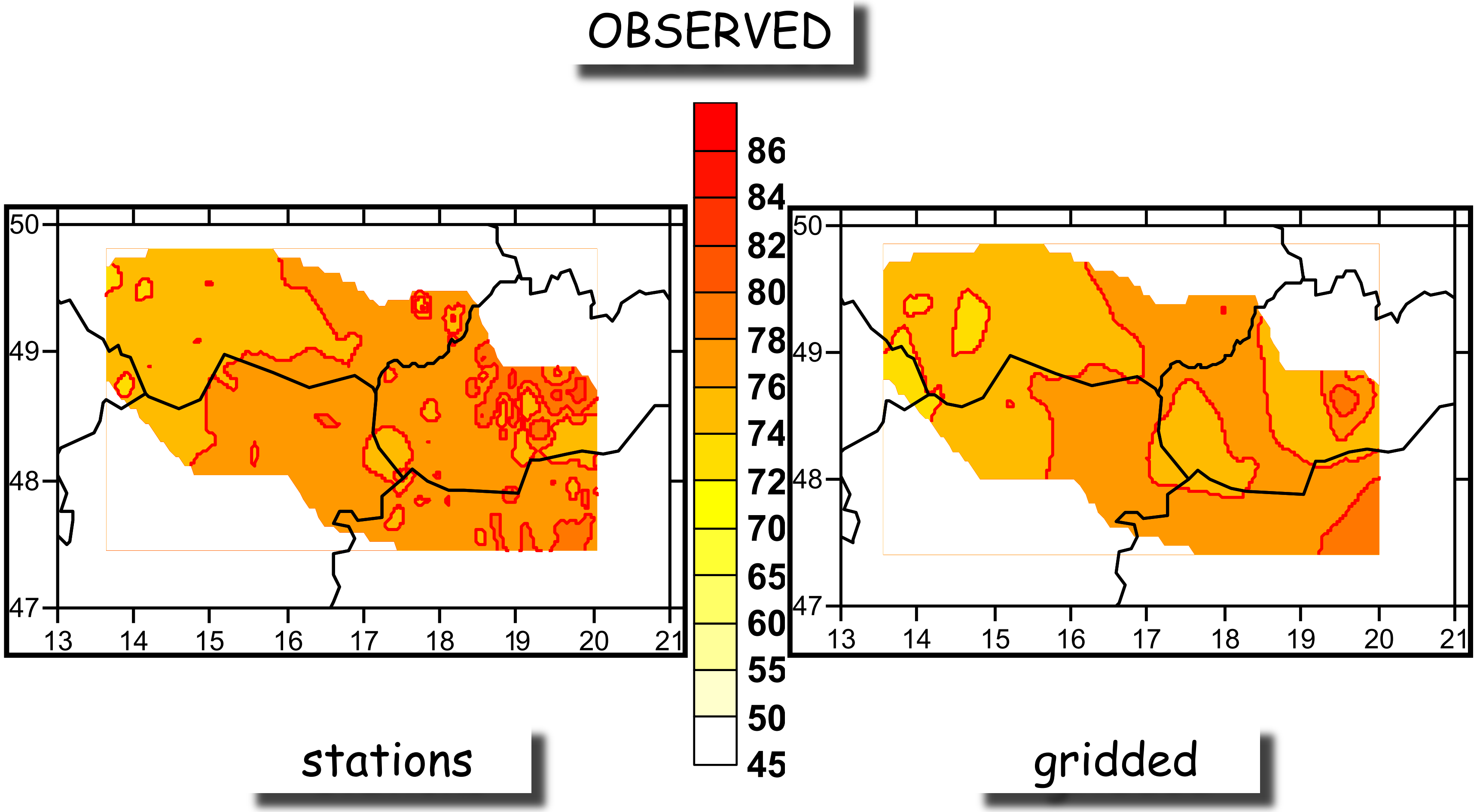
# Examples

- four examples to illustrate validation of short-term variability
- Huth et al., *J. Climate* 2001
  - 6 stations in central Europe
  - SDS
    - linear regression
    - different ways of accounting for missing variance
  - 2 variants of weather generator
  - 2 GCMs
- Huth, *J. Climate* 2002
  - 39 stations in central & western Europe
  - various linear SDS methods (MLR, CCA, SVD, …) with various combinations of predictor fields
- Huth et al., *Int. J. Climatol.,* 2008
  - 8 stations in Europe
  - linear & nonlinear SDS methods
- Huth et al., *Theor. Appl. Climatol.* 2015
  - dense network (stations & grid) in central Europe (CZ, AT, HU, SK borders)
  - SDS
    - linear regression
    - 4 non-linear methods (analogs, local linear models, 2 neural networks)
  - 2 RCMs
    - ALADIN-Climate/CZ – 10 km grid
    - Reg CM3 – 25 km grid

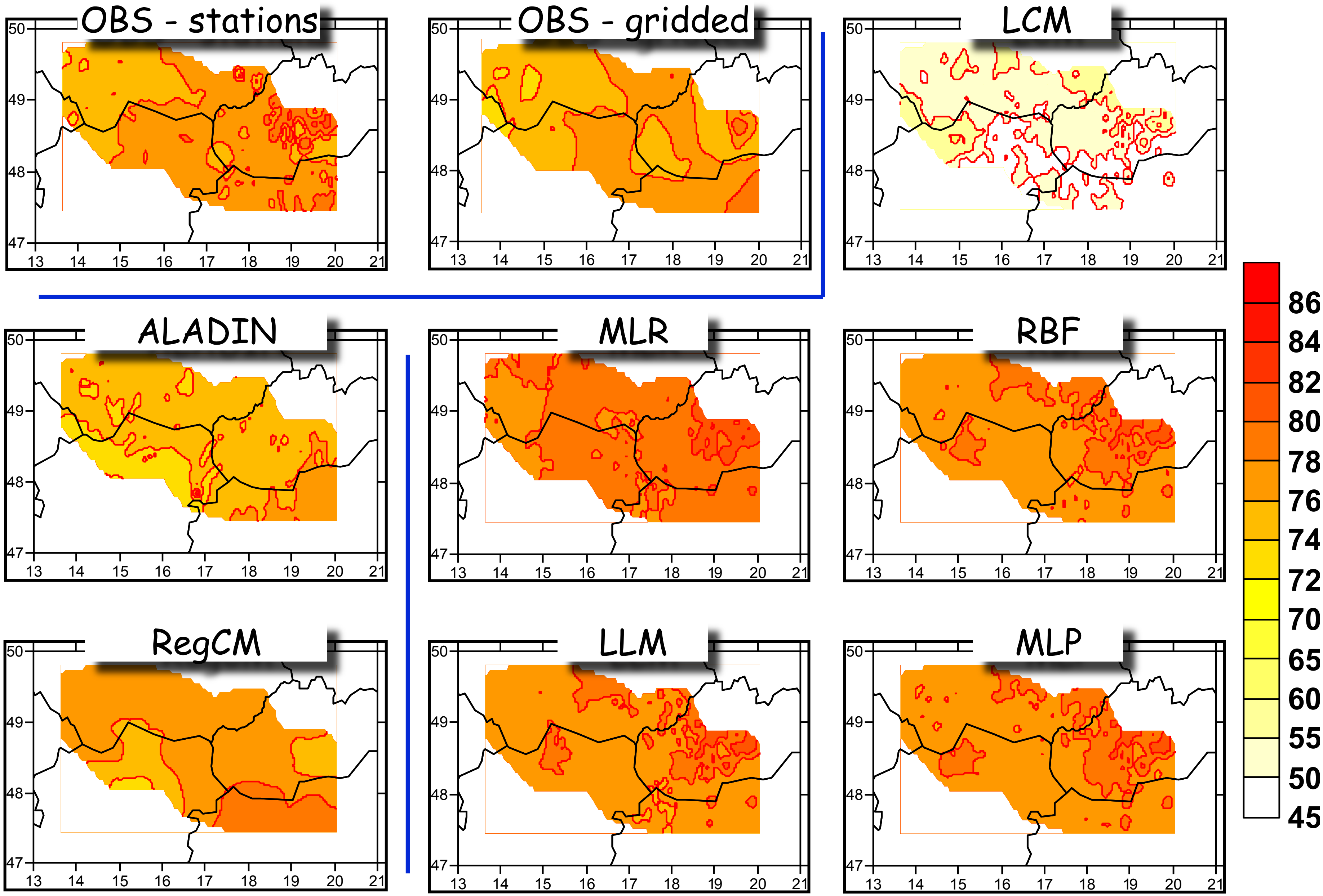
# Persistence

- lag-1 day autocorrelation
- simple, important, but only rarely evaluated
- note: does not account for the magnitude of day-to-day variability
- note: post-processing (bias correction) methods cannot affect it

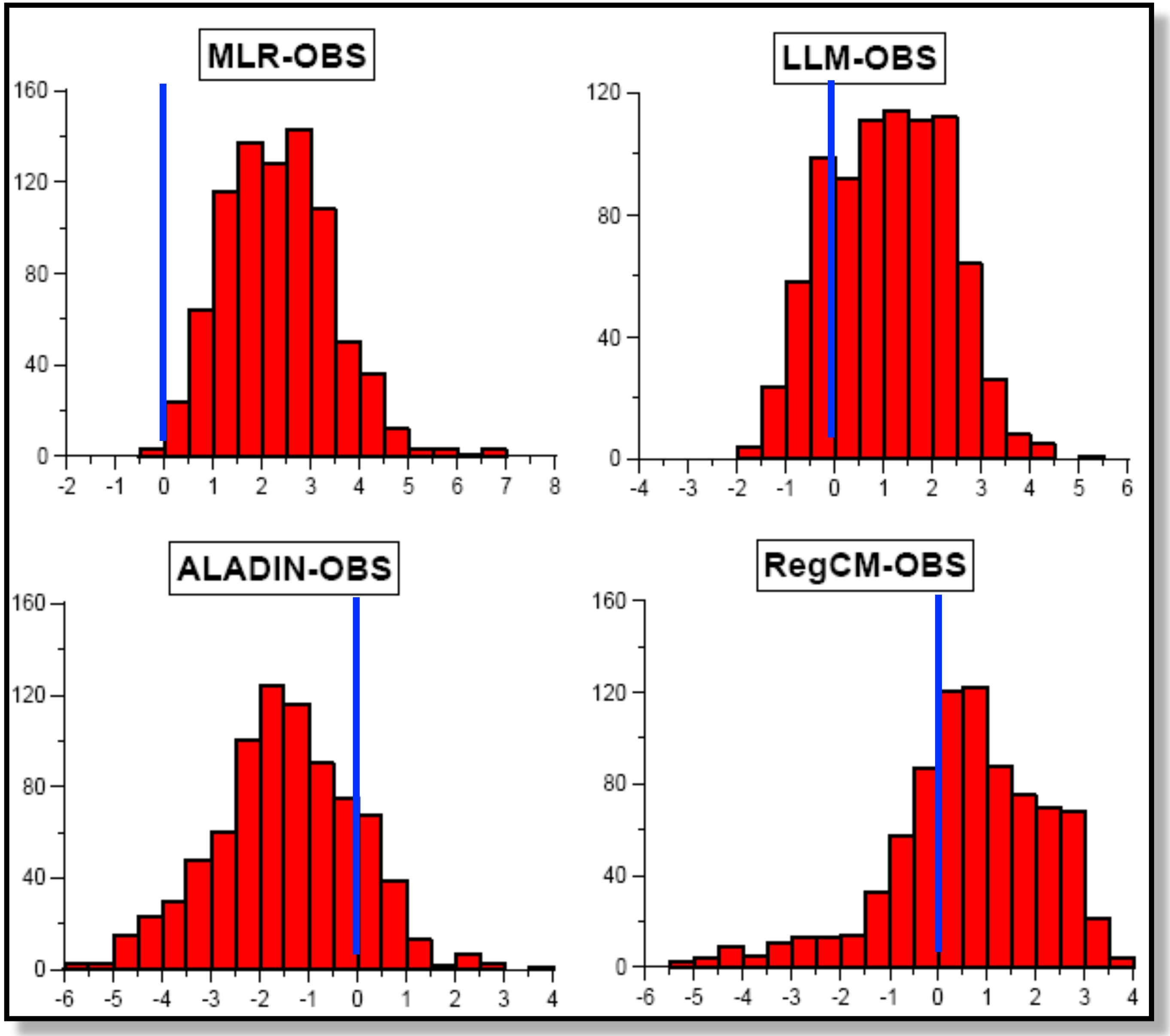
# Tmax, 1-day lag persistence, whole year



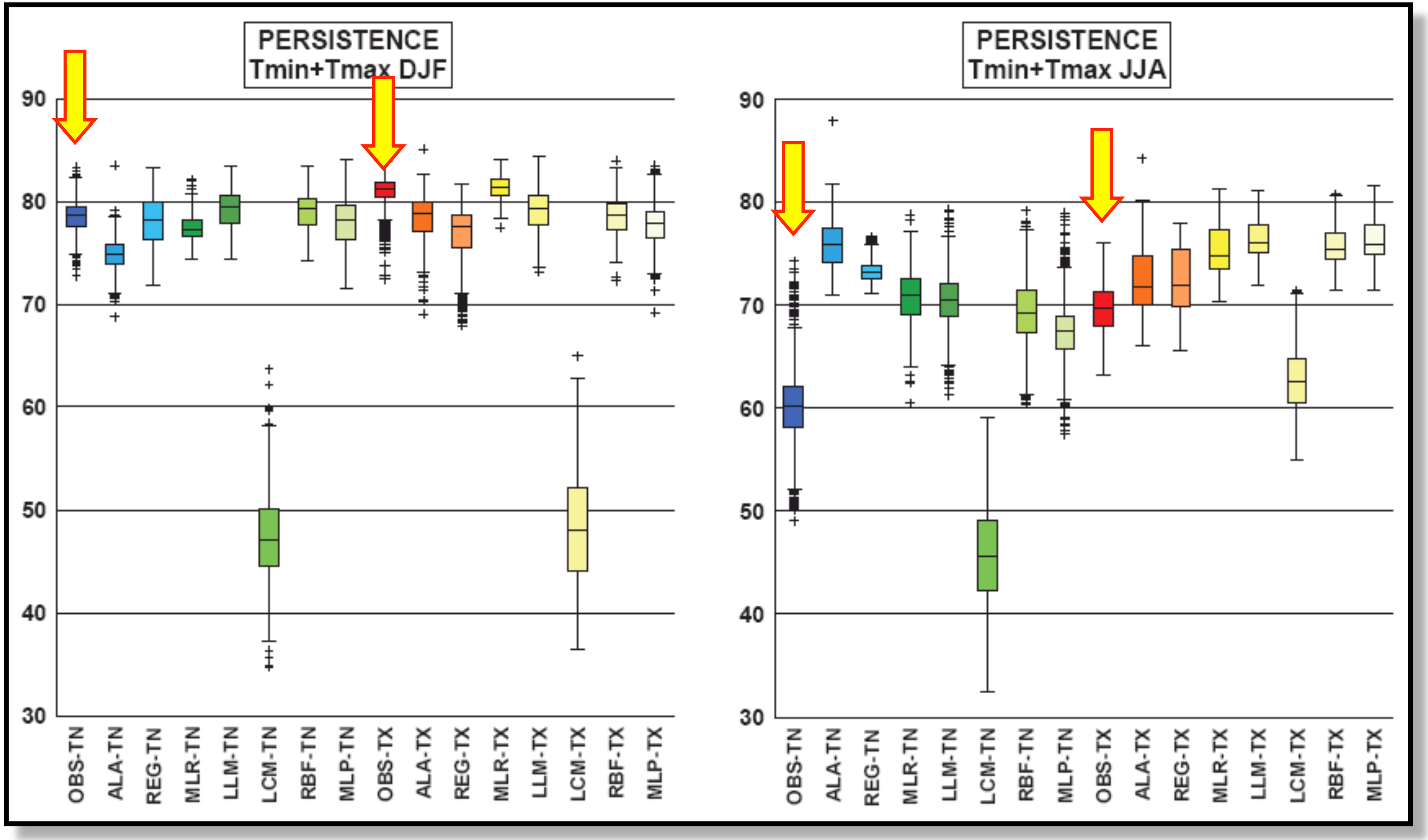
# Tmax, 1-day lag persistence, whole year



# Tmax, 1-day lag persistence, whole year
## difference from OBS, x100



# Tmax & Tmin, 1-day lag persistence, DJF & JJA



# Day-to-day changes

- different aspect of short-term variability
- time series with identical persistence may have very different distributions of day-to-day changes
- characteristics of statistical distribution (histogram) of day-to-day changes are evaluated, namely
  - standard deviation
  - skewness (asymmetry)
- reflects the ability of models to include (and correctly simulate) various physical processes (radiation, advection, …)

# day-to-day max.temperature change, summer

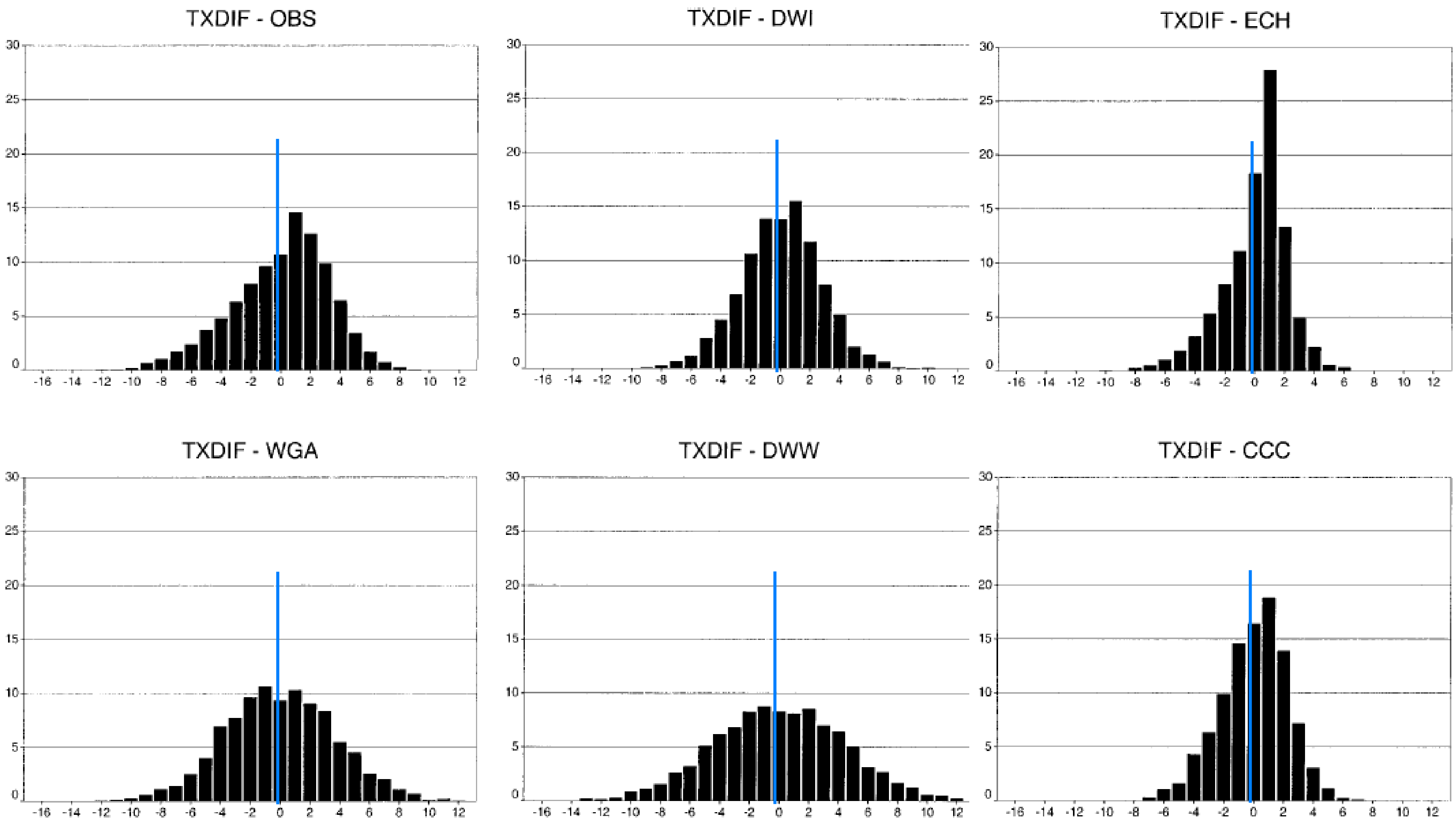



FIG. 7. Histograms of day-to-day change in maximum temperature in summer (binned into 1°C intervals) at Strážnice for selected time series. On the horizontal axis is temperature difference in degrees Celsius, on the vertical axis is frequency in percent. See Table 1 for definitions of time series.

# day-to-day min.temperature change, winter

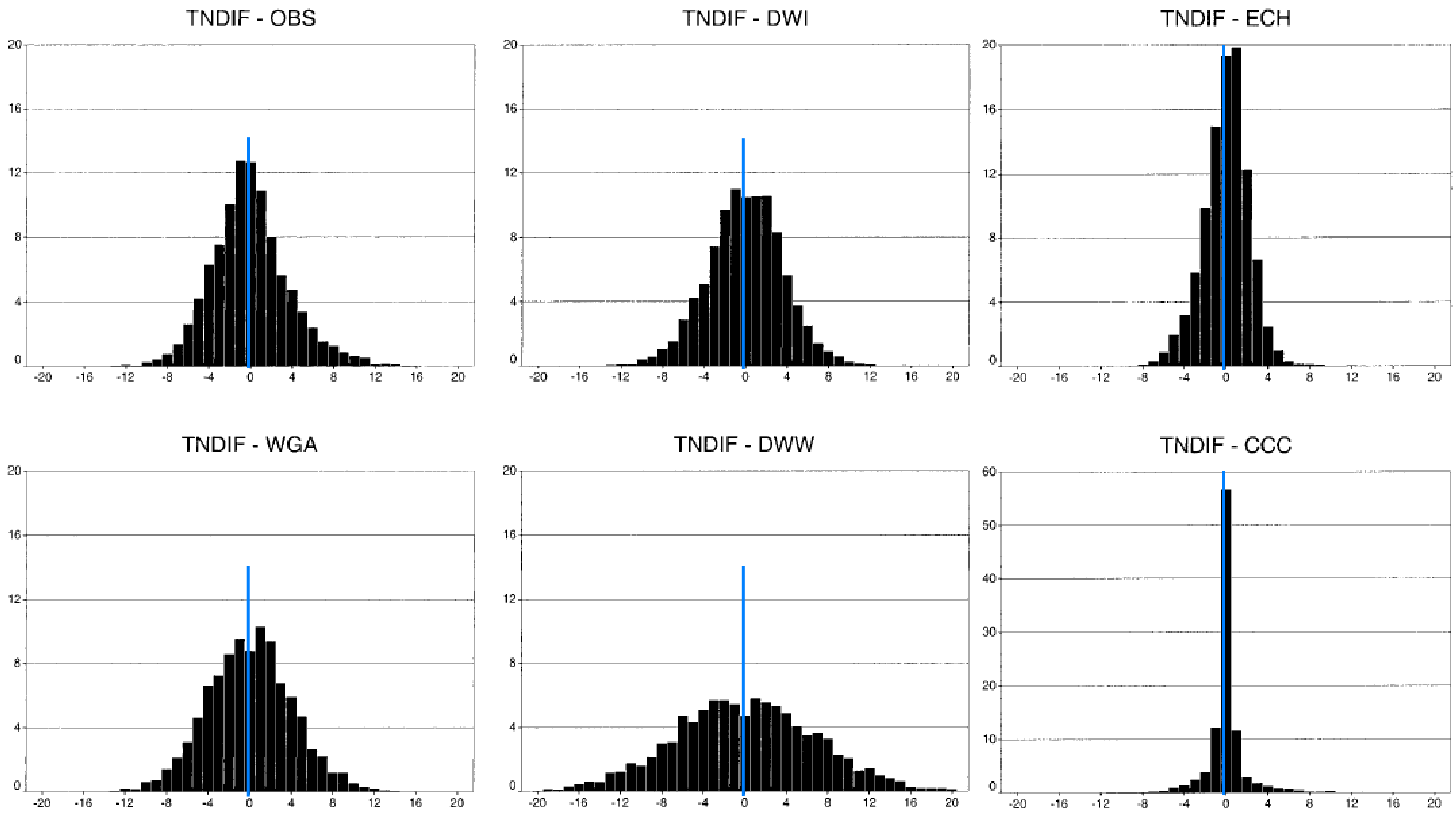



FIG. 8. Same as in Fig. 7 except for minimum temperature in winter. Note a different vertical scale for the CCC series.

FIG. 8. (Continued)

# day-to-day temperature change

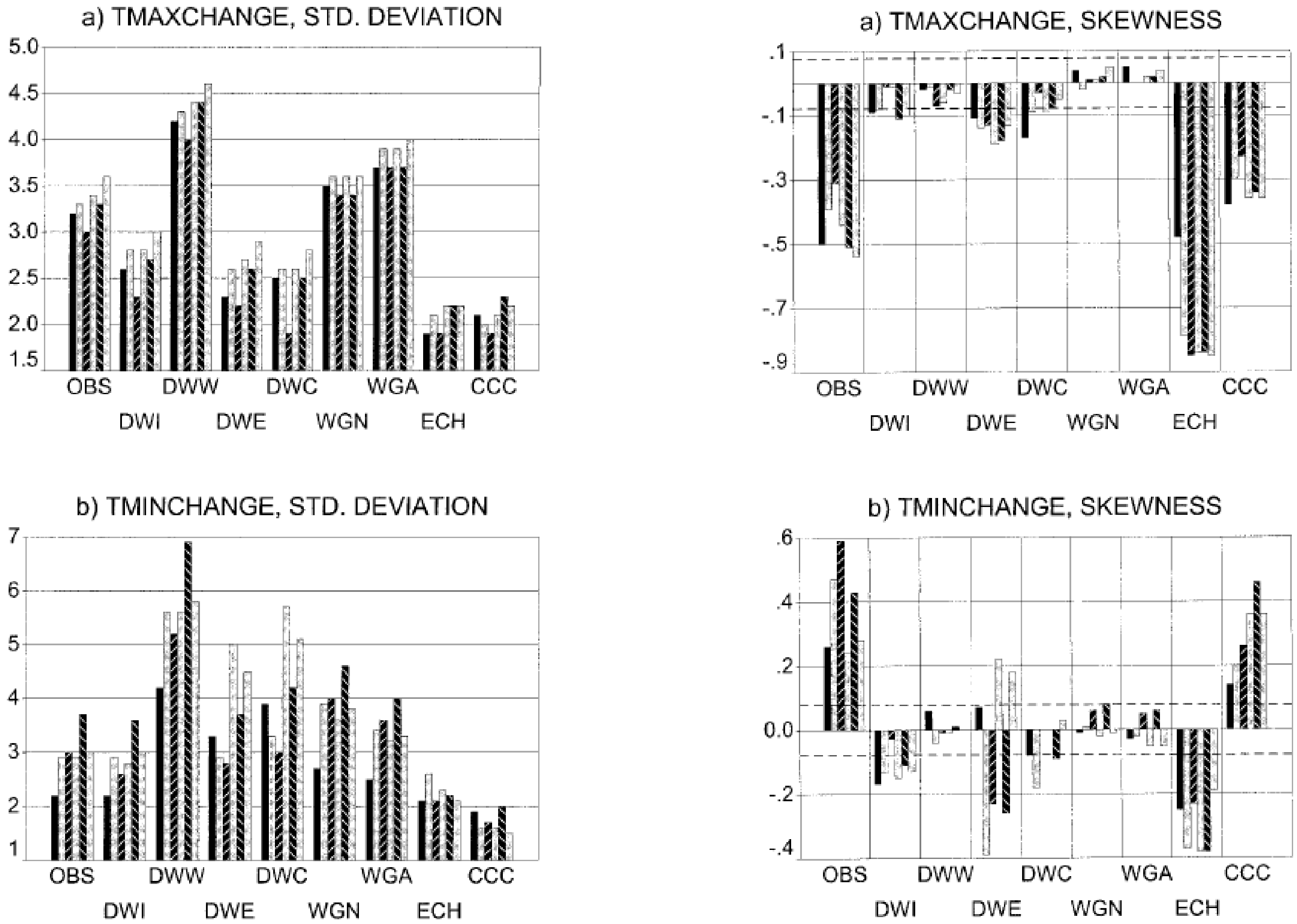



FIG. 6. Same as in Fig. 4 except for skewness. The dashed lines indicate the critical value for skewness to be different from zero at the 95% significance level, assuming autocorrelation of 0.1 (typical observed value in both seasons).

# Extended extreme events

- important characteristics of extreme weather
- potentially big difference if extremes occur individually or in sequences
- examples
  - heat waves
  - cold spells
- typical definition – periods of a certain minimum duration with temperature exceeding a threshold (absolute or percentile-based)
- integral characteristic – integrates different aspects o temperature (extremes, persistence, annual cycle, …)
- possible characteristics to validate
  - frequency
  - duration
  - percentage of extreme days included in extended events (reflects mainly persistence)
  - intensity (highest temperature or highest temperature exceedance over threshold during the event)
  - date of occurrence (reflects the ability to simulate annual cycle)

# heat waves, cold spells

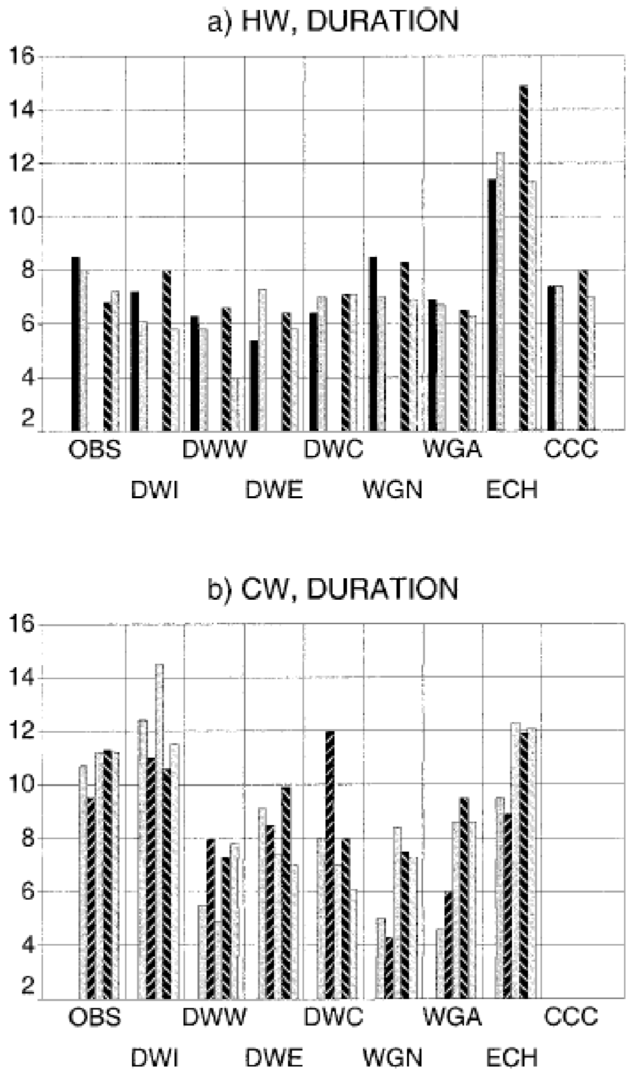



FIG. 10. Mean duration (in days) of (a) heat waves and (b) cold waves. The stations with a low incidence of PETEs are not shown. Legends as in Fig. 3; in (b) the CCC series are omitted.

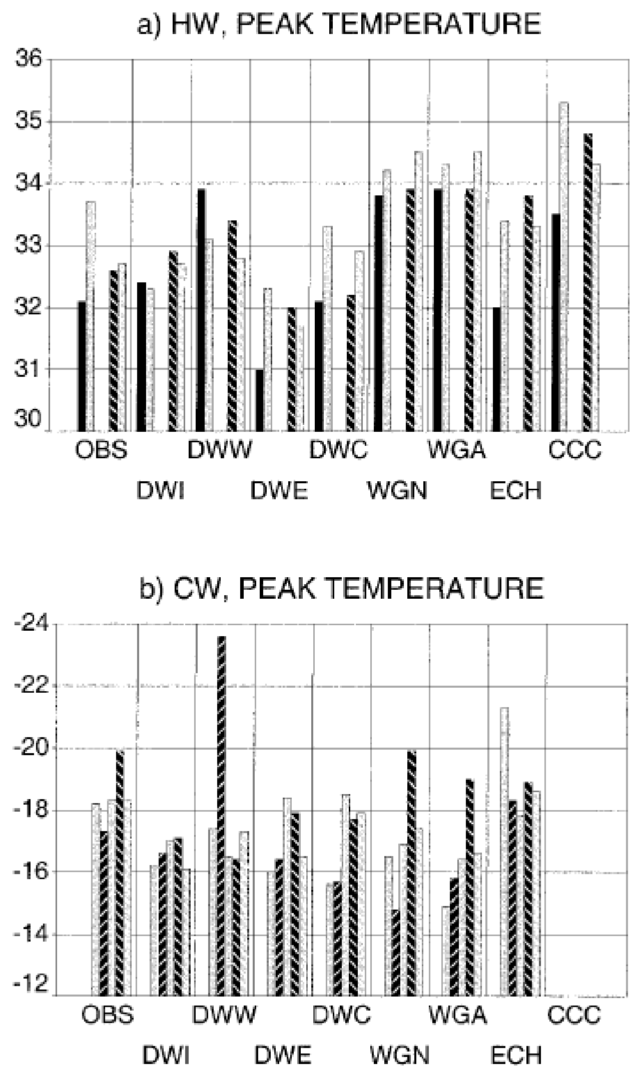



FIG. 11. Mean peak temperature (in °C) for (a) heat waves and (b) cold waves. Otherwise as in Fig. 10.

# heat waves, cold spells

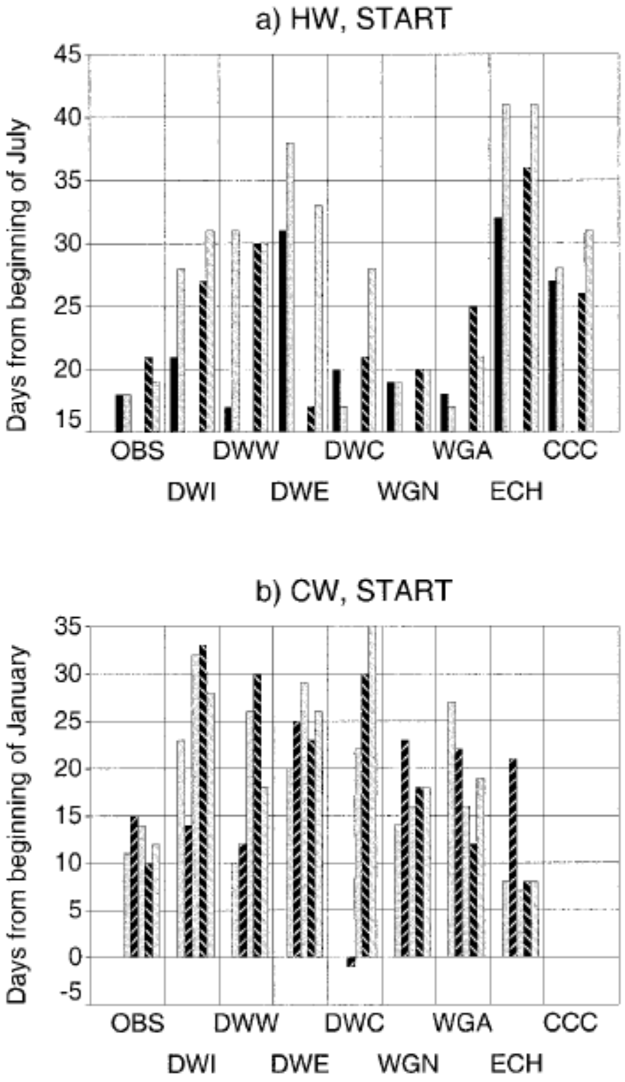



FIG. 12. Mean date of start of (a) heat waves and (b) cold waves. Numbers on the ordinate indicate the number of days from the beginning of Jul and Jan, respectively: e.g., for HWs 20 means 20 Jul. Otherwise as in Fig. 10.

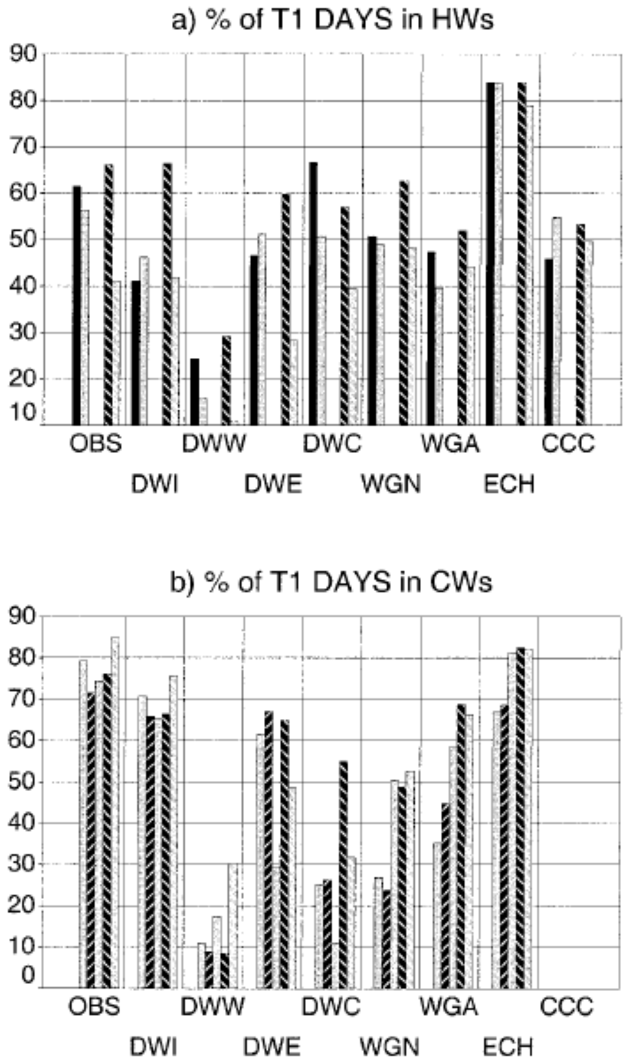
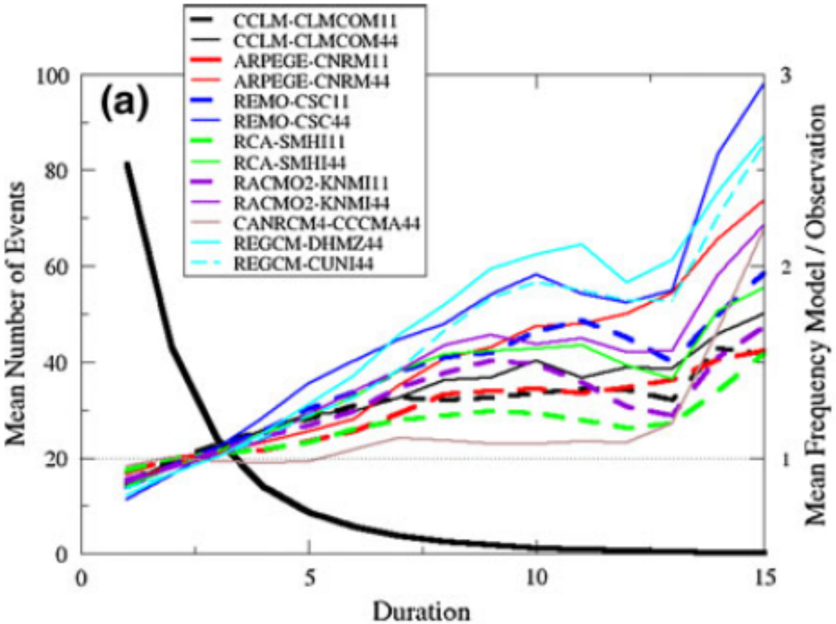



FIG. 13. Inclusion of $T_1$ days into (a) HWs and (b) CWs in percent. For definition see the text. Otherwise as in Fig. 10.

# heat waves

- Vautard et al., *Clim. Dyn.* 2013



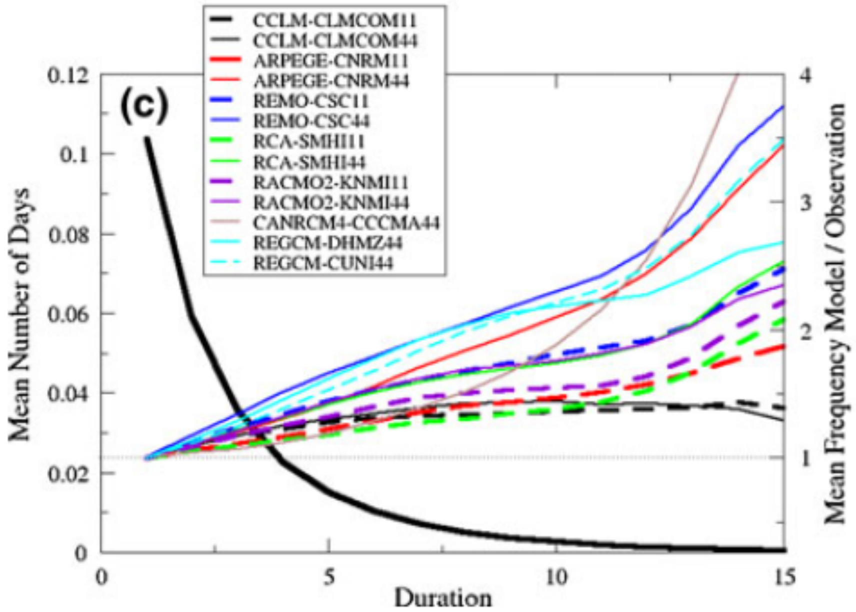
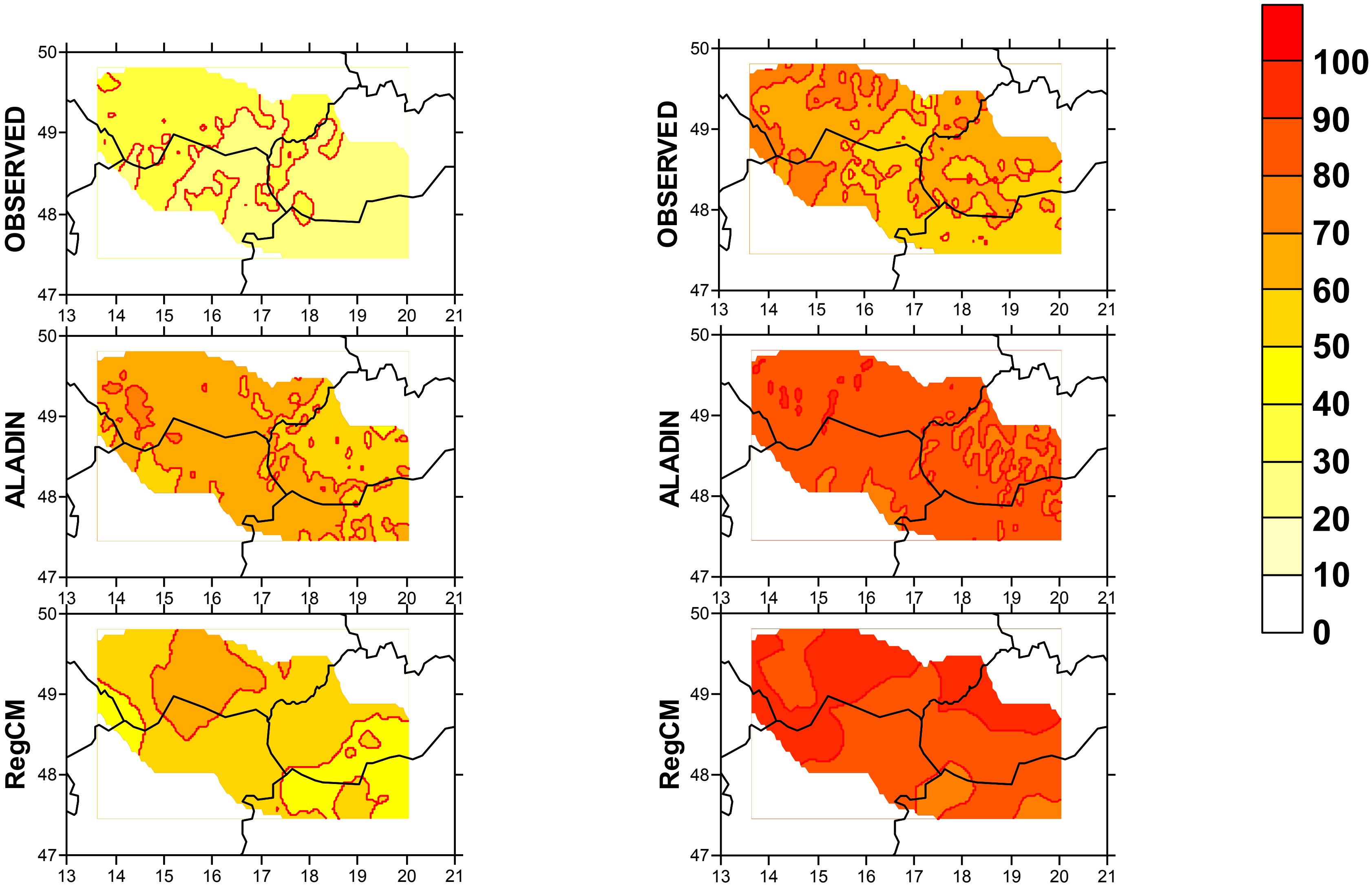
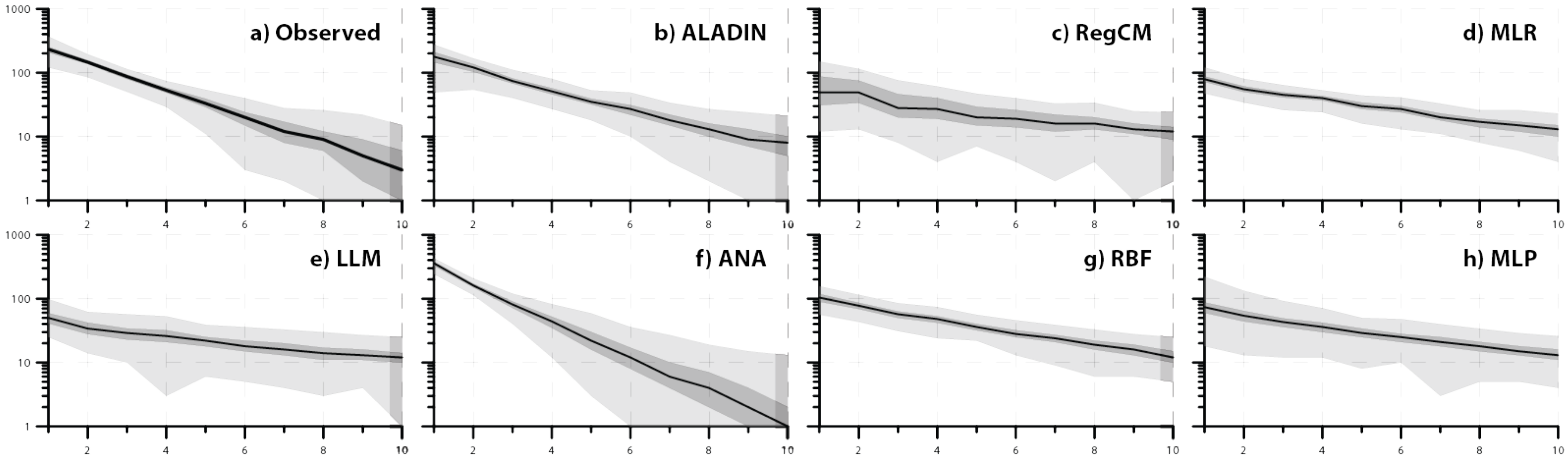




**Fig. 5 a, b** Average observed number of heat wave events of duration larger than a given number of days, as a function of this number of days (*heavy black decreasing curve*). The average is performed over the ECA&D stations lying in the [20 W, 40E: 30 N, 70 N] domain. The figure also shows the ratio of the number of simulated to observed events for each model (*other curves* with model legend given in the *graph* itself). To improve readability the ensemble is split into two sub-ensembles: non-WRF simulations (**a**) and WRF simulations (**b**). **c, d** Same as **a, b** for the frequency of days (instead of number of events) in spells with durations larger or equal to the value in abscissa (instead of the number of events). High-resolution simulations are highlighted with *dashed lines* and have the *same color* as the low-resolution simulations

# Precipitation transition probabilities: dry-wet, wet-wet



# Wet periods




- Number of uninterrupted periods of wet days 1 to 10 days long. Shown are the median value in the set of grid points in the validation domain (bold line), the interquartile range (darker shading) and min-max range (lighter shading).

# Trends (long-term variations)

- long-term variations – essential for climate change assessment, impacts etc.
- if a model is not able to simulate current trends, how can we rely on it for future climate change?
- in spite of it, trend validation studies are scarce
- model's time series must correspond to real time series
- i.e., applicable only if model is driven by observed data (typically represented by reanalysis)
  - RCM nested in reanalysis
  - SDS model trained on reanalysis
  - GCM nudged towards reanalysis (very rarely done so far)
- two possible approaches
  - trends as linear regression fits – variable vs. time
  - differences for contrasting periods (warm vs. cold; wet vs. dry)

# Trends (long-term variations)

- three examples
- all for temperature
- Lorenz & Jacob, *Clim. Res.* 2010
  - 8 European domains
  - 13 RCMs driven by ERA40
  - ENSEMBLES project
- Bukovsky, *J. Climate* 2012
  - North America
  - 6 RCMs driven by NCEP-2
  - NARCCAP programme
- Huth et al., *Theor. Appl. Climatol.* 2015
  - central Europe
  - 2 RCMs driven by ERA40
  - 5 SDS models trained on ERA40
  - CECILIA project

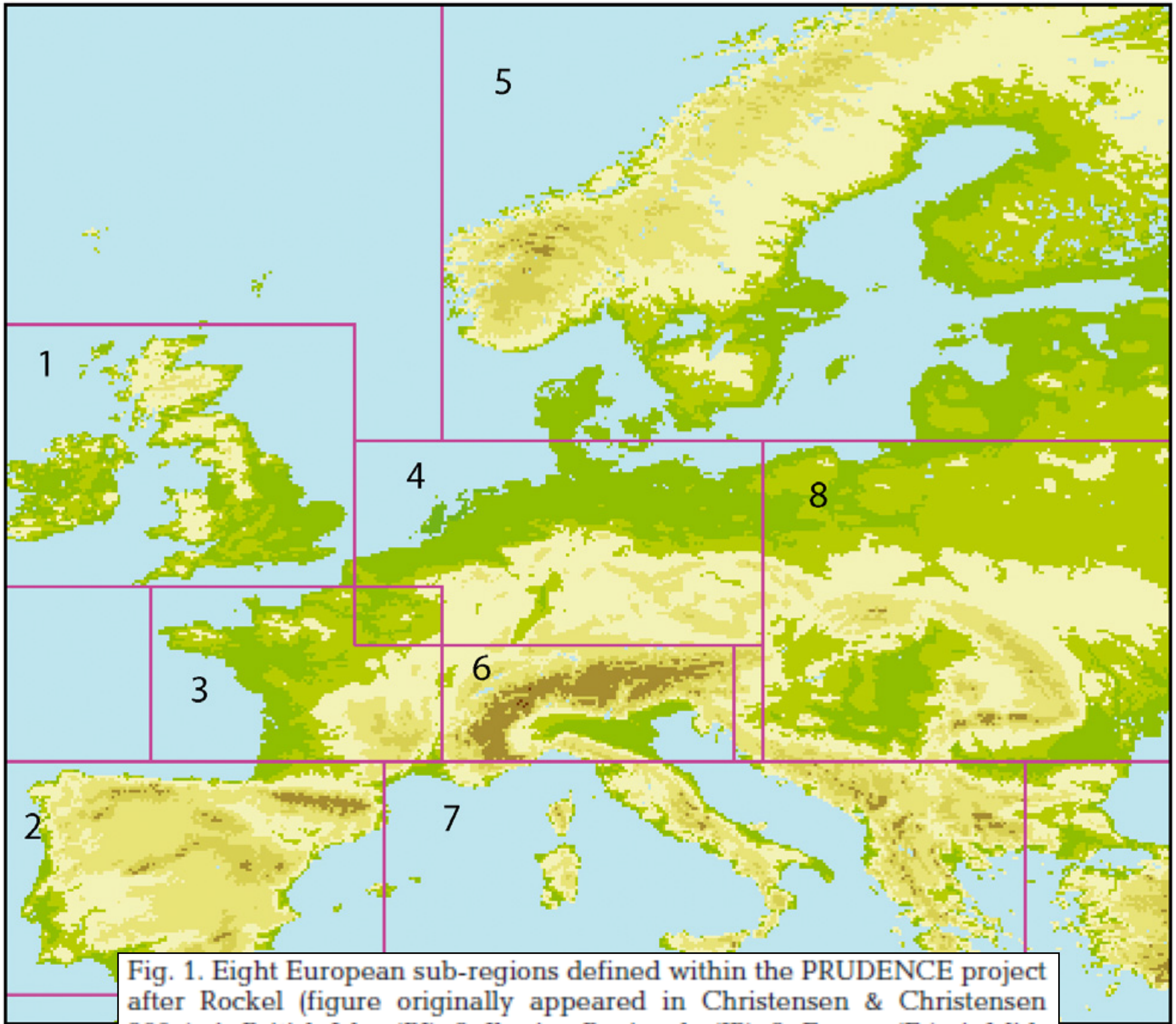
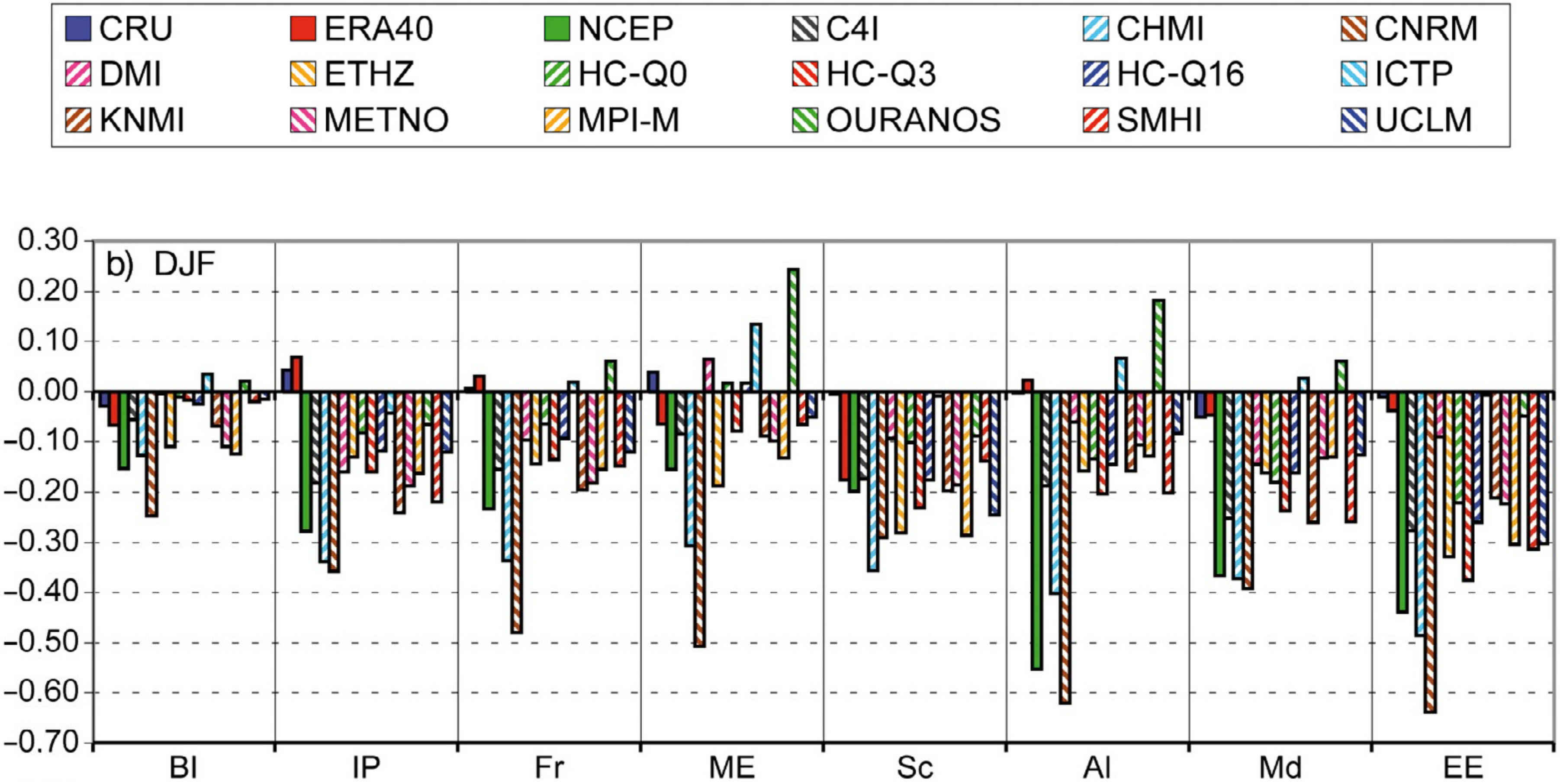
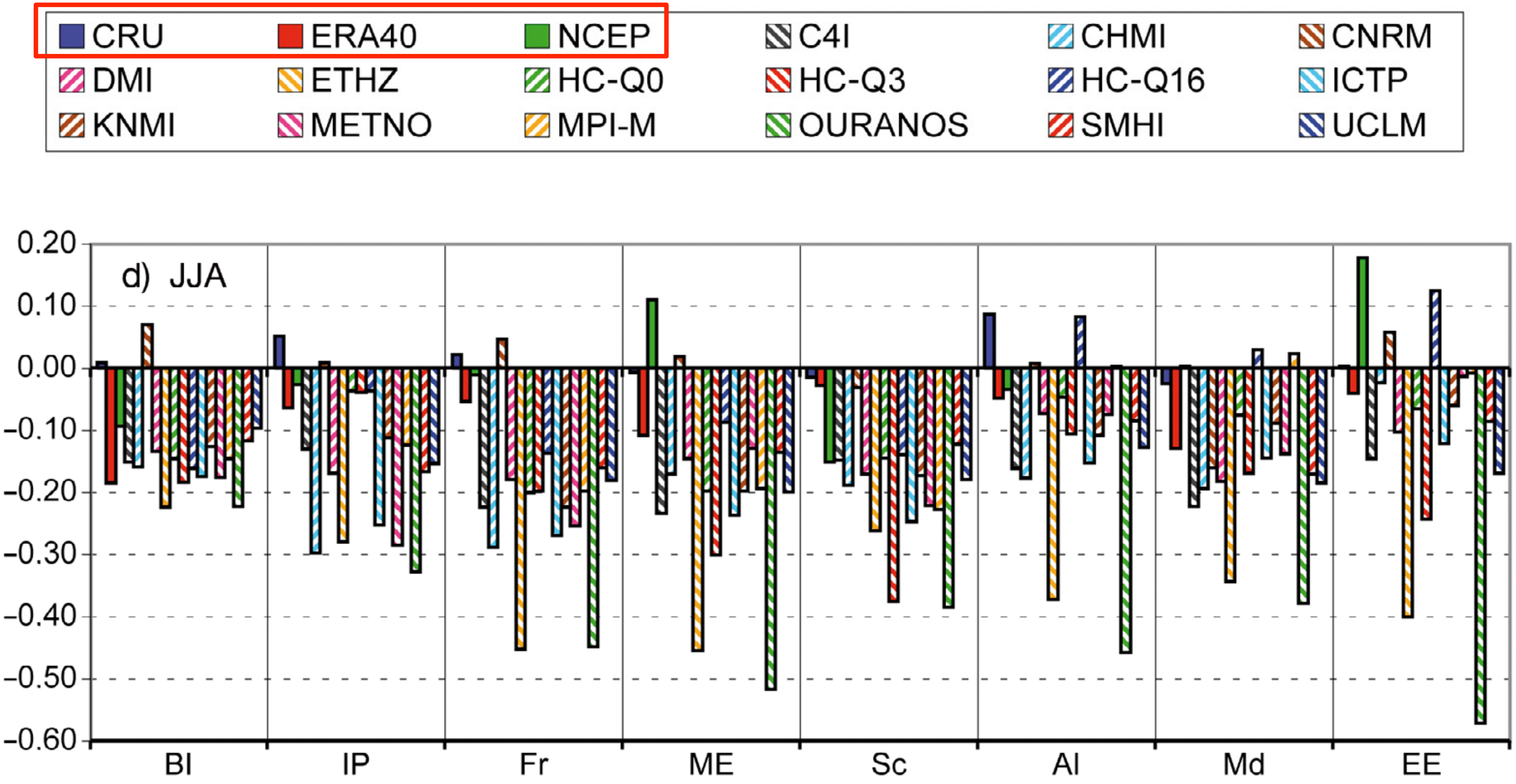



Fig. 1. Eight European sub-regions defined within the PRUDENCE project after Rockel (figure originally appeared in Christensen & Christensen 2007). 1: British Isles (BI), 2: Iberian Peninsula (IP), 3: France (Fr), 4: Mid-Europe (ME), 5: Scandinavia (Sc), 6: Alps (Al), 7: Mediterranean (Md), 8: Eastern Europe (EE)

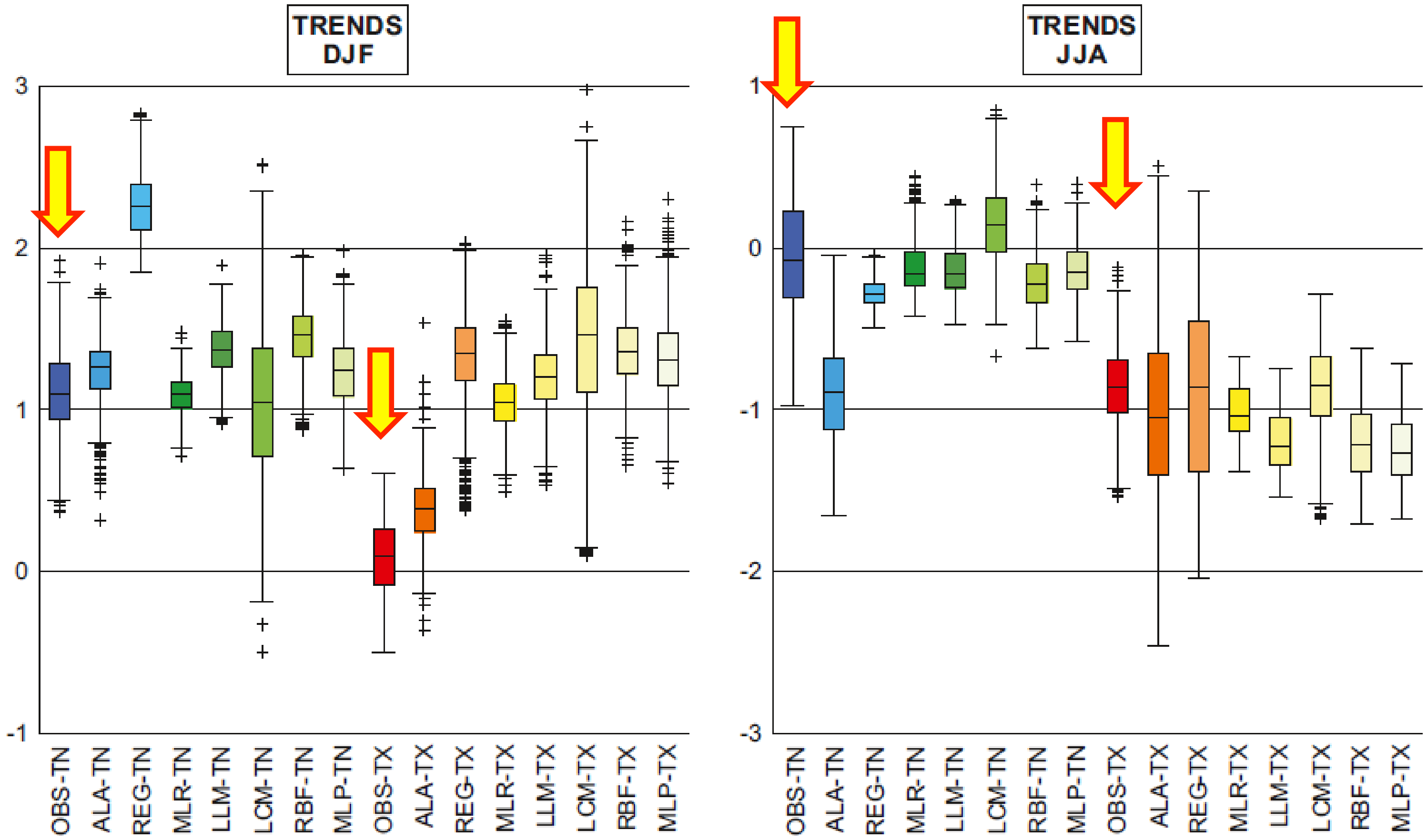
- trend difference (in °C / decade) from E-OBS
- note discrepancies between observed data / reanalyses



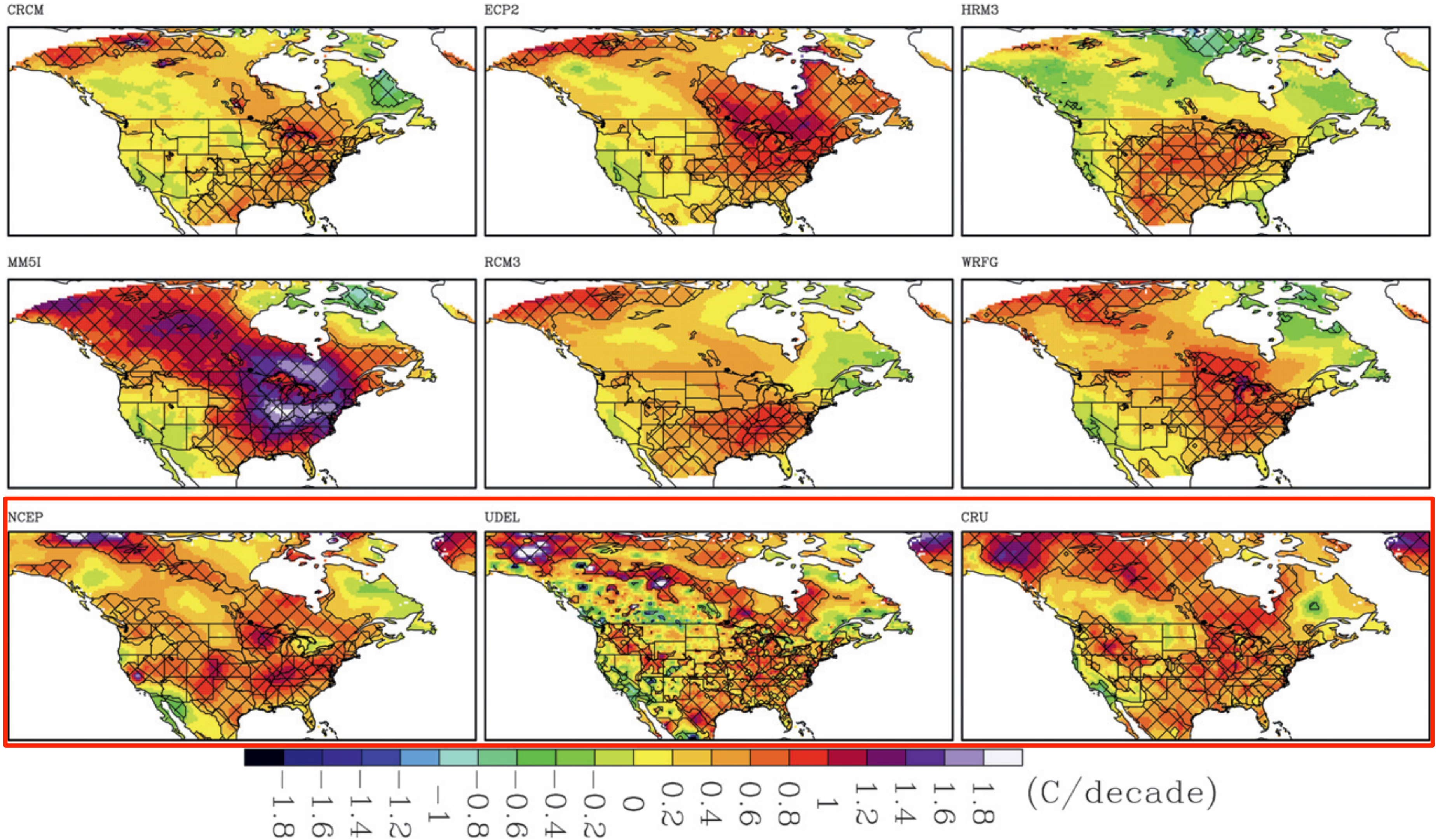
- trend difference (in °C / decade) from E-OBS
- note discrepancies between observed data / reanalyses



- trends (in °C / decade)



- trends (in °C / decade)
- DJF



- trends (in °C / decade)
- JJA

North American "warming hole"

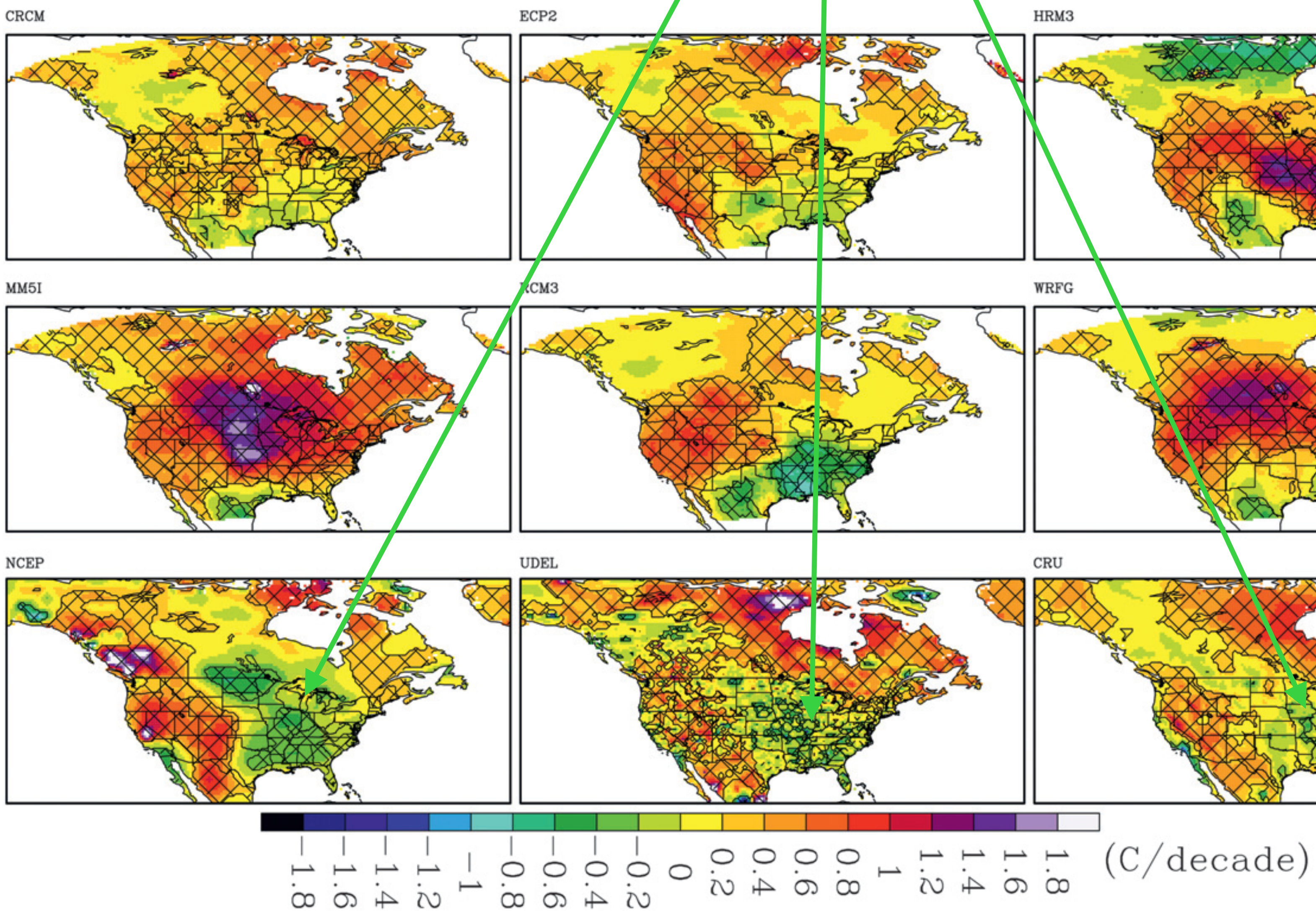


TABLE 1. Percentage of domain with trends that are statistically different from those in CRU.

| | DJF | MAM | JJA | SON |
|---|---|---|---|---|
| CRCM | 10.69 | 8.86 | 26.89 | 18.78 |
| ECP2 | 12.10 | 12.69 | 26.68 | 15.69 |
| HRM3 | 25.92 | 51.07 | 59.08 | 23.22 |
| MM5I | 26.66 | 36.12 | 53.25 | 19.08 |
| RCM3 | 7.31 | 7.02 | 40.52 | 34.00 |
| WRFG | 11.14 | 30.69 | 49.98 | 32.01 |
| NCEP-2 | 11.26 | 23.82 | 46.06 | 19.45 |
| UDEL | 11.21 | 12.71 | 28.15 | 14.37 |

- not a great success, is it …
- where do the differences from reality come from?
  - problems inside the models
  - imprecise reference climate data (trends differ between databases / reanalyses)
  - problems in the driving reanalyses (e.g. presence of artificial trends in upper level fields)
  - sampling variations
- difficult to distringuish model errors from other potential error sources

# persistence, DJF, Tmean

TABLE 7. Lag-1 autocorrelations ($\times 1000$) averaged over all stations for different downscaling methods (persistence) and their difference from observations (bias).

| Method | Persistence | Bias |
|---|---|---|
| Observed | 863 | — |
| Pointwise regression, inflation | 855 | −8 |
| Pointwise regression, white noise | 656 | −207 |
| Full regression, 3 PCs | 939 | +76 |
| Full regression, 5 PCs | 939 | +76 |
| Full regression, 7 PCs | 922 | +59 |
| Full regression, 11 PCs | 909 | +46 |
| CCA, 3/9 PCs, 3 modes | 889 | +26 |
| CCA, 5/9 PCs, 3 modes | 889 | +26 |
| CCA, 7/9 PCs, 4 modes | 884 | +21 |
| CCA, 11/9 PCs, 4 modes | 883 | +20 |
| CCA, 11/4 PCs, 4 modes | 899 | +36 |
| CCA, 11/7 PCs, 4 modes | 905 | +42 |
| CCA, 11/9 PCs, 4 modes | 883 | +20 |
| CCA, 11/9 PCs, 1 mode | 885 | +22 |
| CCA, 11/9 PCs, 2 modes | 875 | +12 |
| CCA, 11/9 PCs, 3 modes | 888 | +25 |
| CCA, 11/9 PCs, 4 modes | 883 | +20 |
| CCA, 11/9 PCs, 5 modes | 904 | +41 |
| CCA, 11/9 PCs, 6 modes | 900 | +37 |
| CCA, 11/9 PCs, 7 modes | 898 | +35 |

# persistence, DJF, Tmax

Table VIII. One day lag autocorrelations (×1000) for maximum temperature. Values within +/−0.040 from the observations are in bold; values below (above) this range are in italics (in light print). In the last column, the mean absolute error of the 1-day lag correlations (×1000) is shown for selected models.

| | Soda | Zugs | Sala | Bamb | Hohe | Vale | Smol | Prag | MAE |
|---|---|---|---|---|---|---|---|---|---|
| Observed | **717** | **737** | **737** | **825** | **778** | **707** | **810** | **850** | 0.0 |
| Pointwise regr. | **748** | 793 | **769** | **822** | **804** | **746** | **795** | **827** | 28.1 |
| 4 PCs regr. | 920 | 921 | 914 | 928 | 925 | 927 | 923 | 931 | – |
| 12 PCs regr. | 902 | 896 | 888 | 922 | 903 | 904 | 904 | 921 | – |
| 20 PCs regr. | 846 | 855 | 891 | 887 | 859 | 846 | 898 | 892 | – |
| NN pointwise | 825 | 855 | **756** | **796** | **783** | 767 | **840** | **838** | 47.6 |
| NN 20 PCs | 837 | – | 855 | – | 835 | – | – | **869** | – |
| T-mode, Z5, 4 cl. | **736** | 782 | **737** | **781** | **786** | **730** | *750* | *786* | 32.9 |
| T-mode, Z5, 11 cl. | *672* | 761 | *605* | *698* | **764** | **689** | *712* | *709* | – |
| T-mode, Z5, 18 cl. | *653* | 749 | *558* | *641* | **743** | *662* | *660* | *666* | – |
| T-mode, Z0, 4 cl. | **730** | 790 | **727** | *777* | **795** | **731** | *764* | *765* | 37.0 |
| k-means, Z5, 4 cl. | **730** | 780 | **721** | *779* | **789** | **712** | **781** | *786* | 28.4 |
| k-means, Z0, 4 cl. | **744** | 784 | **721** | *766* | **795** | **713** | 768 | *778* | 35.7 |