VALIDATION OF REGIONAL CLIMATE MODELS

Sven Kotlarski
Institute for Atmospheric and Climate Science, ETH Zurich

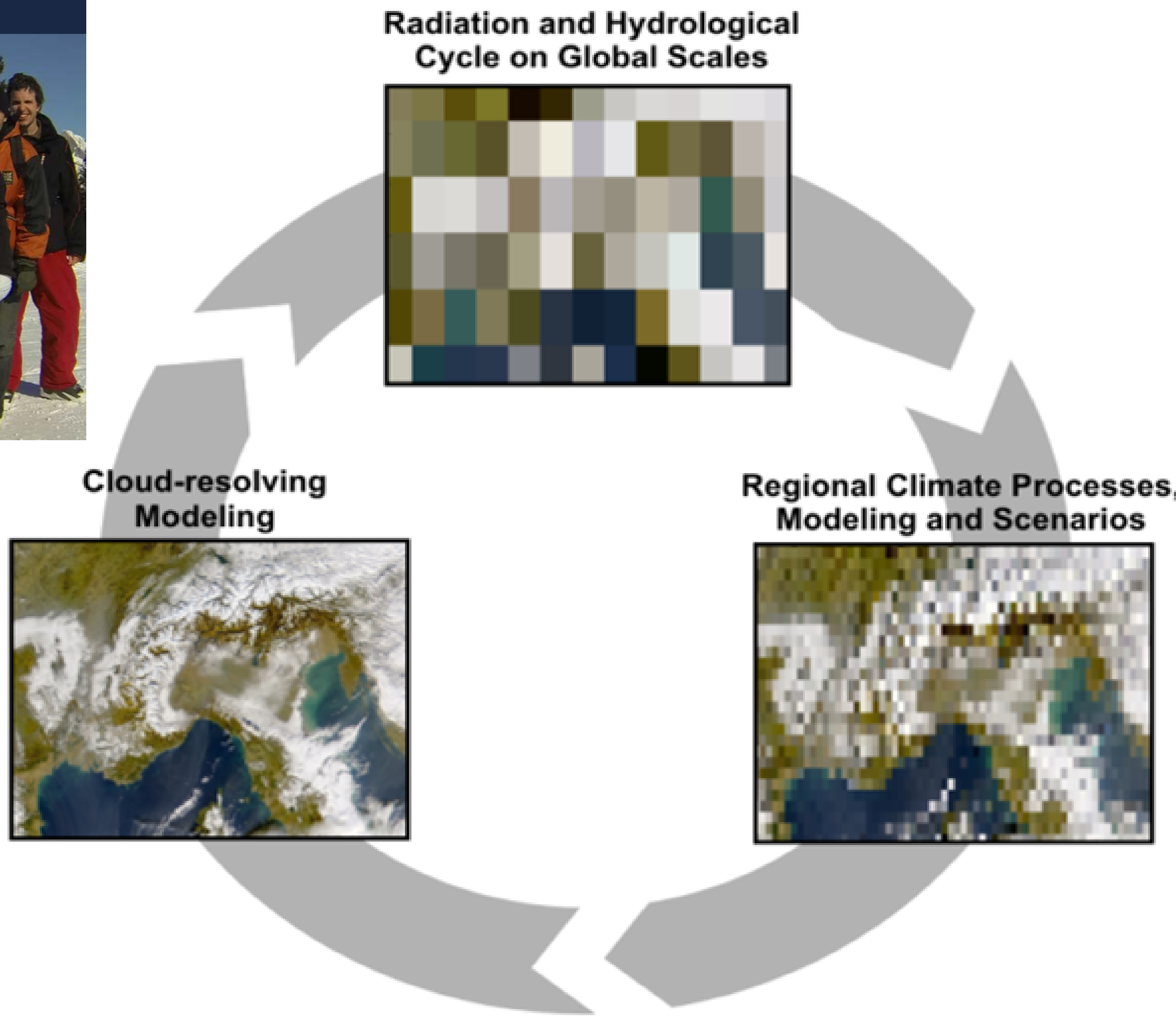
sven.kotlarski@env.ethz.ch

3rd VALUE Training School:
Spatial and Temporal Variability in Statistical and Dynamical Downscaling

ICTP Trieste, 03-08 November 2014

The «Climate and Water Cycle» Group at ETH Zurich (C. Schär)

http://www.iac.ethz.ch/groups/schaer



Main modelling tools:

ECHAM / MPI-ESM (global)

COSMO-CLM (regional)

Climate model evaluation as an important component of model development and application.

OUTLINE

1	REGIONAL CLIMATE MODELLING (WRAP-UP)	BLOCK 1
2	MODEL EVALUATION: THE RATIONALE	
3	APPROACHES	BLOCK 2
4	PERFORMANCE METRICS	
5	TO CONSIDER!	BLOCK 3
6	MODEL WEIGHTING	
7	EXAMPLE	BLOCK 4
8	SUMMARY & CONCLUSIONS	

OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE ME

5 TO C

DEL WEIGHTING

7 EXAMPLE

8 SUMMARY & CONCLUSIONS

BLOCK 1

CK 2

BLOCK 3

BLOCK 4

Questions welcome ANY TIME

OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE METRICS

5 TO CONSIDER!

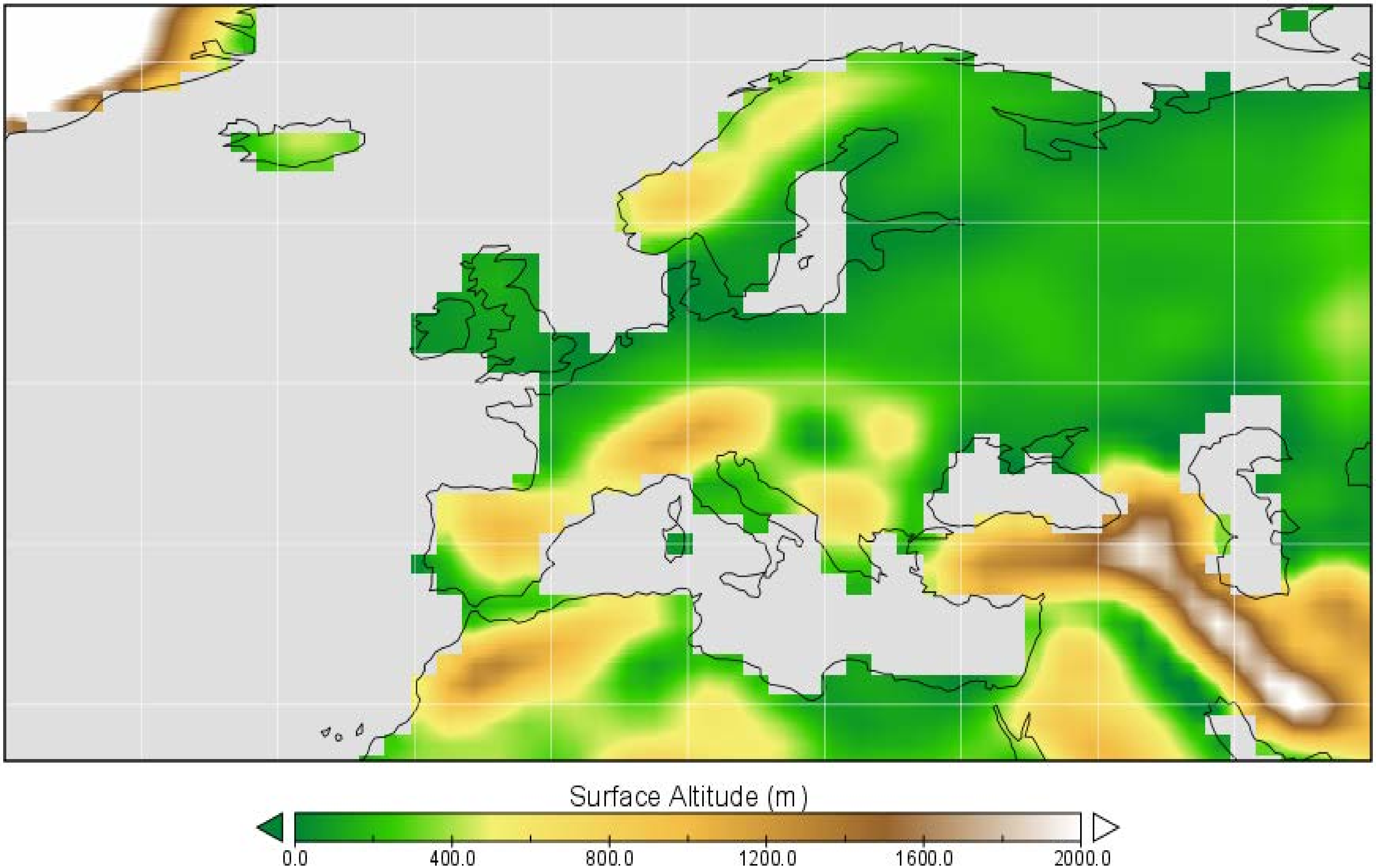
6 MODEL WEIGHTING

7 EXAMPLE

8 SUMMARY & CONCLUSIONS

GCM OROGRAPHY HadGEM2-ES, 1.875° x 1.25°, approx. 140 km

1 Regional Modelling
2
3
4
5
6
7
8



Piz Daint

CSCS Manno

Cray XC30

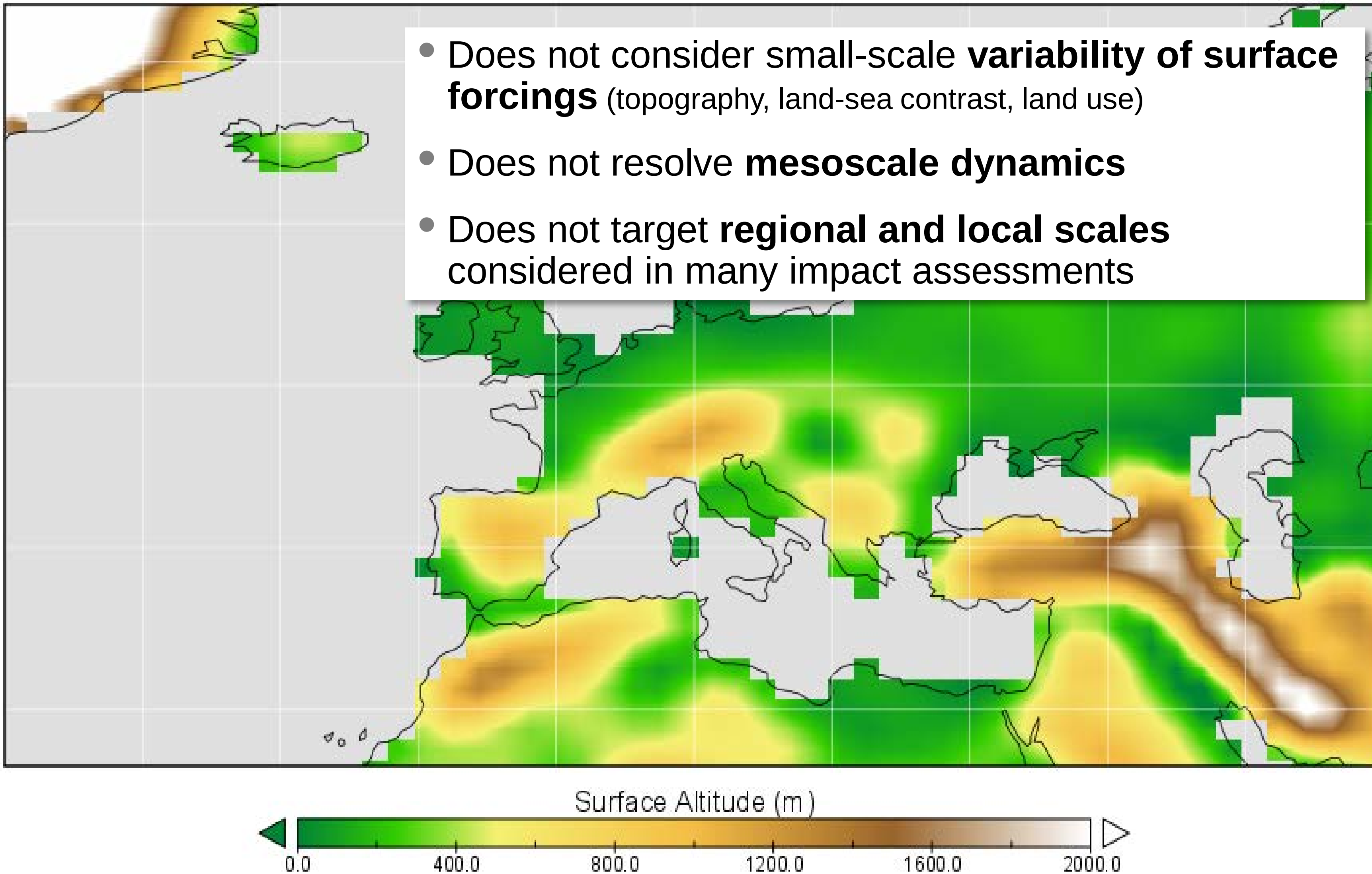
blizzard

DKRZ Hamburg

IBM power6

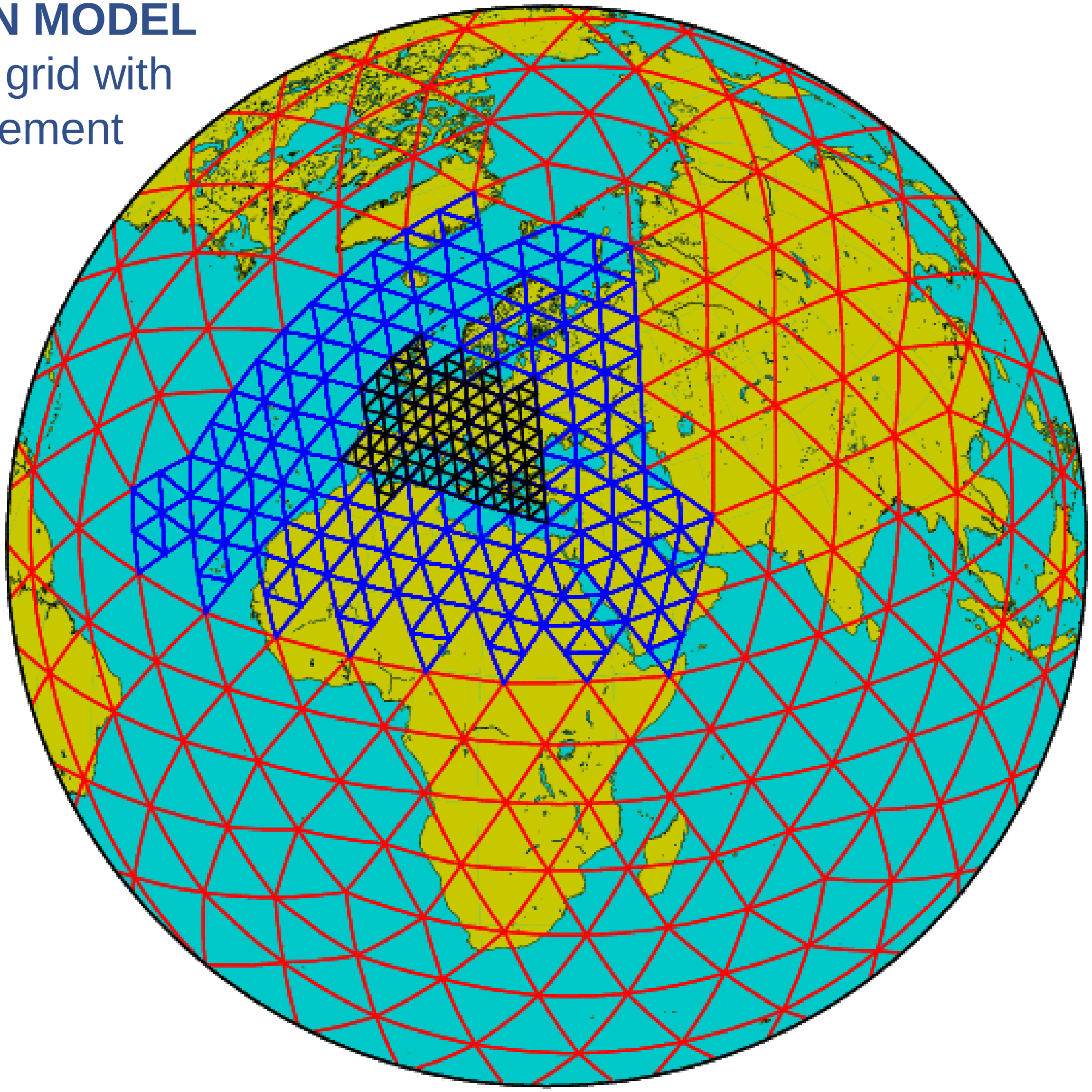
GCM OROGRAPHY HadGEM2-ES, 1.875° x 1.25°, approx. 140 km

- Does not consider small-scale **variability of surface forcings** (topography, land-sea contrast, land use)
- Does not resolve **mesoscale dynamics**
- Does not target **regional and local scales** considered in many impact assessments



VARIABLE-RESOLUTION GLOBAL CLIMATE MODEL

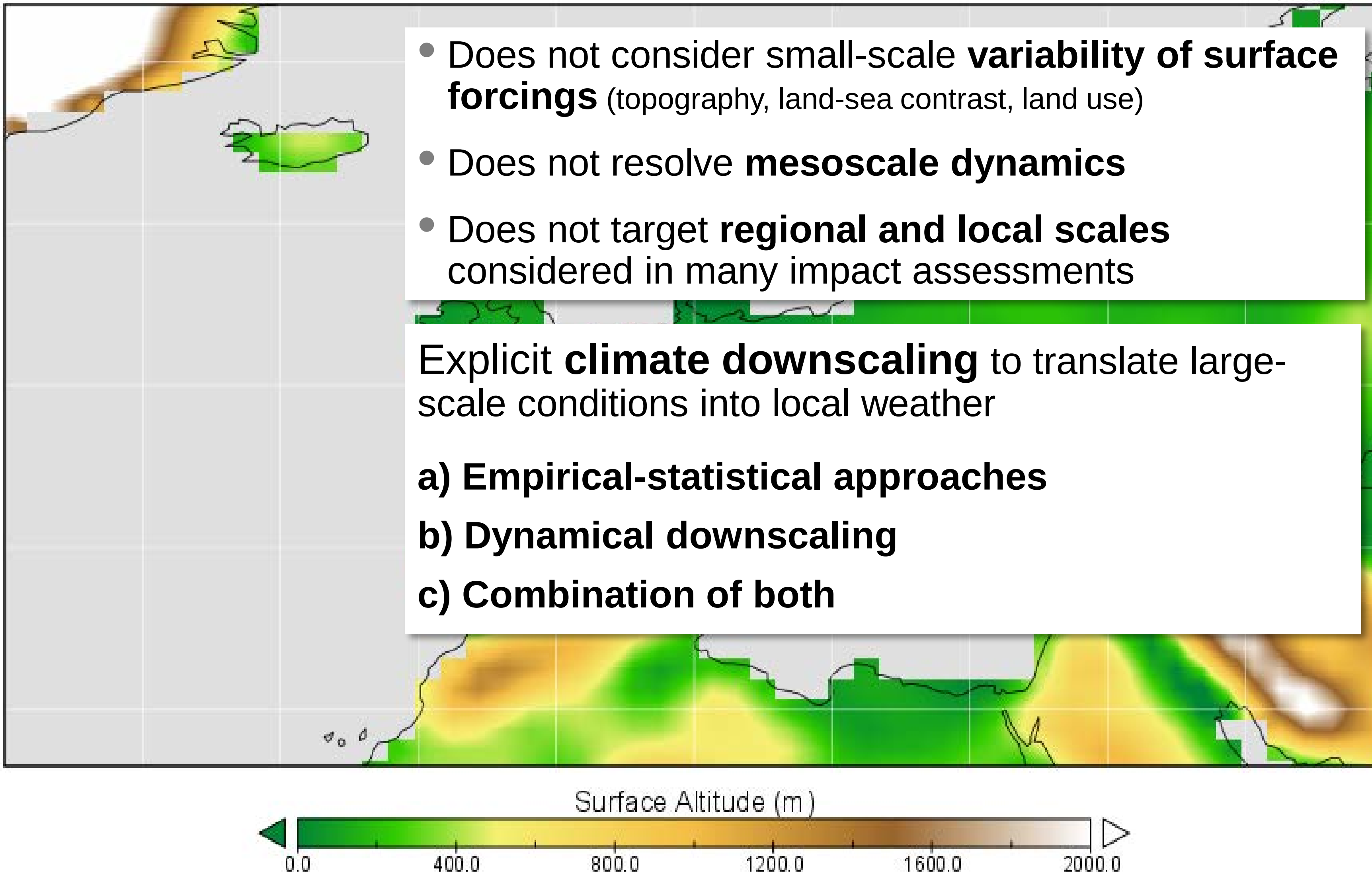
THE ICON MODEL
triangular grid with local refinement



Source: ***MPI-M Hamburg / ICON***

GCM OROGRAPHY HadGEM2-ES, 1.875° x 1.25°, approx. 140 km

1 Regional Modelling
2
3
4
5
6
7
8



- Does not consider small-scale **variability of surface forcings** (topography, land-sea contrast, land use)
- Does not resolve **mesoscale dynamics**
- Does not target **regional and local scales** considered in many impact assessments

Explicit **climate downscaling** to translate large-scale conditions into local weather

a) Empirical-statistical approaches

b) Dynamical downscaling

c) Combination of both

EMPIRICAL-STATISTICAL DOWNSCALING (classical view)

Large scale (flow) conditions

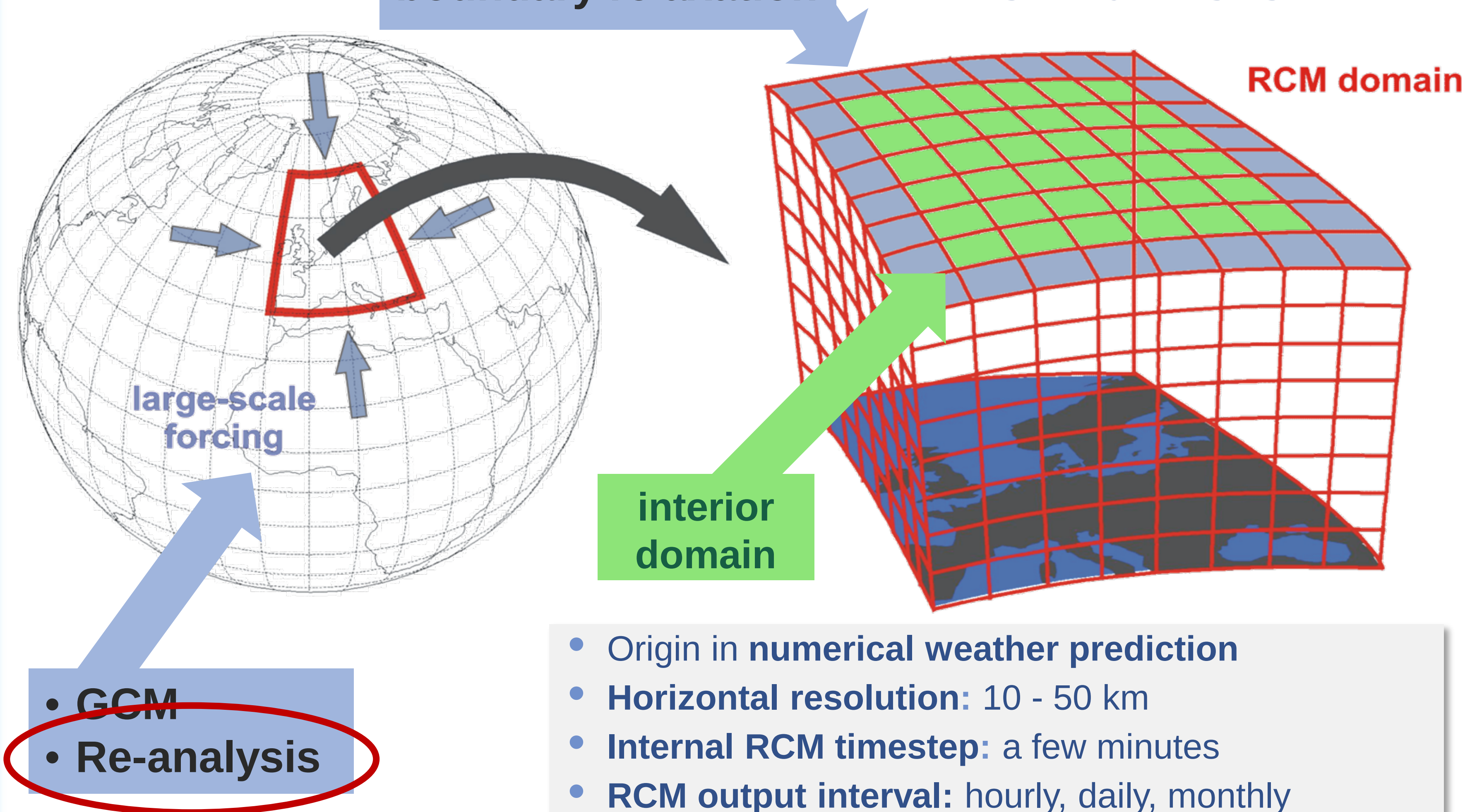
Regional / local scale conditions corresponding to large scale

1) Empirically derive a transfer function

2) Extrapolate transfer function into future

REGIONAL CLIMATE MODELLING

Apply a limited area model (**regional climate model, RCM**) as a "**magnifying glass**"...



- Origin in **numerical weather prediction**
- **Horizontal resolution:** 10 - 50 km
- **Internal RCM timestep:** a few minutes
- **RCM output interval:** hourly, daily, monthly
- Typically **one-way nesting** only

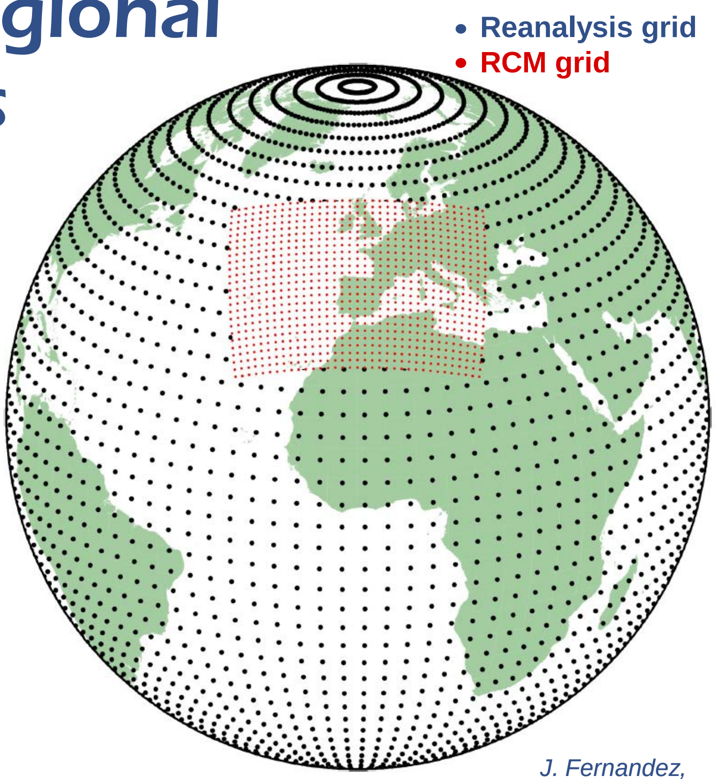
RE-ANALYSES: BASICS

- Systematic approach to produce data sets for climate monitoring and research
- Idea: Continuously assimilate observations (surface, radio sondings, remote sensing) into a weather/climate model and run this model forward in time -> reprocessing observational data spanning an extended historical period using a consistent modern analysis system
- Apply unchanging assimilation schemes and models («frozen»)
- Most reanalyses are global, but regional products at higher resolutions exist as well
- Besides atmospheric reanalyses further types exist (e.g., oceanic reanalyses)

RE-ANALYSES: PURPOSE

1
Regional Modelling
2
3
4
5
6
7
8

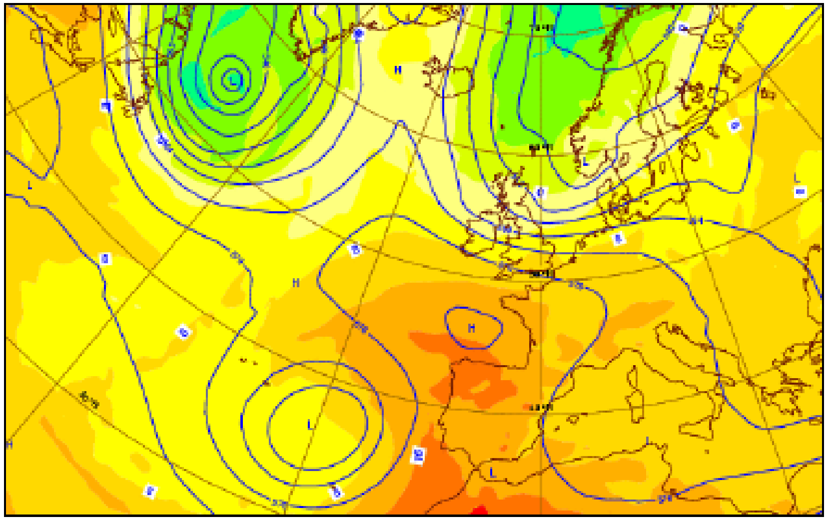
- Initialization of operational weather forecasts
- Climate analysis over historical periods
- Provision of initialization and boundary data for atmospheric limited area models (e.g., RCMs)
- **Validation of global and regional climate model experiments**
- Provision of atmospheric boundary conditions for, e.g., hydrological models
- e.g., ERA40, ERA-Interim, ERA-20C, JRA-55, MERRA, NCEP/NCAR
- «PERFECT BOUNDARIES»
- But: Re-analysis uncertainty!



J. Fernandez, Univ. Santander

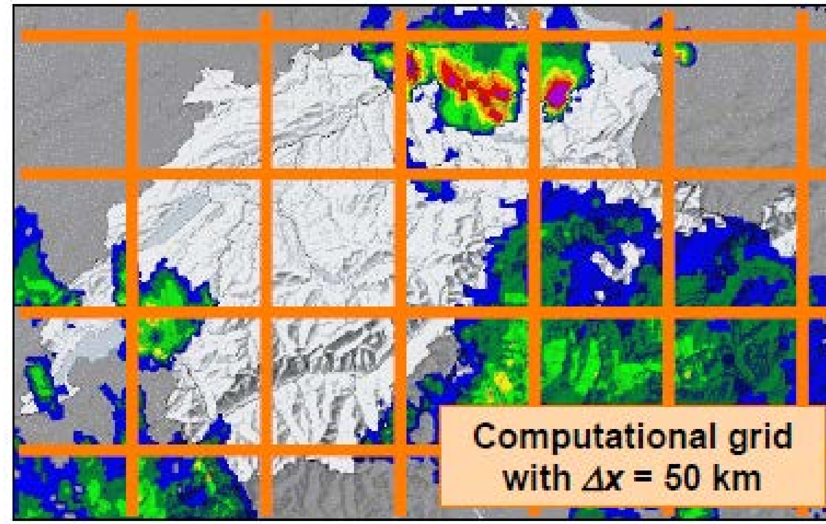
MODEL COMPONENTS

DYNAMICS



- Address the **resolved part** of atmospheric dynamics and thermodynamics.
- Solution of the **governing equations of fluid motion on a computational grid**
- Examples of resolved structures: **general circulation of atmosphere, low and high pressure systems, mountain flows**

PHYSICS



- Representation of **unresolved scales** by **parameterizations** (sub-grid)
- Typically contain **empirical** components and are to some extent **tuned/calibrated**
- Major source of **model uncertainty**
- Examples of parameterized processes: **boundary layer, convection, precipitation, clouds, land surface**

Courtesy: C. Schär, ETH Zurich

PROS & CONS

1
Regional Modelling
2
3
4
5
6
7
8

+

- Physically consistent response, including climate feedbacks
- Application of models for future periods possible (in principle)

–

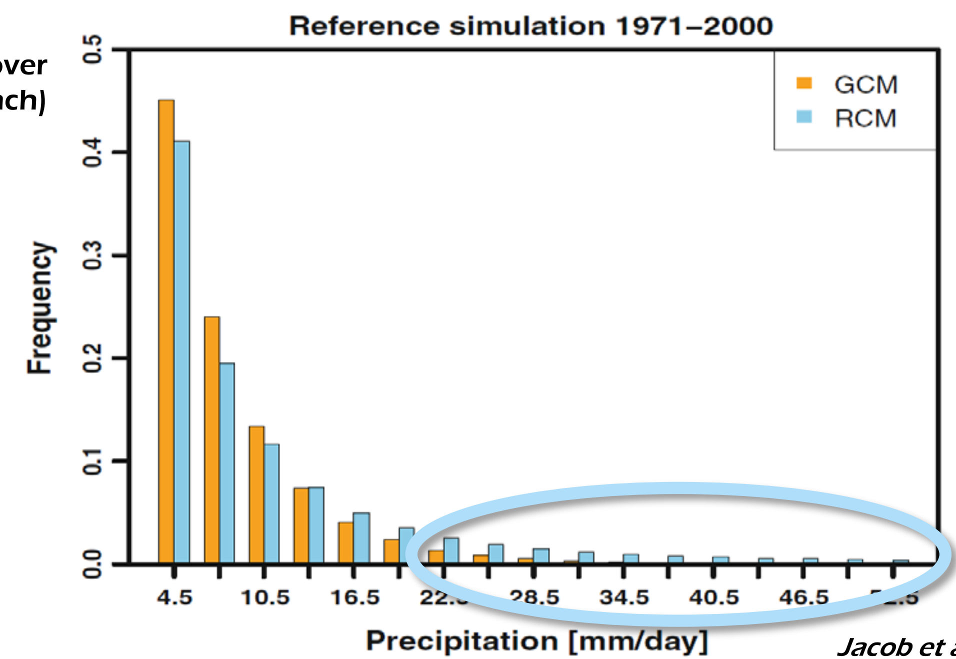
- Computationally expensive
- Advanced expertise required
- Limited number of realizations
- Limited spatial resolution (does not target the site scale)
- Strongly depends on driving GCM (*garbage in – garbage out*)
- "Added value" wrt. GCM not always apparent (found, e.g., in high-order statistics reflecting intense and localized events)

THE ADDED VALUE

1 Regional Modelling
2
3
4
5
6
7
8

- *An RCM won't improve all aspects of a GCM simulation*
- *Added value often hard to find for time-averaged quantities or on large spatial scales*
- *Most likely in **frequency distributions and high-order statistics reflecting intense and localized events** (e.g. tails of daily precipitation intensity distribution; e.g. Jacob et al. 2013)*
- *Added value on scales that are common to both the RCM and the driving GCM?*

Frequencies of daily precipitation intensities over central Europe (ensembles of five simulations each)



Jacob et al. 2013

TYPES OF RCM EXPERIMENTS

boundary forcing (global)

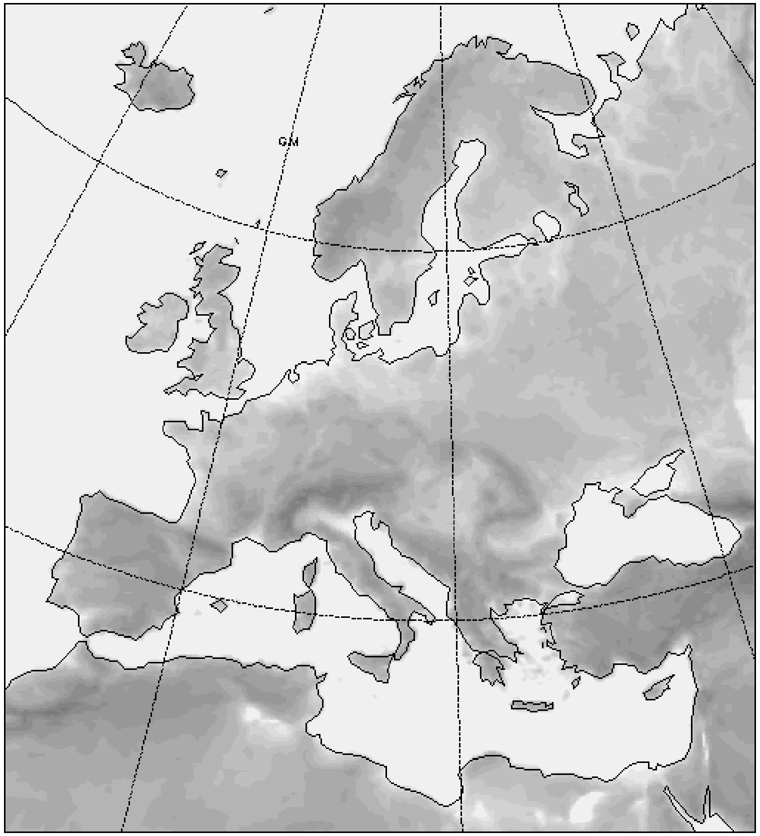
Re-analysis ***(perfect boundaries)***

GCM historical GHG

GCM future GHG

Re-analysis/ GCM Idealized setups

RCM



EVALUATION RUN

Evaluation of downscaling

CLIMATE SCENARIO

Evaluation of GCM-RCM chain

Climate change

SENSITIVITY RUN

Sensitivities, process understanding

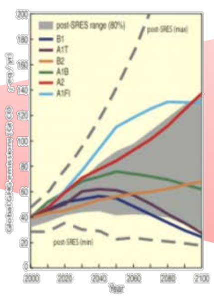
THE UNCERTAINTY CASCADE

1 **Regional Modelling**

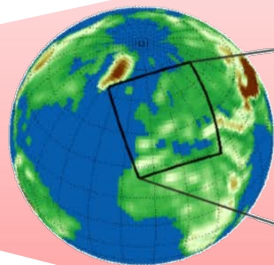
2
3
4
5
6
7
8

Statistical-empirical methods

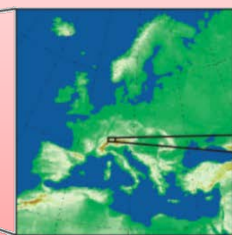
Greenhouse gas and aerosol scenario



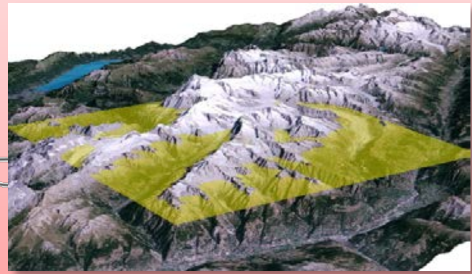
GCM
(choice and setup)



RCM
(choice and setup)



Interface to impact model
(further downscaling and postprocessing)



Impact model

Completeness?
External forcings?
Internal variability

Ensemble approaches
to quantify and constrain uncertainties

CLIMATE MODEL ENSEMBLES

EMISSION SCENARIO ENSEMBLES

- **Carry out multiple projections assuming different emission scenarios**

MULTI MODEL ENSEMBLES

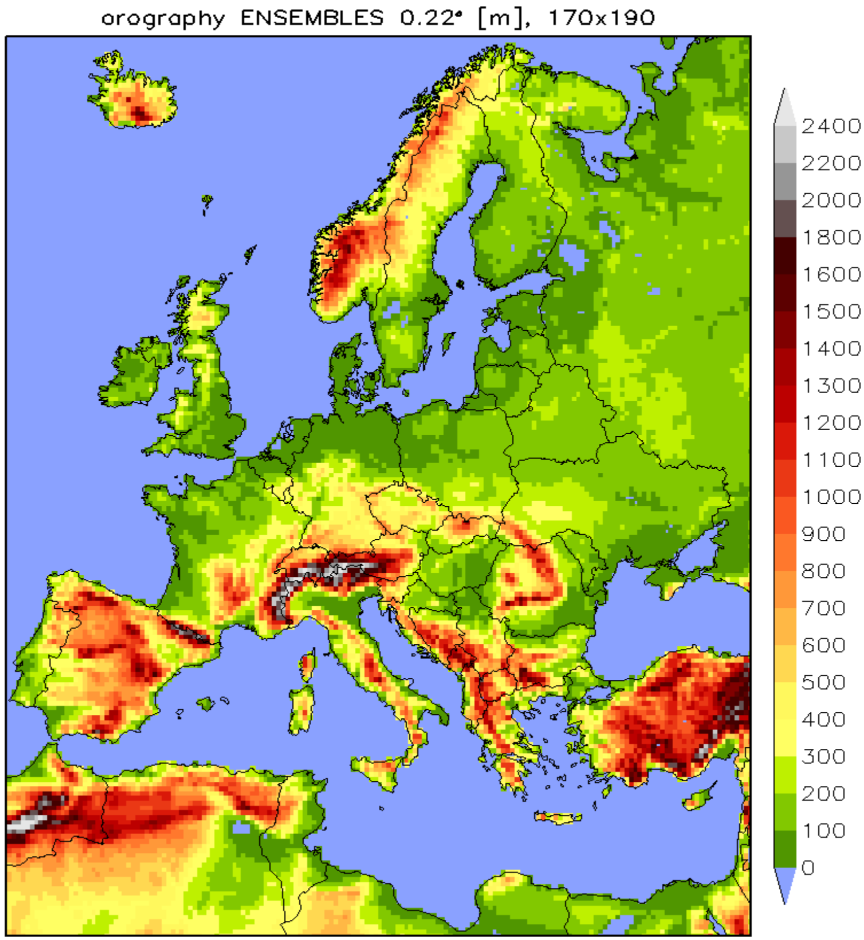
- **Combine multiple projections from different models**
- **Ideally: models independent of each other** (typically not given!)
- **Intermodel variability as a measure of uncertainty** (spread of projections)

PERTURBED PHYSICS ENSEMBLES

- **Combine different simulations of the <u>same</u> model but with perturbed versions of the original model physics**
- **More systematic sampling possible** (multi model ensembles: *opportunistic* ensembles)
- **Intramodel variability as a measure of uncertainty**
- **e.g. *climateprediction.net***

THE ENSEMBLES PROJECT

- EU FP6 research programme 2004-2009
  http://ensembles-eu.metoffice.com
- Setup of an ensemble prediction system for climate change in Europe
- Regional component: Application of 17 RCMs at 25 and 50 km resolution
- **ERA40-driven evaluation runs**
- **Regional climate scenarios (multi GCM-multi RCM)**

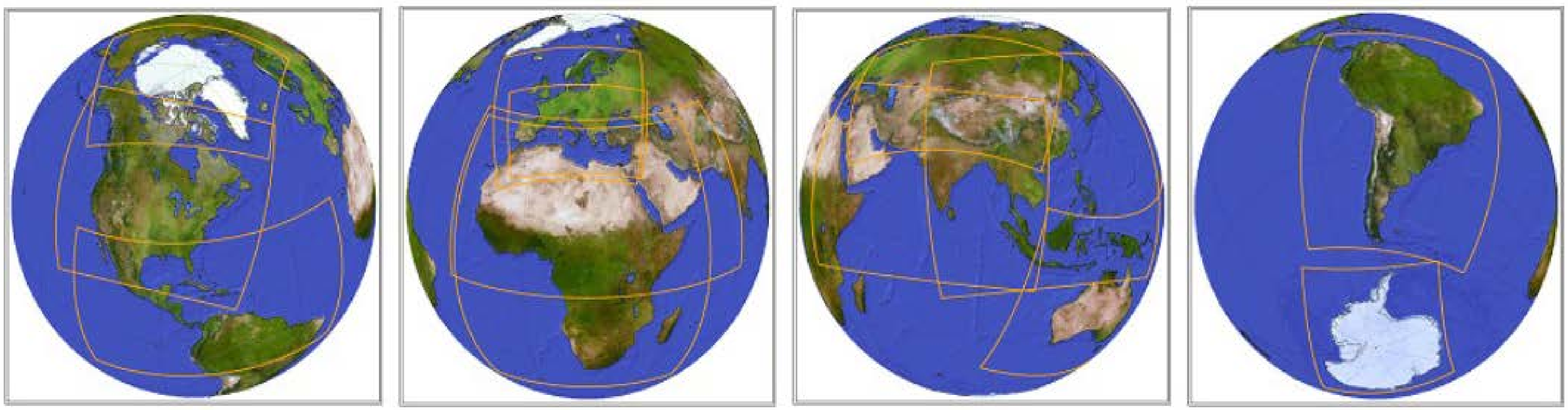


- Rather few scenarios (15 until 2100)
- Only SRES A1B considered
- Horizontal resolution too coarse for many applications

CORDEX

Coordinated Regional Climate Downscaling Experiment

International framework for next generation of regional climate change projections for all terrestrial
regions of the globe *(http://wcrp-cordex.ipsl.jussieu.fr)*



http://wcrp-cordex.ipsl.jussieu.fr

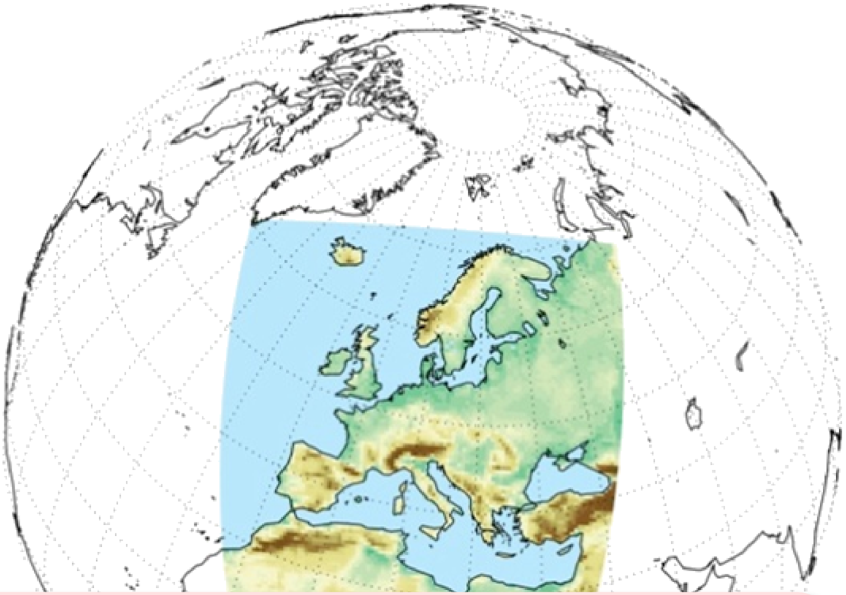
The CORDEX community has grown to now include 14 domains;

- Includes dynamical and statistical downscaling approaches
- Forcing: **CMIP5 GCMs** assuming *Representative Concentration Pathways* **(RCPs)**
- Common RCM resolution: **50 km**, focus domain: **Africa**

EURO-CORDEX

1
Regional Modelling
2
3
4
5
6
7
8

- European branch of CORDEX
  http://www.euro-cordex.net
- Community currently consists of **29 modelling centers** applying **10 different RCMs**
- Experiments at **50 km** and **12 km** for European domain



Evaluation runs

forcing:
ERA-Interim **(1989-2008)**

Climate scenarios

forcing: CMIP5 GCMs **(1951-2100)**

50km: 66 simulations (10 RCMs, 12 GCMs, 3 RCPs)

12 km: 42 simulation (9 RCMs, 7 GCMs, 3 RCPs)

About 1/3 of experiments currently available **on ESGF archive**
(e.g. *http://esgf-data.dkrz.de*)

ENSEMBLES vs. EURO-CORDEX

ENSEMBLES

EURO-CORDEX

- ***Updated GCMs/ESMs***
- ***Updated RCMs***
- ***RCPs***
- ***Higher grid resolution (for 12 km)***
- ***Much larger ensembles***

- High resolution **versus** ensemble size
- **150 years on 12 km: 2.5 mio CPUh**
- **150 years on 50 km: *1/33* of it**

- Uncertainties (still) not fully sampled

OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE METRICS

5 TO CONSIDER!

6 MODEL WEIGHTING

7 EXAMPLE

8 SUMMARY & CONCLUSIONS

WHY SHOULD WE VALIDATE AN RCM?
(or a climate model in general)

WHY RCM EVALUATION?

Note: *Also applies to GCMs and many other kinds of models!*

DOES THE MODEL WORK FOR THE PURPOSE IT HAS BEEN BUILT FOR?

- RCMs as (INCOMPLETE!) *mathematical representations* of the regional climate system
- Based on physical principles, but subject to **structural and parametric uncertainties**
- Check: **Can the model approximately reproduce past climatologies and past climatic trends?**
- **Basic requirement for trust in regional climate scenarios** (that potentially have major implications)
- But: *Any* performance threshold is subjective; only broad picture can be provided

WHY RCM EVALUATION? (cont'd)

1
2 Rationale
3
4
5
6
7
8

MODEL SELECTION AND WEIGHTING

- Basis for weighting individual models in multi-model ensembles
- But: Weighting usually based on subjective performance scores and potentially dangerous (Weigel et al., 2010)
- If several models are available but only one (or a few) can be afforded to run: Evaluation can inform selection to some extent
- Basis for excluding models with major deficiencies

MODEL SETUP AND CALIBRATION

- Chosing a specific setup of an RCM for a given application (domain, timestep, parameterization options, ...)
- Parameter calibration within a specific setup («model tuning»)

WHY RCM EVALUATION? (cont'd)

ADDED VALUE ANALYSIS

- RCM application costly (both manpower and computing time)
- Especially true for high spatial resolution:
  1:33 for 12 km vs. 50 km
- Model evaluation to inform decision whether this investment is reasonable or not
- But: Evaluation does not tell the entire story (added value might only appear in the scenarios)

IDENTIFICATION OF MODEL DEFICIENCIES

- Evaluation can highlight deficiencies of a particular model
- But: Does not necessarily highlight the physical reasons for biases

WHY RCM EVALUATION? (cont'd)

MODEL DEVELOPMENT

- Evaluation of newly introduced model components
- Usual procedure: (1) stand-alone / idealized mode
  (2) fully interactive mode
- Might require re-calibration!

RCM VERSUS SD EVALUATION

Compared to SD evaluation, RCM evaluation ...

... should not be carried out a the point scale but at the RCM grid cell scale or coarser (**scale mismatch**)

... should always be carried out for a larger area (multiple grid cells or entire model domain; **«global» calibration**)

... can typically not be carried out event-wise

... should – if possible – target **physical relationships** in order to be informative

OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE METRICS

5 TO CONSIDER!

6 MODEL WEIGHTING

7 EXAMPLE

8 SUMMARY & CONCLUSIONS

TYPES OF RCM EXPERIMENTS

«Run your model for some period in the past and check the performance.»
(**Not as trivial as it seems!**)

boundary forcing (global)

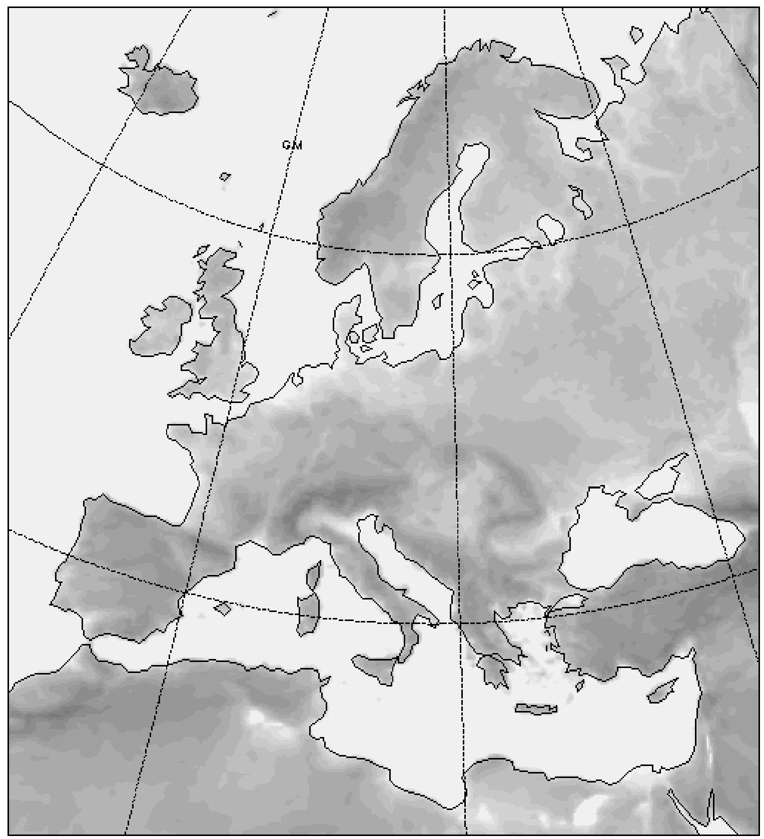
Re-analysis
(perfect boundaries)

GCM
historical GHG

GCM
future GHG

Re-analysis/ GCM
Idealized setups

RCM



EVALUATION RUN
Evaluation of downscaling

CLIMATE SCENARIO
Evaluation of GCM-RCM chain

Climate change

SENSITIVITY RUN
Sensitivities, process understanding

TYPES OF EVALUATION

1
2
3 Approa-ches
4
5
6
7
8

EVALUATION RUN
(Re-analysis driven)

- Assumption of «perfect boundaries»
- Separation of downscaling performance from biases due to erroneous large-scale forcing
- Temporal correspondence on large temporal and spatial scales

SCENARIO RUN
(GCM-driven historical)

- Evaluation of combined GCM-RCM chain
- RCM results strongly influenced by errors in the boundary forcing («garbage in – garbage out»)
- **No temporal corresponden-ce!** (especially if driven by AOGCM)

SENSITIVITY RUN

- Scope of evaluation strongly depends on specific setup
- Typically physical-based evaluation
- Reference: often another simulation of the same model

REFERENCE

THE REFERENCE

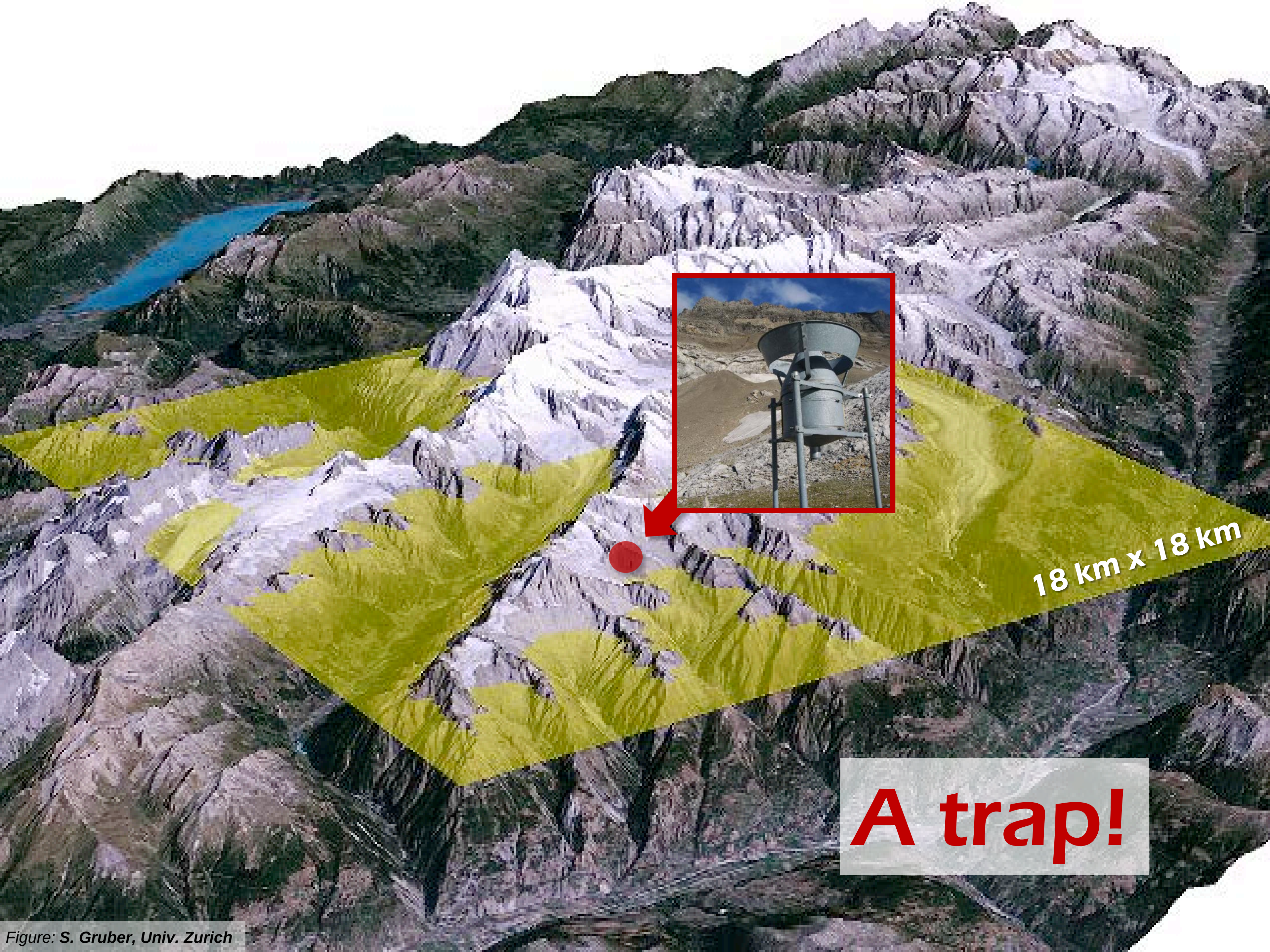
1
2
3 Approa-ches
4
5
6
7
8

- «Observations» in historical periods
  (typically involves models and assumptions)

- A different model that you trust in
  (could be, for instance, a re-analysis or a model based on first physical principles)

- A reconstruction of the historical climate (especially applies to paleoclimate studies)

- A reference simulation of the same model


18 km x 18 km
A trap!
Figure: S. Gruber, Univ. Zurich

SCALE MISMATCH

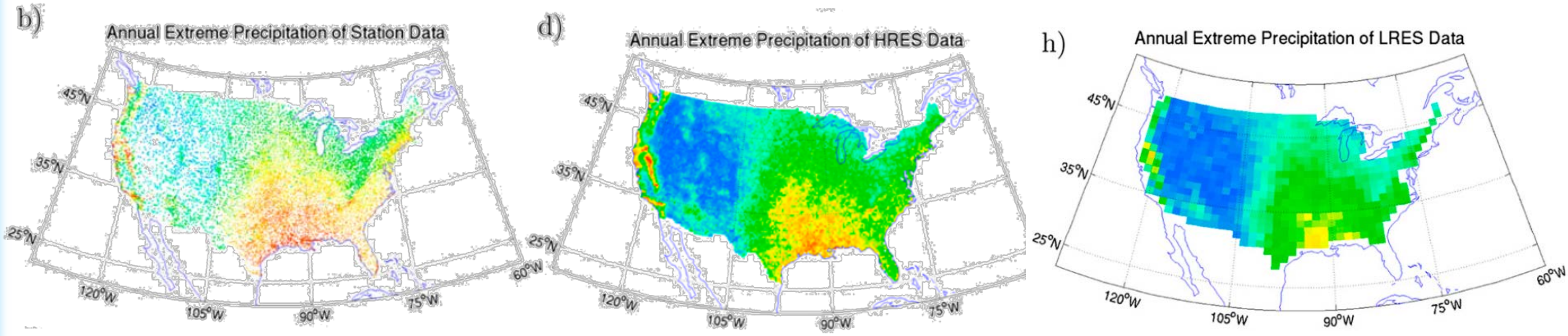
1
2
3
Approa-
ches
4
5
6
7
8

- RCMs operate on grid cell scale
- Output typically needs to be interpreted as «mean over grid cell area»

- **Scale mismatch** when comparing gridded model results to measurements at individual stations
  - Smoothing of spatial variability
  - Smoothing of (localized) extremes, especially precipitation and winds
  - Elevation and slope effects in topographic terrain
  - Neglect of subgrid variability (as, for instance, introduced by land surface characteristics)

SCALE MISMATCH (cont'd)

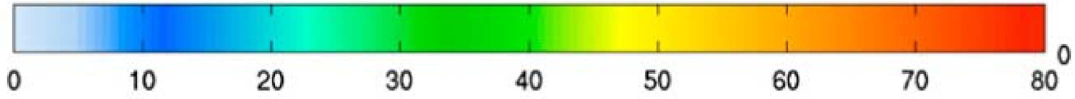
97th percentile of wet-day precipitation (1979-2003):
Stations vs. grids



GHCN stations

Gridded to 0.25°
(Cressman interp.)

Remapped to 0.9° x 1.25°
(Conservative remapping)



Gervais et al., 2014

GRIDDED REFERENCE DATA

Use of gridded reference data

A) Station measurements interpolated onto a regular grid

- Measurements and interpolation subject to considerable uncertainties! (see later)

B) Re-analysis products

- Observations only indirectly represented (data assimilation)
- Uncertainties due to assimilation scheme, re-analysis model and changing mix of underlying observational data
- For instance: introduction of satellite data in 1970s

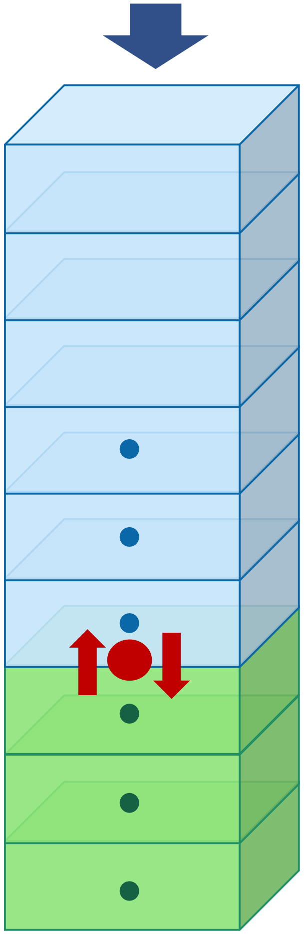
C) Remote sensing products

- Also involve models and assumptions (e.g. radiative transfer)
- Good spatial, but typically limited temporal coverage

EXCEPTION: SINGLE-COLUMN MODES

1
2
3 Approa-ches
4
5
6
7
8

For development / refinement of parameterizations an RCM (or parts of it) is often operated in a special **single-column mode** (just one single soil/ atmospheric column is considered, no horizontal dependencies)



Single-station reference data often useful!
(soil temperatures, snow cover, surface fluxes, air temperatures at different heights, ...)

- Especially applies to land-surface parameterization schemes
- «Controlled» boundary conditions
- Idealized prescribed (observed) forcing of column
- Physiographic parameters close to those observed

GRIDDED REFERENCE DATA

Use of gridded reference data

A) Station measurements interpolated onto a regular grid

- Measurements and interpolation subject to considerable uncertainties! (see later)

B) Re-analysis products

- Observations only indirectly represented (data assimilation)
- Uncertainties due to assimilation scheme, re-analysis model and changing mix of underlying observational data
- For instance: introduction of satellite data in 1970s

C) Remote sensing products

- Also involve models and assumptions (e.g. radiative transfer)
- Good spatial, but typically limited temporal coverage

Regridding still necessary in most cases!
(matching RCM and reference data resolution)

1
2
3
Approa-
ches
4
5
6
7
8

THE E-OBS GRIDDED DATASET

http://www.ecad.eu/download/ensembles/ensembles.php

- Daily gridded **mean/mix/max temperature**, **precipitation**, **sea level pressure** for Europe
- Developed within the EU ENSEMBLES project
- 1950 - 2013 (v10)
- Four different resolutions:
  - 0.25° regular
  - 0.5° regular
  - 0.22° rotated
  - 0.44° rotated

  ENSEMBLES RCM grid
- Underlying station time series available for most parts
- Standard reference for RCM evaluation over Europe

Shortcomings:
- Low station density over many areas (nominal resolution < effective resolution; smoothing of extremes)
- Even nominal resolution coarser than recent high-resolution EURO-CORDEX runs (0.11°)
- Temporal inhomogeneities

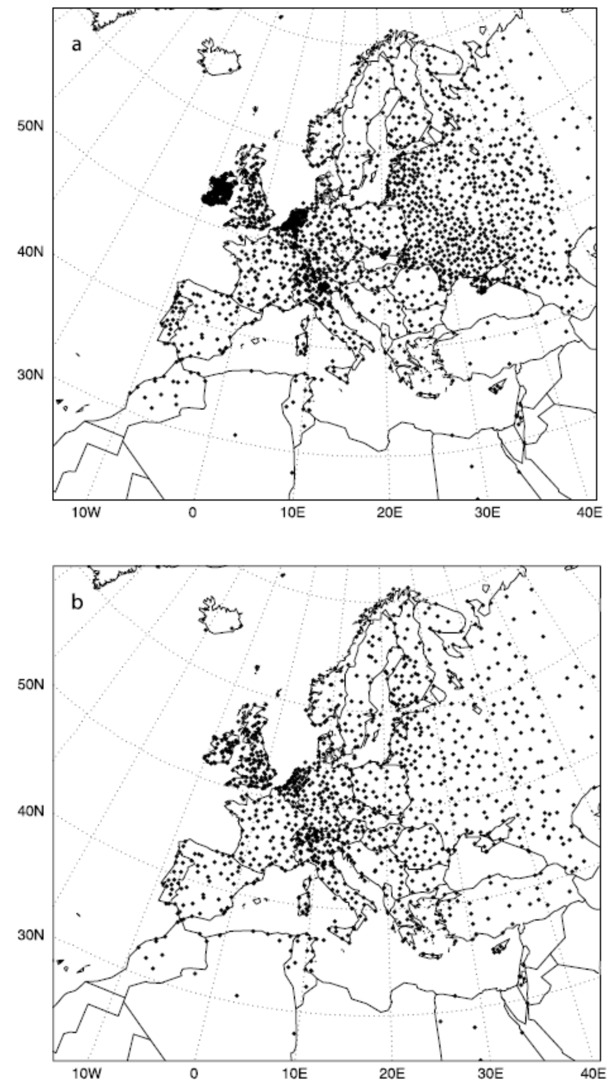
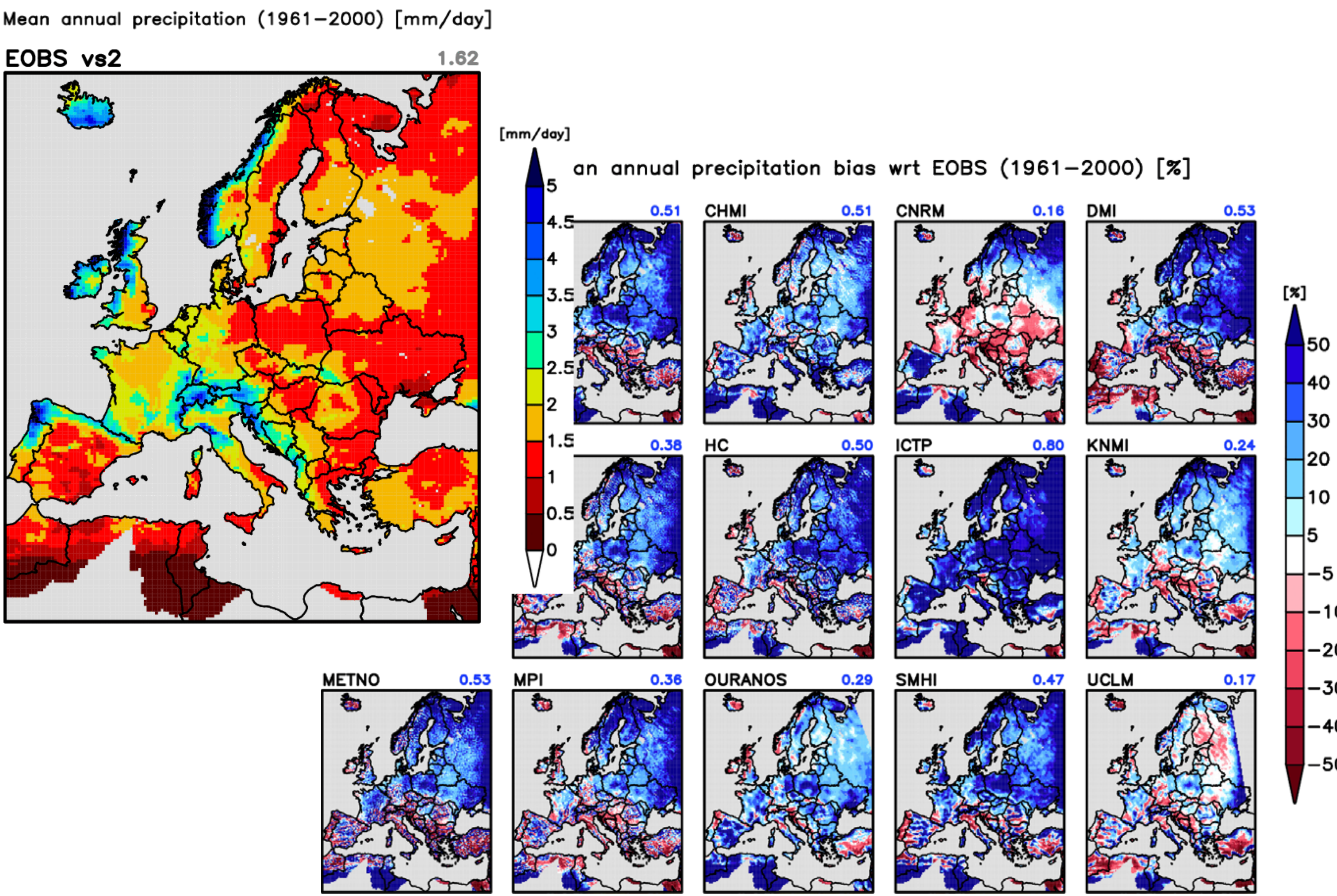


Figure 1. The complete gridding region (land-only), showing the station network for (a) precipitation and (b) mean temperature.

Haylock et al., 2008

THE E-OBS GRIDDED DATASET (cont'd)

1
2
3
Approa-
ches
4
5
6
7
8



OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE METRICS

«Performance measures», «Skill scores», «Performance score», «Evaluation metrics»

5 TO CONSIDER!

6 MODEL WEIGHTING

7 EXAMPLE

8 SUMMARY & CONCLUSIONS

SCOPE

SIMULATION → Comparison ← REFERENCE

↓

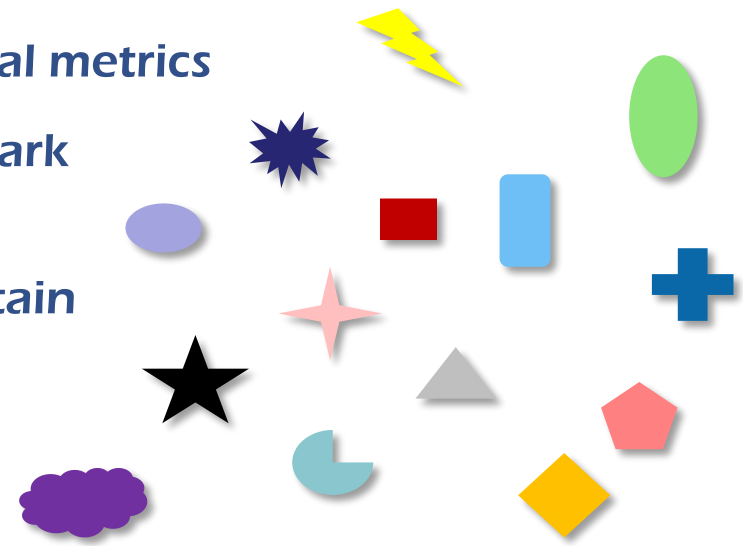
Performance metric

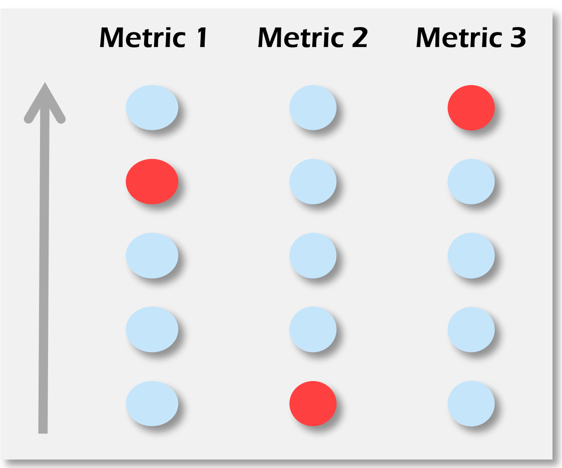
- Metrics should measure/quantify the model performance against a given reference dataset for a specific aspect: «Is the model able to simulate things we have observed?»
- Combined scores (accounting for several aspects / variables) possible
- Usually not desgined to diagnose reasons for model errors
- Ideally, a metric should allow a comparison of the performance of different models («good performance» … -> … «bad performance»): scalar quantity
- Also: Assessment of temporal and spatial variability of performance of a given model
- Also: Assessment of performance of different setups of a given model

METRIC SELECTION

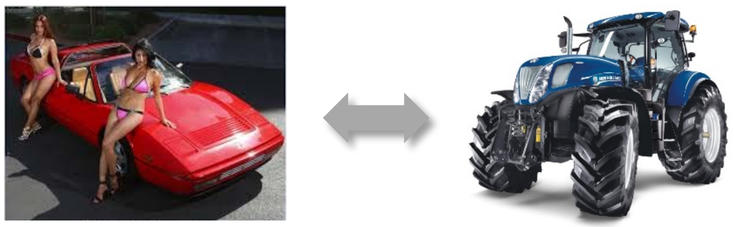
1
2
3
4
Measures
5
6
7
8

- «Metric Zoo»: Infinite number of potential metrics
- No well-defined common set of benchmark metrics; but several «standard» metrics
- One single metric ALWAYS neglects certain aspects of model performance
- RCM: Metrics typically consider climatology or trend!





- Subjective choice
- Outcome of evaluation exercise typically strongly depends on metric
- Concept of one-best model is ill-defined! (but there may be a best model for a given purpose)



METRIC SELECTION (cont'd)

APPLICATION-DRIVEN

«I'm only interested in mean annual temperature, therefore my metric should only consider performance wrt. mean annual temperature.»

«I'm only interested in the Alps, therefore my metric only needs to consider model performance in this region»

Often easy to carry out.

But potentially dangerous:
Compensating errors might indicate good model performance.

Provides little evidence whether or not the physics are well represented.

PHYSICS- AND PROCESS-RELATED

Assess model performance with respect to the representation of physical processes.

Typically requires to include more than one variable.

Typically more relevant for obtaining trust in a model especially wrt.

Probably more relevant for climate change signals.

Often limited availability of reference data.

Final scoring can be tricky (=uncertain!)

METRIC SELECTION (cont'd)

To be informative, performance metric(s) in model evaluation should …

… cover a wide range of aspects of model performance

… consider several variables
(standard: T and P only)

… consider a larger domain

… consider observational uncertainty

… be transparent

GENERIC AND APPLICATION-SPECIFIC ASPECTS

GENERIC ASPECTS OF MODEL PERFORMANCE

- Capturing a mean climatology
- Capturing trends

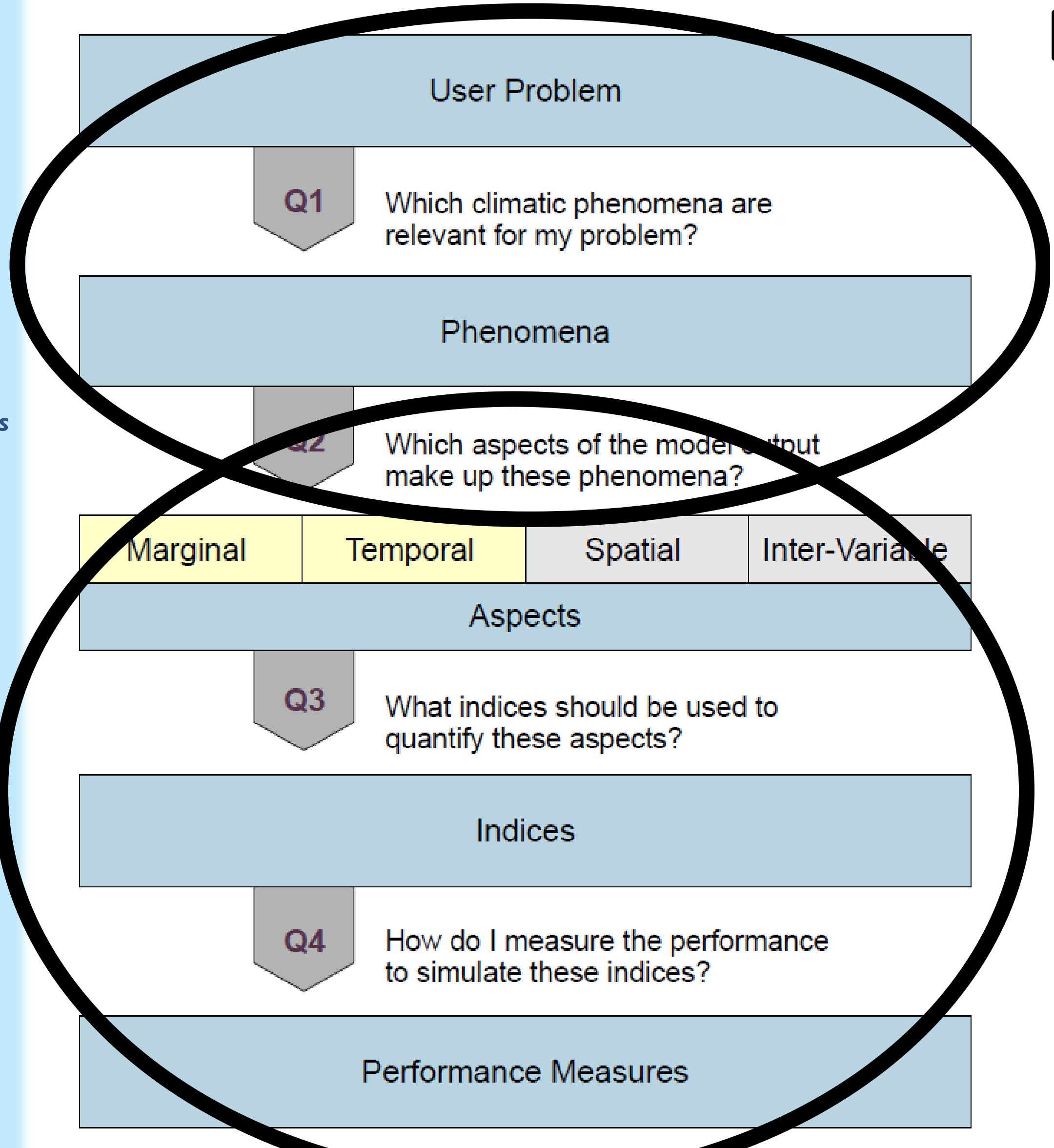
APPLICATION-SPECIFIC / USER-SPECIFIC ASPECTS OF MODEL PERFORMANCE

- Capturing temporal variability
- Capturing spatial structures
- Capturing extremes
- ...

Include both!

EXAMPLE 1: THE VALUE VALIDATION FRAMEWORK

1
2
3
4
Measures
5
6
7
8



Maraun et al., submitted

Index	Performance Measure
Marginal Aspects	
mean	bias/relative error
variance	relative error
20 season/year return level	bias/relative error
number of threshold exceedances	bias
Temporal Aspects	
time series	mean squared error/ correlation
ACF lag 1, 2,3	N.A.
median of spell length distribution	bias
90th percentile of spell length distrib.	bias
minimum/maximum of annual cycle	bias/relative error
Spatial Aspects	
decorrelation length	relative error
variogram range	relative error
decay length of tail dependence	relative error
Multivariate Aspects	
Pearson/rank correlation	N.A.
probability of joint exceedances	N.A.
indices conditional on (no) exceedance	as above

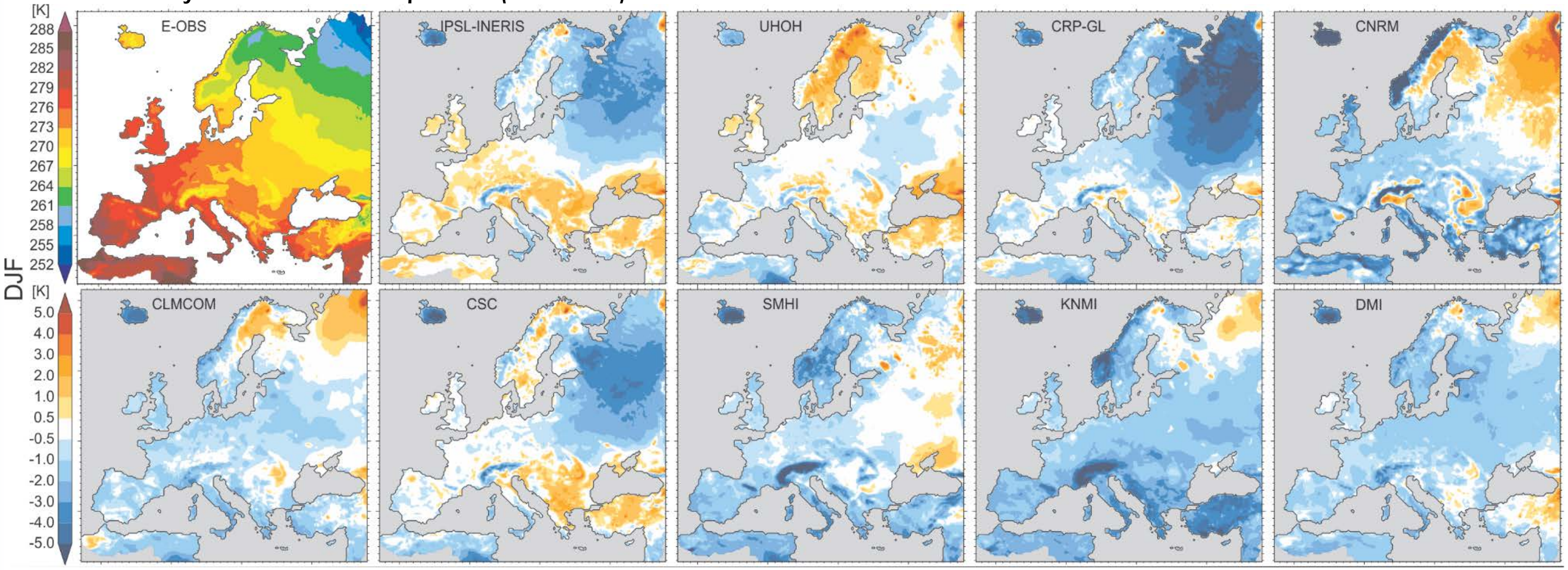
EXAMPLE 2: EURO-CORDEX STANDARD EVALUATION

1
2
3
4
Measures
5
6
7
8

Kotlarski et al., GMD, 2014

1. Seasonal mean biases at grid point scale for entire RCM domain

Bias of 20-year mean winter temperature (1989-2008)

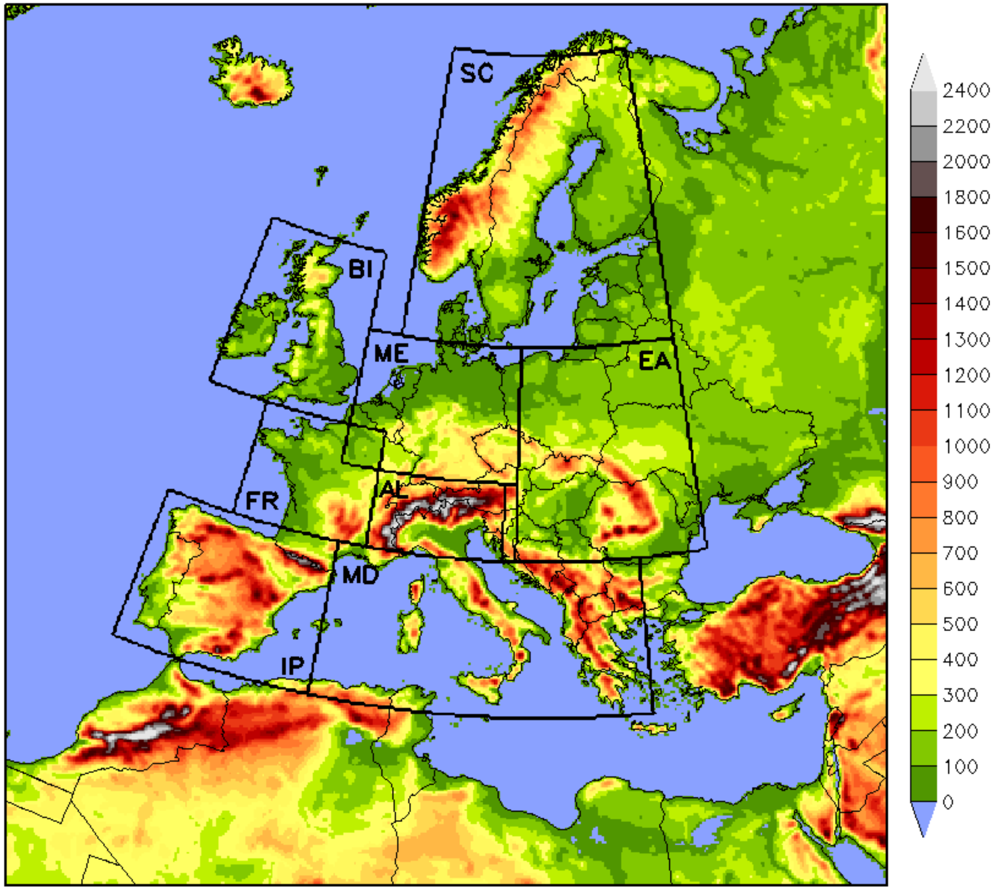


EXAMPLE 2: EURO-CORDEX STANDARD EVALUATION

1
2
3
4
Measures
5
6
7
8

Kotlarski et al., GMD, 2014

1. Seasonal mean biases at grid point scale for entire RCM domain
2. Eight metrics applied to eight different analysis regions, describing different aspects of model performance



EXAMPLE 2: EURO-CORDEX STANDARD EVALUATION

Kotlarski et al., GMD, 2014

1
2
3
4 Measures
5
6
7
8

1. Seasonal mean biases at grid point scale for entire RCM domain
2. Eight metrics applied to eight different analysis regions, describing different aspects of model performance

Temporal and spatial means

BIAS Difference (model - observations) of climatol. annual and seasonal mean values (regional averages)

Spatial variability

95%-P 95th percentile of all absolute grid point differences (model - observations) based on climatological annual and seasonal mean values

PACO Pattern correlation between modeled and observed climatological annual and seasonal mean values at all grid points

RSV Ratio of spatial variances of all grid points (model over observations) of climatological annual and seasonal mean values

Temporal variability

RIAV Ratio of interannual variance (model over observations) of time series of annual and seasonal mean values (regional averages)

TCOIAV Correlation between modeled and observed time series of annual and seasonal mean values (regional averages)

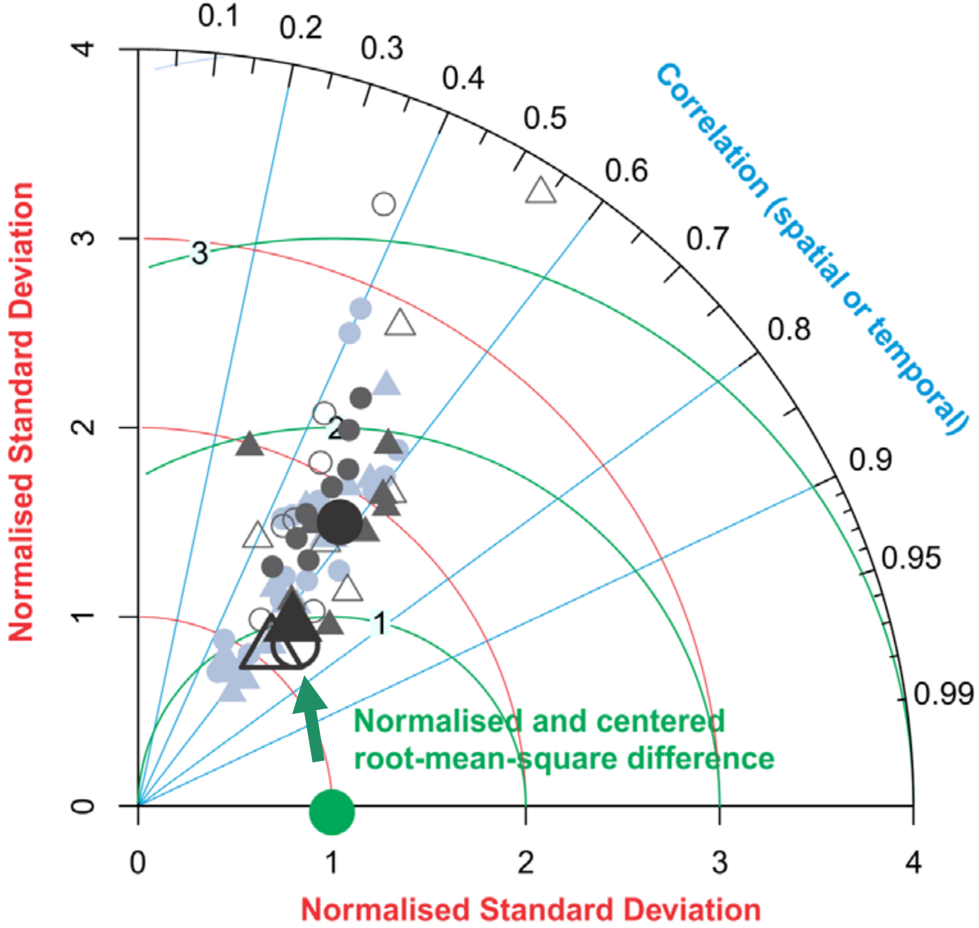
Mean annual cycle

CRCO Spearman rank correlation between modeled and observed climatological monthly mean values (regional averages)

ROYA Ratio (model over observations) of yearly amplitudes (difference between maximum and minimum) of climatological monthly mean values (regional averages)

THE TAYLOR DIAGRAMS

- Provide a way of graphically summarizing different aspects of model performance
- Here: Similarity of spatial or temporal patterns (model versus reference)
- Possible due to inter-relation of several metrics
- Different variants



Distance from origin: «Normalized and centered RMS difference»

Does **not** take into account mean bias! (but this can be color-coded)

TAYLOR DIAGRAM: EXAMPLE

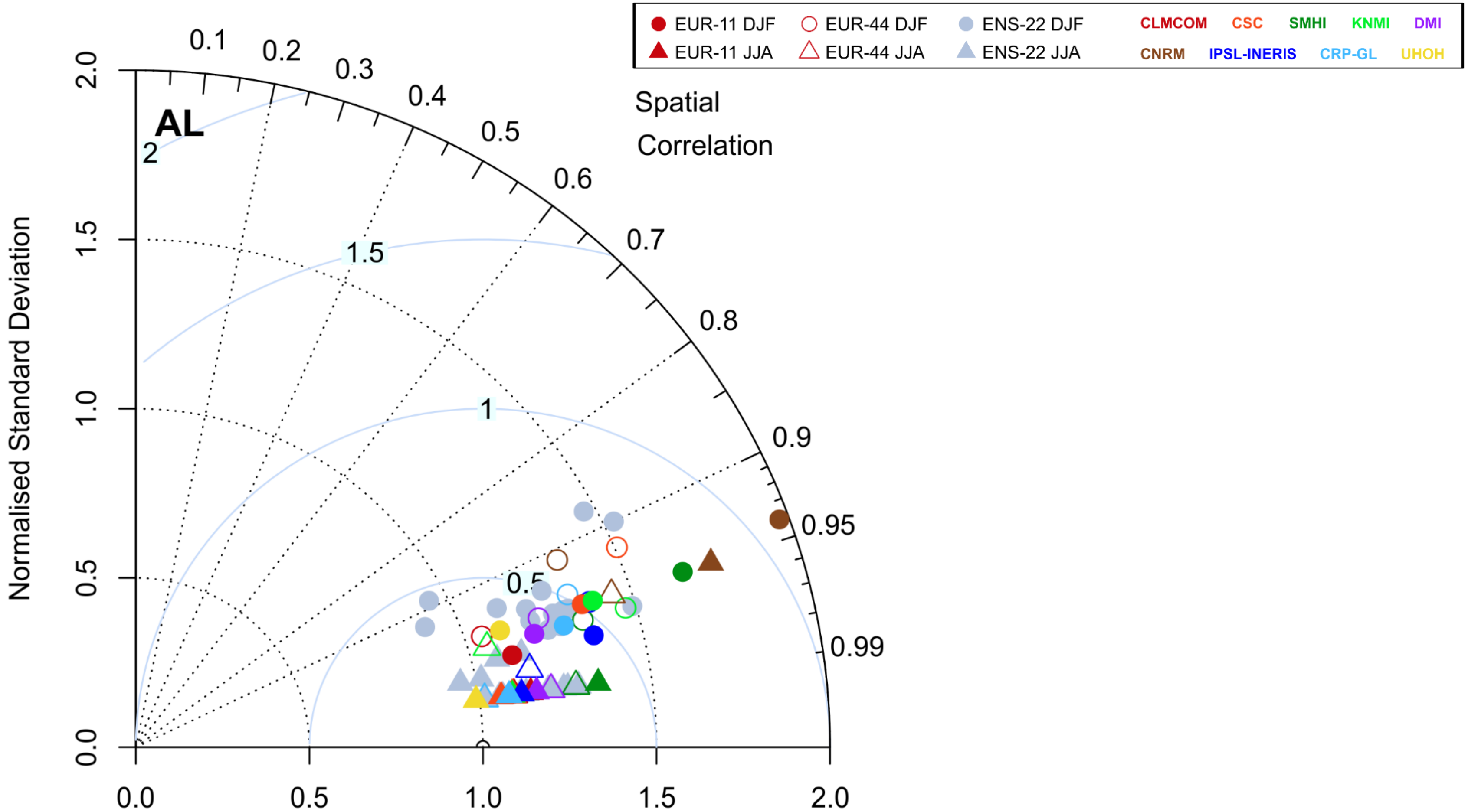


Figure B5. Spatial Taylor diagrams exploring the model performance with respect to the spatial variability of mean winter (circles) and mean summer (triangles) temperature within subdomains AL, BI, FR and MD (see Fig. 9 for subdomains EA, IP, ME and SC). Filled markers: EUR-11 ensemble, nonfilled markers: EUR-44 ensemble, gray markers: ENS-22 ensemble. The diagrams combine the spatial pattern correlation (PACO, cos(azimuth angle)) and the ratio of spatial variability (RSV, radius). The distance from the 1–1 location corresponds to the normalized and centered root-mean-square difference (which does not take into account the mean model bias), expressed as multiples of the observed standard deviation. Note the different number of underlying grid cells per subdomain in the individual ensembles.

Kotlarski et al., 2014

EXAMPLE 3: MULTIVARIATE SCORE

Bellprat et al., 2012

$$\mathrm{PI} = \frac{1}{VRTY}\sum_{v}^{V}\sum_{r}^{R}\sum_{t}^{T}\sum_{y}^{Y}\frac{\sqrt{(m_{v,r,t,y}-o_{v,r,t,y})^2}}{(\sigma_{o_{v,r,t}}+\sigma_{iv_{v,r,t,y}}+\sigma_{\epsilon_{v,r,t,y}})}. \quad (1)$$

Here, $V = 3$ is the number of model variables (T2M, PR, CLCT), $R = 8$ is the number of analysis regions (PRUDENCE regions), $T = 12$ is the number of temporal means (months), and Y is the number of years evaluated, which depends upon the ensemble considered. The variables m and o denote simulated and observed monthly means for the respective variable and region, σ_o is the standard deviation of the interannual variations derived from the observations, σ_{iv} is the standard deviation of the internal variability of the regional model derived from ensemble IV, and σ_ϵ is the standard deviation of the observational error derived from different reference datasets. For each variable (T2M, PR, and CLCT) we use three independent datasets, listed in Table 3, to estimate the observational error.

PI=0 -> perfect match

VALIDATION OF TRENDS

GLOBAL CHANGE

Forecasting Regional Climate Change Flunks Its First Test

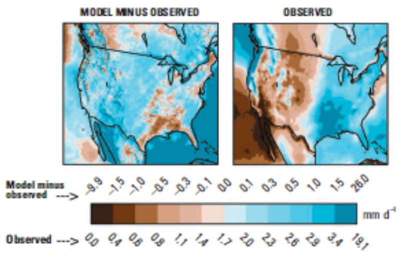
The strengthening greenhouse is warming the world, but what about your backyard, or at least your region? It's hard to say, climate researchers concede. Modelers have sharpened their tools enough to project declining grape yields in a warmer, drier California wine country and to forecast that the Mediterranean region will be getting drier in coming decades. But just how reliable such localized projections might be remains unclear.

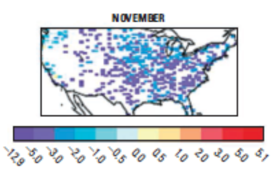
Now, a group of global, rather than regional, modelers has tested a widely used regional model by simulating climate change, not just static past climate. That's how these researchers say all regional models should be tested, but aren't. Preliminary results show that the model improved little if at all on the fuzzy view of future climate provided by a globe-spanning model. More such testing "is something that the [regional modeling] community needs to do to get its act together," says global climate modeler Gavin Schmidt of NASA's Goddard Institute for Space Studies (GISS) in New York City who was not involved in the work.

Regional modeling was the natural next step for climate modelers. To predict next week's weather, much less the next century's climate, models have to take in the whole planet. But after several decades of development, those planet-scale models still couldn't provide detailed, reliable climate projections for, say, West Texas versus East Texas. So modelers began embedding a detailed, higher resolution climate model spanning, for example, much of North America, in a global climate model. The global model would calculate broad changes and feed them into the embedded regional model, which would then compute more-detailed (and, presumably, more-accurate) simulations of smaller atmospheric features, such as storms and fronts, as well as better rendering of the atmospheric effects of surface features such as coastlines and mountains.

But was regional modeling doing any better than global modeling at making regional predictions? Pavan Racherla, Drew Shindell, and Gregory Faluvegi, all of GISS, tackled the question while doing regional modeling for a climate impacts study. "The first thing we wanted to do was evaluate the [model] output," Shindell says. "That's what we do with [global] models." But they found that regional modelers were checking model performance only when simulating climatology, the average climate for a given period of time. "That was strange," Shindell says. "The key thing we look at is climate change. That hasn't been the standard in regional climate modeling."

If a regional model does a better job simulating climatology, the three wondered, will it also do a better job forecasting climate change? To at least begin to find out, they considered a widely used regional model—the Weather Research and Forecasting (WRF) model—embedded in the global GISS-ModelE2 over the continental United States. They simulated the climate of two periods, 1968 through 1978 and 1995 through 2005, to see how WRF did at simulating climatology. Then they subtracted the earlier period from the later one to see how WRF handled climate change.





Too blue. When simulating fall precipitation (*above, left*), a regional model was mostly too wet and when simulating daily maximum temperatures (*left*) it was too cool.

WRF did not shine. "Skill capturing climatology does not translate into skill capturing climate change," Shindell concludes, echoing the group's paper of late last year in the *Journal of Geophysical Research: Atmospheres*. "There is modest improvement over the [global] model, but it's not so large." And most of that improvement came only when the global model was periodically allowed to "nudge" the wandering regional model back toward a more realistic broad-scale pattern of climate.

The GISS group does not identify why WRF failed to improve significantly on their global model, but Shindell suspects it has something to do with what the two kinds of models are best at. Climatology is determined mainly by the interplay of the land, sea, and topography with the atmosphere, which is regional models' forte, he says, but those aren't changing. Climate change, on the other hand, progresses by changes in physical properties such as the migration of jet streams, changes in cloud cover, and shifts in precipitation patterns, which a global model handles pretty well on its own.

Other researchers say the GISS group has tackled a problem that has too long been neglected. "It's an intriguing, very thought-provoking paper," says climate scientist Robert Wilby of Loughborough University in the United Kingdom. "It's a first step that's to be applauded." But it's far from a knockout punch to regional modeling. Instead, "it really highlights how tricky it is to show the value added" by regional modeling, Wilby says.

Regional modeler Lai-Yung Ruby Leung of Pacific Northwest National Laboratory in Richland, Washington, agrees. "The questions they ask are the right questions, but it would be much better if done in a multimodel experiment," she says: Different models in different combinations could identify model strengths to be exploited and weaknesses to be avoided.

Leung is a co–principal investigator of the North American Regional Climate Change Assessment Program (NARCCAP), which ran six regional models in 12 possible combinations with four global models. But the climate changes of recent decades were so small and the natural variations of climate so large, NARCCAP models were not tested against past climate change as the GISS team did. That, Leung says, is really a job for a larger, more international program.

–RICHARD A. KERR

CREDITS: P. RACHERLA, D. SHINDELL, AND G. FALUVEGI/NASA GISS (2)

Controversial «paper» with severe shortcomings that – nevertheless – highlighted an important aspect that is often overlooked:

RCMs are typically applied in climate change studies, i.e., need to capture the regional climate response to a given forcing -> validation of trends!

Problems:

- Lack of temporally homogeneous reference and boundary forcing data (re-analysis)
- Observational / re-analysis period short and rather small GHG forcing -> natural variability often dominates
- Period to be analyzed further shortened by granting spin-up time

OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE METRICS

5 TO CONSIDER!

6 MODEL WEIGHTING

7 EXAMPLE

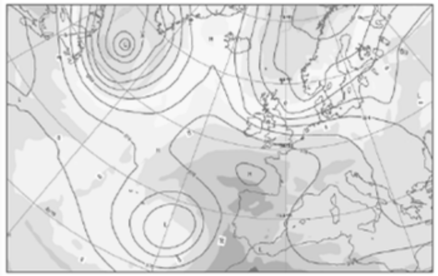
8 SUMMARY & CONCLUSIONS

MODEL CALIBRATION

THE ROLE OF MODEL CALIBRATION

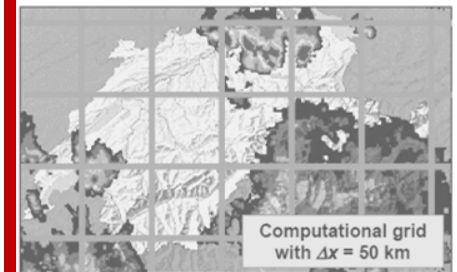
1
2
3
4
5
To Consider!
6
7
8

DYNAMICS



- Address the **resolved part** of atmospheric dynamics and thermodynamics.
- Solution of the **governing equations of fluid motion on a computational grid**
- Examples of resolved structures: **general circulation of atmosphere, low and high pressure systems, mountain flows**

PHYSICS



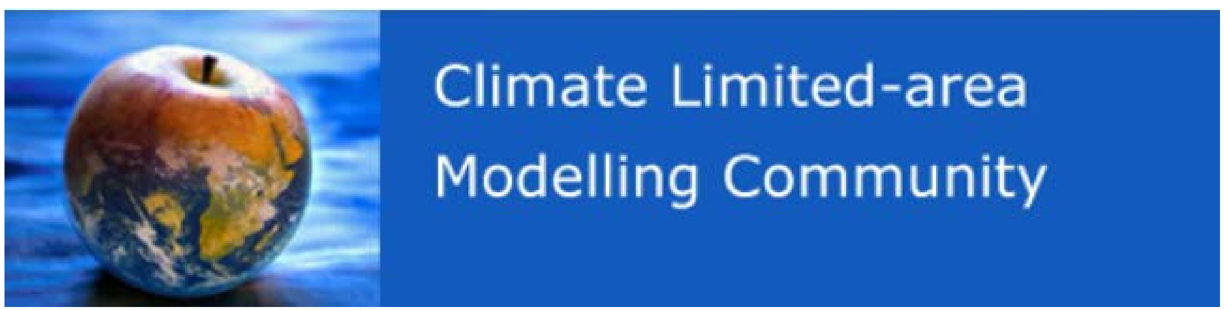
- Representation of **unresolved scales** by **parameterizations** (sub-grid)
- Typically contain **empirical** components and are to some extent **tuned/calibrated**
- Major source of **model uncertainty**
- Examples of parameterized processes: **boundary layer, convection, precipitation, clouds, land surface**

- Model physics typically include a large number of non-constrained parameters that need to be calibrated («tuning»)
- Calibration will affect model performance!
- The same is true for further choices concerning model setup (domain size, time step, relaxation procedure, horizontal and vertical resolution, etc.)
- Calibration is typically INTRANSPARENT!

Evaluation might not be «independent» (if the same evaluation period, reference data and performance measures were used during calibration)

1st EXAMPLE: MODEL CALIBRATION

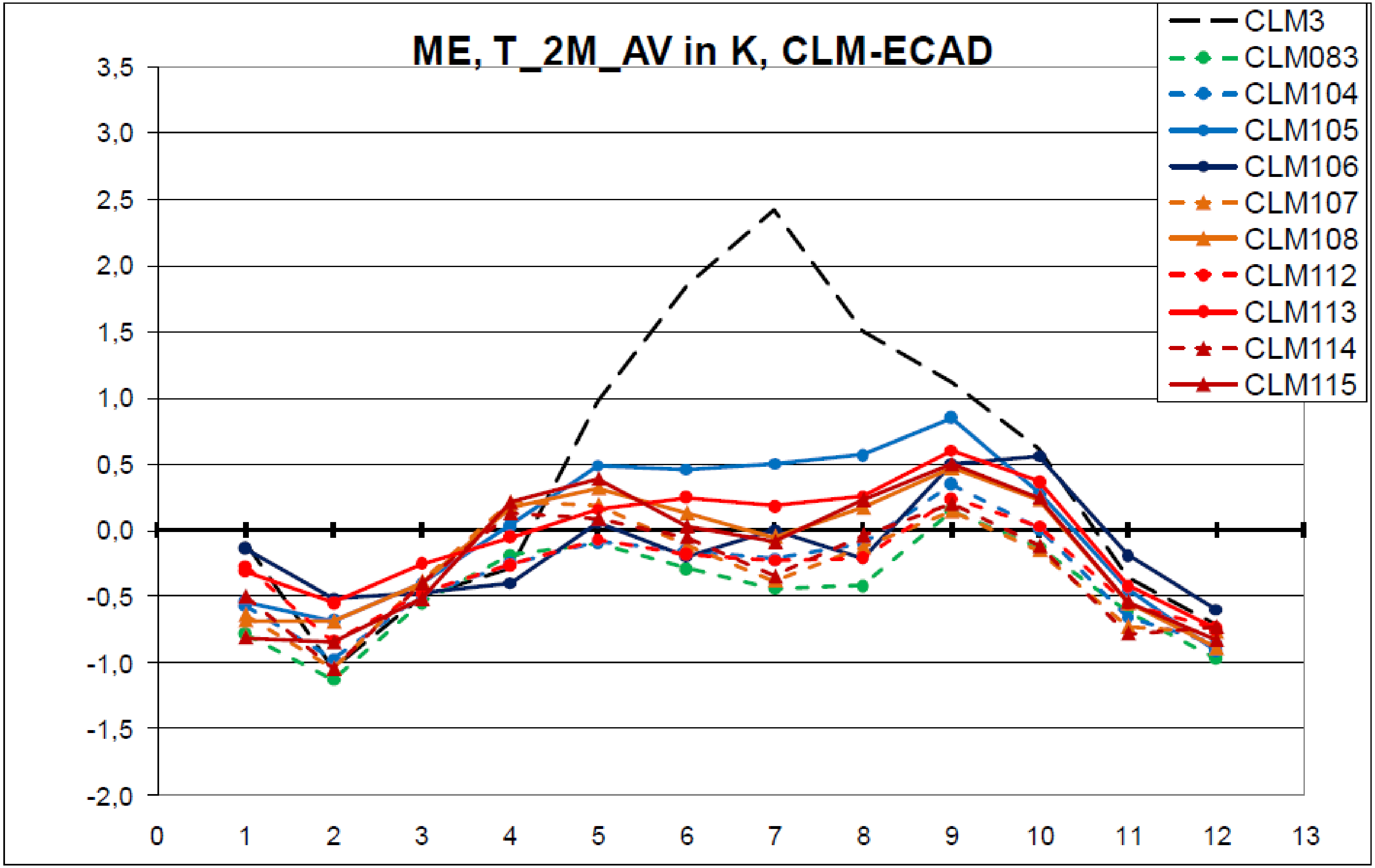
- Setup of a CORDEX reference version for the RCM COSMO-CLM
- Testing of a large number of model setups (parameter settings in physics, parameterisation schemes, time step, preprocessing scheme, etc.)
- 10-year long simulations driven by the ERA40 re-analysis



1st EXAMPLE: MODEL CALIBRATION (cont'd)

1
2
3
4
5
To Consider!
6
7
8

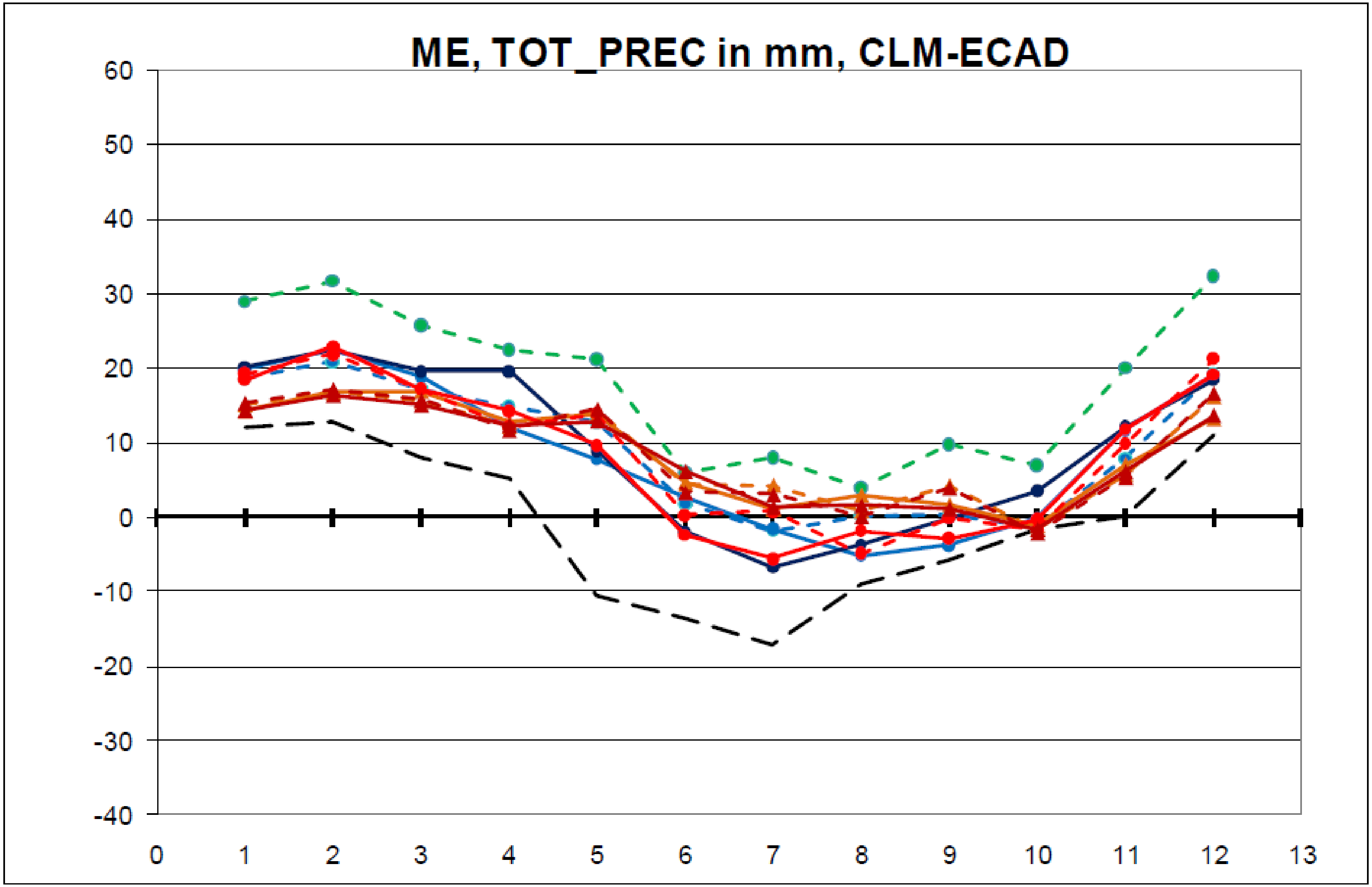
Mean monthly temperature bias (1991-2000) over mid-Europe [K]



1st EXAMPLE: MODEL CALIBRATION (cont'd)

1
2
3
4
5
To Consider!
6
7
8

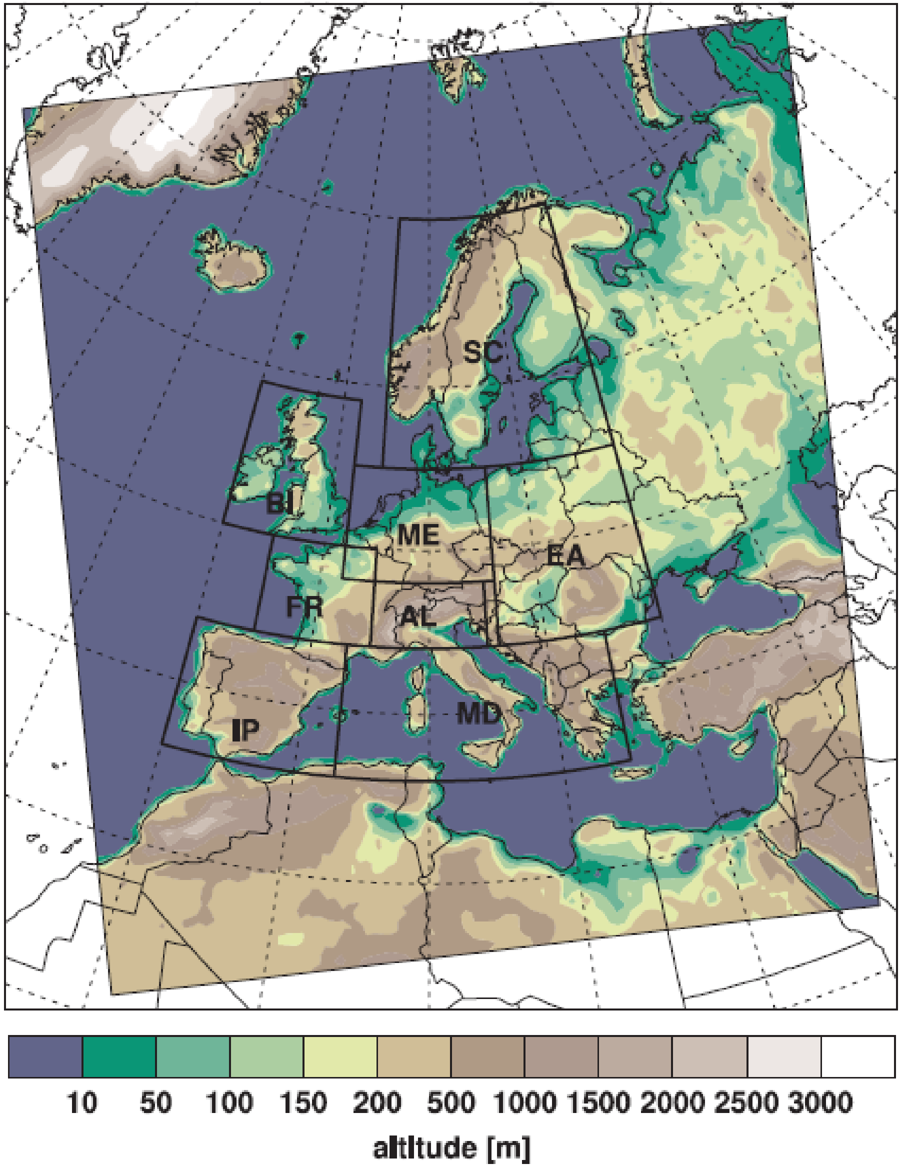
Mean monthly precipitation bias (1991-2000) over mid-Europe [mm]



2nd EXAMPLE: MODEL CALIBRATION

1
2
3
4
5
To Consider!
6
7
8

- Development of an objective calibration scheme for the RCM COSMO-CLM
  Bellprat et al. 2012a,
  Bellprat et al. 2012b
- >40 parameters varied in a «perturbed physics ensemble)
- 1-year long and 10-year long simulations
- Evaluation over eight European sub-domains («PRUDENCE regions», «Rockel regions») -> see EURO-CORDEX standard evaluation -> comparability!



2nd EXAMPLE: MODEL CALIBRATION (cont'd)

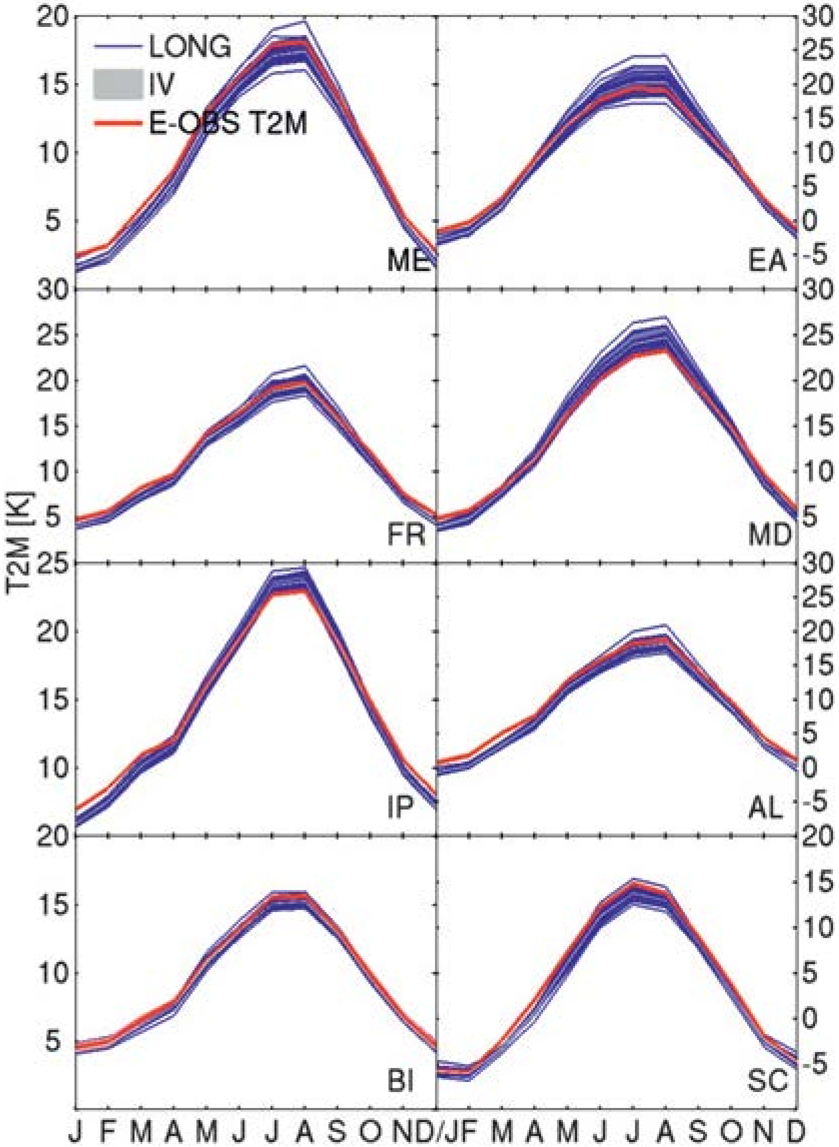
TABLE 2. Perturbed parameters in (top) SHORT and (bottom) LONG. The bold entries denote the default value in CCLM. For all parameters a minimum and maximum bound is tested, while some individual parameters have been tested more extensively in addition to that. SHORT is designed to identify important model parameters. In LONG multiple parameters are changed at a time, summarized in the second part of the table.

Parameter/property	Acronym	Value
	SHORT	
Turbulence		
Minimal diffusion coefficients for heat ($m^2 s^{-1}$)	Tkhmin	[0, **1**, 2]
Minimal diffusion coefficients for momentum ($m^2 s^{-1}$)	Tkmmin	[0, **1**, 2]
Turbulent length scale (m)	turb_len	[100, **500**, 1000]
Factor for turbulent heat dissipation	d_heat	[12, 15, **10.1**]
Factor for turbulent momentum dissipation	d_mom	[12, 15, **16.6**]
Factor for turbulent diffusion of TKE	c_diff	[0.01, **0.2**, 10]
Land surface		
Scalar for laminar boundary layer roughness	rlam_heat	[0.1, **1**, 3, 5, 10]
Scalar for laminar boundary layer roughness sea	rat_sea	[1, 10, **20**, 50, 100]
Factor for canopy height	rat_can	[0, **1**, 10]
Ratio of laminar boundary layer thickness for *q* and *h*	rat_lam	[0.1, **1**, 10]
Surface area index of the waves over sea	c_sea	[1, **1.5**, 5, 10]
Surface area index of the (evaporative) soil	c_soil	[0, **1**, 10]
Surface area index of grid points over land	c_lnd	[1, **2**, 10]
Roughness length of a typical synoptic station (m)	z0m_dia	[0.001, **0.1**, 10]
Length scale of subscale surface patterns over land (m)	patlen	[10, 100, **500**, 1000]
Exponent to get the effective surface area	e_surf	[0.1, **1.5**, 10]
Stomata resistance	crsmin	[50, 200, **300**]
Convection		
Fractional mass flux for downdrafts at LFS	rmfdeps	[0.2, **0.35**, 0.5]
Assumed convective cloud cover (%)	rcucov	[0.01, **0.05**, 0.5]
Factor for the time scale for cape closure	rtau	[0.5, **1**, 1.5]
Coefficient for determining conversion from cloud water to rain	rprcon	[0.000 15, 0.001, **0.0015**, 0.002, 0.015]
Penetrative entrainment rate ($1\ m^{-1}$)	entrpen	[4e-5, **8e-5**, 12e-5]
Midlevel entrainment rate ($1\ m^{-1}$)	entrmid	[4e-5, **8e-5**, 12e-5]
Entrainment rate for shallow convection ($1\ m^{-1}$)	entrsc	[5e-5, 1e-4, **3e-4**, 1e-3, 2e-3]
Microphysics		
Cloud droplet concentration ($1\ m^{-3}$)	cloud_num	[5e7, **5e8**, 1e9]
Cloud water threshold for autoconversion	qi0	[**0**, 0.000 01, 0.0001, 0.001, 0.01]
Separating mass between cloud and rain (kg)	zxstar	[3.36e-11, **2.6e-10**, 7.25e-09]
Factor for fall velocity of snow	zv0s	[10, **15**, 30]
Radiation		
Subgrid-scale cloud height scalar	uc1	[0.2, **0.5**, 0.8]
Critical value for normalized oversaturation	q_crit	[1, **4**, 7, 10]
Cloud cover at saturation in statistical cloud diagnostic	clc_diag	[0.2, **0.5**, 0.8]
Interval (in time steps) between two calls of the radiation scheme	hincrad	[0.5, 0.75, **1**]
Convective subgrid cloud scalar	conv_clc	[0.7, **1**, 1.3]
	LONG	
Physics		
Convection scheme type	iconv_type	IFS, **Tiedtke**
Subgrid-scale orography	lsso	**On**, off
Transport of rain and snow	ltrans_prec	**On**, off
Prognostic rain and snow	lprogprec	**On**, off
Cloud water and cloud ice	itype_gscp	**On**, off
Stomata resistance ($s\ m^{-1}$)	crsmin	[**150**, 300]
Length scale of subscale surface patterns over land (m)	patlen	[200, **500**]
Numerics		
Numerical scheme	LF, RK	**Leapfrog** or Runge–Kutta
Asselin filter	alphaass	[0.5, 0.7, **1.0**]
Correction factor for horizontal diffusion of moisture	hd_corr_q	[0, 0.25, **0.5**]
Correction factor for horizontal diffusion of temperature	hd_corr_t	[0, 0.25, 0.375, **0.75**]
Correction factor for horizontal diffusion of *u*, *v*, *w*	hd_corr_u	[0.25, 0.375, 0.75, **1**]
Interval running the convection scheme	ninconv	[**1**, 2]

Bellprat et al., 2012a

2nd EXAMPLE: MODEL CALIBRATION (cont'd)

Mean annual cycle (1991-2000) of **temperature** in the perturbed physics ensemble over the eight sub-domains

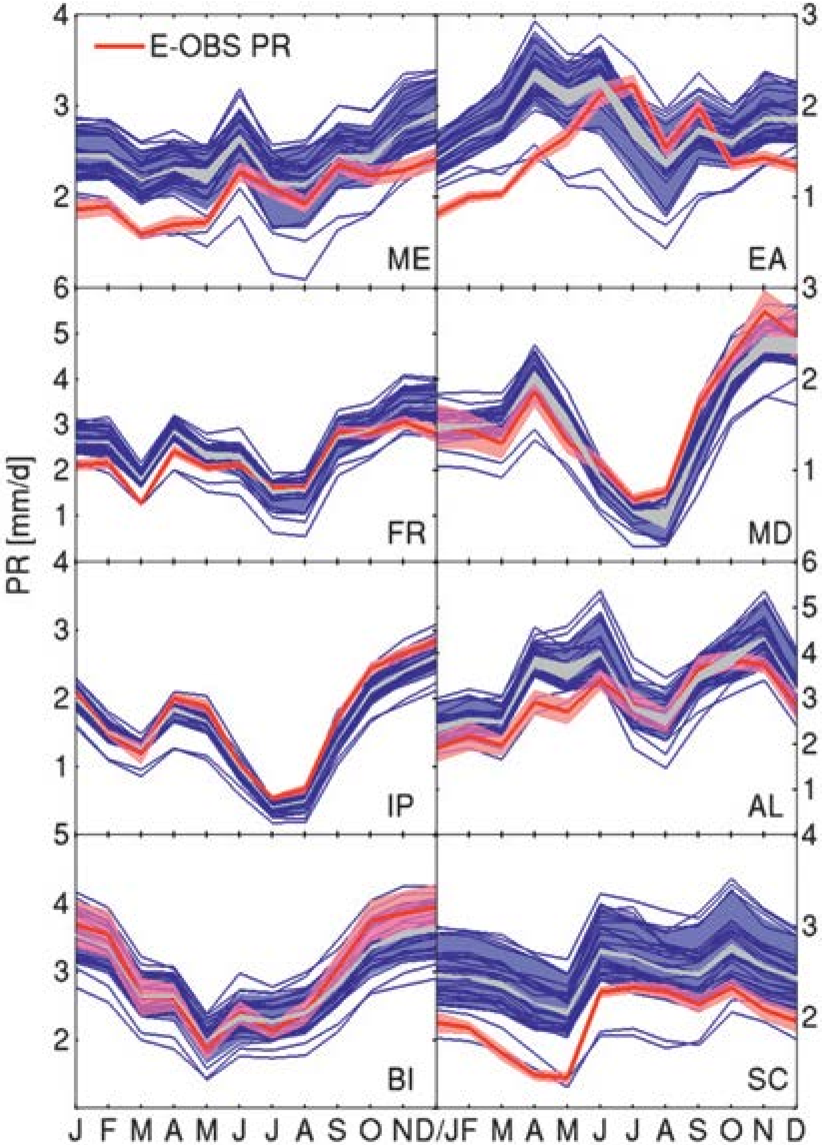


Bellprat et al., 2012a

2nd EXAMPLE: MODEL CALIBRATION (cont'd)

1
2
3
4
5
To Consider!
6
7
8

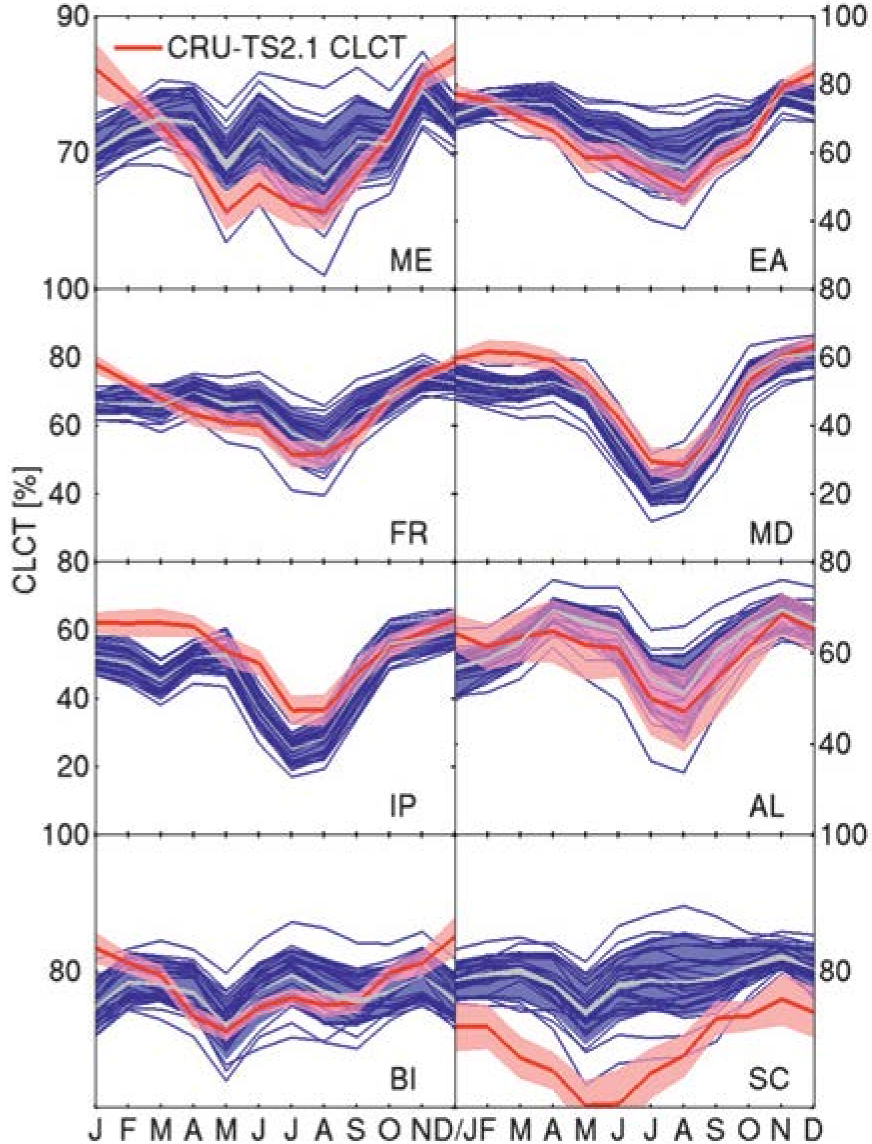
Mean annual cycle (1991-2000) of **precipitation** in the perturbed physics ensemble over the eight sub-domains



Bellprat et al., 2012a

2nd EXAMPLE: MODEL CALIBRATION (cont'd)

Mean annual cycle (1991-2000) of **cloud cover** in the perturbed physics ensemble over the eight sub-domains



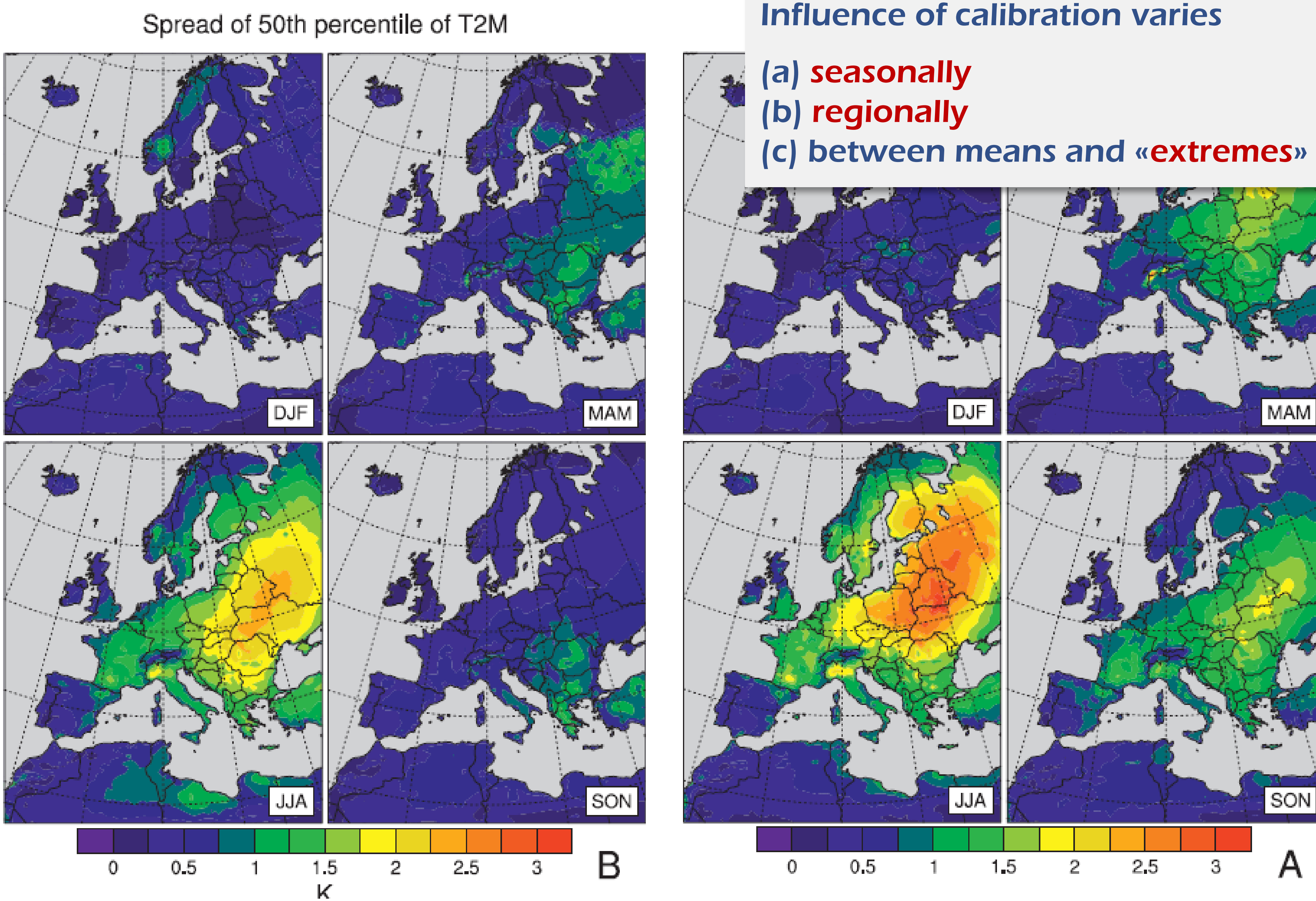
Bellprat et al., 2012a

2nd EXAMPLE: MODEL CALIBRATION (cont'd)

1
2
3
4
5
To Consider!
6
7
8

Influence of calibration varies

(a) seasonally
(b) regionally
(c) between means and «extremes»



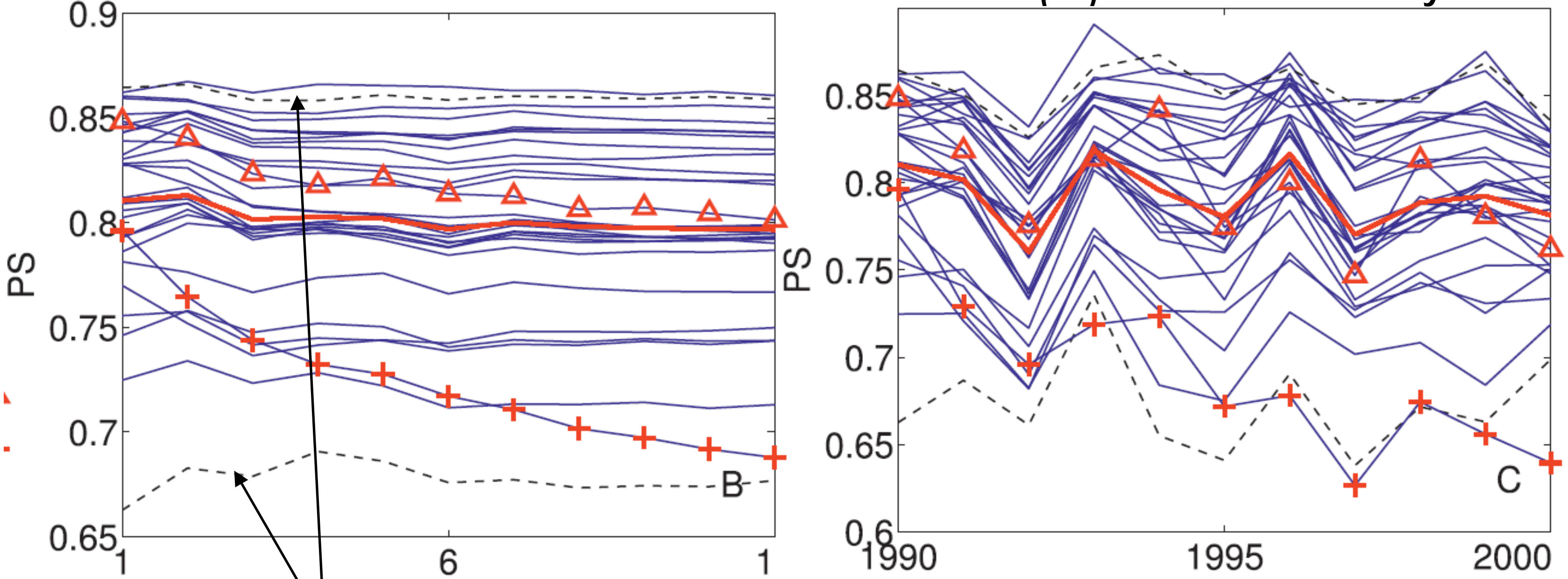
Bellprat et al., 2012a

2nd EXAMPLE: MODEL CALIBRATION (cont'd)

1
2
3
4
5 To Consider!
6
7
8

Skill (PS) when sequentially adding additional years to the score

Skill (PS) for each individual year



Bellprat et al., 2012a

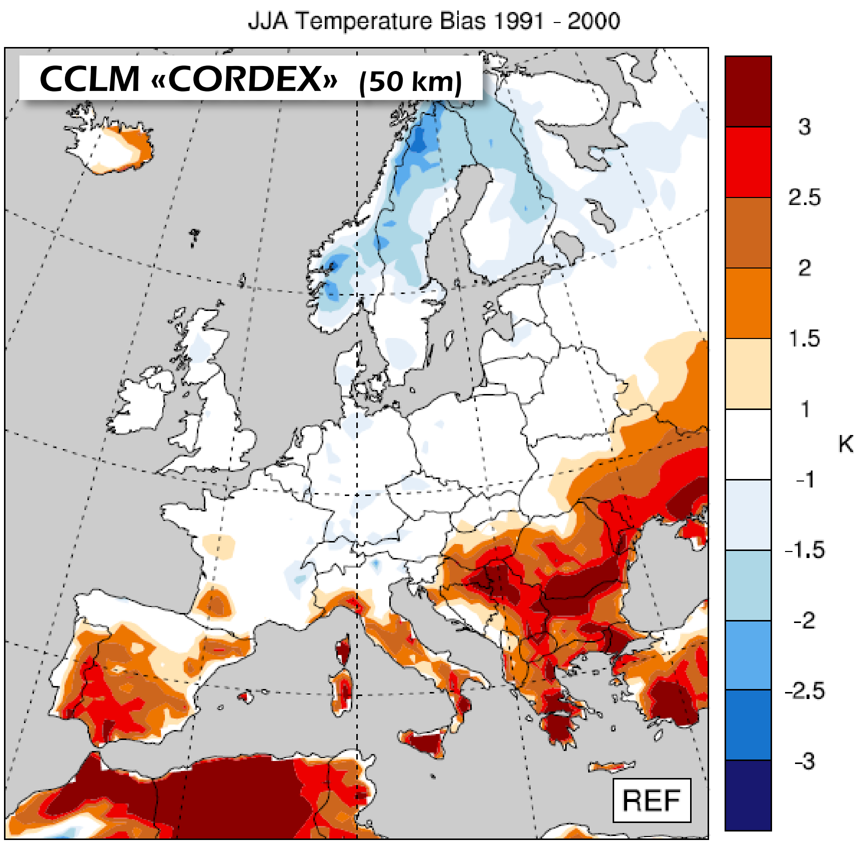
Worst / best simulation in the first year

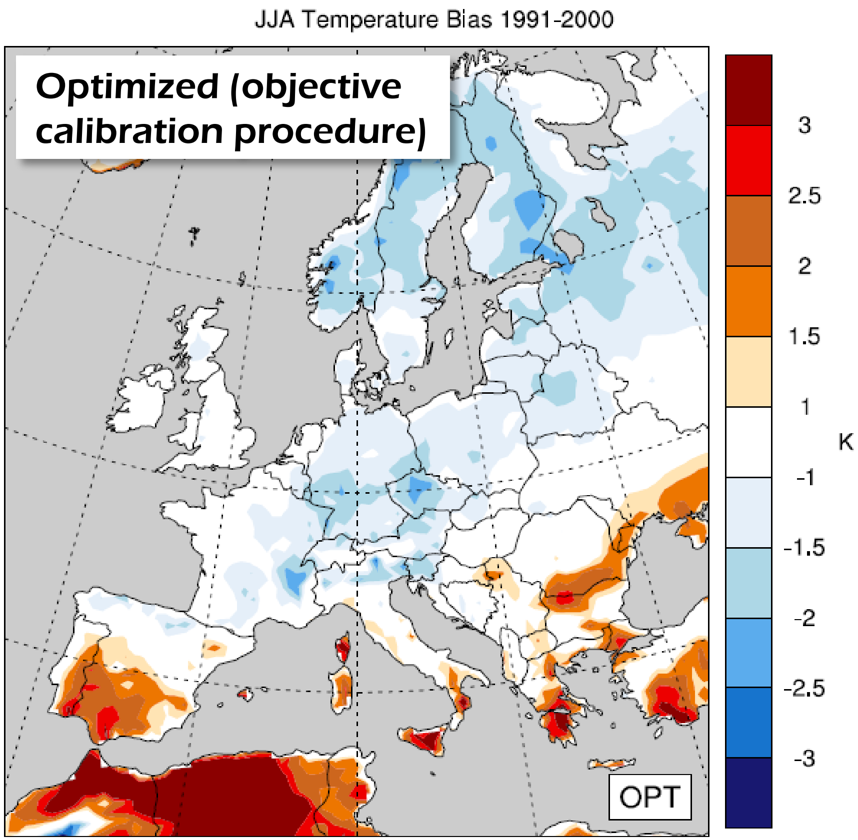
Relative performance not stable in time!

(here: mostly due to spin-up of soil water content)

2nd EXAMPLE: MODEL CALIBRATION (cont'd)

1
2
3
4
5
To Consider!
6
7
8





Bellprat et al., in prep.

In this case:

Model calibration is transparent, published and «objective» (usually NOT the case)

SUMMARY: EFFECT OF MODEL CALIBRATION

(IPCC AR5, WG1, Chapter 9, 2014)

- Model tuning directly influences the evaluation of climate models
- The quantities that are tuned cannot be used in model evaluation
- Quantities closely related to those tuned will provide only weak tests of model performance
- Model quality is tested most rigorously through the concurrent use of many model quantities, evaluation techniques, and performance metrics that together cover a wide range of emergent (or un-tuned) model behaviour.

INTERNAL VARIABILITY

INTERNAL CLIMATE VARIABILITY

1
2
3
4
5
To Consider!
6
7
8

Internal climate variability (IV):
Unforced random variability in climatic parameters due to internal non-linear processes in the climate system

Consequence for comprehensive (deterministic!) climate models:
Results (e.g., temporal and spatial patterns) can strongly depend on slight perturbations of the (typically not well constrained) initial conditions

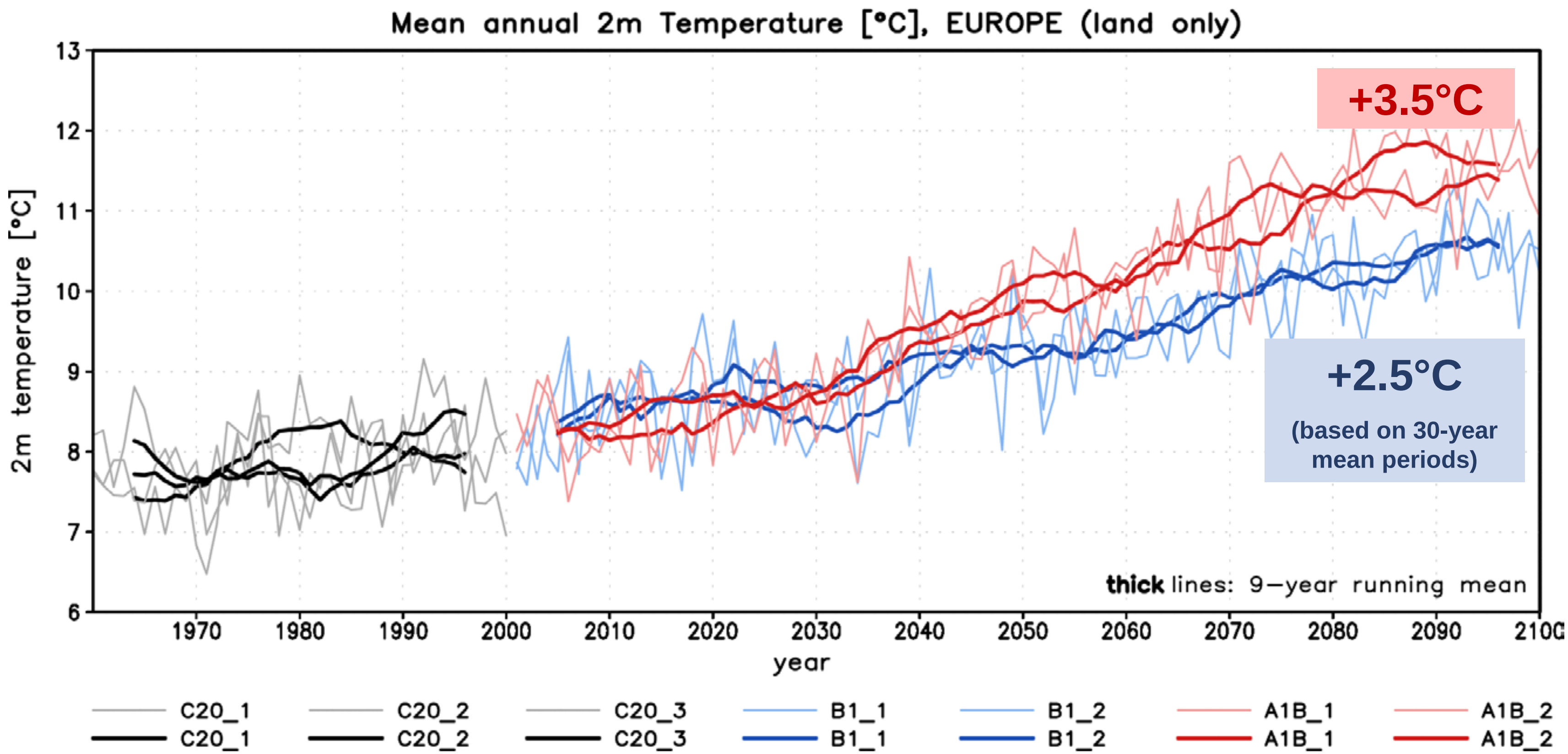
This introduces uncertainty in both climate scenarios and model evaluation, especially on **regional/local scales**, for **short analysis periods** and for **extremes**!

“Side-effect”: A free-running GCM initialized at some historical point in time will have **no temporal correspondence** with reality!

THE «CLM CONSORTIAL RUNS»

1
2
3
4
5
To Consider!
6
7
8

One RCM (COSMO-CLM) driven by several **realizations** of one GCM (ECHAM5)



Each realization is equally likely!

IV IN REGIONAL CLIMATE SIMULATIONS

- Even with identical boundary forcing, slightly differently initialized / perturbed RCM experiments with exactly the same setup will differ from each other to some extent (and therefore also performance measures)
- This effect is random!!

IV influence is

...larger for short analysis periods (partly averages out on longer tiem scales)

...larger for small analysis domains (e.g., individual grid cells)

...typically larger for precipitation than for temperature

...larger for (rare) extremes

...typically larger in summer (RCM solution less constrained by boundary forcing)

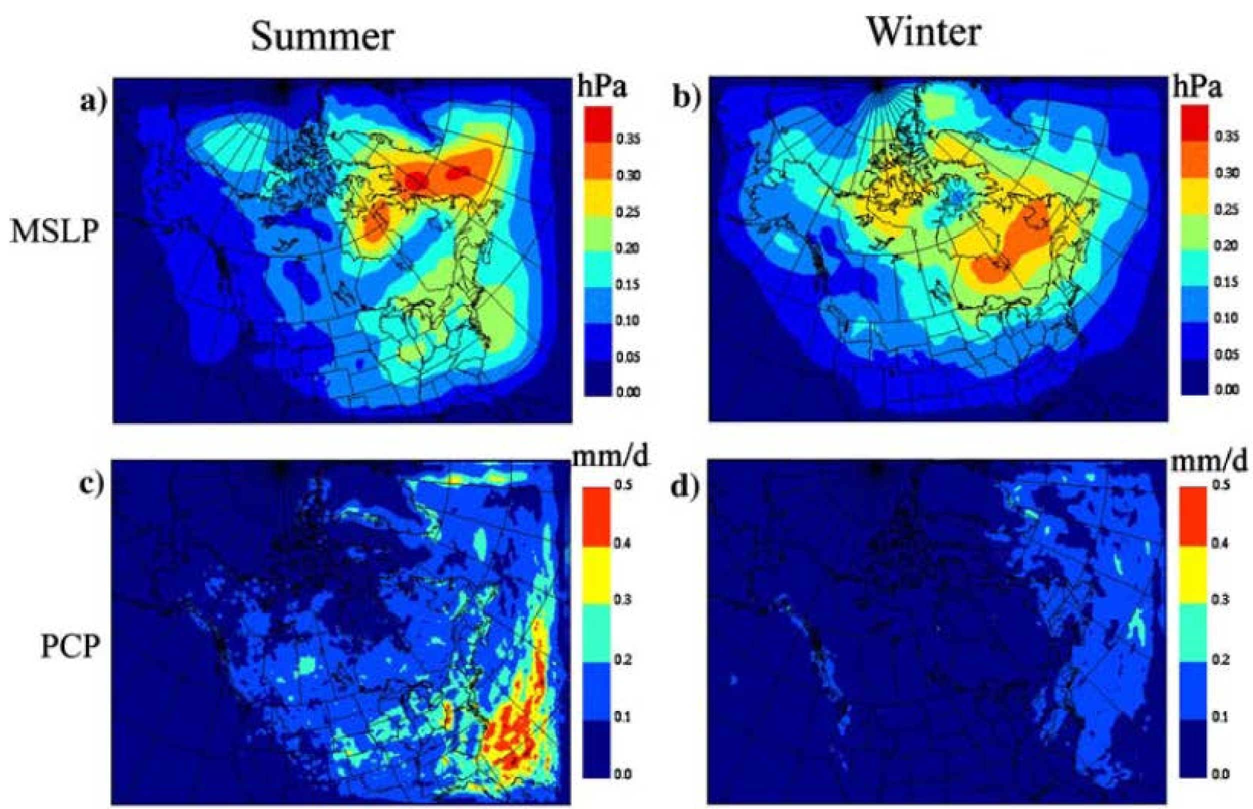
...larger towards the outflow boundary (RCM solution less constrained by boundary forcing)

EXAMPLE: INFLUENCE OF IV ON 10-YEAR RCM CLIMATE

1
2
3
4
5
To Consider!
6
7
8

10 CRCM simulations for 1980-1989 driven by NCEP/NCAR re-analysis with slightly perturbed initial conditions

Fig. 7 Square root of the variance between the 10-year climate of each member of the ensemble $\left(\sqrt{\sigma^2_{\bar{X}}}\right)$ from 1980 to 1989 for the mean-sea-level pressure (MSLP; hPa) with ten members in **a** summer and **b** winter. Computation is repeated for **c–d** the precipitation (PCP; mm/day) and **e–f** the screen temperature (ST; °C)



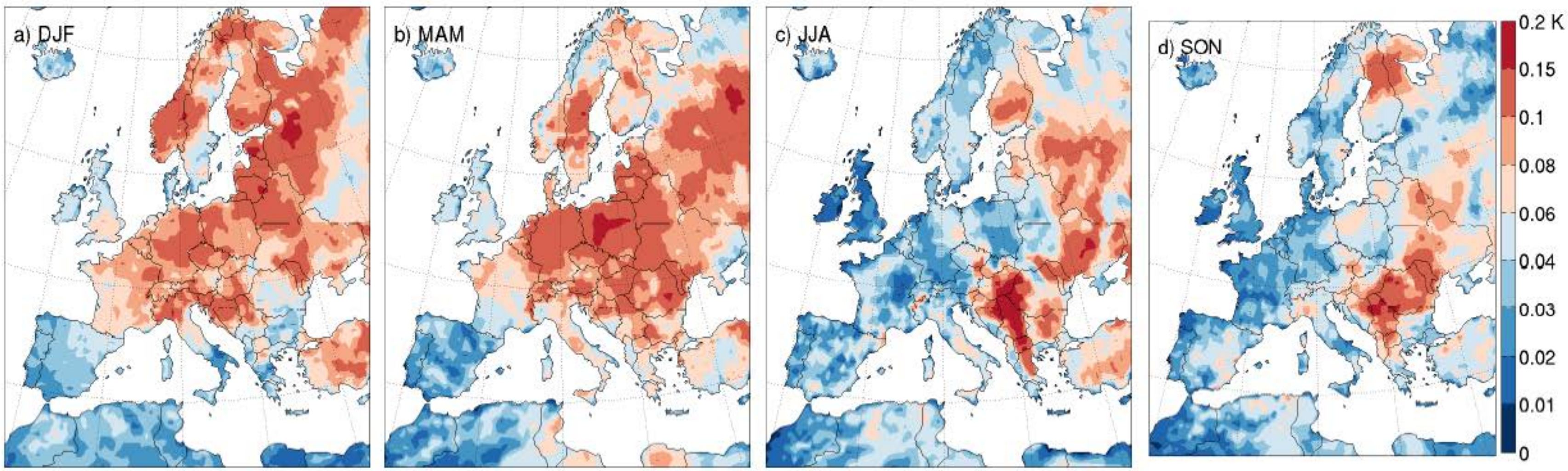
Lucas-Picher et al., 2008

IV influence larger in summer than in winter and larger in the East than in the West.

EXAMPLE: INFLUENCE OF IV ON 42-YEAR RCM CLIMATE

4 COSMO-CLM simulations for 1958-2000 driven by ERA40 re-analysis with slightly shifted start dates

Mean seasonal temperature difference (42-year means) between the sensemble members



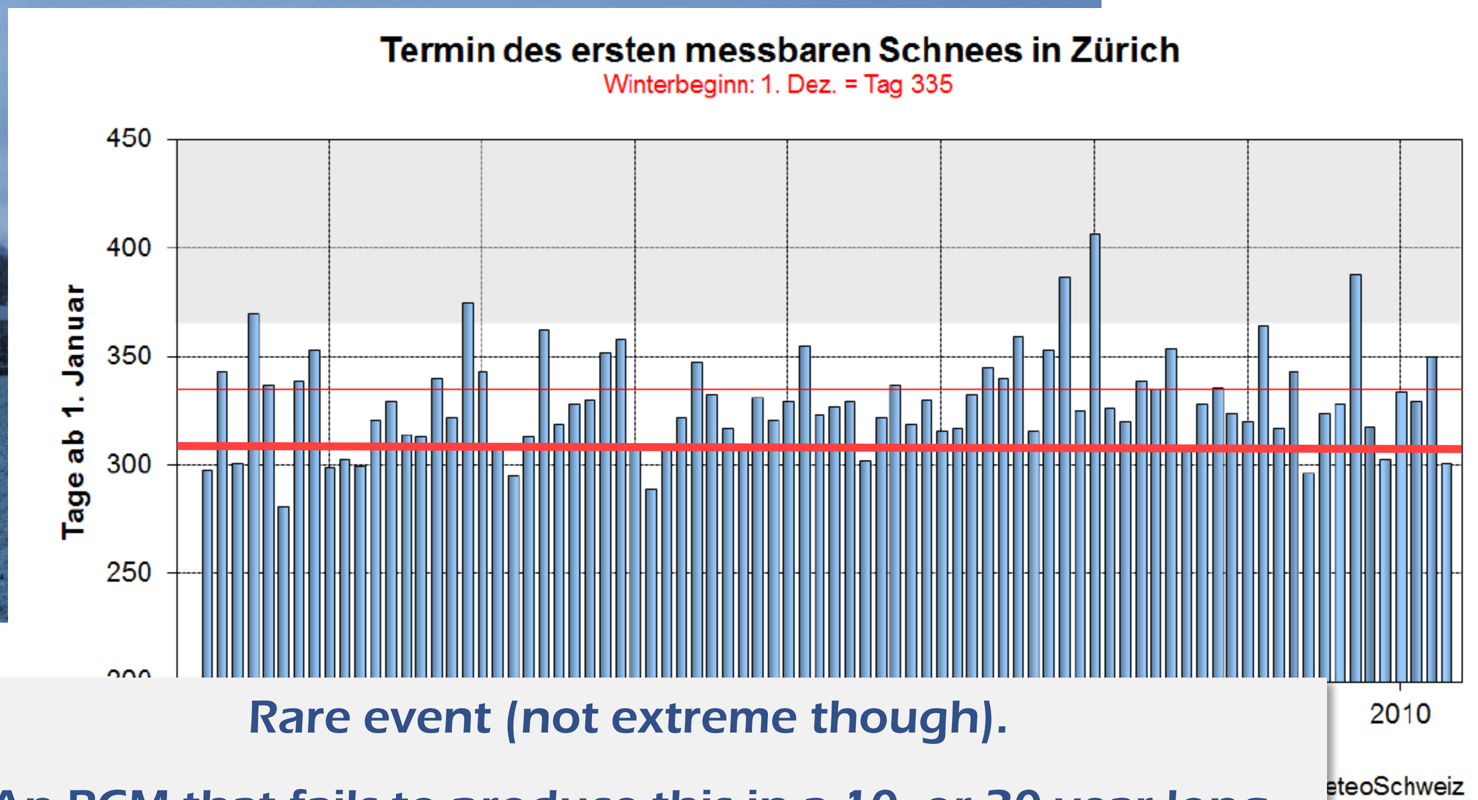
Roesch et al., 2008

> "It can thus be concluded that the model's performance in predicting climate extremes cannot be properly evaluated using only one model simulation"

YESTERDAY MORNING...

1
2
3
4
5
To Consider!
6
7
8

Lachen (Switzerland), 408 m a.s.l., 6th November 2014



Rare event (not extreme though).

An RCM that fails to produce this in a 10- or 20-year long simulation is NOT necessarily deficient!

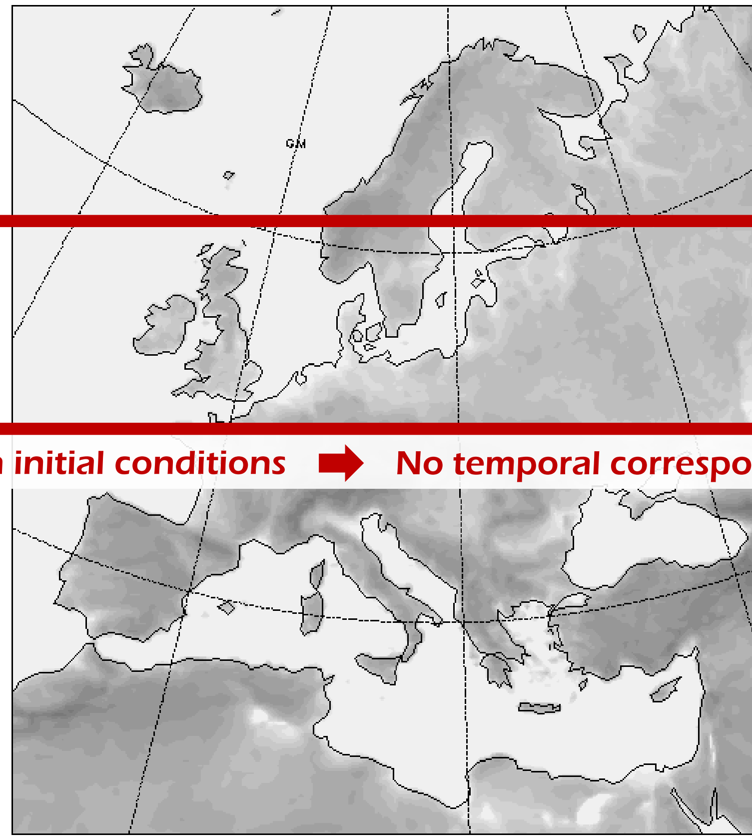
TEMPORAL AND SPATIAL CORRESPONDENCE

TYPES OF RCM EXPERIMENTS

1
2
3
4
5 To Consider!
6
7
8

boundary forcing (global)

RCM



Re-analysis *(perfect boundaries)* → EVALUATION RUN: **Evaluation of downscaling**

GCM **historical GHG** → CLIMATE SCENARIO: **Evaluation of GCM-RCM chain**

Internal variability and uncertain initial conditions ➡ No temporal correspondence with «real-world»

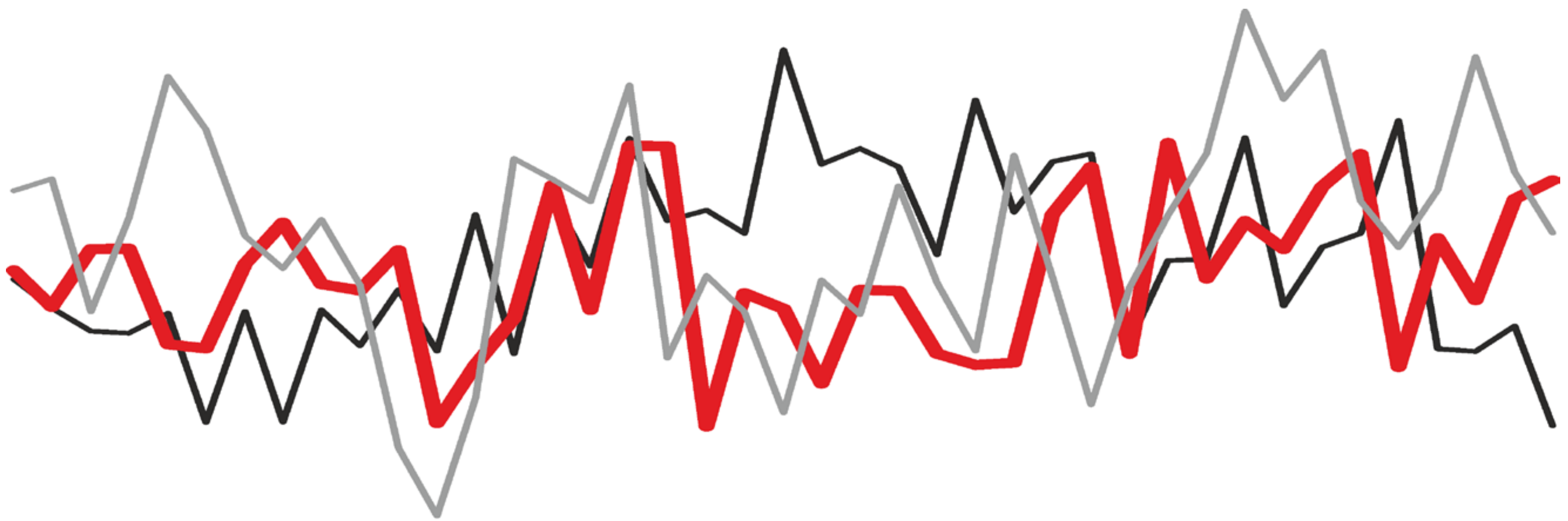
future GHG → change

Re-analysis/ GCM Idealized setups → SENSITIVITY RUN: Sensitivities, process understanding

REALIZATIONS OF THE SAME CLIMATE

1
2
3
4
5
To Consider!
6
7
8

(annual mean temperatures, average over European land surface)



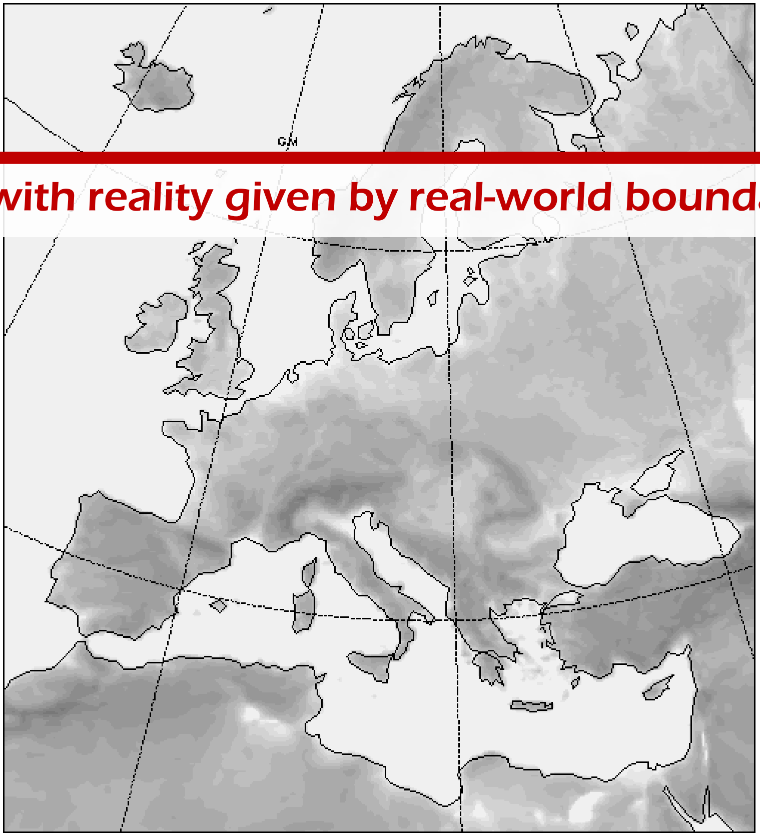
Observations
GCM-driven run #1
GCM-driven run #2

- No temporal correspondence (year-to-year, day-to-day)
- Evaluation has to be carried out with respect to climatologies
- But: Evaluation of (forced) trends, interannual variability, transition probabilities, etc. possible

TYPES OF RCM EXPERIMENTS

boundary forcing (global)

RCM



EVALUATION RUN

Re-analysis
(perfect boundaries)

Evaluation of downscaling

Temporal correspondence with reality given by real-world boundary forcing

CLIMATE SCENARIO

GCM
historical GHG

Evaluation of GCM-RCM chain

GCM
future GHG

Climate change

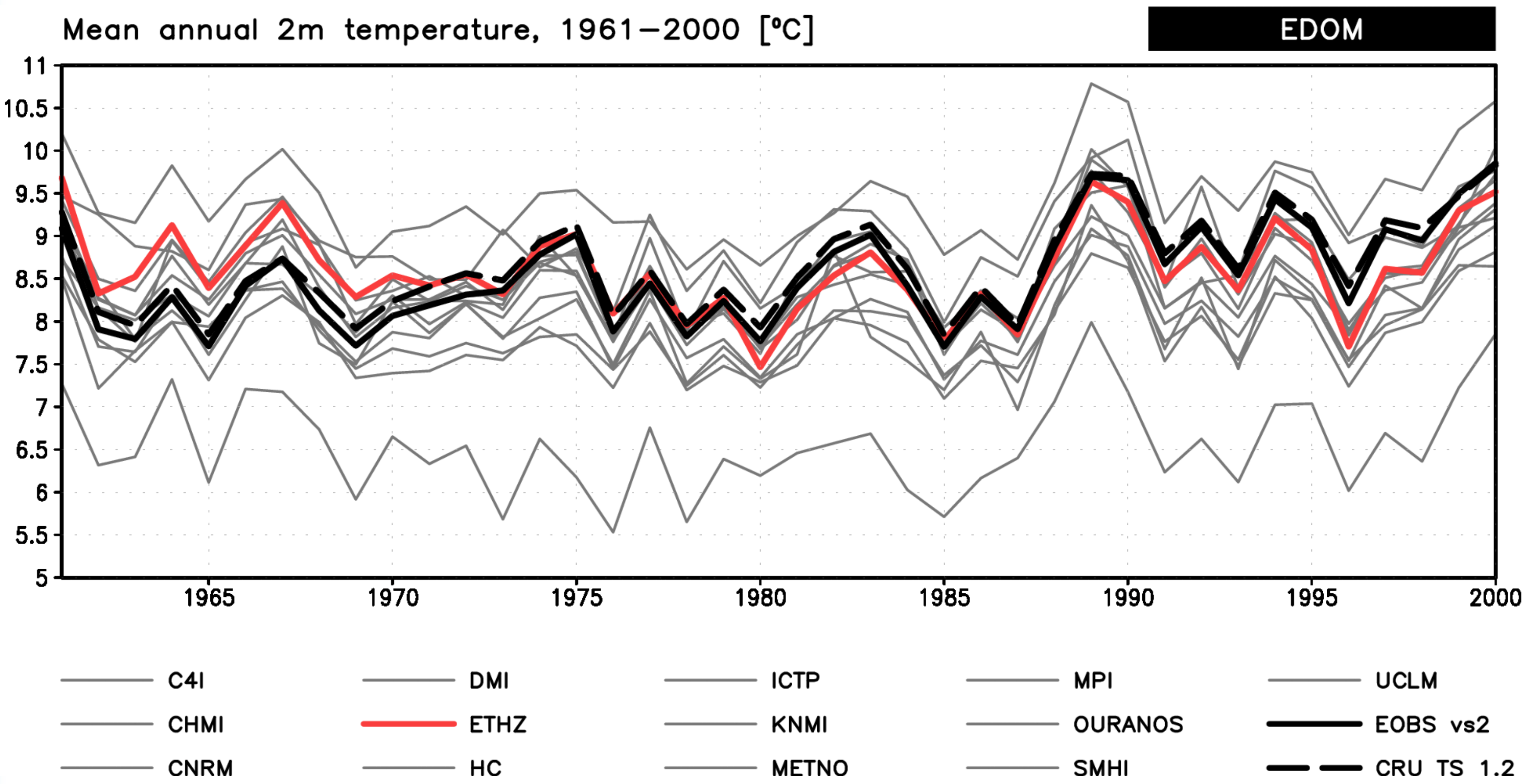
SENSITIVITY RUN

Re-analysis/ GCM
Idealized setups

Sensitivities, process understanding

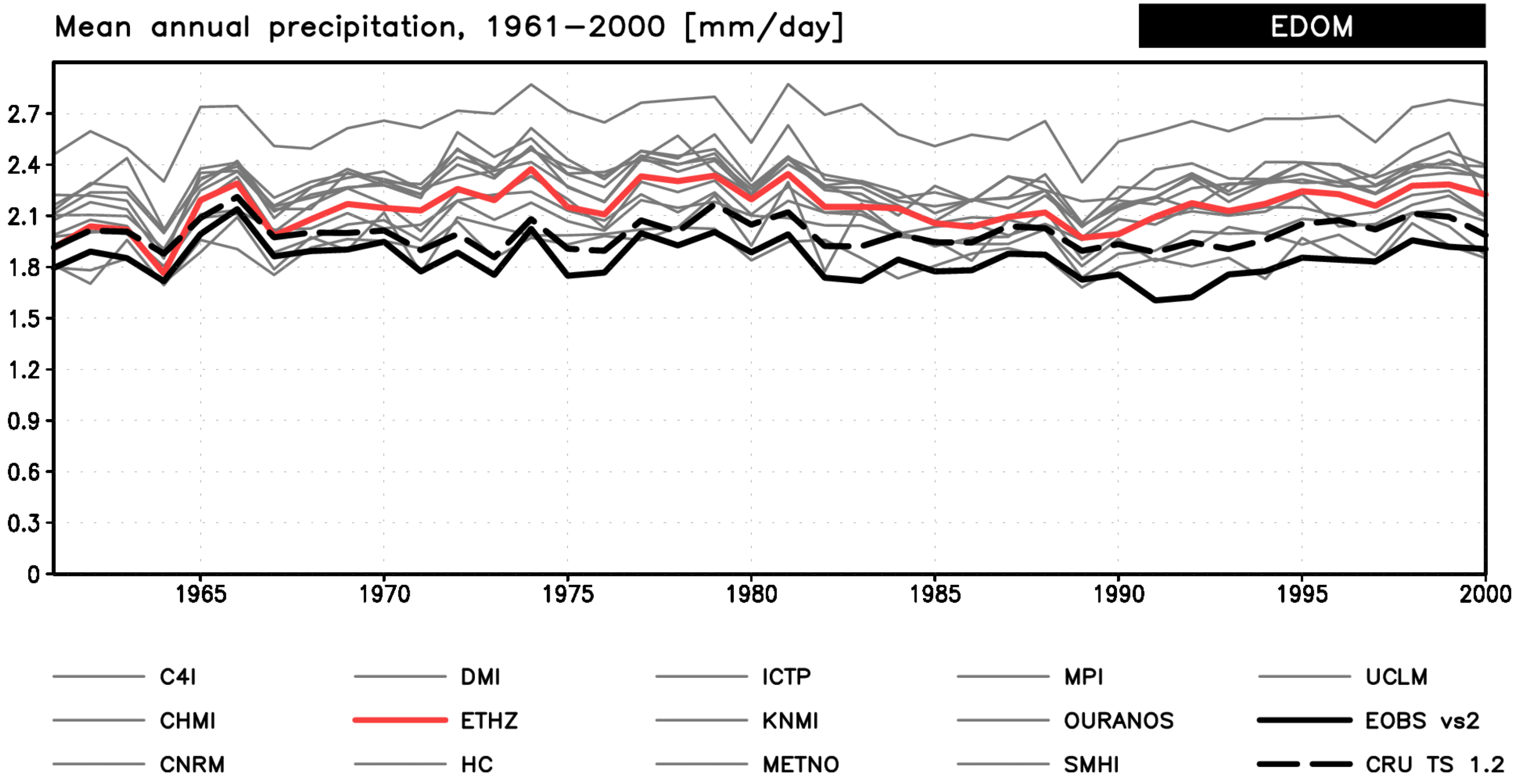
MEAN ANNUAL TEMPERATURE (OBS. AND ERA40-driven RCMs)

(mean over European land surface)



MEAN ANNUAL PRECIPITATION (OBS. AND ERA40-driven RCMs)

(mean over European land surface)

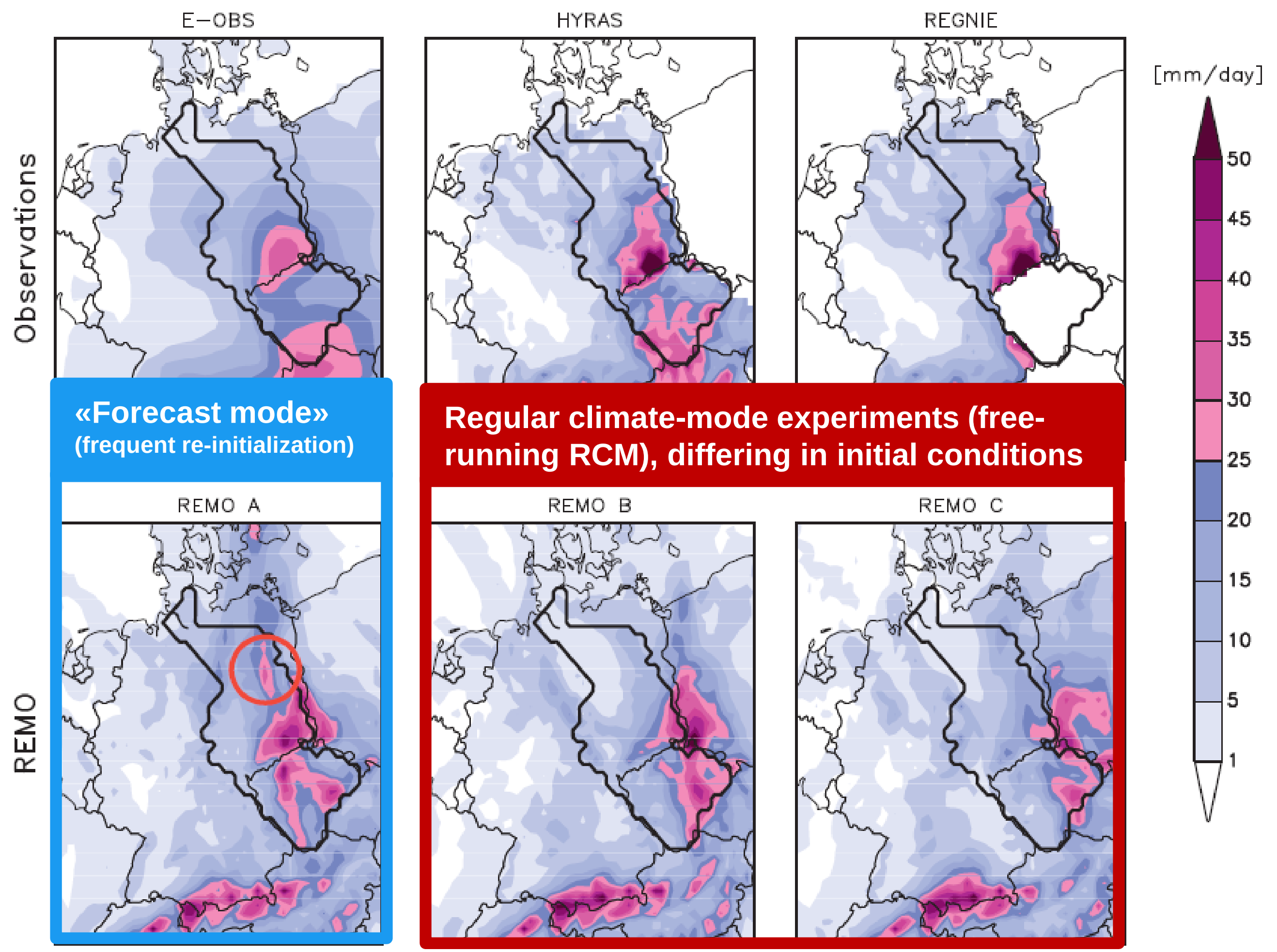


LIMITATIONS OF EVENT-WISE VALIDATION

IV: Direct correpondence diminishes on short temporal and small spatial scales and also for features that are influenved by the memory of the land surface.

Elbe flooding 2002:

Mean precipitation (10th-13th August) in three different observational reference datasets and three different RCM experiments driven by the ECMWF analysis



Kotlarski et al., 2012

FORECAST-MODE AND NUDGING

FORECAST-MODE

Frequently (e.g., every 24 hours) re-initialize the RCM's prognostic fields with the interpolated boundary forcing (analysis, re-analysis) in the interior domain -> poor-man's «assimilation» of observations into the system.

Used to construct regional re-analyses.

Reduces the degrees of freedom of the RCM and prevents a large influence of IV.

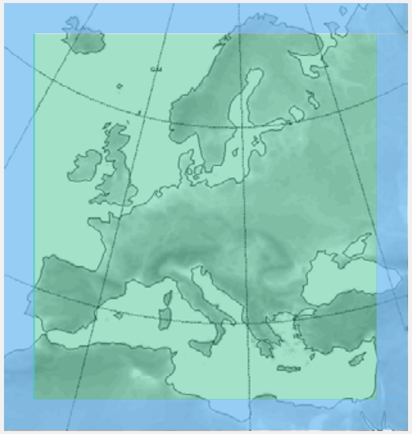
Cannot be applied in scenario context! Evaluation less informative!

(SPECTRAL) NUDGING

Apply large-scale boundary forcing also in the interior domain (but only for large scales and in upper atmosphere and without full replacement of RCM solution).

Keeps RCM flow close to boundary forcing; reduces IV by reducing degrees of freedom for RCM.

Can be applied in scenario context.



«Regular» boundary forcing (relaxation)

Additional nudging

Special model setups that increase the temporal correspondence with the real world (by reducing IV), but that are less informative wrt. model quality

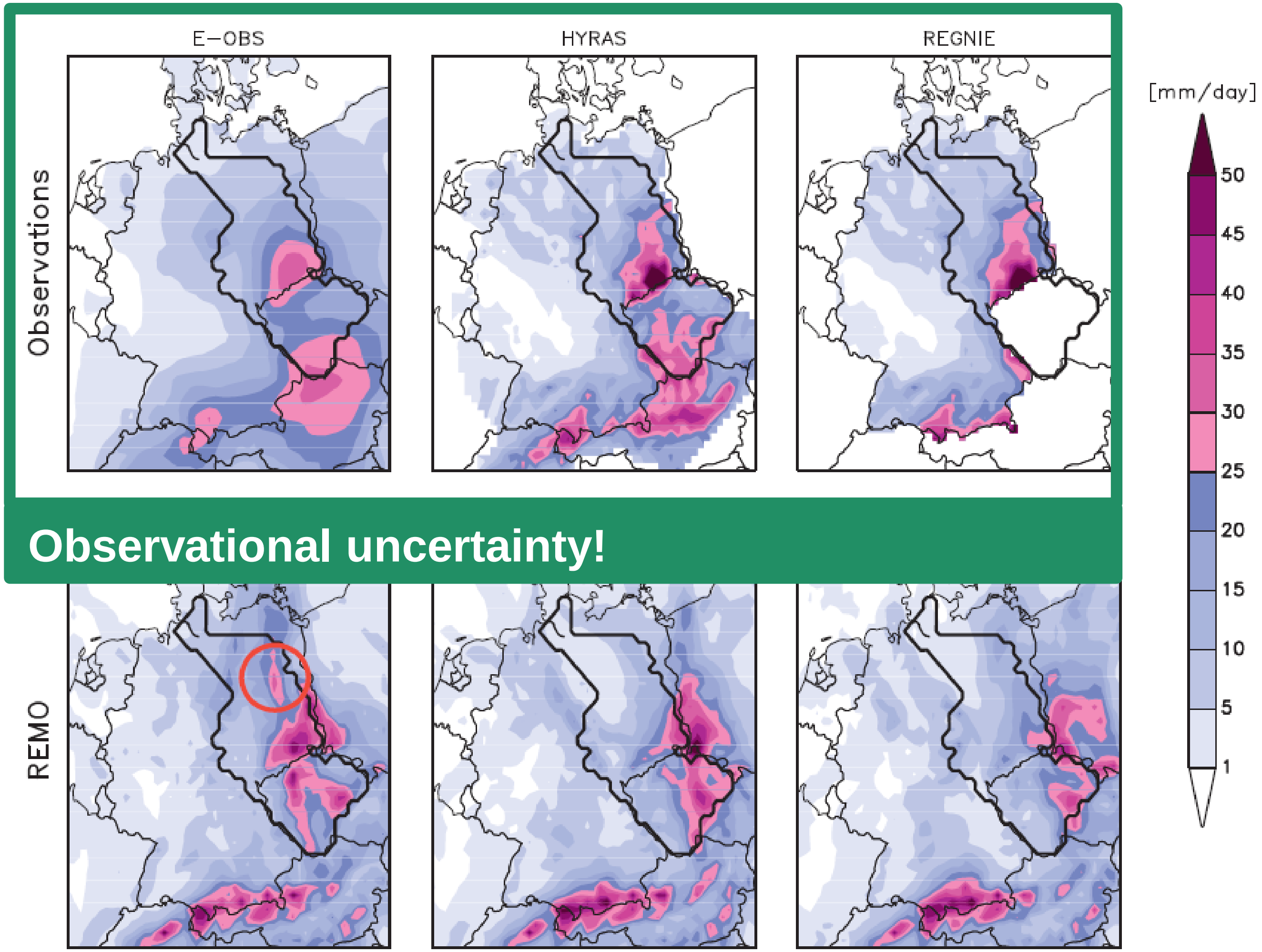
OBSERVATIONAL UNCERTAINTY

OBSERVATIONAL UNCERTAINTY - OVERVIEW

Model evaluation typically affected by uncertainties and errors of the observational reference.

Elbe flooding 2002:

Mean precipitation (10th-13th August) in three different observational reference datasets and three different RCM experiments driven by the ECMWF analysis



Kotlarski et al., 2012

OBSERVATIONAL UNCERTAINTY - ORIGINS

- **Measurement errors** (e.g., automatic weather stations)
- **Deficient translation of measured quantities into validation parameters** (e.g. radiances to temperatures, cloud coverage or precipitation rates)
- **Inappropriate gridding procedure and/or target resolution**
- **Spatial and/or temporal inhomogeneities of underlying station dataset**
- **Representativeness errors, including physiographic effects** (Does a grid point of an observational grid really represent an areal mean value?)

MEASUREMENT ERRORS: PRECIPITATION

1
2
3
4
5
To Consider!
6
7
8

- Systematic undercatch of rain gauges due to deformation of wind field and evaporative losses
- Strongly depends on site characteristics, ambient weather conditions and measurement device
- Most important for snowfall and during strong winds (less than 50% of true precipitation)
- Usually not corrected for in gridded products (e.g., E-OBS)

A wet model bias of 10-20% can well be explained by deficient observations!

Only of minor importance for statistical downscaling

Complicates comparison of SD and RCM performance

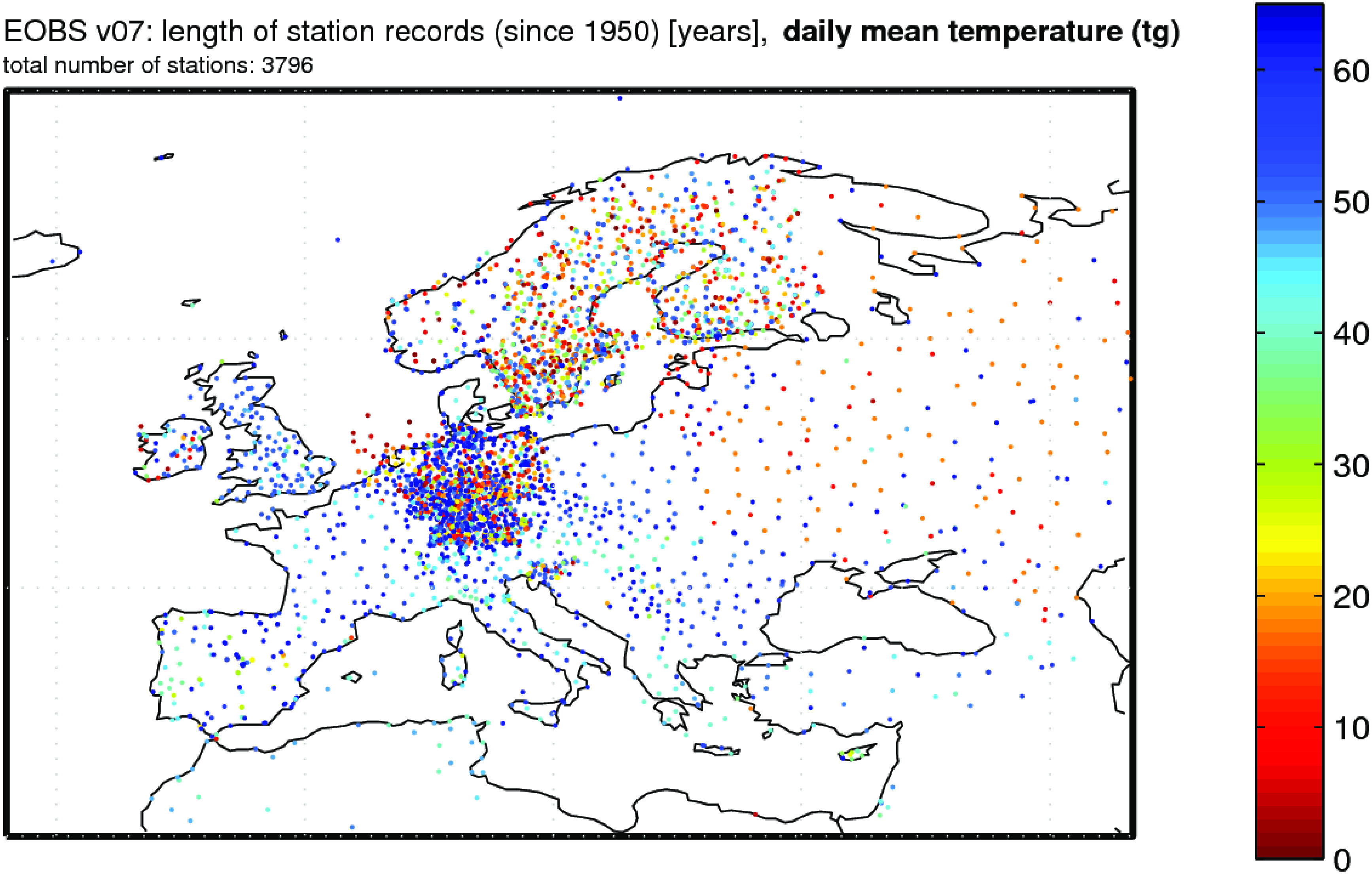
TEMPORAL AND SPATIAL INHOMOGENEITIES

1
2
3
4
5
To Consider!
6
7
8

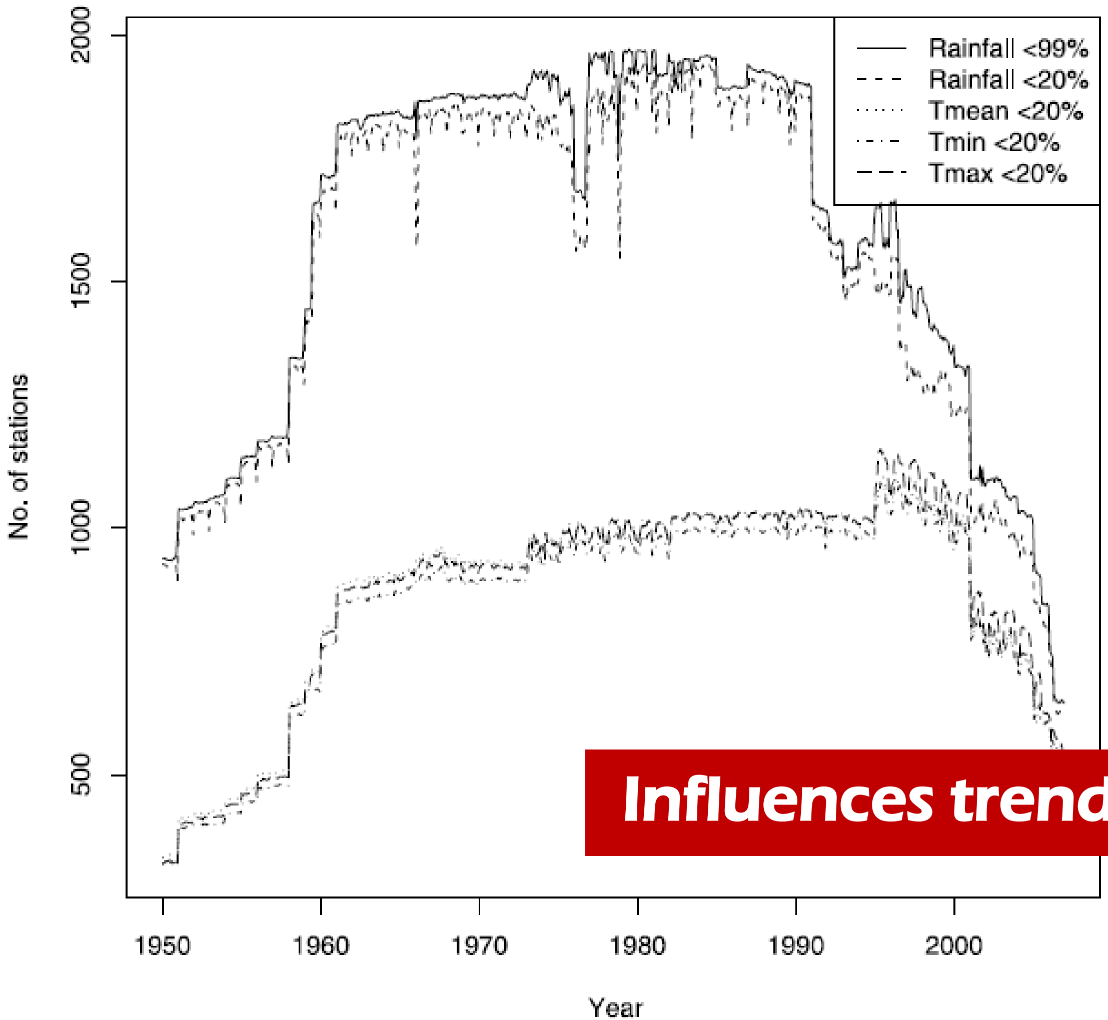
*E-OBS is based on les than **3.000** stations, spread unevenly across approximately **18.000** 0.22 grid-boxes..*

EOBS v07: length of station records (since 1950) [years], **daily mean temperature (tg)**

total number of stations: 3796



TEMPORAL AND SPATIAL INHOMOGENEITIES (cont'd)



Influences trend estimates!

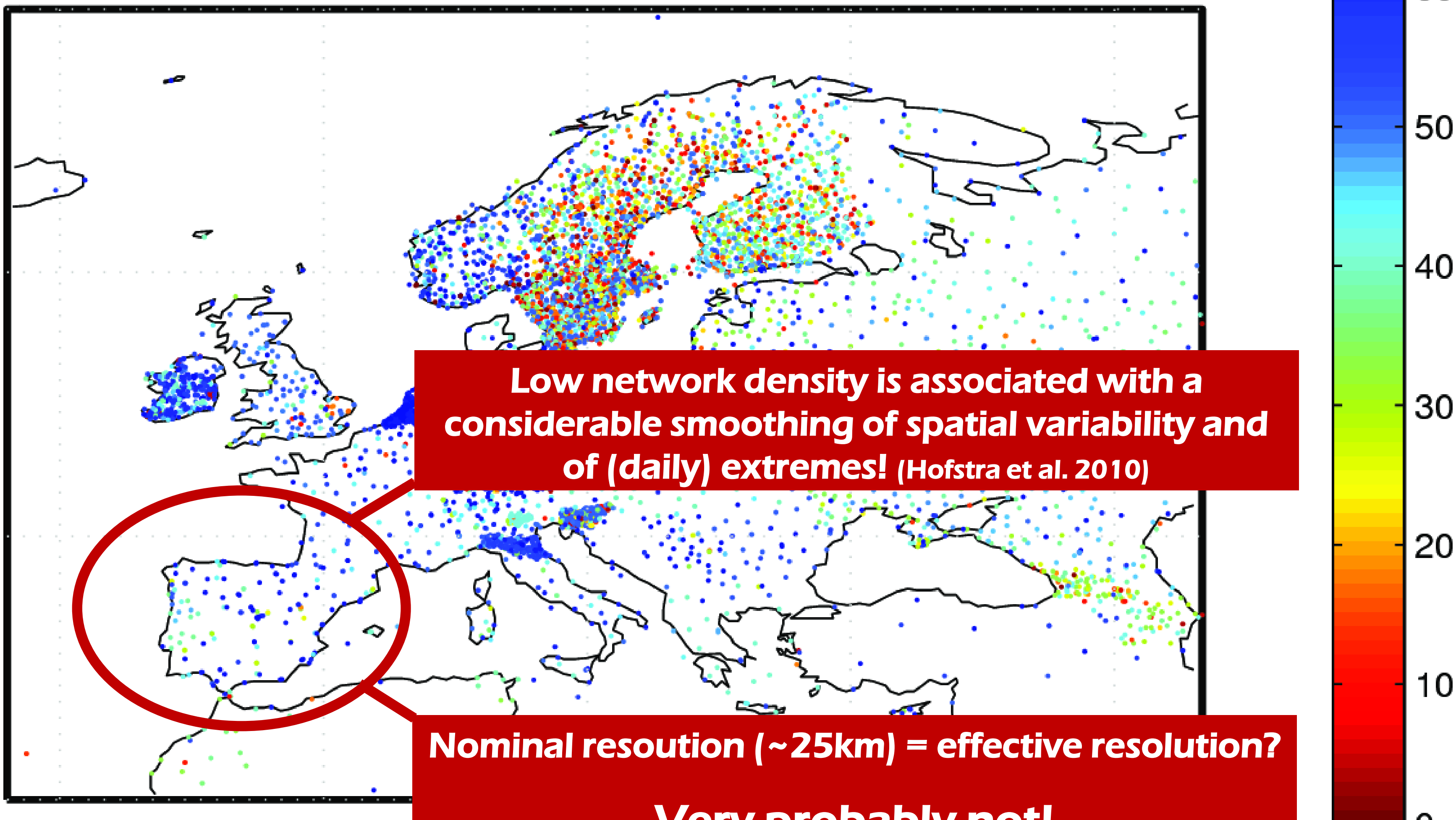
Figure 2. The number of stations with less than 99% and 20% missing observations for each month.

Haylock et al., 2008

TEMPORAL AND SPATIAL INHOMOGENEITIES (cont'd)

1
2
3
4
5
To Consider!
6
7
8

EOBS v07: length of station records (since 1950) [years], **precipitation (rr)**
total number of stations: 7067



Low network density is associated with a considerable smoothing of spatial variability and of (daily) extremes! (Hofstra et al. 2010)

Nominal resoution (~25km) = effective resolution?

Very probably not!

INFLUENCE ON MODEL EVALUATION

RCMs versus **national grid** with high underlying network density

RCMs versus **E-OBS**

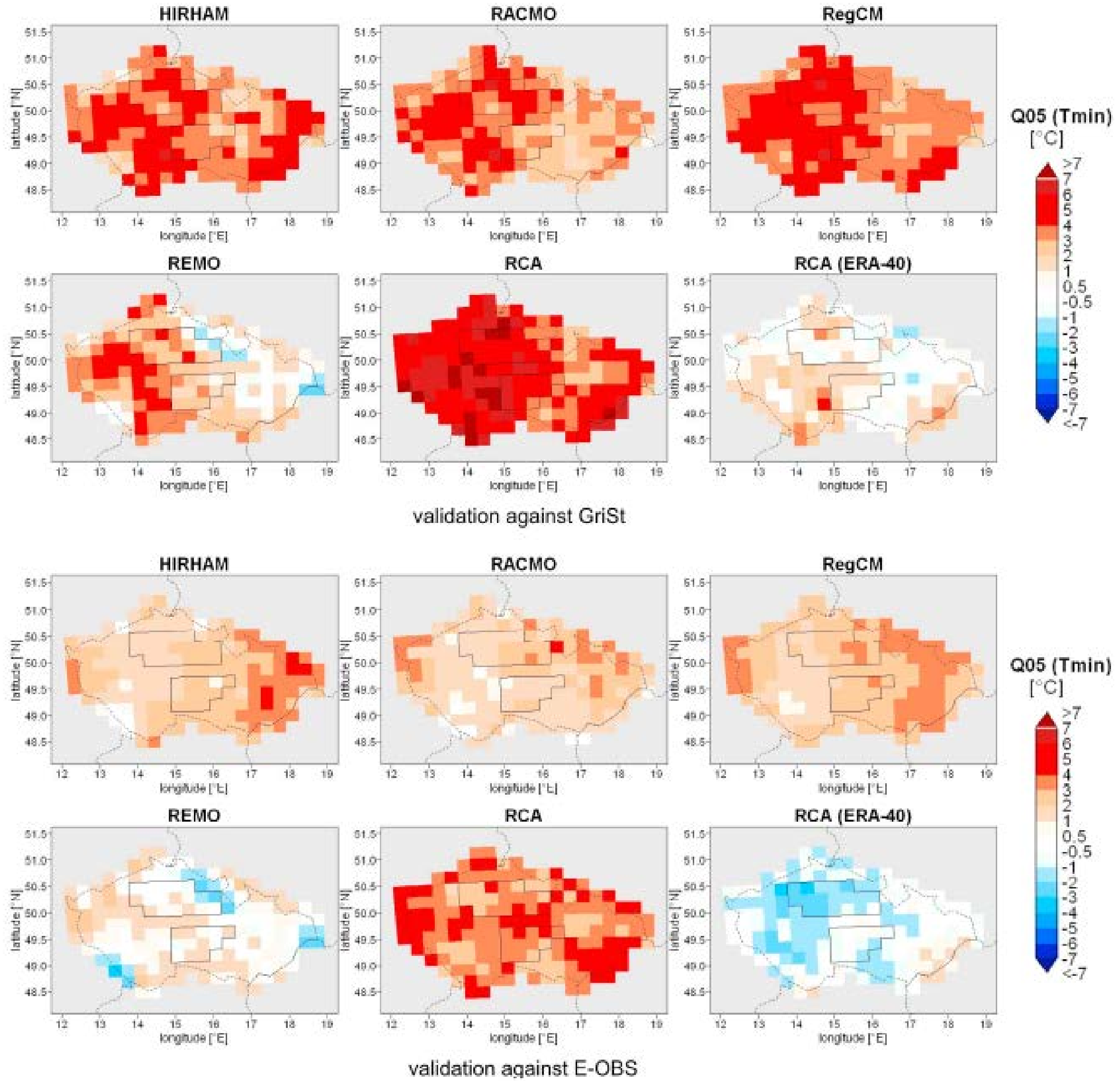
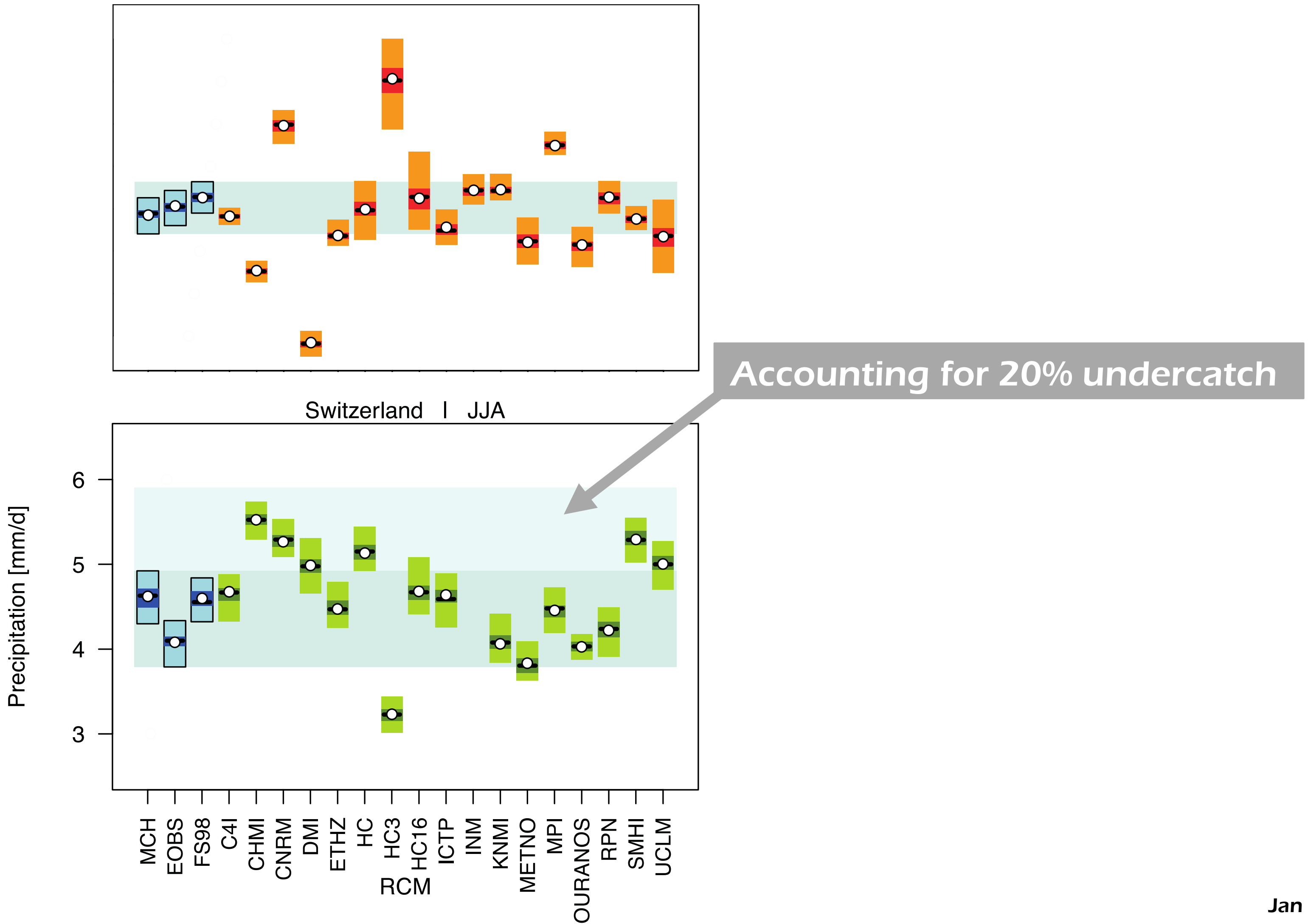


Figure 6. Differences in the 5% quantile of T_{min} in DJF between control simulations of RCMs (1961–1990) and gridded observed data for (top) GriSt and (bottom) E-OBS.

Kysely and Plavcova, 2010

INFLUENCE ON MODEL EVALUATION

Evaluation of mean annual temperature (1981-2000) and mean annual precipitaton (1971-1998) over Switzerland: 3 observational references and 17 ERA40-driven RCMs

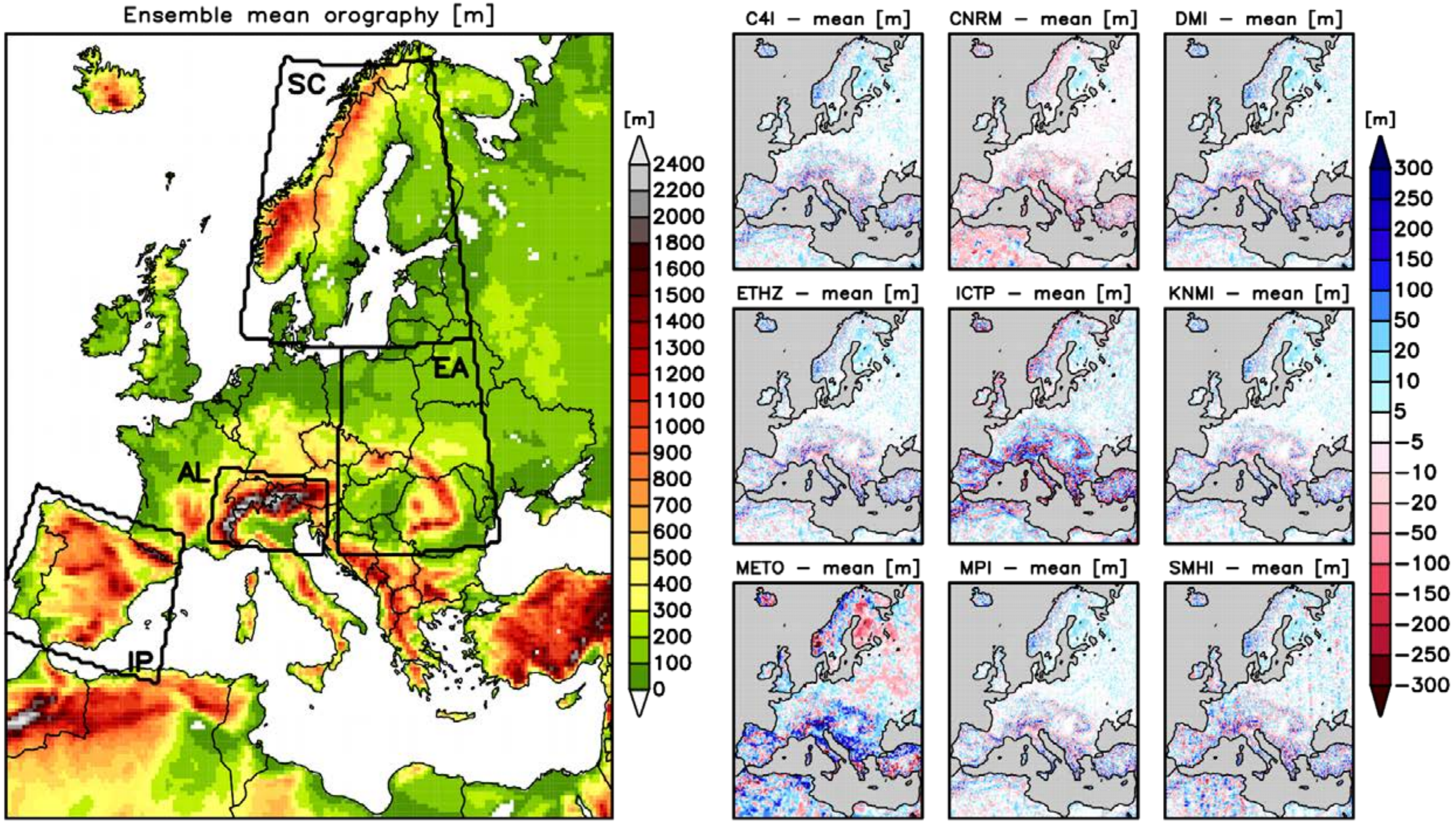


Jan Rajczak, IAC ETH

SYSTEMATIC TOPOGRAPHIC EFFECTS

1
2
3
4
5
To Consider!
6
7
8

Temperature validation on grids has to account for different surface orographies!



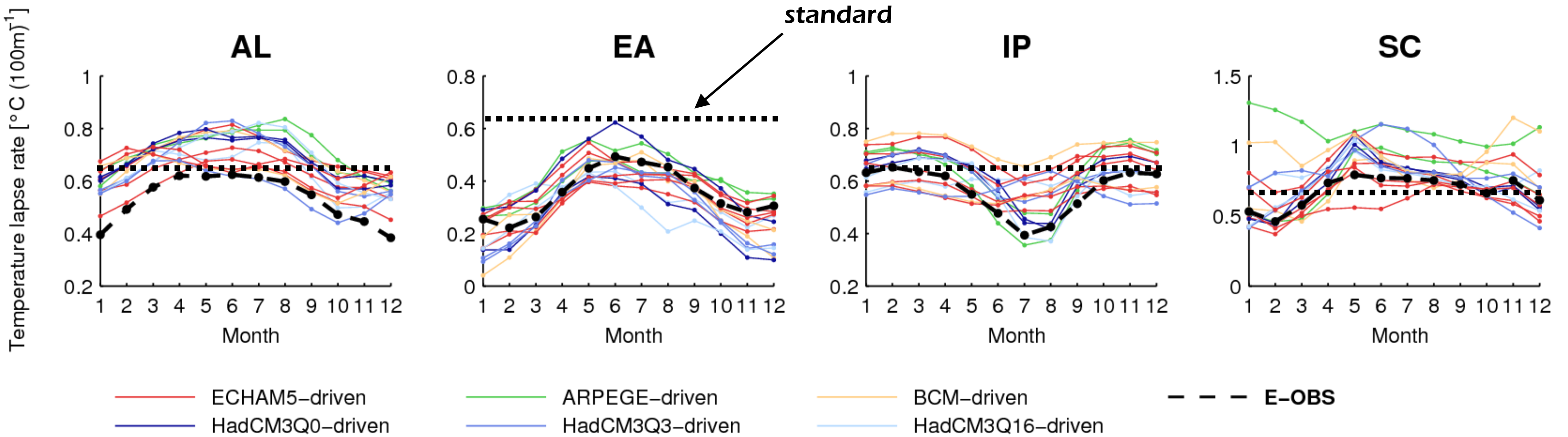
Height-correction is required before comparison/evaluation.
Introduces additional uncertainty into evaluation procedure.
No problem for SD.

SYSTEMATIC TOPOGRAPHIC EFFECTS (cont'd)

1
2
3
4
5
To Consider!
6
7
8

A standard lapse rate of 0.6 or 0.65 K/100m is often applied -> not appropriate in most cases due to regional and seasonal variation of lapse rate!

Mean monthly lapse rates (1961-2000) over the Alps (AL), Eastern Europe (EA), the Iberian Peninsula (IP), and Scandinavia (SC) in E-OBS and the GCM-driven ENSEMBLES RCMs



Alternative (more complicated!):
Apply observed/simulated regional lapse rates for height correction.

PRESENT DAY PERFORMANCE vs. CLIMATE CHANGE SIGNAL

&

NON-STATIONARY MODEL BIASES

OVERVIEW

1
2
3
4
5
To Consider!
6
7
8

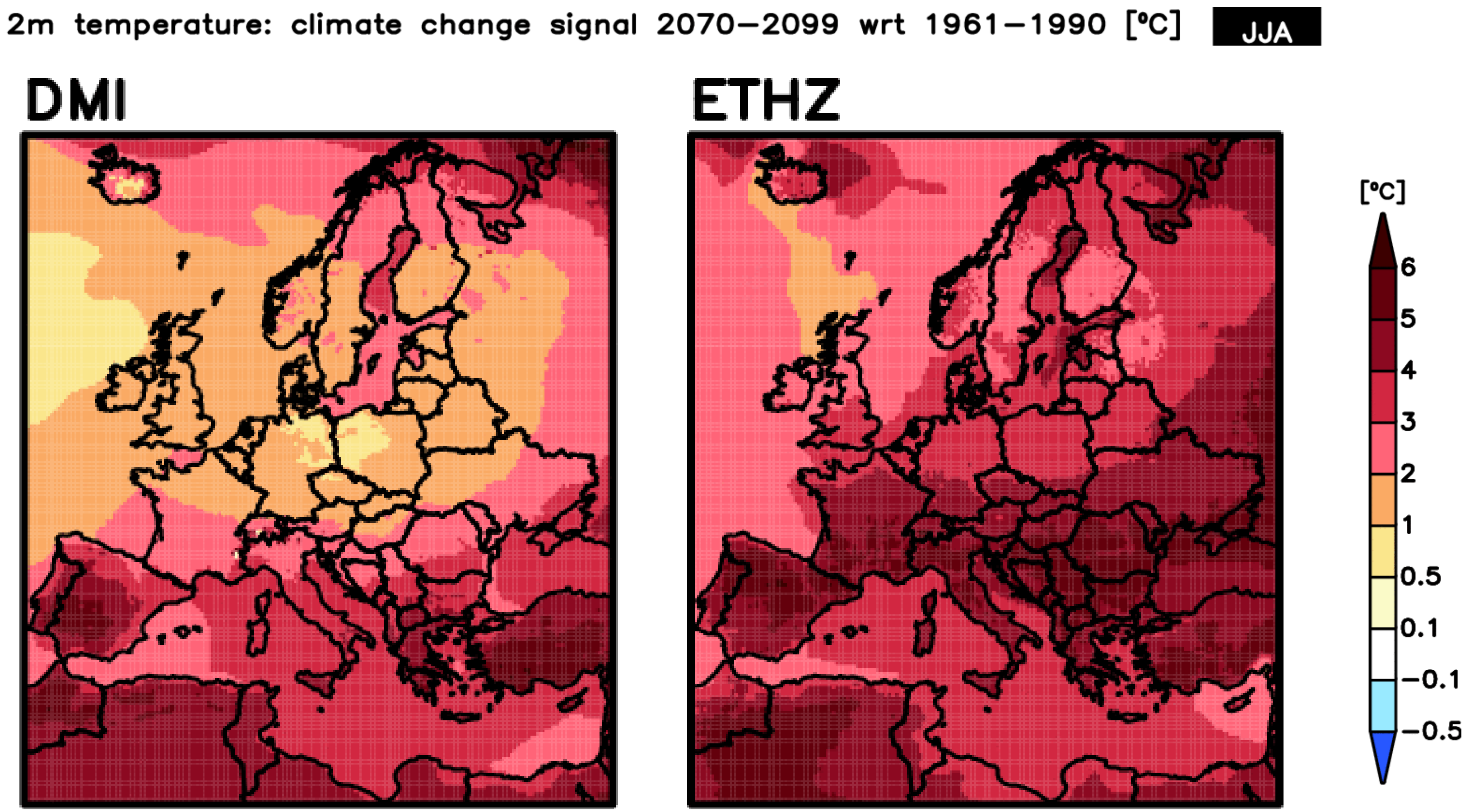
- **Problem**: Model bias cannot necessarily be assumed to be stationary in time, particularly if two different climatic states are considered
- If model biases are non-stationary: Limited significance of evaluating performance in historical periods; bias changes will distort simulated climate change signal!
- Observational and historical simulation record typically too short to diferentiate between two climatic states
- No future observations available for assessing future model biases

Indeed…

… clear relation between skill in present-day climate and simulated climate change signal usually not found

… strong indications for non-stationary biases

1
2
3
4
5
To Consider!
6
7
8



Do these models show a stationary temperature bias on the spatial and temporal scales considered?

TEMPERATURE DEPENDENCY OF RCM BIAS

1
2
3
4
5
To Consider!
6
7
8

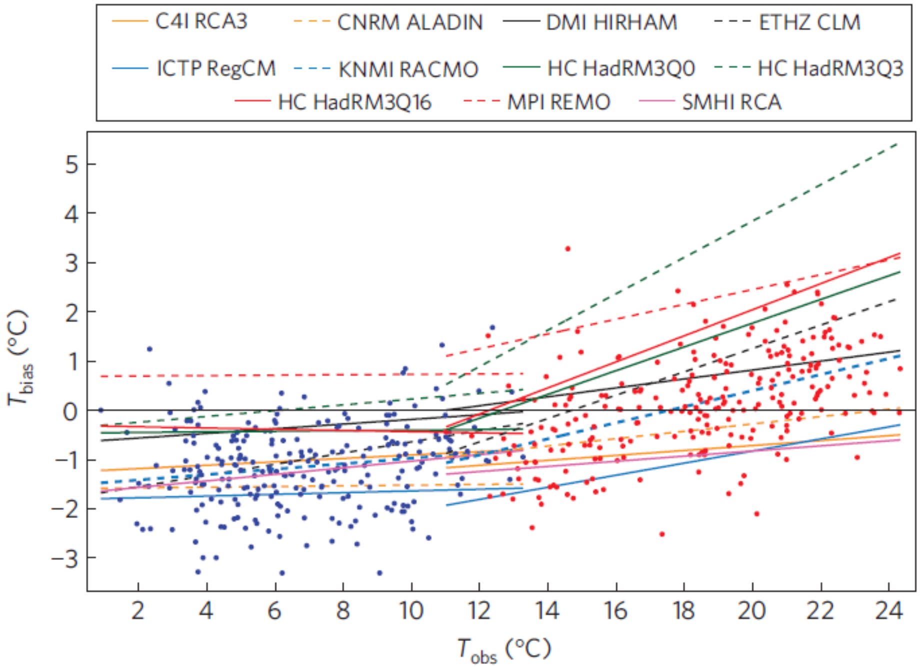
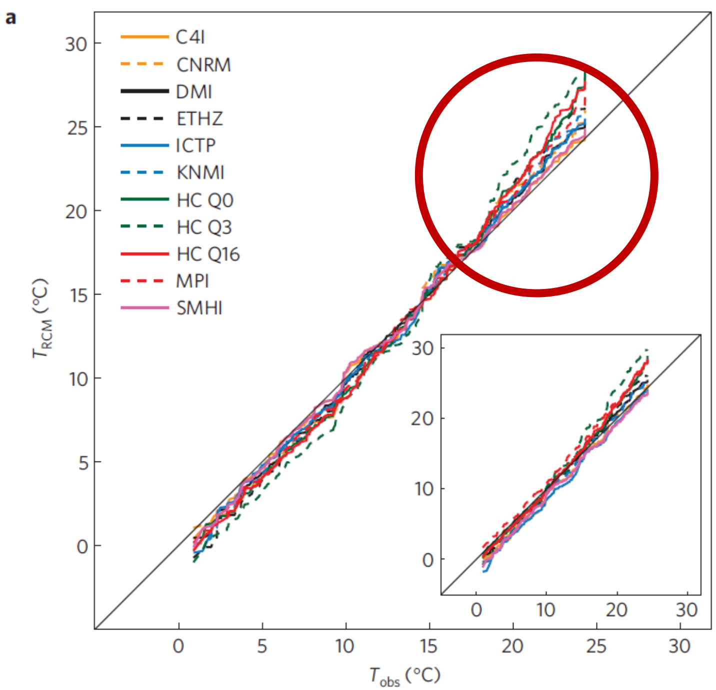


Figure 1 | Model temperature biases. Linear fits to ERA40-driven monthly mean model temperature biases versus observed monthly mean temperatures for the Mediterranean subregion for 1961–2000. Points denote monthly KNMI RACMO values (blue for the cold period NDJFMA and red for the warm period MJJASO) and the dashed blue lines (see legend) are best fits based on these points.

Mediterranean: Most RCMs show an increasing warm summer bias with temperature



Centred Q-Q-Plot for GCM-driven RCMs (1961-2000)

Warmer months with steadily higher temperature bias

Boberg and Christensen, 2012

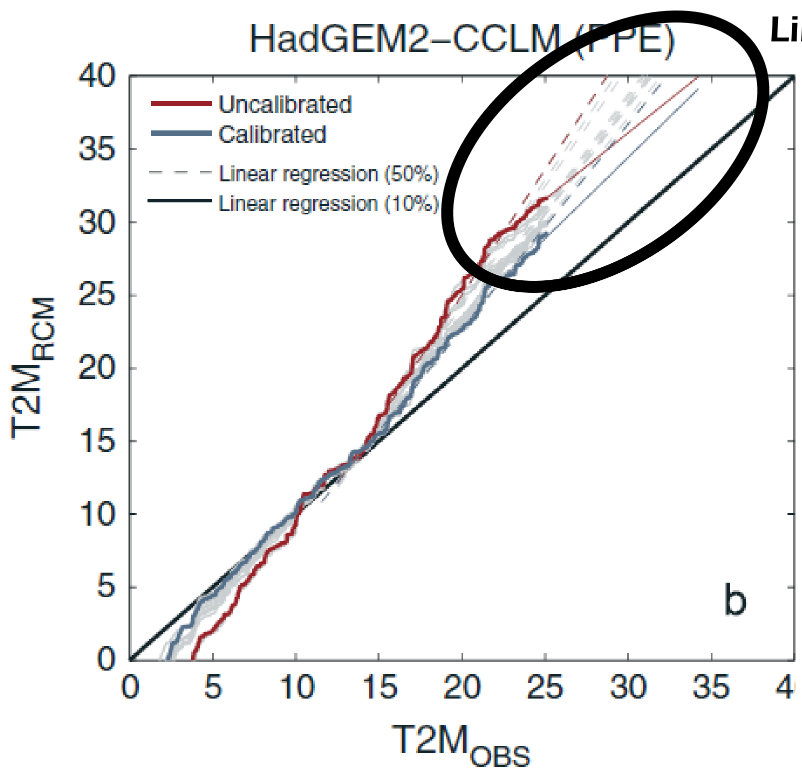
EVALUATION vs. CLIMATE CHANGE SIGNAL

Should we nevertheless focus on the best-performing models?

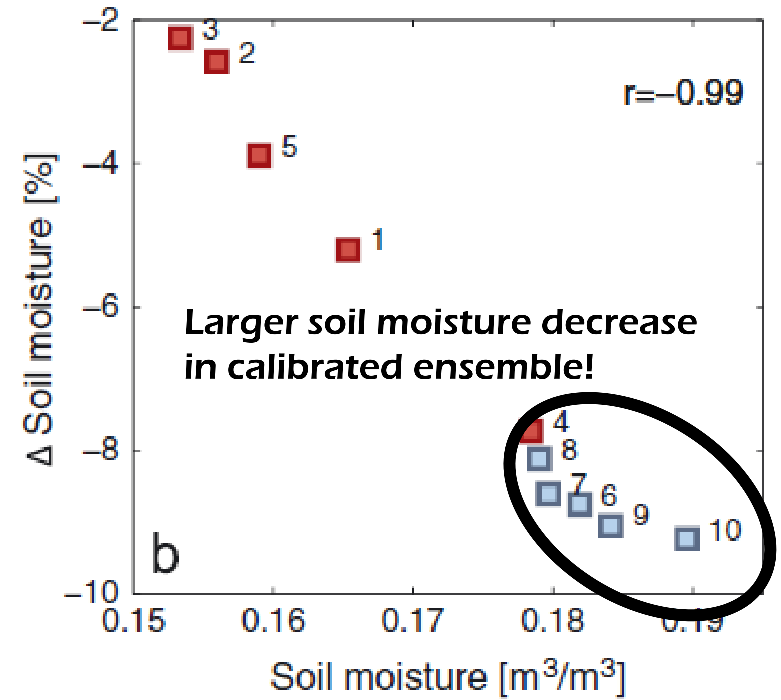
Are bias changes smallest fore these models?

NO!

Linearly increasing bias?



Empirical Q-Q-plot of observed and simulated (COSMO-CLM PPE ensemble) JJA temperatures over the Meditteranean



Soil moisture in control period (x-axis) versus soil moisture changes

Bellprat et al., 2013

EVALUATION vs. CLIMATE CHANGE SIGNAL (cont'd)

1
2
3
4
5
To Consider!
6
7
8

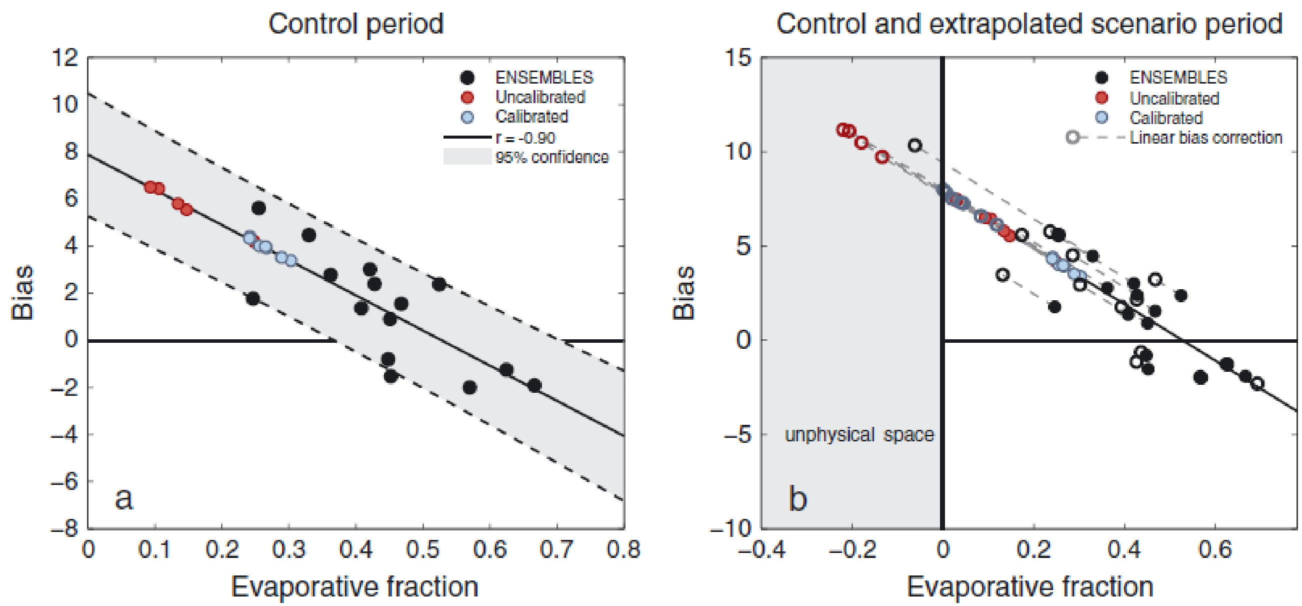


Figure 4. (a) Summer temperature biases for all PPE and MME simulations compared to the respective evaporative fraction over the Mediterranean region. The simulations show a linear relationship (significant at 5%) with a 95% uncertainty range in grey. (b) Filled circles for control period as in (a); in addition, empty circles show extrapolated temperature biases and evaporative fractions for the scenario period, assuming a linearly increasing temperature bias as proposed by BC12, and a linear relationship between bias and evaporative fraction as determined from (a).

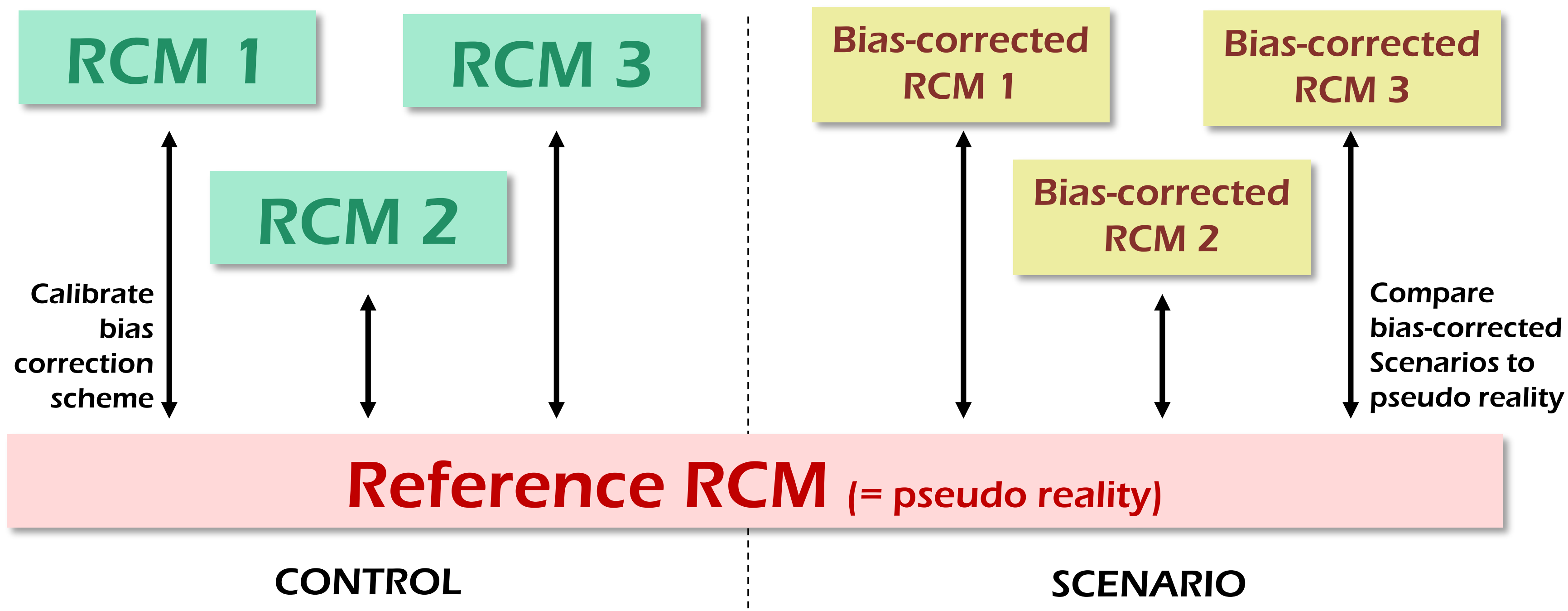
- Translation of linearly increasing model bias to constant model bias in calibrated ensemble, i.e., smaller bias changes in uncalibrated ensemble!
- Regular delta change methodology: uncalibrated ensemble provides the better estimate of climate change signal!

Bellprat et al., 2013

PSEUDO REALITIES

Further indications of non-stationary model biases:

«Pseudo reality» frameworks (e.g., Vrac et al. 2007, Maraun 2012, Bellprat et al. 2013)



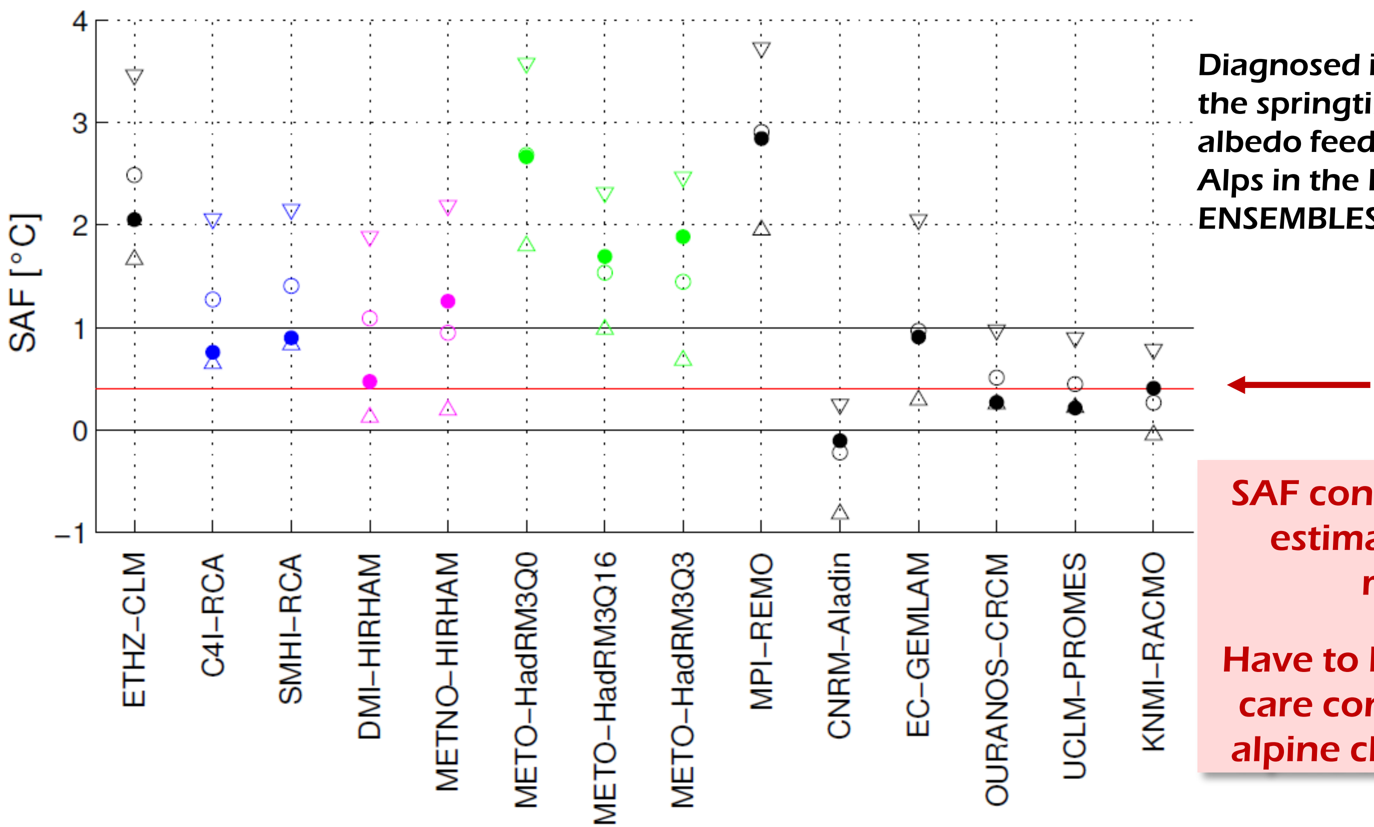
- Cannot uncover all kinds of bias non-stationarities (common non-stationaries possible)
- But: Provides strong evidence for bias non-stationarities over some regions and for some parameters

VALIDATION OF PHYSICAL RELATIONS

PHYSICAL RELATIONS

Very relevant question: **Does a (physically-based) climate model properly represent physical inter-variable relations?**

e.g.:
- Connection between large-scale airflow and local precipitation (Maraun et al.,
- Influence of snow cover on 2m temperatures (snow-albedo feedback!)



Diagnosed intensity of the springtime snow-albedo feedback in the Alps in the ERA40-driven ENSEMBLES RCMs

SAF considerably over-estimated by most models!

Have to be treated with care concerning high-alpine climate change!

Winter et al., 2014

SKILLFUL SCALE

Can a climate model really be analysed and evaluated at ist nominal spatial resolution?
(Several grid cells are required to represent atmospheric phenomena!)

REPRESENTATIVENESS

Should we assume that the simulated location of some phenomenon
is identical to the «true» location?
(or are there systematic spatial shifts in the climate model output)

QUALITY OF BOUNDARY FORCING

The skill of an RCM depends on the quality of the supplied boundary forcing!

SPATIAL CORRELATION OF MODEL BIAS

Biases at individual grid cells cannot be assumed to be independent of each other
(important for hypthesis testing)

OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE METRICS

5 TO CONSIDER!

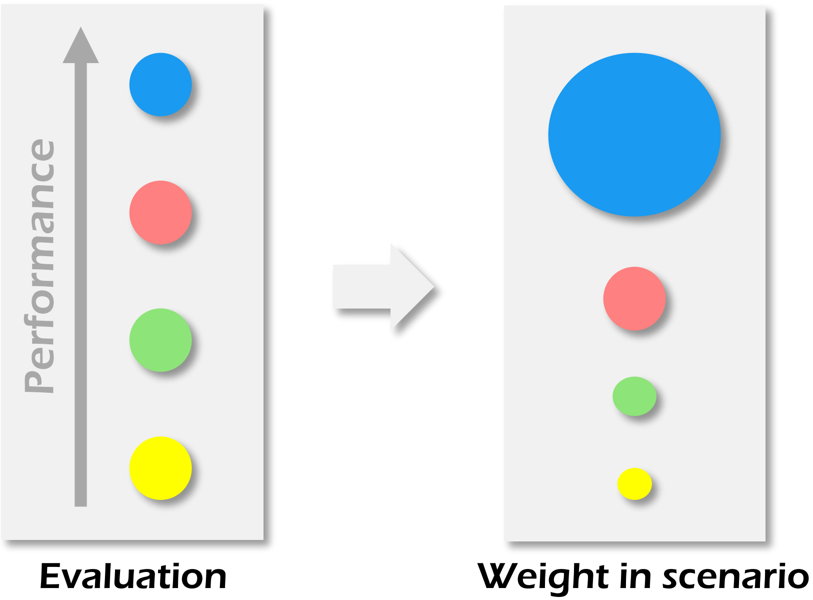
6 MODEL WEIGHTING

7 EXAMPLE

8 SUMMARY & CONCLUSIONS

MODEL WEIGHTING: RATIONALE

- Weight/rank models in multi-model climate projections according to their performance in present day/past climates -> performance-based weights
- Underlying assumption: Models that perform better in present day/past climates are more reliable
- Good performance -> high weight, bad performance -> low weight
- Simplest scheme: equal weight (one vote for each model)
- Extreme but widespread variant: Selection or exclusion of individual models



MODEL WEIGHTING: RISKS

1
2
3
4
5
6 Weighting
7
8

Risks:

- Weighting schemes potentially have a strong influence on final outcome
- But: Any weighting scheme includes subjective elements, related to both the selection of the input information and the way in which this information is used
- Choice of metric can impact weight/rank!
- Some chance that weights are inappropriate (e.g. non-stationary biases) -> equal weighting preferable
- Eliminating models can be risky!
- «Unless there is a clear relation between what we observe and what we predict, the risk of reducing the projection accuracy by inappropriate weights appears to be higher than the prospect of improving it by optimum weights.» (Weigel et al., 2010)
- RCMs: Driving GCM determines large-scale climate change signals to a considerable extent and also needs to be considered

THE ENSEMBLES WEIGHTING SPECIAL ISSUE: OVERVIEW

1
2
3
4
5
6
Weighting
7
8

Vol. 44: 179–194, 2010
doi: 10.3354/cr00916

CLIMATE RESEARCH
Clim Res

Published December 10

Contribution to CR Special 23 'Regional Climate Model evaluation and weighting'

OPEN ACCESS

Weight assignment in regional climate models

Jens Hesselbjerg Christensen[1,*], Erik Kjellström[2], Filippo Giorgi[3], Geert Lenderink[4], Markku Rummukainen[2,5]

- Series of papers in a 2010 special issue
- Exploratory performance-based weighting of the ENSEMBLES models (ERA40-driven evaluation runs)
- Combination of 6 specifically designed performance metrics
- Exploitation of different aggregation/combination procedures of the individual weights

THE ENSEMBLES WEIGHTING SPECIAL ISSUE: METRICS

Metric	Variables	Period	Reference data set	Data type	Seasons	Area	Source
f_1	Z_{500hPa}	1961–2000	ERA40	Daily	DJFM, JJAS	Minimum domain	Sanchez-Gomez et al. (2008)
f_2	P, T	1961–2000	CRU TS1.2	Monthly	DJF, MAM, JJA, SON	EUR	Coppola et al. (2010)
f_3	P, T_{min}, T_{max}	1961–1990	EOBS2.0	Daily	DJF, MAM, JJA, SON	EUR	Kjellström et al. (2010)
f_4	P, T_{min}, T_{max}	1971–2000	EOBS2.0	Daily	DJF, MAM, JJA, SON	EUR	Lenderink (2010), Buonomo (unpubl.)
f_5	T	1961–2000	EOBS2.0	Monthly	DJF, MAM, JJA, SON, ANN	Average of 8 subdomains	Lorenz & Jacob (2010)
f_6	P, T	1961–2000	EOBS2.0	Monthly		EUR	Halenka et al. (unpubl.)

Christensen et al., 2010

f_1 **Large-scale circulation based on a weather regime classification**
Are the RCMs able to reproduce observed weather regimes?

f_2 **Meso-scale signal based on seasonal mean temperature and precipitation**
Validation of spatial patterns after removing large-scale component by spatial filter.

f_3 **Probability density distributions of daily and monthly temperature and precip.**
Evaluation of PDFs of daily temperature and precipitation.

f_4 **Extremes in terms of re-occurrence periods for temperature and precipitation**
Evaluation of daily temperature and precipitation extremes (empirical PDF and GEV).

f_5 **Long-term trends in temperature**
Evaluation of linear temperature trends during the ERA40 period.

f_6 **Annual cycle in temperature and precipitation**
Representation of the mean annual cycle of temperature and precipitation.

THE ENSEMBLES WEIGHTING SPECIAL ISSUE: COMBINATION

- Combination of the six individual metrics into one final weight for each RCM
- Several options tested

(1) W_{PROD}: Simple multiplication of weights

$$W_{\mathrm{PROD}} = \prod_{i=1}^{6} f_i^{\,n_i} \qquad (1)$$

where all the individual weights are first normalized to yield a value between 0 and 1 (see accompanying papers in this issue, and references in Table 2) before entering Eq. (1). The final weight (W_{PROD}) for each model is also normalized across the models in order to facilitate application to the model ensemble. The simple multiplication can be refined by allowing for the exponent n_i in Eq. (1) to be chosen as any positive number. Assuming $n_i = 1$ implies weighting the various metrics equally, whereas choosing any positive value different from 1 shifts the emphasis across the individual metrics (a value of 0 would imply equal weighting of the RCMs). This latter approach would be warranted if some metrics were considered to be more fundamental than others, for example when applying the method to a specific impact sector or if some of the metrics were not independent from each other. Other methods could be introduced based on

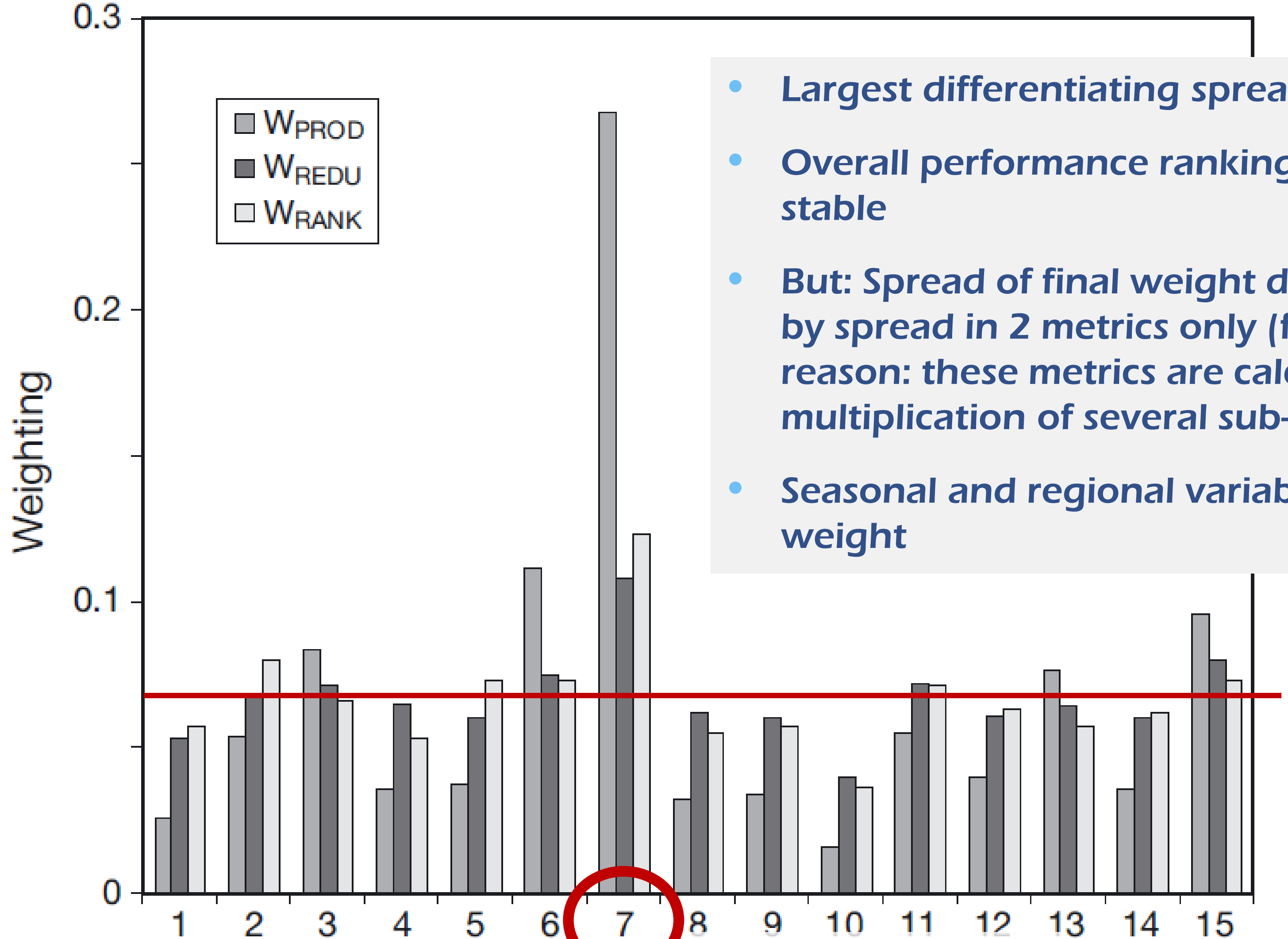
(2) W_{REDU}

Spread of weights is reduced by varying n_i (ratio between highest and lowest individual model weight = 1.2 for each weight)

(3) W_{RANK}

Models are first ranked for each individual metric. Ranks are summed and transformed into final weight.

THE ENSEMBLES WEIGHTING SPECIAL ISSUE: RESULTS



- Largest differentiating spread for W_{PROD}
- Overall performance ranking rather stable
- But: Spread of final weight dominated by spread in 2 metrics only (f_2, f_4); reason: these metrics are calculated by multiplication of several sub-metrics
- Seasonal and regional variability of final weight

Fig. 2. W_{PROD}, W_{REDU} and W_{RANK} for each of the regional climate models (RCMs). See Table 1 for model numbers and Table 3 for definition of weights

Christensen et al., 2010

THE ENSEMBLES WEIGHTING SPECIAL ISSUE: RESULTS

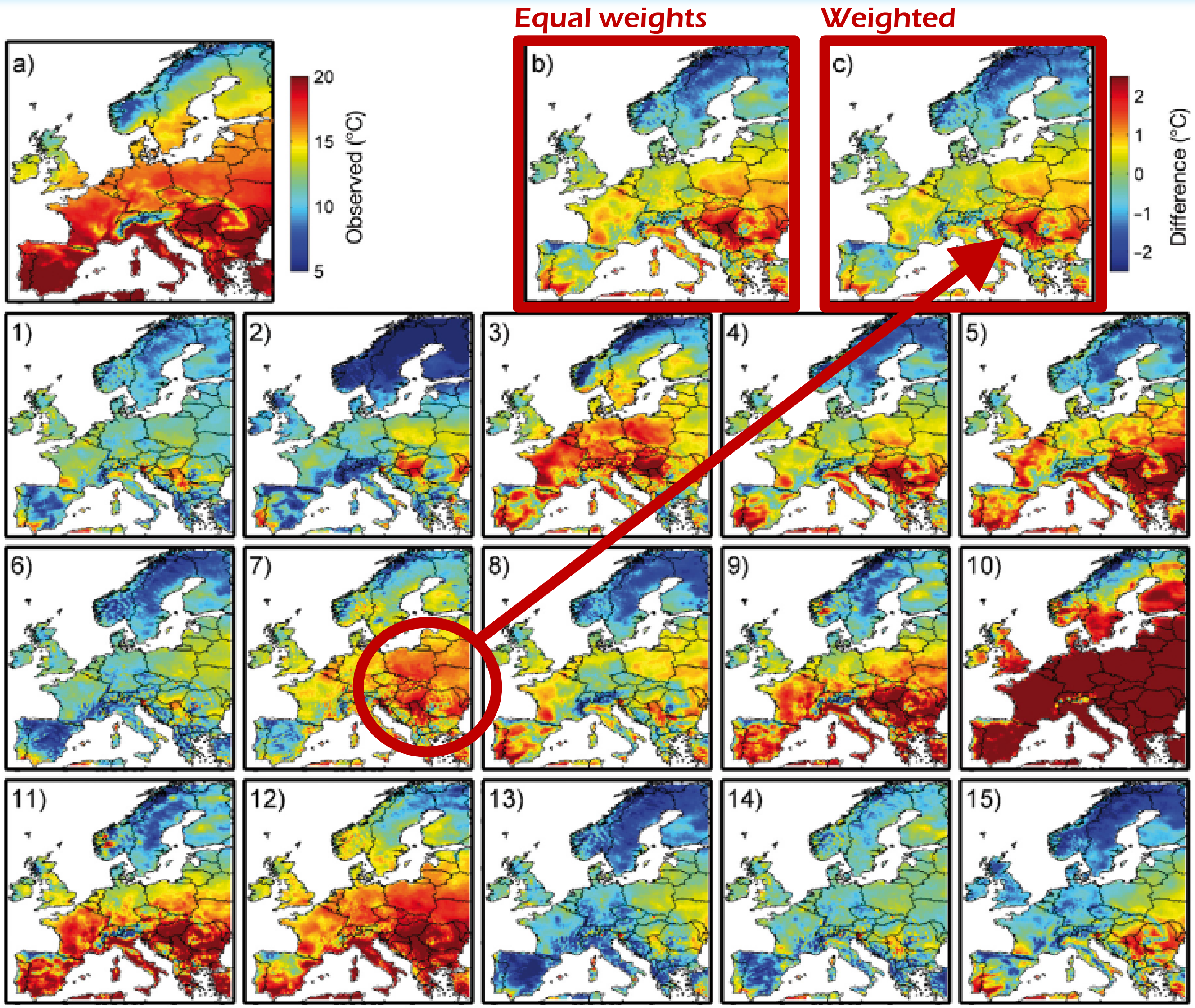


Fig. 3. Summer (JJA) temperature at 2 m (T_{2m}; °C). (a) E-OBS; (b) difference between unweighted ensemble mean and E-OBS; and (c) difference between weighted ensemble mean and E-OBS. Panels 1–15 show the difference between model and E-OBS for each individual regional climate model (see Table 1 for model numbers). The left-most color scale applies to panel (a) only; the right-most color scale applies to all other panels

Christensen et al., 2010

THE ENSEMBLES WEIGHTING SPECIAL ISSUE: CONCLUSIONS

- Combination scheme only slightly impacts final ranking
- Quality of weighted ensemble mean NOT consistently superior to equal weighting
- Model weights are relative and only apply to exactly this ensemble
- Metrics do not consider all aspects of model quality!
- Ranking for individual metrics partly different from overall weight
- Intrinsic uncertainty: Quality of reference observations (E-OBS)!
- Correlation between individual metrics
- Choice of metrics and their combination subjective!
- Systematic GCM biases in scenario studies not yet considered…
- Non-stationary model biases not considetred

OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE METRICS

5 TO CONSIDER!

6 MODEL WEIGHTING

7 **EXAMPLE**

8 SUMMARY & CONCLUSIONS

VALIDATION EXAMPLE

Geosci. Model Dev., 7, 1297–1333, 2014
www.geosci-model-dev.net/7/1297/2014/
doi:10.5194/gmd-7-1297-2014

Geoscientific Model Development
Open Access

Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble

S. Kotlarski[1], K. Keuler[2], O. B. Christensen[3], A. Colette[4], M. Déqué[5], A. Gobiet[6], K. Goergen[7,8], D. Jacob[9,10], D. Lüthi[1], E. van Meijgaard[11], G. Nikulin[12], C. Schär[1], C. Teichmann[9,10], R. Vautard[13], K. Warrach-Sagi[14], and V. Wulfmeyer[14]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[2]Chair of Environmental Meteorology, Brandenburg University of Technology (BTU), Cottbus-Senftenberg, Germany
[3]Danish Meteorological Institute, Copenhagen, Denmark
[4]Institut National de l'Environnement Industriel et des Risques (INERIS), Verneuil-en-Halatte, France
[5]Météo-France/CNRM, CNRS/GAME, Toulouse, France
[6]Wegener Center for Climate and Global Change, University of Graz, Graz, Austria
[7]Centre de Recherche Public – Gabriel Lippmann, Belvaux, Luxembourg
[8]Meteorological Institute, University of Bonn, Bonn, Germany
[9]Climate Service Center 2.0, Hamburg, Germany
[10]Max Planck Institute for Meteorology, Hamburg, Germany
[11]KNMI Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[12]Swedish Meteorological and Hydrological Institute, Norrköping, Sweden
[13]Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CEA/CNRS/UVSQ, Gif-sur-Yvette, France
[14]Institute of Physics and Meteorology, University of Hohenheim, Stuttgart, Germany

Correspondence to: S. Kotlarski (sven.kotlarski@env.ethz.ch)

Received: 2 December 2013 – Published in Geosci. Model Dev. Discuss.: 14 January 2014
Revised: 9 May 2014 – Accepted: 13 May 2014 – Published: 10 July 2014

Abstract. EURO-CORDEX is an international climate downscaling initiative that aims to provide high-resolution climate scenarios for Europe. Here an evaluation of the ERA-

The analysis confirms the ability of RCMs to capture the basic features of the European climate, including its variability in space and time. But it also identifies nonnegligible

OBJECTIVES

Document skill of EURO-CORDEX RCM ensemble in reproducing present-day climate over Europe under "perfect boundary conditions" (ERA-Interim forcing)

Highlight common model deficiencies and areas of necessary model improvements

Assess possible progress wrt. ENSEMBLES experiments

Establish a quality standard for future model developments

OVERVIEW ON PERFORMANCE

- Consider both the 12 km and the 50 km ensemble
- Focus on temperature and precipitation
- Focus on monthly, seasonal and annual mean statistics
- Apply simple and reproducible metrics

➡
- Study is of a descriptive nature and does not try to ultimately explain individual model biases
- Potential benefits of higher resolution not explicitly addressed

DATA AND METHODS

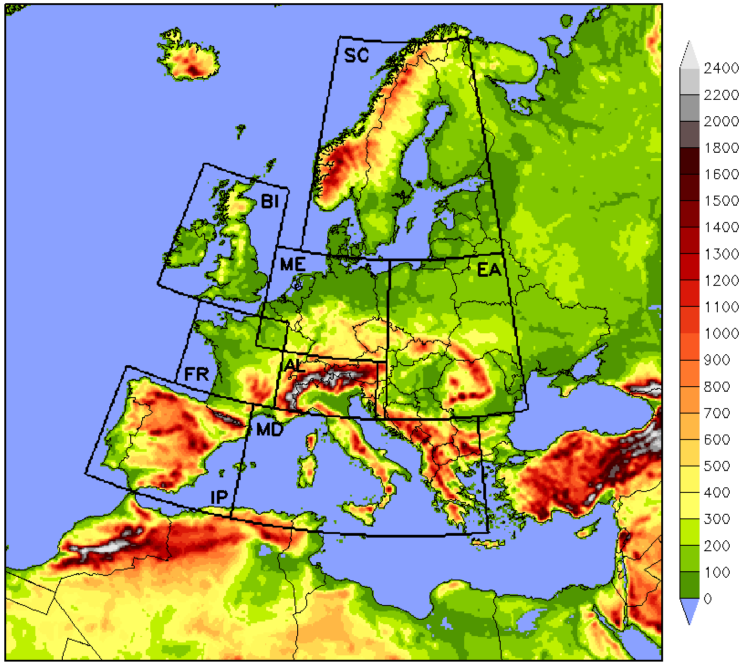
RCM Data

- **ERA-Interim driven EURO-CORDEX ensemble (1989-2008)**
- **EUR-44 (50 km): 8 experiments**
- **EUR-11 (12 km): 9 experiments**
- **6 different RCMs, 1 global model**

CCLM 4.8.17 (CLMCOM)
REMO 4.8.17 (CSC)
RCA 4 (SMHI)
RACMO 2.2 (KNMI)
HIRHAM 5 (DMI)

WRF 3.3.1 (IPSL-INERIS)
WRF 3.3.1 (CRP-GL)
WRF 3.3.1 (UHOH)
ARPEGE 5.1 (CNRM)

- **16 ERA40-driven ENSEMBLES runs (25 km, 1981-2000)**



Observational Reference

- **E-OBS v07, 0.22° (25 km), daily resolution**
- **EUR-44 evaluated on 50 km, EUR-11 evaluated on 25 km**

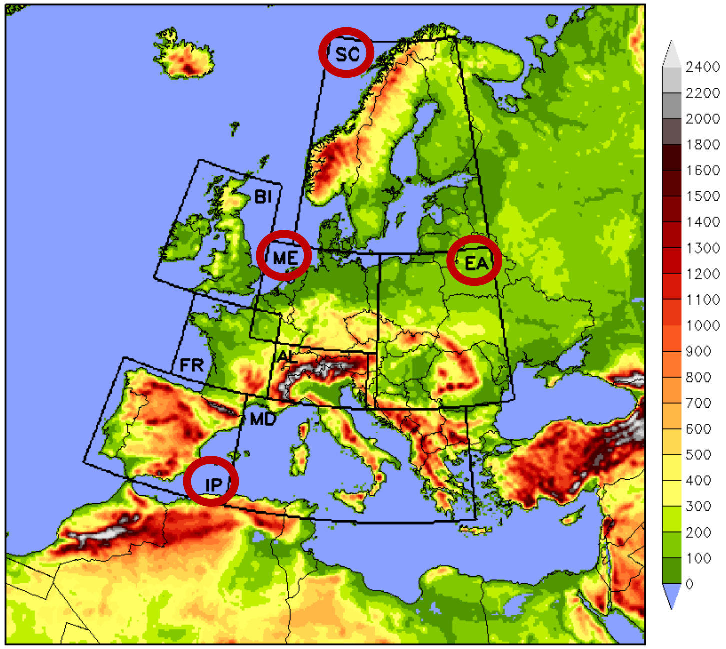
DATA AND METHODS (cont'd)

1
2
3
4
5
6
7
Example
8

Evaluation Metrics

1. Seasonal mean biases at grid point scale for entire Euro-CORDEX domain (EUR-11)

2. Eight metrics applied to eight analysis regions, describing different aspects of model performance (EUR-11 and EUR-44)
   - Temporal and spatial means
   - Spatial variability
   - Temporal variability
   - Mean annual cycle



EVALUATION METRICS

1

2 **BIAS** Difference (model - observations) of climatol. annual and seasonal mean values (regional averages) — Temporal and spatial means

3 **95%-P** 95th percentile of all absolute grid point differences (model - observations) based on climatological annual and seasonal mean values

4 **PACO** Pattern correlation between modeled and observed climatological annual and seasonal mean values at all grid points

5 **RSV** Ratio of spatial variances of all grid points (model over observations) of climatological annual and seasonal mean values

Spatial variability

6 **RIAV** Ratio of interannual variance (model over observations) of time series of annual and seasonal mean values (regional averages)

7 Example **TCOIAV** Correlation between modeled and observed time series of annual and seasonal mean values (regional averages)

Temporal variability

8 **CRCO** Spearman rank correlation between modeled and observed climatological monthly mean values (regional averages)

ROYA Ratio (model over observations) of yearly amplitudes (difference between maximum and minimum) of climatological monthly mean values (regional averages)

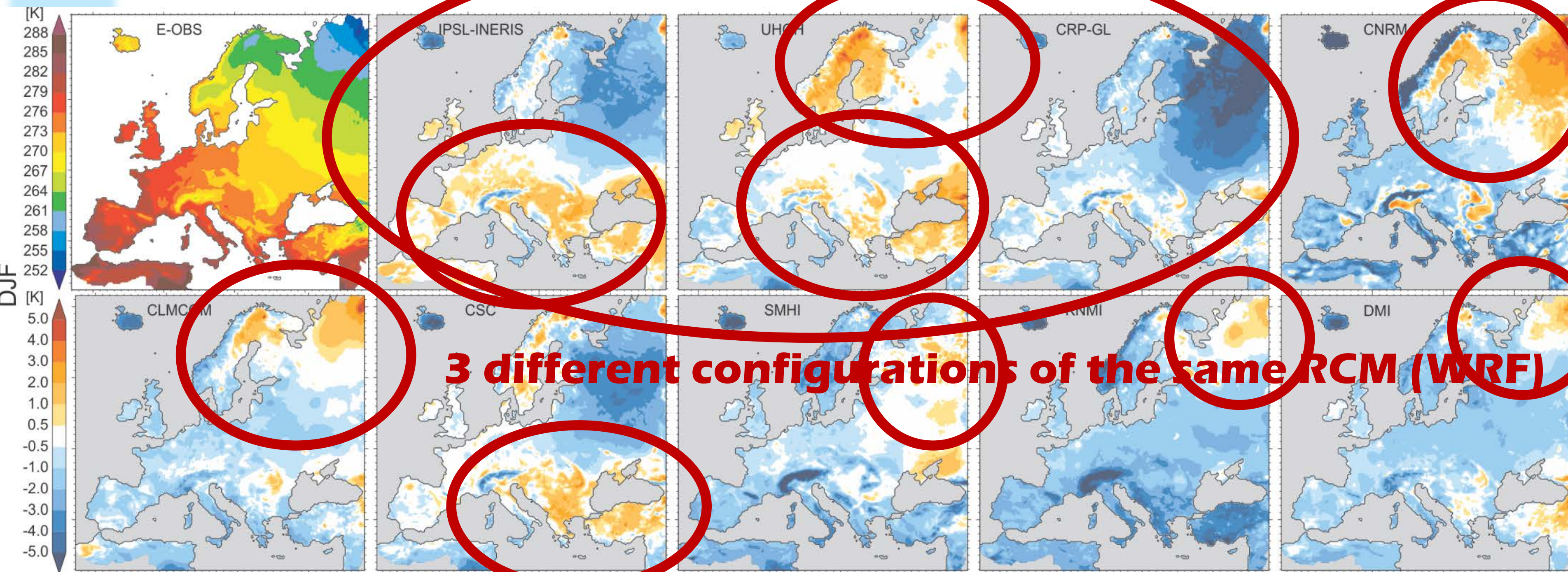
Mean annual cycle

Only selection shown ...

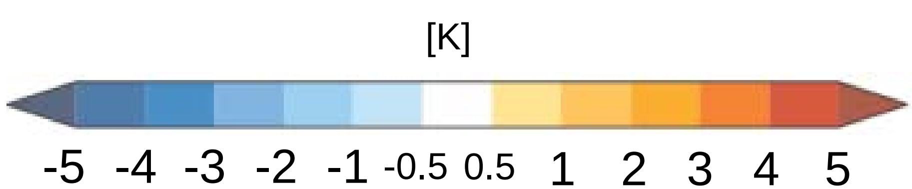
WINTER (DJF) TEMPERATURE BIAS mean 1989-2008, EUR-11

Warm bias over N and NE Europe

Large variety of bias patterns, but cold bias dominates



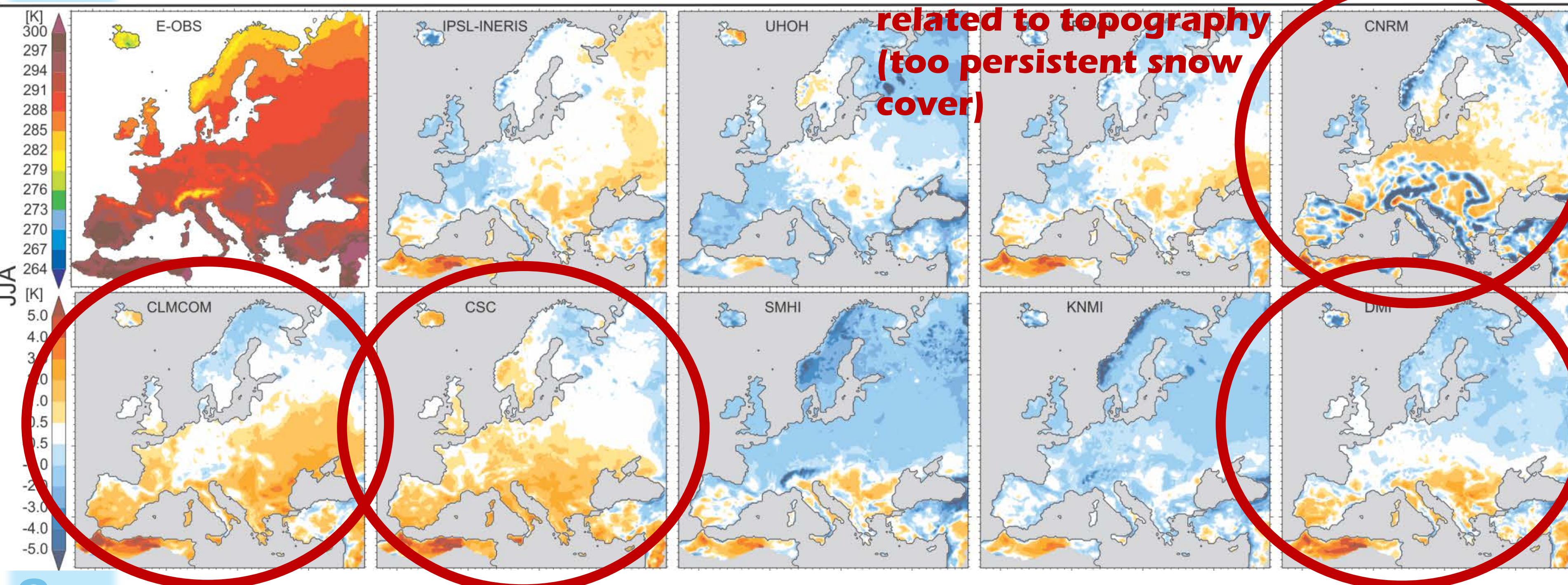
Warm bias over S and SE Europe



SUMMER (JJA) TEMPERATURE BIAS mean 1989-2008, EUR-11

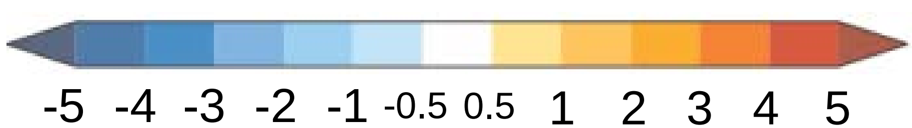
1

Bias pattern strongly related to topography (too persistent snow cover)



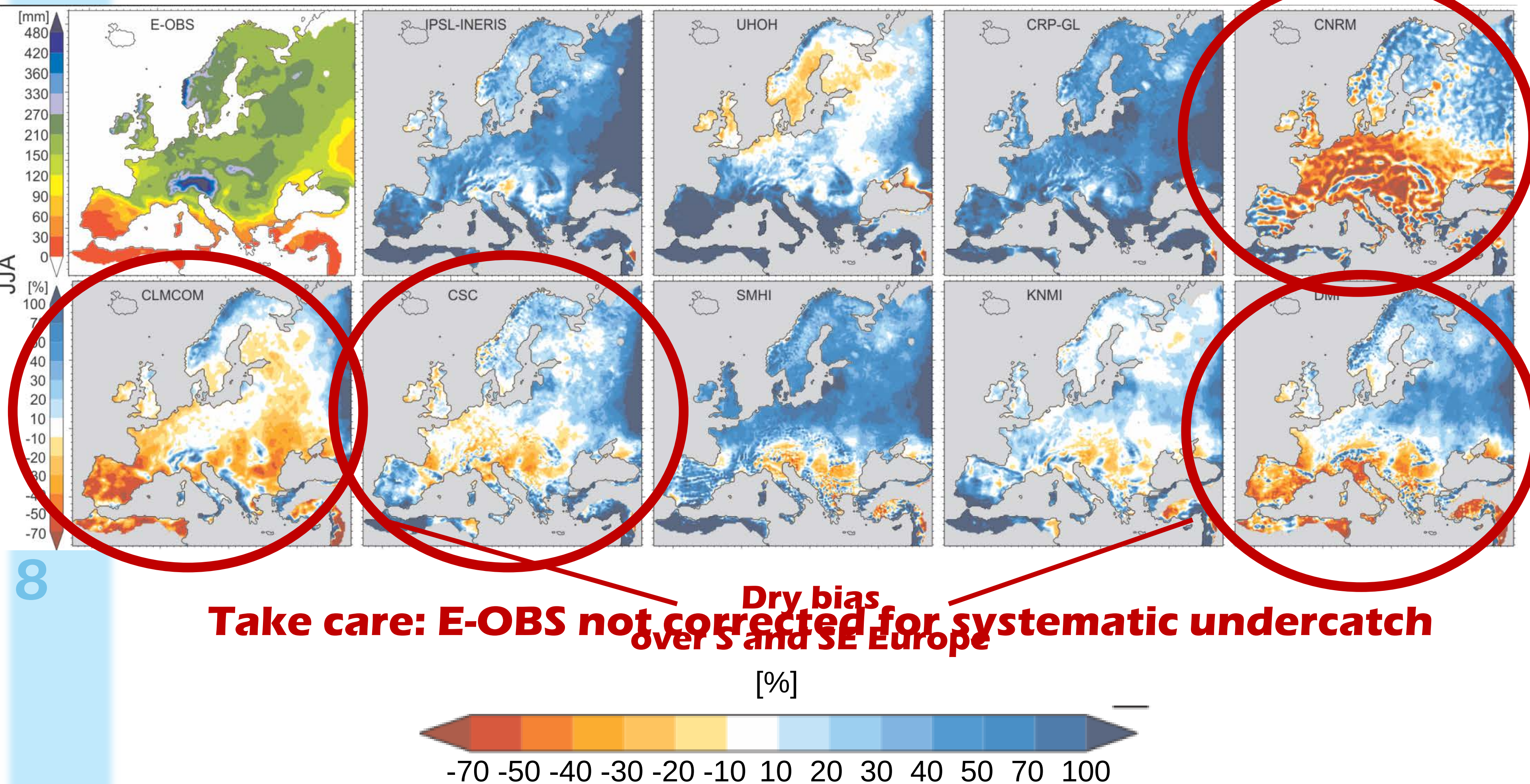
8

Pronounced warm summer bias (S and SE Europe)



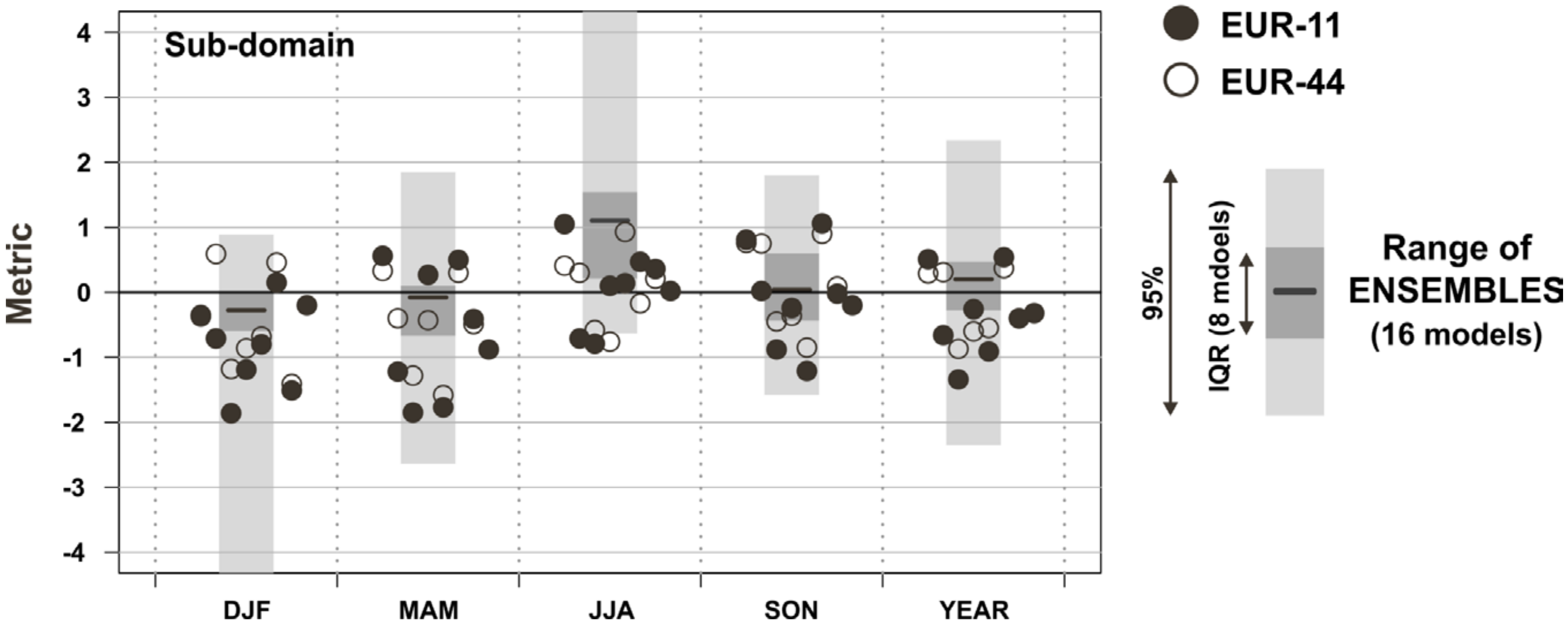
SUMMER (JJA) PRECIPITATION BIAS mean 1989-2008, EUR-11

Wet bias dominates



Take care: E-OBS not corrected for systematic undercatch

EXPLANATION OF OVERVIEW FIGURES

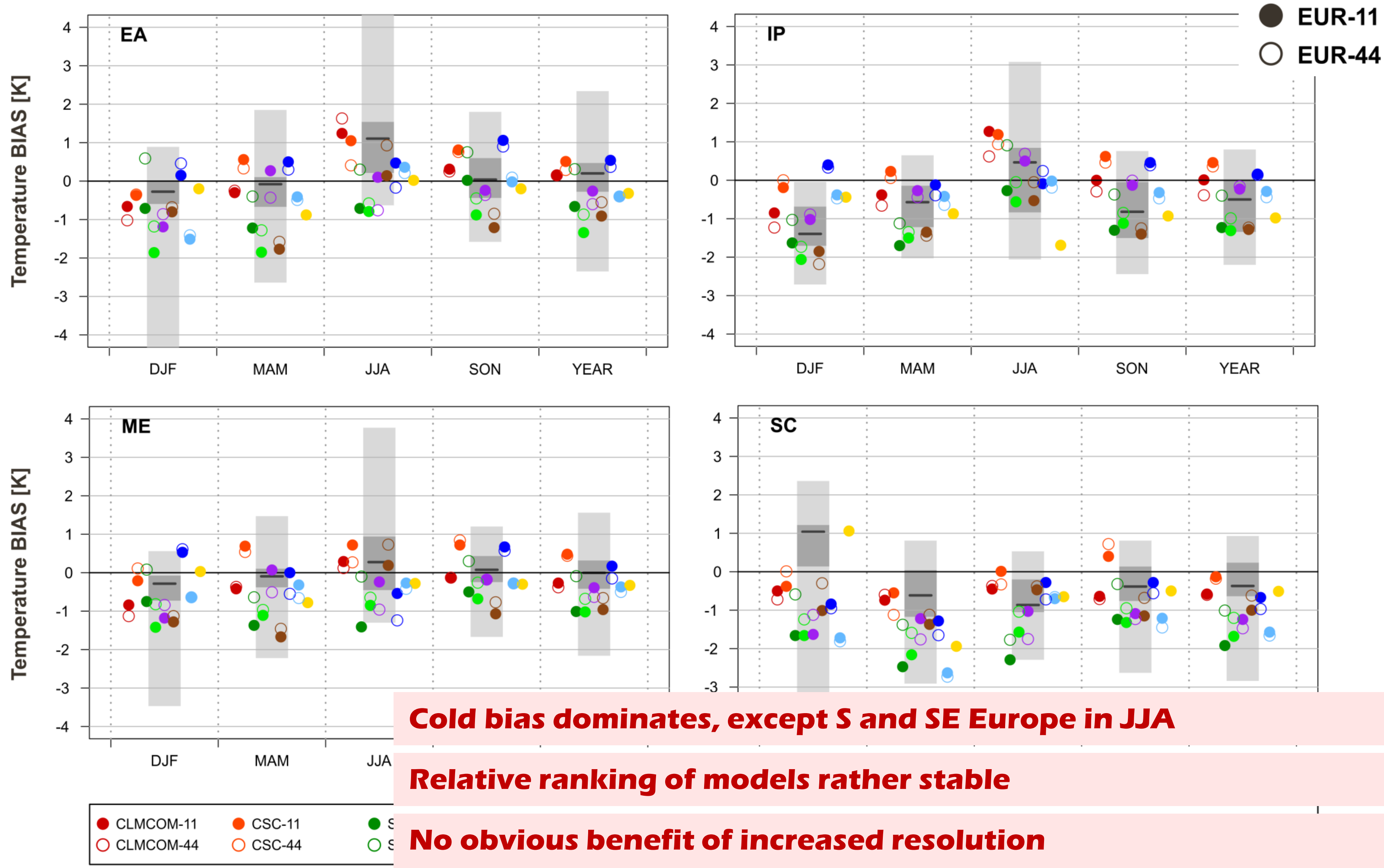


Sub-domains shown:

Eastern Europe (EA), Iberian Peninsula (IP), Mid-Europe (ME), Scandinavia (SC)

TEMPERATURE: Regional and temporal mean bias [K]

1
2
3
4
5
6
7
Example
8



Cold bias dominates, except S and SE Europe in JJA

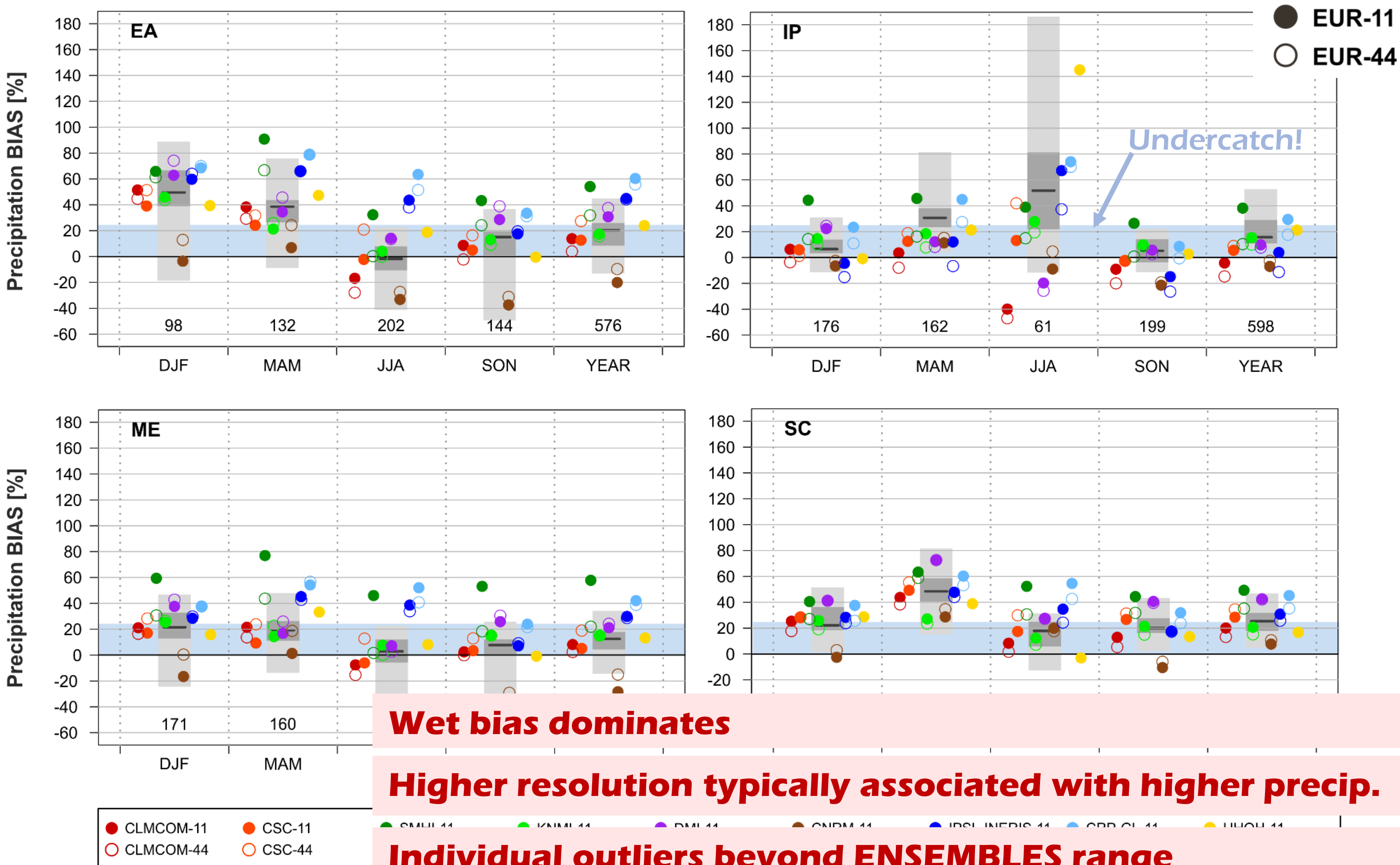
Relative ranking of models rather stable

No obvious benefit of increased resolution

Warm summer bias and cold winter bias reduced wrt. ENSEMBLES

PRECIPITATION: Regional and temporal mean bias [%]

1
2
3
4
5
6
7
Example
8



Wet bias dominates

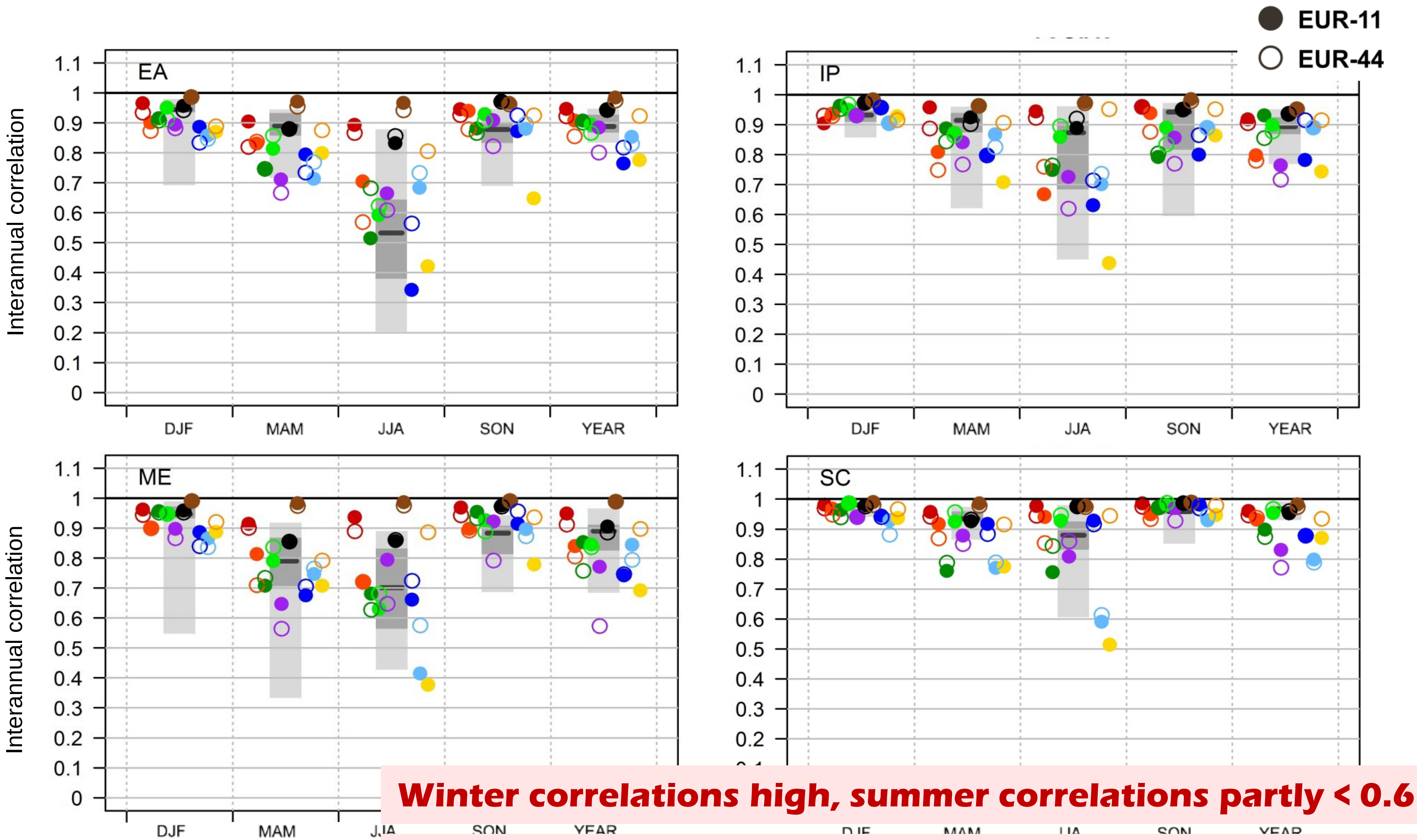
Higher resolution typically associated with higher precip.

Individual outliers beyond ENSEMBLES range

Biases often located in undercatch range

TEMPERATURE: Inter-annual correlation

1
2
3
4
5
6
7
Example
8



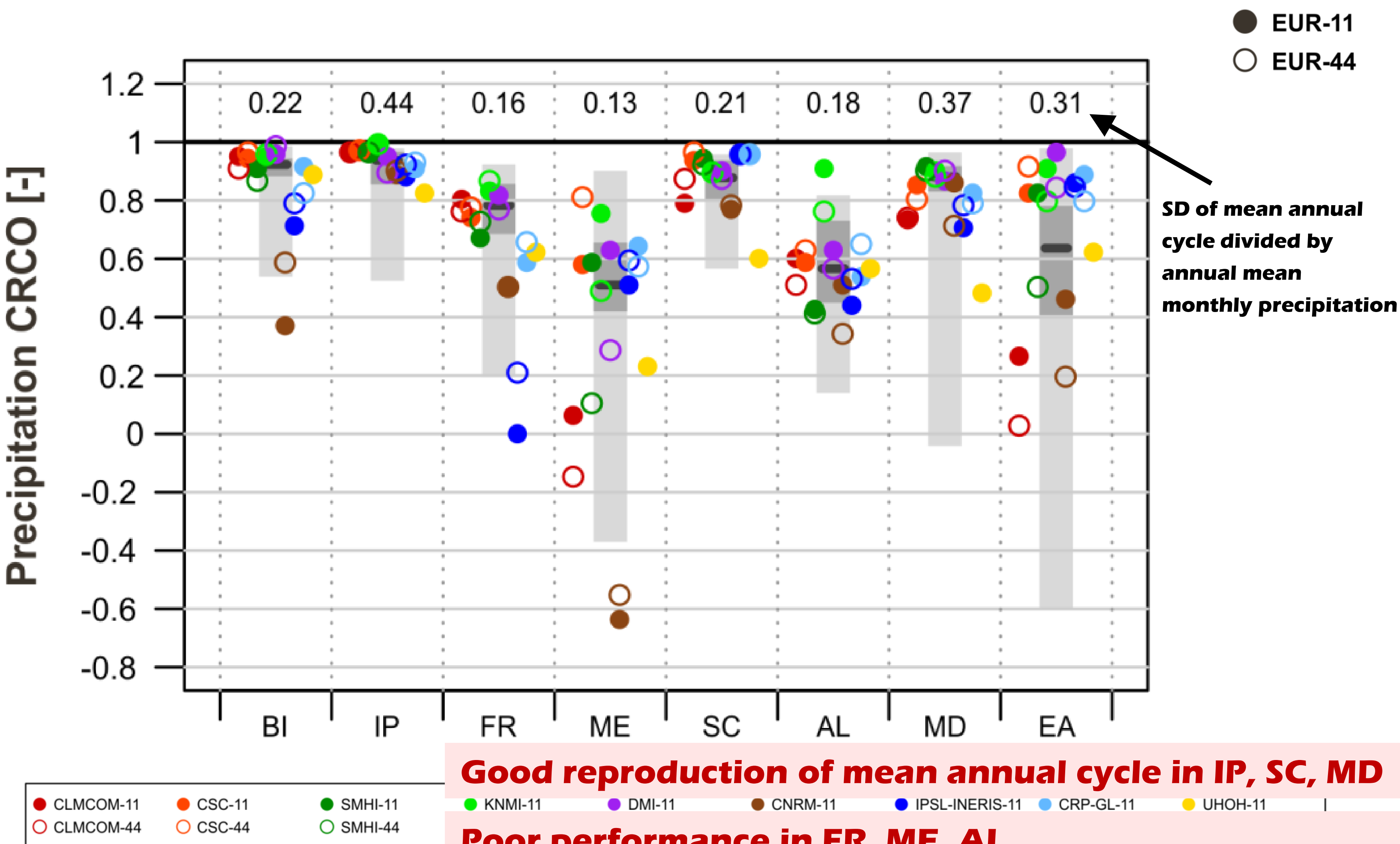
Winter correlations high, summer correlations partly < 0.6

Correlations close to 1 for CNRM

Higher resolution not beneficial

PRECIPITATION: Rank correlation of mean annual cycle

1
2
3
4
5
6
7
Example
8

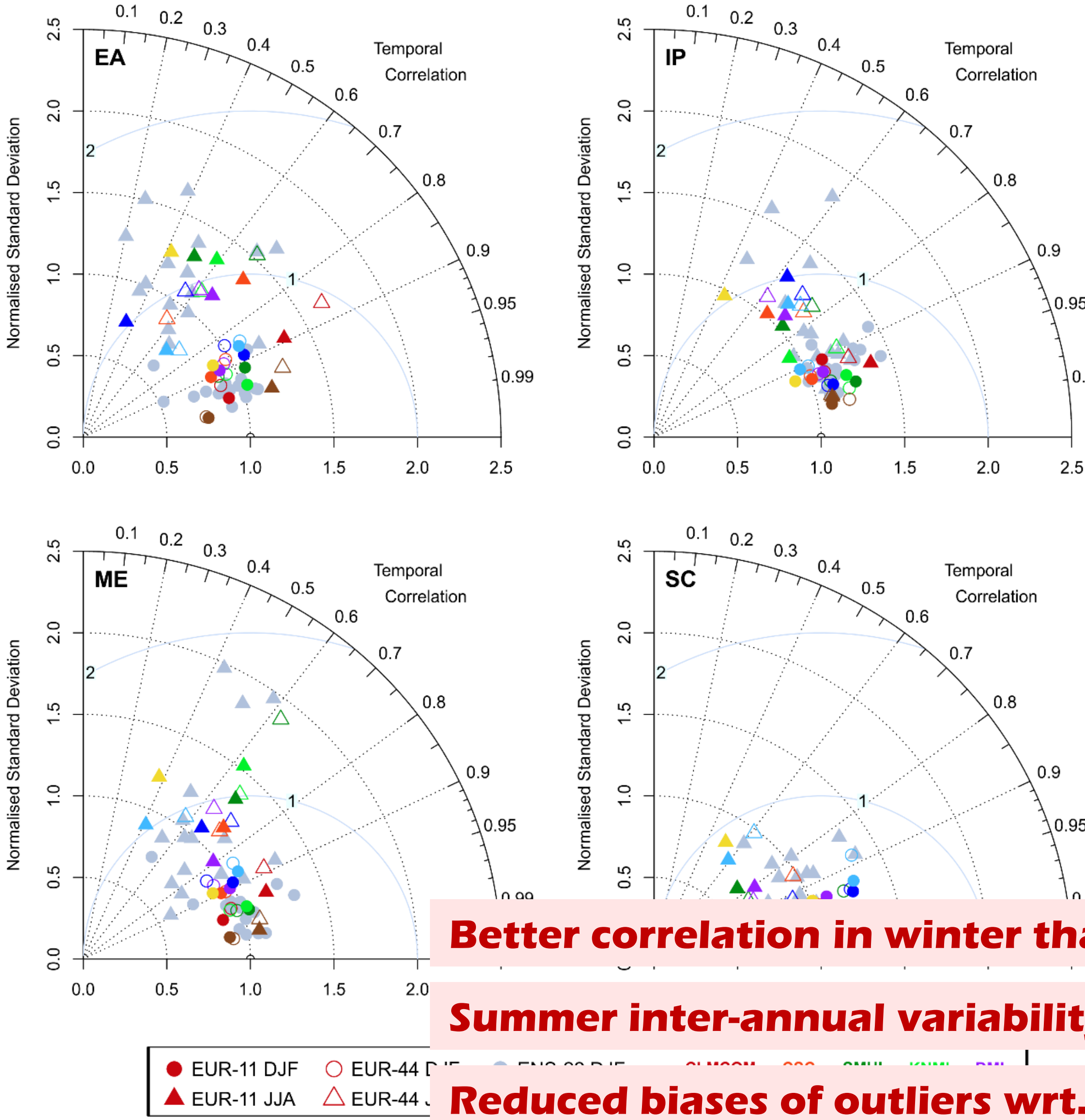


Good reproduction of mean annual cycle in IP, SC, MD

Poor performance in FR, ME, AL

But: latter regions show a weak annual cycle only

TEMPERATURE: TEMPORAL TAYLOR DIAGRAMS



Better correlation in winter than in summer

Summer inter-annual variability mostly overestimated

Reduced biases of outliers wrt. ENSEMBLES

SUMMARY: EURO-CORDEX STANDARD EVALUATION

1
2
3
4
5
6
7
Example
8

Evaluation…

… confirms the ability of current RCMs to represent to basic spatio-temporal features of the European climate under perfect boundary conditions

… highlights considerable shortcomings for selected metrics, seasons and regions and a considerable range of model biases for an identical forcing

… reveals common model biases found in majority of experiments (e.g., predominant wet and cold bias over most parts of Europe, warm and dry summer bias in Southern Europe)

… indicates progress wrt. ENSEMBLES for a few aspects (e.g. reduced warm summer bias over Southern Europe), but no general improvement

… does not reveal an obvious benefit of an increased spatial resolution at the temporal and spatial scales considered

→ Suite of accompanying studies to investigate potential added value of EUR-11 and to provide in-depth analysis of bias characteristics

OUTLINE

1 REGIONAL CLIMATE MODELLING (WRAP-UP)

2 MODEL EVALUATION: THE RATIONALE

3 APPROACHES

4 PERFORMANCE METRICS

5 TO CONSIDER!

6 MODEL WEIGHTING

7 EXAMPLE

8 SUMMARY & CONCLUSIONS

SUMMARY & CONCLUSIONS

1
2
3
4
5
6
7
8
Summary & Conclusions

- (Regional) Climate model evaluation as an important component of model development and application
- Important to provide trust into models and their scenarios

- Infinite number of evaluation schemes!
- Choice of scheme can strongly determine final outcome
- RCM evaluation ALWAYS has a subjective component
- Large number of issues to consider during evaluation exercise and interpretation of results

SUMMARY & CONCLUSIONS (cont'd)

1
2
3
4
5
6
7
8
Summary & Conclusions

NOTE!

- **Skill in the present does not imply skill in the future**
- But: A model has to **reflect the behaviour of the real system** in order to be suitable for scenario development (minimum requirement)

OUTLOOK: CLOUD-RESOLVING SCENARIOS

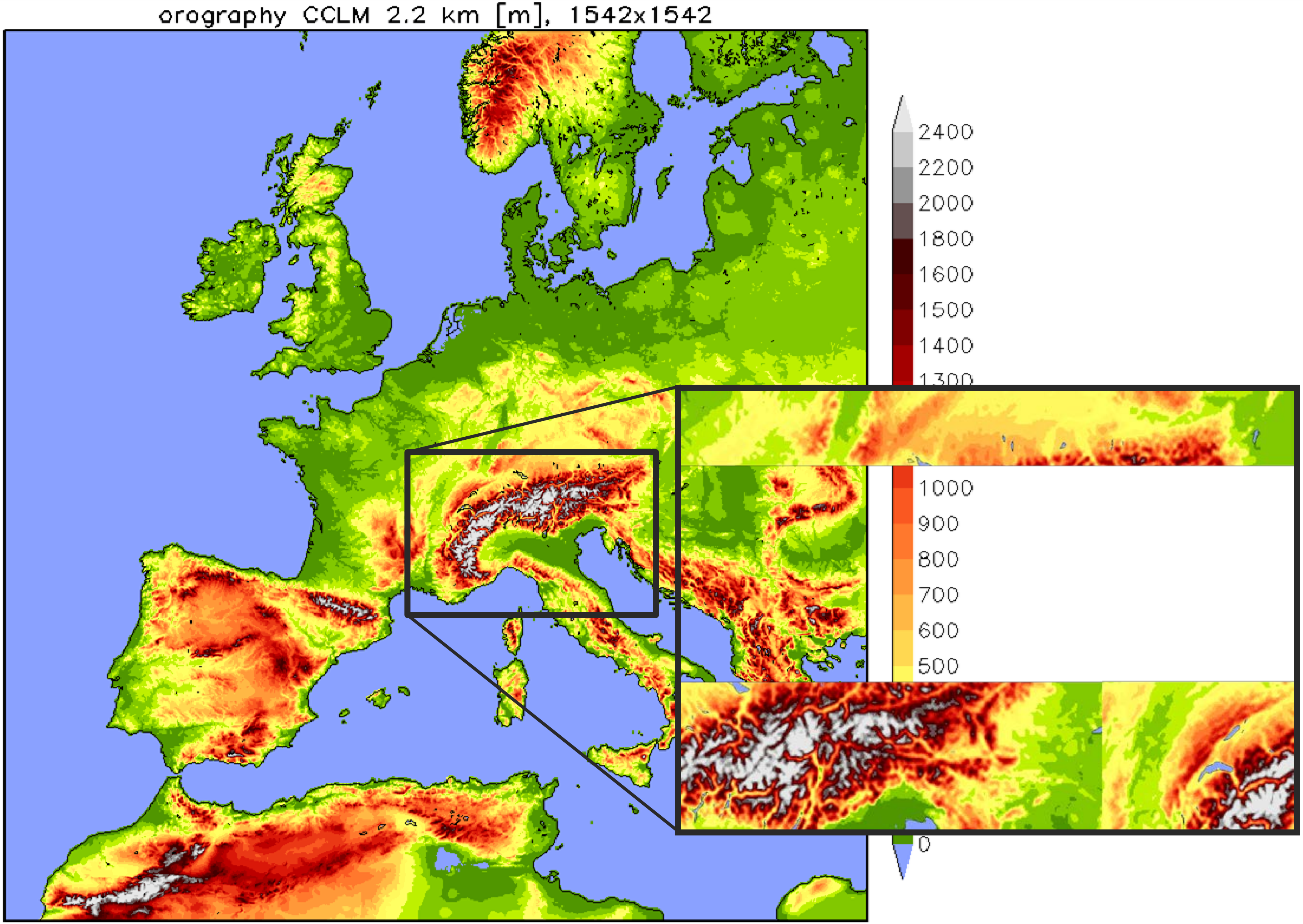
Spatial resolution of regional climate scenarios limited by available computing power (currently 10-50 km for larger ensembles).

Cloud-resolving scenarios at the kilometer-scale now becoming more and more feasible.

These scenarios can to some extent explicitly resolve moist convection and convection parameterizations can at least partly be switched off.

OUTLOOK: CLOUD-RESOLVING SCENARIOS (cont'd)

1
2
3
4
5
6
7
8
Summary & Conclusions

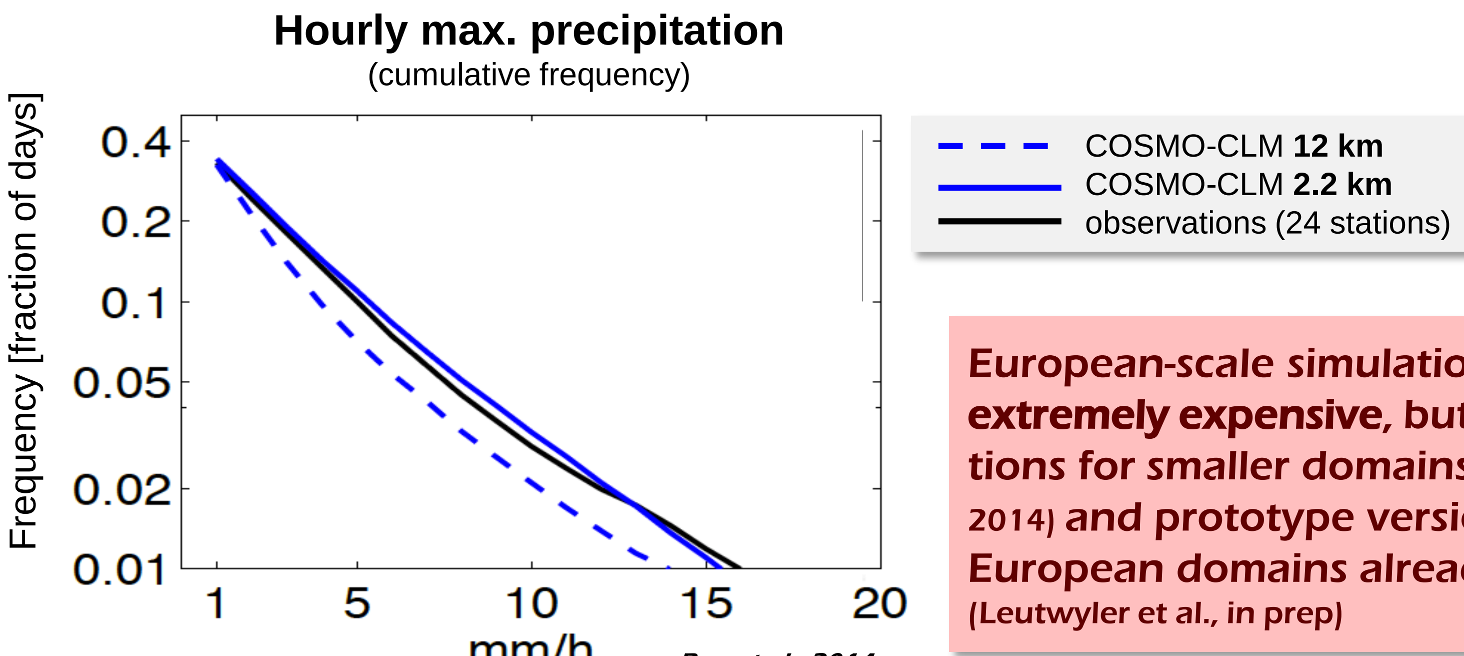


OUTLOOK: CLOUD-RESOLVING SCENARIOS (cont'd)

Added value of **cloud resolving simulations** to be expected in many cases:

- Diurnal cycle of summer convection (Hohenegger et al. 2008)
- Soil moisture – precipitation feedback (Froidevaux et al. 2014)
- Spatial precipitation patterns, precipitation extremes (Prein et al. 2014, Kendon et al. 2014, Ban et al. 2014)

Evaluation (!): Sub-daily precipitation statistics Switzerland (1998-2007)



Ban et al.. 2014

European-scale simulations extremely expensive, but simula-tions for smaller domains (e.g., Ban et al. 2014) and prototype versions for European domains already available (Leutwyler et al., in prep)

FURTHER LITERATURE…

1
2
3
4
5
6
7
8
Summary & Conclusions

Regional climate modelling (in general)

- McGregor, Meteorol. Atmos. Phys., 1997
- Laprise, J. Computat. Phys., 2008
- Giorgi, J. Phys. IV France, 2006

Evaluation

- IPCC AR5, WG1, Chapter 09 («Climate model evaluation»)
- Gleckler et al., J. Geophys. Res., 2008
- Christensen et al., Clim. Change, 2007
- Christensen et al., Clim Res., 2010
- Kotlarski et al., Geosci. Model Dev., 2014
- Boberg and Christensen, Nature Clim. Change, 2012

Email: *sven.kotlarski@env.ethz.ch*

THANK YOU FOR LISTENING!