CONSTRAINING TEMPERATURE VARIABILITY PROJECTIONS USING SMILES THAT BEST REPRESENT OBSERVATIONS

Nicola Maher[1,2], **Laura Suarez-Gutierrez**[3,2], and Sebastian Milinski[4,2]
1) CIRES/CU Boulder, CO, USA, 2) Max Planck Institute for Meteorology, Hamburg, Germany, 3) ETH Zurich, Zurich, Switzerland, 4) ECMWF, Bonn, Germany

Want to talk?
laura.suarez@env.ethz.ch

Photography allowed

Constraining variability changes under warming by using best-performing models shows larger variability increases, and more inter-model agreement in the strength of the changes than the full ensemble.

However, constraining only marginally affects the multi-model mean or the inter-model agreement on the sign of future changes.

1. Constraining variability changes

Why: Improving our understanding of how variability changes under warming is **crucial to determine changes in extremes and adaptability ranges** as the world warms.

What: We generate the **best available estimates** of these how variability changes by identifying the **Single Model Initial-Condition Large Ensembles** (SMILEs) that most accurately represent this variability in observations. By using SMILEs we ensure a **robust sampling of the internal variability as well as a more reliable model evaluation.**

How: With the framework in Suarez-Gutierrez et al., 2021, we evaluate **11 CMIP5 and CMIP6 models** over 9 ocean and 24 land regions for **both December, January, February** (DJF) and **June, July, August** (JJA), against ERSST and GISTEMP observations, then use best-performing models for variability projections.

2. Simple Evaluation Framework [1]

1. **Observations are distributed evenly across the ensemble** → cluster within the central 75th percentile less than 85% of the months
2. **Observations are mostly within the ensemble spread** → fall outside the ensemble limits during less than 10% of the months

This framework highlights **biases in** how models simulate the **internal variability and forced response** in observations without separating both signals.

Blue OR Red → model over- or underestimates forced response

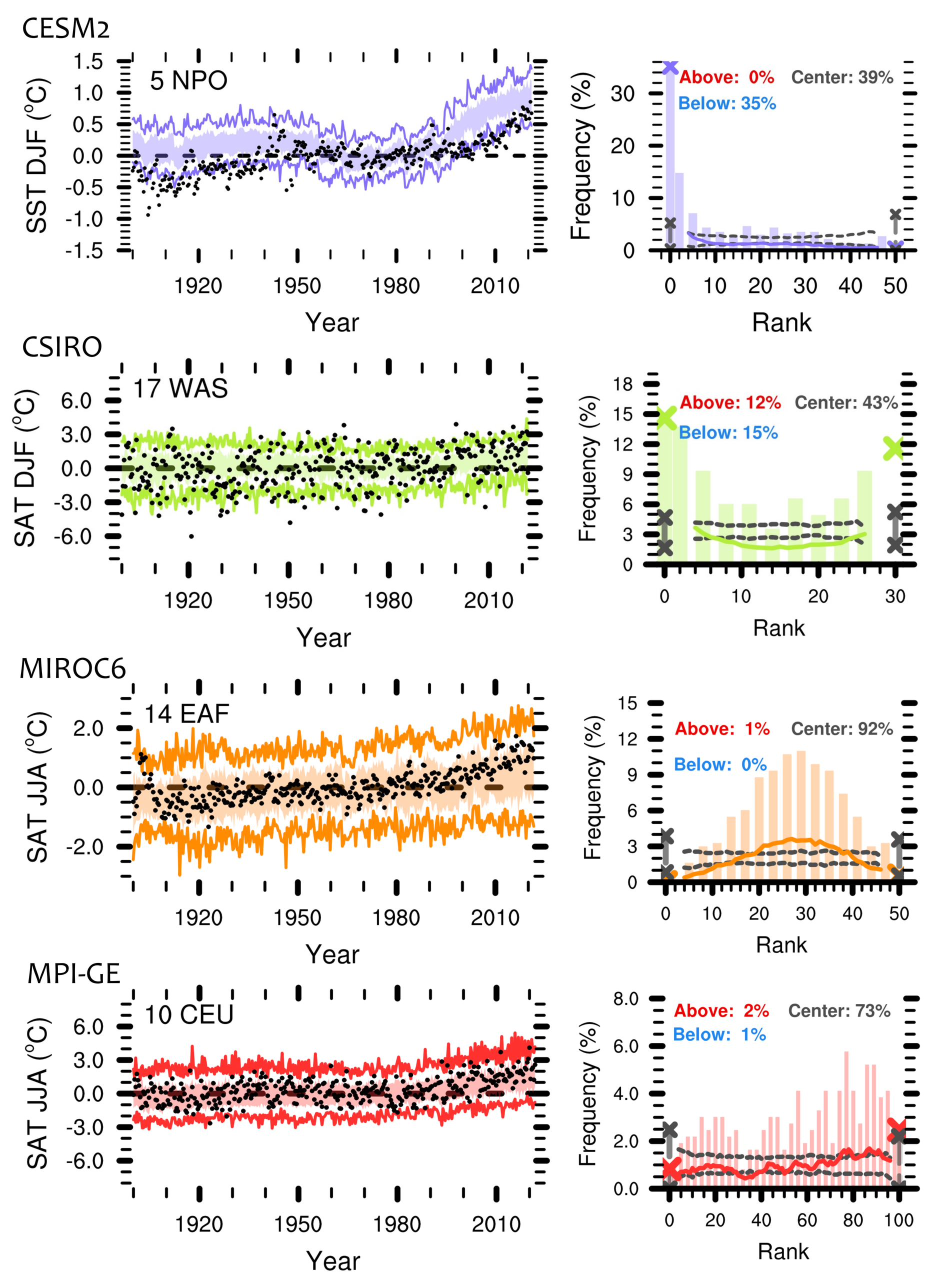
Blue AND Red → model underestimates **internal variability** AND/OR over/underestimates **forced** response

Gray → model overestimates **internal variability**

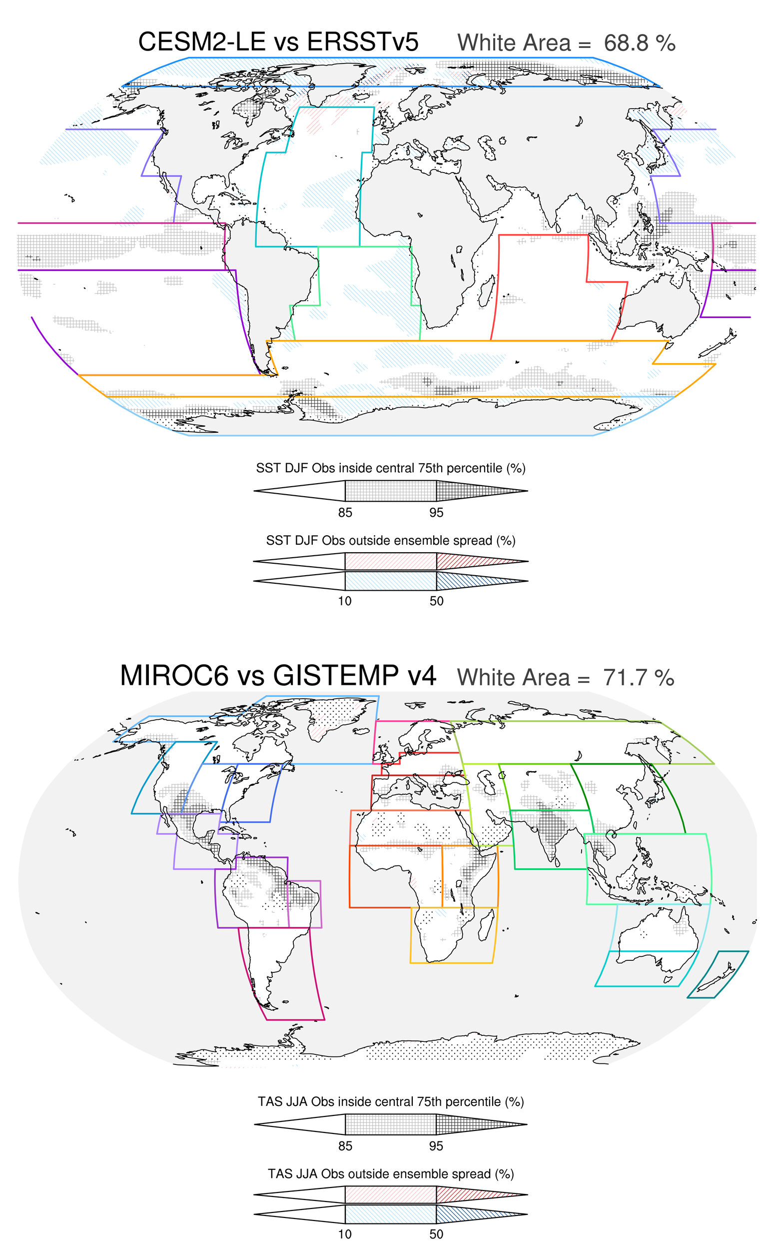
White → model shows **no significant biases**

Best-performing models = 1/3 or more grid cells without biases per region

[1] Suarez-Gutierrez, Milinski, & Maher (2021). Exploiting large ensembles for a better yet simpler climate model evaluation. Clim. Dyn.

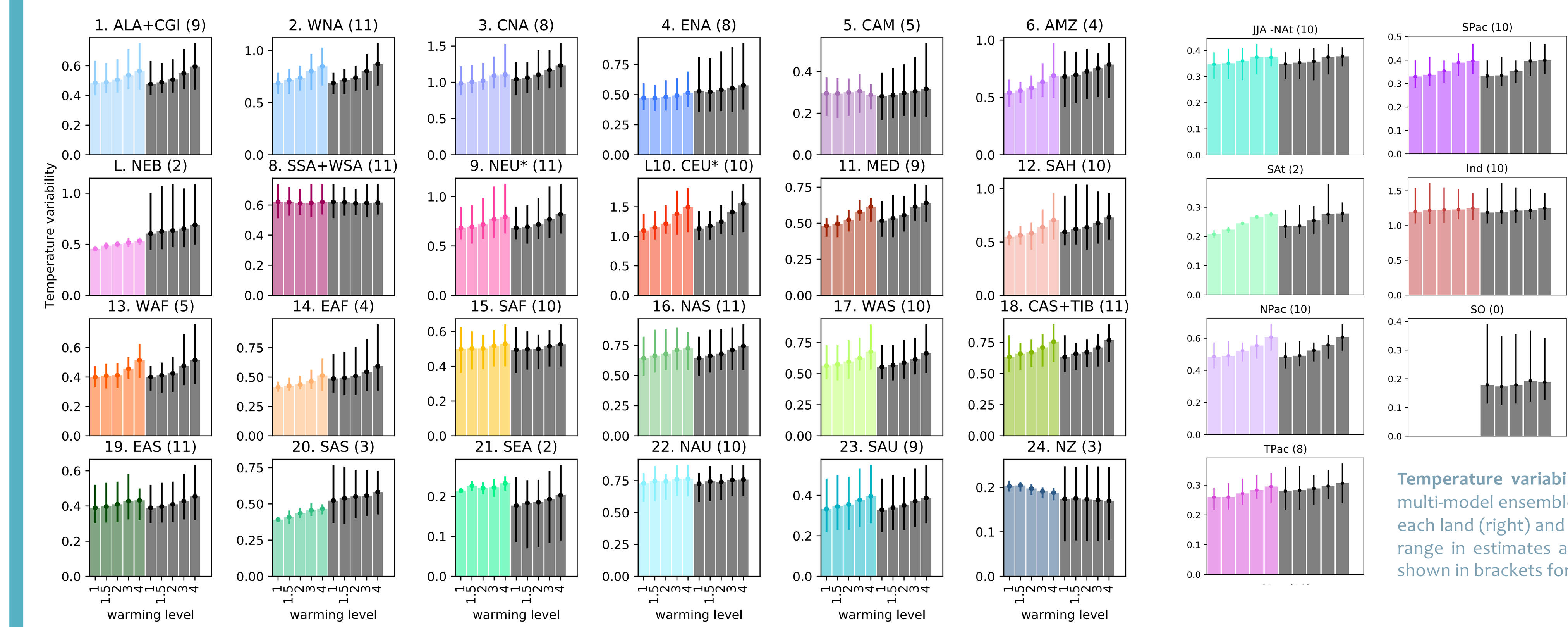


Time series and rank histograms for selected SMILEs and regions. Ensemble maxima and minima (lines) and 75th central ensemble spread (shading; 12.5-87.5th percentiles) against observations (black dots). All anomalies are monthly means relative to the period of 1961-1990. Rank histograms represents the place that observations take in a list of ascending ensemble members (color bars, line 6-rank mean). Crosses mark the frequency of minimum (0) and maximum (number of members) ranks. Perfect Model Rank Range (gray, 5-95th percentiles) represents the range in places each ensemble member takes in a list of the remaining ensemble members, as if it were the observations. If a model adequately represents observations its rank histogram should be roughly flat and within its perfect model range.



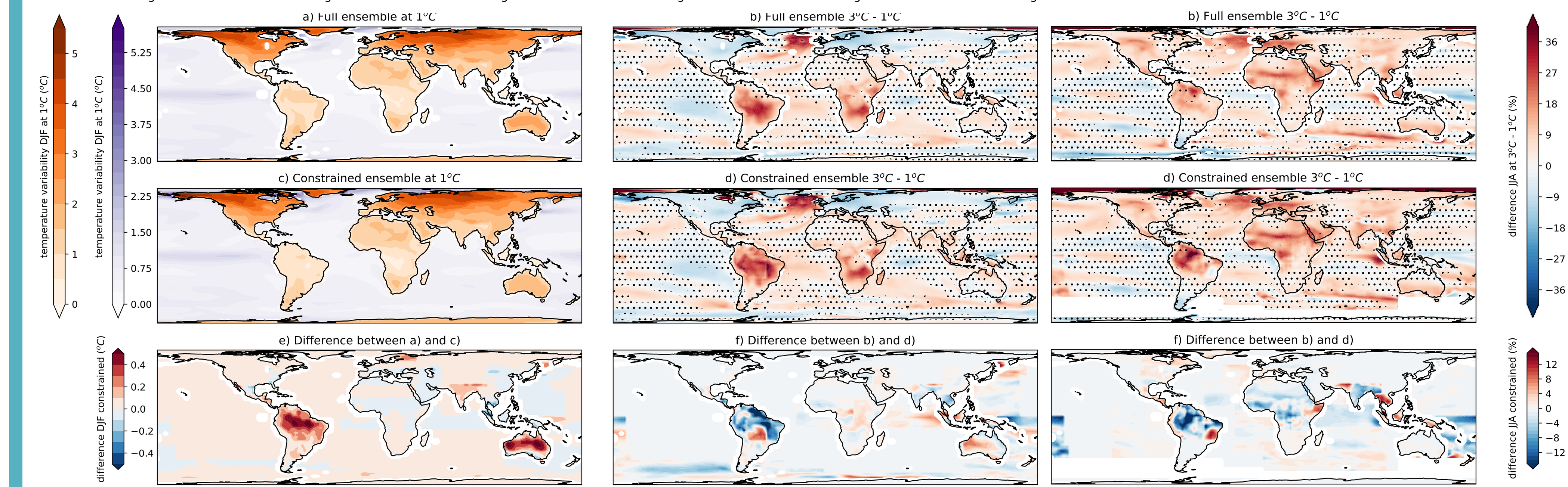
Spatial evaluation of variability and forced response in monthly mean temperature anomalies for ocean (top) or land (bottom) regions showing where observations are larger than the ensemble maximum (red) or smaller than the ensemble minimum (blue) for more than 10% to 50% of the months, and regions where observations cluster within the central 75th percentile ensemble spread (gray) more than 85% or 95% of the months. Dotted areas represent regions where observations are available for less than 30 months. Simulated data are regridded to match the observational grid.

3. Variability Changes under Warming: Full versus Constrained Estimates



Temperature variability under different warming levels per region for JJA in the full multi-model ensemble (black) and in the constraint best-performing ensemble (colors), for each land (right) and ocean (left) regions. Wide bars show ensemble mean, lines show full range in estimates across different models. The number of best-performing models is shown in brackets for each region.

- **Performance constraining leads to increased model agreement** in variability estimates over several regions, especially over land, e.g., Mediterranean (L11) and Sahara (L12) regions, East Africa (L14), or Southern Asia (L20, L21)
- However **performance constraining only minimally changes multi-model mean estimates** over several regions, e.g. West Africa (L13) or Southern Asia (L20)



Temperature Variability Estimates for Full vs. Constraint Ensembles. a) Full Ensemble temperature variability at 1C for DJF. b) Full Ensemble temperature variability change between 3C and 1C for DJF (left) and JJA (right). c) Constraint Ensemble temperature variability at 1C. d) Constraint Ensemble temperature variability change between 3C and 1C of warming 1C for DJF (left) and JJA (right). e) Effect of performance constraint in 1C temperature variability. f) Effect of performance constraint in temperature variability change between 3C and 1C of warming 1C for DJF (left) and JJA (right). Stippling shows regions where at least 80% of the models agree on the sign of variability change. All for seasonal mean temperature.

- Largest effects of **performance constraining are over tropical land areas,** especially Amazon and Australia, and over **South Pacific and Southern Oceans**
- **Constrained Estimates show larger variability increases** than the full ensemble over large regions in both seasons
- However, over large land areas **models in the constrained ensemble still disagree on the sign** of variability changes

4. Multi-Model Evaluation: Best-performing SMILEs

- The SMILEs that best capture observations over the larger number of regions are **MPI-GE, CESM2-LE, and GFDL-SPEAR-MED**
- **Several regions are not well represented by most models,** e.g., sea-ice adjacent oceans (Arctic O1, Antarctic O2, and JJA Southern Ocean O9), or complex and narrow orography (Central America L7, South East Asia L 21)

	ACCESS		CanESM2		CanESM5		CESM-LE		CESM2-LE		CSIRO		GFDL-ESM2M		GFDL-SPEAR-MED		MIROC6		MPI-GE		MPI-GE 6		TOTAL		
	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	
O1	29,7	22,1	23,0	33,3	34,1	34,7	4,0	27,9	22,1	37,2	0,5	12,6	20,4	39,4	27,3	35,4	13,1	33,4	8,1	23,7	9,3	23,4	0	0	O1
O2	18,4	7,9	15,3	0,2	17,0	5,8	10,7	0,0	27,8	11,1	40,6	0,2	28,7	10,6	40,1	11,4	30,7	4,1	36,0	5,6	43,6	6,0	0	0	O2
O3	42,1	55,3	44,6	82,1	60,3	77,4	61,3	78,1	77,6	84,2	67,3	78,5	58,0	79,7	77,0	89,4	80,2	88,9	72,2	82,3	85,9	88,2	6	10	O3
O4	50,2	52,4	49,8	60,9	51,5	60,0	40,1	48,0	66,0	43,7	48,7	31,9	44,2	62,0	91,5	76,0	67,2	53,1	78,6	79,9	56,7	55,2	3	2	O4
O5	84,9	82,4	75,2	80,0	77,7	79,1	89,3	90,9	71,9	87,1	75,1	65,4	81,6	94,1	82,4	85,5	91,9	89,2	80,1	85,4	92,8	94,7	11	10	O5
O6	83,9	89,0	70,5	76,6	68,2	86,8	80,1	87,4	44,4	52,0	59,6	81,1	80,8	65,0	91,8	89,5	91,1	48,0	77,3	68,7	83,0	80,3	9	8	O6
O7	91,7	81,6	85,1	81,4	80,6	73,7	94,3	85,6	85,0	78,0	87,6	62,3	89,2	76,8	91,3	83,7	90,3	82,9	80,5	90,3	86,7	87,6	11	10	O7
O8	96,1	74,0	77,6	79,0	94,6	92,9	74,6	84,1	92,5	91,5	85,5	56,1	82,6	77,3	93,8	84,7	90,2	85,2	67,4	86,3	84,2	90,1	11	10	O8
O9	60,2	26,1	69,9	45,2	49,1	35,1	69,2	47,0	66,9	32,7	55,8	20,4	65,6	25,9	66,4	33,9	56,9	29,8	75,1	54,0	66,7	41,6	6	0	O9
L1	86,2	55,8	75,0	73,2	86,2	69,0	85,0	83,0	90,0	67,6	83,8	59,6	81,4	68,4	91,7	69,6	88,7	80,4	91,2	80,9	85,4	73,4	11	9	L1
L2	79,9	87,8	88,2	86,6	92,5	84,1	78,9	89,0	92,8	87,7	74,4	89,2	94,3	71,5	93,6	85,8	84,2	67,2	88,0	81,4	92,2	90,8	11	11	L2
L3	90,8	41,0	87,8	84,3	95,9	85,9	92,5	82,0	100,0	94,9	90,9	85,4	100,0	59,7	97,9	68,6	95,1	66,8	86,9	66,8	96,9	65,0	11	8	L3
L4	81,8	66,5	61,8	65,1	63,2	85,7	73,4	62,6	86,3	73,6	58,6	54,3	88,8	74,8	84,0	66,9	75,5	79,7	84,9	86,3	78,3	66,6	8	8	L4
L5	54,9	50,2	48,0	66,6	59,2	83,1	51,2	41,2	71,0	74,3	62,1	67,2	58,0	41,9	64,7	60,0	59,8	54,2	72,9	71,3	60,6	53,9	2	5	L5
L6	62,0	58,4	48,9	55,9	58,5	70,3	68,1	74,8	74,1	87,1	37,9	49,9	43,1	58,6	58,0	74,9	48,4	53,1	69,7	59,9	62,5	62,8	3	4	L6
L7	45,1	64,3	32,8	60,3	64,3	61,7	50,6	36,9	70,0	64,3	38,0	54,6	33,1	46,6	69,8	60,2	47,7	47,4	72,4	80,6	46,3	68,3	3	2	L7
L8	65,9	76,2	76,5	76,7	62,7	74,4	74,4	85,4	79,5	80,1	61,0	68,7	57,9	84,4	71,8	81,3	38,3	71,8	73,2	84,2	70,2	82,8	6	11	L8
L9	68,1	83,0	64,4	83,5	73,7	95,3	63,6	92,8	81,9	87,1	72,3	83,5	92,1	86,0	97,9	78,9	75,2	93,4	88,6	100,0	82,3	90,2	9	11	L9
L10	86,7	71,5	86,0	25,1	92,3	100,0	81,7	100,0	93,1	100,0	98,9	67,8	98,0	96,8	98,9	96,8	95,9	91,5	97,9	89,2	94,9	100,0	11	10	L10
L11	75,1	69,5	62,3	63,2	61,4	93,0	92,6	80,6	90,7	82,0	67,4	69,3	95,3	62,5	91,3	83,9	87,0	72,5	95,3	68,9	87,3	86,8	9	9	L11
L12	83,0	68,0	76,5	85,4	74,0	81,5	86,5	80,8	87,7	74,3	82,1	81,1	90,2	56,4	83,4	75,8	92,8	78,5	99,0	79,7	92,3	84,4	11	10	L12
L13	74,6	82,5	73,1	67,7	85,4	60,9	66,6	64,4	82,4	82,9	59,9	44,8	81,4	63,4	75,6	75,5	83,9	67,5	79,9	60,1	63,2	58,3	9	5	L13
L14	68,3	67,7	68,2	65,2	76,6	63,5	78,4	66,5	79,6	79,7	62,9	46,2	73,5	61,1	62,2	68,9	68,8	54,6	76,5	40,7	67,7	48,5	9	4	L14
L15	70,9	81,1	79,0	73,8	85,8	74,6	66,2	81,2	74,3	93,7	77,8	65,2	75,4	70,1	81,7	79,1	56,3	74,3	75,0	75,9	78,5	79,2	10	10	L15
L16	74,2	92,6	93,9	88,4	92,6	96,7	85,8	94,5	93,9	89,8	80,5	87,7	99,3	80,5	89,7	88,6	88,8	96,1	98,9	79,7	96,5	93,5	11	11	L16
L17	89,2	76,3	89,4	79,3	83,8	77,3	89,3	80,9	97,0	80,1	50,6	59,9	94,4	72,1	81,6	80,6	94,9	81,8	94,8	66,6	96,4	79,2	10	10	L17
L18	90,9	82,6	87,1	81,9	97,4	79,4	82,6	90,4	92,9	95,5	69,1	90,8	88,7	82,1	88,1	94,6	73,3	68,2	89,3	78,6	92,9	91,6	11	11	L18
L19	61,9	68,1	57,0	75,4	79,7	66,2	95,5	80,9	92,4	80,0	73,0	78,1	94,5	79,3	85,5	66,9	75,7	81,8	94,5	88,6	93,3	82,3	9	11	L19
L20	75,6	54,7	41,6	41,0	59,6	70,5	62,7	67,2	82,5	60,6	71,3	45,0	59,4	48,2	70,4	68,5	60,0	46,4	61,7	60,2	82,0	63,9	5	3	L20
L21	45,1	50,8	46,2	46,9	39,7	21,9	42,1	42,4	62,9	59,2	73,0	51,1	62,5	65,0	54,3	34,9	73,7	68,4	81,8	70,8	57,1	39,5	3	2	L21
L22	70,1	76,7	26,2	65,0	54,7	68,2	61,1	80,0	80,6	93,0	74,9	71,2	35,0	81,3	49,7	83,2	38,5	78,9	40,1	87,8	50,5	79,8	3	10	L22
L23	63,9	86,5	55,5	68,3	63,4	81,3	54,9	91,2	77,1	97,1	53,5	54,8	52,5	59,9	68,7	76,1	44,5	69,9	59,2	99,0	86,7	92,9	3	9	L23
L24	41,4	34,2	72,6	20,7	14,1	7,1	41,3	44,1	86,3	54,4	46,0	3,6	8,5	16,1	55,1	32,7	42,7	67,8	78,4	100,0	63,1	82,0	3	3	L24
TOTAL	20	20	18	19	17	24	21	23	28	24	18	14	19	16	25	26	21	23	28	26	23	22			

Multi-model evaluation for monthly DJF and JJA temperature anomalies. Best-performing models are those with 66,6% or more of non-biased area per region and season, highlighted in green. Regions with less than 3 best-performing models are highlighted in red. Bottom row shows total number of regions with good performance per model

CIRES
University of Colorado Boulder
MAX-PLANCK-INSTITUT FÜR METEOROLOGIE
ETH zürich
ECMWF
Co-funded by the European Union