Large potential of performance-based model weighting to improve decadal climate forecast skill

Verjans, V.[1], Donat, M.[1,2], Delgado-Torres, C.[1], and DelSole, T.[3]

1: Barcelona Supercomputing Center, Barcelona, Spain
2: Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
3: George Mason University, Fairfax, VA, USA

CL4.10
Climate predictions from seasonal to multi-decadal timescales and their applications

Verjans et al. (2026), *npj Climate and Atmospheric Sciences*

How can we validate added skill from model weighting?

How can we validate added skill from model weighting?

Geophysical Research Letters

RESEARCH LETTER
10.1002/2016GL072012

A climate model projection weighting scheme accounting for performance and interdependence

Reto Knutti[1,2], Jan Sedláček[1], Benjamin M. Sanderson[2], Ruth Lorenz[1], Erich M. Fischer[1], and Veronika Eyring[3]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland, [2]National Center for Atmospheric Research, Boulder, Colorado, USA, [3]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Key Points:
- Model weighting can constrain future projections
- Ensemble projections must also account for model interdependence
- Finding appropriate metrics to weight models remains challenging

communications earth & environment

ARTICLE
https://doi.org/10.1038/s43247-023-01009-8 OPEN

Check for updates

Bayesian weighting of climate models based on climate sensitivity

Elias C. Massoud[1✉], Hugo K. Lee[2], Adam Terando[3,4] & Michael Wehner[5]

Geosci. Model Dev., 16, 4715–4747, 2023
https://doi.org/10.5194/gmd-16-4715-2023
© Author(s) 2023. This work is distributed under the Creative Commons Attribution 4.0 License.

Geoscientific Model Development
Open Access
EGU

Climate model Selection by Independence, Performance, and Spread (ClimSIPS v1.0.1) for regional applications

Anna L. Merrifield[1], Lukas Brunner[2], Ruth Lorenz[1], Vincent Humphrey[1], and Reto Knutti[1]

Earth Syst. Dynam., 14, 457–483, 2023
https://doi.org/10.5194/esd-14-457-2023
© Author(s) 2023. This work is distributed under the Creative Commons Attribution 4.0 License.

Earth System Dynamics
Open Access
EGU

Performance-based sub-selection of CMIP6 models for impact assessments in Europe

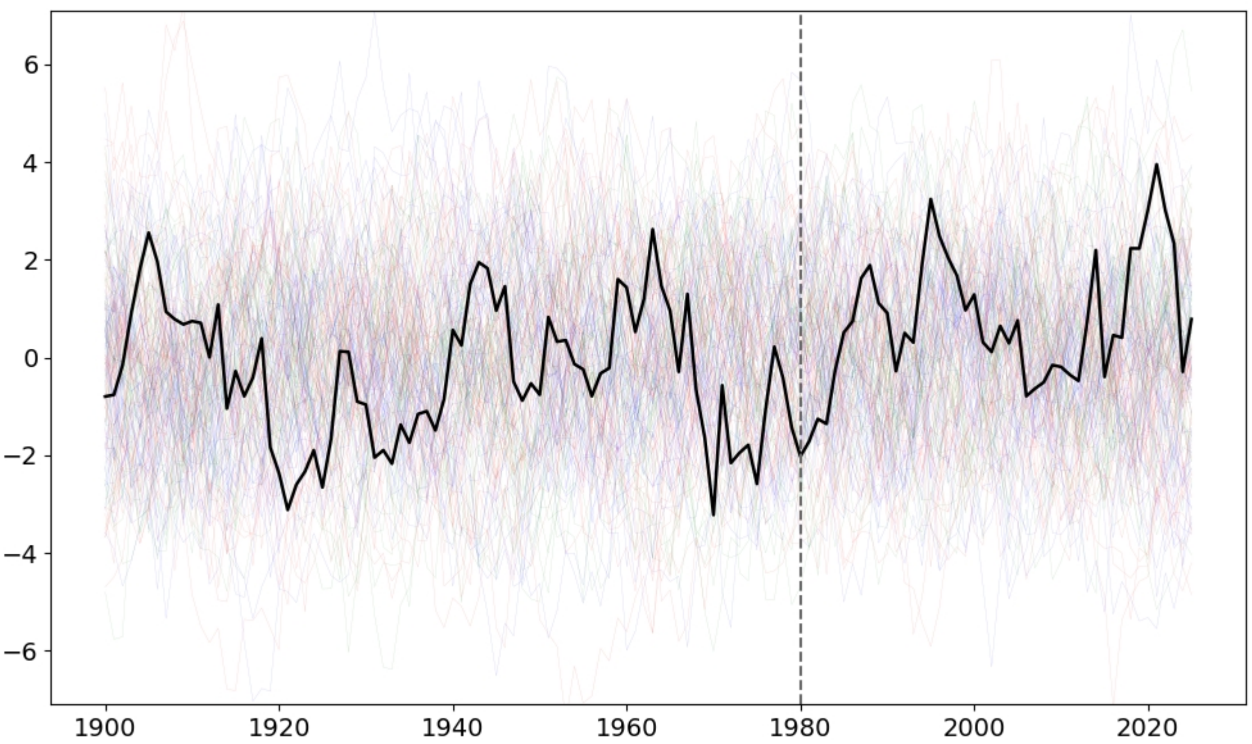
Tamzin E. Palmer[1], Carol F. McSweeney[1], Ben B. B. Booth[1], Matthew D. K. Priestley[2], Paolo Davini[3], Lukas Brunner[4], Leonard Borchert[5,6], and Matthew B. Menary[1,6]

- Validation assumptions
  - Models matching past observations ⇒ better future skill (?)
  - Models with better pseudo-observation predictive skill ⇒ better real-world skill (?)

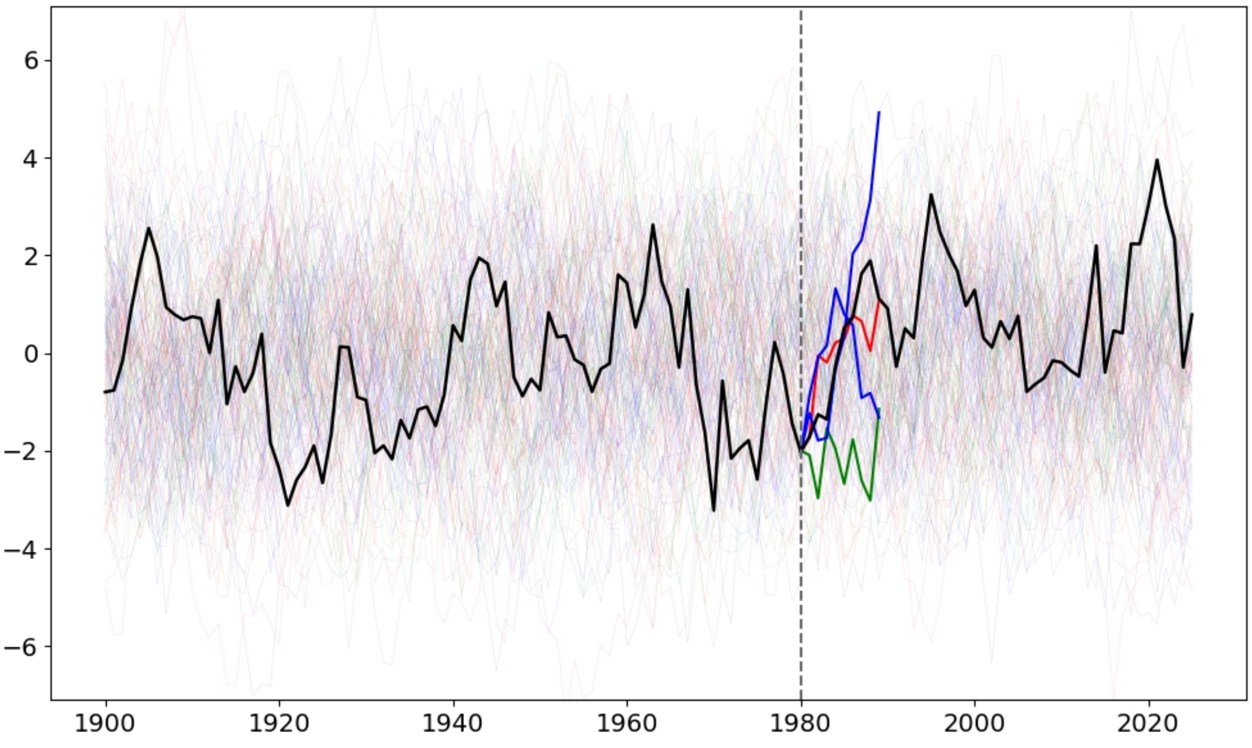
- Decadal predictions: we can make hindcasts!

- Analogue-based initialisation (ABI)
  - select analogues from large ensembles by minimising distance to observations
  - track this sub-ensemble for the next decade
  - repeat for any start date

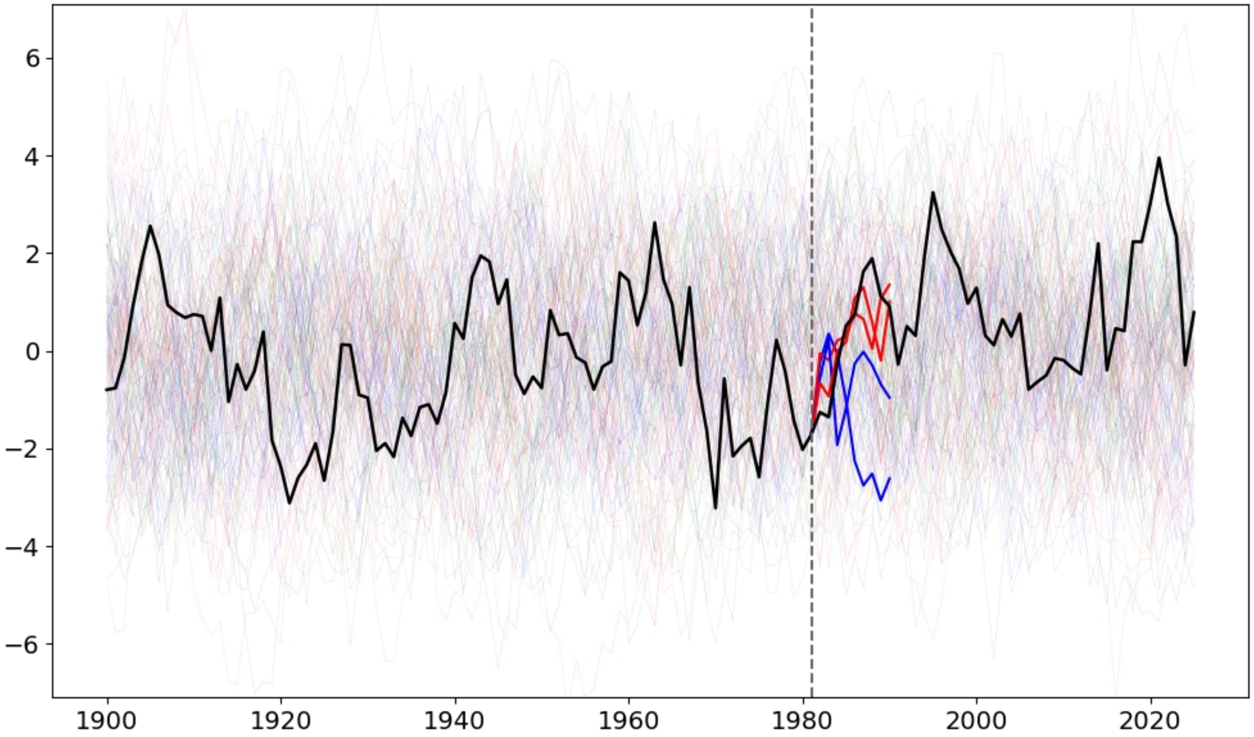
- Analogue-based initialisation (ABI)
  - select analogues from large ensembles by minimising distance to observations
  - track this sub-ensemble for the next decade
  - repeat for any start date



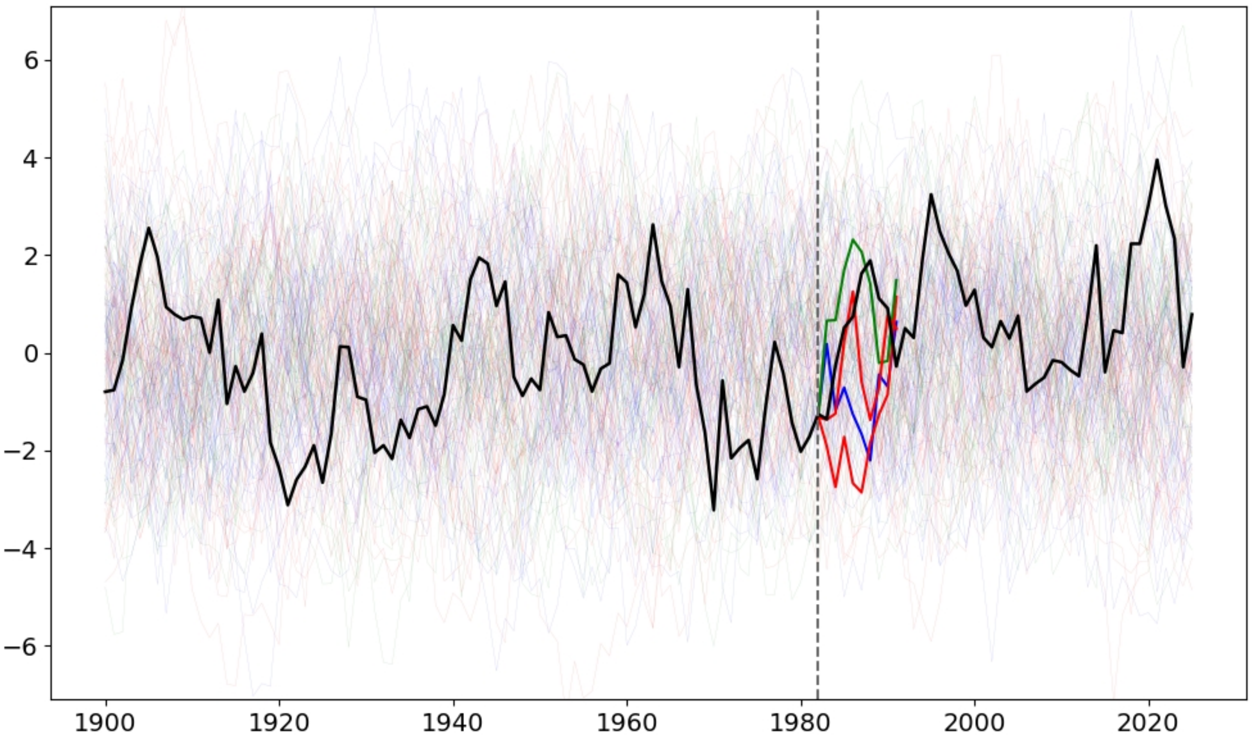
- Analogue-based initialisation (ABI)
  - select analogues from large ensembles by minimising distance to observations
  - track this sub-ensemble for the next decade
  - repeat for any start date



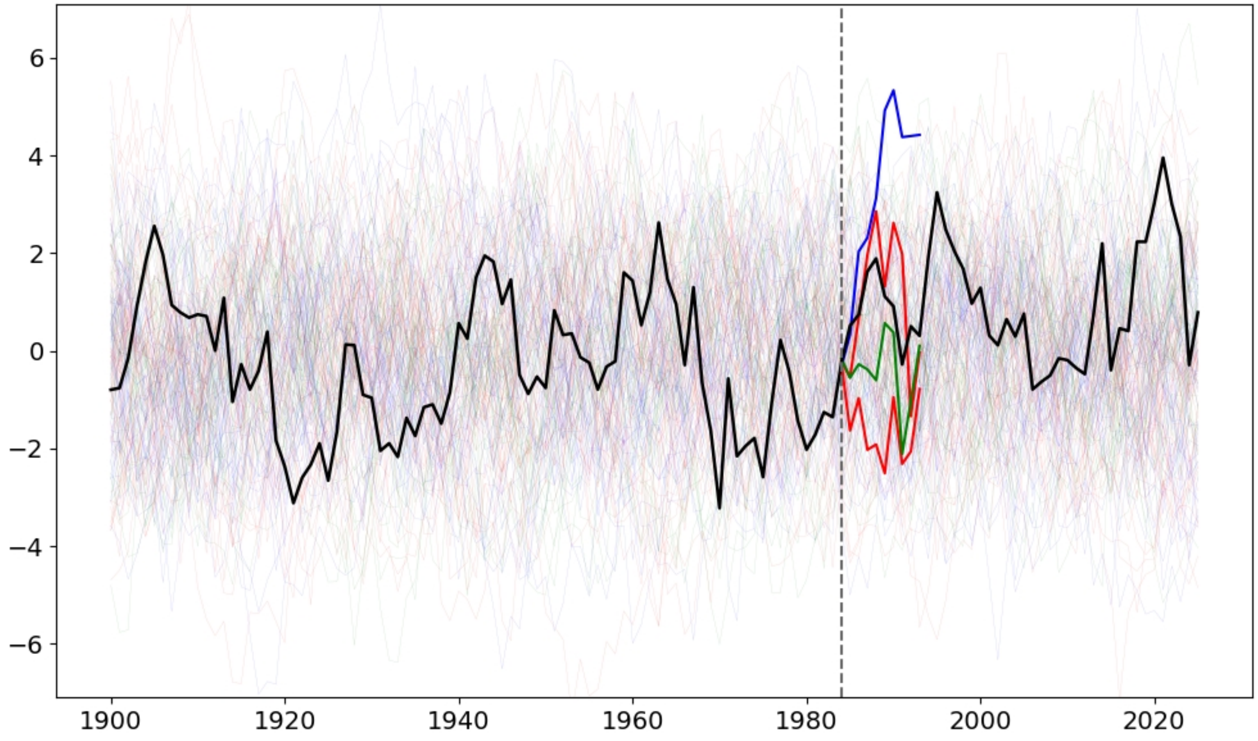
- Analogue-based initialisation (ABI)
  - select analogues from large ensembles by minimising distance to observations
  - track this sub-ensemble for the next decade
  - repeat for any start date



- Analogue-based initialisation (ABI)
  - select analogues from large ensembles by minimising distance to observations
  - track this sub-ensemble for the next decade
  - repeat for any start date

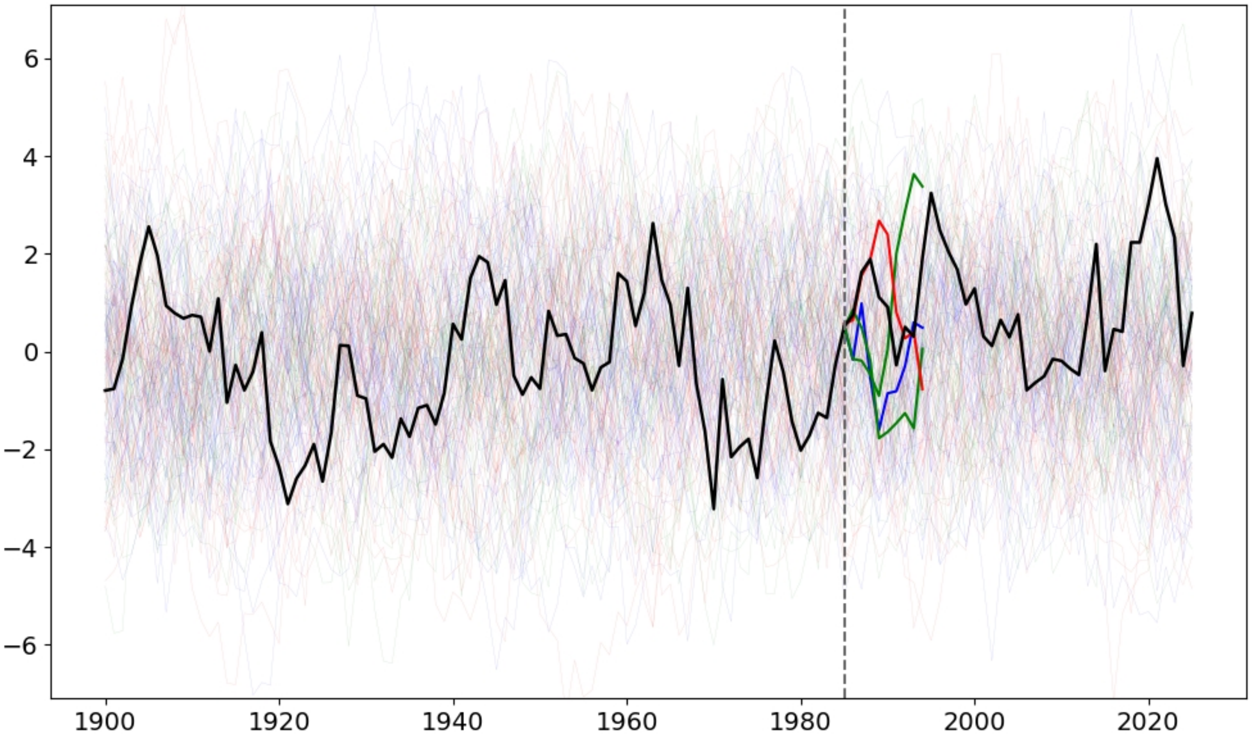


- Analogue-based initialisation (ABI)
  - select analogues from large ensembles by minimising distance to observations
  - track this sub-ensemble for the next decade
  - repeat for any start date

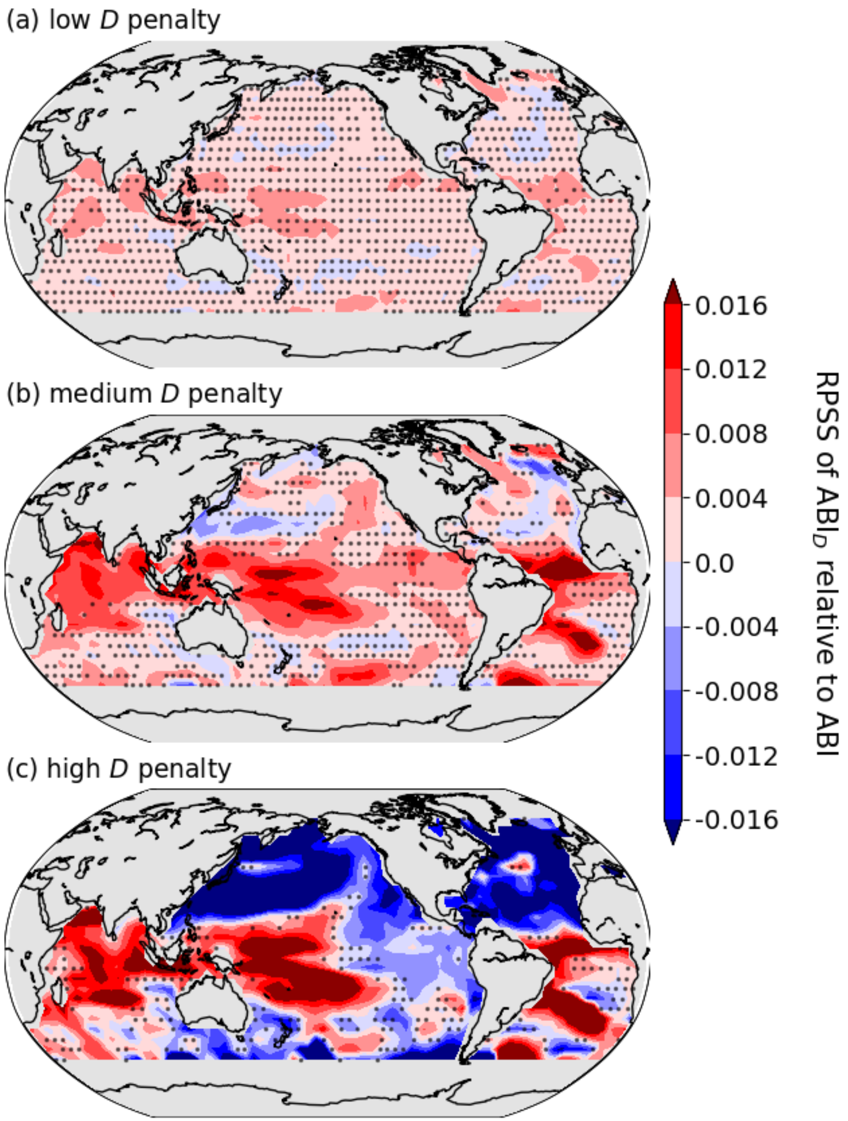


- Analogue-based initialisation (ABI)
  - select analogues from large ensembles by minimising distance to observations
  - track this sub-ensemble for the next decade
  - repeat for any start date
- Combine (1) initial error minimisation
  (2) performance criteria

  ⇒ sub-ensemble selection



- Focus on decadal (1-10 year) SST predictions.
- A **novel performance-weighting metric**: the deviance (D)
  $\Rightarrow$ ABI vs. ABI_D
- Trade-off: initial error vs. performance
- We calibrate the deviance penalties in pseudo-observation experiments across 316 CMIP6 ensemble members
- Avoids **overfitting** observations + Ensures **robustness** of calibration
- Larger skill differences in zones with **more decadal predictability**:
  - Tropics (strongly positive)
  - High-latitudes
- Low skill differences in zones with **little decadal predictability**:
  - Mid-latitudes



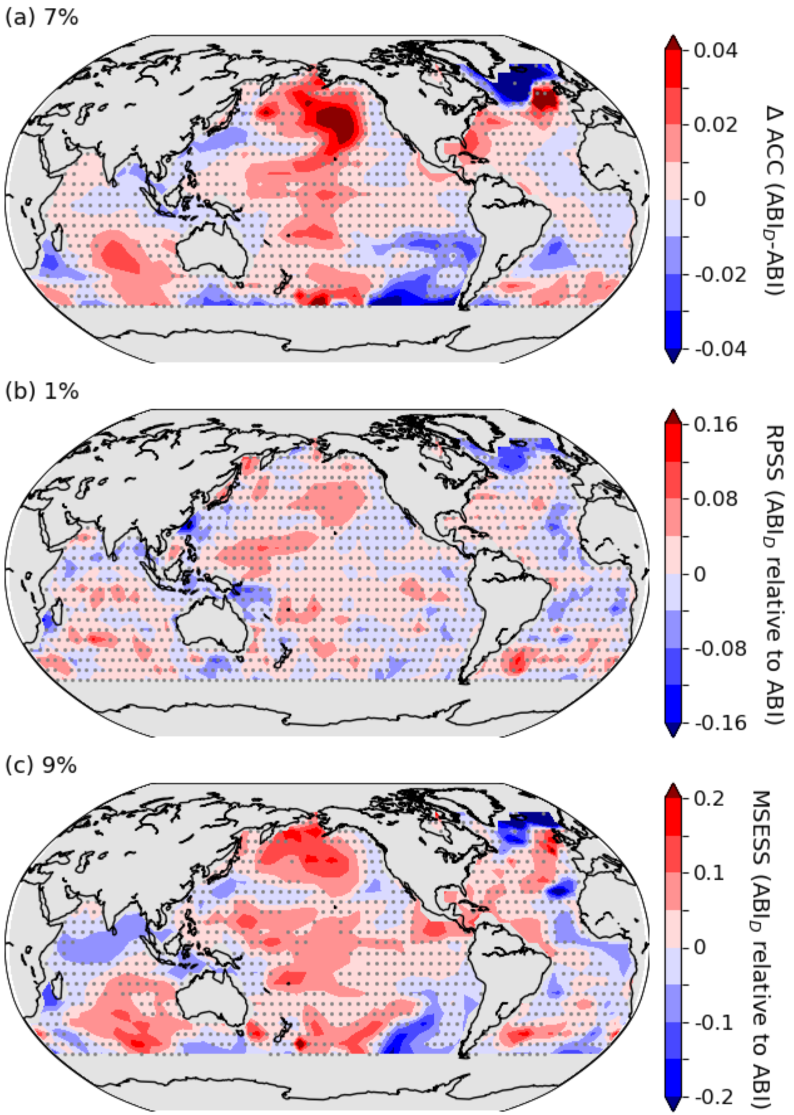
- Hindcasting observational record 1930-2024 (86 different hindcasts start dates)

- Using model-calibrated performance weights

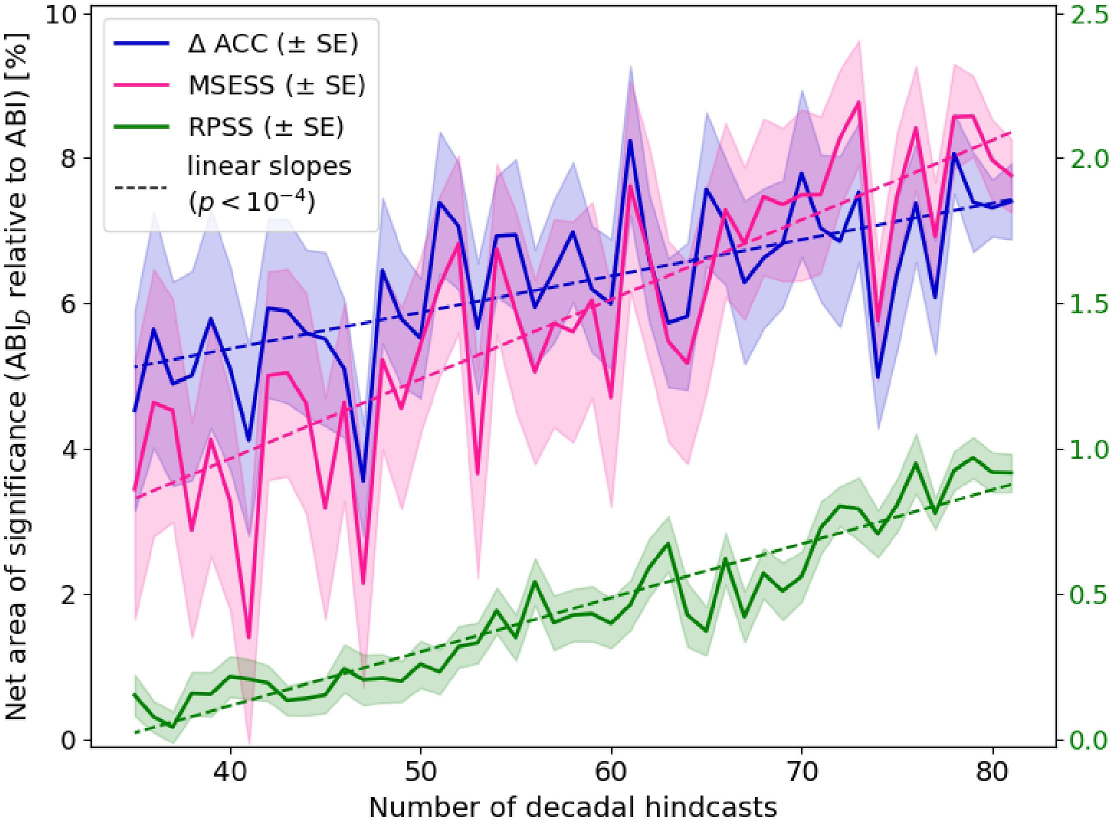
- On average, slightly improved prediction skill from model weighting
  ⇒ **contrasts with pseudo-observations**

- Here also, skill differences concentrate in tropics and at high-latitudes

- Performance-weighted and unweighted ensembles share, on average, **23.7/30 analogues** per hindcast



- Sub-sampling the 86 possible hindcasts
- **Does skill gain from model weighting depend on the number of hindcast verifications?**
  - Tendency of increasing skill gain
  - High variability (one can be unlucky)



Lessons drawn regarding model weighting

(A) Pseudo-observation skill does not necessarily translate into increased real-world skill.

(B) Robustly identifying skill gains from model weighting is challenging.
- (1) Limits on intrinsic climate predictability
- (2) Inherent similarity of unweighted and weighted ensembles
- (3) Skill evaluation metrics are prone to sampling uncertainties

<u>Questions and comments are welcome</u>

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Questions/Collaboration interests:
vincent.verjans@bsc.es

Publication link
Verjans et al. (2026),
npj Climate and Atm. Sc.

EXPECT
Funded by the European Union

Extra slides

$$\begin{cases} \boldsymbol{y}_o(t) = \boldsymbol{c}_o + \sum_{i=1}^{p} \mathbf{B}_{o,i}^{\mathrm{T}} \boldsymbol{y}_o(t-i) + \boldsymbol{a}_o z(t) + \boldsymbol{\epsilon}_o(t) \\ \boldsymbol{y}_m(t) = \boldsymbol{c}_m + \sum_{i=1}^{p} \mathbf{B}_{m,i}^{\mathrm{T}} \boldsymbol{y}_m(t-i) + \boldsymbol{a}_m z(t) + \boldsymbol{\epsilon}_m(t) \end{cases}$$

$$\begin{array}{l} \Omega_0 : \boldsymbol{\Sigma}_o, \boldsymbol{\Sigma}_m \text{ unrestricted, } \mathbf{B}_o, \mathbf{B}_m \text{ unrestricted and } \boldsymbol{a}_o, \boldsymbol{a}_m \text{ unrestricted} \\ \Omega_1 : \boldsymbol{\Sigma}_o = \boldsymbol{\Sigma}_m,\ \mathbf{B}_o, \mathbf{B}_m \text{ unrestricted and } \boldsymbol{a}_o, \boldsymbol{a}_m \text{ unrestricted} \\ \Omega_2 : \boldsymbol{\Sigma}_o = \boldsymbol{\Sigma}_m,\ \mathbf{B}_o = \mathbf{B}_m \text{ and } \boldsymbol{a}_o, \boldsymbol{a}_m \text{ unrestricted} \\ \Omega_3 : \boldsymbol{\Sigma}_o = \boldsymbol{\Sigma}_m,\ \mathbf{B}_o = \mathbf{B}_m,\ \boldsymbol{a}_o = \boldsymbol{a}_m \end{array},$$

$$\begin{cases} \tilde{D}_{0:1} = (\nu_o + \nu_m) \log |\hat{\boldsymbol{\Sigma}}_{\Omega_1}| - \nu_o \log |\hat{\boldsymbol{\Sigma}}_o| - \nu_m \log |\hat{\boldsymbol{\Sigma}}_m| \\ \tilde{D}_{1:2} = (\nu_o + \nu_m) \left(\log |\hat{\boldsymbol{\Sigma}}_{\Omega_2}| - \log |\hat{\boldsymbol{\Sigma}}_{\Omega_1}| \right) \\ \tilde{D}_{2:3} = (\nu_o + \nu_m) \left(\log |\hat{\boldsymbol{\Sigma}}_{\Omega_3}| - \log |\hat{\boldsymbol{\Sigma}}_{\Omega_2}| \right) \end{cases}$$

$$\begin{array}{c} (\nu_o + \nu_m) \log |\hat{\boldsymbol{\Sigma}}_{\Omega_3}| > (\nu_o + \nu_m) \log |\hat{\boldsymbol{\Sigma}}_{\Omega_2}| > \\ (\nu_o + \nu_m) \log |\hat{\boldsymbol{\Sigma}}_{\Omega_1}| > \nu_o \log |\hat{\boldsymbol{\Sigma}}_o| + \nu_m \log |\hat{\boldsymbol{\Sigma}}_m| \end{array},$$

$$s_k(t) = \frac{1}{\overline{\mathrm{RMSE}}(t)_k} - \gamma_0 D_{0:1,k} - \gamma_1 D_{1:2,k} - \gamma_2 D_{2:3,k},$$

$$\mathrm{MIC}(\boldsymbol{y}, M, p) = T \log\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\mathbf{C}}_{\boldsymbol{y}}|}\right) + \kappa,$$

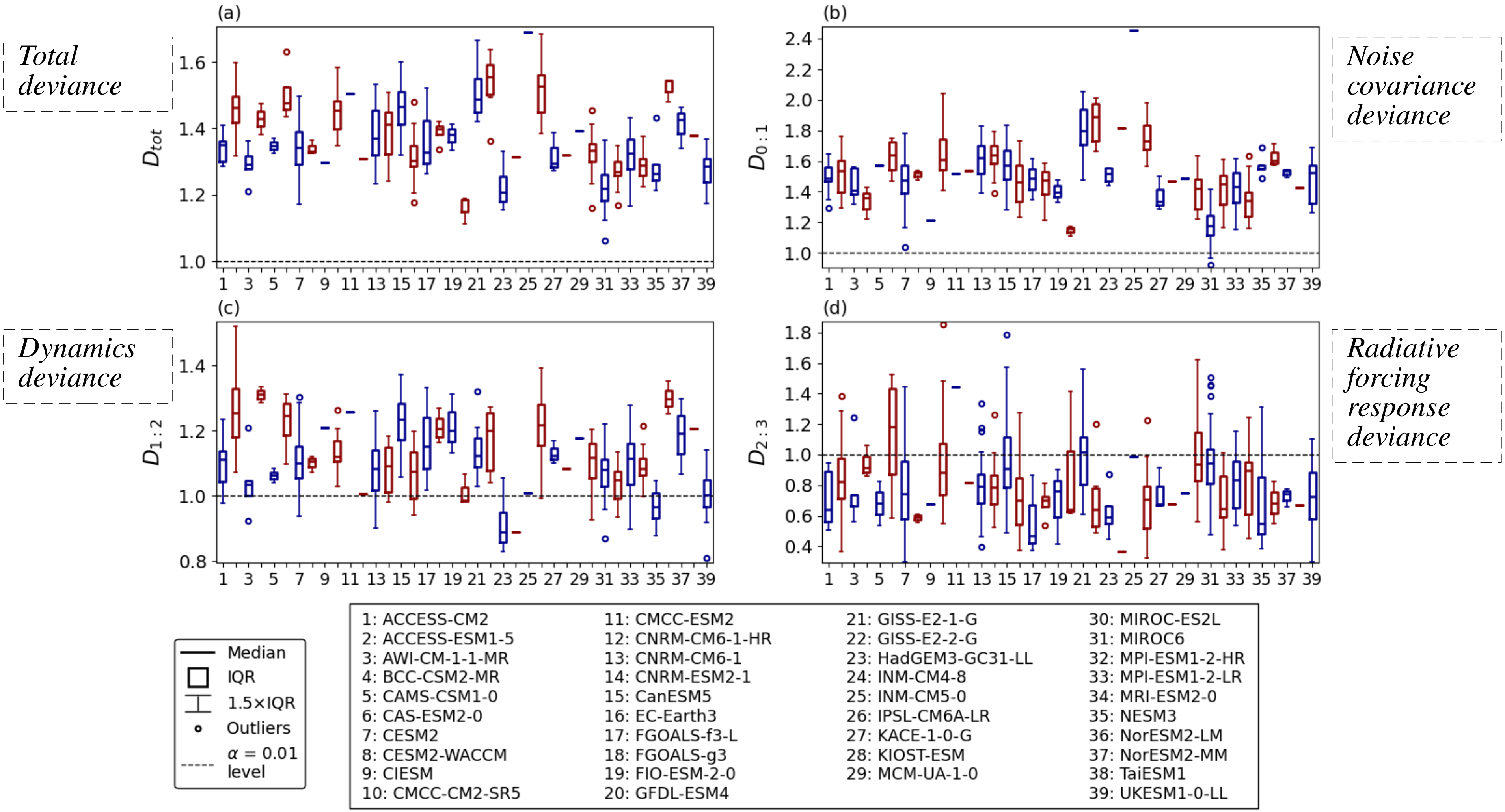
$$\kappa = T(T+1)\left(\frac{M(p+1)}{T - M(p+1) - 2} - \frac{M}{T - M - 2} - \frac{pM}{T - pM - 2}\right).$$

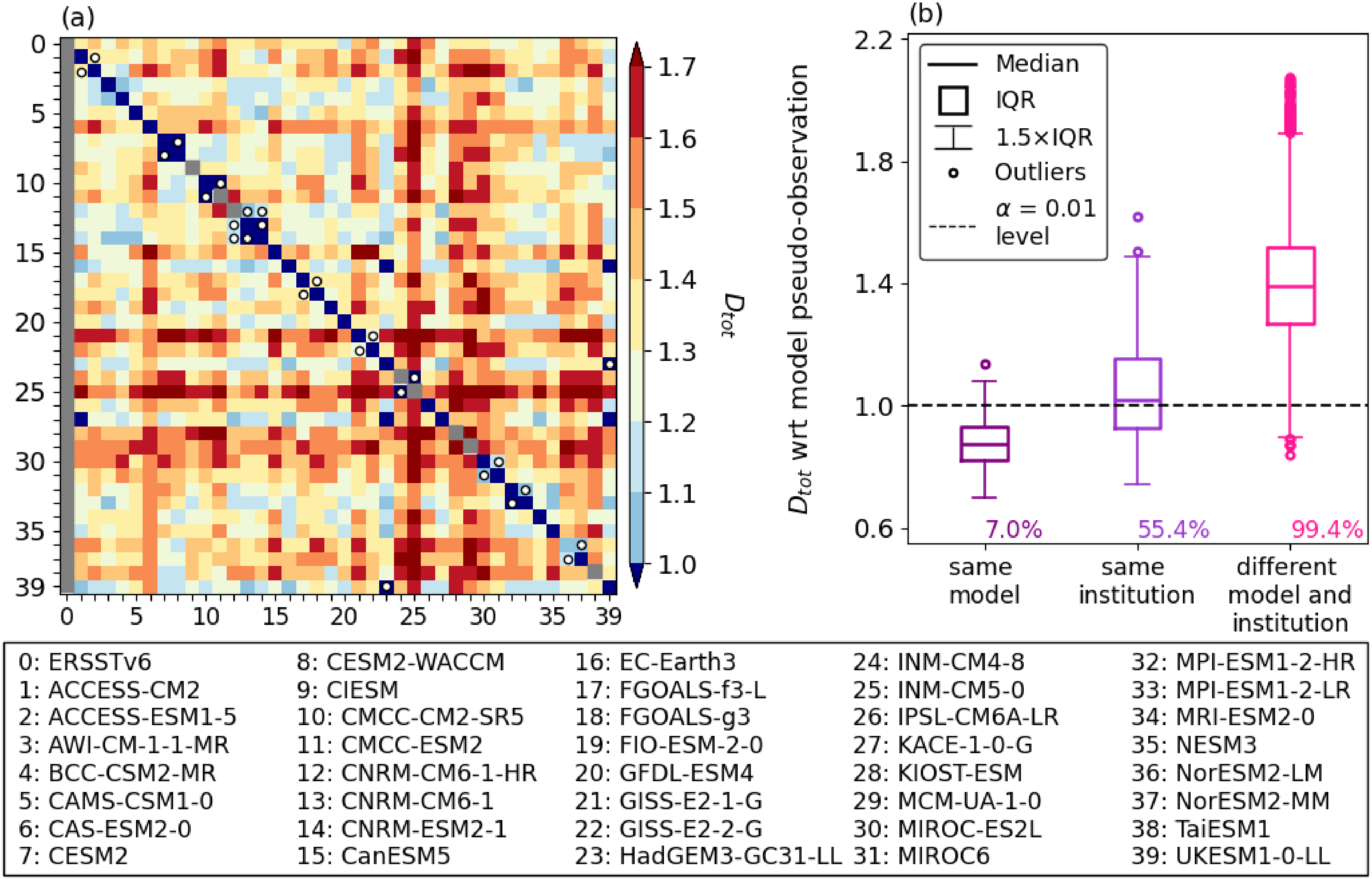
Table 1 Climate models used and ensemble size. For each climate model, the number of uninitialized ensemble members used in deviance computations (N(D)) and for hindcasting (N(hindcast)) are provided. Deviance computation uses historical (1850-2014) simulations, which are extended by the SSP2-4.5 scenario [15] for hindcasting, where available (1850-2024).

Model name	N(D)	N(hindcast)	Model name	N(D)	N(hindcast)
ACCESS-CM2	10	5	GISS-E2-1-G	10	10
ACCESS-ESM1-5	40	40	GISS-E2-2-G	6	5
AWI-CM-1-1-MR	5	1	HadGEM3-GC31-LL	4	1
BCC-CSM2-MR	3	1	INM-CM4-8	1	1
CAMS-CSM1-0	2	2	INM-CM5-0	1	1
CAS-ESM2-0	4	2	IPSL-CM6A-LR	32	11
CESM2	100	4	KACE-1-0-G	3	3
CESM2-WACCM	3	3	KIOST-ESM	1	1
CIESM	1	1	MCM-UA-1-0	1	1
CMCC-CM2-SR5	10	1	MIROC-ES2L	30	30
CMCC-ESM2	1	1	MIROC6	50	50
CNRM-CM6-1-HR	1	1	MPI-ESM1-2-HR	10	1
CNRM-CM6-1	30	6	MPI-ESM1-2-LR	30	30
CNRM-ESM2-1	11	10	MRI-ESM2-0	10	5
CanESM5	50	50	NESM3	5	2
EC-Earth3	22	16	NorESM2-LM	3	3
FGOALS-f3-L	3	1	NorESM2-MM	3	2
FGOALS-g3	4	4	TaiESM1	1	1
FIO-ESM-2-0	3	3	UKESM1-0-LL	16	5
GFDL-ESM4	3	1	Total	523	316

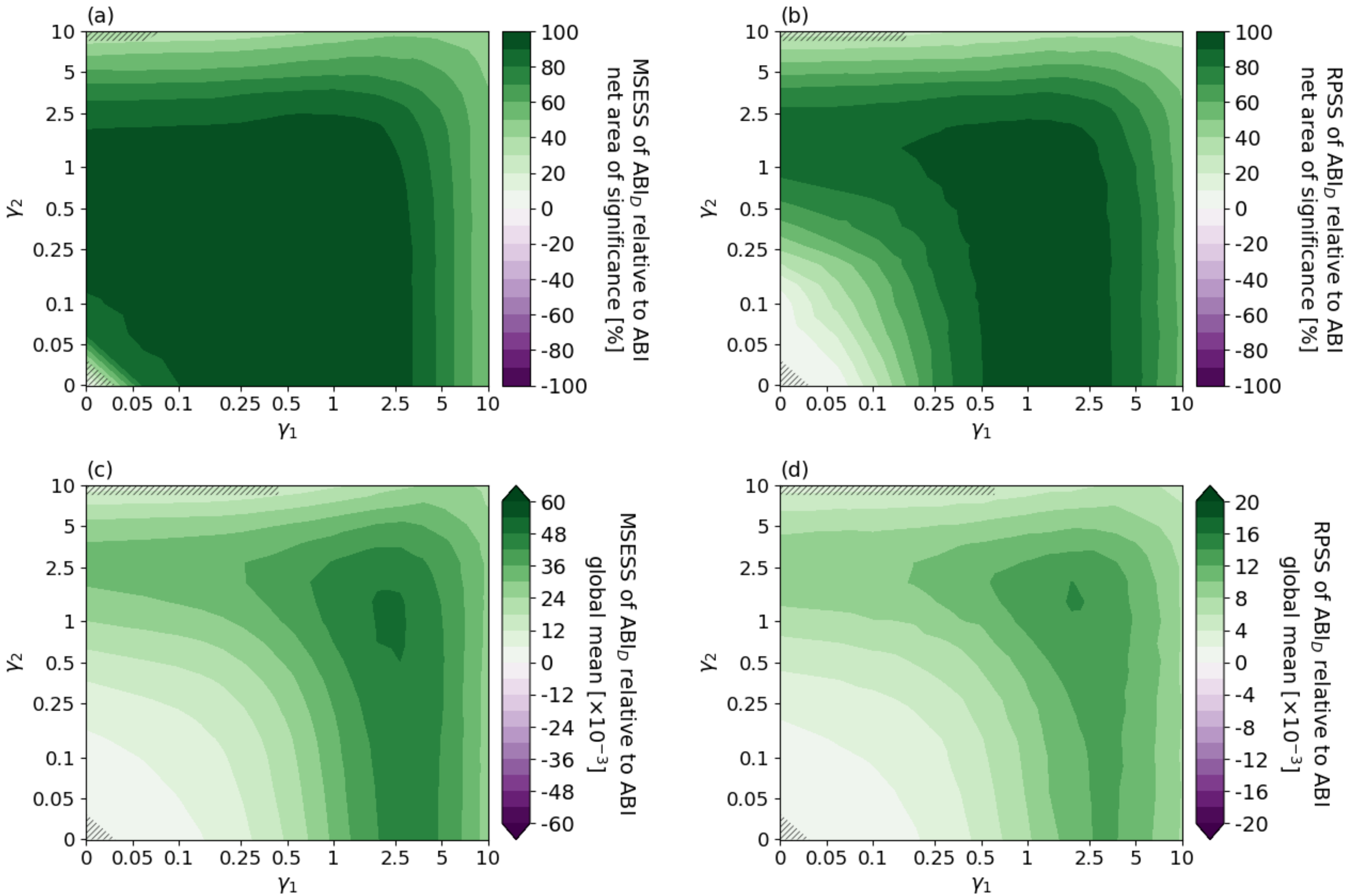
Model name	N
CanESM5	20
CMCC-CM2-SR5	10
EC-Earth3	10
HadGEM3-GC31-MM	9
MIROC6	10
MPI-ESM1-2-LR	16
Total	75

Applying deviance analysis to SST realisations from 523 ensemble members (39 models)

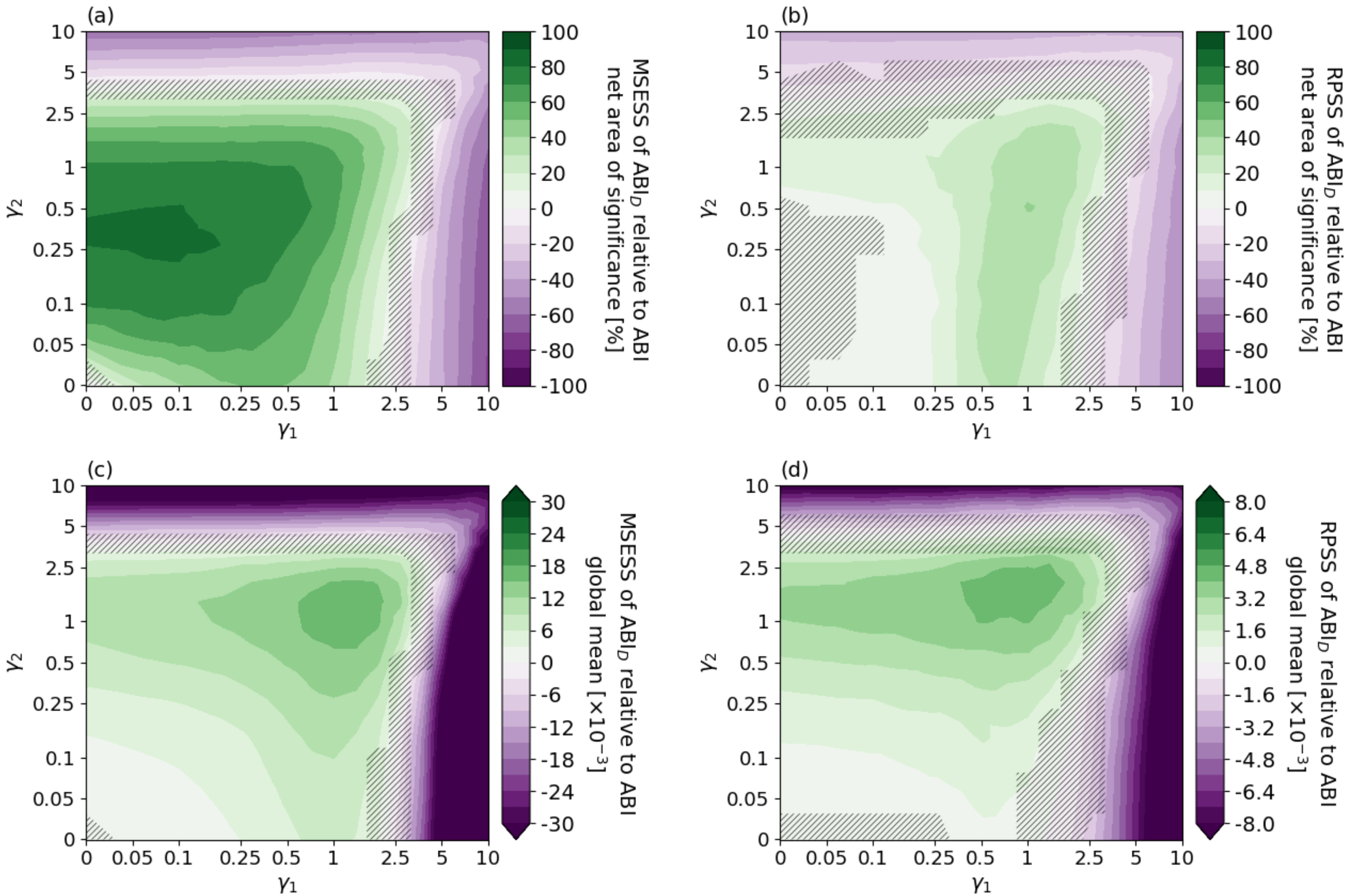


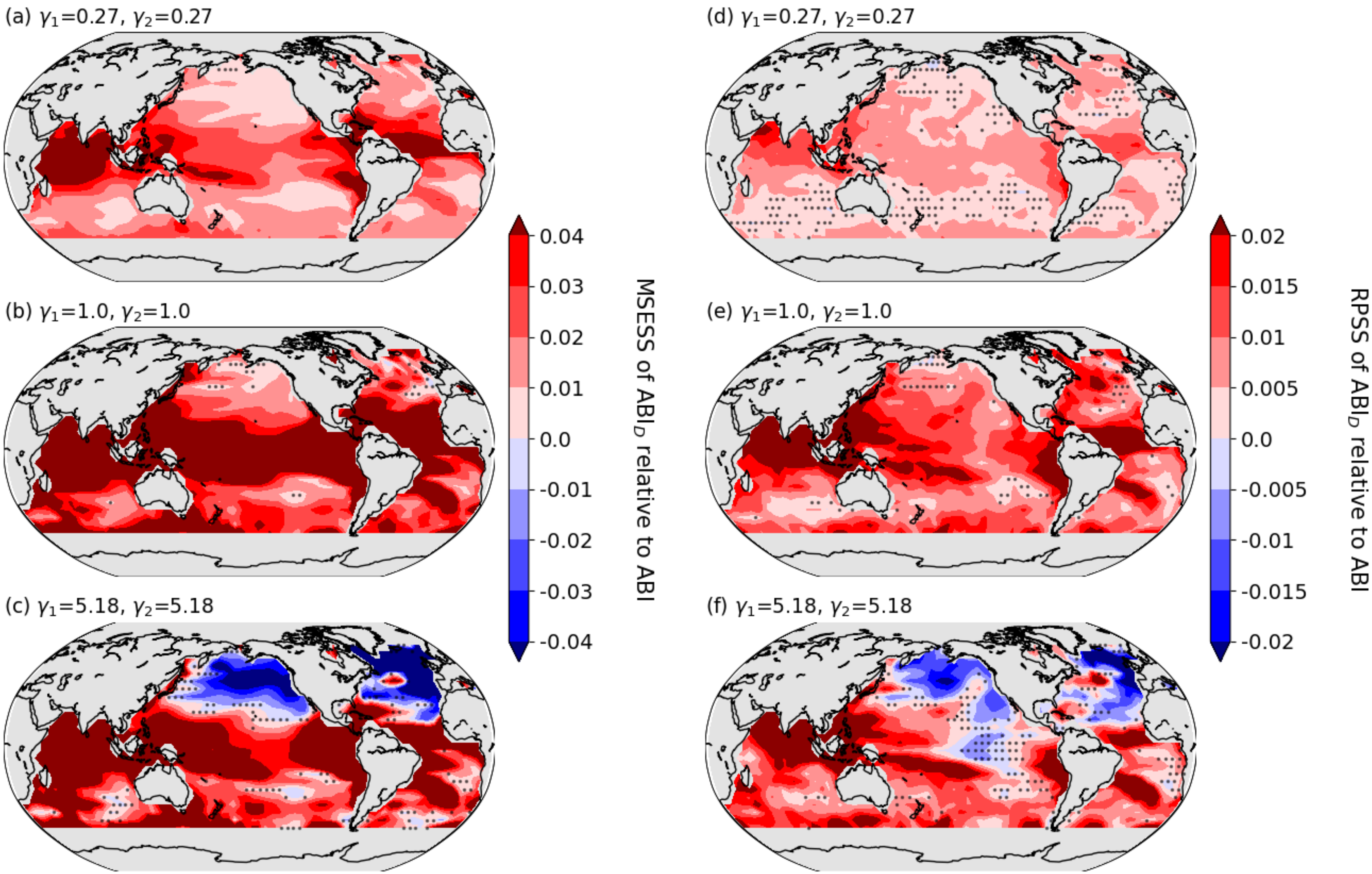


(a)
0
5
10
15
20
25
30
35
39
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
D_{tot}
(b)
2.2
1.8
1.4
1.0
0.6
Median
IQR
1.5×IQR
Outliers
α = 0.01 level
D_{tot} wrt model pseudo-observation
7.0%
55.4%
99.4%
same model
same institution
different model and institution
0: ERSSTv6
1: ACCESS-CM2
2: ACCESS-ESM1-5
3: AWI-CM-1-1-MR
4: BCC-CSM2-MR
5: CAMS-CSM1-0
6: CAS-ESM2-0
7: CESM2
8: CESM2-WACCM
9: CIESM
10: CMCC-CM2-SR5
11: CMCC-ESM2
12: CNRM-CM6-1-HR
13: CNRM-CM6-1
14: CNRM-ESM2-1
15: CanESM5
16: EC-Earth3
17: FGOALS-f3-L
18: FGOALS-g3
19: FIO-ESM-2-0
20: GFDL-ESM4
21: GISS-E2-1-G
22: GISS-E2-2-G
23: HadGEM3-GC31-LL
24: INM-CM4-8
25: INM-CM5-0
26: IPSL-CM6A-LR
27: KACE-1-0-G
28: KIOST-ESM
29: MCM-UA-1-0
30: MIROC-ES2L
31: MIROC6
32: MPI-ESM1-2-HR
33: MPI-ESM1-2-LR
34: MRI-ESM2-0
35: NESM3
36: NorESM2-LM
37: NorESM2-MM
38: TaiESM1
39: UKESM1-0-LL

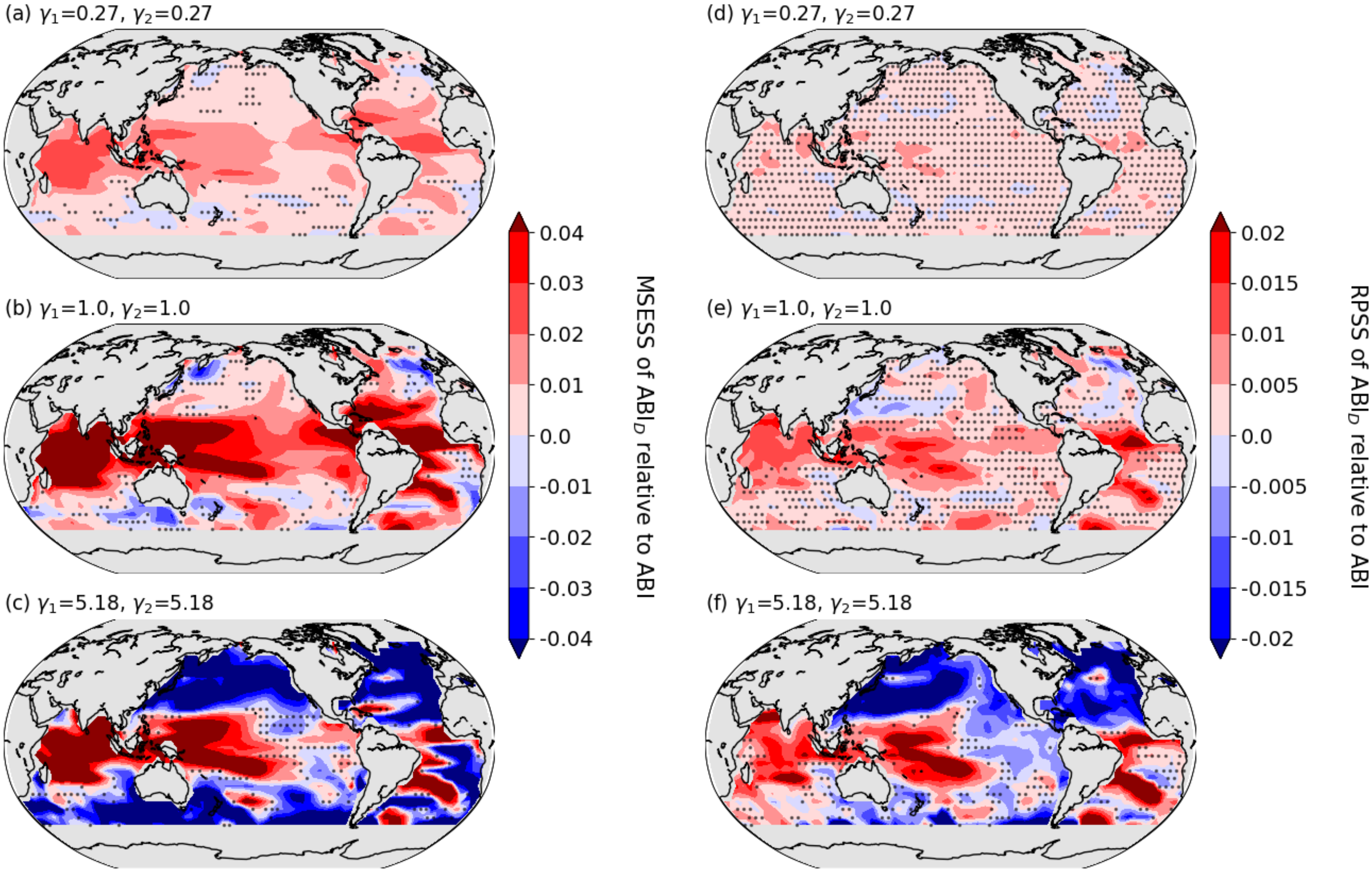


(a)
(b)
(c)
(d)
γ_1
γ_2
0 0.05 0.1 0.25 0.5 1 2.5 5 10
MSESS of ABI$_D$ relative to ABI net area of significance [%]
RPSS of ABI$_D$ relative to ABI net area of significance [%]
MSESS of ABI$_D$ relative to ABI global mean [×10^{-3}]
RPSS of ABI$_D$ relative to ABI global mean [×10^{-3}]
100 80 60 40 20 0 -20 -40 -60 -80 -100
60 48 36 24 12 0 -12 -24 -36 -48 -60
20 16 12 8 4 0 -4 -8 -12 -16 -20

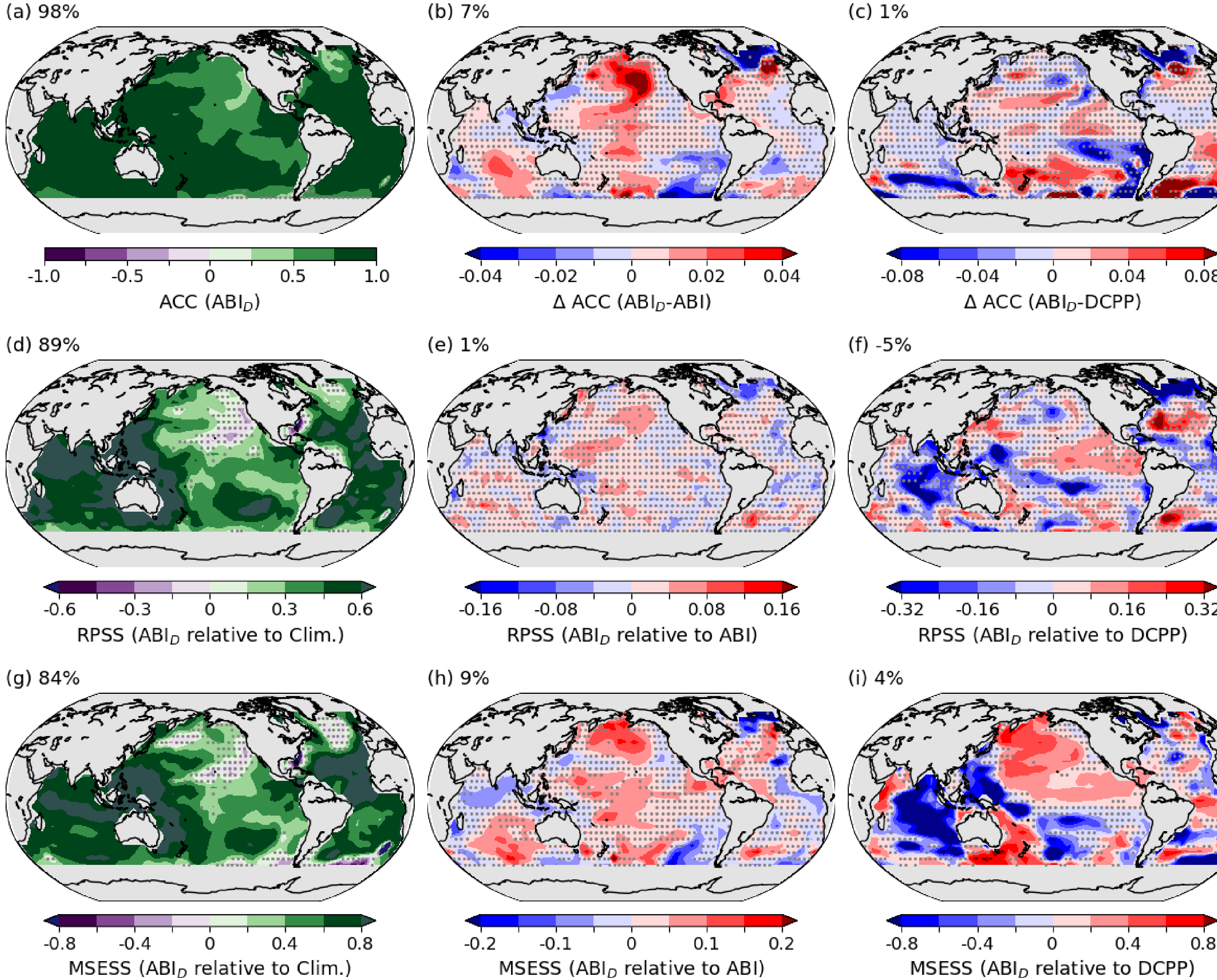

(a)
(b)
(c)
(d)
γ_1
γ_2
0
0.05
0.1
0.25
0.5
1
2.5
5
10
MSESS of ABI$_D$ relative to ABI net area of significance [%]
RPSS of ABI$_D$ relative to ABI net area of significance [%]
MSESS of ABI$_D$ relative to ABI global mean [$\times 10^{-3}$]
RPSS of ABI$_D$ relative to ABI global mean [$\times 10^{-3}$]
100
80
60
40
20
0
-20
-40
-60
-80
-100
30
24
18
12
6
0
-6
-12
-18
-24
-30
8.0
6.4
4.8
3.2
1.6
0.0
-1.6
-3.2
-4.8
-6.4
-8.0



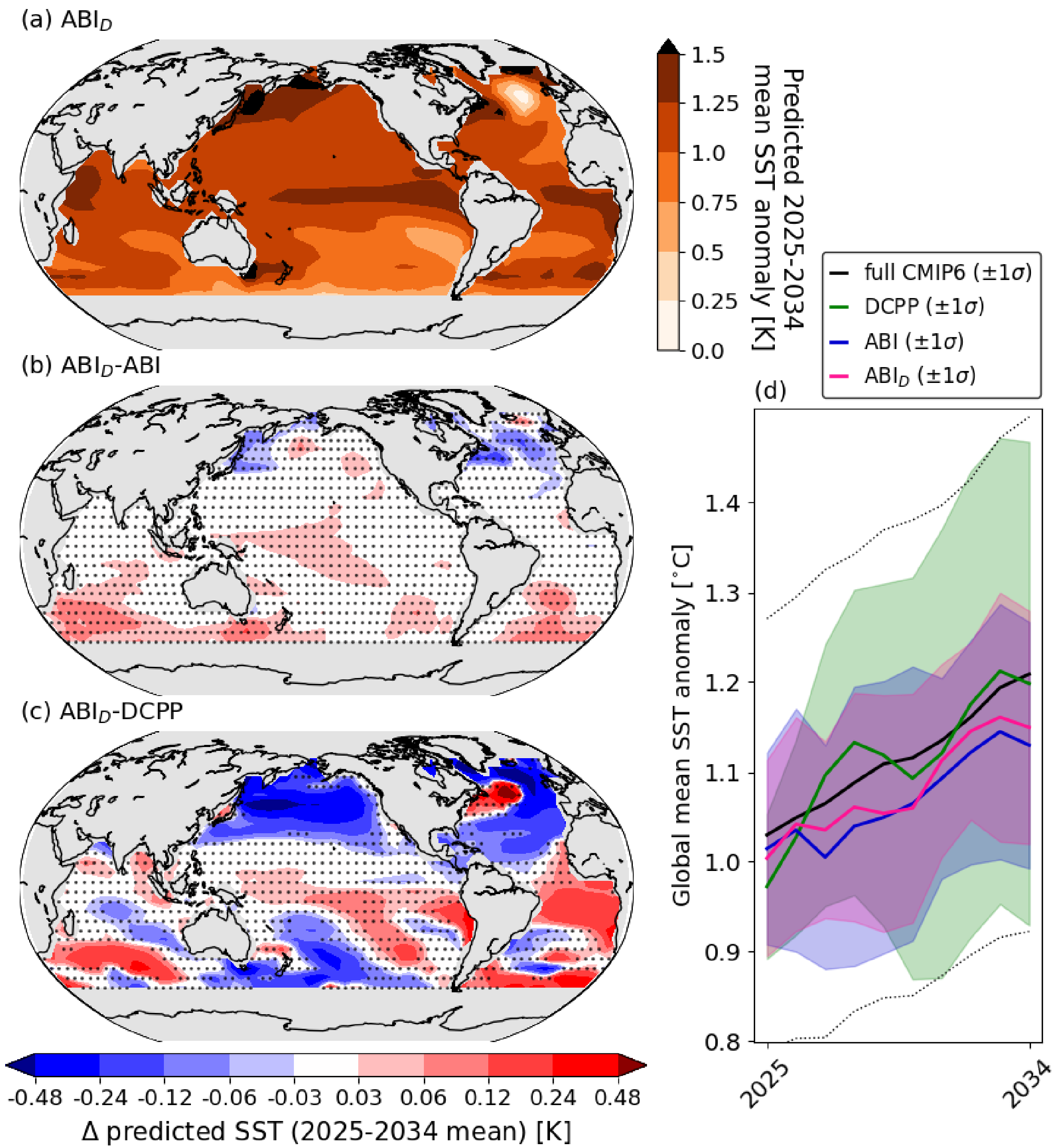
(a) γ_1=0.27, γ_2=0.27
(b) γ_1=1.0, γ_2=1.0
(c) γ_1=5.18, γ_2=5.18
0.04
0.03
0.02
0.01
0.0
-0.01
-0.02
-0.03
-0.04
MSESS of ABI_D relative to ABI
(d) γ_1=0.27, γ_2=0.27
(e) γ_1=1.0, γ_2=1.0
(f) γ_1=5.18, γ_2=5.18
0.02
0.015
0.01
0.005
0.0
-0.005
-0.01
-0.015
-0.02
RPSS of ABI_D relative to ABI



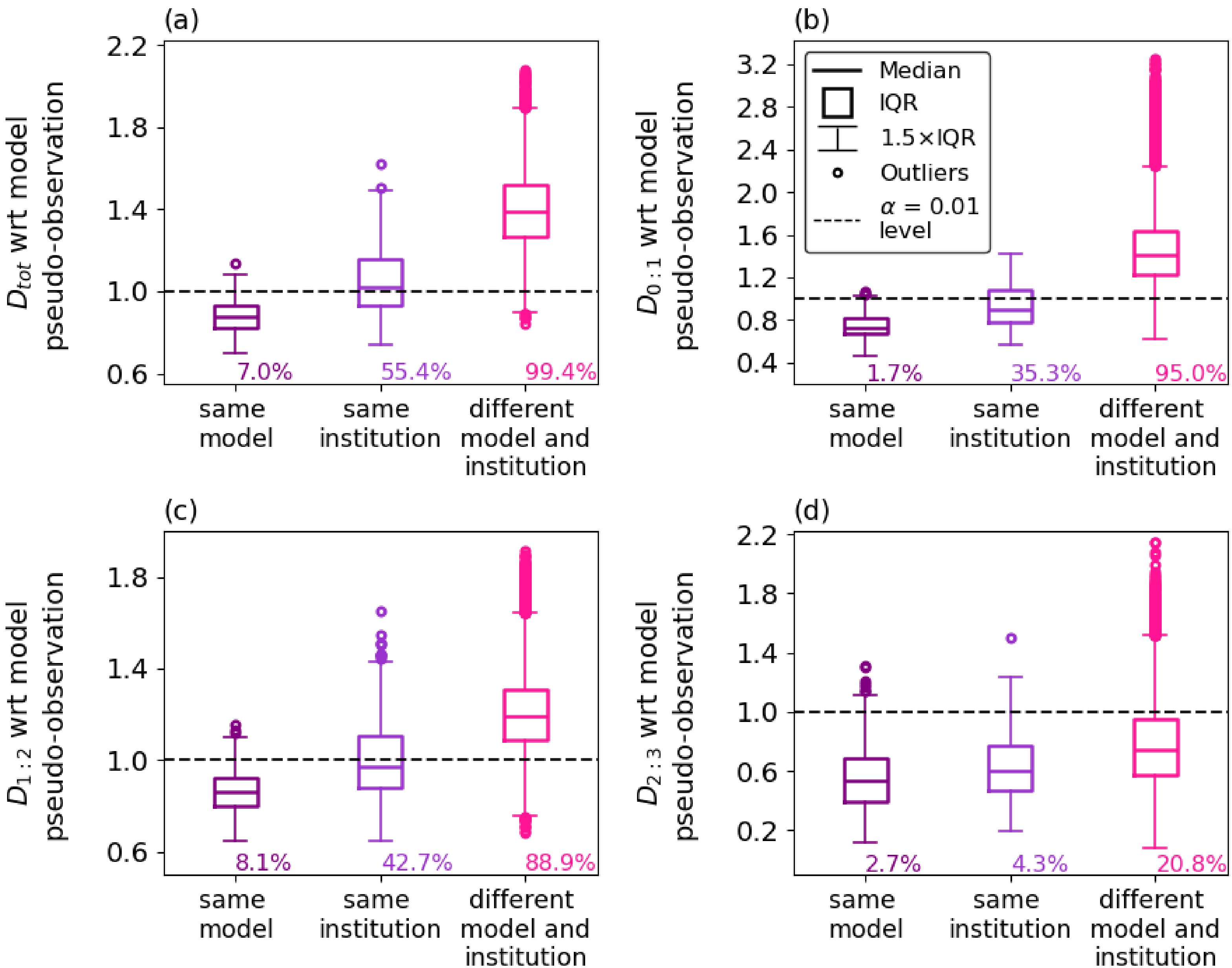
(a) γ_1=0.27, γ_2=0.27
(b) γ_1=1.0, γ_2=1.0
(c) γ_1=5.18, γ_2=5.18
0.04
0.03
0.02
0.01
0.0
-0.01
-0.02
-0.03
-0.04
MSESS of ABI_D relative to ABI
(d) γ_1=0.27, γ_2=0.27
(e) γ_1=1.0, γ_2=1.0
(f) γ_1=5.18, γ_2=5.18
0.02
0.015
0.01
0.005
0.0
-0.005
-0.01
-0.015
-0.02
RPSS of ABI_D relative to ABI

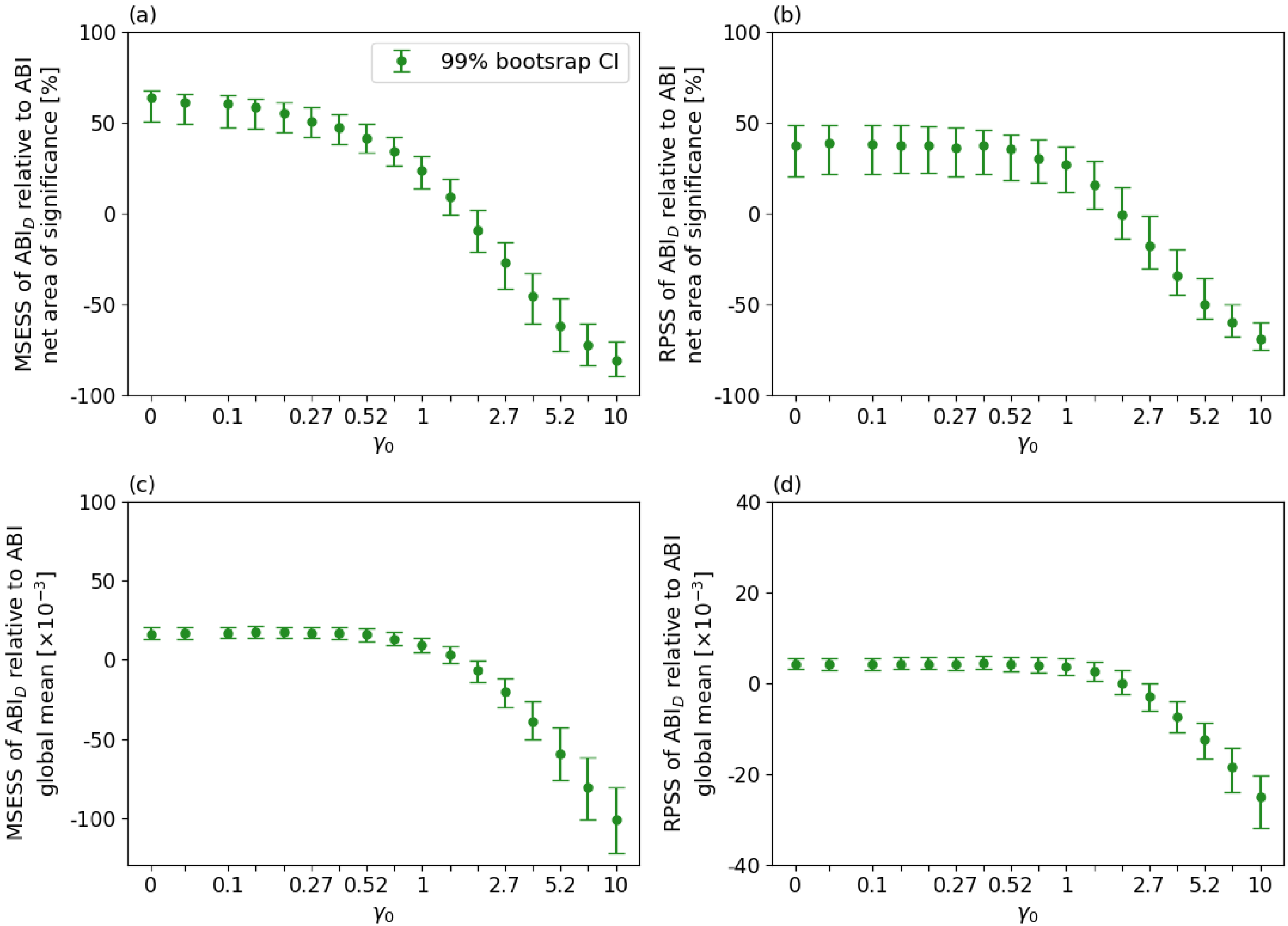


(a) 98%
-1.0 -0.5 0 0.5 1.0
ACC (ABI$_D$)
(b) 7%
-0.04 -0.02 0 0.02 0.04
Δ ACC (ABI$_D$-ABI)
(c) 1%
-0.08 -0.04 0 0.04 0.08
Δ ACC (ABI$_D$-DCPP)
(d) 89%
-0.6 -0.3 0 0.3 0.6
RPSS (ABI$_D$ relative to Clim.)
(e) 1%
-0.16 -0.08 0 0.08 0.16
RPSS (ABI$_D$ relative to ABI)
(f) -5%
-0.32 -0.16 0 0.16 0.32
RPSS (ABI$_D$ relative to DCPP)
(g) 84%
-0.8 -0.4 0 0.4 0.8
MSESS (ABI$_D$ relative to Clim.)
(h) 9%
-0.2 -0.1 0 0.1 0.2
MSESS (ABI$_D$ relative to ABI)
(i) 4%
-0.8 -0.4 0 0.4 0.8
MSESS (ABI$_D$ relative to DCPP)

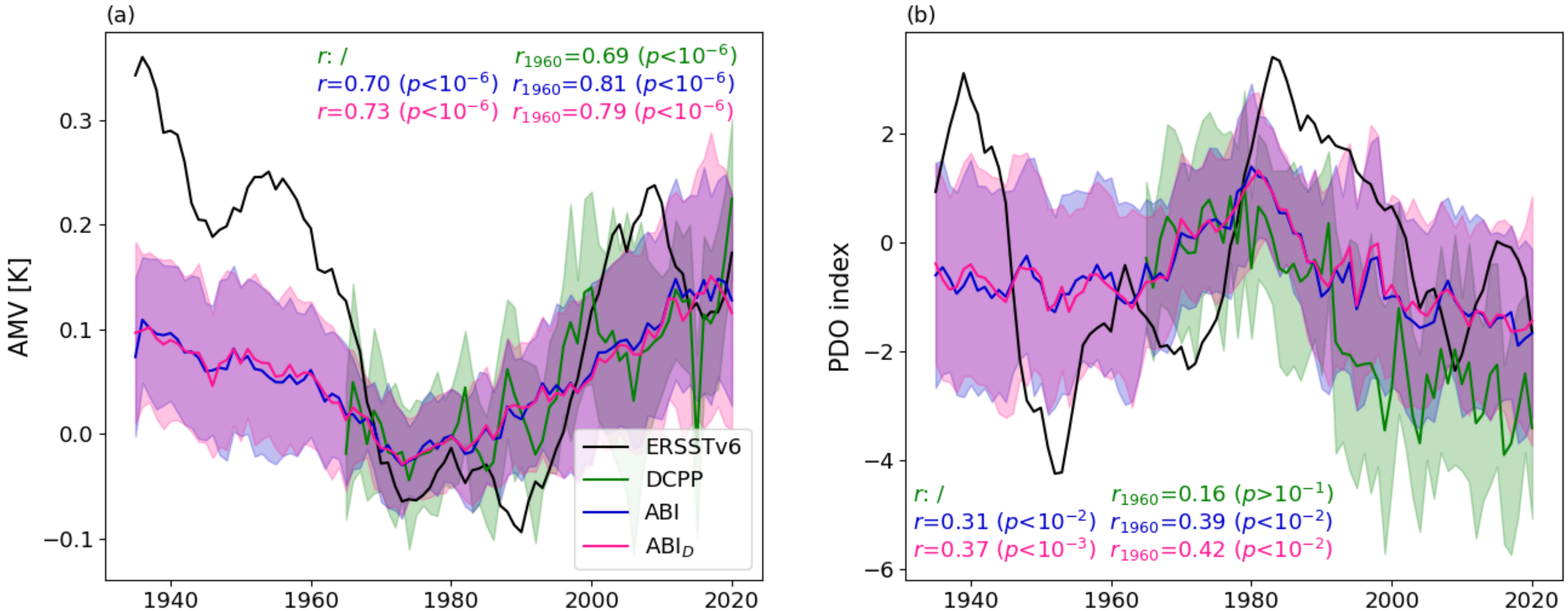


(a) ABI$_D$
Predicted 2025-2034 mean SST anomaly [K]
0.0
0.25
0.5
0.75
1.0
1.25
1.5
(b) ABI$_D$-ABI
(c) ABI$_D$-DCPP
-0.48 -0.24 -0.12 -0.06 -0.03 0.03 0.06 0.12 0.24 0.48
Δ predicted SST (2025-2034 mean) [K]
full CMIP6 (±1σ)
DCPP (±1σ)
ABI (±1σ)
ABI$_D$ (±1σ)
(d)
Global mean SST anomaly [°C]
0.8
0.9
1.0
1.1
1.2
1.3
1.4
2025
2034

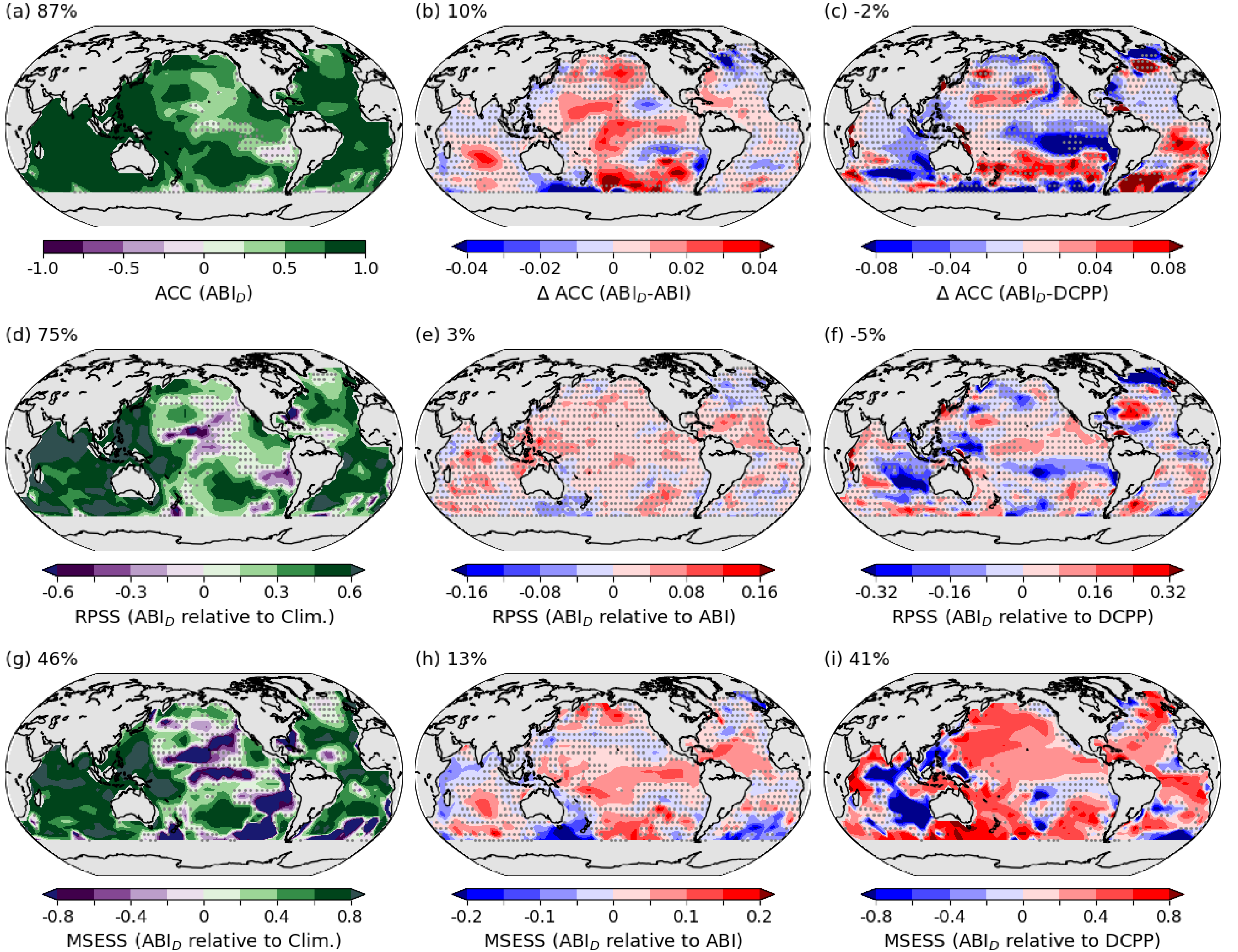

(a)
2.2
1.8
1.4
1.0
0.6
D_{tot} wrt model pseudo-observation
7.0%
55.4%
99.4%
same model
same institution
different model and institution
(b)
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4
$D_{0:1}$ wrt model pseudo-observation
Median
IQR
1.5×IQR
Outliers
α = 0.01 level
1.7%
35.3%
95.0%
same model
same institution
different model and institution
(c)
1.8
1.4
1.0
0.6
$D_{1:2}$ wrt model pseudo-observation
8.1%
42.7%
88.9%
same model
same institution
different model and institution
(d)
2.2
1.8
1.4
1.0
0.6
0.2
$D_{2:3}$ wrt model pseudo-observation
2.7%
4.3%
20.8%
same model
same institution
different model and institution



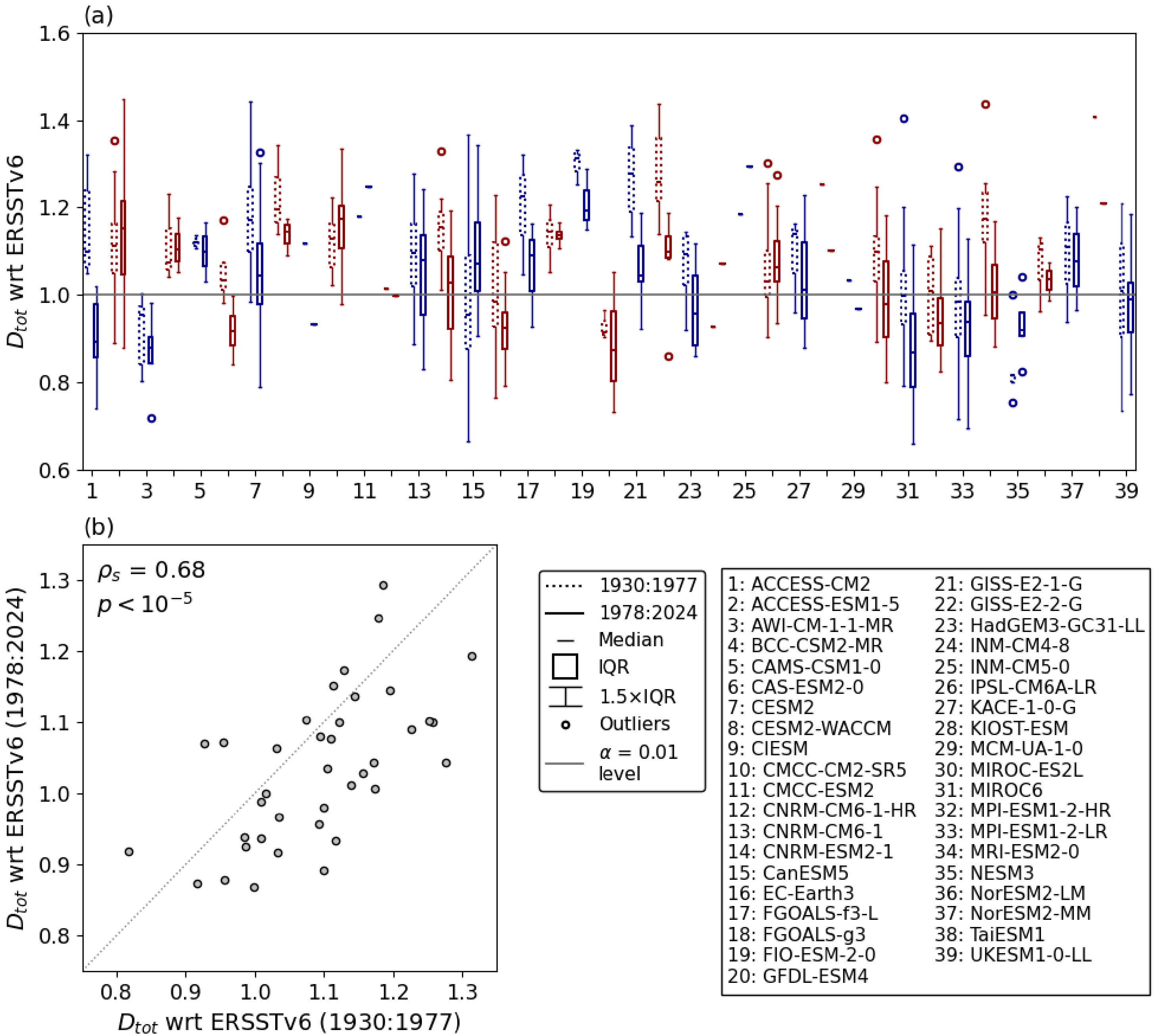
(a)
(b)
(c)
(d)
99% bootsrap CI
MSESS of ABI_D relative to ABI net area of significance [%]
RPSS of ABI_D relative to ABI net area of significance [%]
MSESS of ABI_D relative to ABI global mean [$\times 10^{-3}$]
RPSS of ABI_D relative to ABI global mean [$\times 10^{-3}$]
γ_0
0
0.1
0.27
0.52
1
2.7
5.2
10
100
50
0
-50
-100
40
20
-20
-40



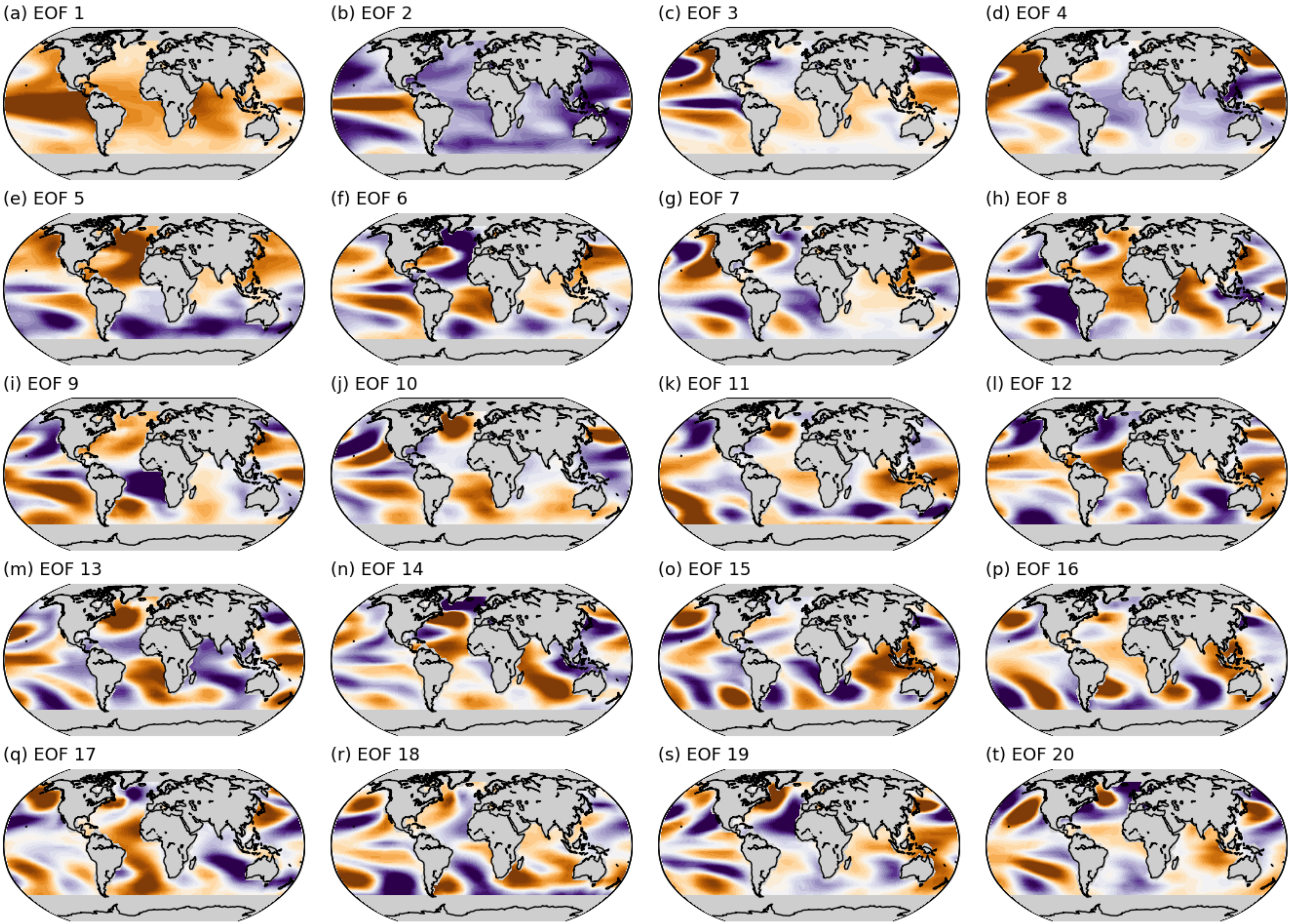
(a)
r: /
r=0.70 (p<10^−6)
r=0.73 (p<10^−6)
r_1960=0.69 (p<10^−6)
r_1960=0.81 (p<10^−6)
r_1960=0.79 (p<10^−6)
AMV [K]
ERSSTv6
DCPP
ABI
ABI_D
(b)
PDO index
r: /
r=0.31 (p<10^−2)
r=0.37 (p<10^−3)
r_1960=0.16 (p>10^−1)
r_1960=0.39 (p<10^−2)
r_1960=0.42 (p<10^−2)

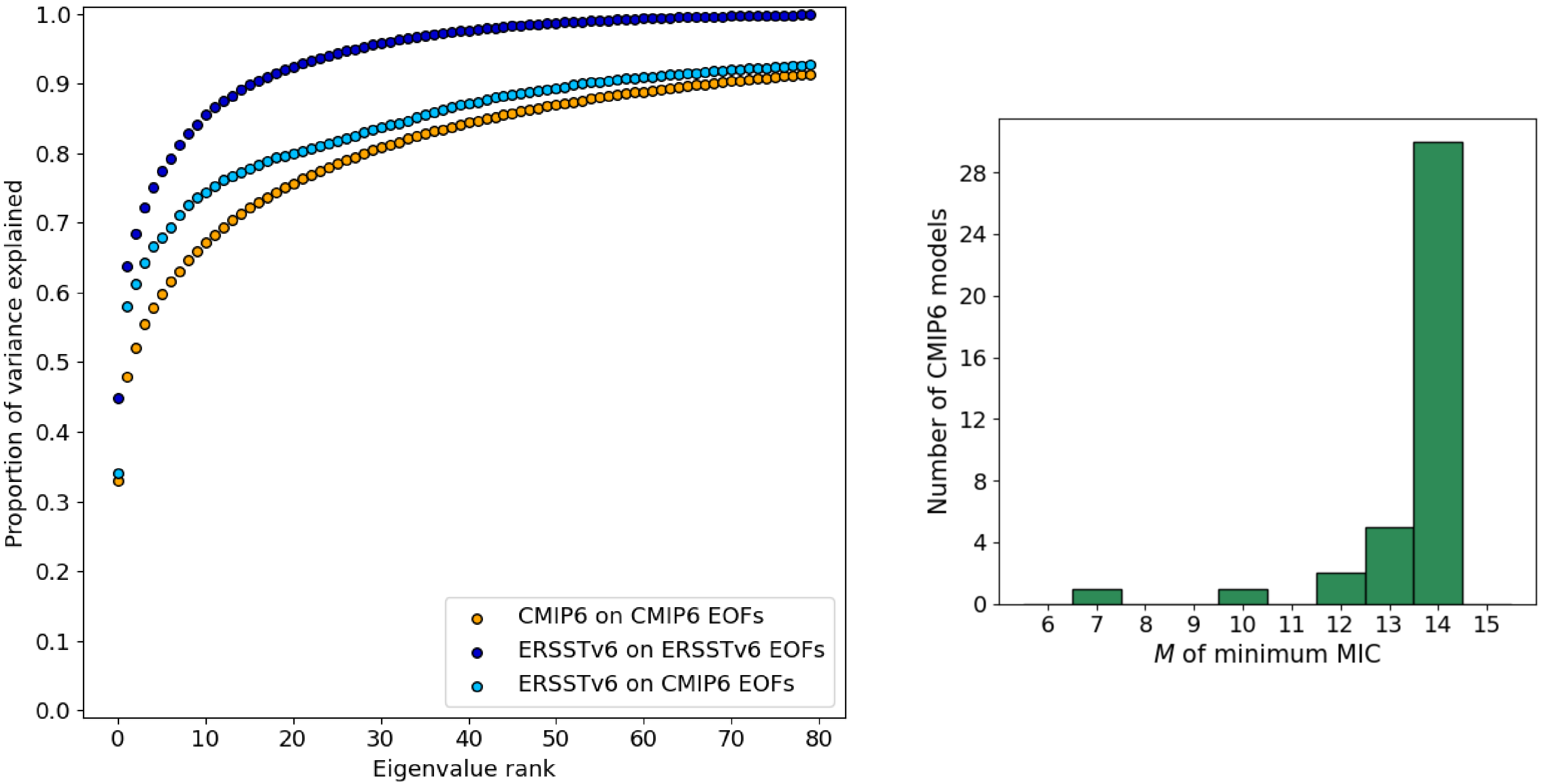


(a) 87%
(b) 10%
(c) -2%
-1.0 -0.5 0 0.5 1.0
ACC (ABI_D)
-0.04 -0.02 0 0.02 0.04
Δ ACC (ABI_D-ABI)
-0.08 -0.04 0 0.04 0.08
Δ ACC (ABI_D-DCPP)
(d) 75%
(e) 3%
(f) -5%
-0.6 -0.3 0 0.3 0.6
RPSS (ABI_D relative to Clim.)
-0.16 -0.08 0 0.08 0.16
RPSS (ABI_D relative to ABI)
-0.32 -0.16 0 0.16 0.32
RPSS (ABI_D relative to DCPP)
(g) 46%
(h) 13%
(i) 41%
-0.8 -0.4 0 0.4 0.8
MSESS (ABI_D relative to Clim.)
-0.2 -0.1 0 0.1 0.2
MSESS (ABI_D relative to ABI)
-0.8 -0.4 0 0.4 0.8
MSESS (ABI_D relative to DCPP)



(a)
D_tot wrt ERSSTv6
1.6
1.4
1.2
1.0
0.8
0.6
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
35
37
39
(b)
ρ_s = 0.68
p < 10^-5
D_tot wrt ERSSTv6 (1978:2024)
D_tot wrt ERSSTv6 (1930:1977)
0.8
0.9
1.0
1.1
1.2
1.3
1930:1977
1978:2024
Median
IQR
1.5×IQR
Outliers
α = 0.01 level
1: ACCESS-CM2
2: ACCESS-ESM1-5
3: AWI-CM-1-1-MR
4: BCC-CSM2-MR
5: CAMS-CSM1-0
6: CAS-ESM2-0
7: CESM2
8: CESM2-WACCM
9: CIESM
10: CMCC-CM2-SR5
11: CMCC-ESM2
12: CNRM-CM6-1-HR
13: CNRM-CM6-1
14: CNRM-ESM2-1
15: CanESM5
16: EC-Earth3
17: FGOALS-f3-L
18: FGOALS-g3
19: FIO-ESM-2-0
20: GFDL-ESM4
21: GISS-E2-1-G
22: GISS-E2-2-G
23: HadGEM3-GC31-LL
24: INM-CM4-8
25: INM-CM5-0
26: IPSL-CM6A-LR
27: KACE-1-0-G
28: KIOST-ESM
29: MCM-UA-1-0
30: MIROC-ES2L
31: MIROC6
32: MPI-ESM1-2-HR
33: MPI-ESM1-2-LR
34: MRI-ESM2-0
35: NESM3
36: NorESM2-LM
37: NorESM2-MM
38: TaiESM1
39: UKESM1-0-LL


(a) EOF 1
(b) EOF 2
(c) EOF 3
(d) EOF 4
(e) EOF 5
(f) EOF 6
(g) EOF 7
(h) EOF 8
(i) EOF 9
(j) EOF 10
(k) EOF 11
(l) EOF 12
(m) EOF 13
(n) EOF 14
(o) EOF 15
(p) EOF 16
(q) EOF 17
(r) EOF 18
(s) EOF 19
(t) EOF 20


Proportion of variance explained
Eigenvalue rank
CMIP6 on CMIP6 EOFs
ERSSTv6 on ERSSTv6 EOFs
ERSSTv6 on CMIP6 EOFs
Number of CMIP6 models
M of minimum MIC