



cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

20 Years

UNIBO GUEST LECTURES

Expanding CMCC Seasonal Prediction System v3.5 applications to the local scale through STATISTICAL DOWNSCALING TECHNIQUES

Leonardo Aragão¹, Andrea Borrelli² & Silvio Gualdi¹

¹Earth SYstem modeling and Data Assimilation (ESYDA)

²CLimate VARIability and Predictions (CLIVAP)

www.cmcc.it



A satellite image of Earth showing the African continent, Europe, and the Middle East. The image is taken from a high angle, showing the curvature of the planet. The landmasses are visible in shades of brown and green, while the oceans are a deep blue. Swirling white clouds are scattered across the scene, particularly over the Atlantic and Indian Oceans. The text "Seasonal Prediction" is overlaid in the bottom left corner in a bold, grey, sans-serif font.

Seasonal Prediction



Seasonal Prediction

provide a long-range outlook of changes in the Earth system over periods of a few weeks or months, as a result of predictable changes in some of the slow-varying components of the system.



Seasonal Prediction

provide a **long-range outlook** of changes in the **Earth system** over periods of a few weeks or months, as a result of **predictable** changes in some of the **slow-varying** components of the system.



to forecast  to predict
weather climate

Seasonal Prediction

provide a **long-range outlook** of changes in the **Earth system** over periods of a few weeks or months, as a result of predictable changes in some of the **slow-varying** components of the system.



to forecast  to predict

weather

product

time-stamped

climate

process

outlook

Seasonal Prediction

provide a **long-range** outlook of changes in the **Earth system** over periods of a few weeks or months, as a result of predictable changes in some of the **slow-varying** components of the system.

to forecast  to predict

weather

climate

product

process

time-stamped

outlook

short-range

long-range

event

slow-varying

deterministic model 

Earth system

Seasonal Prediction

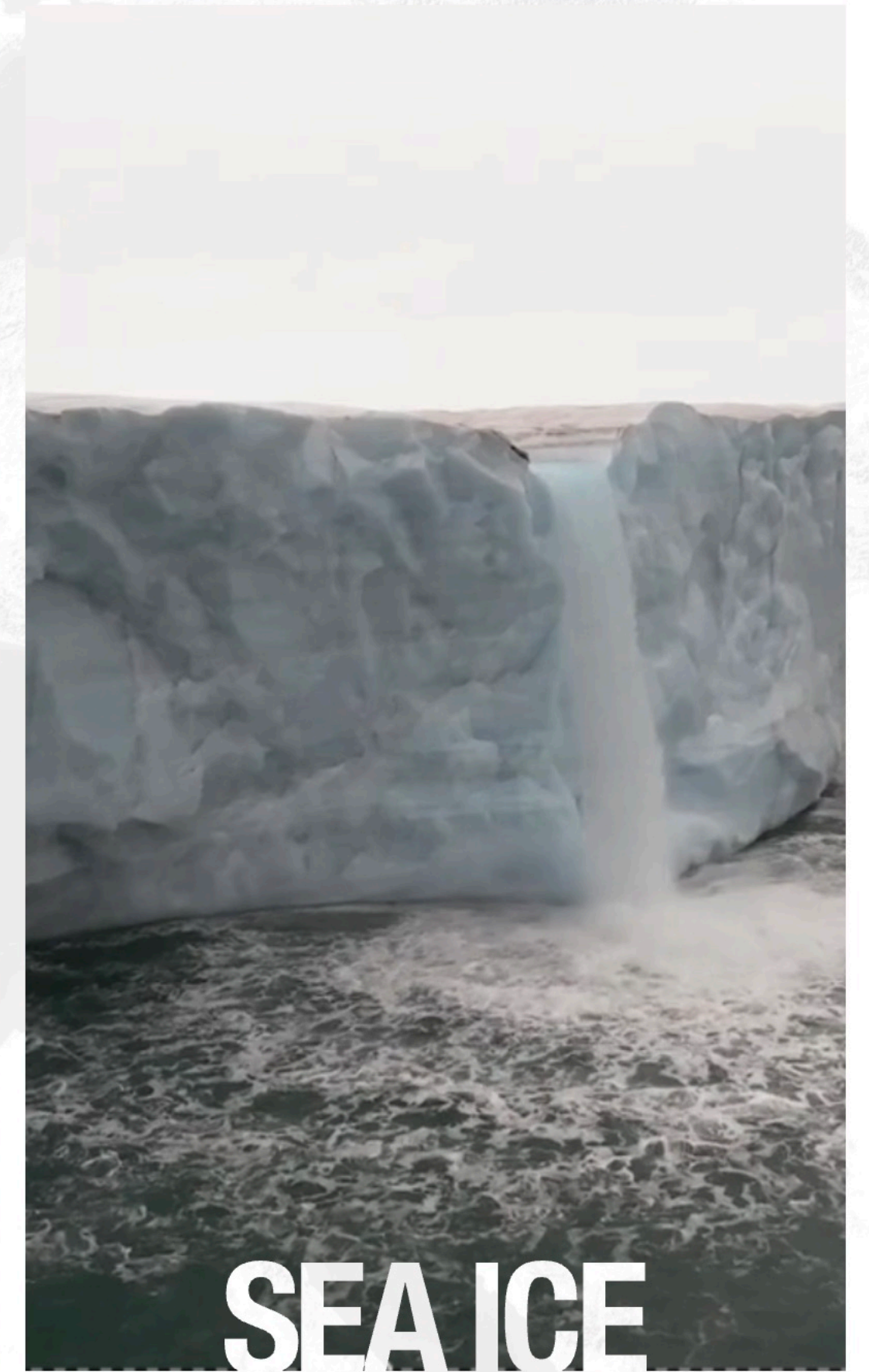
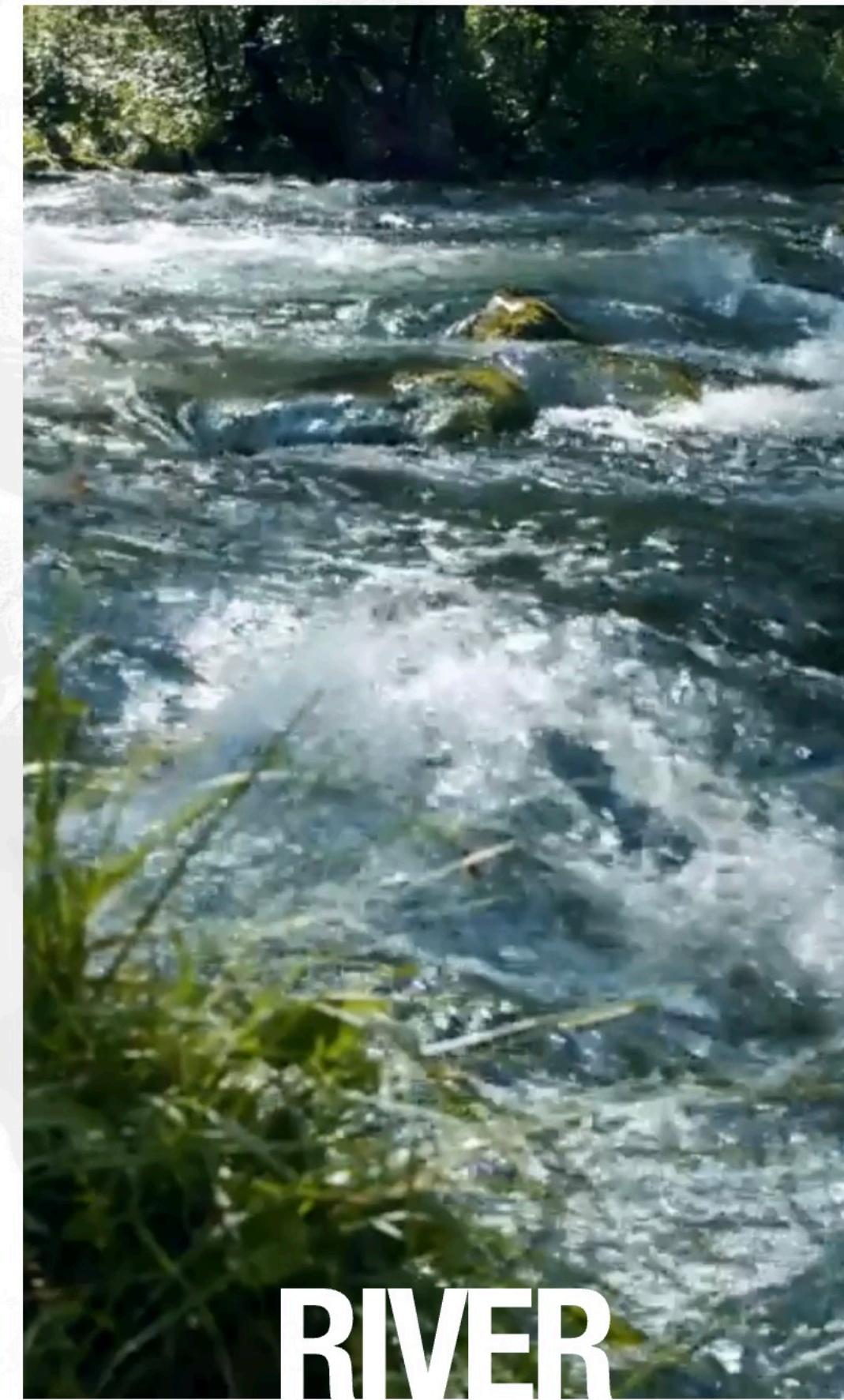
provide a long-range outlook of changes in the Earth system over periods of a few weeks or months, as a result of predictable changes in some of the slow-varying components of the system.



EARTH SYSTEM

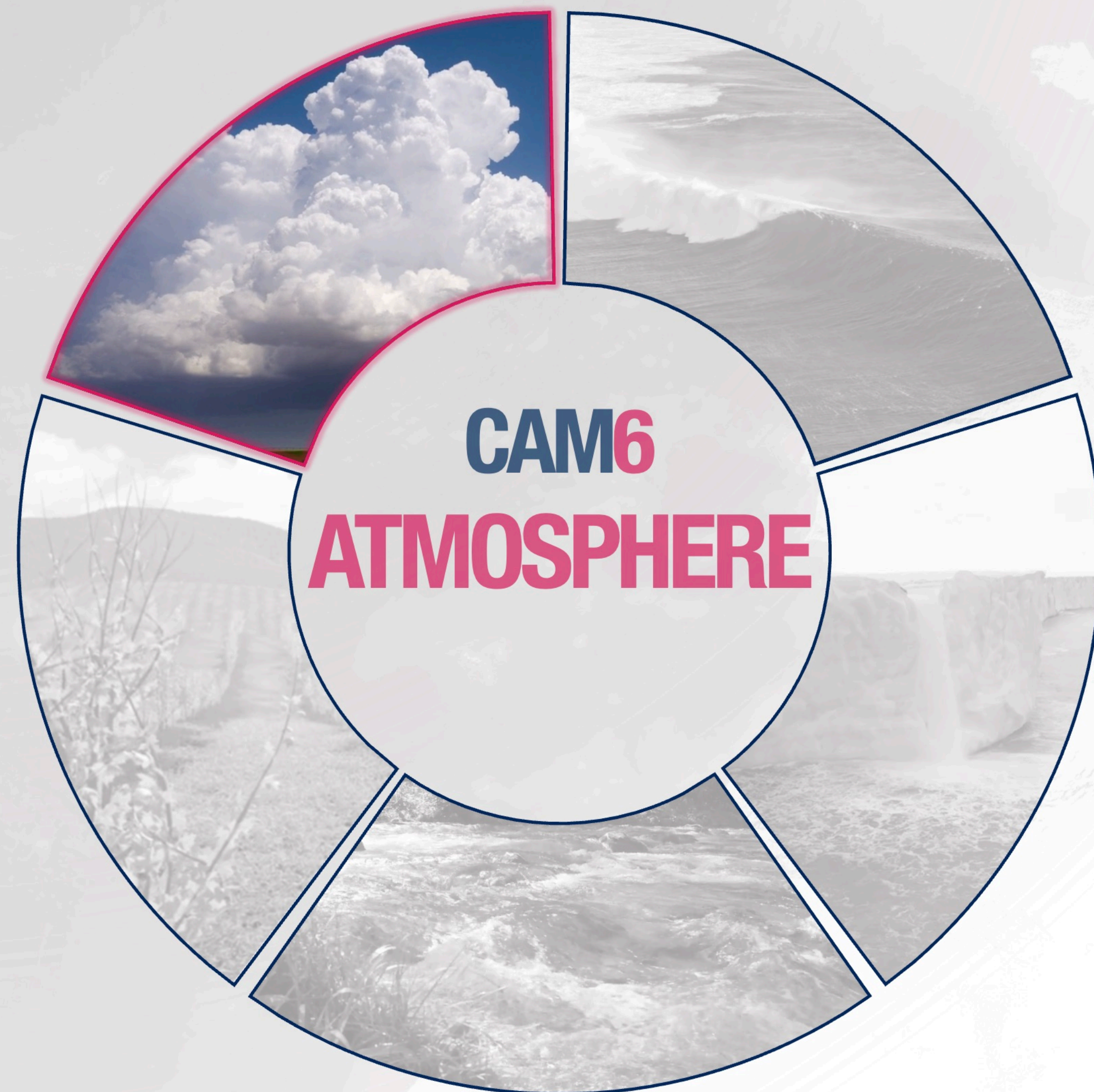
EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4



EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)



Community Atmosphere Model version 6 (5.3)

Craig *et al.* (2021)

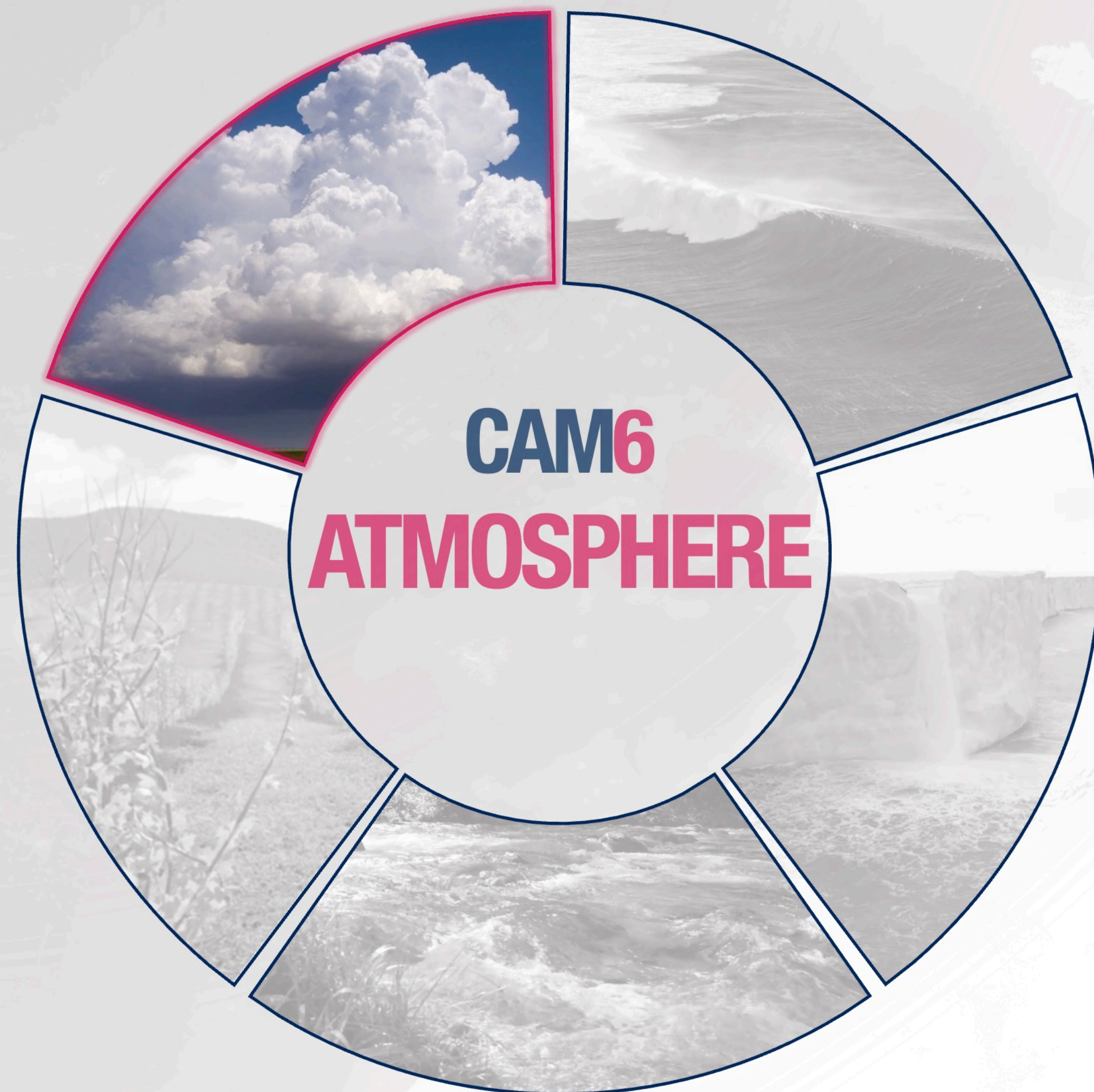
- Atmospheric component of the NCAR's Community Earth System Model (CESM)
- Regular grid of 0.47° (latitude) \times 0.63° (longitude)
- 83 vertical hybrid levels
- Troposphere and Stratosphere up to 0.01 hPa



More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)

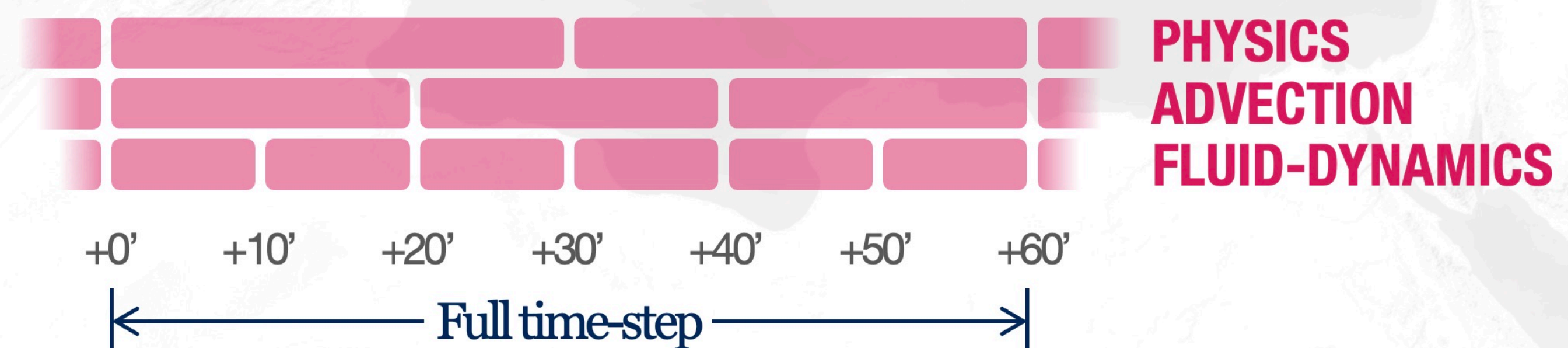


Community Atmosphere Model version 6 (5.3)

Craig *et al.* (2021)

- Atmospheric component of the NCAR's Community Earth System Model (CESM)
- Regular grid of 0.47° (latitude) \times 0.63° (longitude)
- 83 vertical hybrid levels
- Troposphere and Stratosphere up to 0.01 hPa

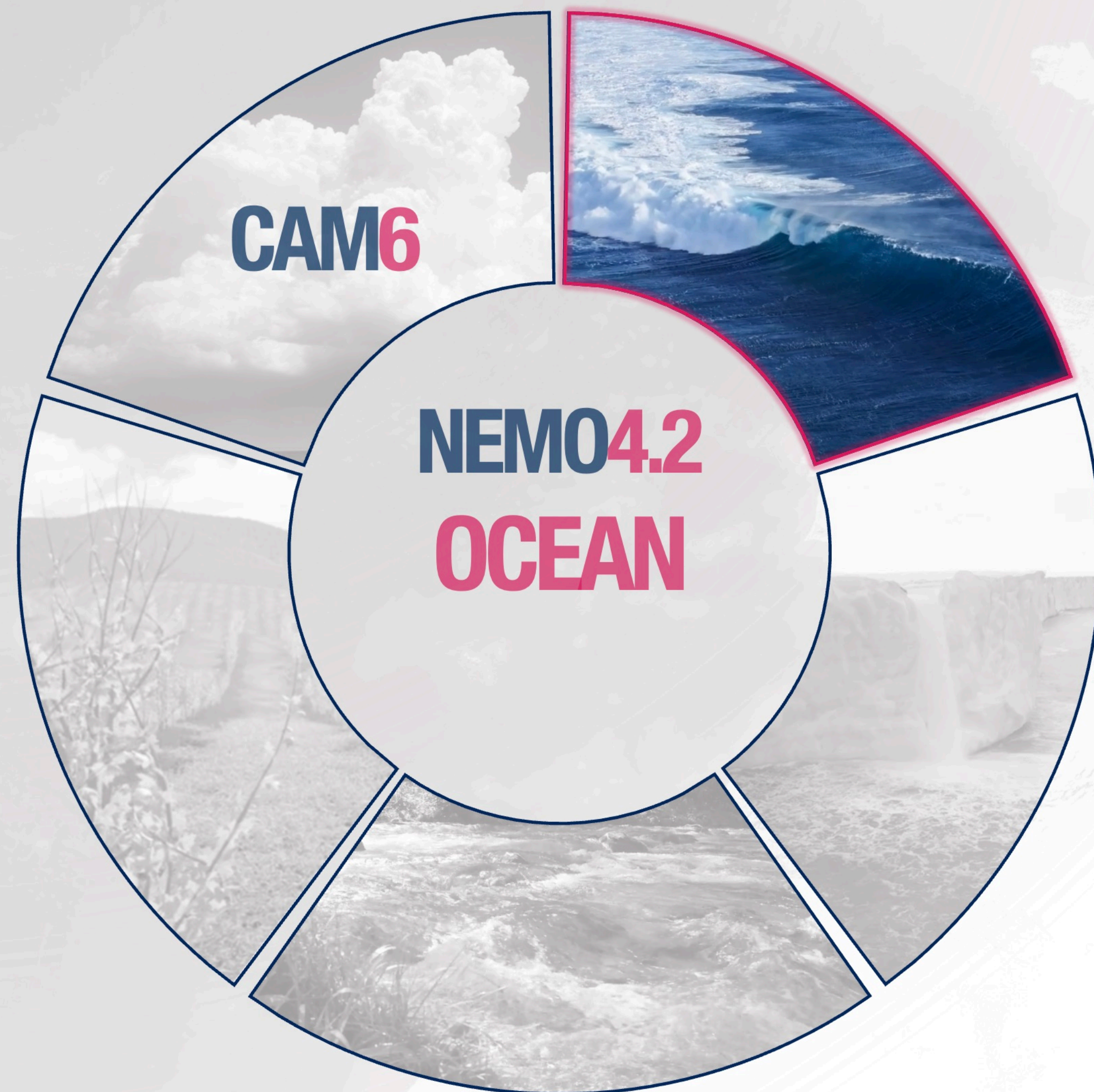
Time-steps:



More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)



Nucleus for European Modelling of the Ocean model version 4.2 (3.4)

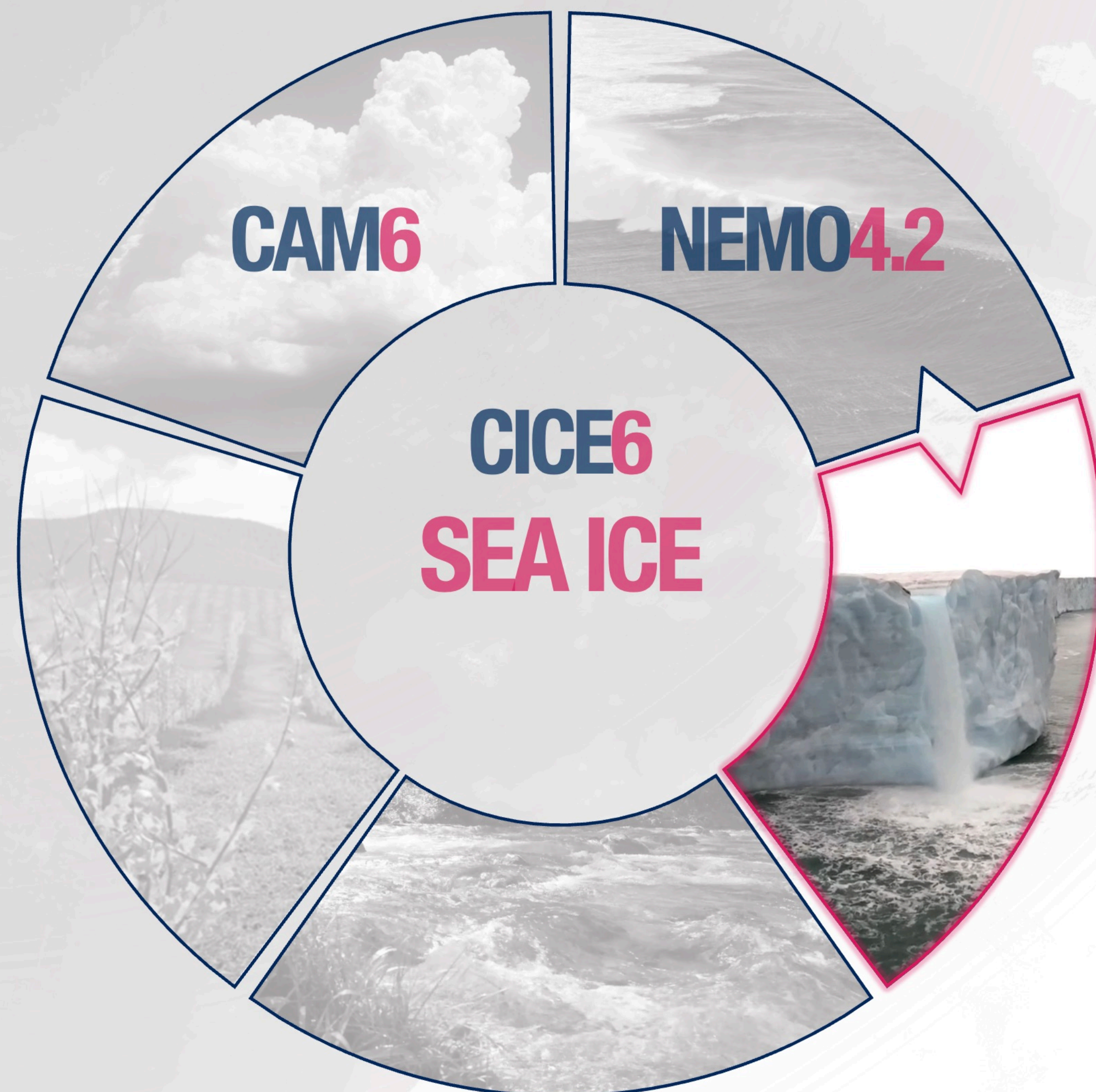
Madec *et al.* (2022)

- Tripolar ORCA grid of 0.25°
- 75 vertical levels
- 20-minute time-step
- Prognostic variables: three velocity components, the sea surface height, the potential temperature, and salinity

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)



Los Alamos sea ICE model version 6 (4)

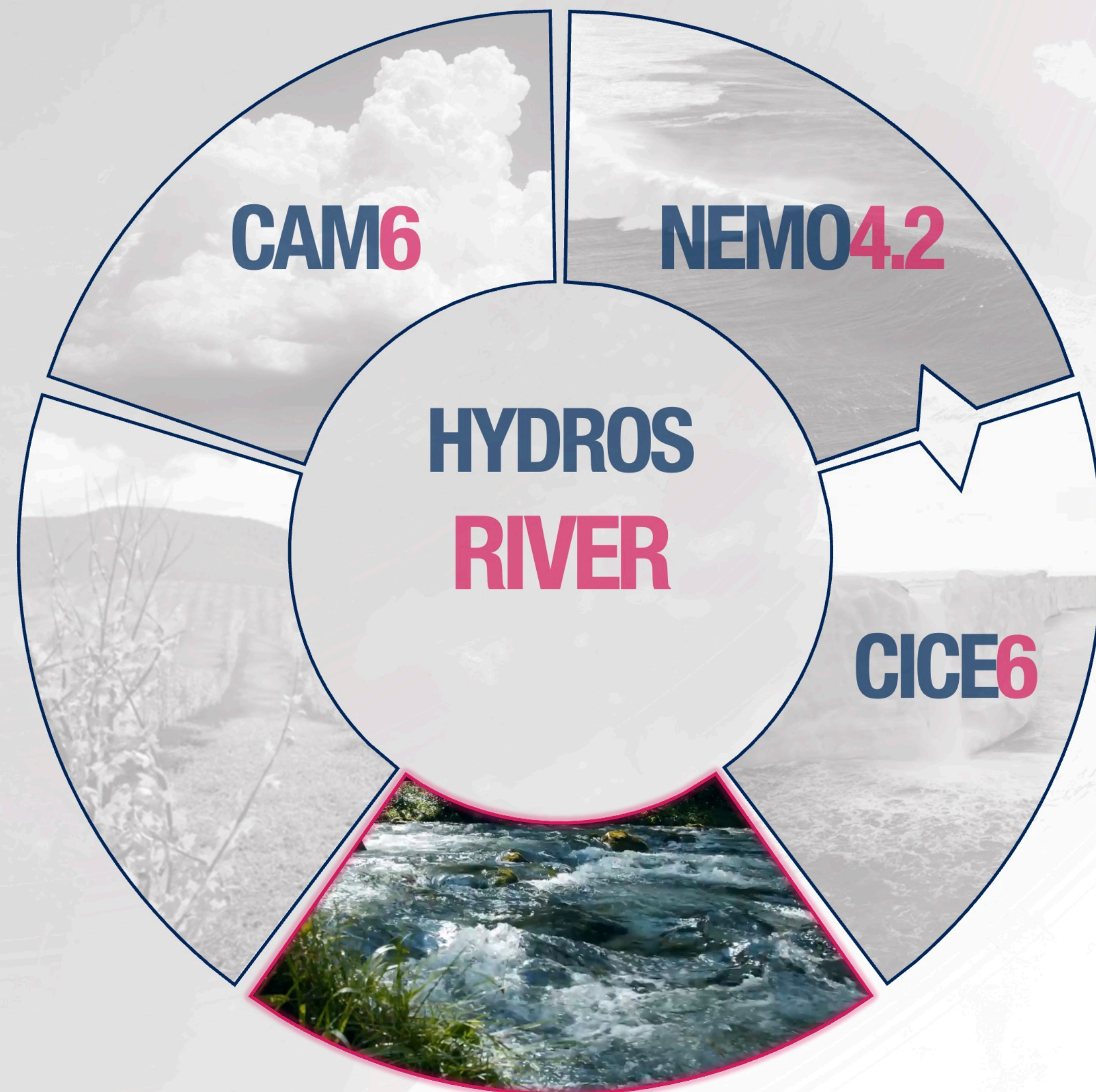
Hunke *et al.* (2018)

- Consortium ICE
- Integrated with NEMO (coupled): Tripolar ORCA grid of 0.25° 20-minute time-step
- 8 ice layers
- 3 snow layers
- 5 ice categories (single)
- No wave model

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)



HYdro-Dynamic ROuting Scheme (RTM)

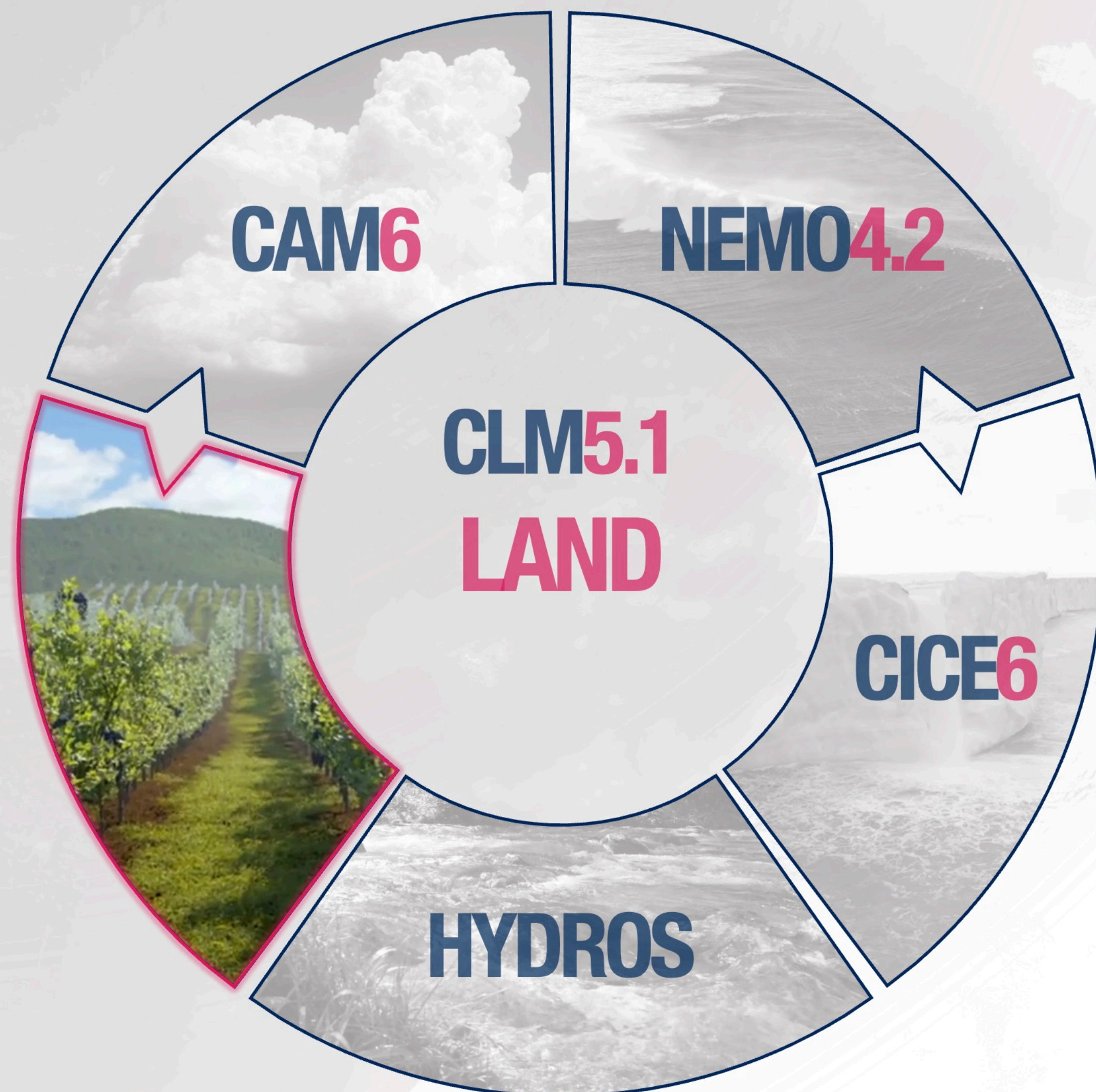
Materia *et al.* (2020)

- Replaces the River Transport Model (RTM, Graham *et al.* 1999)
- A more complex scheme for computing the flow velocity
- Generates time-dependent flow velocity influenced by orography and water availability
- Regular grid of 0.5°
- 3-hour time step
- Receive input data from the land surface model

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)



Community Land Model version 5.1 (4.5)

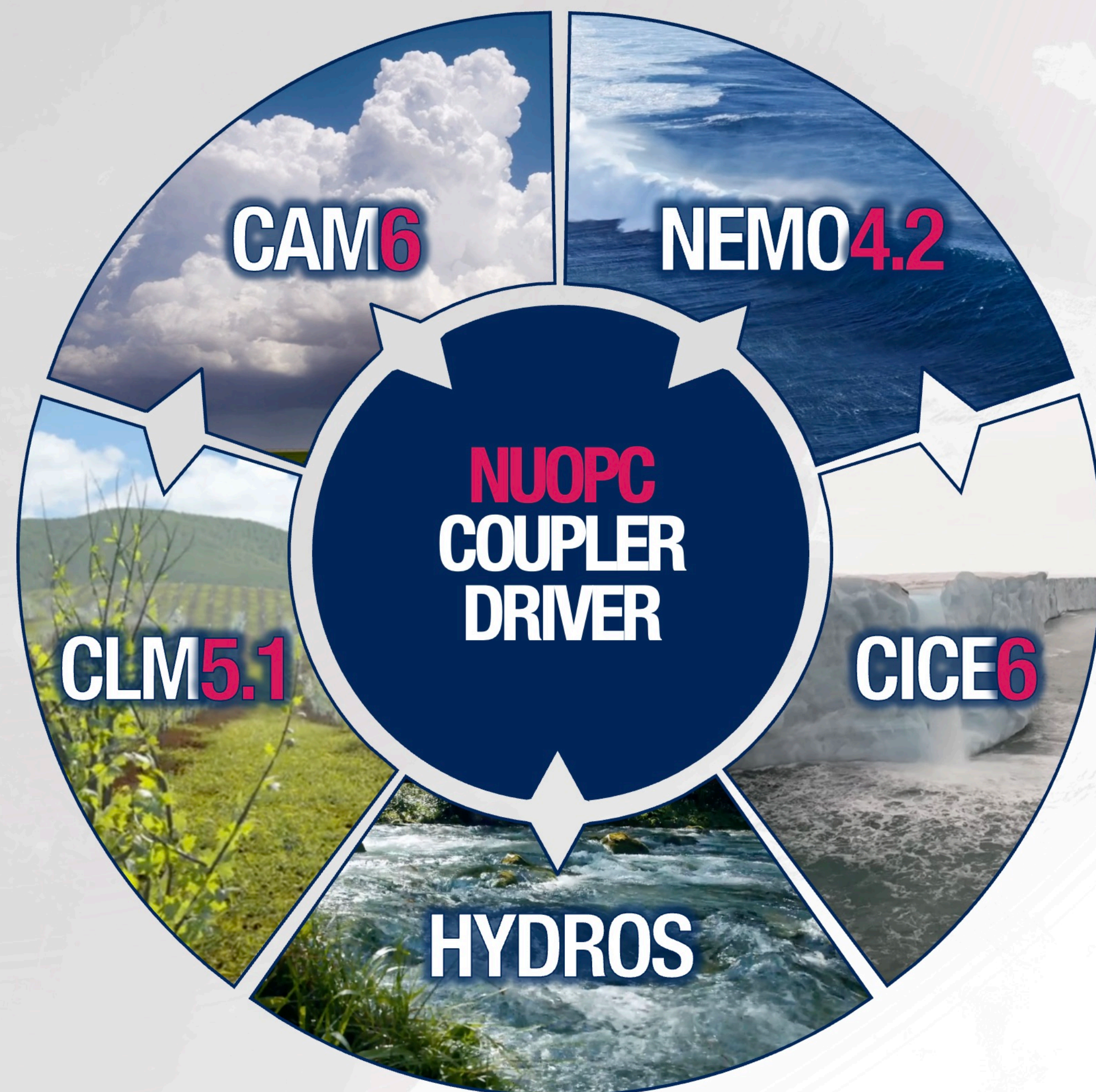
Lawrence *et al.* (2019)

- Land/Surface component of the NCAR's Community Earth System Model (CESM)
- Integrated with CAM6 (coupled): Regular grid of 0.47° (latitude) \times 0.63° (longitude)
30-minute time step
- 20 soil layers (down to 20 m)
- 5 bedrock layers (down to 42 m)
- Snow density parameterisation accounting for a tuning factor to adjust the snow ageing
- Biogeochemical cycles included

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)



National Unified Operational Prediction Capability

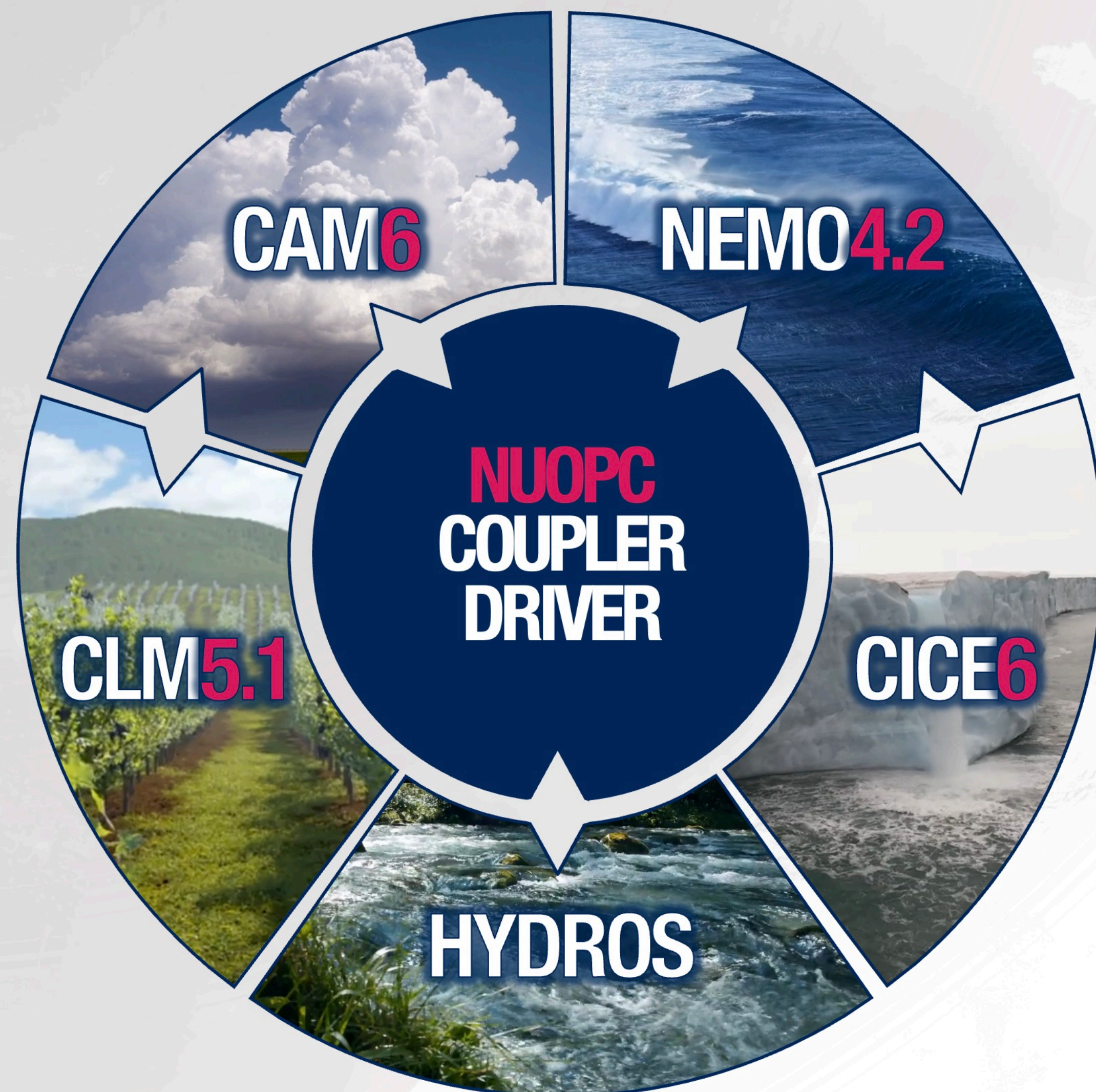
Earth System Modelling Framework

- Coupler component of the NCAR's Community Earth System Model (CESM)
- Atmosphere-Ocean 60 minutes (full time-step of the atmospheric model)
- Atmosphere-Land 30 minutes (internally in CAM6)
- Ocean-Sea Ice 20 minutes (internally in NEMO4.2)
- Land-River 180 minutes (3x full time-step of the atmospheric model)

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)



INITIALISATIONS & ENSEMBLE

Copernicus Climate Change Services (C3s) framework

Perturbed Initial Conditions

$$\begin{array}{ccccc} 10 & \times & 3 & \times & 9 & = & 270 \\ \text{Atmosphere} & & \text{Land} & & \text{Ocean} & & \text{Possibilities} \end{array}$$

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)



INITIALISATIONS & ENSEMBLE

Copernicus Climate Change Services (C3s) framework

270
Possibilities



50
Ensemble Members*

*Hindcast mode (30 in forecast mode).

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)

INITIALISATIONS & ENSEMBLE

Copernicus Climate Change Services (C3s) framework

270
Possibilities



50
Ensemble
Members*

*Hindcast mode (30 in forecast mode).

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

The screenshot shows a web browser window with the URL cds.climate.copernicus.eu/datasets/seasonal-original-single-lev.... The page header includes the European Union flag, 'PROGRAMME OF THE EUROPEAN UNION', 'Copernicus Europe's eyes on Earth', 'Climate Change Service', 'ECMWF', and the user name 'Leonardo Aragão'. The main navigation bar contains 'Climate Data Store', 'Datasets', 'Applications', 'User guide', 'Forum', 'Live', and a 'Your requests' button. The title of the dataset is 'Seasonal forecast daily and subdaily data on single levels'. Below the title are four tabs: 'Overview' (selected), 'Download', 'Quality', and 'Documentation'. The 'Overview' tab contains the following text: 'This entry covers **single-level data** and **soil-level data** at the **original time resolution** (once a day, or once every 6 hours, depending on the variable). Seasonal forecasts provide a long-range outlook of changes in the Earth system over periods of a few weeks or months, as a result of predictable changes in some of the slow-varying components of the system. For example, ocean temperatures typically vary slowly, on timescales of weeks or months; as the ocean has an impact on the overlaying atmosphere, the variability of its properties (e.g. temperature)'. To the right of the text is a circular diagram showing data flow from 'ECMWF', 'Met Office', 'METEO FRANCE', 'DWD', and 'CMCC' to the 'Climate Change Service'. A small penguin icon is visible in the bottom right corner of the page content.

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)

INITIALISATIONS & ENSEMBLE

Copernicus Climate Change Services (C3s) framework

270
Possibilities



50
Ensemble
Members*

*Hindcast mode (30 in forecast mode).

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)

The screenshot shows a web browser window with the URL `cds.climate.copernicus.eu/datasets/seasonal-original-single-lev...`. The page is titled 'Climate Data Store' and features a navigation bar with links to 'Datasets', 'Applications', 'User guide', 'Forum', and 'Live'. A red button labeled 'Your requests' is also present. The main content area is titled 'Data description' and contains a table with the following information:

Data type	Gridded
Projection	Regular latitude-longitude grid
Horizontal coverage	Global
Horizontal resolution	1° x 1°
Temporal coverage	1993 to 2016 (hindcasts); 2017 to present (forecasts)
Temporal resolution	Subdaily (6h) and daily
File format	GRIB
Update frequency	Real-time forecasts are released once per month on the 6th at 12UTC for ECMWF and on the 10th at 12 UTC for the other originating centres.

A small red circular icon with a white penguin is located in the bottom right corner of the page.

EARTH SYSTEM MODELS

CMCC Seasonal Prediction System v4 (v3.5)

INITIALISATIONS & ENSEMBLE

Copernicus Climate Change Services (C3s) framework

270
Possibilities



50
Ensemble
Members*

Seasonal forecast daily and s x +

cds.climate.copernicus.eu/datasets/seasonal-original-single-lev...

Climate Data Store Datasets Applications User guide Forum Live Your requests

Data description

Data type	Gridded
Projection	Regular latitude-longitude grid
Horizontal coverage	Global
Horizontal resolution	1° x 1°
Temporal coverage	1993 to 2016 (hindcasts); 2017 to present (forecasts)
Temporal resolution	Subdaily (6h) and daily
File format	GRIB
Update frequency	Real-time forecasts are released once per month on the 6th at 12UTC for ECMWF and on the 10th at 12 UTC for the other originating centres.

*Hindcast mode (30 in forecast mode).

More information in: Sanna *et al.* (2025) // DOI [10.25424/cmcc-dkcv-fs25](https://doi.org/10.25424/cmcc-dkcv-fs25)



STATISTICAL DOWNSCALING

OUR SCOPE

“The general objective is to develop a computational algorithm in Python language that allows downscaling low-resolution meteorological data (historical or forecast) using machine learning techniques.”

OUR SCOPE

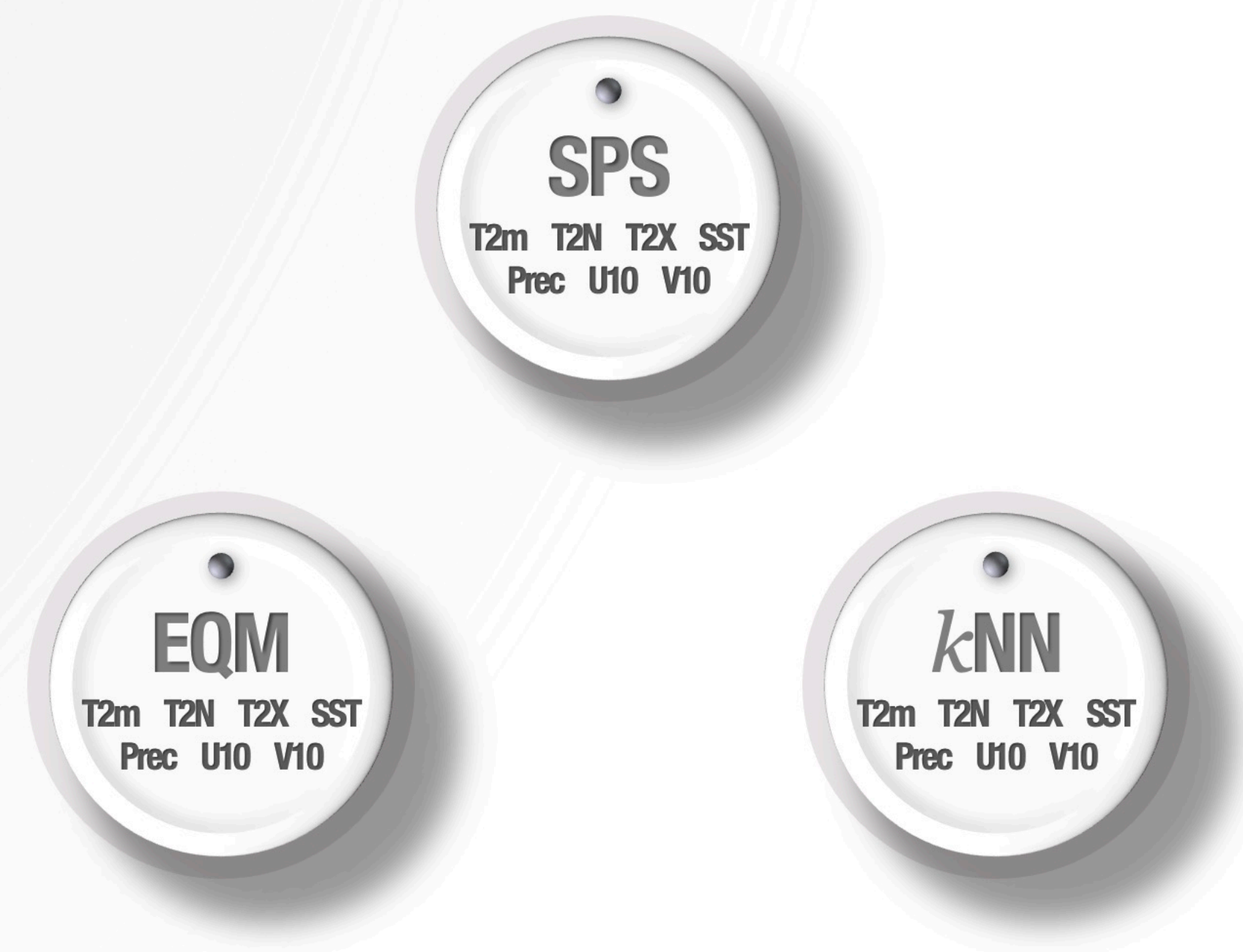
The general objective is to develop a computational algorithm in Python language that allows downscaling low-resolution meteorological data (historical or forecast) using machine learning techniques.

EXPERIMENT

“ To evaluate two different machine learning methods, Empirical Quantile Mapping (EQM) and k-Nearest Neighbours (kNN), in downscaling seven meteorological fields over the Italian Peninsula based on the 24-year hindcasts (1993-2016) of the CMCC Seasonal Prediction System v3.5 (SPS). ”

DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)

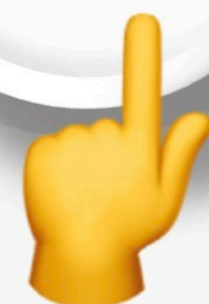
REG
DOM
RES
PER

Global
180x360
1.00° Daily
1993-2016



SPS

T2m T2N T2X SST
Prec U10 V10



ERA5

T2m T2N T2X
SST U10 V10



CHIRPS

Prec



kNN

T2m T2N T2X SST
Prec U10 V10



EQM

T2m T2N T2X SST
Prec U10 V10

- T2m** Daily average 2m Temperature
- T2N** Daily minimum 2m Temperature
- T2X** Daily maximum 2m Temperature
- SST** Daily average Sea Surface Temperature
- Prec** Daily Precipitation (accumulated)
- U10** Daily average zonal wind component
- V10** Daily average meridional wind component

DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)

REG Global
DOM 180x360
RES 1.00° Daily
PER 1993-2016

SPS

T2m T2N T2X SST
Prec U10 V10

REG Global
DOM 720x1440
RES 0.25° Hourly
PER 1940-Now

ERA5

T2m T2N T2X
SST U10 V10

REG 50°N–50°S
DOM 2000x7200
RES 0.05° Daily
PER 1981-Now

CHIRPS

Prec

kNN

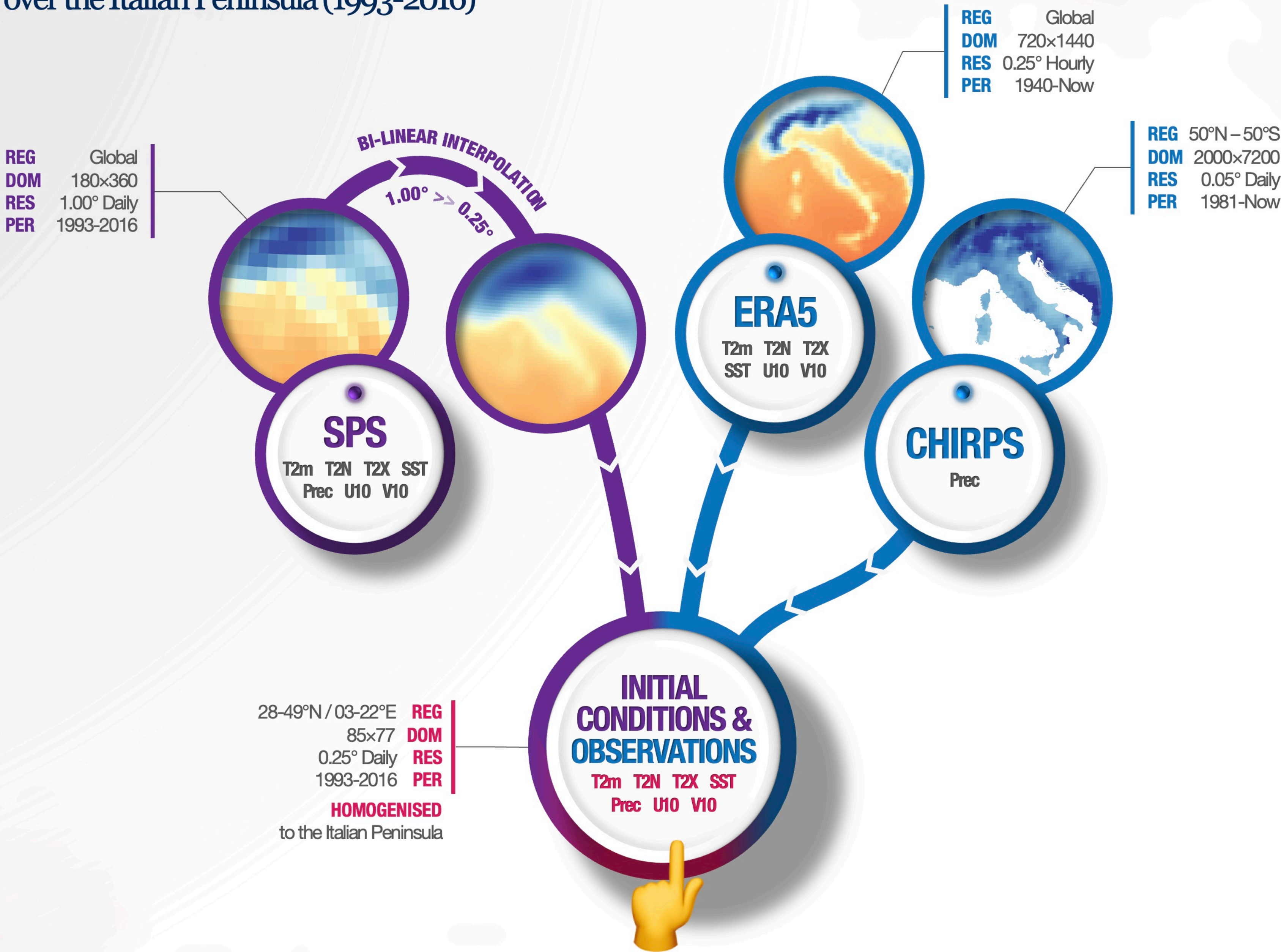
T2m T2N T2X SST
Prec U10 V10

EQM

T2m T2N T2X SST
Prec U10 V10

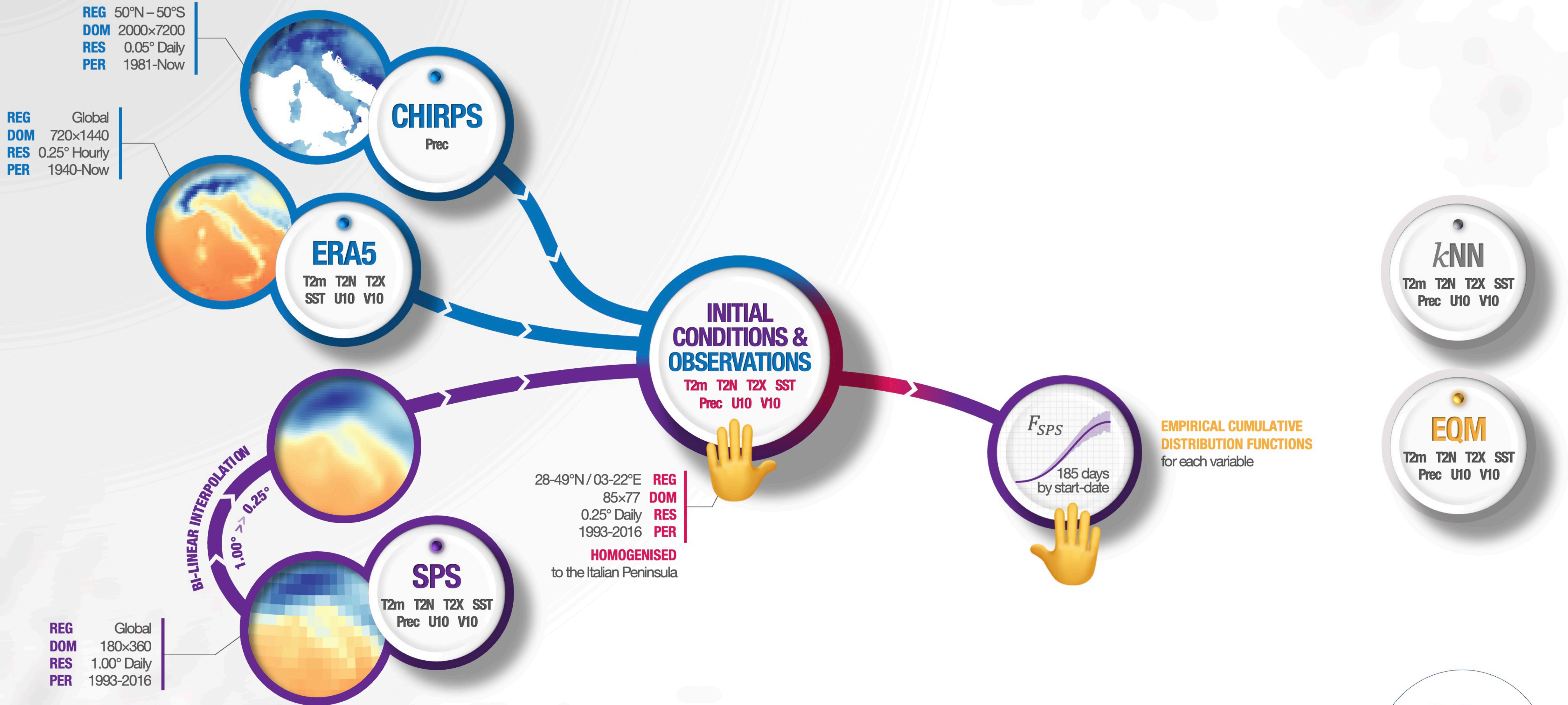
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



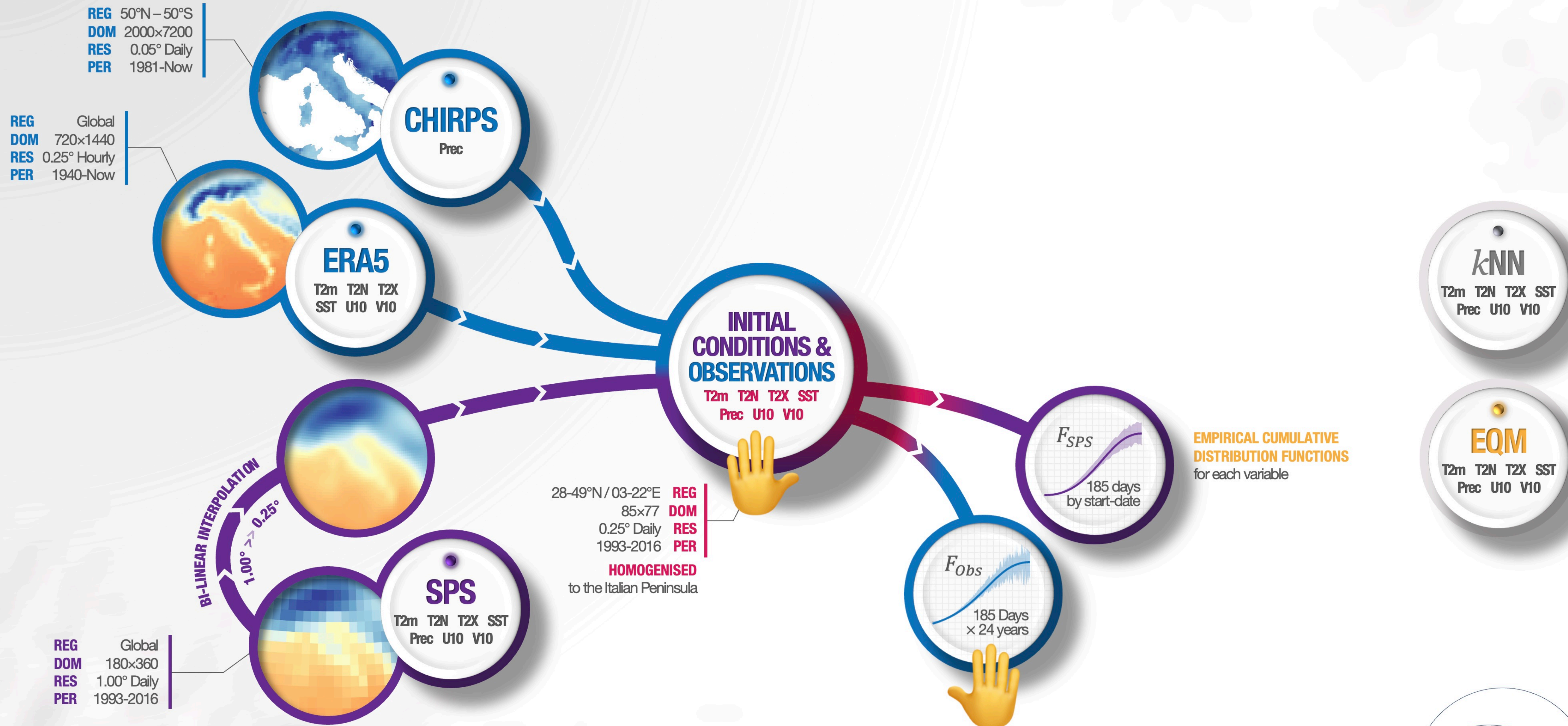
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



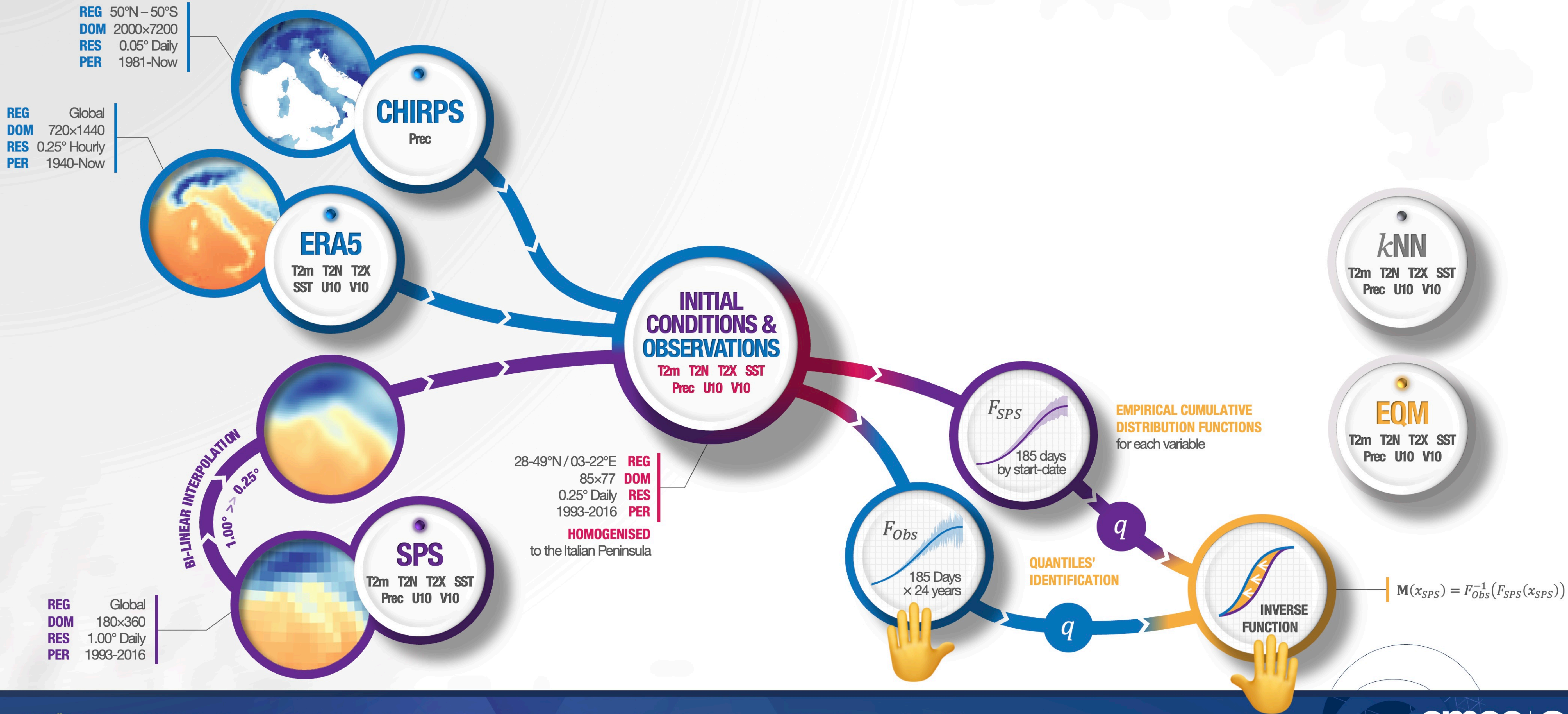
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



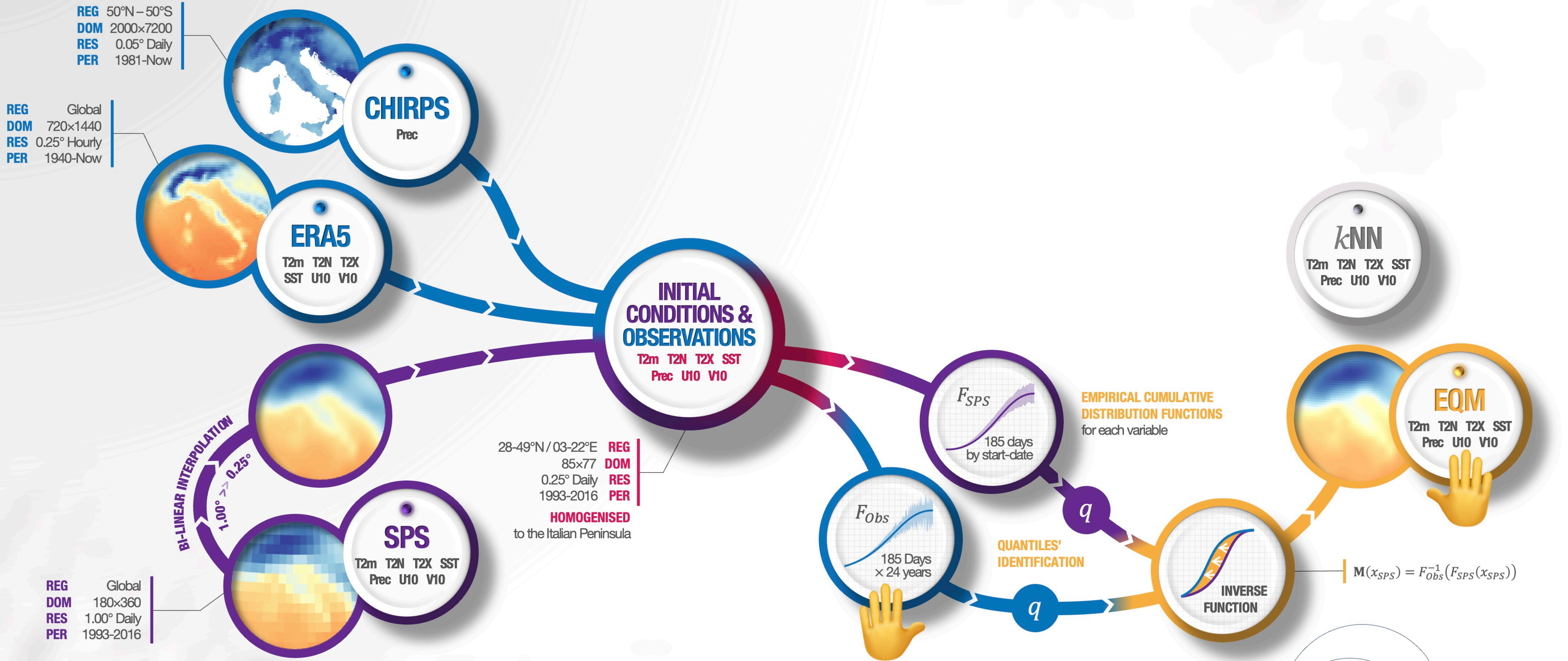
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



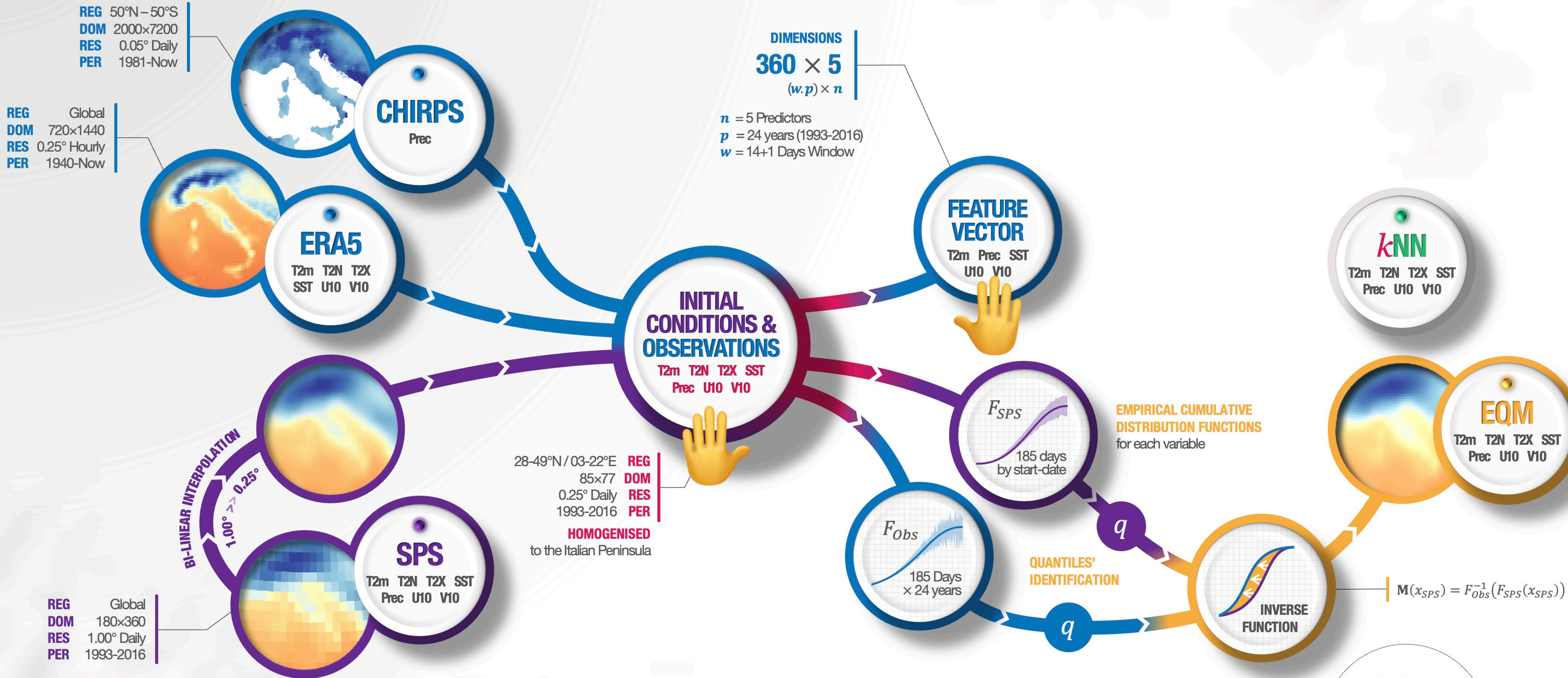
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



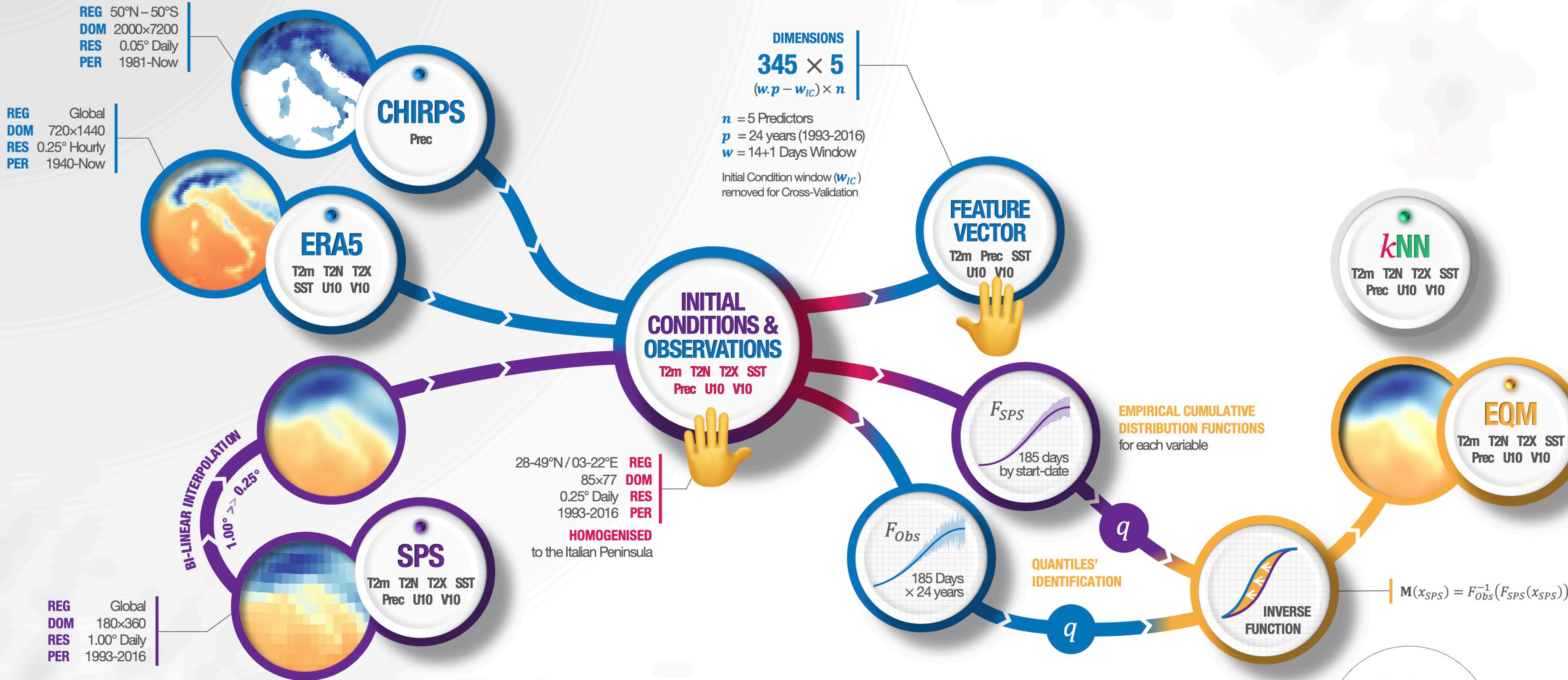
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



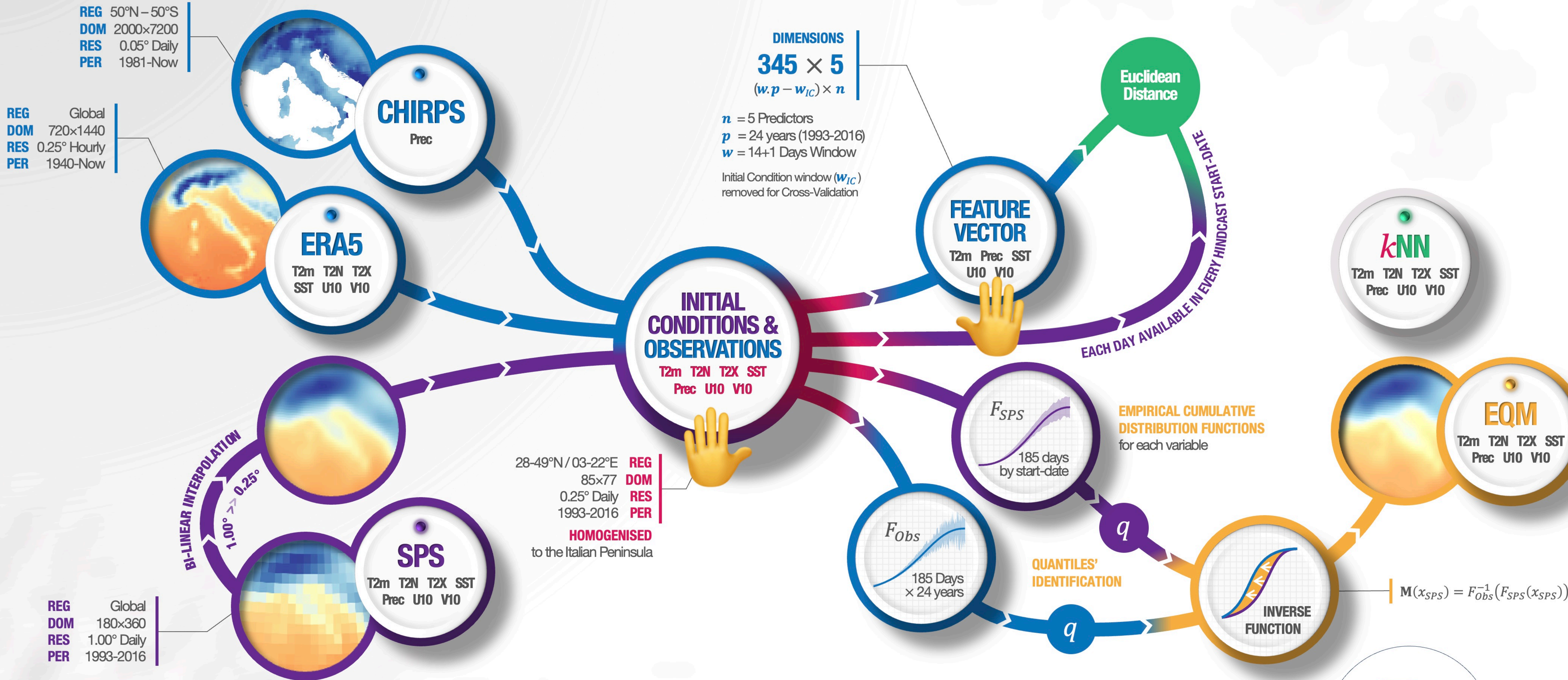
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



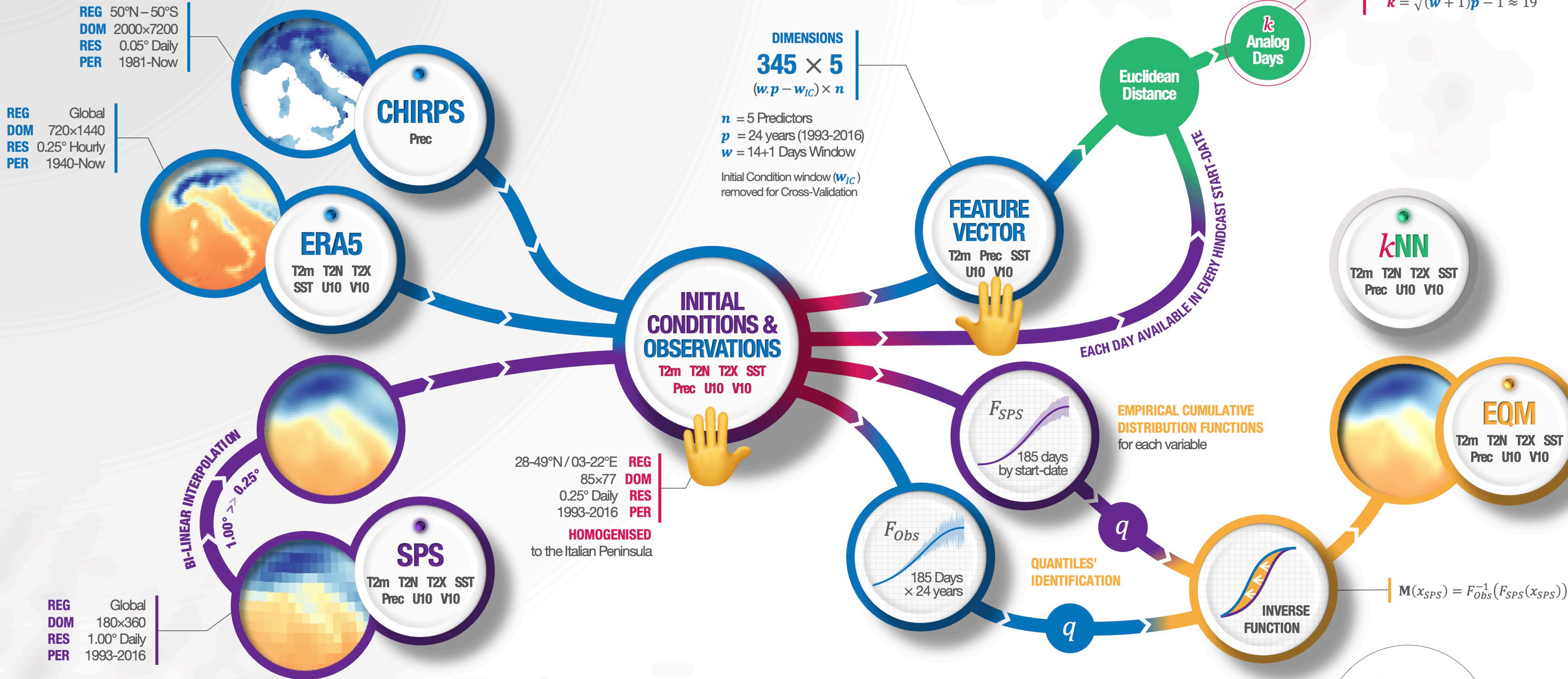
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



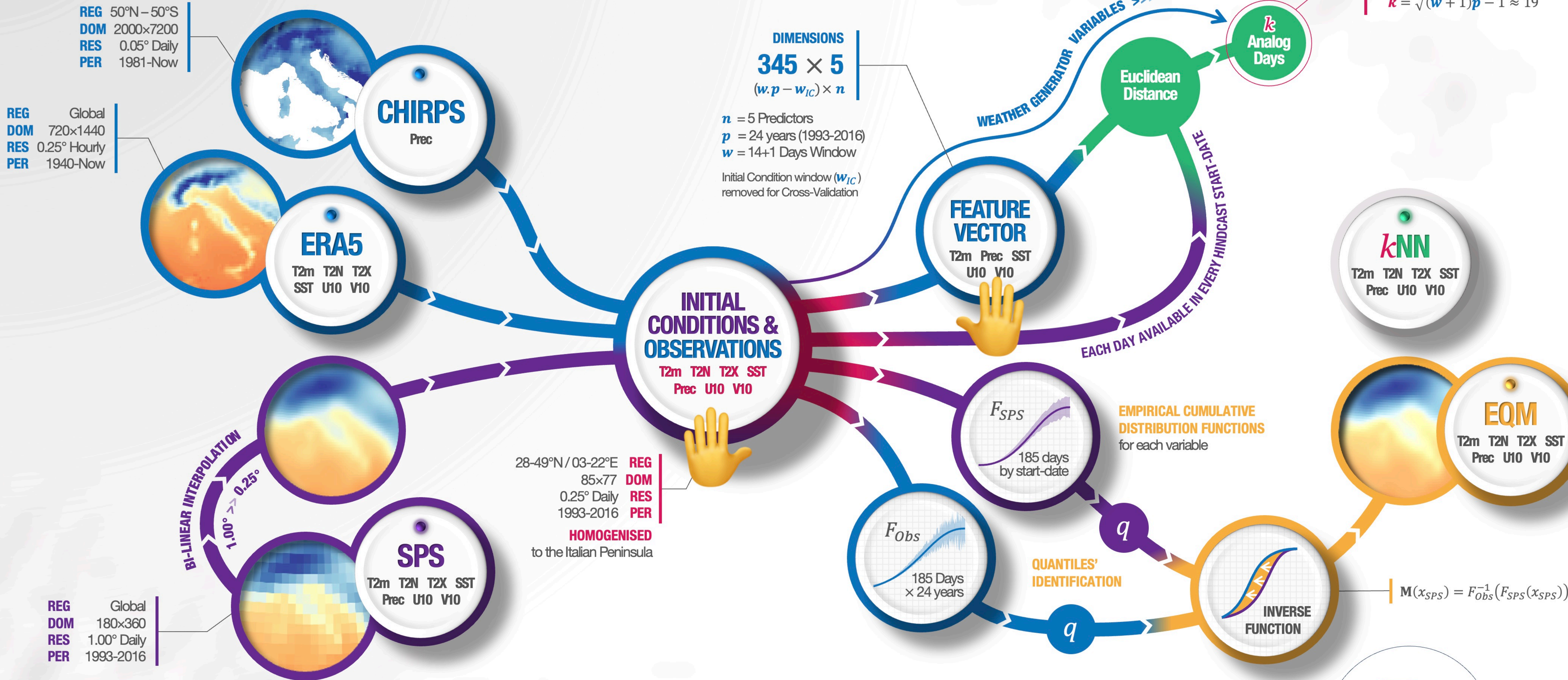
DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



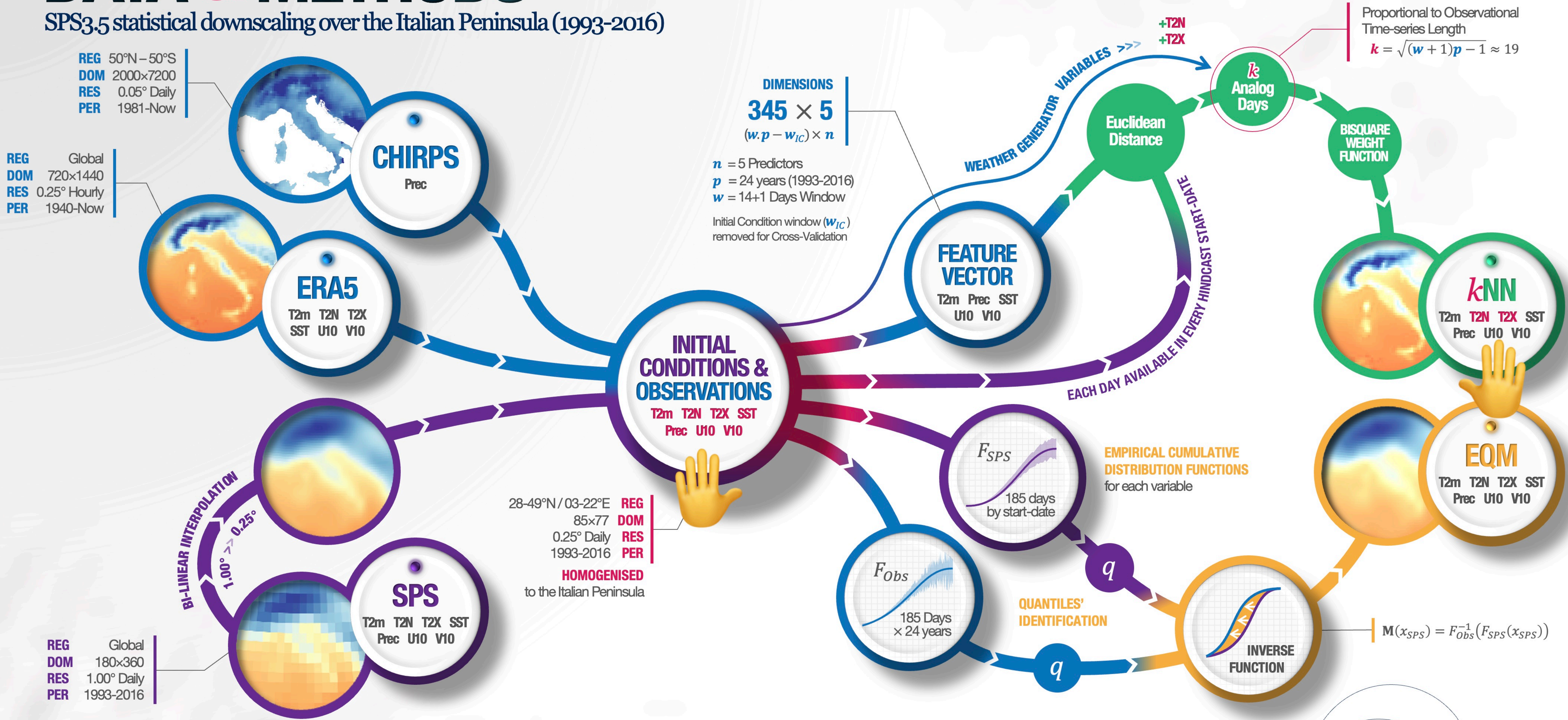
DATA & METHODS


SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)



DATA & METHODS

SPS3.5 statistical downscaling over the Italian Peninsula (1993-2016)

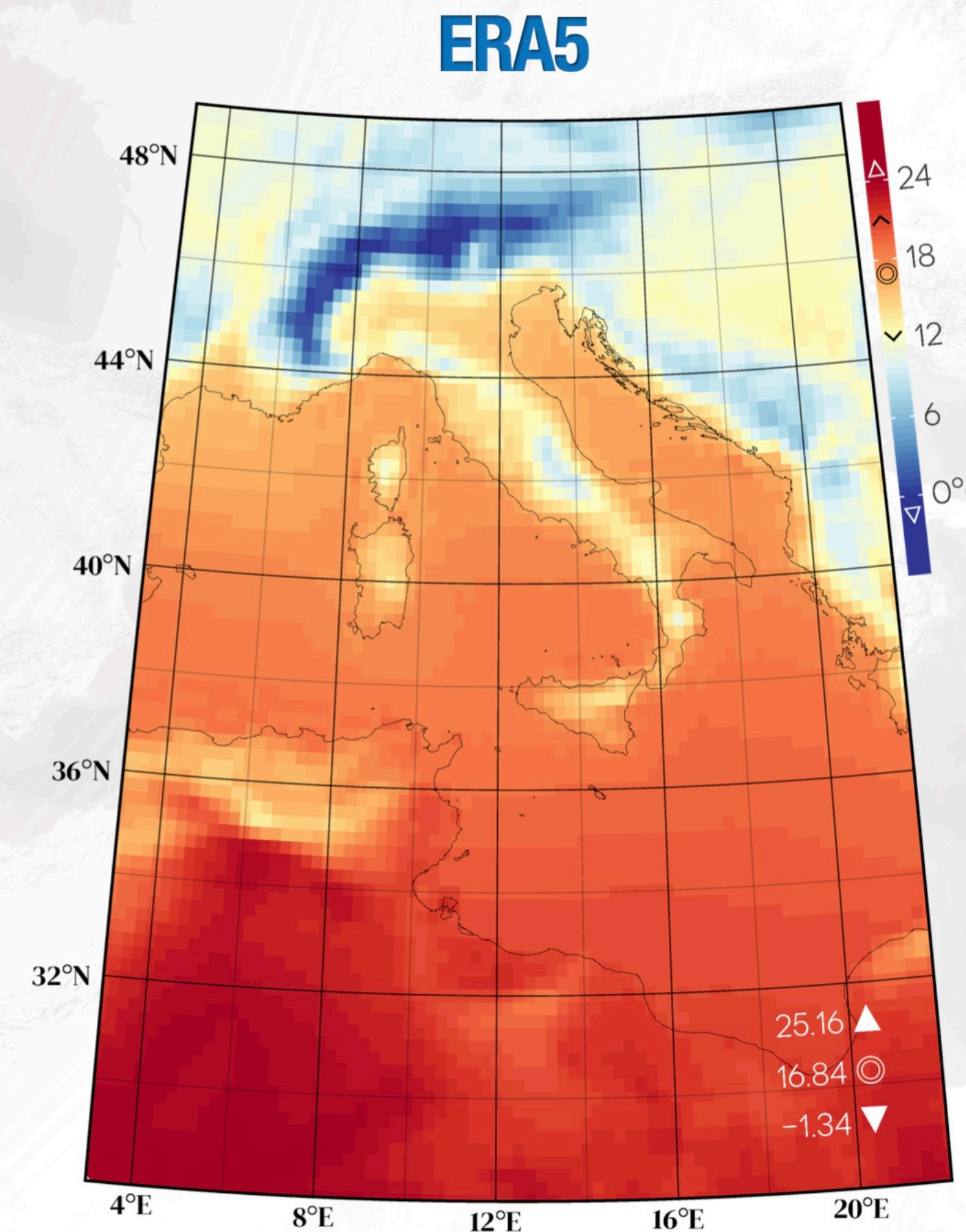




SUMMARISED RESULTS

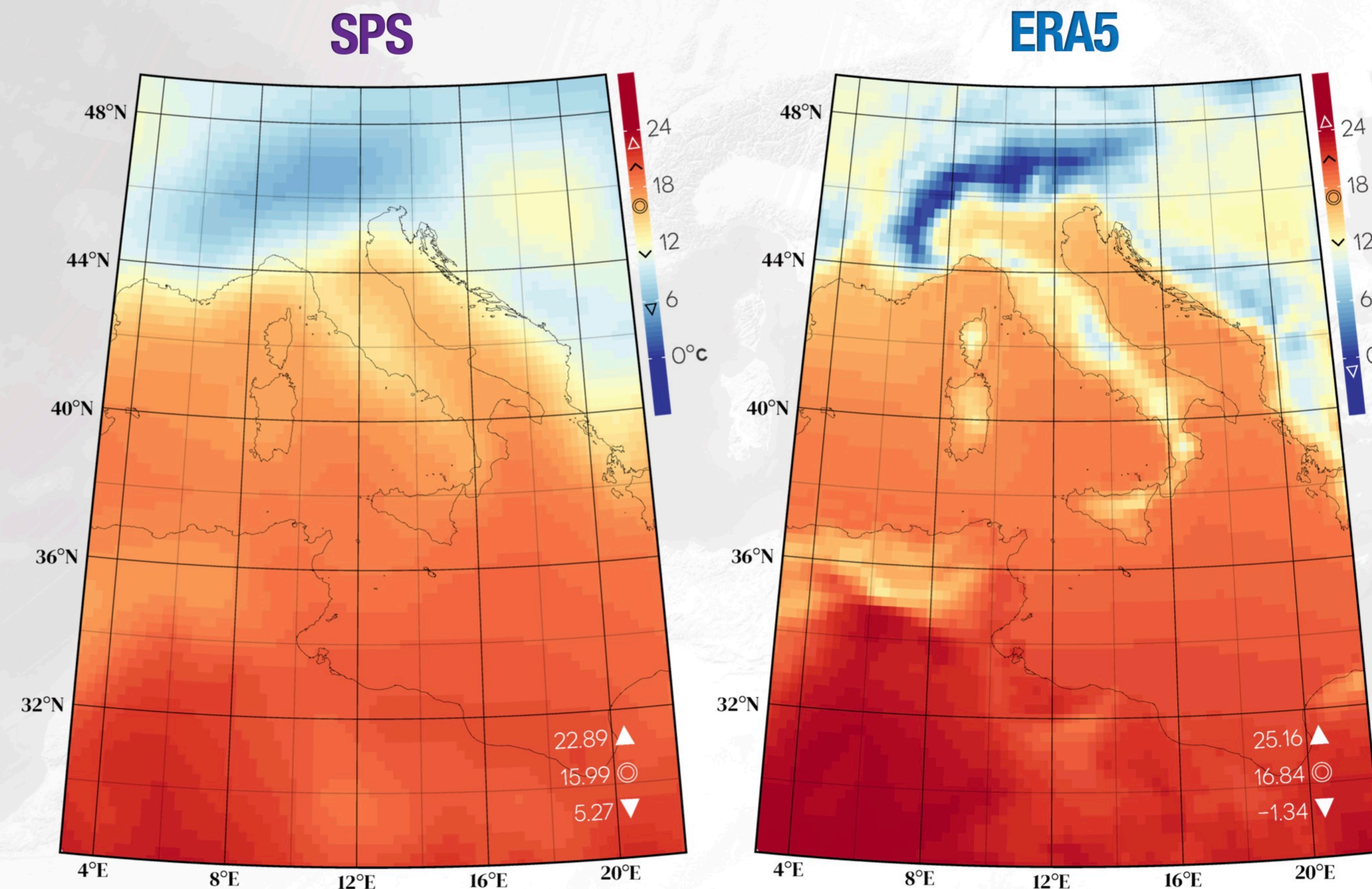
SUMMARISED RESULTS

Ensemble annual mean (lead month 1) for the daily average 2m temperature (1993-2016)



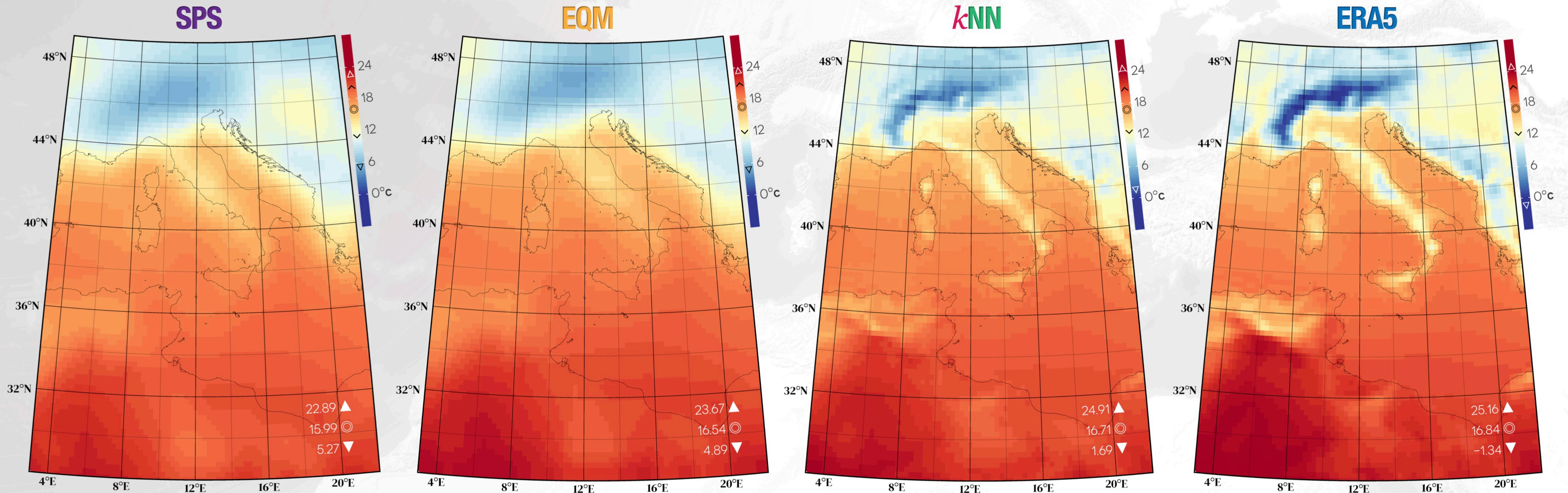
SUMMARISED RESULTS

Ensemble annual mean (lead month 1) for the daily average 2m temperature (1993-2016)



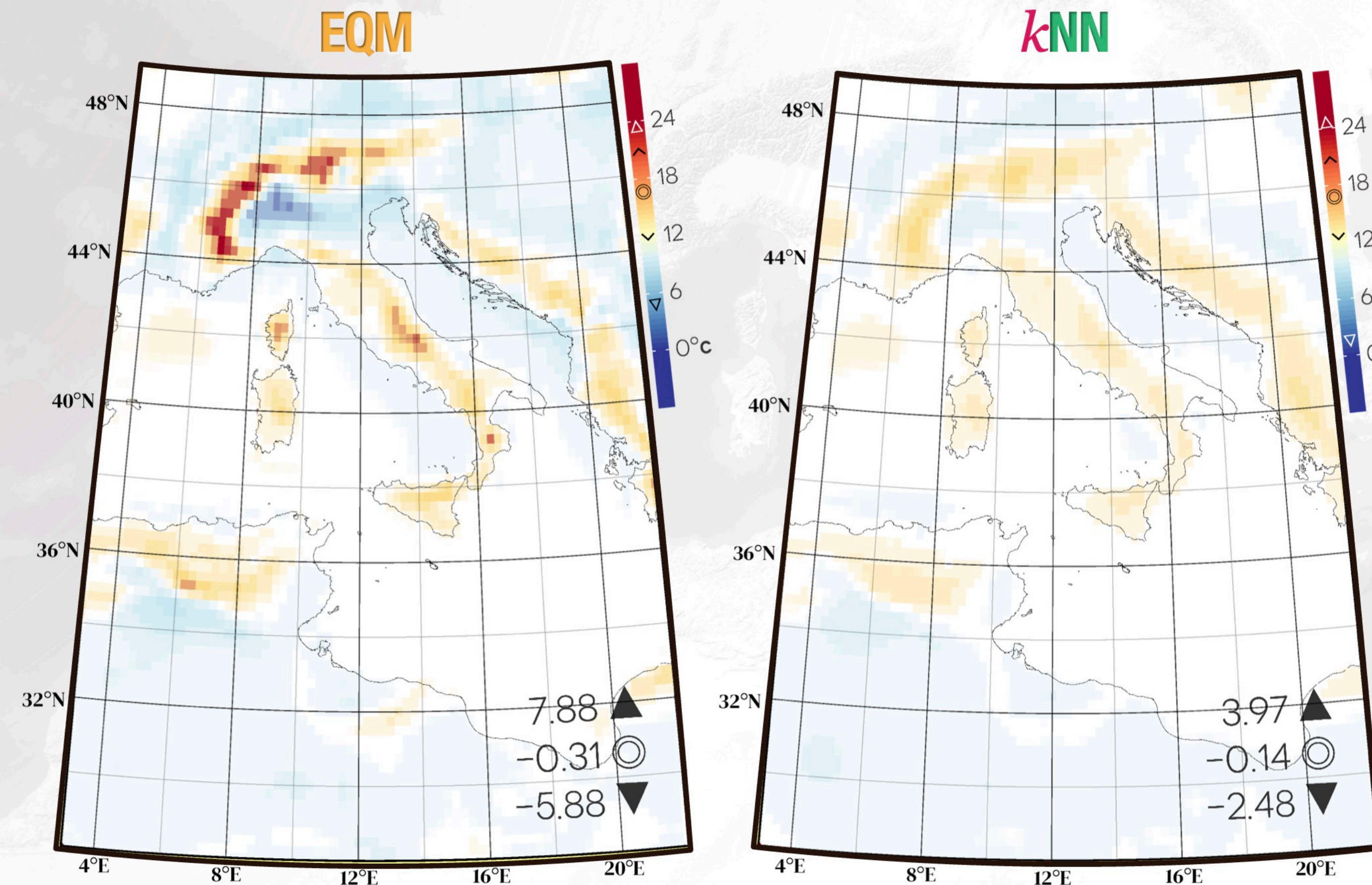
SUMMARISED RESULTS

Ensemble annual mean (lead month 1) for the daily average 2m temperature (1993-2016)



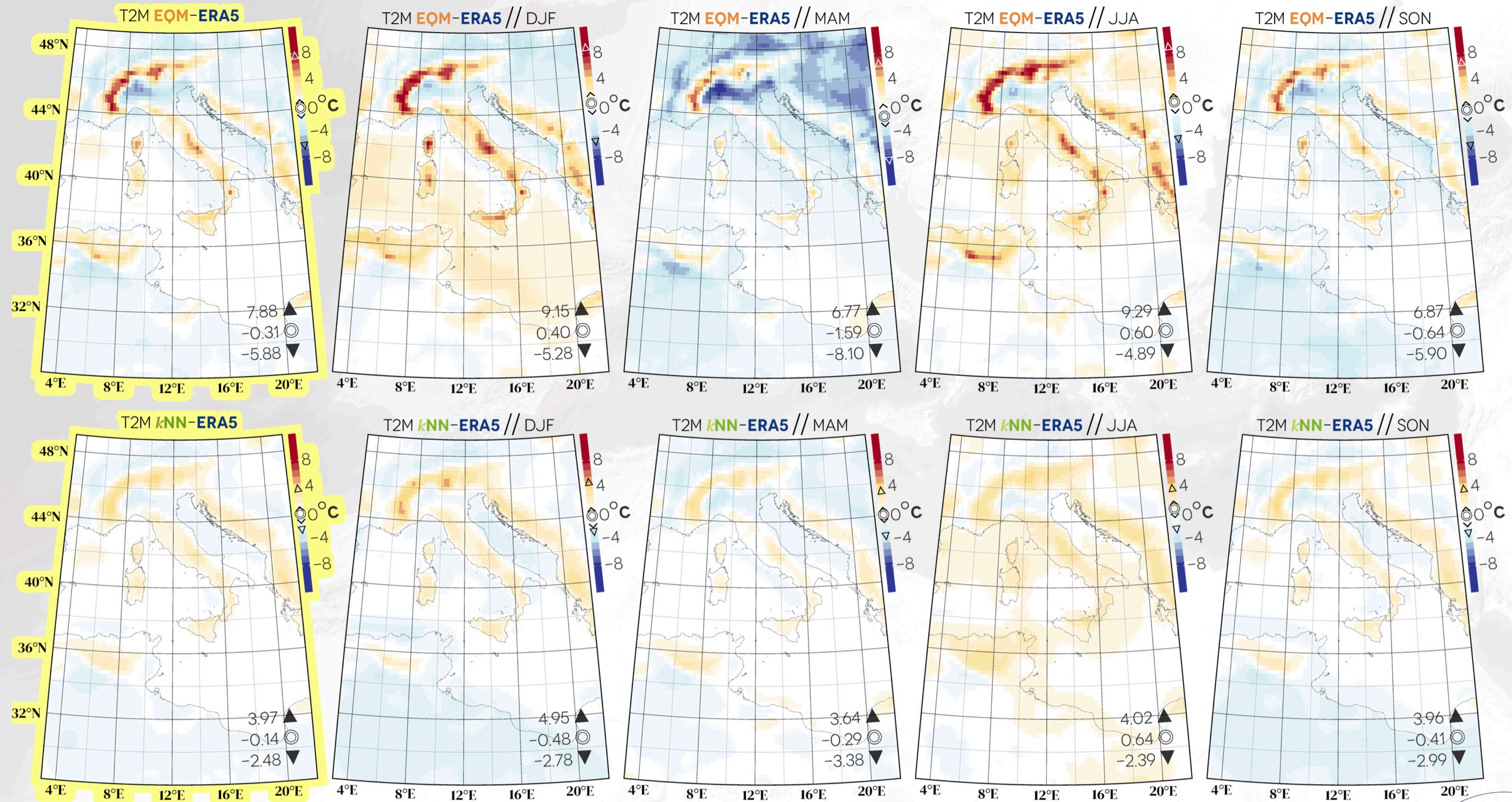
SUMMARISED RESULTS

Ensemble annual mean (lead month 1) for the daily average 2m temperature Bias (1993-2016)



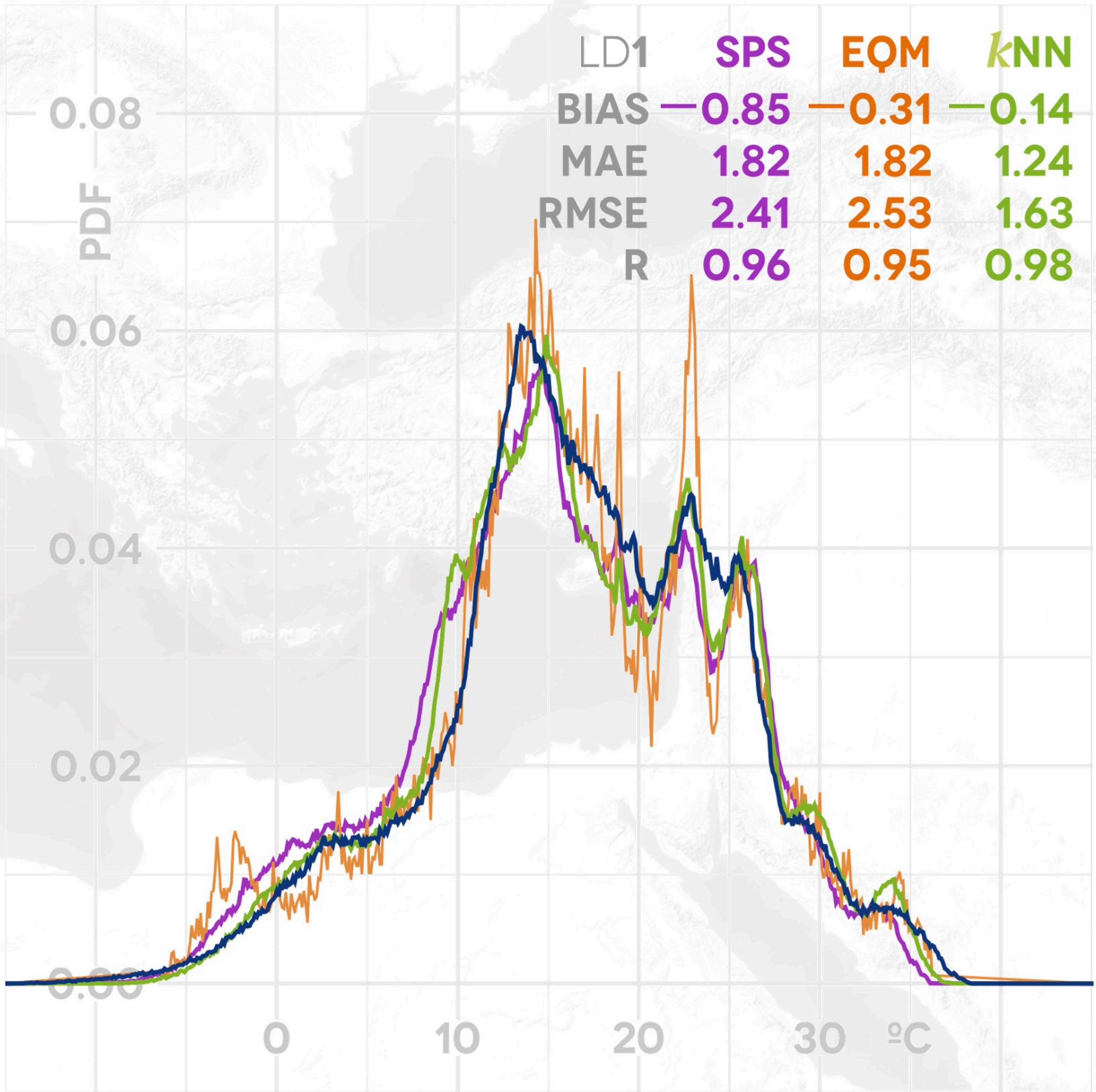
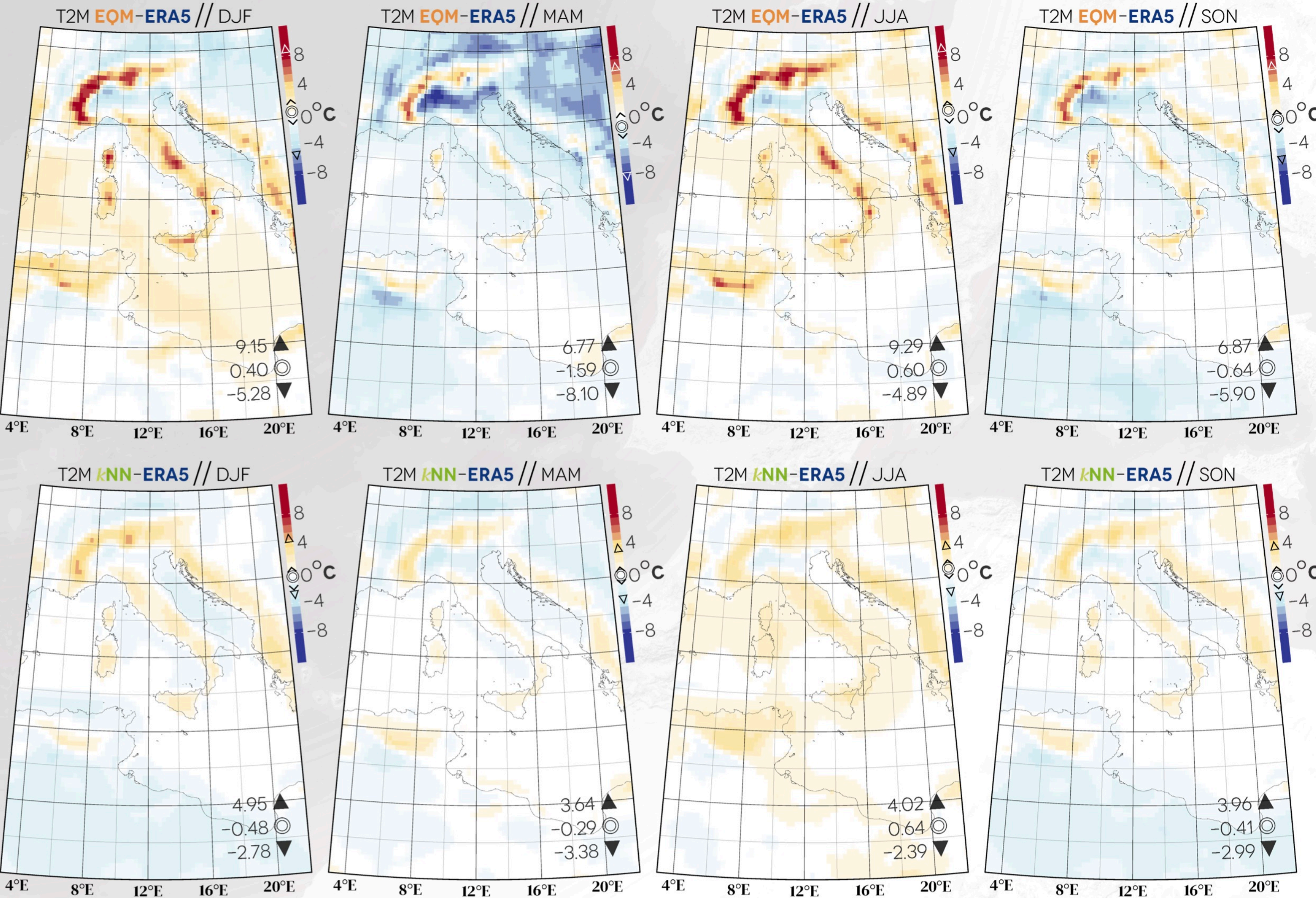
SUMMARISED RESULTS

Ensemble Seasonal mean (lead month 1) for the daily average 2m temperature Bias (1993-2016)



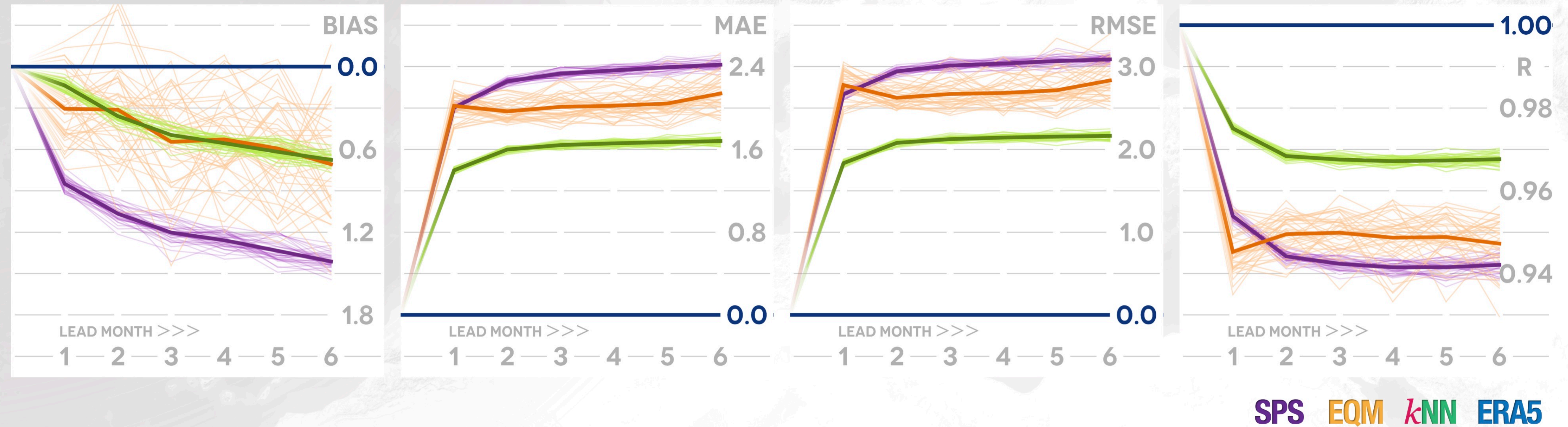
SUMMARISED RESULTS

Ensemble Seasonal mean (lead month 1) for the daily average 2m temperature Bias (1993-2016)



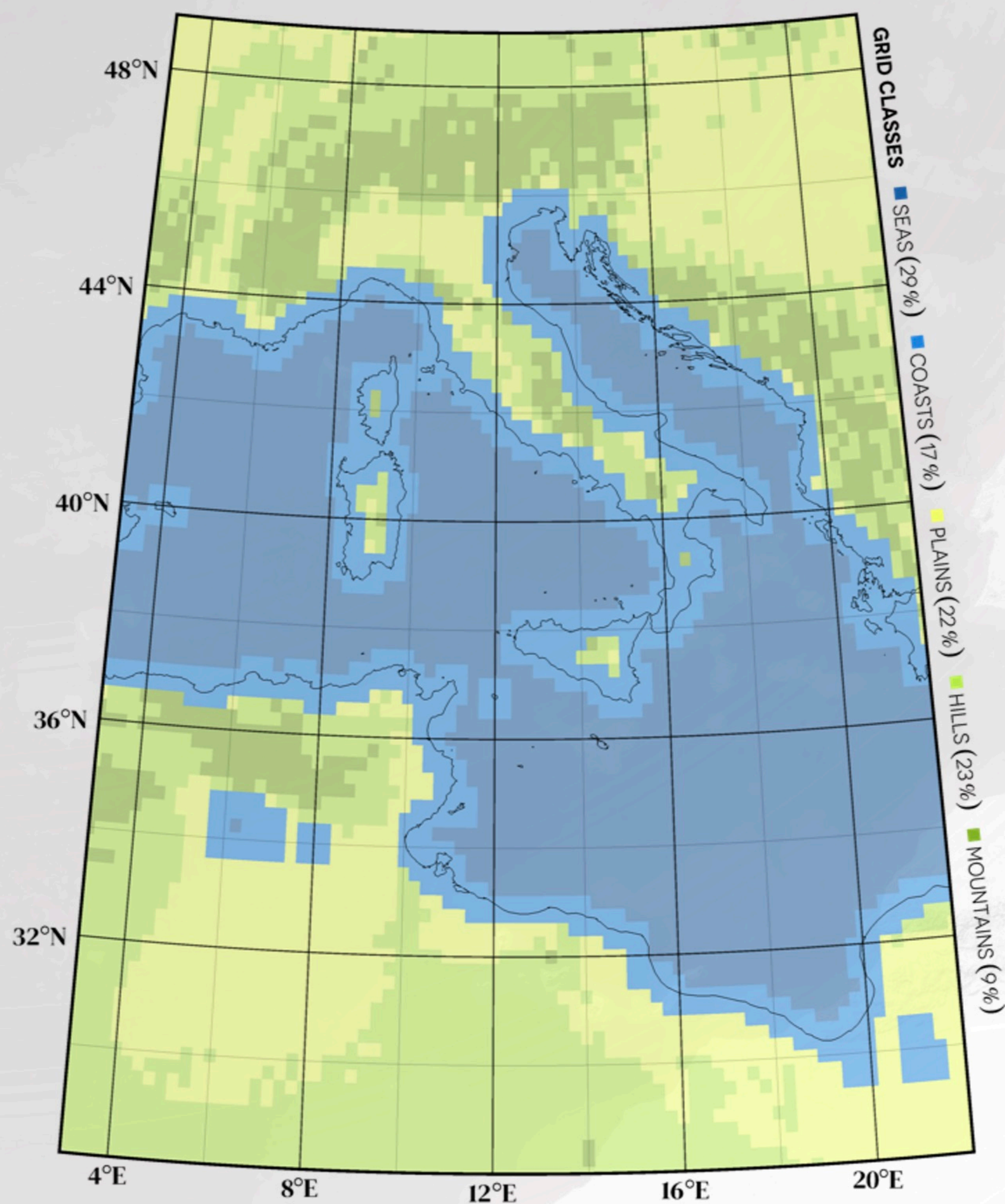
SUMMARISED RESULTS

Statistical indices by lead month of the daily average 2m temperature (1993-2016)



SUMMARISED RESULTS

Statistical indices by environment (lead month 1) of the daily average 2m temperature (1993-2016)



	DOMAIN	SEAS	COASTS	PLAINS	HILLS	MOUNTAINS
Bias	SPS	−0.85	−0.7	−1.12	−1.65	−1.03
	EQM	−0.31	−0.07	−0.25	−1.25	−0.62
	kNN	−0.14	−0.08	0.04	−0.72	−0.41
MAE	SPS	1.82	1.15	2.06	2.12	1.94
	EQM	1.82	1.03	1.67	2.22	2.11
	kNN	1.24	0.69	1.1	1.49	1.52
RMSE	SPS	2.41	1.67	2.74	2.66	2.38
	EQM	2.53	1.36	2.16	3.06	2.75
	kNN	1.63	0.9	1.39	1.84	1.89
R	SPS	0.96	0.97	0.94	0.98	0.98
	EQM	0.95	0.96	0.94	0.96	0.96
	kNN	0.98	0.98	0.98	0.98	0.98

SEAS
Gridpoints at the sea level
(Altitude $h = 0\text{m}$)

COASTS
Gridpoints at the sea level
($h = 0\text{m}$) surrounded by at least
one gridpoint with $h > 0\text{m}$, plus
those surrounded by gridpoints
at the sea level.

PLAINS
Gridpoints with $0\text{m} < h \leq 300\text{m}$

HILLS
Gridpoints with $300\text{m} < h \leq 800\text{m}$

MOUNTAINS
Gridpoints with $h > 800\text{m}$

SUMMARISED RESULTS

Cross-validation of the ensemble mean within the SPS3.5 hindcast period over the Italian domain (1993-2016)

		T2m	Prec	SST	U10	V10	T2N	T2X
Bias	SPS	-0.85	-4.39	-0.44	0.15	-0.04		
	EQM	-0.31	8.76	0.45	-0.06	0.13	-0.44	-0.20
	kNN	-0.14	-20.48	0.06	0.01	-0.06	-0.24	-0.08
MAE	SPS	1.82	19.89	0.78	0.81	0.72		
	EQM	1.82	27.35	0.89	0.86	0.78	1.92	1.95
	kNN	1.24	24.72	0.49	0.73	0.63	1.23	1.33
RMSE	SPS	2.41	33.81	1.15	1.09	0.95		
	EQM	2.53	42.91	1.12	1.17	1.05	2.62	2.74
	kNN	1.63	41.45	0.65	1.01	0.86	1.63	1.79
R	SPS	0.96	0.76	0.97	0.74	0.61		
	EQM	0.95	0.68	0.97	0.69	0.52	0.95	0.95
	kNN	0.98	0.73	0.99	0.77	0.69	0.98	0.98



CONCLUSIONS & NEXT STEPS

CONCLUSION

*The analysis of spatial variability shows that **kNN** produces more detailed and realistic **T2m** fields across Italy than **EQM**. Unlike **EQM**'s smooth and generalised fields, **kNN** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**. Statistically, **kNN** achieves better performance with a 0.98 correlation and 1.63°C RMSE, outperforming **EQM**.*

***T2m** seasonal bias assessment reveals that **EQM** has higher winter biases (up to 8°C in the Alps), while **kNN** shows smoother and more accurate seasonal cycles. **kNN** better reflects **ERA5** data across all seasons, especially in marine regions. **EQM**'s apparent low marine bias is misleading, caused by seasonal over- and underestimations balancing each other.*

*Environmental analysis across Italy shows that **kNN** outperforms **EQM** in all regions, especially in coastal and mountainous areas where **EQM** fails to capture contrasts. Both methods perform better over marine areas, but **kNN** shows lower errors inland. The persistent bias in mountainous zones indicates challenges in representing cold extremes.*

*In seasonal forecasting (up to 6 months), **kNN** shows lower bias and RMSE than **EQM** and **SPS**, preserving initial signals while improving local-scale detail. **EQM**'s bias correction introduces uncertainty due to inter-member spread. **kNN**'s ensemble is more consistent, offering better reliability for climate-sensitive applications.*

*Expanding these analyses to other variables yields a similar scenario. While **kNN** tries to align with **ERA5**, preserving the initial conditions from **SPS**, **EQM** shows greater spread and less physical coherence, abruptly forcing **SPS** towards the observations. Thus, **kNN** better replicates observed variability in spatially structured environments.*

*Precipitation downscaling remains difficult for statistical methods. **EQM** overestimates light rain and misses seasonal variability, especially in cold months and orographic zones. **kNN** better captures spatial patterns in spring and autumn but underrepresents convective summer extremes. These challenges demand more sophisticated inputs and frameworks.*

*Due to analogs averaging, **kNN** struggles with extremes, often underestimating cold days and intense summer precipitation, especially in mountainous regions. Minor errors in analogs selection lead to significant rainfall underestimations. Hybrid methods and convective indices could enhance **kNN**'s ability to capture such extremes.*

*When used as a weather generator, **kNN** accurately reproduces daily temperature extremes (**T2N** and **T2X**), aligning closely with **ERA5** in terms of spread and skewness, unlike **EQM**. These results confirm the value of **kNN**'s multivariate coherence and spatial sensitivity. However, improving its treatment of extremes remains key for broader applicability.*

CONCLUSION

The analysis of spatial variability shows that **ERA5** produces more detailed and realistic TSS fields across Italy than **EQM**. **ERA5** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**. Statistically, **ERA5** achieves better performance with a skill correlation and a 40% RMSE, outperforming **EQM**.

The seasonal bias assessment reveals that **EQM** has higher winter bias (up to 4°C in the Alps), while **ERA5** better captures **ERA5** in spring regions. **EQM**'s apparent low winter bias is an over- and underestimation balancing.

Environmental analysis across Italy shows **ERA5** captures extremes better in all regions, especially in coastal and mountainous areas. Both methods struggle to capture extremes. The persistent bias in reconstruction areas indicates challenges in representing cold extremes.

In seasonal forecasting (up to 6 months), **ERA5** shows lower bias and RMSE than **EQM** and **ERA5**, preserving initial signals while improving local scale detail. **EQM**'s bias correction introduces uncertainty due to later number spread. **ERA5**'s ensemble is more consistent, offering better reliability for climate-sensitive applications.

Expanding these analyses to other variables yields a similar scenario: while **ERA5** tries to align with **ERA5**, preserving the initial conditions from **ERA5**, **EQM** shows greater spread and less physical coherence, diverging further from the observations. Thus, **ERA5** better replicates observed variability in spatially structured environments.

Precipitation forecasting remains difficult for statistical methods. **ERA5** better captures spatial patterns in spring, while **EQM** shows excessive spread extremes. These findings highlight the need for improved inputs and frameworks.

ERA5 struggles with extremes, often underestimating winter extremes and overestimating summer extremes, especially in coastal areas. These errors in model selection lead to significant model underestimations. Hybrid methods and ensemble methods could enhance **ERA5**'s ability to capture such extremes.

When used as a weather generator, **ERA5** accurately reproduces daily temperature extremes (Tmax and Tmin), aligning closely with **ERA5** in terms of spread and skewness, unlike **EQM**. These results confirm the value of **ERA5** multivariate coherence and spatial consistency. However, improving its treatment of extremes remains key for broader applicability.

“Unlike **EQM**'s smooth and generalised fields, **kNN** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**.”

CONCLUSION

*The analysis of spatial variability shows that **kNN** produces more detailed and realistic **T2m** fields across Italy than **EQM**. Unlike **EQM**'s smooth and generalised fields, **kNN** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**. Statistically, **kNN** achieves better performance with a 0.98 correlation and 1.63°C RMSE, outperforming **EQM**.*

***T2m** seasonal bias assessment reveals that **EQM** has higher winter biases (up to 8°C in the Alps), while **kNN** shows smoother and more accurate seasonal cycles. **kNN** better reflects **ERA5** data across all seasons, especially in marine regions. **EQM**'s apparent low marine bias is misleading, caused by seasonal over- and underestimations balancing each other.*

*Environmental analysis across Italy shows that **kNN** outperforms **EQM** in all regions, especially in coastal and mountainous areas where **EQM** fails to capture contrasts. Both methods perform better over marine areas, but **kNN** shows lower errors inland. The persistent bias in mountainous zones indicates challenges in representing cold extremes.*

*In seasonal forecasting (up to 6 months), **kNN** shows lower bias and RMSE than **EQM** and **SPS**, preserving initial signals while improving local-scale detail. **EQM**'s bias correction introduces uncertainty due to inter-member spread. **kNN**'s ensemble is more consistent, offering better reliability for climate-sensitive applications.*

*Expanding these analyses to other variables yields a similar scenario. While **kNN** tries to align with **ERA5**, preserving the initial conditions from **SPS**, **EQM** shows greater spread and less physical coherence, abruptly forcing **SPS** towards the observations. Thus, **kNN** better replicates observed variability in spatially structured environments.*

*Precipitation downscaling remains difficult for statistical methods. **EQM** overestimates light rain and misses seasonal variability, especially in cold months and orographic zones. **kNN** better captures spatial patterns in spring and autumn but underrepresents convective summer extremes. These challenges demand more sophisticated inputs and frameworks.*

*Due to analogs averaging, **kNN** struggles with extremes, often underestimating cold days and intense summer precipitation, especially in mountainous regions. Minor errors in analogs selection lead to significant rainfall underestimations. Hybrid methods and convective indices could enhance **kNN**'s ability to capture such extremes.*

*When used as a weather generator, **kNN** accurately reproduces daily temperature extremes (**T2N** and **T2X**), aligning closely with **ERA5** in terms of spread and skewness, unlike **EQM**. These results confirm the value of **kNN**'s multivariate coherence and spatial sensitivity. However, improving its treatment of extremes remains key for broader applicability.*

CONCLUSION

*The analysis of spatial variability shows that **kNN** produces more detailed and realistic **T2m** fields across Italy than **EQM**. Unlike **EQM**'s smooth and generalised fields, **kNN** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**. Statistically, **kNN** achieves better performance with a 0.98 correlation and 1.63°C RMSE, outperforming **EQM**.*

***T2m** seasonal bias assessment reveals that **EQM** has higher winter biases (up to 8°C in the Alps), while **kNN** shows smoother and more accurate seasonal cycles. **kNN** better reflects **ERA5** data across all seasons, especially in marine regions. **EQM**'s apparent low marine bias is misleading, caused by seasonal over- and underestimations balancing each other.*

*Environmental analysis across Italy shows that **kNN** outperforms **EQM** in all regions, especially in coastal and mountainous areas where **EQM** fails to capture contrasts. Both methods perform better over marine areas, but **kNN** shows lower errors inland. The persistent bias in mountainous zones indicates challenges in representing cold extremes.*

*In seasonal forecasting (up to 6 months), **kNN** shows lower bias and RMSE than **EQM** and **SPS**, preserving initial signals while improving local-scale detail. **EQM**'s bias correction introduces uncertainty due to inter-member spread. **kNN**'s ensemble is more consistent, offering better reliability for climate-sensitive applications.*

*Expanding these analyses to other variables yields a similar scenario. While **kNN** tries to align with **ERA5**, preserving the initial conditions from **SPS**, **EQM** shows greater spread and less physical coherence, abruptly forcing **SPS** towards the observations. Thus, **kNN** better replicates observed variability in spatially structured environments.*

*Precipitation downscaling remains difficult for statistical methods. **EQM** overestimates light rain and misses seasonal variability, especially in cold months and orographic zones. **kNN** better captures spatial patterns in spring and autumn but underrepresents convective summer extremes. These challenges demand more sophisticated inputs and frameworks.*

*Due to analogs averaging, **kNN** struggles with extremes, often underestimating cold days and intense summer precipitation, especially in mountainous regions. Minor errors in analogs selection lead to significant rainfall underestimations. Hybrid methods and convective indices could enhance **kNN**'s ability to capture such extremes.*

*When used as a weather generator, **kNN** accurately reproduces daily temperature extremes (**T2N** and **T2X**), aligning closely with **ERA5** in terms of spread and skewness, unlike **EQM**. These results confirm the value of **kNN**'s multivariate coherence and spatial sensitivity. However, improving its treatment of extremes remains key for broader applicability.*

CONCLUSION

The analysis of spatial variability shows that **EQM** produces more detailed and realistic TSS fields across Italy than **ORF**, **ORF** (**ORF**) smooths and generalised fields. **EQM** captures key geographic features like the Po Valley and the Alps, aligning closely with **ORF**. Statistically, **EQM** achieves better performance with a correlation of 0.82 and a RMSE of 0.0002, outperforming **ORF**.

The seasonal bias assessment reveals that **ORF** has higher winter bias (up to 40% in some regions) than **EQM**, which shows lower errors across all seasons. **EQM** captures extremes better, especially in cold months and orographic zones.

Extremes analysis shows that **EQM** captures extremes better than **ORF**, which shows lower errors in cold months. The persistent bias in reconstruction areas indicates challenges in representing cold extremes.

In seasonal forecasting (up to 6 months), **EQM** shows lower bias and RMSE than **ORF** and **ORF**, preserving initial signals while improving local scale detail. **ORF**'s bias correction introduces uncertainty due to later number spread. **EQM**'s ensemble is more consistent, offering better reliability for climate-sensitive applications.

Expanding these analyses to other variables yields a similar scenario. While **EQM** tries to align with **ORF**, preserving the initial conditions from **ORF**, **ORF** shows greater spread and less physical coherence, diverging from **ORF** towards the observations. Thus, **EQM** better replicates observed variability in spatially structured environments.

Precipitation forecasting remains difficult for statistical methods. **EQM** shows lower errors than **ORF** and **ORF**, capturing extremes better. **EQM** captures extremes better than **ORF**, which shows lower errors in cold months. The persistent bias in reconstruction areas indicates challenges in representing cold extremes.

Extremes analysis shows that **EQM** captures extremes better than **ORF**, which shows lower errors in cold months. The persistent bias in reconstruction areas indicates challenges in representing cold extremes.

When used as a weather generator, **EQM** accurately reproduces daily temperature extremes (**ORF** and **ORF**), aligning closely with **ORF** in terms of spread and skewness, unlike **ORF**. These results confirm the value of **EQM** in multivariate coherence and spatial consistency. However, improving its treatment of extremes remains key for broader applicability.

“**EQM** overestimates light rain and misses seasonal variability, especially in cold months and orographic zones. **kNN** better captures spatial patterns in spring and autumn but underrepresents convective summer extremes.”

CONCLUSION

*The analysis of spatial variability shows that **kNN** produces more detailed and realistic **T2m** fields across Italy than **EQM**. Unlike **EQM**'s smooth and generalised fields, **kNN** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**. Statistically, **kNN** achieves better performance with a 0.98 correlation and 1.63°C RMSE, outperforming **EQM**.*

***T2m** seasonal bias assessment reveals that **EQM** has higher winter biases (up to 8°C in the Alps), while **kNN** shows smoother and more accurate seasonal cycles. **kNN** better reflects **ERA5** data across all seasons, especially in marine regions. **EQM**'s apparent low marine bias is misleading, caused by seasonal over- and underestimations balancing each other.*

*Environmental analysis across Italy shows that **kNN** outperforms **EQM** in all regions, especially in coastal and mountainous areas where **EQM** fails to capture contrasts. Both methods perform better over marine areas, but **kNN** shows lower errors inland. The persistent bias in mountainous zones indicates challenges in representing cold extremes.*

*In seasonal forecasting (up to 6 months), **kNN** shows lower bias and RMSE than **EQM** and **SPS**, preserving initial signals while improving local-scale detail. **EQM**'s bias correction introduces uncertainty due to inter-member spread. **kNN**'s ensemble is more consistent, offering better reliability for climate-sensitive applications.*

*Expanding these analyses to other variables yields a similar scenario. While **kNN** tries to align with **ERA5**, preserving the initial conditions from **SPS**, **EQM** shows greater spread and less physical coherence, abruptly forcing **SPS** towards the observations. Thus, **kNN** better replicates observed variability in spatially structured environments.*

*Precipitation downscaling remains difficult for statistical methods. **EQM** overestimates light rain and misses seasonal variability, especially in cold months and orographic zones. **kNN** better captures spatial patterns in spring and autumn but underrepresents convective summer extremes. These challenges demand more sophisticated inputs and frameworks.*

*Due to analogs averaging, **kNN** struggles with extremes, often underestimating cold days and intense summer precipitation, especially in mountainous regions. Minor errors in analogs selection lead to significant rainfall underestimations. Hybrid methods and convective indices could enhance **kNN**'s ability to capture such extremes.*

*When used as a weather generator, **kNN** accurately reproduces daily temperature extremes (**T2N** and **T2X**), aligning closely with **ERA5** in terms of spread and skewness, unlike **EQM**. These results confirm the value of **kNN**'s multivariate coherence and spatial sensitivity. However, improving its treatment of extremes remains key for broader applicability.*

CONCLUSION

The analysis of spatial variability shows that **ECMWF** produces more detailed and realistic TSS fields across Italy than **ERA5** (better **ERA5**) datasets and generalized fields. **ECMWF** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**. Statistically, **ECMWF** achieves better performance with a correlation and a R^2 of 0.88, outperforming **ERA5**.

The seasonal bias assessment reveals that **ERA5** has higher winter biases (up to 8°C in the Alps), while **ECMWF** shows smoother and more accurate extremes (better **ERA5**). **ECMWF** better captures **ERA5** extremes, **ERA5** appears less accurate in capturing extremes and underestimates extremes.

Environmental analysis across all regions, especially in coastal and mountainous areas, shows lower errors inland. The persistent bias in mountainous areas indicates challenges in representing cold extremes.

In seasonal forecasting (up to 6 months), **ECMWF** shows lower bias and R^2 than **ERA5** and **ERA5**, preserving initial signals while improving local scale detail. **ERA5** has correction introducing uncertainty due to later number spread. **ECMWF** ensemble is more consistent, offering better reliability for climate sensitive applications.

Expanding these analyses to other variables yields a similar scenario. While **ECMWF** tries to align with **ERA5**, preserving the initial conditions from **ERA5**, **ERA5** shows greater spread and less physical coherence, diverging from **ERA5** towards the observations. Thus, **ECMWF** better replicates observed variability in spatially structured environments.

Precipitation downscaling remains difficult for statistical methods. **ERA5** overestimates both rain and snow seasonal variability, especially in cold regions. **ECMWF** better captures spatial patterns in spring and summer extremes. These findings highlight the need for improved frameworks.

When errors in analog selection lead to significant underestimations, hybrid methods and conservative indices could enhance **ECMWF**'s ability to capture such extremes.

When used as a weather generator, **ECMWF** accurately reproduces daily temperature extremes (**ERA5** and **ERA5**), aligning closely with **ERA5** in terms of spread and skewness, unlike **ERA5**. These results confirm the value of **ECMWF** multivariate coherence and spatial consistency. However, improving its treatment of extremes remains key for broader applicability.

“Due to analogs averaging, **kNN** struggles with extremes, often underestimating cold days and intense summer precipitation, especially in mountainous regions.”

CONCLUSION

*The analysis of spatial variability shows that **kNN** produces more detailed and realistic **T2m** fields across Italy than **EQM**. Unlike **EQM**'s smooth and generalised fields, **kNN** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**. Statistically, **kNN** achieves better performance with a 0.98 correlation and 1.63°C RMSE, outperforming **EQM**.*

***T2m** seasonal bias assessment reveals that **EQM** has higher winter biases (up to 8°C in the Alps), while **kNN** shows smoother and more accurate seasonal cycles. **kNN** better reflects **ERA5** data across all seasons, especially in marine regions. **EQM**'s apparent low marine bias is misleading, caused by seasonal over- and underestimations balancing each other.*

*Environmental analysis across Italy shows that **kNN** outperforms **EQM** in all regions, especially in coastal and mountainous areas where **EQM** fails to capture contrasts. Both methods perform better over marine areas, but **kNN** shows lower errors inland. The persistent bias in mountainous zones indicates challenges in representing cold extremes.*

*In seasonal forecasting (up to 6 months), **kNN** shows lower bias and RMSE than **EQM** and **SPS**, preserving initial signals while improving local-scale detail. **EQM**'s bias correction introduces uncertainty due to inter-member spread. **kNN**'s ensemble is more consistent, offering better reliability for climate-sensitive applications.*

*Expanding these analyses to other variables yields a similar scenario. While **kNN** tries to align with **ERA5**, preserving the initial conditions from **SPS**, **EQM** shows greater spread and less physical coherence, abruptly forcing **SPS** towards the observations. Thus, **kNN** better replicates observed variability in spatially structured environments.*

*Precipitation downscaling remains difficult for statistical methods. **EQM** overestimates light rain and misses seasonal variability, especially in cold months and orographic zones. **kNN** better captures spatial patterns in spring and autumn but underrepresents convective summer extremes. These challenges demand more sophisticated inputs and frameworks.*

*Due to analogs averaging, **kNN** struggles with extremes, often underestimating cold days and intense summer precipitation, especially in mountainous regions. Minor errors in analogs selection lead to significant rainfall underestimations. Hybrid methods and convective indices could enhance **kNN**'s ability to capture such extremes.*

*When used as a weather generator, **kNN** accurately reproduces daily temperature extremes (**T2N** and **T2X**), aligning closely with **ERA5** in terms of spread and skewness, unlike **EQM**. These results confirm the value of **kNN**'s multivariate coherence and spatial sensitivity. However, improving its treatment of extremes remains key for broader applicability.*

CONCLUSION

The analysis of spatial variability shows that **ERA5** produces more detailed and realistic TSS fields across daily than **EQM**. Unlike **EQM**'s smooth and generalized fields, **ERA5** captures key geographic features like the Po Valley and the Alps, aligning closely with **ERA5**. Statistically, **ERA5** achieves better performance with a skill correlation and a 40% RMSE, outperforming **EQM**.

The seasonal bias assessment reveals that **EQM** has higher winter biases (up to 4°C in the Alps), while **ERA5** shows smoother and more accurate seasonal cycles. **ERA5** better captures **ERA5**'s apparent low winter over- and underestimations better.

Environmental analysis across all regions, especially in coastal and urban areas, shows that **ERA5** captures extremes. Both methods perform well in capturing extremes, but **ERA5** shows lower errors related. The persistent bias in underestimating winter indicates challenges in representing cold extremes.

In seasonal forecasting (up to 6 months), **ERA5** shows lower bias and RMSE than **EQM** and **ERA5**, preserving initial signals while improving local-scale detail. **EQM**'s bias correction introduces uncertainty due to later number spread. **ERA5**'s ensemble is more consistent, offering better reliability for climate-sensitive applications.

Expanding these analyses to other variables yields a similar scenario: while **ERA5** tries to align with **ERA5**, preserving the initial conditions from **ERA5**, **EQM** shows greater spread and less physical coherence, diverging from **ERA5** towards the observations. Thus, **ERA5** better replicates observed variability in spatially structured environments.

Precipitation downscaling remains difficult for statistical methods. **EQM** overestimates both rain and snow seasonal variability, especially in cold regions. While **ERA5** captures extreme patterns in spring and summer, these extremes are not fully represented.

Overall, **ERA5** outperforms **EQM** in capturing extremes, offers more realistic seasonal cycles, and maintains coherence in precipitation, especially in cold regions. While errors in model selection lead to significant underestimations, hybrid methods and ensemble models could enhance **ERA5**'s ability to capture such extremes.

When used as a weather generator, ERA5 accurately reproduces daily temperature extremes (T2N and T2X), aligning closely with ERA5 in terms of spread and skewness, unlike EQM. These results confirm the value of **ERA5**'s multivariate coherence and spatial smoothness. However, improving its treatment of extremes remains key for broader applicability.

“When used as a weather generator, **kNN** accurately reproduces daily temperature extremes (**T2N** and **T2X**), aligning closely with **ERA5** in terms of spread and skewness, unlike **EQM**.”

NEXT STEPS

SCIENTIFIC PUBLICATION

The full version of this experiment considering the whole set of variables, analysis and discussions

OPERATIONAL VERSION

*Integrate the **kNN** tool to **SPS4** to provide statistical downscaled predictions monthly*

MedCORDEX DOMAIN

6 times faster is not enough

TOWARDS EXTREMES

*Review the list of predictors to include large-scale variables and enhance the **kNN** skill to a daily level*

ARAGÃO, Borrelli & Gualdi (2025)

*CMCC Seasonal Prediction System v3.5 hindcasts downscaled from 1-degree to 0.25-degree using Empirical Quantile Mapping and k-Nearest Neighbours.
CMCC Data Delivery System, DOI [10.25424/cmcc-4brr-ym18](https://doi.org/10.25424/cmcc-4brr-ym18)*

ARAGÃO, Borrelli & Gualdi (2025) ▫ **Under Review**

*Expanding CMCC Seasonal Prediction System v3.5 applications to the local scale through statistical downscaling techniques.
Earth System Science Data, DOI [10.5194/essd-2025-371](https://doi.org/10.5194/essd-2025-371)*



cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

20 Years

Expanding CMCC Seasonal Prediction System v3.5 applications to the local scale through
STATISTICAL DOWNSCALING TECHNIQUES

Thank you!

Leonardo Aragão, Andrea Borrelli & Silvio Gualdi
leonardo.aragao@cmcc.it

www.cmcc.it

