6th WGNE workshop on systematic errors in weather and climate models (ECMWF - 2022)

Climate forecast reliability by artificial intelligence classification

JR Garcia, Paulo Nobre, Chou Sin-Chan, Jorge Gomes, Emanuel Giarolla, Manoel Baptista
INPE - National Institute for Space Research - Brazil

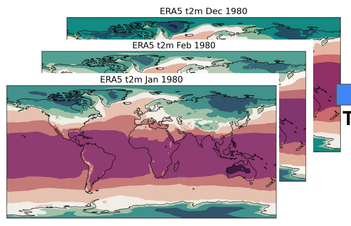
Goal	Estimate monthly 2m temp forecast reliability by AI from a given initial condition

METHODOLOGY	
Truth (assumption)	t2m forecast near normal
Independent variables	All atmospheric model initial conditions variables at all levels (predictors)
Dependent variables	t2m (2m temperature) (target variable)
Both independent and dependent variables were categorized into classes below normal, near normal and above normal, based on the correspondent monthly climatology	

BESM Brazilian Earth System Model

DATA	
Resolution	TQ0062L042 grid 96 x 192 x 42 pressure levels
Time	1980-2009 hindcast (30 years)
Initial Conditions	Aug, Sep, Oct, Nov (4 x 30ys = 120 initial conditions)
Forecasts	12 months for each IC (12m x 30ys = 360 monthly forecasts)

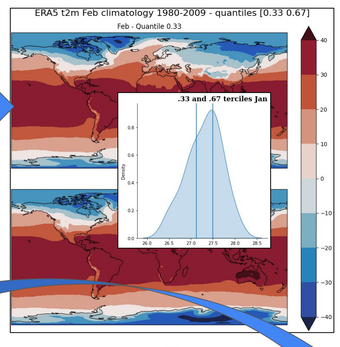
ERA5 t2m monthly 1980-2009
30 years x 12 mon = 360 months



1

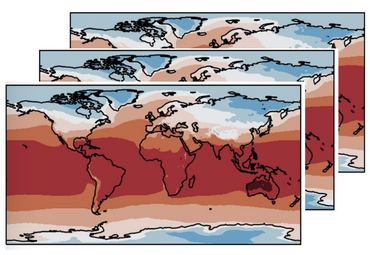
TERCILES

ERA5 t2m monthly .33 and .67 quantiles per grid point (12 months)



2

ERA5 t2m monthly 1980-2 climatology (12 month

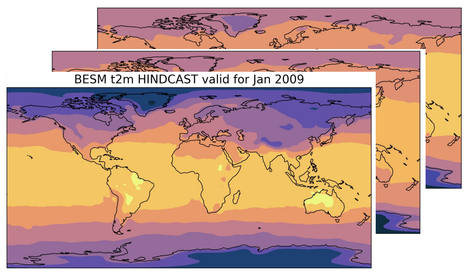


3

3-CLASS CATEGORIZATION

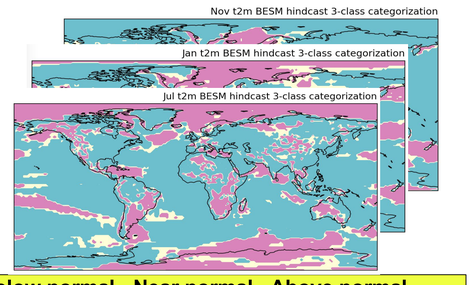
BUILDING THE TARGET

BESM t2m monthly hindcast 1980-2009
30 years x 12 mon x 4 init dates = 1440 months

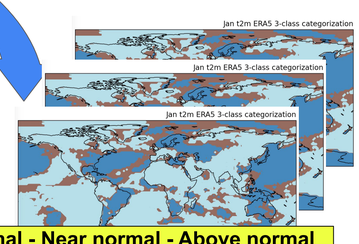


4

3-CLASS CATEGORIZATION



Below normal - Near normal - Above normal
(-1) (0) (1)



Below normal - Near normal - Above normal
(-1) (0) (1)

5

5

```
NetCDF → Xarray → to_dataframe() → to_csv()
~26M rows, 7 cols
```

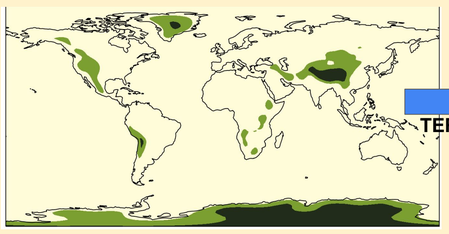
lat	lon	mon_init	ltime	cat_fct	cat_obs	categ
88.572	0.000	8	1	-1	1	-2
88.572	1.875	8	1	-1	1	-2
88.572	3.750	8	1	-1	1	-2
88.572	5.625	8	1	-1	1	-2
88.572	7.500	8	1	-1	1	-2
...	...	...	...	...	...	...
88.572	350.625	11	2	1	1	0
88.572	352.500	11	2	1	1	0
88.572	354.375	11	2	1	1	0
88.572	356.250	11	2	1	1	0
88.572	358.125	11	2	1	1	0

CATEG = ERROR (FCT-OBS)

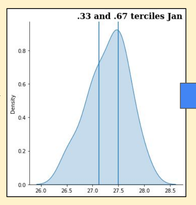
- 0 = High confidence
- 1 = Moderate
- 2 = Low

Signal informs direction

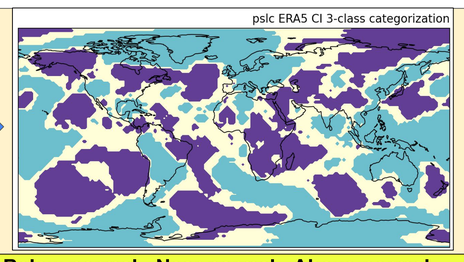
BUILDING PREDICTORS



TERCILES



3-CLASS CATEGORIZATION



Below normal - Near normal - Above normal
(-1) (0) (1)

MERGED

```
NetCDF → Xarray → to_dataframe() → to_csv()
All 42 pressure levels were expanded
169 cols
```

lat	lon	mon_init	ltime	cat_fct	cat_obs	categ
88.572	0.000	8	1	-1	1	-2
88.572	1.875	8	1	-1	1	-2
88.572	3.750	8	1	-1	1	-2
88.572	5.625	8	1	-1	1	-2
88.572	7.500	8	1	-1	1	-2
...	...	...	...	...	...	...
88.572	350.625	11	2	1	1	0
88.572	352.500	11	2	1	1	0
88.572	354.375	11	2	1	1	0
88.572	356.250	11	2	1	1	0
88.572	358.125	11	2	1	1	0

pslc	tvir_002.042	tvir_007.647	tvir_013.831	tvir_020.938	...	vvel_921.760	vvel_938.427	vvel_953.154	vvel_966.105	vvel_977.447	vvel_987.352	vvel_995.983
					...							
0	1	1	0	1	...	1	1	1	1	1	1	1
0	1	1	0	1	...	1	1	1	1	1	1	1
1	1	1	0	1	...	-1	-1	-1	-1	-1	-1	-1
1	1	1	0	1	...	-1	-1	-1	-1	-1	-1	-1
1	1	0	0	1	...	-1	-1	-1	-1	-1	-1	-1

INITIAL CONDITIONS
ERA5 monthly pressure level
Reanalysis 1980-2009
Months: Aug, Sep, Oct, Nov
30y x 4 m = 120 ICs

pslc	**Surface Pressure (hPa)**
tvir	**Virt Temperature (K)**
umes	**Specific Humidity (kg/kg)**
uvel	**Zonal Wind (m/s)**
vvel	**Meridional Wind (m/s)**

AI MODELING

DECISION TREE CLASSIFIER

- **Predictors:** <all_169_meteo_vars> + lat + lon + cat_fct
- **Target:** categ
- Loop in init_month and lead time → so data were separated at each iteration

AI RESULTS