

# SEASONAL FORECASTS BASED ON STATISTICAL ALGORITHMS

Massimiliano Pasqui

Institute of Biometeorology National Research Council Rome, Italy.

[m.pasqui@ibimet.cnr.it](mailto:m.pasqui@ibimet.cnr.it)

with contributions from:

V. Pavan, F.J. Doblas-Reyes, J. Primicerio, S. Quaresima, E. Di  
Giuseppe, F. Guarnieri, L. Genesio, R. Benedetti, A. Crisci

**REGIONAL TRAINING ON LONG-RANGE FORECASTS, CLIMATE WATCH RELATED  
ASPECTS AND CLIMATE SCENARIOS - LRF Training  
13 - 16 November 2013, Belgrade, Serbia**

## BEST PRACTICES IN SEASONAL FORECASTING

The following requirements for producing, using, and assessing seasonal forecasts were agreed upon during and in discussions subsequent to the First WCRP Workshop on Seasonal Prediction.

- Forecast error must be addressed by appropriately quantifying dynamical model uncertainty;
- Model output should be recalibrated based on historical model performance;
- Probabilistic forecast information should be issued;
- A description of the forecast process should be made available;
- In retrospective forecast mode, no information about the future should be used;
- Forecast quality information should be provided, including several metrics of quality;
- Regional climate service providers need to work with both the forecasting and application communities to develop tailored downscaled products;
- Users must be encouraged to use all the ensemble members to quantify forecast uncertainty;
- Web-based tools need to be developed to allow users to tailor forecast information;
- Regional mechanisms like Regional Climate Outlook Forums (RCOFs) should be used to develop regional climate outlooks based on the consensus and objective scientific assessment of multiple-prediction outcomes;
- Liaison with users should be promoted to understand their climate information needs in decision making and also to raise their awareness of the uncertainty aspects of seasonal forecasting;
- Regional/national ownership of seasonal forecasts should be promoted through effective and sustained capacity building and infrastructural support.

Kirtman, Ben, Anna Pirani, 2009: The State of the Art of Seasonal Prediction: Outcomes and Recommendations from the First World Climate Research Program Workshop on Seasonal Prediction. *Bull. Amer. Meteor. Soc.*, 90, 455–458. doi: <http://dx.doi.org/10.1175/2008BAMS2707.1>

### The State of the Art of Seasonal Prediction

Outcomes and Recommendations from the First World Climate Research Program Workshop on Seasonal Prediction

BY BEN KIRTMAN AND ANNA PIRANI

# Interesting Timescales

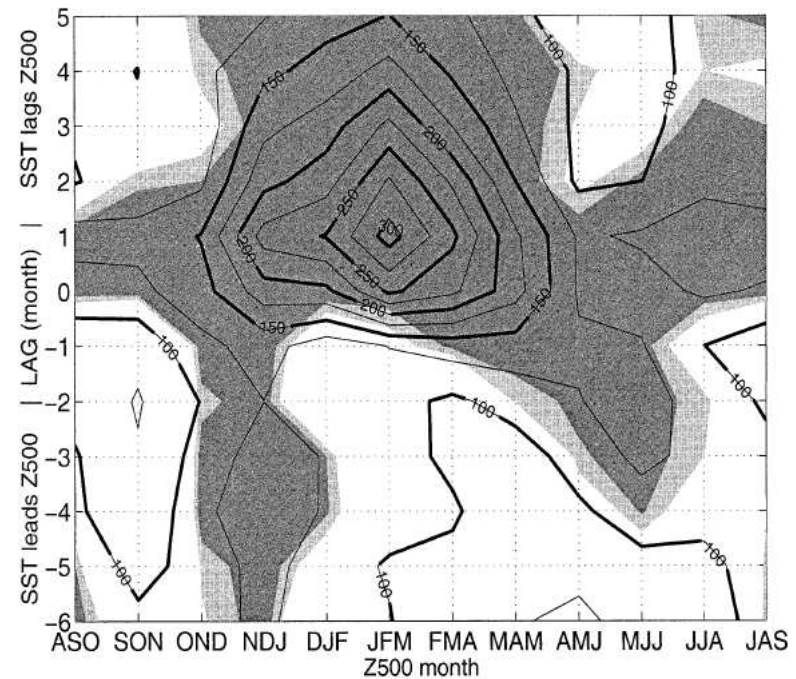
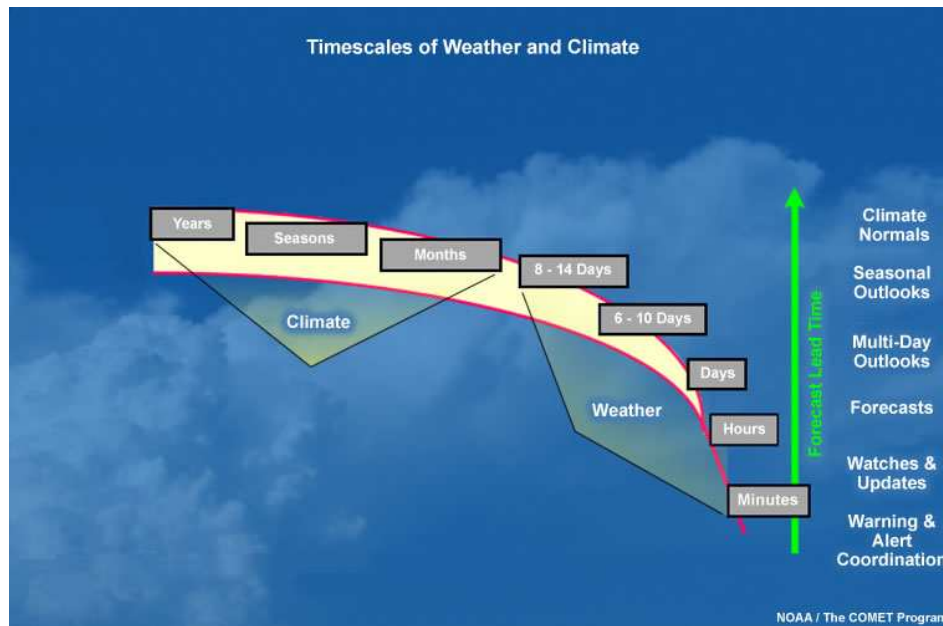
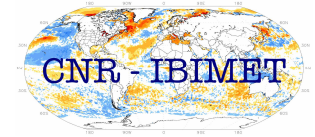
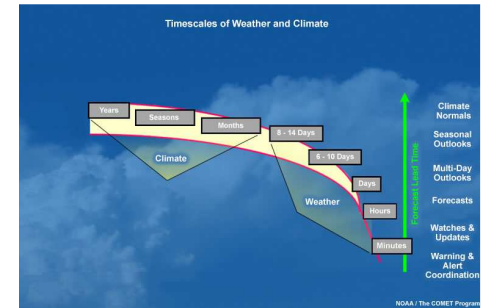


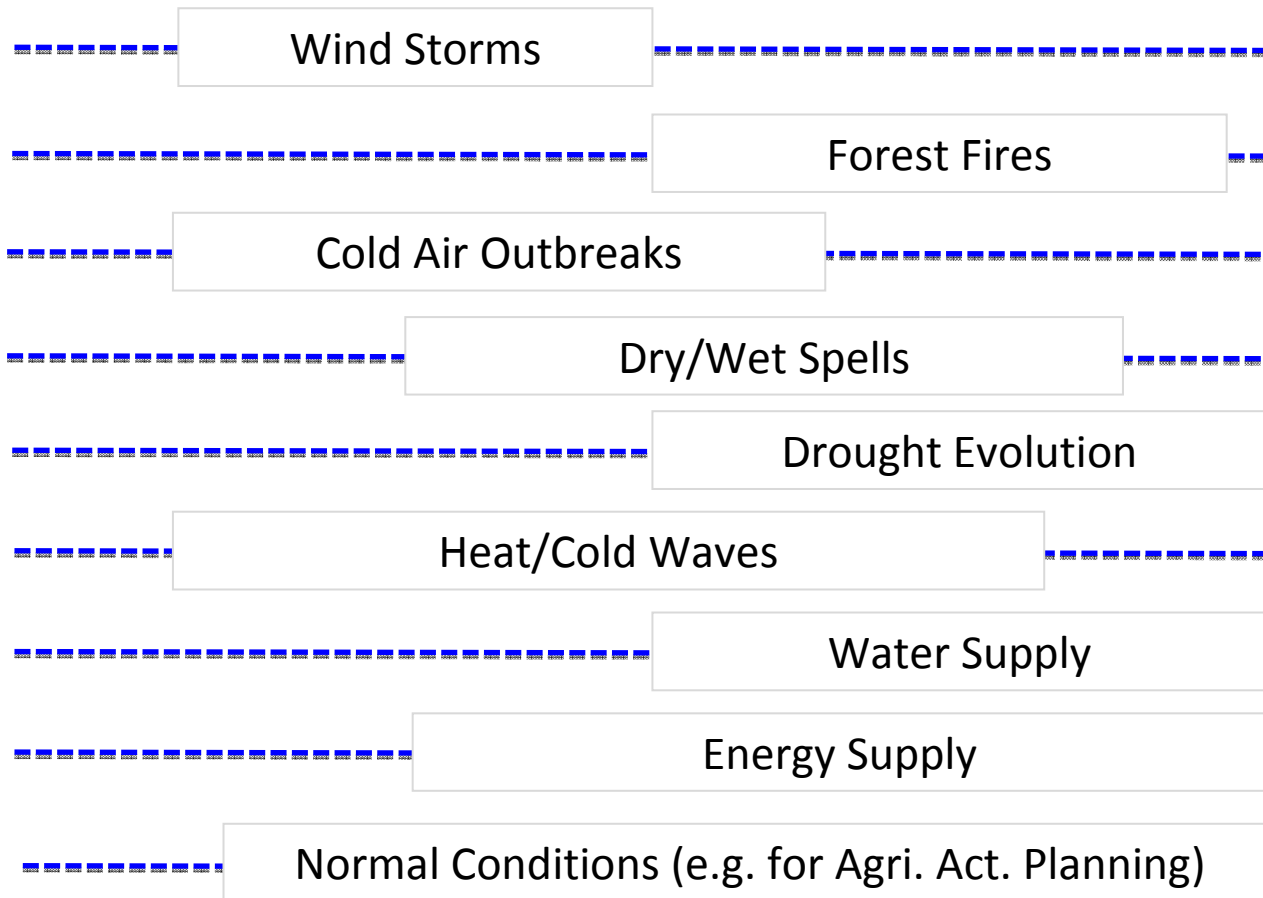
FIG. 1. The SC of the first MCA mode between Pan-Atlantic SST and  $Z_{500}$  anomalies (the SCs are dimensionless as the SST and height fields have been normalized—see section 2). SST leads  $Z_{500}$  at negative lags indicated (in months) on the y axis while the x axis denotes the months assigned to  $Z_{500}$ . The shaded area indicates where the covariance is statistically significant at the 5% and 10% level (dark and light shading, respectively).

Czaja and Frankignoul, 2002

# Interesting Timescales #2



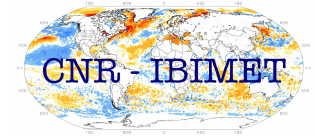
Day s    Week s    Month    Season    → Decadal?



- Seasonal forecasting represents the attempt to predict the spatial and temporal distribution of atmospheric anomalies few months in advance.
- Even though the detailed dynamical evolution of atmospheric system is not predictable on those time scales, some of its statistical features and behaviors can be forecasted.
- In particular it is possible to infer the mean behavior over a month or season and how much the PDFs of such values, or anomalies, differ from "climatology". Or in other words it can be highlighted a specific regimes.

- In the Mediterranean basin where a changing climate seems acting towards a remarkable footprint both for rainfall and more pronounced on heat wave related events, the water management should be adopted at seasonal time scale for critical scenario modelling.
- Nowadays seasonal forecast systems should still cover a dynamical scale gap to reach local scale and sub monthly climatic variability.
- E.g. drought is a composition of different physical mechanisms: *large scale climatic anomalies* and *local environmental factors* and it is of prime interest for civil, agricultural and industrial activities.
- Seasonal forecasting deals mainly with large scale climatic phenomena.

- Dynamical models become a major tool in seasonal prediction and have demonstrated a significant progress in tropical seasonal to inter-annual predictions (e.g., ENSO dynamic), but their direct outputs seem to be much less satisfactory for middle and higher latitudes;
- On the other side statistical models can give comparable and even higher skills for specific regions and for some variables, and their results can be easily applied for practical use without a full comprehension of physical mechanisms behind.
- Finally statistical models need little computational resources being much more simple to be used than dynamical ones.
- But they requires some careful attentions since reliability is obtained through a balance between goodness-of-fit and stability of the model.



# IBIMET on SeaFor

Main Activities related:

- Early – Warnings
- Monitoring and Survey



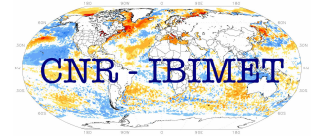
## Institutional End Users

• National Civil Protection Department: Monthly and Seasonal forecast group. *It meets monthly, provides outlook for one to three months ahead mainly for the management of water scarcity crisis [it started in winter 2007]. It is composed by IBIMET, ARPA-SIM, CNMCA, CRA-CMA (ex UCEA), ISAC-CNR.*

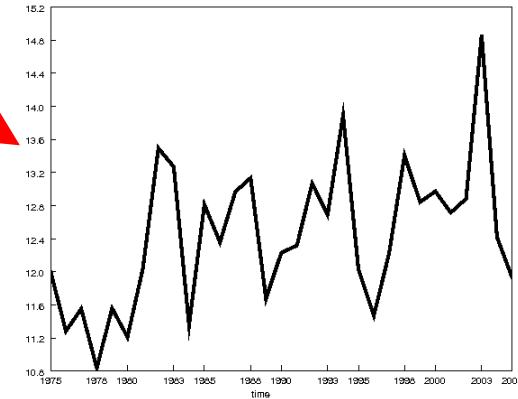
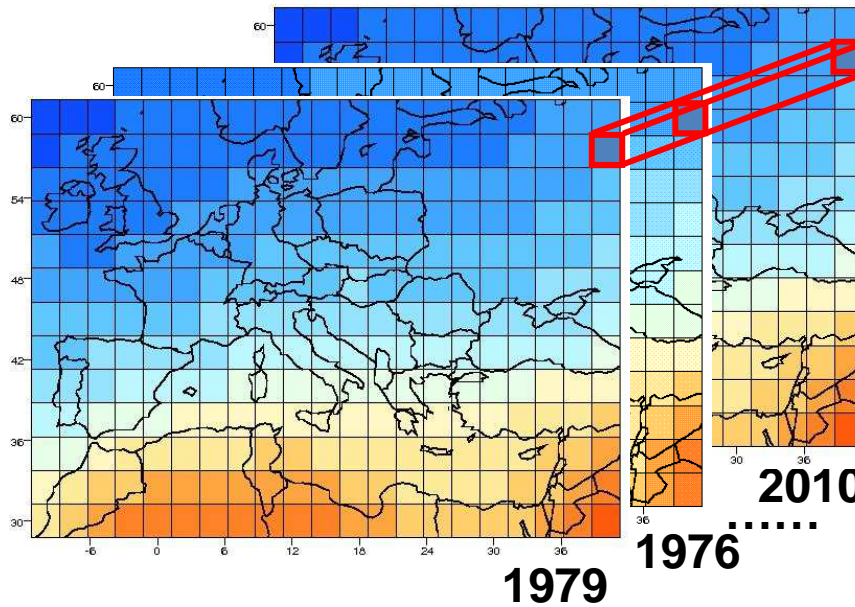
• Arno river basin Water Safeguard Commission: *monthly meetings*



# Seasonal Forecast Overview



**Observed  
Temperature  
Matrix**



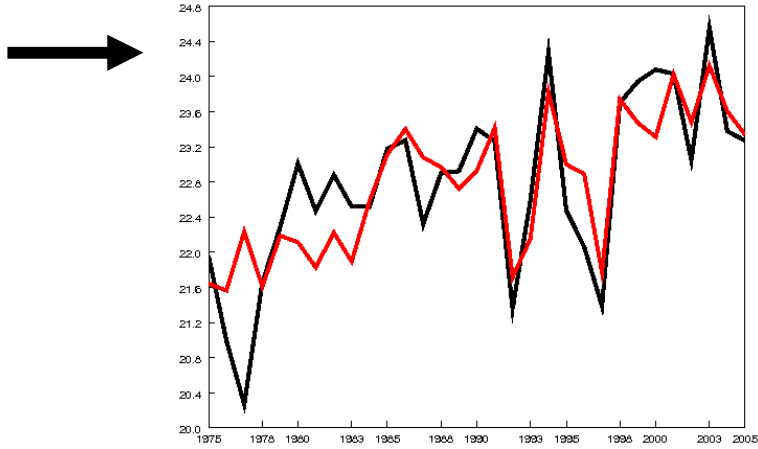
**Temperature at location  
(x,y)**

**Predictors  
Matrix**

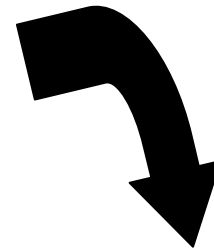
0	3	2	5	4	7	6	9	8	11	10	13	12	15	14	17	16	19	18	21	20	23	22	25	24
3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2	1	0	3	2	5	4	7	6	9	8	11	10	13	12	15	14	17	16	19	18	21	20	23	22
5	2	3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
4	3	2	1	0	3	2	5	4	7	6	9	8	11	10	13	12	15	14	17	16	19	18	21	20
7	4	5	2	3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
6	5	4	3	2	1	0	3	2	5	4	7	6	9	8	11	10	13	12	15	14	17	16	19	18
9	6	7	4	5	2	3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
8	7	6	5	4	3	2	1	0	3	2	5	4	7	6	9	8	11	10	13	12	15	14	17	16
11	8	9	6	7	4	5	2	3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
10	9	8	7	6	5	4	3	2	1	0	3	2	5	4	7	6	9	8	11	10	13	12	15	14
13	10	11	8	9	6	7	4	5	2	3	0	1	2	3	4	5	6	7	8	9	10	11	12	13
12	11	10	9	8	7	6	5	4	3	2	1	0	3	2	5	4	7	6	9	8	11	10	13	12
15	12	13	10	11	8	9	6	7	4	5	2	3	0	1	2	3	4	5	6	7	8	9	10	11
14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	3	2	5	4	7	6	9	8	11	10
17	14	15	12	13	10	11	8	9	6	7	4	5	2	3	0	1	2	3	4	5	6	7	8	9
16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	3	2	5	4	7	6	9	8
19	16	17	14	15	12	13	10	11	8	9	6	7	4	5	2	3	0	1	2	3	4	5	6	7
18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	3	2	5	4	7	6
21	18	19	16	17	14	15	12	13	10	11	8	9	6	7	4	5	2	3	0	1	2	3	4	5
20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	3	2	5	4
23	20	21	18	19	16	17	14	15	12	13	10	11	8	9	6	7	4	5	2	3	0	1	2	3
22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	3	2
25	22	23	20	21	18	19	16	17	14	15	12	13	10	11	8	9	6	7	4	5	2	3	0	1
24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

**Multivariate Linear  
Regression algorithm**

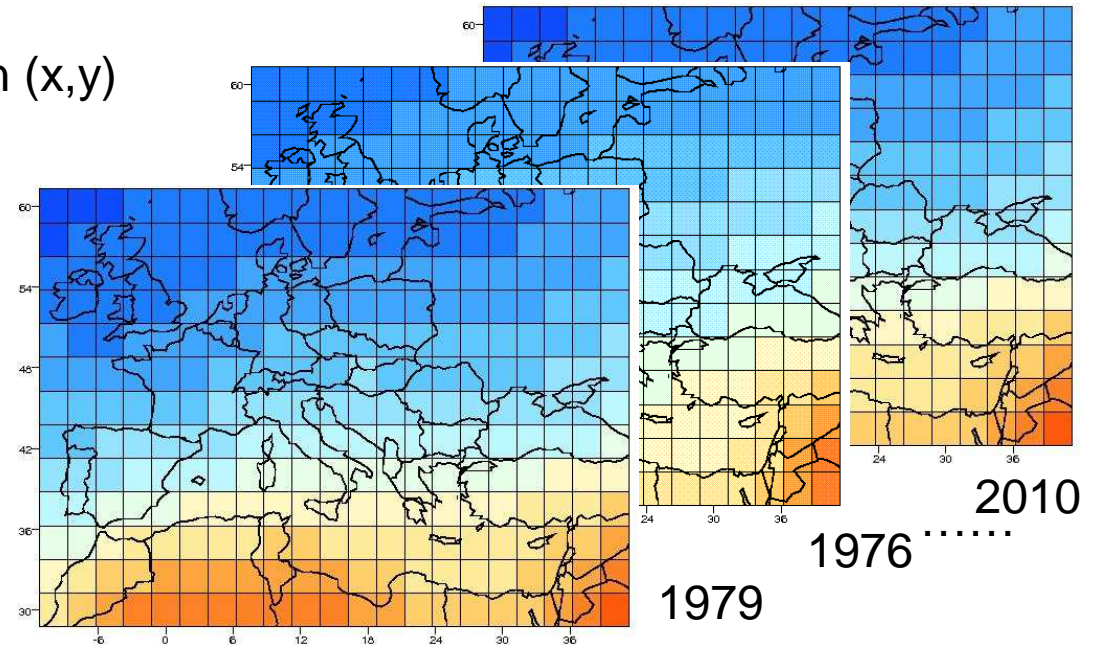
**Forecasted Temperature at  
(x,y) for time (t) expressed as  
probability of occurrence**



Hindcasted Temperature at location (x,y)



### Hindcast Temperature Matrix



Multiregressive

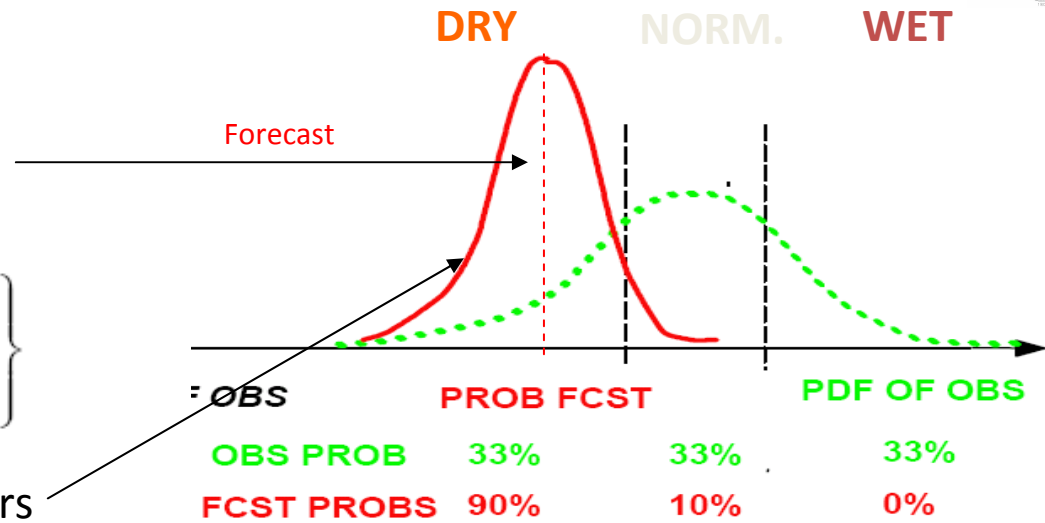
$$\hat{y}_i = b_0 + \sum_{k=1}^K b_k x_{ik}$$

$$\left\{ \begin{array}{l} y_i \quad i = 1, \dots, n \\ x_{i,k} \quad i = 1, \dots, n, \quad k = 1, \dots, K \end{array} \right\}$$

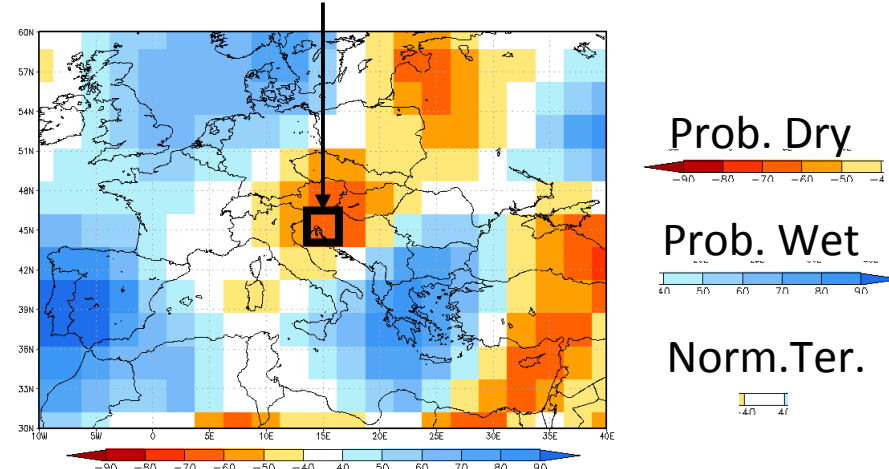
Multiregressive Errors

$$s_\epsilon^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - K - 1} = \frac{SSE}{n - K - 1}$$

$s_\epsilon^2$  : Forecast Error Variance  
 SSE: Sum Square Errors  
 n : Sample amount  
 K : Predictors



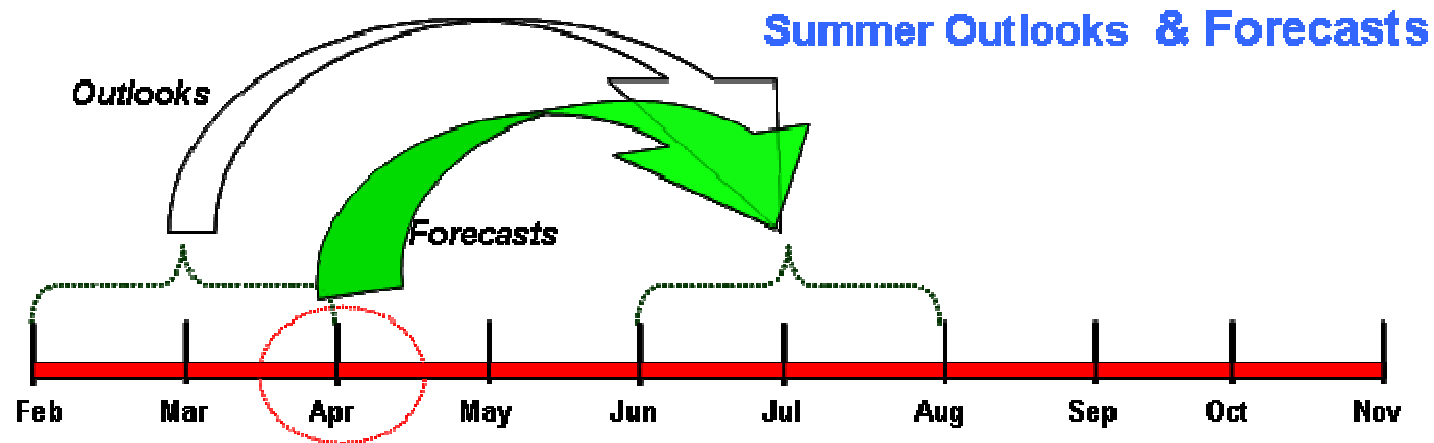
$$P_{forecast(D,N,W)} \equiv P_{Fcst} \perp P_{Obs}_{D,N,W}$$



The multi-regressive method based on physical atmospheric indices and sea surface anomalies, at monthly time scale.

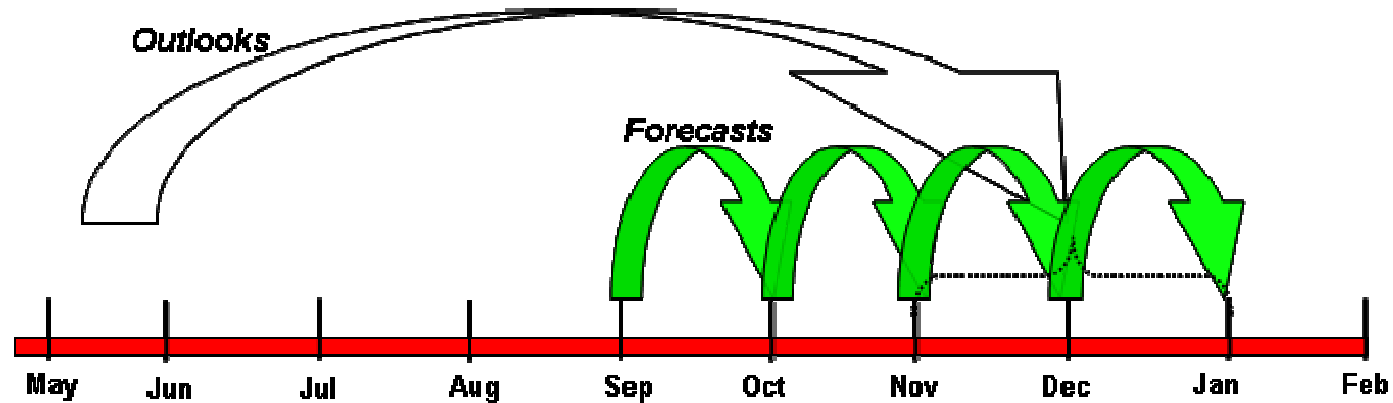
Lead – time choices are made on physical basis and on a maximization of the regression values between observed and forecasted field anomalies thus performing an “*adaptation*” for the best choice of predictors.

Short Name	Full Name	Source
MED	Mediterranean Sea 1 <sup>st</sup> EOF SST (JUNG ET AL., 2006)	CPC <sup>1b</sup>
AMO	Atlantic <u>Multidecadal</u> Oscillation (Enfield et al., 2001)	CDC <sup>1b</sup>
MEI	Multivariate ENSO Index (Wolter & Timlin, 1993)	CDC
NAO	North Atlantic Oscillation ( <u>Barnston</u> and <u>Livezey</u> , 1987)	CPC
SV-NAM	Seasonally Varying Northern Hemisphere Annular Mode ( <u>Ogi</u> et al., 2003, 2004)	HOK <sup>1b</sup>
MZI	Modified Zonal Index (J. P. Li & Wang, 2003)	IRI <sup>1b</sup>
TRI	Atlantic <u>Tripole</u> 1 <sup>st</sup> EOF SST ( <u>Deser</u> & <u>Timlin</u> , 1997)	IRI
GUI	Guinea Gulf 1 <sup>st</sup> EOF SST <sup>2b</sup>	IRI
IND	Indian Sea 1 <sup>st</sup> EOF SST <sup>2b</sup>	IRI
NASCI	North American Snow Cover Index <sup>4b</sup>	IRI
SISCI	Siberian Snow Cover Index <sup>2b</sup>	IRI
PDO	Pacific <u>Decadal</u> Oscillation (Mantua et al., 1997)	JISAO

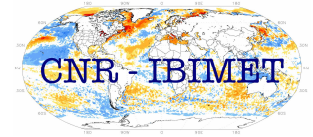


Summer variability could be associated to winter atmospheric variability (SV – NAM, snow cover, etc.) and to Atlantic SST anomalies (Tripole and 1<sup>st</sup> EOF Guinea).

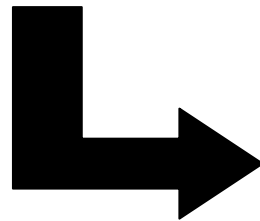
## Winter / Spring Outlooks & Forecasts



Fall to Winter variability could be associated to Spring to SST anomalies through reemerging Atlantic SSTA mechanism (OUTLOOKS) plus the atmospheric fall – winter variability (MONTHLY).



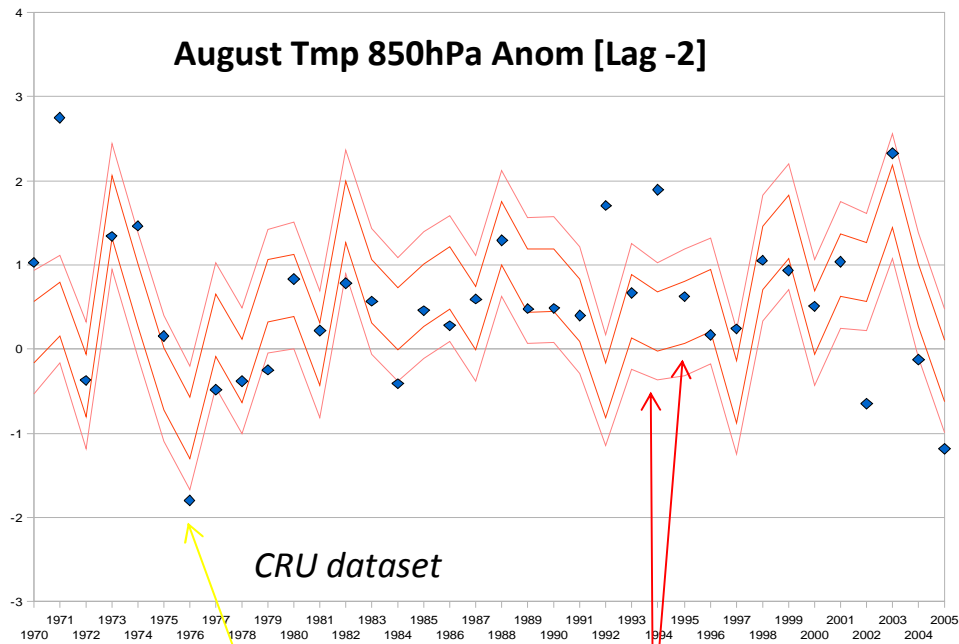
<http://web.fi.ibimet.cnr.it/seasonal/>



### Seasonal Rainfall and Temperature Forecast

#### Adaptive Multiregressive Method

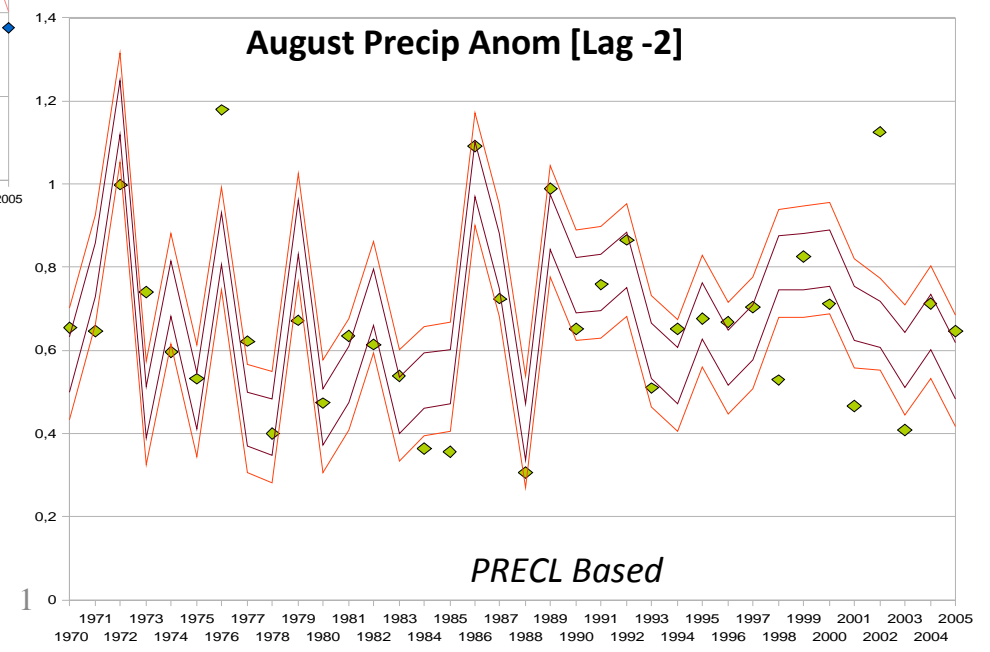
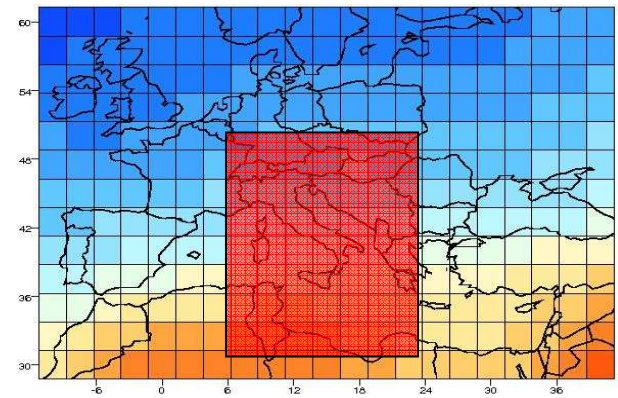
Month / Forecast	Geopotential Anomaly at 850 hPa	Precipitation Probability Forecast	Temperature Probability Forecast
<b>Nov 2013</b>	<p>November 2013</p> <p>Best Date category for 850 hPa geopotential height Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 850 hPa geopotential height Forecast error is 5.000</p>	<p>November 2013</p> <p>Best Date category for 1 month accumulated precipitation Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 1 month accumulated precipitation Forecast error is 5.000</p>	<p>November 2013</p> <p>Best Date category for 850 hPa temperature Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 850 hPa temperature Forecast error is 5.000</p>
<b>Dec 2013</b>	<p>December 2013</p> <p>Best Date category for 850 hPa geopotential height Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 850 hPa geopotential height Forecast error is 5.000</p>	<p>December 2013</p> <p>Best Date category for 1 month accumulated precipitation Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 1 month accumulated precipitation Forecast error is 5.000</p>	<p>December 2013</p> <p>Best Date category for 850 hPa temperature Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 850 hPa temperature Forecast error is 5.000</p>
<b>Jan 2014</b>	<p>January 2014</p> <p>Best Date category for 850 hPa geopotential height Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 850 hPa geopotential height Forecast error is 5.000</p>	<p>January 2014</p> <p>Best Date category for 1 month accumulated precipitation Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 1 month accumulated precipitation Forecast error is 5.000</p>	<p>January 2014</p> <p>Best Date category for 850 hPa temperature Forecast error is 5.000</p> <p>Forecast Method: Adaptive Multiregressive Method</p> <p>Best Date category for 850 hPa temperature Forecast error is 5.000</p>



*CRU dataset*

*Obs surface Anom [°C]*

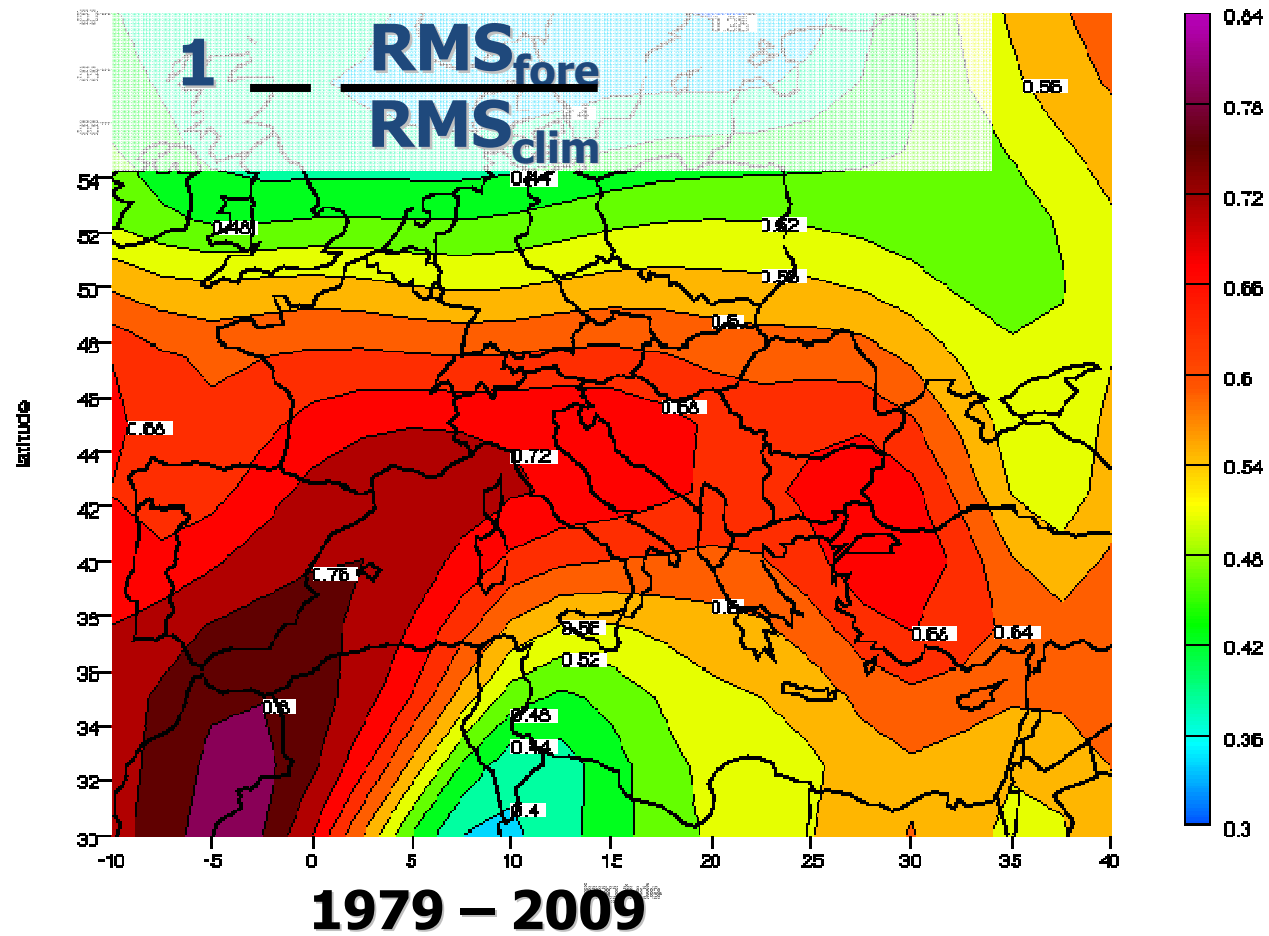
*Forecasted tmp 850hPa [°C]*  
 *$\pm 0.5 / \pm 1$  sigma band*

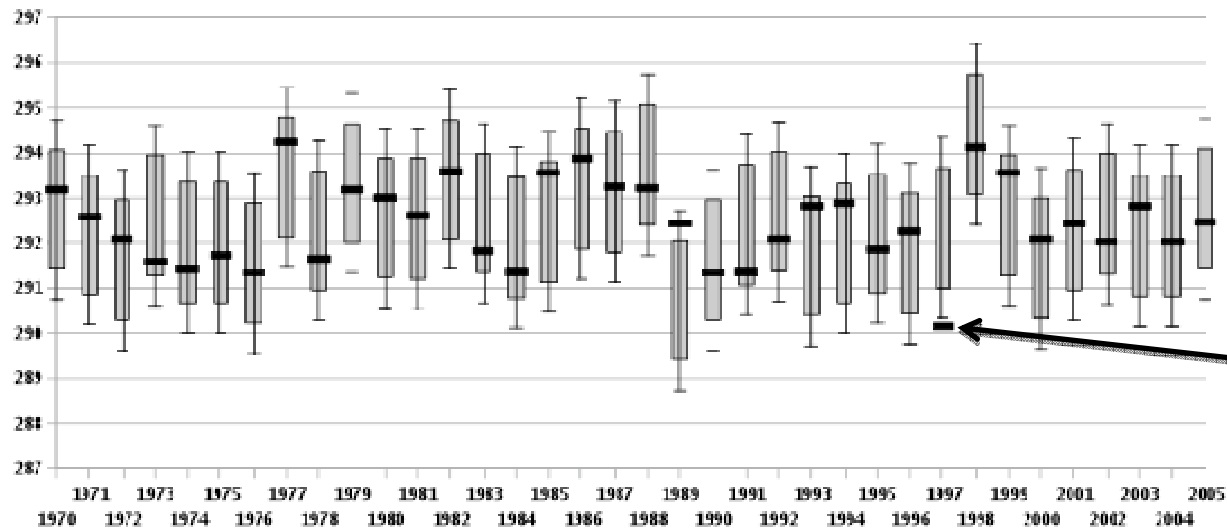
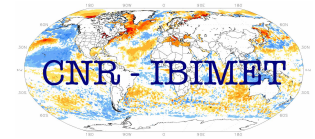


*PRECL Based*



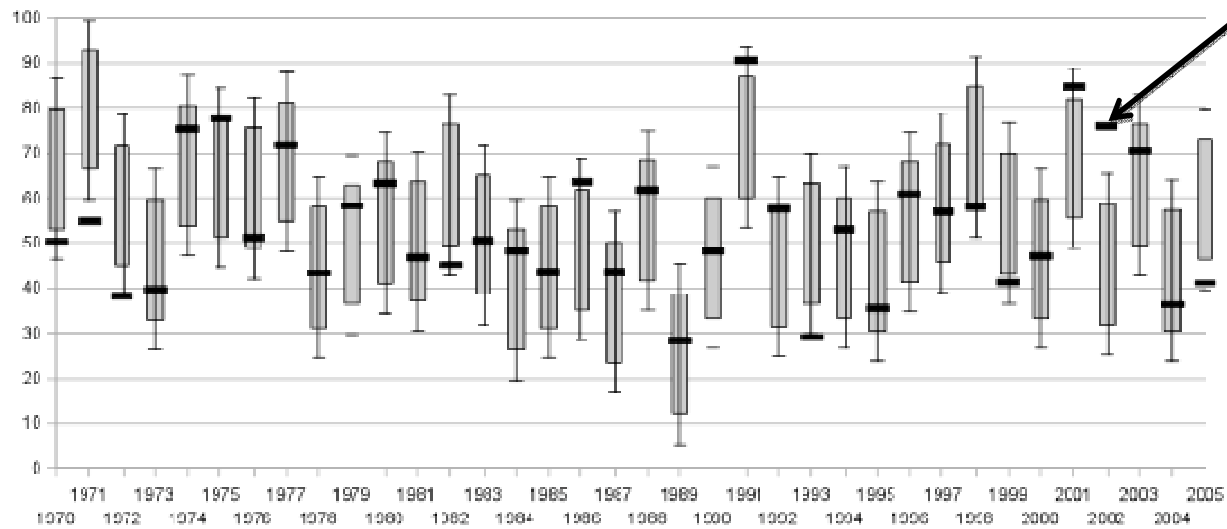
Es.: Temp 850hPa JJA issued in April





Aug Temp EasternMed

Reference Values



Dec Pcp EasternMed

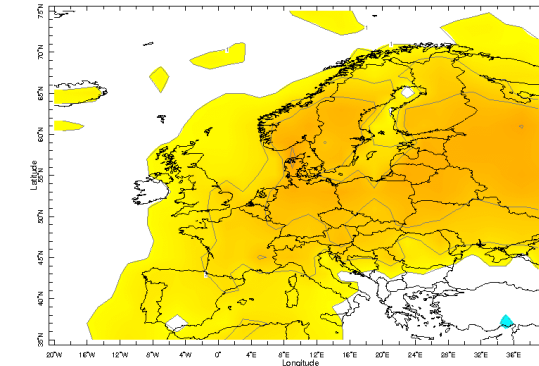
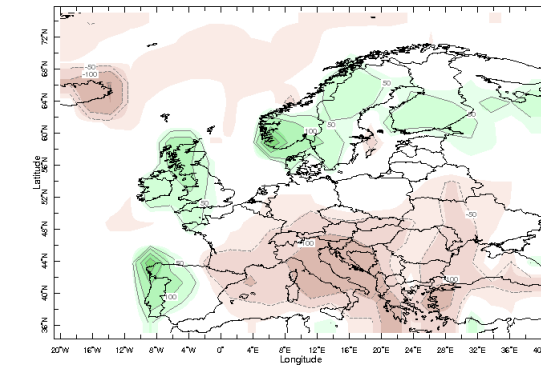
Reference Values

An exceptional (\*) warm period in Europe and a long dry period in southern Europe.

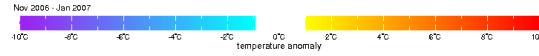
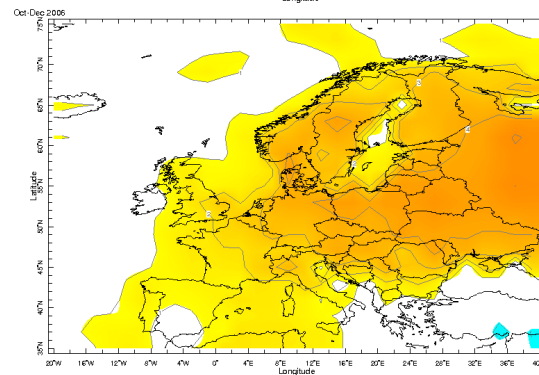
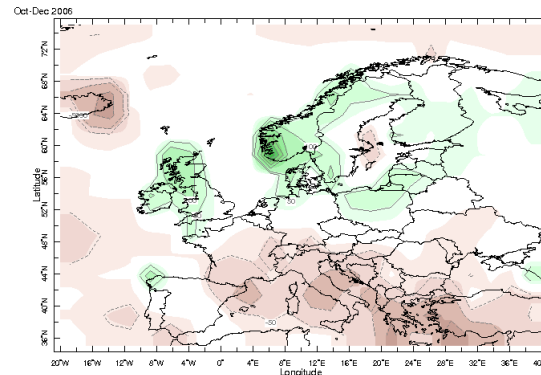
In Jan 2007 Italian Civil Protection (DPC) established a technical working group for monthly to seasonal forecasts for drought emergency. It consist of 5 groups: ARPA-SIMC, AM, CRA-CMA, CNR-ISAC, CNR-IBIMET.

3-months seasonal forecast from ARPA-SIMC & CNR – IBIMET are shown here.

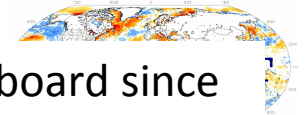
Oct – Dec 2006



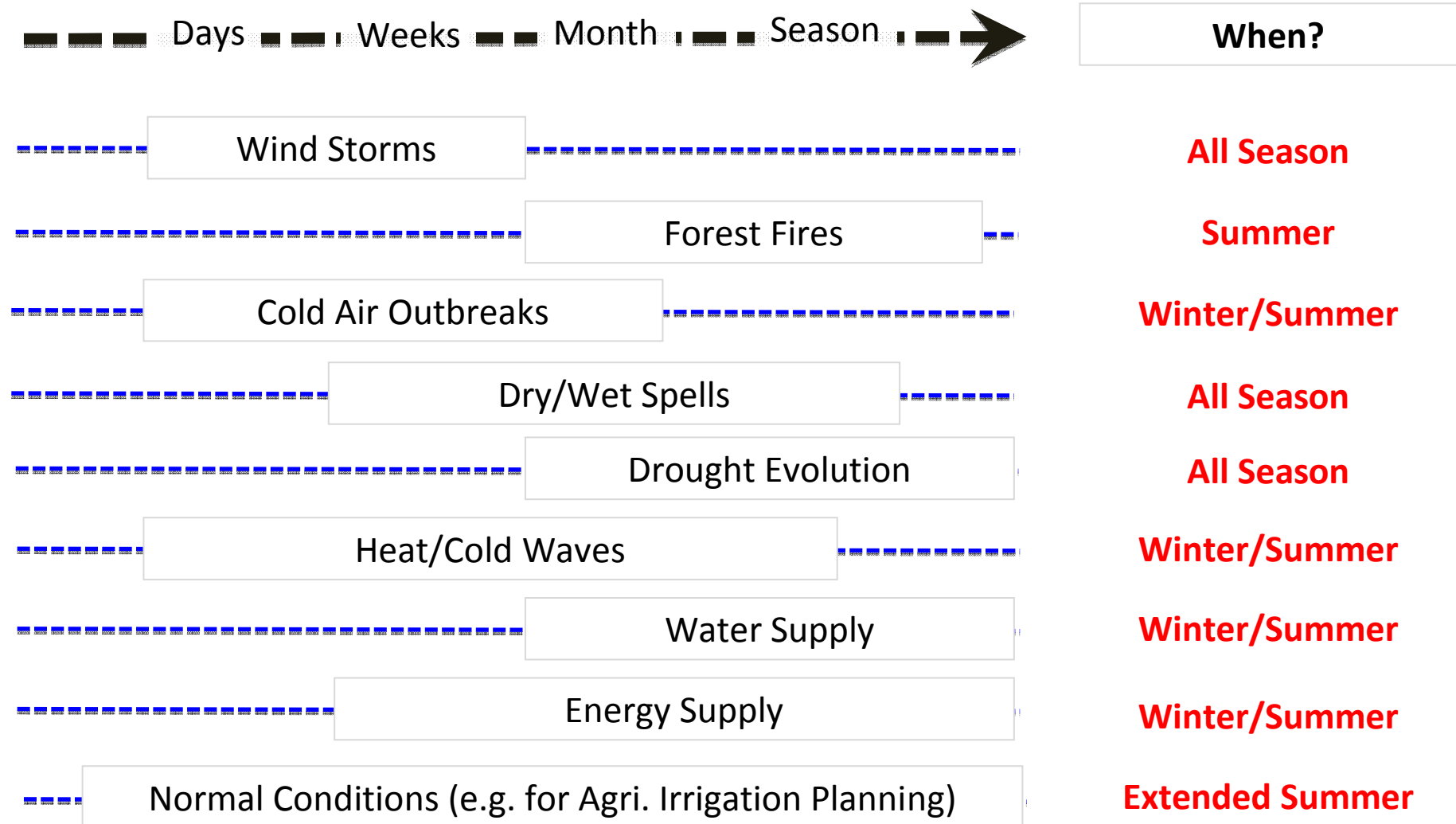
Nov 2006 – Jan 2007



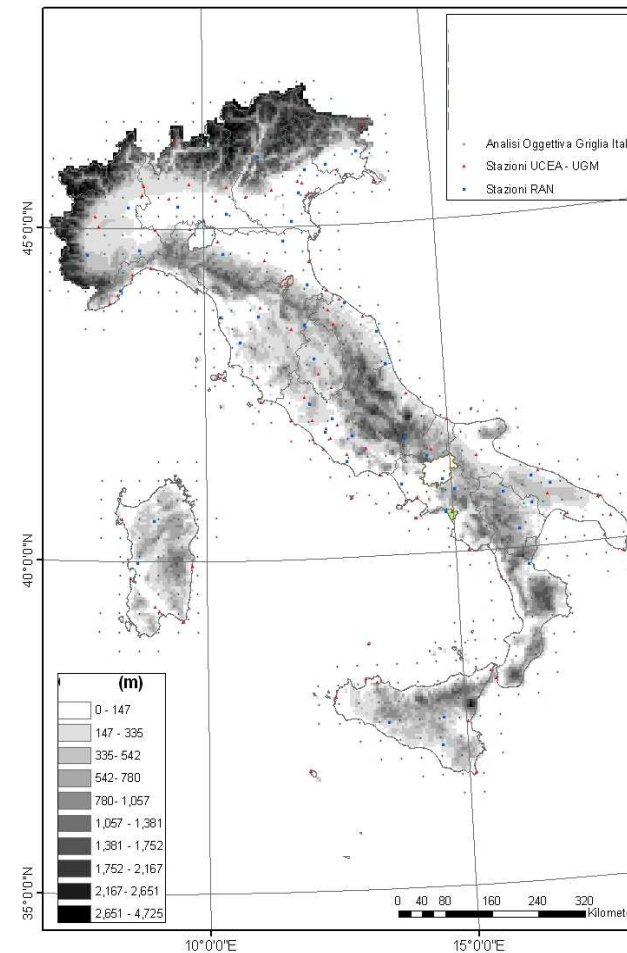
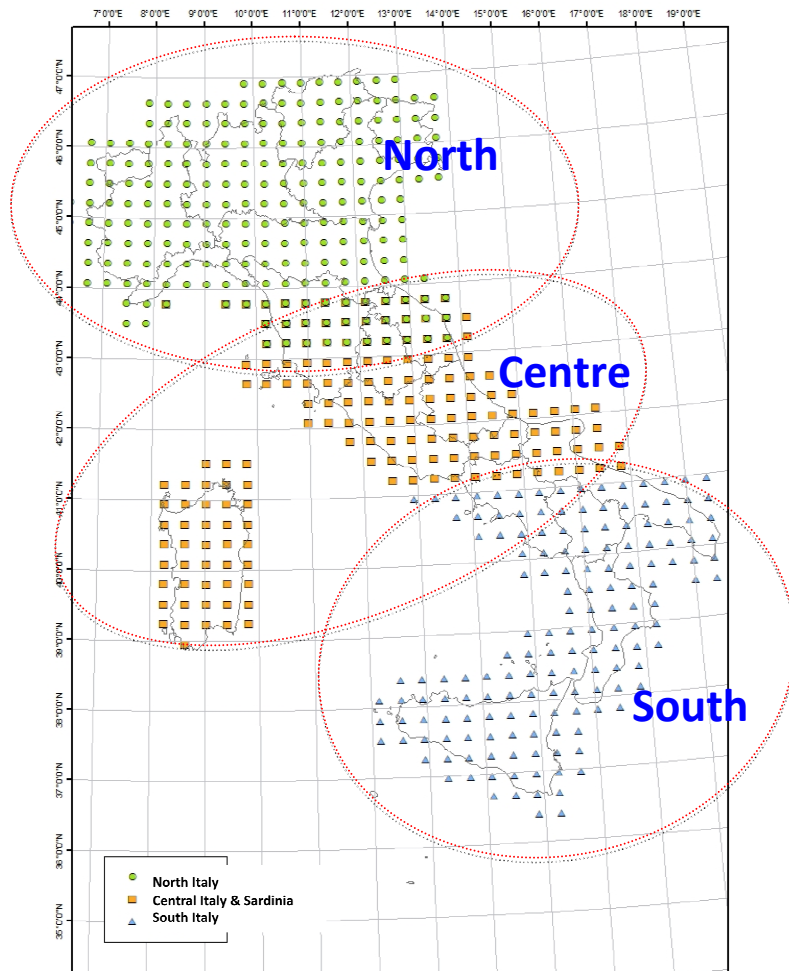
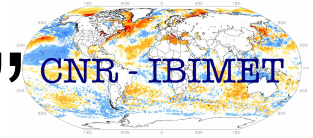
(\*) Luterbacher J.; Liniger, M. A.\*; Menzel, A.; Estrella, N.; Della-Marta, P. M.\*; Pfister, C.; Rutishauser, T.; Xoplaki, 2007 Exceptional European warmth of autumn 2006 and winter 2007: Historical context, the underlying dynamics, and its phenological impacts Geophys. Res. Letter, 34, L12704



The Italian Civil Protection started a monthly to seasonal permanent forecast board since the (Extreme Dry/Warm) Winter 2006-2007 with monthly meeting regarding major critical environment and planning issue: *next three months along with a focus on the following month.*

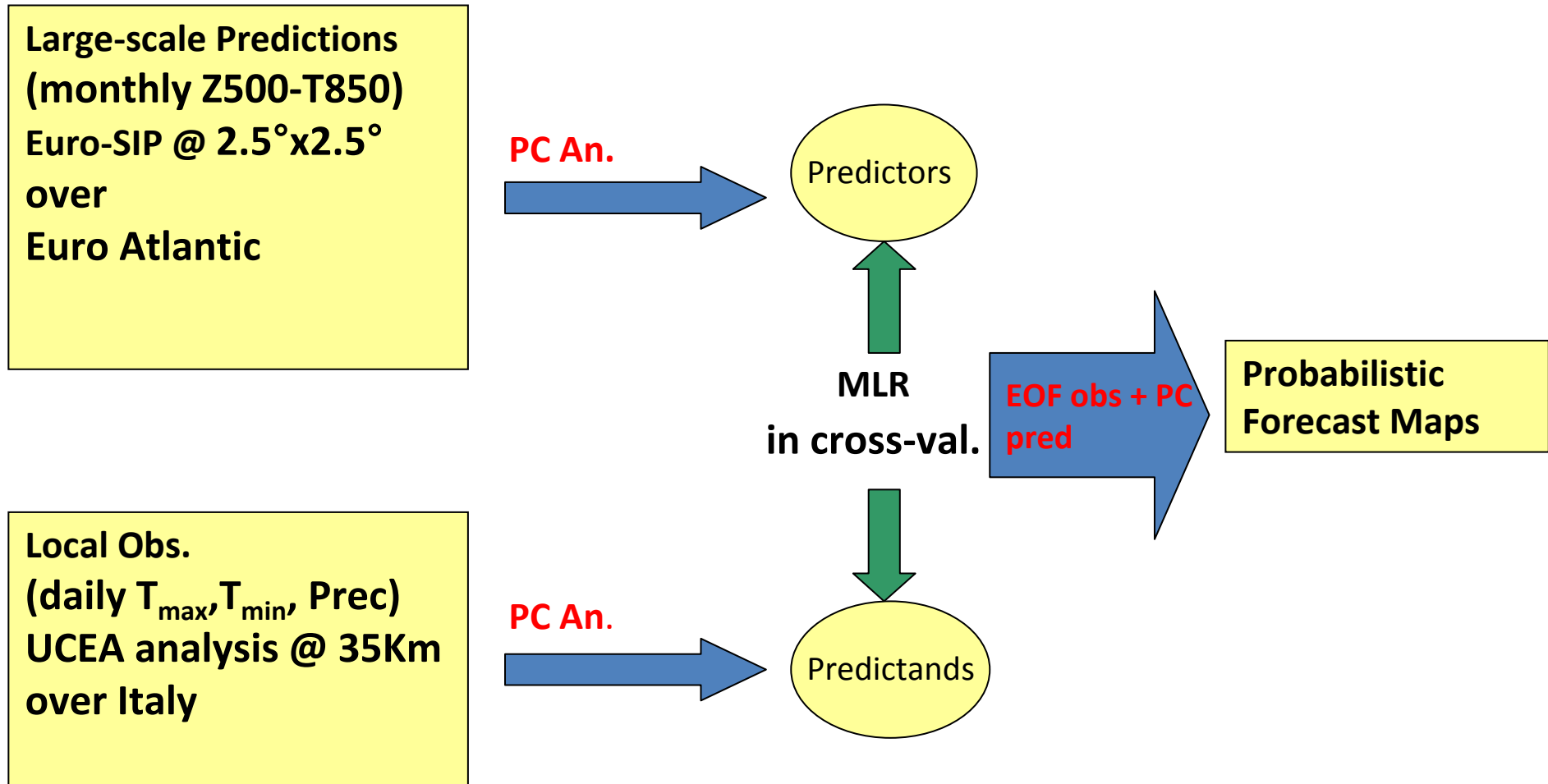


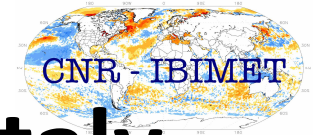
# A reference “ground truth”



- Three reference areas were established according to the DPC needs of observation, long term forecast and a consensus basis, thus not purely climatic.
- The CRA-CMA Analysis data-set consists of an objective analysis of station data (AM & CRA-CMA) covering the whole Italian territory with a regular 35Km grid from 1951 to present and includes daily minimum, maximum temperature and precipitation.

# ARPA-SIMC Calibration Scheme



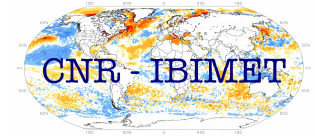


# Skill of $T_{\max}$ JJA predictions over Italy

Results refer to ENSEMBLES-STREAMII multi-model seasonal forecasts (1971-2005)

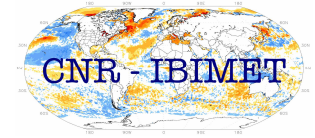
Preprocessing	DMO		Calibrated	
	Cor	BSS	Cor	BSS
Full fields	0.38*	0.14	0.49**	0.21
Detrended	-0.01	0.004	0.27	0.11
Full fields - 2003	<b>0.45</b>	<b>-0.08</b>	<b>0.60**</b>	<b>0.17</b>
(Full fields – 2003) detrended	<b>0.09</b>	-0.03	<b>0.49</b>	0.09

Pavan and Doblas-Reyes (2013)

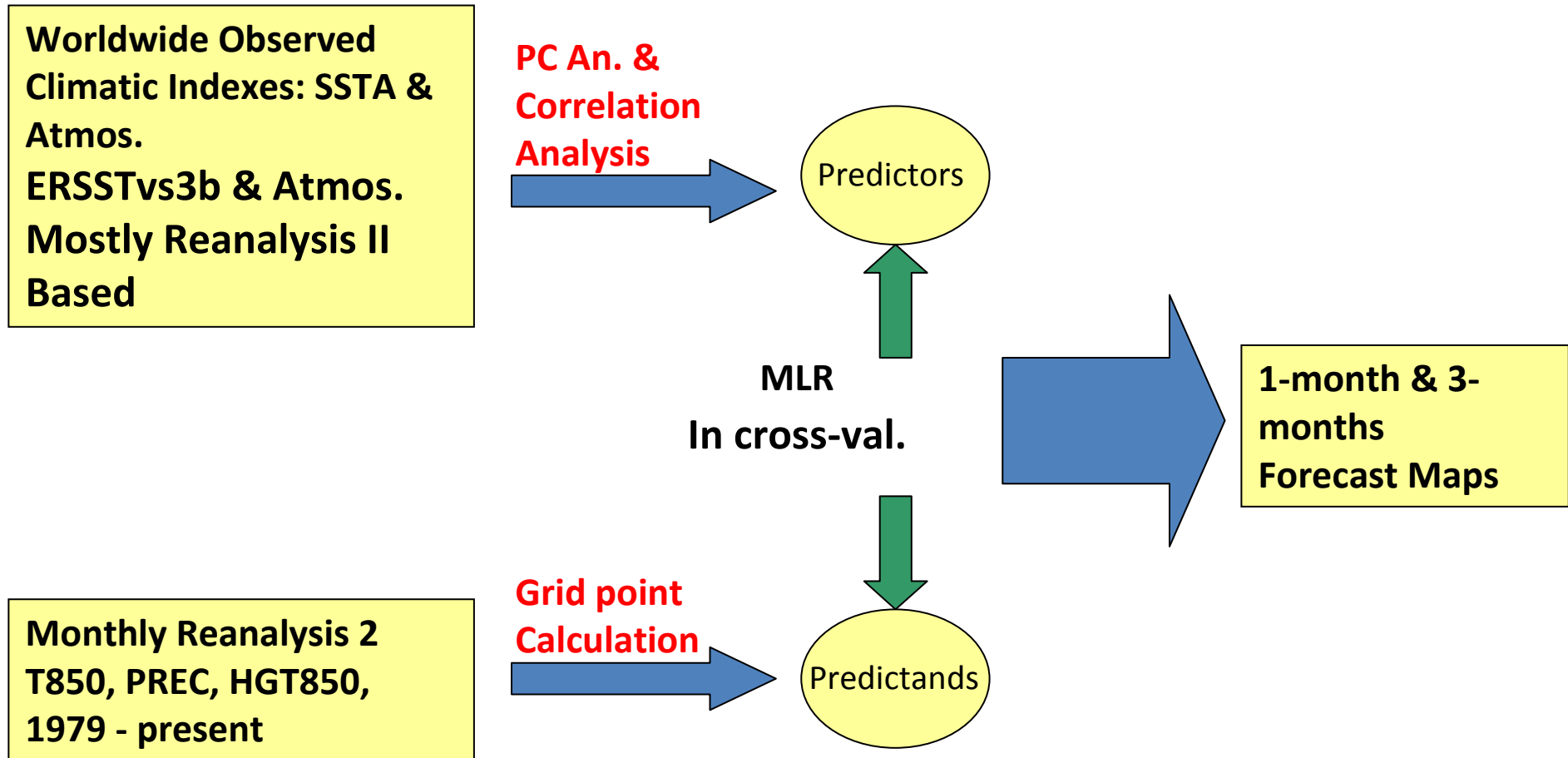


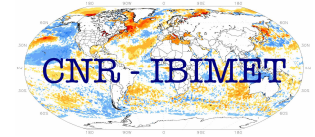
- Calibrated seasonal multi-model seasonal predictions have better skills than DMO multi-model predictions
- Over Italy and in the summer season the DMO  $T_{\max}$  predictions skill is mostly due to the correct representation of linear trends.
- Calibration can still produce skilful predictions even if trends are removed.
- The presence of extreme events in the data-set has a strong influence on the evaluation of the skill of the system and on the calibration process.





# IBIMET – Statistical Forecast System





## SeaFor Validation Results: ARPA-SIMC & IBIMET

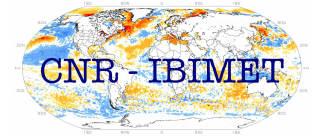
**Ranking primarily the BSS and considering POD & FAR:**

**... general considerations:**

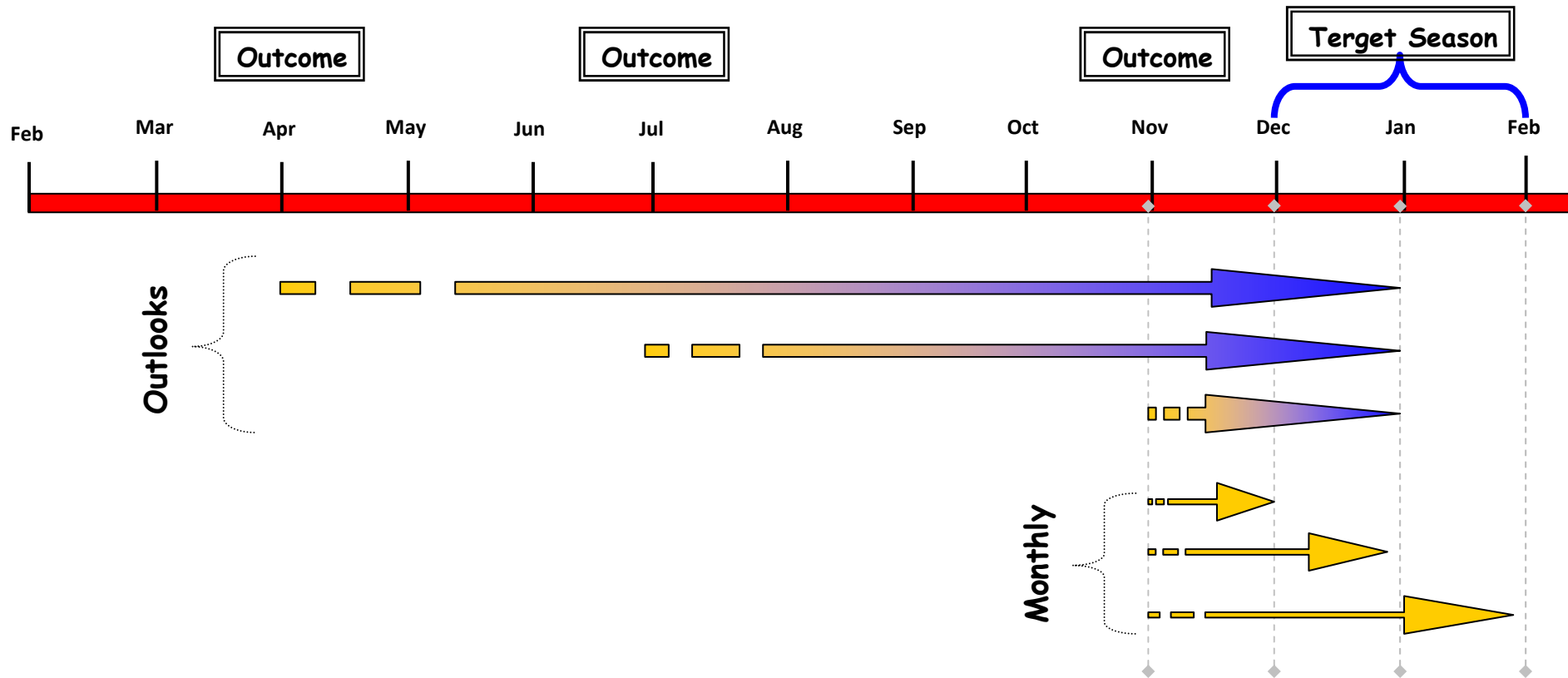
- both 2-classes & 3-classes performances are similar good;
- JJA results are slightly better than DJF;
- T is slightly better than P;
- No major differences between P and T for same quarter year, areas;
- the southern area precipitation is the most critical;

**... and specific considerations**

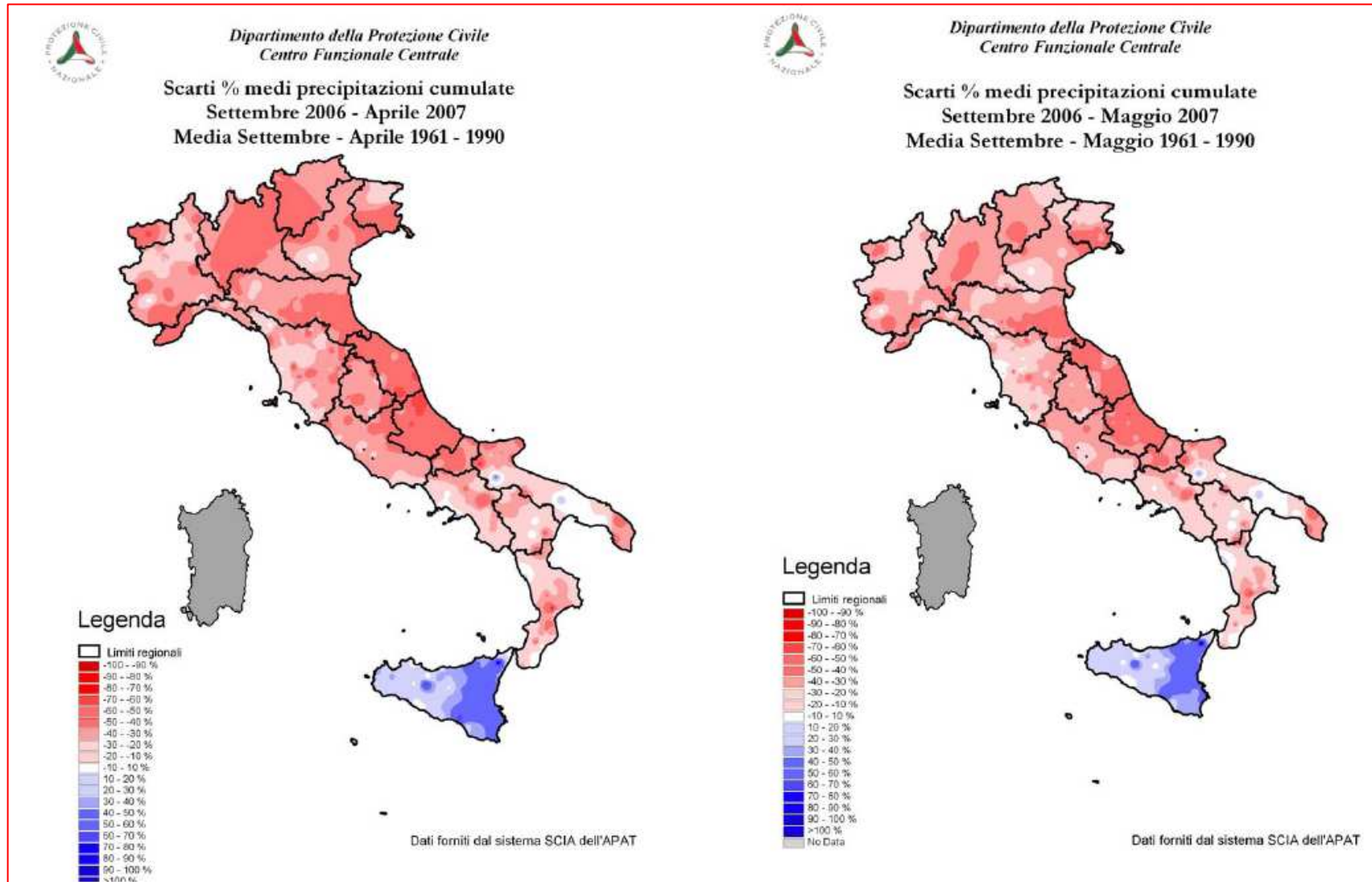
- overall ARPA-SIMC performances are clearly better than IBIMET;
- ... it is clear in DJF, less evident in JJA;
- IBIMET forecasts are poor in DJF, especially for T (even at 850hPa);
- ARPA-SIMC T&P very good for both JJA and DJF in N-C areas;

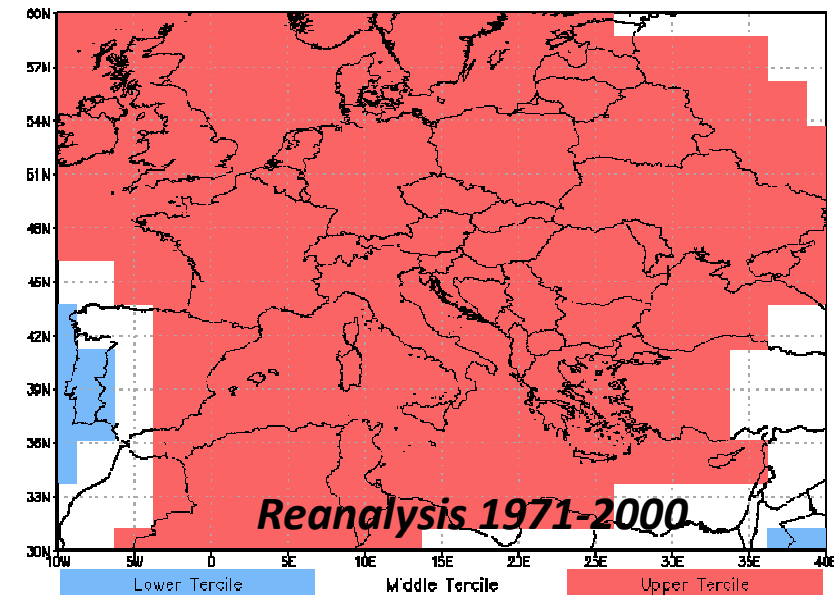
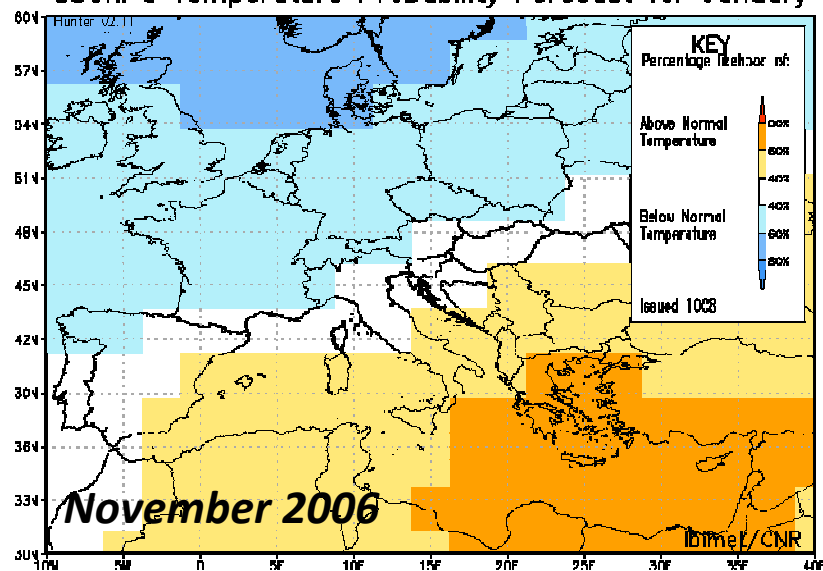
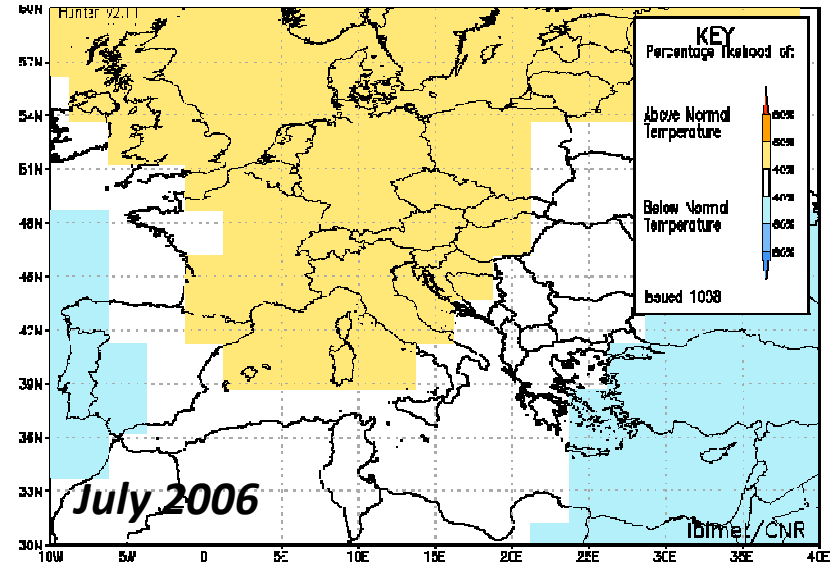
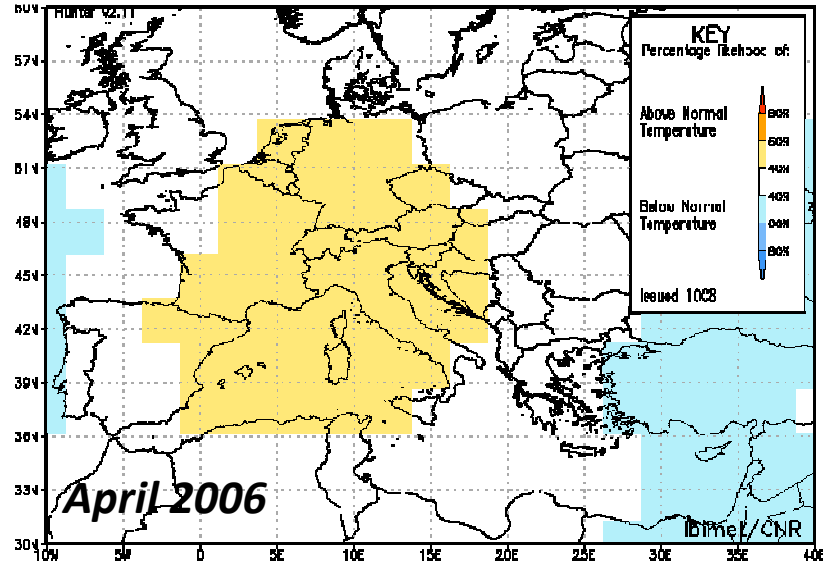


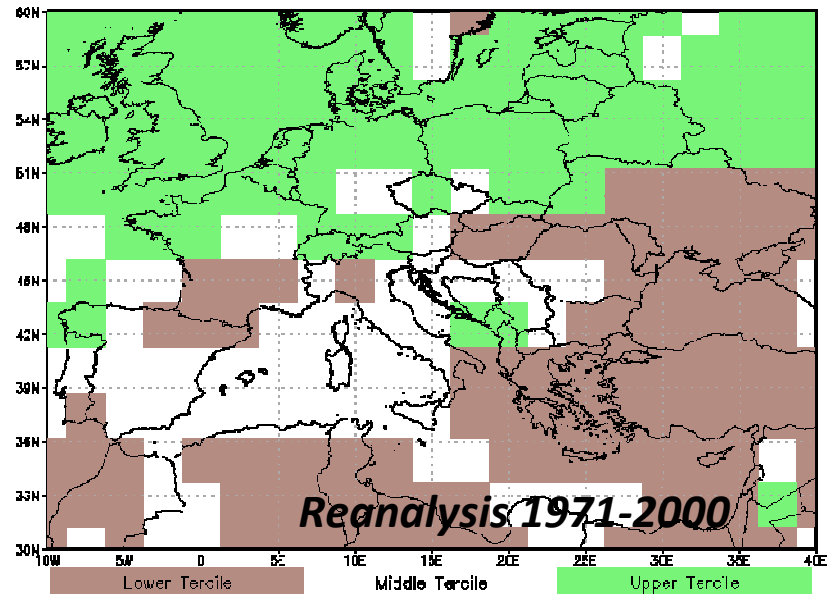
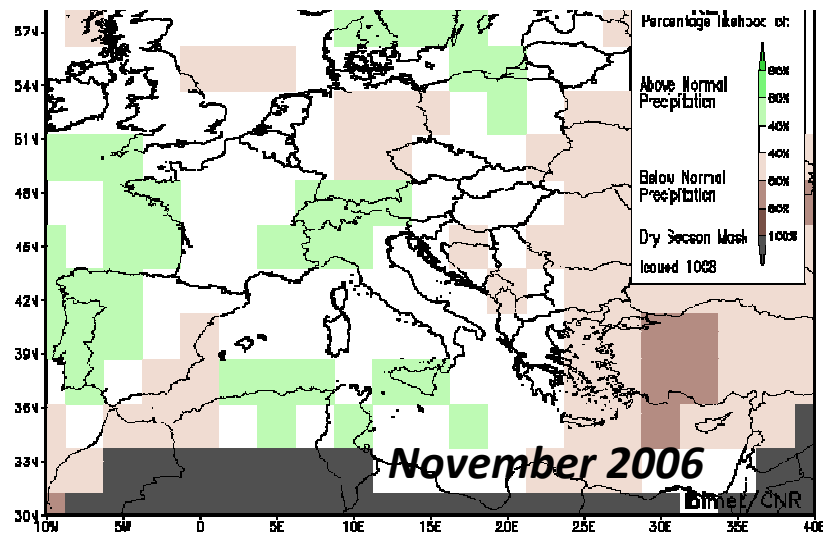
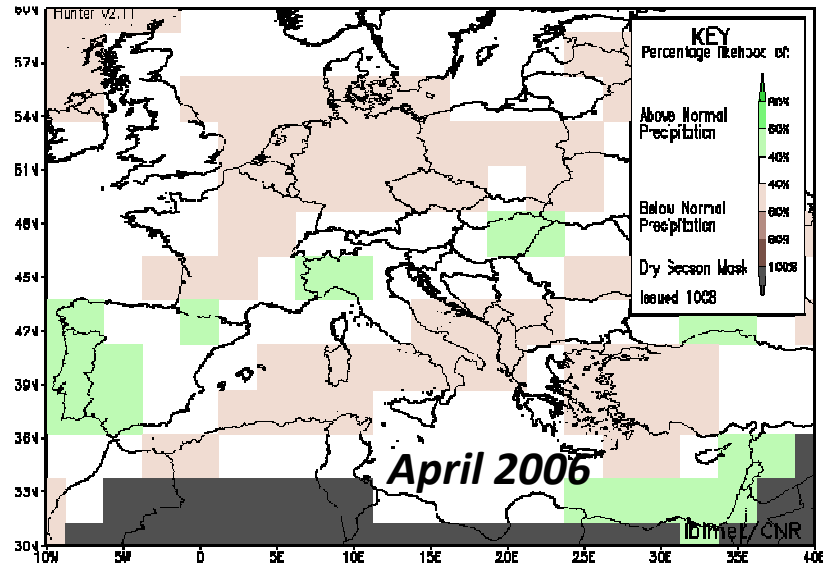
# Winter 2006/2007 Case Study



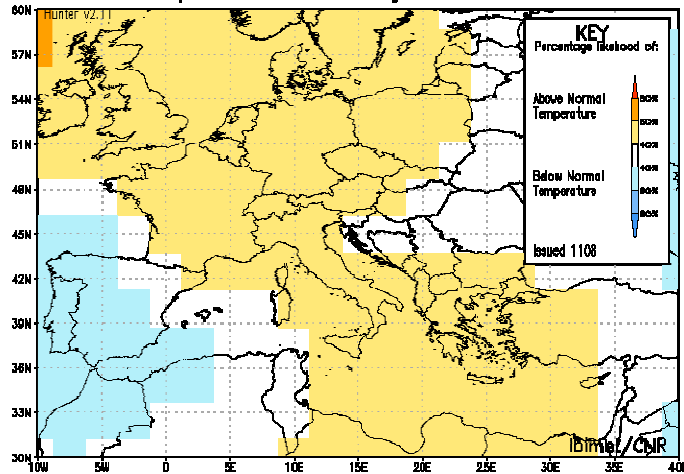
- **Outlooks:** Temp / Precip
- **Monthly:** Temp / Precip / SPI.3



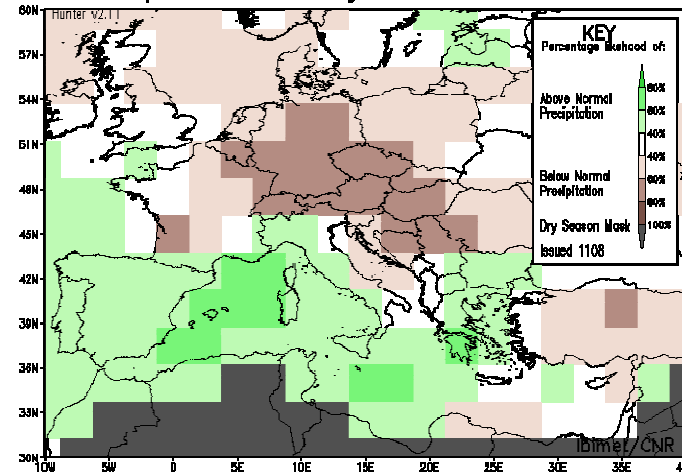




850hPa Temperature Probability Forecast for December

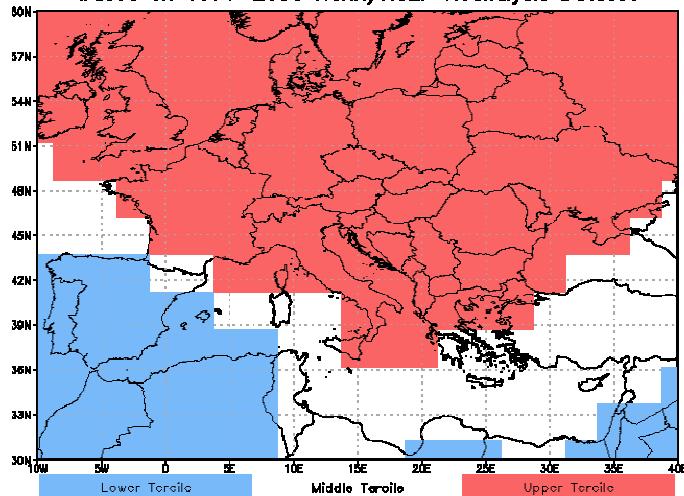


Precipitation Probability Forecast for December

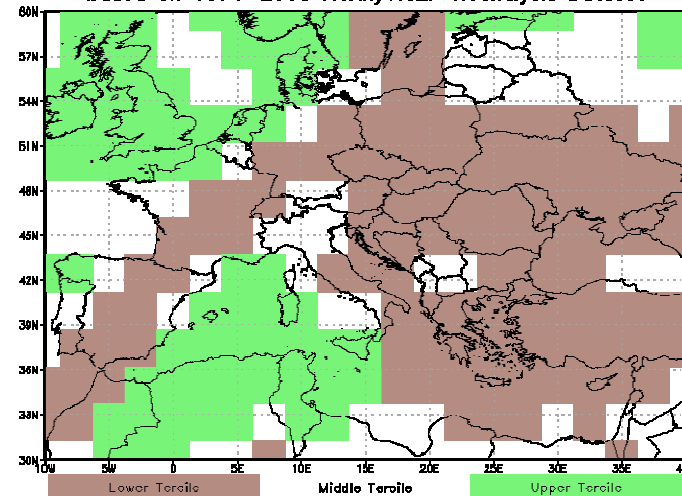


*Data from Nov 2006*

850 hPa Temperature Class for December 2006  
based on 1971–2000 NCAR/NCEP Reanalysis Dataset

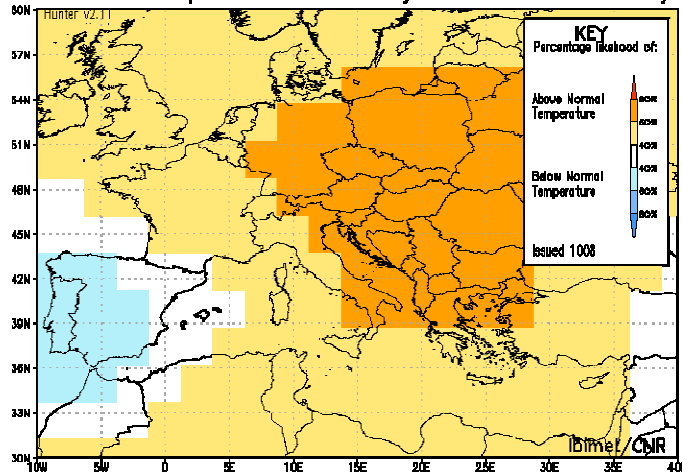


Precipitation Rate Class for December 2006  
based on 1971–2000 NCAR/NCEP Reanalysis Dataset

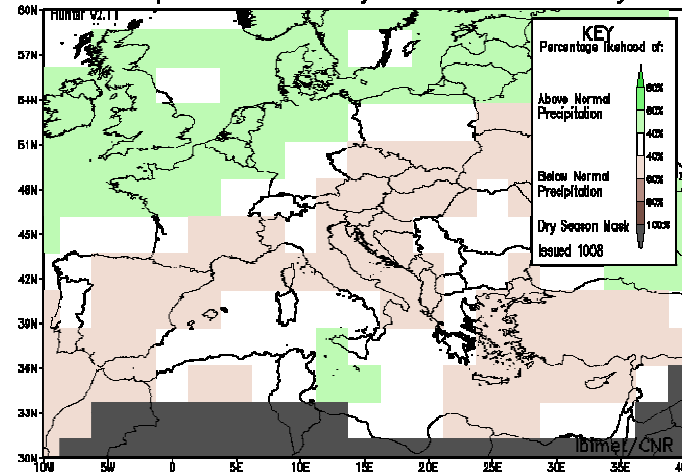


*Reanalysis 1971-2000*

850hPa Temperature Probability Forecast for January

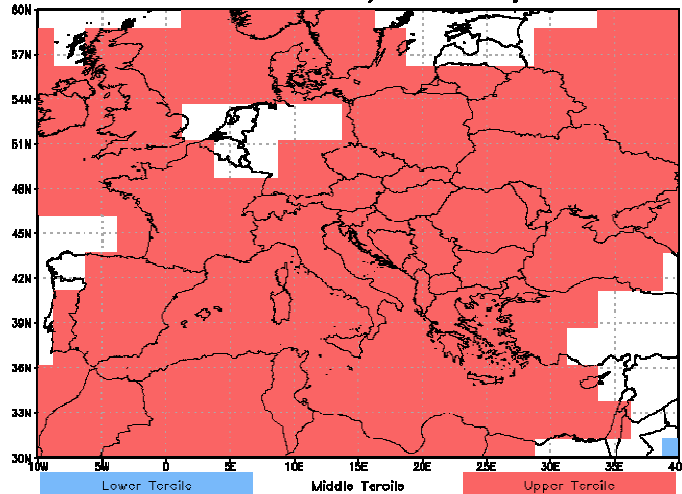


Precipitation Probability Forecast for January

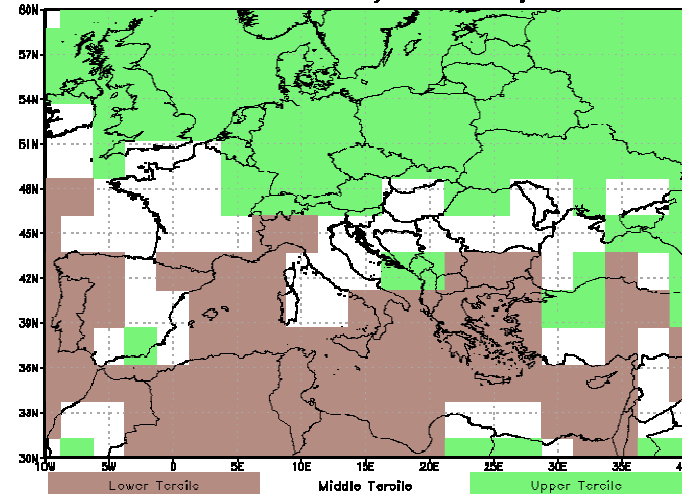


*Data from Dec 2006*

850 hPa Temperature Class for January 2007  
based on 1971–2000 NCAR/NCEP Reanalysis Dataset



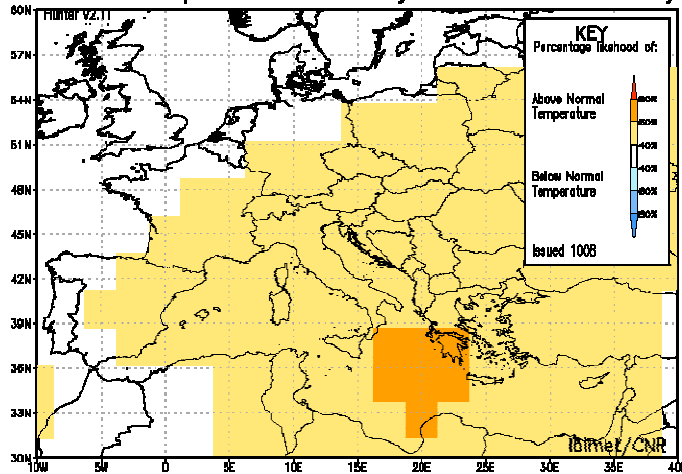
Precipitation Rate Class for January 2007  
based on 1971–2000 NCAR/NCEP Reanalysis Dataset



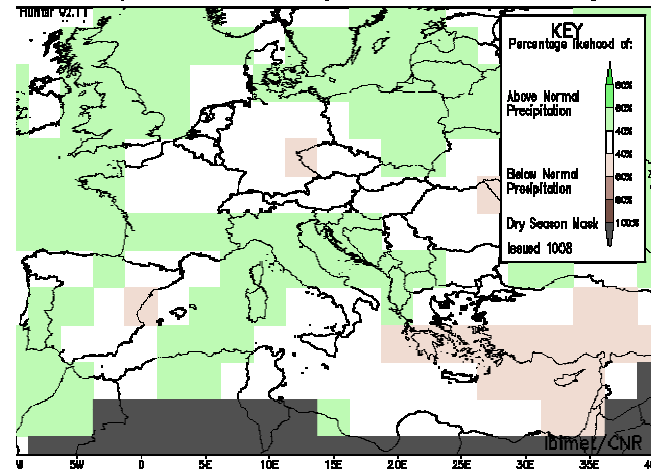
*Reanalysis 1971-2000*



**850hPa Temperature Probability Forecast for February**

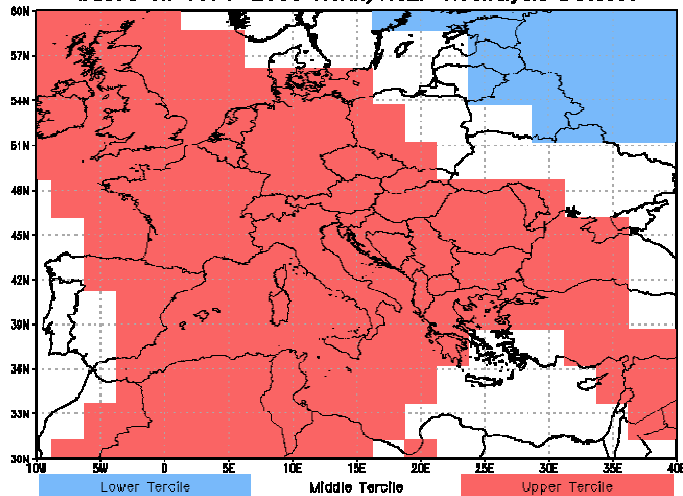


**Precipitation Probability Forecast for February**

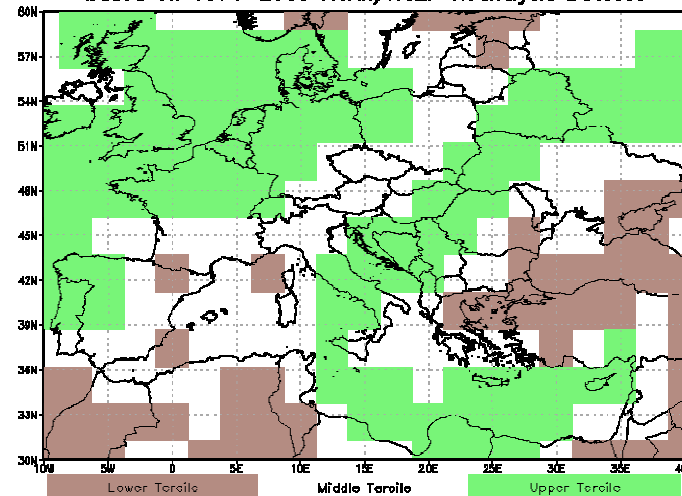


*Data from Jan 2006*

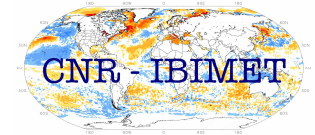
**850 hPa Temperature Class for February 2007  
based on 1971-2000 NCAR/NCEP Reanalysis Dataset**



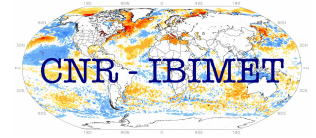
**Precipitation Rate Class for February 2007  
based on 1971-2000 NCAR/NCEP Reanalysis Dataset**



*Reanalysis 1971-2000*



- This seasonal prediction method shows encouraging results over the Mediterranean area: large scale features and correspondent anomalies location have a reliable skill level.
- The successful key for summer season seems to be the choice of physical predictors and their long range memory which is linked to the Winter North Hemisphere Snow Cover extent and Spring melting.
- In particular winter SV-NAM and the Spring Atl. Tripole & 1<sup>st</sup> EOF Guinea, seem to produce the “core” Summer variability over that area.
- Even winter atmospheric anomalies present a reliable skill level.
- Further investigations will be focused on downscaling techniques based on statistical methods.



... some sub-seasonal  
(synoptic/regime)  
characterization ...

# Motivations

- Monthly and quarterly climate anomalies, issued by seasonal forecast systems, even if reliable, could be of marginal use without an associated synoptic regime characterization.
- Circulation classifications, along with a canonical correlation analysis, represent a promising and practical way to reduce the existing temporal gap, between monthly or quarterly climate anomalies and synoptic regime characterization.

# Motivations

- Canonical (which means: reduced to the simplest or clearest possible scheme) Correlation Analysis approach is wide used in climatology and in particular it is applied, since late '80s for extracting information among predictors. (e.g. Barnett and Preisendorfer, MWR, 1987)
- Two papers, Moron et al. (part I and part II, J.Climate, 2008), an inter – annual variability characterization of Sahel precipitation based on the WT inter – annual regime analysis from historical and forecast GCMs.

# Basic Idea

- The basic idea is to find a (linear) relationship, at month or season time scale, between two daily datasets: **temperature anomalies** and a **WTC daily catalogue**.
- In order to move from daily to monthly or quarterly time scale, daily data could be classified. A possible choice is computing a class occurrence matrix.
- The seasonal forecasting goal is capability of capturing the inter and intra annual variability of atmospheric fields. Statistical methods are dealing with inter – annual variability, while global models approach mainly deal with intra – annual variability

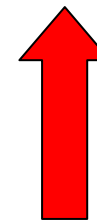
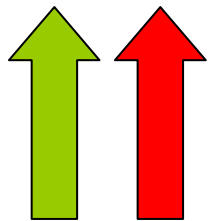
# Basic Idea

A very general hypothesis follows:

$$\text{PDF}(\text{Temp.Anom})_{\text{year}} \Leftrightarrow \text{PDF}(\text{WTC}_{\text{occ}})_{\text{year}} \cdot \text{PDF}(\text{WTC}_{\text{char}})_{\text{year}}$$

A less general hypothesis follows:

$$(\text{Temp.Anom})_{\text{year,class}} = \sum_{\text{type}} (\text{WTC}_{\text{occ}})_{\text{year,type}} \cdot (\text{WTC}_{\text{char}})_{\text{year,type,class}}$$

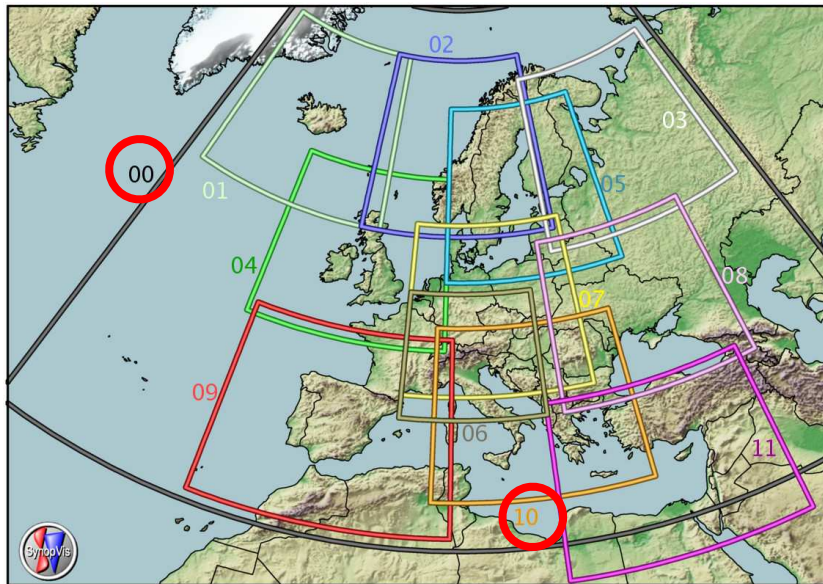


2 ways:

- CCA:  $(\text{Temp.Anom})_{\text{year,class}}$  &  $(\text{WTC}_{\text{occ}})_{\text{year,type}}$
- CCA:  $(\text{Temp.Anom})_{\text{year,class}}$  &  $(\text{WTC}_{\text{occ}})_{\text{year,type,class}}$

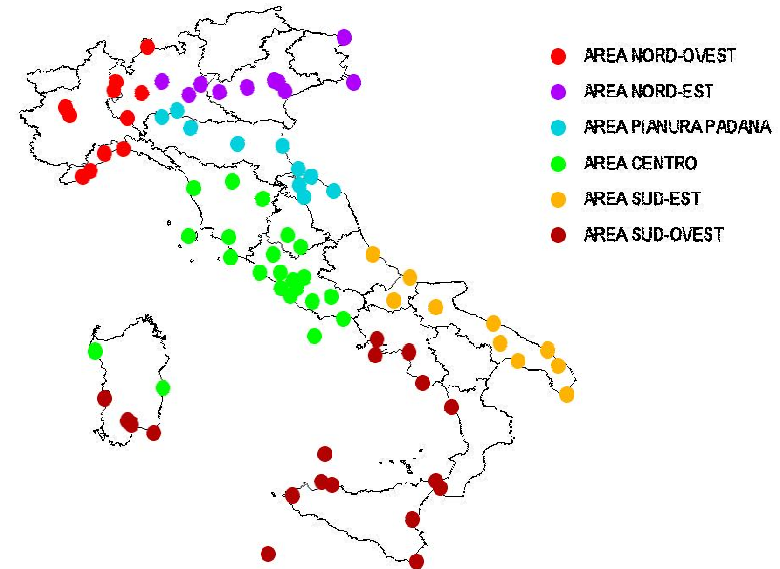
# Data

## Catalogue Candidate



Among the large number of WTC catalogues available in the COST733 repository we select one “candidate” among circulation classifications: TPCA07 (Huth 2000); and 2 domains: D00 and D10.

## Daily temp. anomalies



86 station datasets were collected for the 1961 - 2001 period. After a PCA analysis, these datasets were grouped into 6 homogeneous areas, in the framework of a seasonal forecast research project: TEMPIO.



# Canonical Correlation Analysis

Considering 2 datasets:

$$Y_1, Y_2, \dots, Y_p \text{ and } X_1, X_2, \dots, X_q$$

Construct a linear combination:

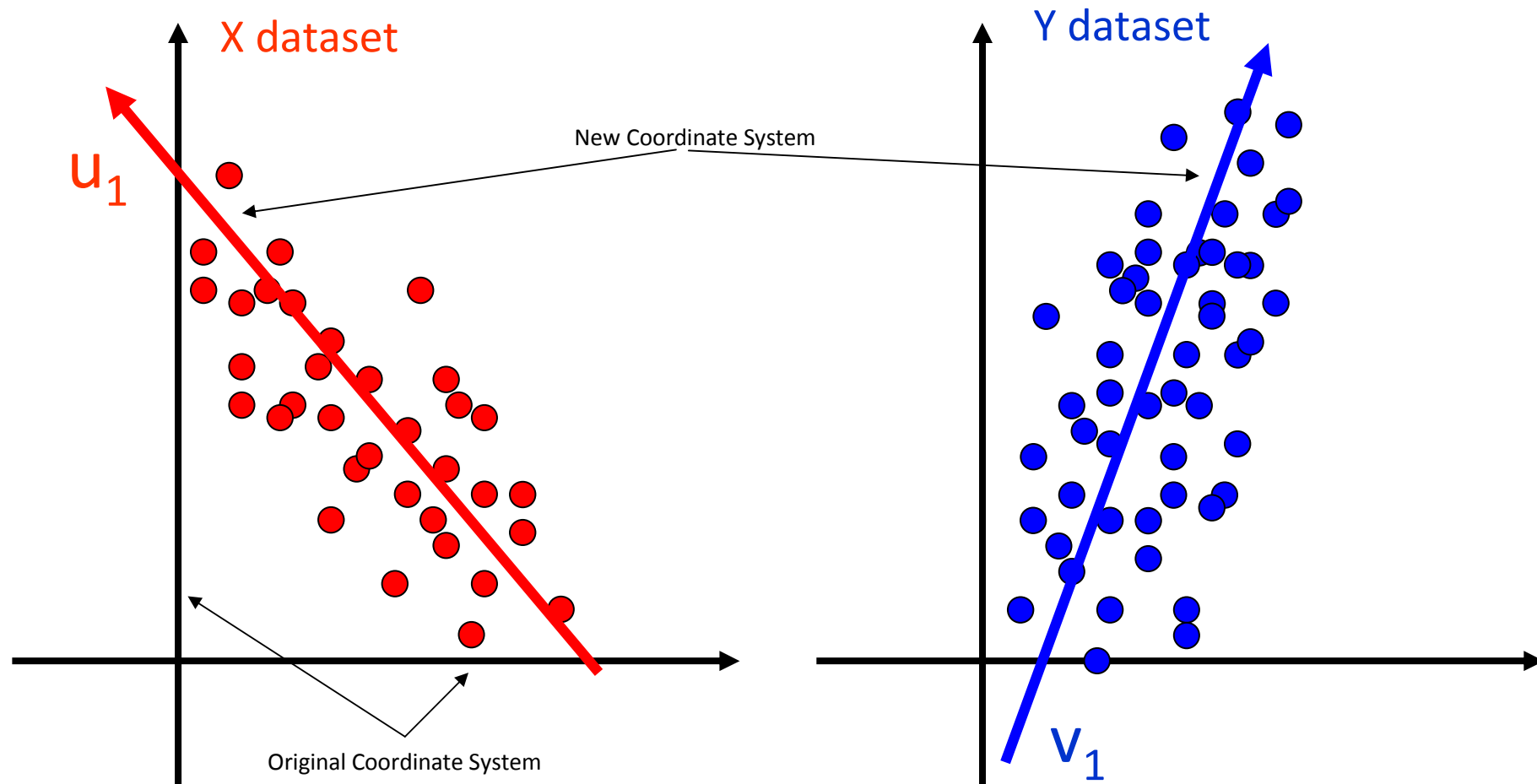
$$U = u_1 Y_1 + u_2 Y_2 + \dots + u_p Y_p \quad \text{with} \quad u' = [u_1, u_2, \dots, u_p]$$
$$V = v_1 X_1 + v_2 X_2 + \dots + v_q X_q \quad \quad \quad v' = [v_1, v_2, \dots, v_q]$$

such that correlation  $R(U, V)$  is a maximum

Let  $S$  be the variance – covariance matrix of  $X$  and  $Y$  datasets

$$S = \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}$$

# Canonical Correlation Analysis



Finding two sets of basis ( $u, v$ ) vectors such that the correlation between the projections of the variables onto these basis vectors is maximized

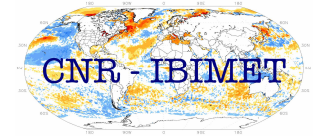
# CCA overview scheme

Original Datasets  
X and Y

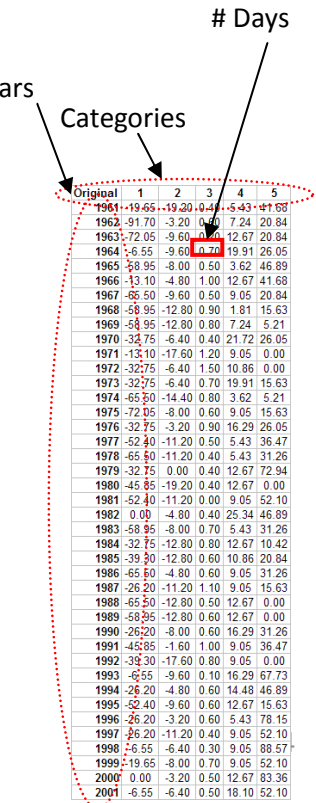
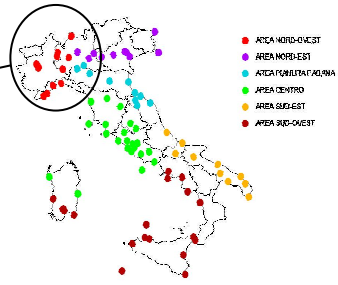
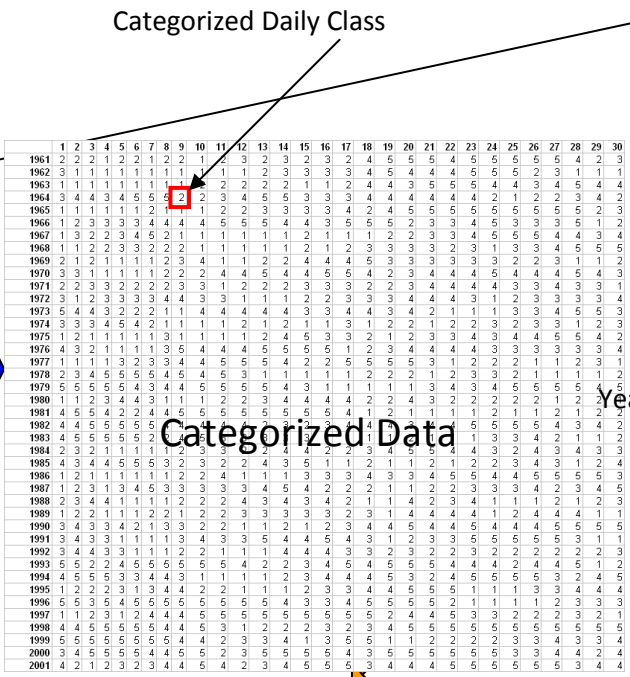
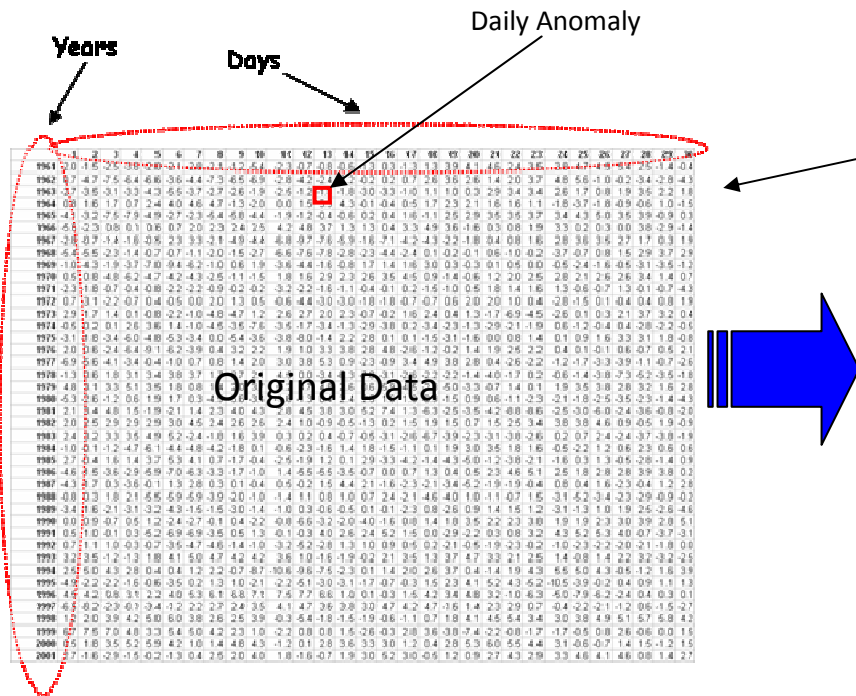
Find  $u$  and  $v$  with CCA and then  
expressing original datasets in terms of the new  
coordinate systems:  $U$  and  $V$

Express each  $V_i$  as a linear  
function of correspondent  $U_i$ :  
$$V_i = \rho_i U_i$$

Going back to the original coordinate system:  
$$Y^* = L(U_1)$$

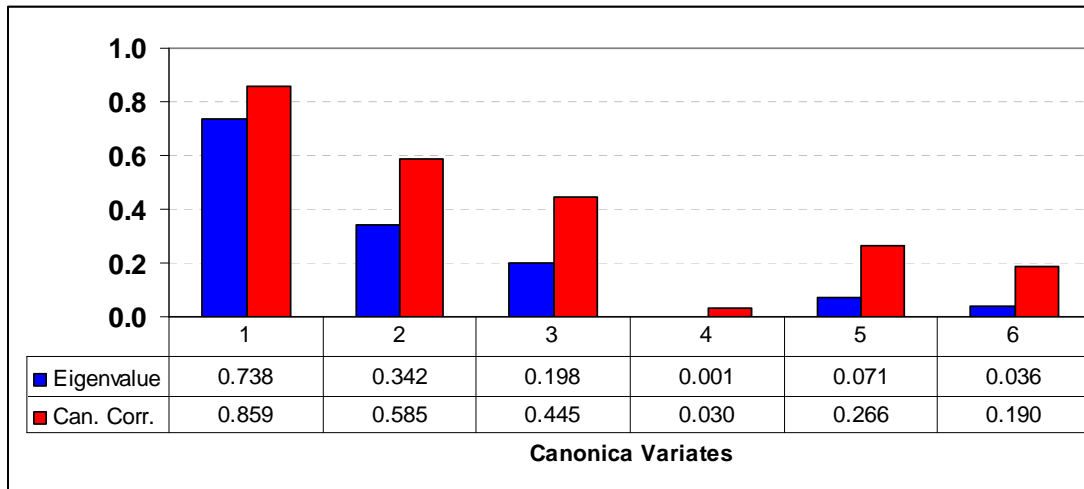


# Data Preparation



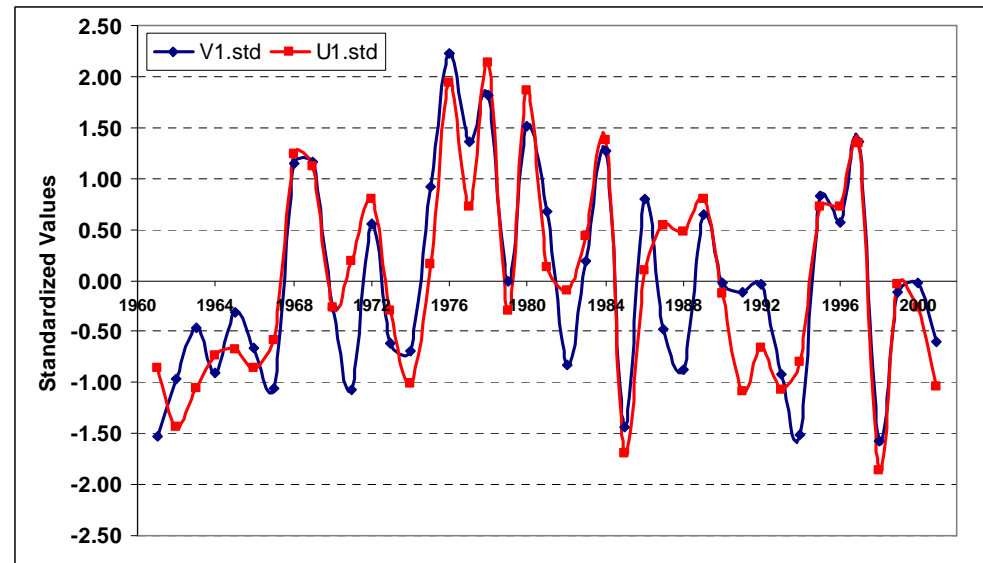
1. A 5 – days smoothed daily time series anomaly (1971 – 2000 based climatology) for each area.
2. Time series is from 1961 – 2001.
3. Reference period is JJA and June, July, August.

# CCA Results

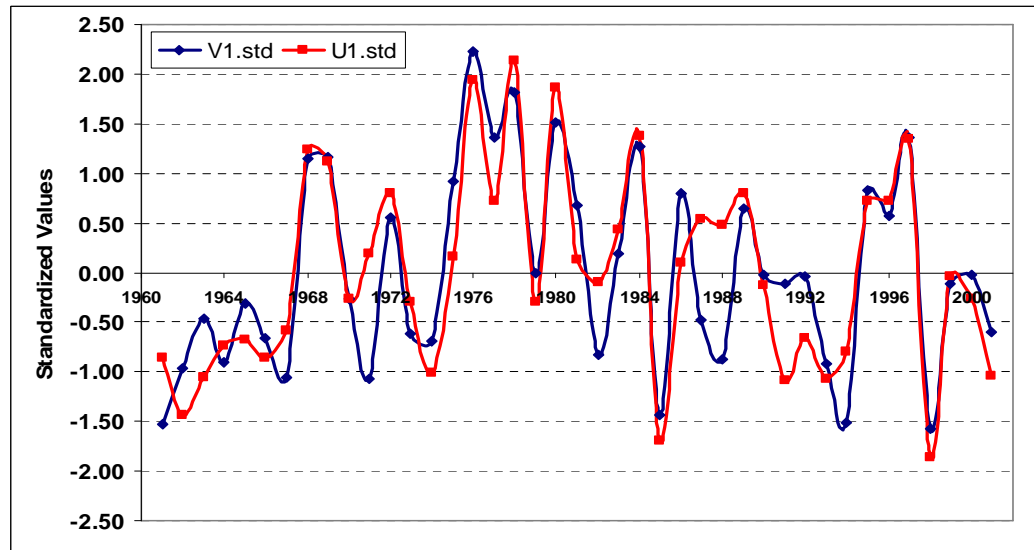


JJA – canonical variates

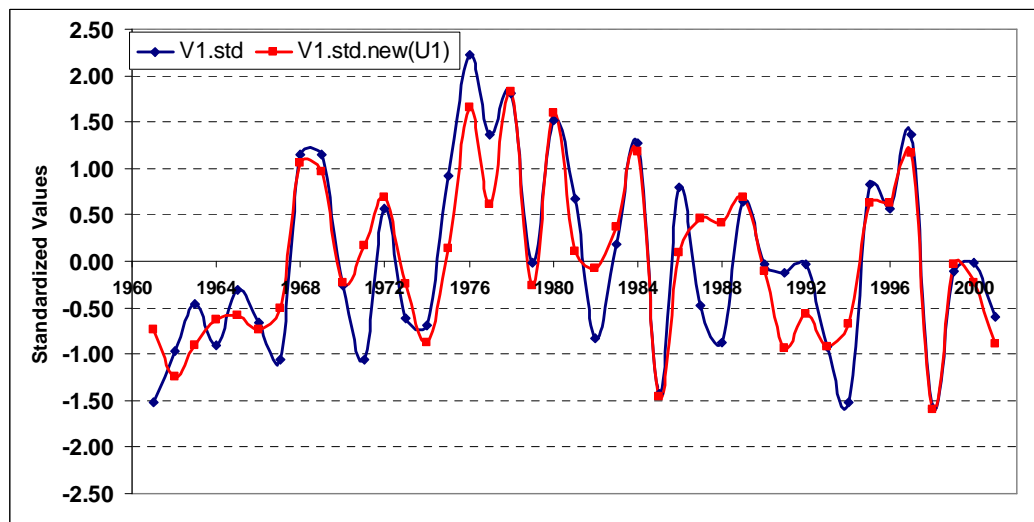
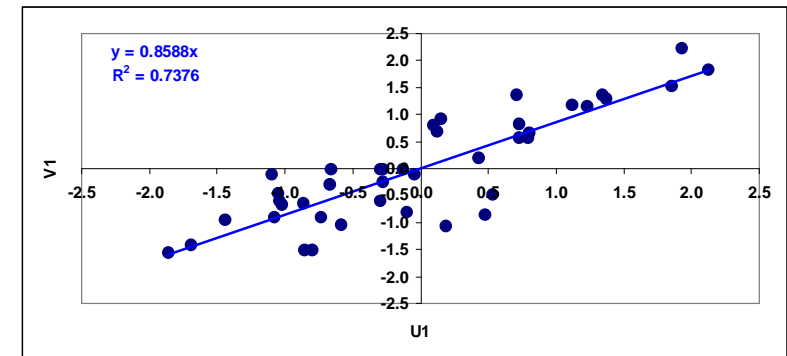
JJA – 1° pair



# CCA Results



JJA - 1°



Expressing  $V_1$  as a linear function of  $U_1$

# CCA Results

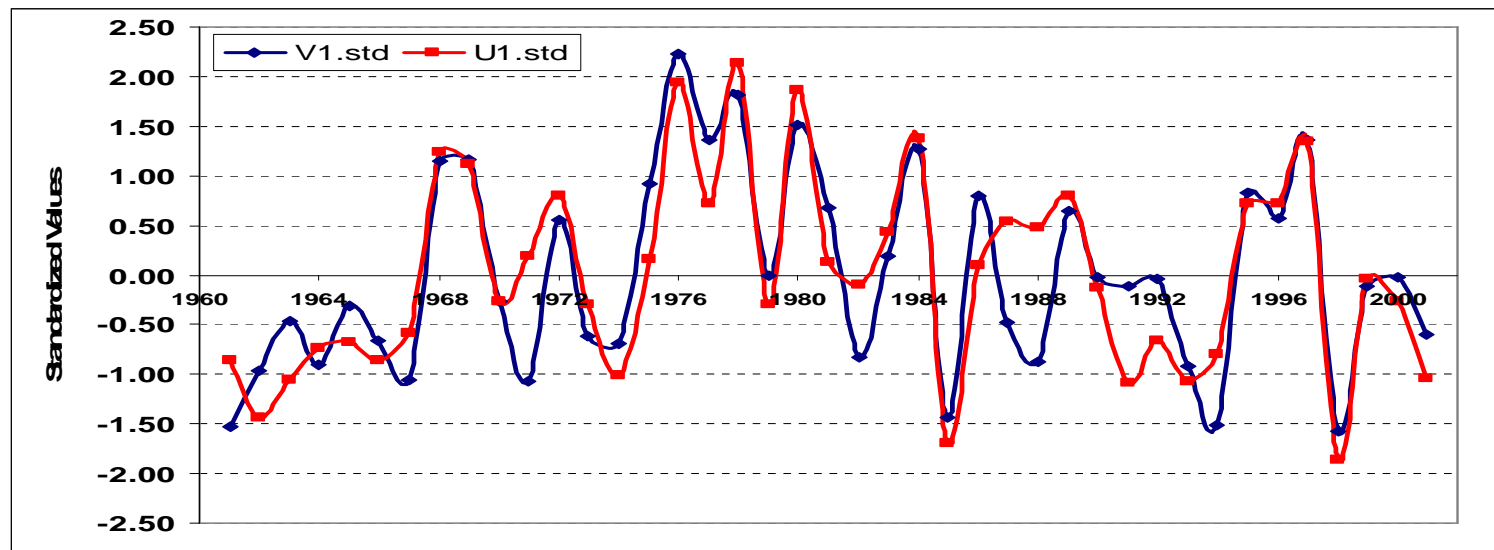
Finally, going back to the original coordinate system, we obtain the initial  $Y$  dataset in terms of the first canonical variate of  $X$  by means:  $Y^* = L(U_1)$  where  $L$  is a linear operator.

$$\begin{pmatrix} Y_{11} & \dots & Y_{1p} \\ Y_{21} & \dots & Y_{2p} \\ \dots & \dots & \dots \\ Y_{n1} & \dots & Y_{np} \end{pmatrix} \qquad \begin{pmatrix} Y^*_{11} & \dots & Y^*_{1p} \\ Y^*_{21} & \dots & Y^*_{2p} \\ \dots & \dots & \dots \\ Y^*_{n1} & \dots & Y^*_{np} \end{pmatrix}$$

We can compare each column of  $Y$  with the correspondent column of  $Y^*$

# Physical Interpretation

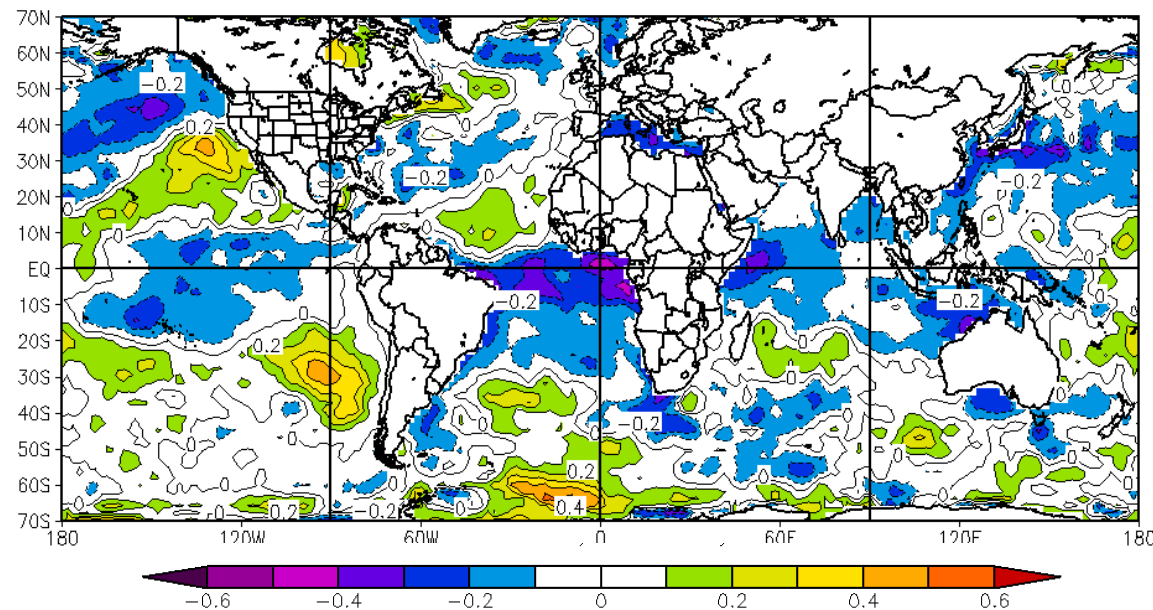
We have now a canonical vectors pair which have the highest correlation between the two datasets. In other words we can express the inter-annual variability of  $Y$  (thermal categories) in terms of  $X$  (WTC occurrence).





# Seasonal Applications

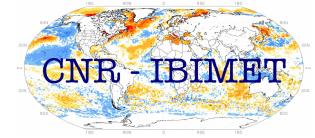
Since  $U_1$  (WT occurrence) is the “essence” of such inter-annual variability correlated with  $V_1$  (thermal categories) we can use it to explore its possible climatic sources. Some examples: SST (April) – WTC occurrence (JJA) relationship.



Thus we can forecast  $U_1$  using SST predictors.

# Seasonal Applications

1. It is possible to forecast  $U_1$ , for example with a multi regressive approach, and then going back to  $V_1$  to get the thermal distribution, instead of forecasting directly the thermal anomaly.
2. We can also reverse the strategy and associate a possible seasonal regime to an thermal seasonal forecast (coming from an independent system).
3. Since the CCA method is quite general, the  $Y$  dataset could be any classified inter – annual observables: e.g. number of rainy days of different thresholds, dry or wet spells of different duration, heat or cool waves of different duration.



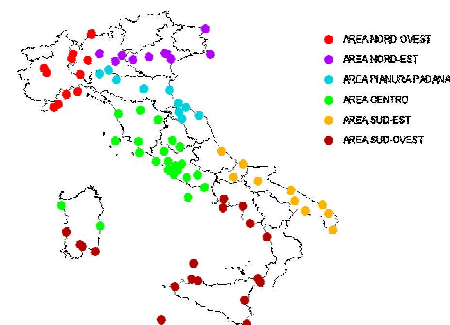
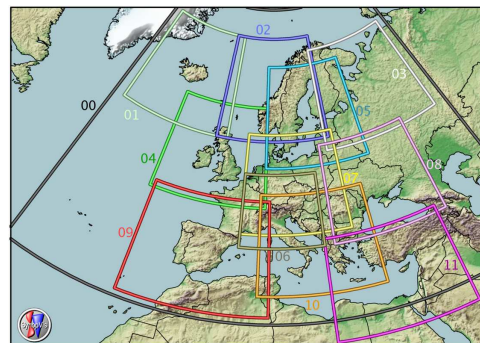
# Seasonal Applications

4. It could be applied to any climatic index after a simple phase classification in terms of strength or duration: Blocking, AO, NAO, etc.
5. Not only the inter – annual variability of WTC occurrence could be used, but also its climatological “character”, by means the thermal character of each WTC.

# CCA and COST 733

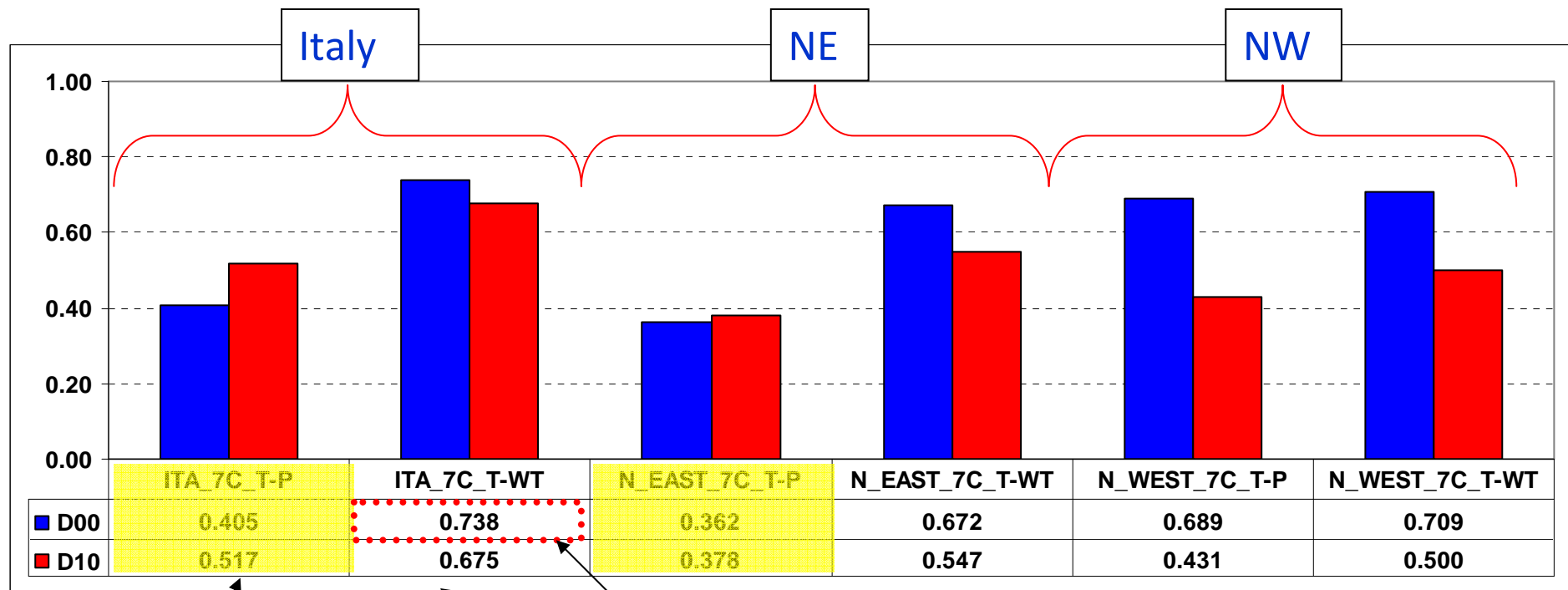
Is there any information, coming from the CCA, that could be used for “selecting” a WTC among the huge number of available catalogues? We think so! Let’s expand the previous example: for each **WTC methods** we have:

- different **number of classes** (original number plus some fixed amounts: 9, 18, 27);
- different **domains** (large, small, often superimposed);
- different **climatic areas**: Italy, North – East, North – West, etc.



# CCA and COST 733

For example applying the CCA to the same WTC, with 7 classes, but for: 2 domains **D00** and **D10**; 3 climatic areas: Italy, N\_EAST, N\_WEST; WT occurrence and characters.



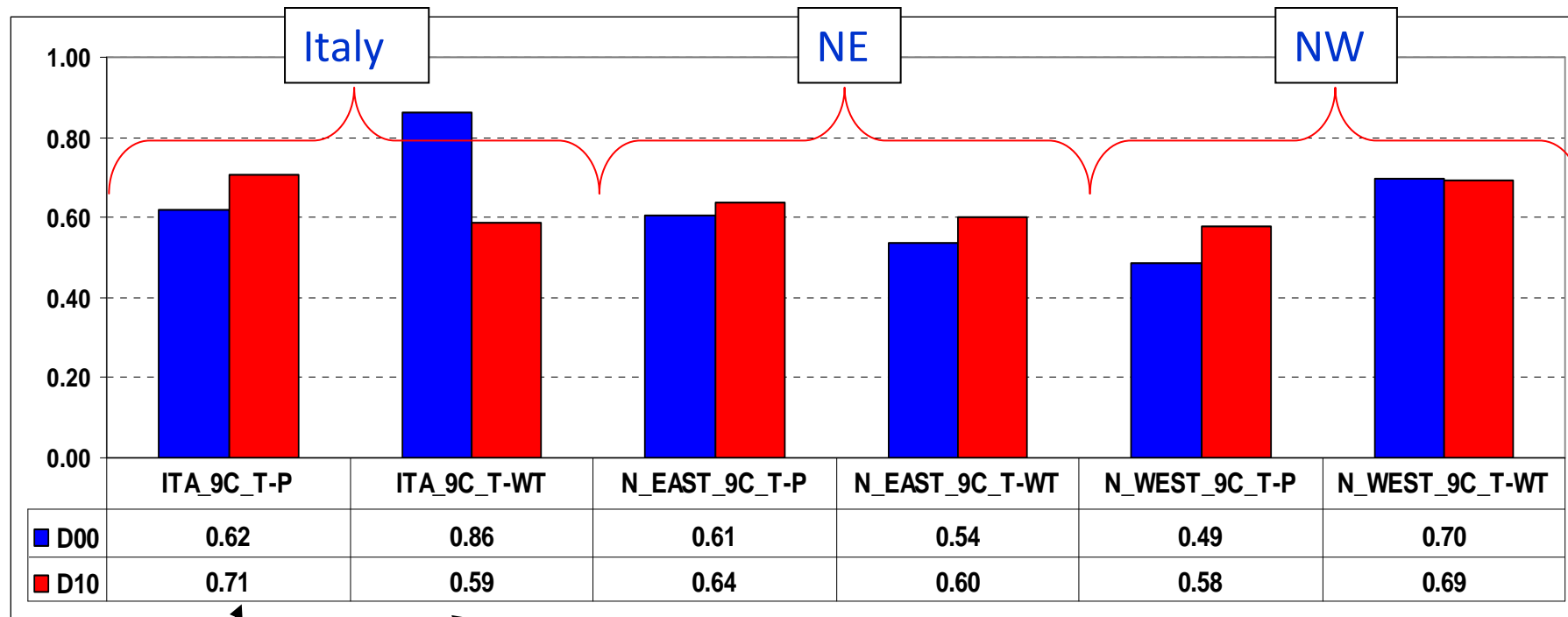
WTC<sub>char</sub> vs. WTC<sub>occ</sub>

Previous Example

JJA – 7 Classes

# CCA and COST 733

As before, but for “9” classes:

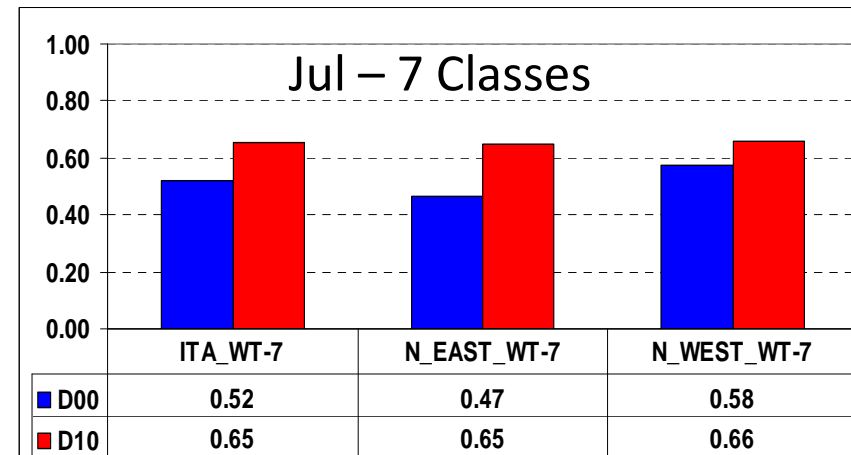
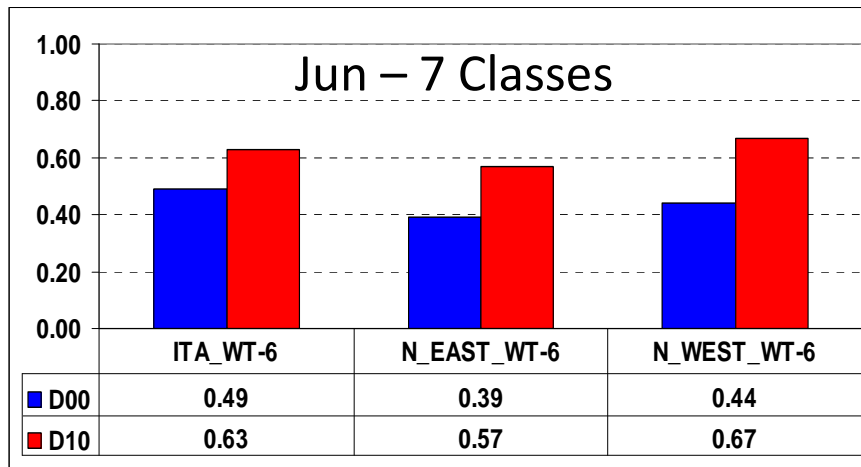


WTC<sub>char</sub> vs. WTC<sub>occ</sub>

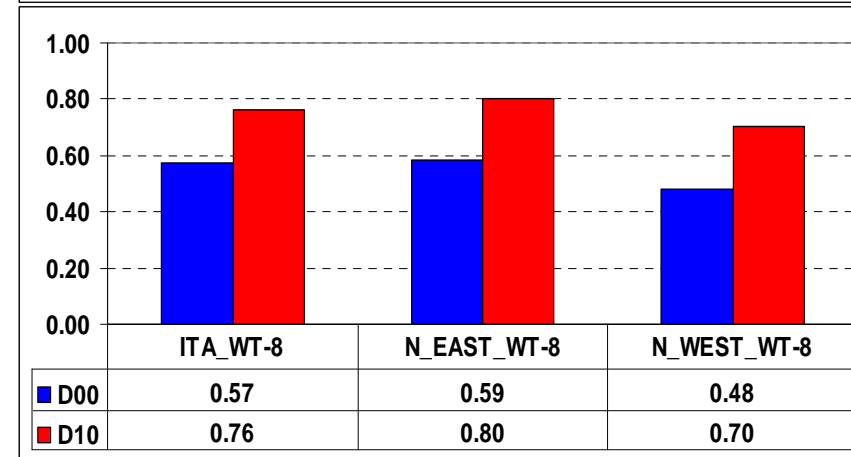
JJA ~ 9 Classes (D10=7 / D00=8)

# CCA and COST 733

Going to the monthly time scale and using the  $WTC_{occ}$  to predict the temperature anomalies:



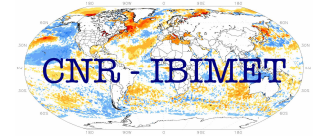
Aug – 7 Classes



1. The CCA method was applied to highlight linear correspondence between inter – annual variability of temperature anomalies and WTC occurrence/ WTC type character variability. Such correspondence exists and depends on number of classes/time scale/areas. Furthermore it is completely reversible, by means that we can linear predict one dataset in terms of the other.
2. Inter – annual WTC occurrence could be used as mid – level description tool between climatic anomalies and temperature or rainfall anomalies, in a seasonal forecast system.



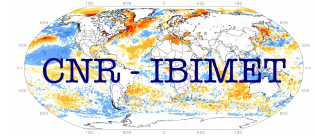
3. From the eigenvalues analysis seems possible to describe both the seasonal and the monthly timescale inter – annual variability. This is quite important because the seasonal forecasting systems based on global models are not able to distinguish among different months in the target season.
4. More catalogues was analysed with CCA and will be available for interested people.



# Thank you for your attention!

## Some References and further reading:

- Barnett, T., and R. Preisendorfer, 1987: Origins and Levels of Monthly and Seasonal Forecast Skill for United States Surface Air Temperatures Determined by Canonical Correlation Analysis. *Mon. Wea. Rev.*, 115, 1825–1850.
- Moron, V., A.W. Robertson, M.N. Ward, and O. Ndiaye, 2008: Weather Types and Rainfall over Senegal. Part I: Observational Analysis. *J. Climate*, 21, 266–287.
- Moron, V., A.W. Robertson, M.N. Ward, and O. Ndiaye, 2008: Weather Types and Rainfall over Senegal. Part II: Downscaling of GCM Simulations. *J. Climate*, 21, 288–307.
- Kirtman, Ben, Anna Pirani, 2009: The State of the Art of Seasonal Prediction: Outcomes and Recommendations from the First World Climate Research Program Workshop on Seasonal Prediction. *Bull. Amer. Meteor. Soc.*, 90, 455–458.
- Benestad, Chen and Hanssen-Bauer: A compendium for statistical downscaling (version 0-9, updated on 16 June 2007): <http://rcg.gvc.gu.se/edu/esd.pdf>
- Pavan, V. and F.J. Doblas-Reyes (2013). Calibrated multi-model ensemble summer temperature predictions over Italy. *Climate Dynamics*, 41, 2115-2132, doi:10.1007/s00382-013-1869-7.
- Gaetani M. and M. Pasqui (2012), Synoptic patterns associated with extreme dust events in the Mediterranean Basin. *Regional Environmental Change*. 10.1007/s10113-012-0386-2
- Di Giuseppe E., G. Jona Lasinio, S. Esposito, M. Pasqui, (2012) Functional clustering for Italian climate zones identification. *Theoretical and Applied Climatology* 10.1007/s00704-012-0801-0
- <http://cost733.geo.uni-augsburg.de/cost733wiki>



... and what about next NDJ ?

Bloomberg News

# If New York Freezes in January Blame Siberian Snow Now

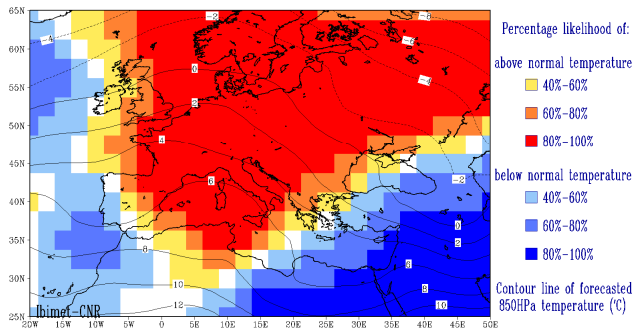
By Brian K. Sullivan | October 25, 2013

<http://www.businessweek.com/news/2013-10-24/if-new-york-is-freezing-in-january-blame-today-s-siberian-snow#p1>

NDJ 2013

Most likely category for 850 HPa mean temperature  
Forecast issued on 8/11/2013

Ibimet-CNR Seasonal Forecast  
multi-regressive model  
Hunter v2.01

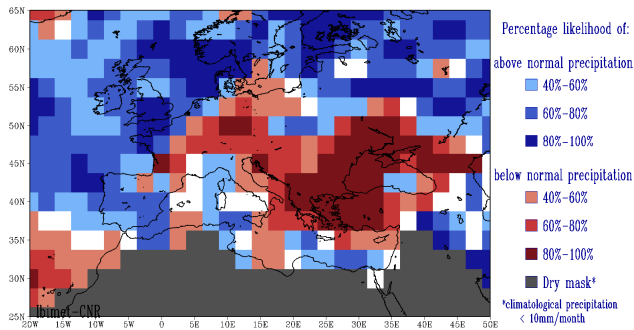


Based on NOAA NCEP-NCAR CDAS-1 monthly dataset at 2.5x2.5 spatial resolution with 1981-2010 climatological reference

NDJ 2013

Most likely category for 3-months cumulated precipitation  
Forecast issued on 8/11/2013

Ibimet-CNR Seasonal Forecast  
multi-regressive model  
Hunter v2.01

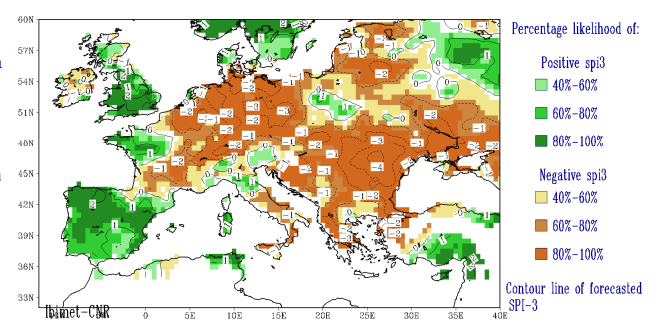


Based on NOAA NCEP-NCAR CDAS-1 monthly dataset at 2.5x2.5 spatial resolution with 1981-2010 climatological reference

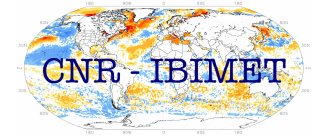
December 2013

Most likely category for SPI-3months [ECAD RainRate]  
Forecast issued on 8/11/2013

Ibimet-CNR Seasonal Forecast  
multi-regressive model  
Hunter v2.01



Based on ECAD - EObs rr monthly dataset at 0.5x0.5 spatial resolution with 1980-2010 climatological reference

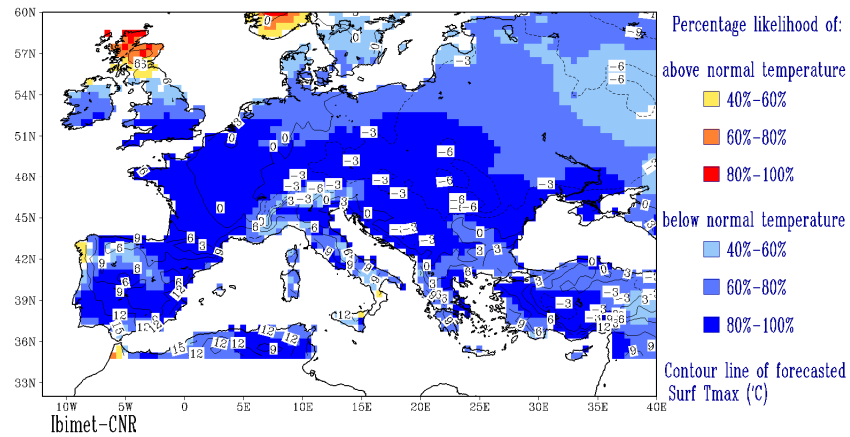


... but ... in January 2014 ...

January 2014

Most likely category for Surf Max Temperature  
Forecast issued on 8/11/2013

Ibimet-CNR Seasonal Forecast  
multi-regressive model  
Hunter v2.01

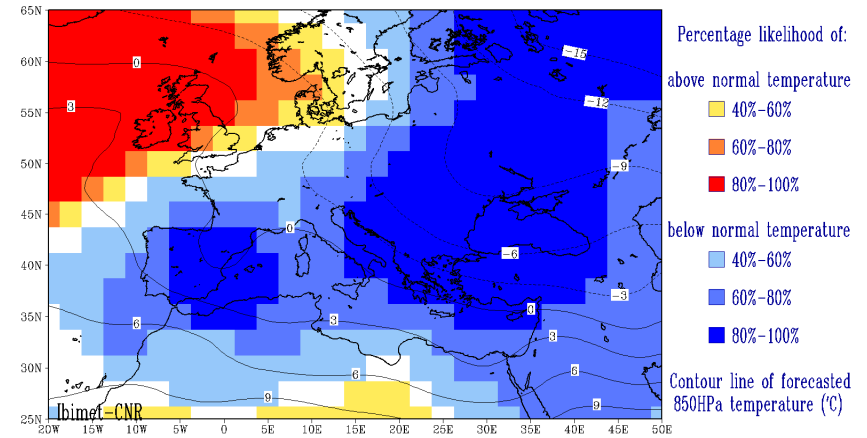


Based on ECAD monthly dataset at 0.25x0.25 spatial resolution with 1981-2010 climatological reference

January 2014

Most likely category for 850 HPa mean temperature  
Forecast issued on 8/11/2013

Ibimet-CNR Seasonal Forecast  
multi-regressive model  
Hunter v2.01



Based on NOAA NCEP-NCAR CDAS-1 monthly dataset at 2.5x2.5 spatial resolution with 1981-2010 climatological reference

... there is still room for a cold period!