www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Reading, 2 March 2016

Quality Assurance for Multi-model Seasonal Forecast Products (QA4Seas)

Francisco J. Doblas-Reyes
BSC Earth Sciences Department

Objective

QA4Seas aims at developing a strategy for the evaluation and quality control (EQC) of the multi-model seasonal forecasts provided by the Copernicus Climate Change Service (C3S) to respond to the needs identified among a wide range of stakeholders.

To achieve the objective the consortium will:

- Consider the evaluation of multi-faceted quality aspects
- Be user driven with a two-stage survey (coordinated with other lots)
- Formulate requirements to the CDS to address user requirements
- Perform a gap analysis of the current information available to users
- Develop a framework and a prototype of the EQC system

Challenges of the QA

- **Addressing adaptation**: it must provide information for climate services and short-term adaptation together because different users approach climate information, where many of them are already familiar with the climate-change problem.
- **Providing consistency**: it must provide information on which trust can be built, requiring a high degree of coherence across products, underlying data sets, processing methods, uncertainty communication, training and guidance, etc.
- **Providing innovation**: it should make operational recent developments from research, with innovative knowledge, methods and technologies to answer real-world issues.
- **Addressing efficiency**: the EQC information should be timely (e.g. available to respond to users' queries with a delay as short as possible), which imposes certain conditions on the approaches and algorithms considered to produce it.

Seasonal forecast QA

- A prediction has no real value without an estimate of its quality based on past performance.
- QA is multifaceted, no single metric is sufficiently comprehensive to characterise the quality of a system or to single-out the best forecast system.
- It addresses administrative (tracking the evolution of systems), scientific (predictability), or socio-economic (users' requirements) questions.
- QA metrics have been developed to study aspects of forecast quality including bias, uncertainty, reliability, resolution, discrimination and sharpness of the forecasts.
- Both deterministic and probabilistic approaches should be considered.
- QA of the European multi-model is not readily available, contrary to the Asian and North American one.

QA and climate services

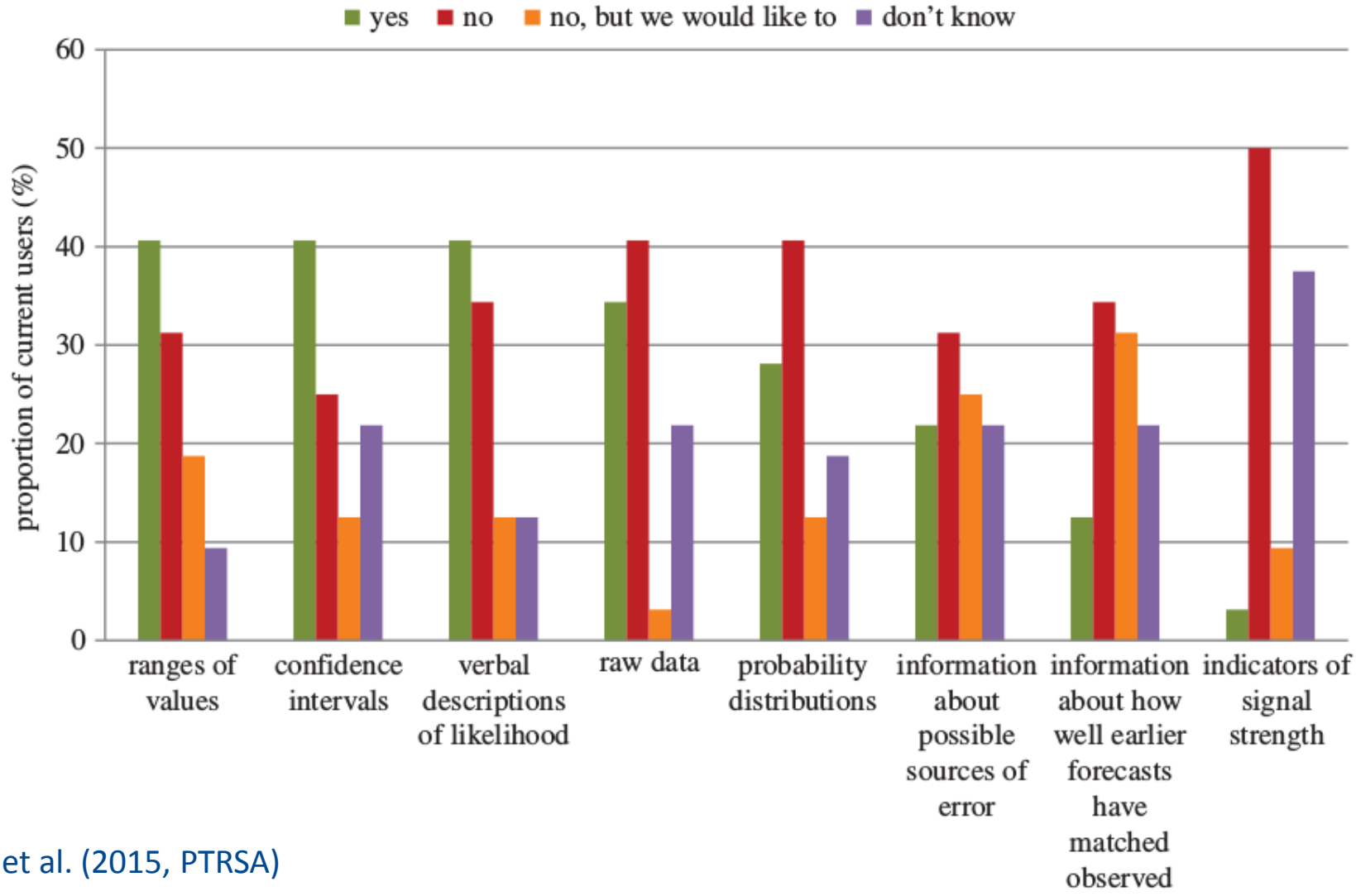
QA should take place in the context of GFCS, the European Research and Innovation Roadmap for Climate Services, and the Ethical Framework for Climate Services.

Reference points:

- Climate service providers should consider the consequences of their actions for those who may use or be affected by the use of climate service products.
- Climate service products should be open to scrutiny and comparison.

Example of gaps

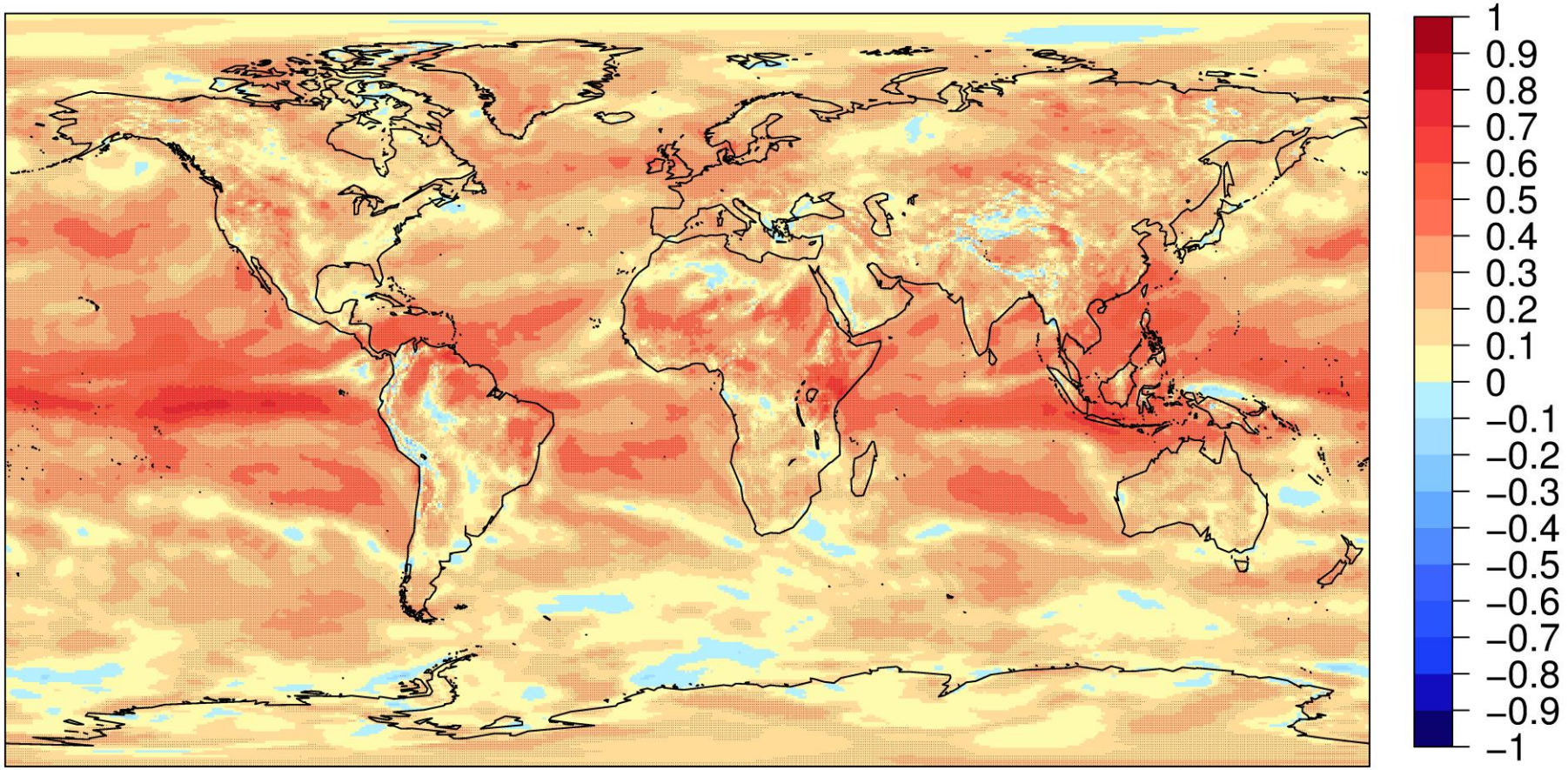
Proportion of users of seasonal forecasts (n = 32) indicating whether they received different forms of information about uncertainty.



Taylor et al. (2015, PTRSA)

QA of wind predictions

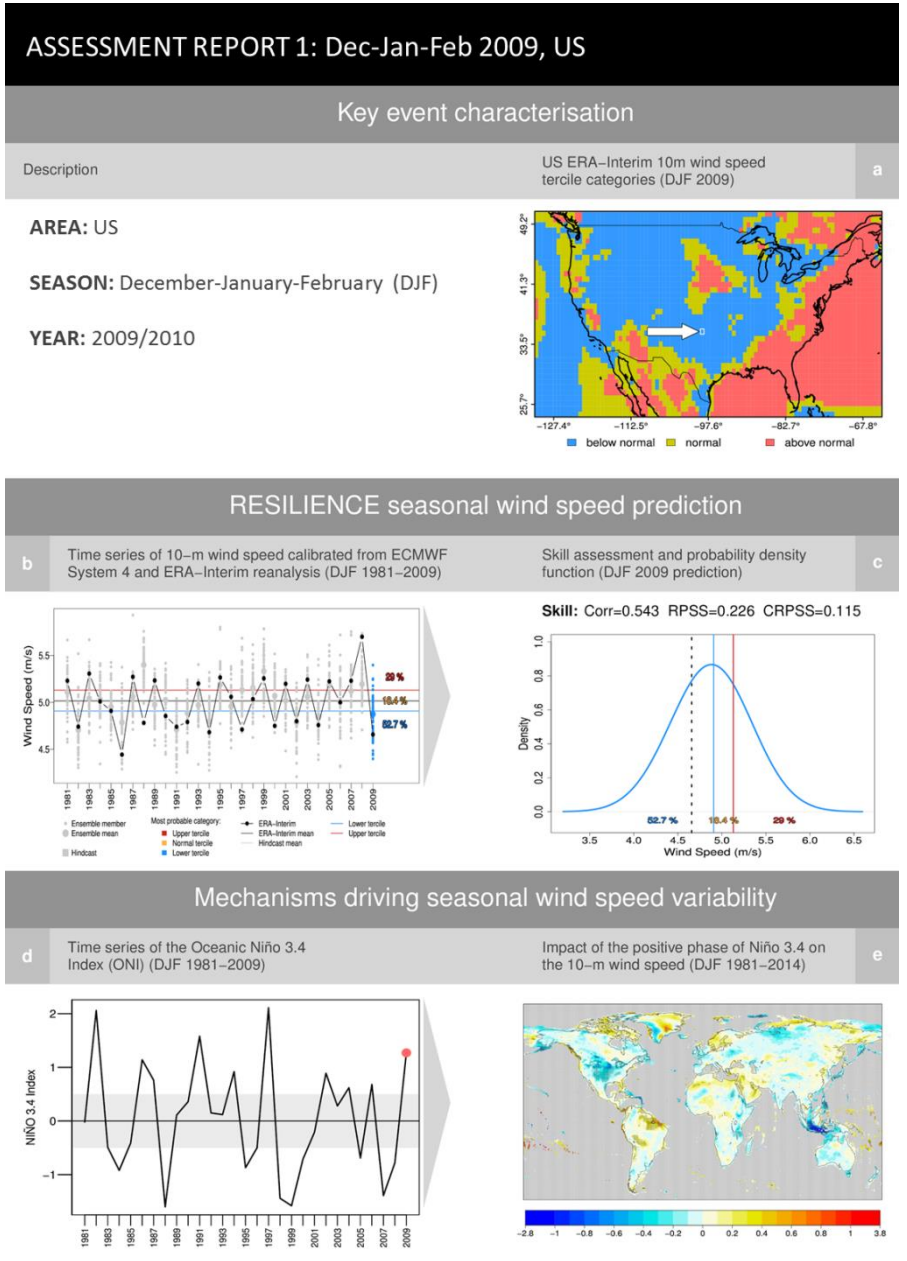
Correlation of ECMWF Monthly Prediction System
10m Wind Speed for Jan_Feb. Forecast time 12–18.



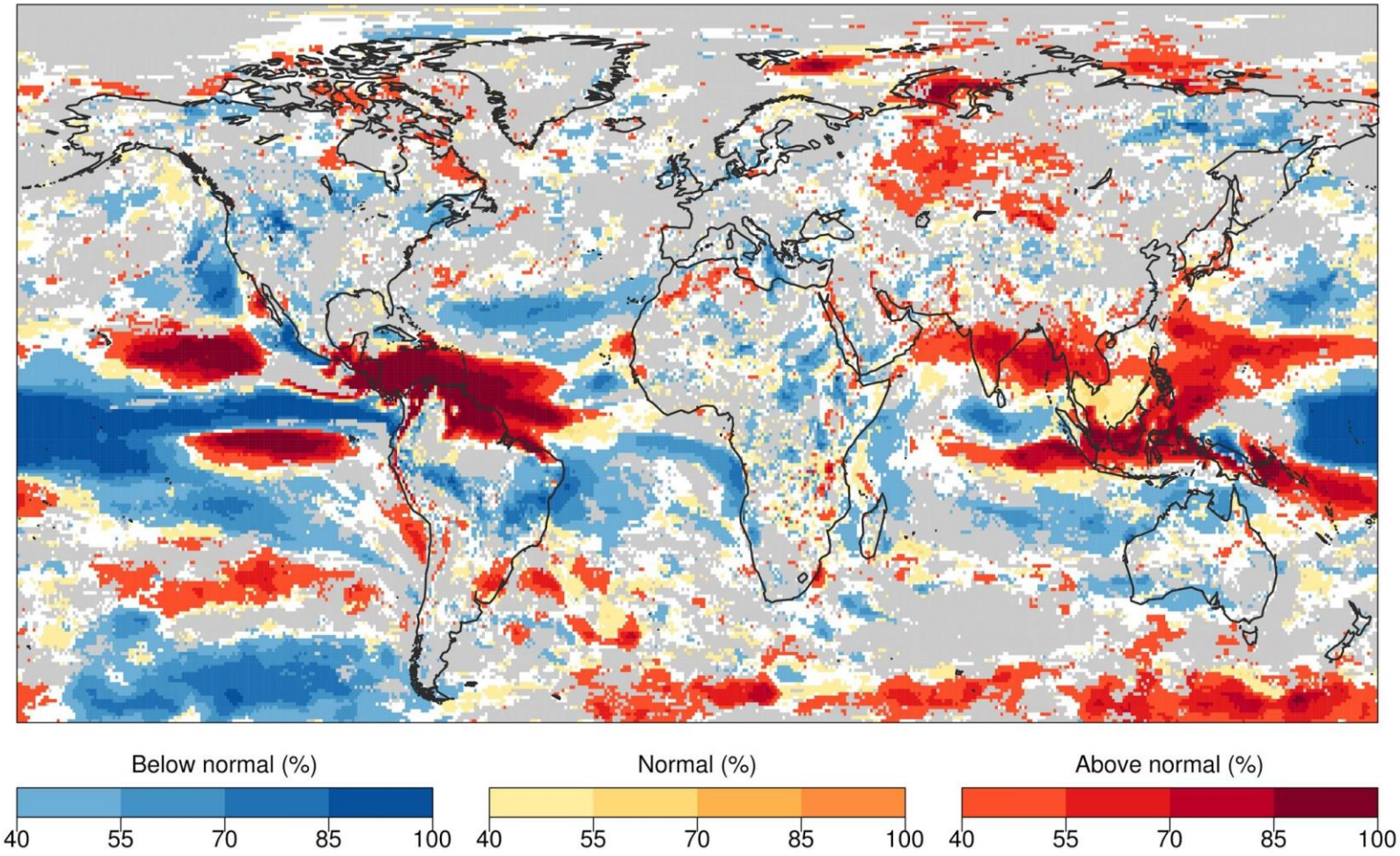
Seasonal wind speed predictions

RESILIENCE PROTOTYPE

- Forecasts from **ECMWF** (European Centre for Medium-Range Weather Forecasts), soon a multi-model
- We assess the global behaviour providing **probabilistic information of the resource**
- Aggregated output in **terciles**:
  - Above normal
  - Normal
  - Below normal
  - Other options possible



ECMWF prediction with verification

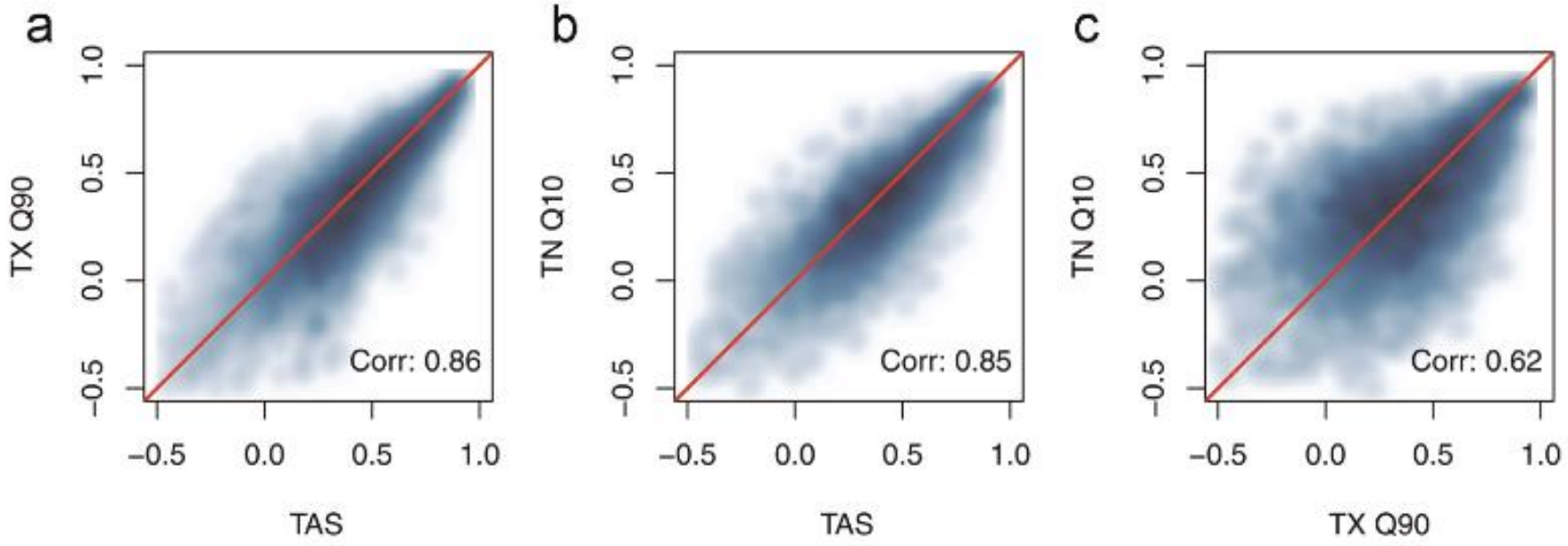


Wind speed prediction for June 1st - August 31st 2015, issued on May 1st 2005.

The most likely wind power category (below normal, normal or above normal), and its percentage probability to occur is shown. "Normal" represents the average of the past. White areas show where the probability is <40% and approximately equal for all three categories. Grey areas show where the climate prediction model does not improve upon the standard and current approach, which projects past climate data into the future.

Verifying “extremes”

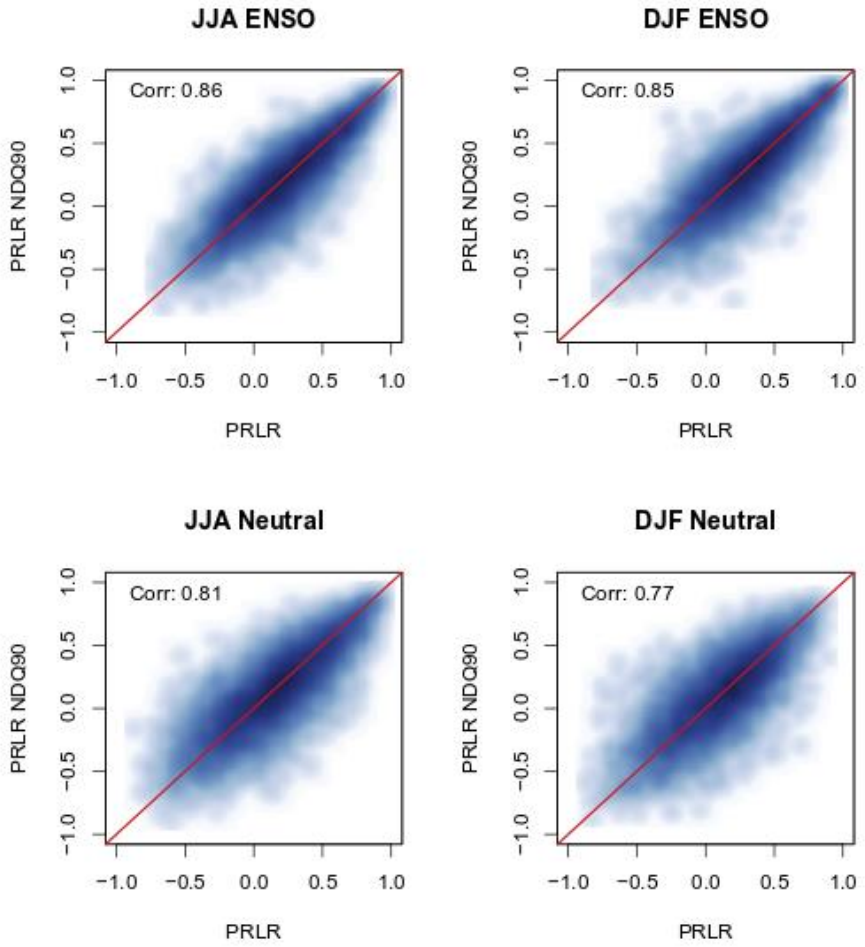
Scatter plot of the correlation of the ENSEMBLES multi-model ensemble mean for seasonal-mean daily temperature (TAS) and the 90th and 10th percentiles of Tmax (TXQ90) and Tmin (TNQ90) over the entire globe during JJA over 1979-2005. The correlation between the two samples is in the upper left corner of each panel.



Peppler et al. (2015, WCE)

Verifying “extremes”

Scatter plot of the correlation of the ENSEMBLES multi-model ensemble mean for seasonal precipitation (PRLR) and the frequency of days the precipitation is above the 90th percentile (PRLRND90) over the globe during JJA over 1979-2005 stratified by the size of the ENSO anomaly.

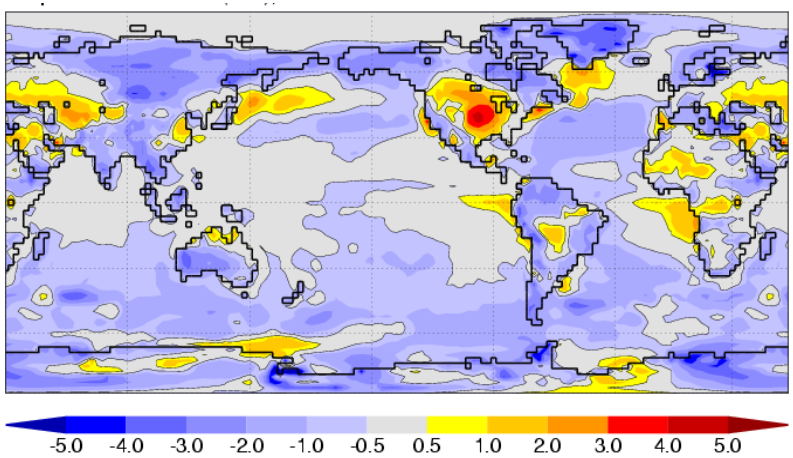


Peppler et al. (2015, WCE)

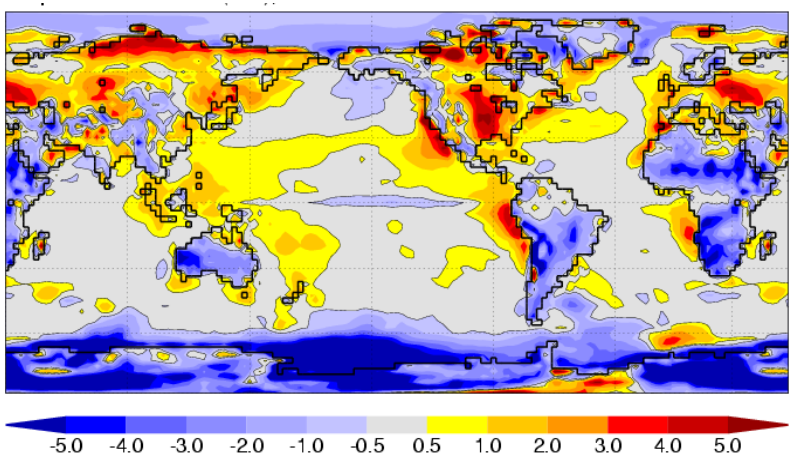
Bias and drift

Mean biases (JJA 2mT over 1993-2005) are often comparable in magnitude to the anomalies which we seek to predict.

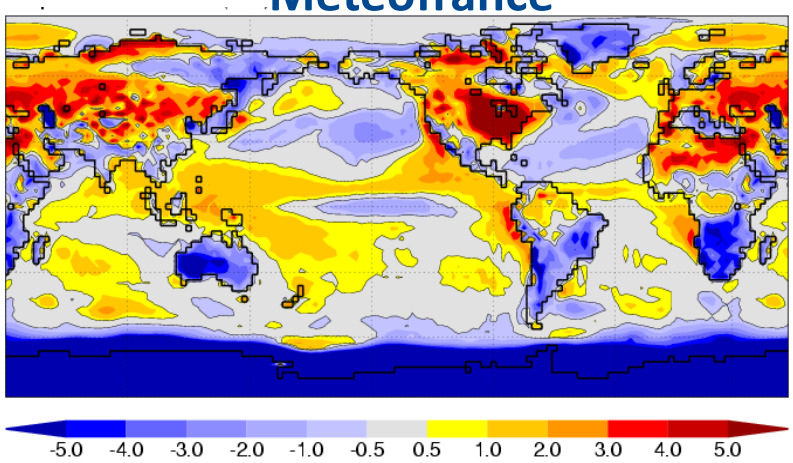
ECMWF



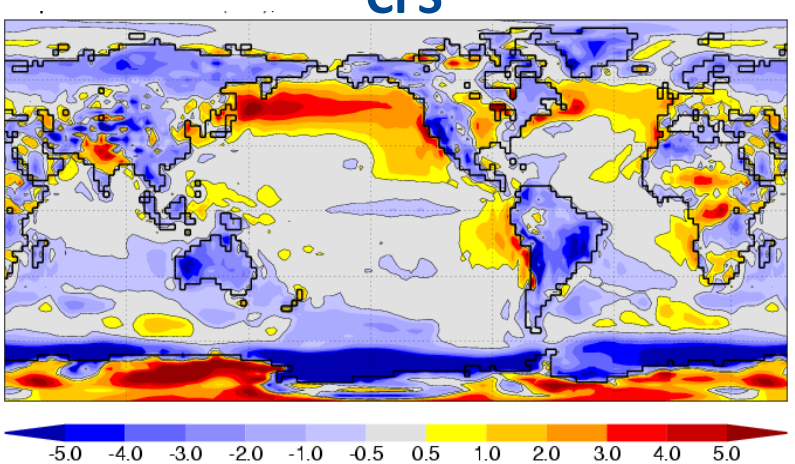
Met Office



Météofrance



CFS



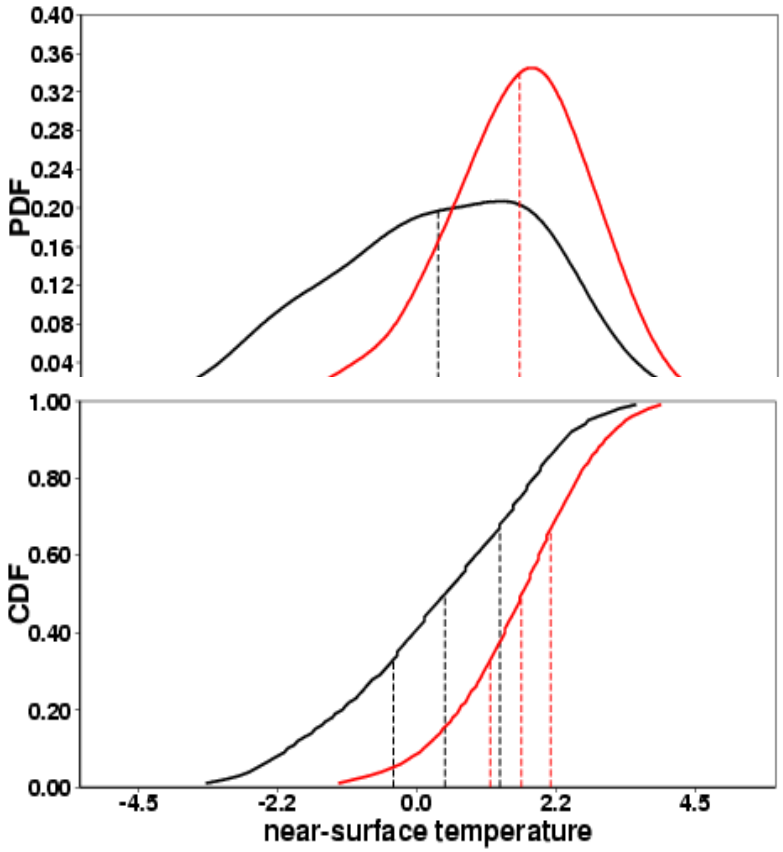
Bias and drift

Climatological PDF of DJF T2m (C) for ERA-40/OPS and ECMWF System4 computed over the period 1960-2005.

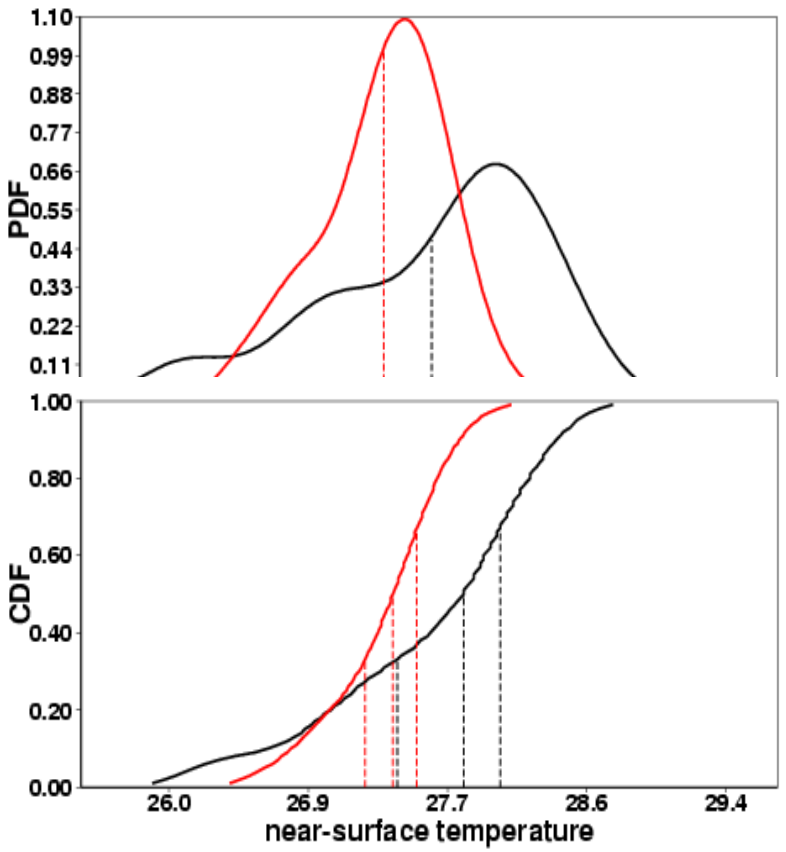
For deterministic forecasts, compute probabilities to the corresponding bin

For probabilistic forecasts, compute probabilities with respect to hindcast and reference thresholds (terciles)

Central Europe (50N, 10E)



Equatorial Pacific (0, 180)



Existing tools under development

Developments from SPECS and EUPORIAS. Tried to solve some of the problems identified in the generic "verification" R package.

Vignettes and documentation exist for all of them.

SCORES

SpecsVerification
- Probabilistic and deterministic scores
- Works on [time x members] arrays

easyVerification
- Applies SpecsVerification scores to arrays of any dimensions, multi-core
- Probabilistic and deterministic scores

FRAMEWORKS

downscaleR + loadeR
- Data retrieval and homogenization
- Bias adjustment, modes, downscaling
- Probabilistic and deterministic scores
- Visualisation of data and results

s2dverification
- Data retrieval and homogenization
- Bias adjustment, filtering, modes
- Probabilistic and deterministic scores
- Visualisation of data and results

Existing tools under development

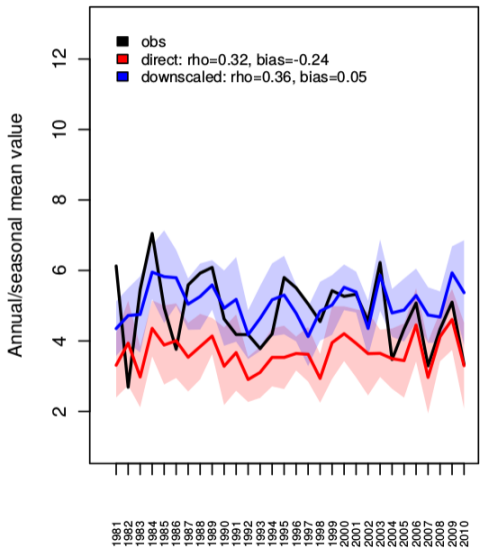
This valuable toolbox should be consolidated:

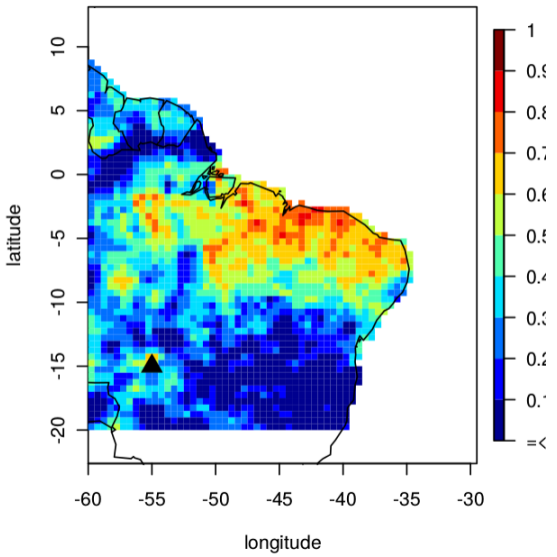
- Select best features of all the packages
- Adapt all scores to work on [time x members] arrays
- easyVerification to apply all the scores
- easyVerification to deal with performance in computation of scores
- Adapt our frameworks to a common, generic data structure (any tool could be applied to all data)
- Fundamental performance issues:
  - RPSS of a 2 GB array on 1 core takes ~2 hours (on a platform P)
  - easyVerification can run on N cores, speed-up=~ N/2 (limited by data transfer), and takes ~30 min. on 8 cores (on same platform P)
  - S2dverification splits the array in chunks of grid points and forecast horizons, gathering results in an 'ff' file, with speed-up =~N (data transfer avoided), and takes ~15 min. on 8 cores (on the same platform P)

Bias adjustment

- downscaleR is a bias adjustment and perfect-prog statistical downscaling application.
- Performs data manipulation: regridding/interpolation, PCA.
- Bias adjustment with cross-validation: scaling, qq-mapping (various forms), parametric.
- Downscaling with analogs, regression (linear and generalized linear), neural networks, weather typing.
- Parallelization under test.

```
out <- downscale(obs = tp.glb,
                 pred = ncep.pca,
                 sim = tp.system4,
                 method = "glm",
                 cross.val = "loocv",
                 n.pcs = 15,
                 parallel = TRUE,
                 ncores = 3)
# Generate an inter-annual diagnostic plot:
quickDiagnostics(obs = tp.glb,
                 sim = tp.system4,
                 downscaled = out,
                 type = "interannual",
                 loc = c(-55,-15))
```

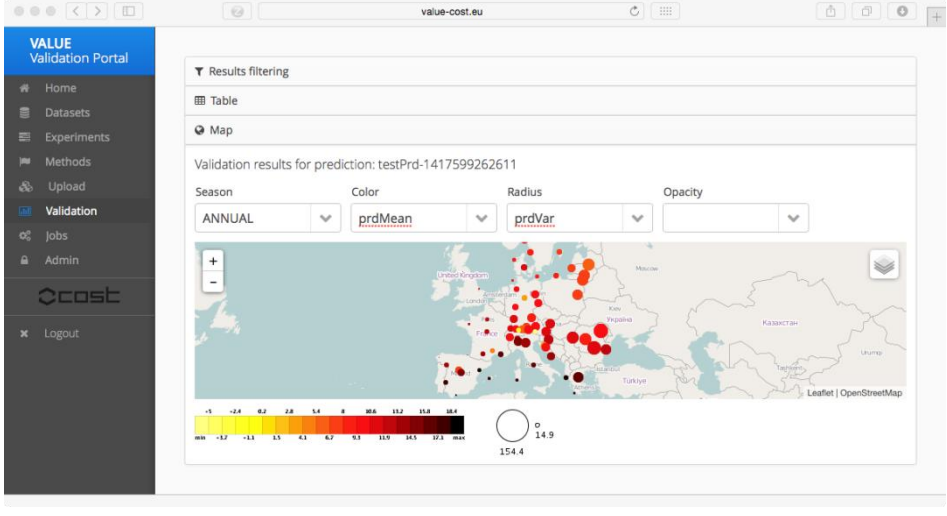




Bias adjustment

- VALUE is an open European network to systematically validate and compare (dynamical and statistical) downscaling methods for climate change research.
- **Metadata** is used in the validation process to decide the suitable validation indices for the different methods. This has been designed in collaboration with ES-DOC and CORDEX-ESD.

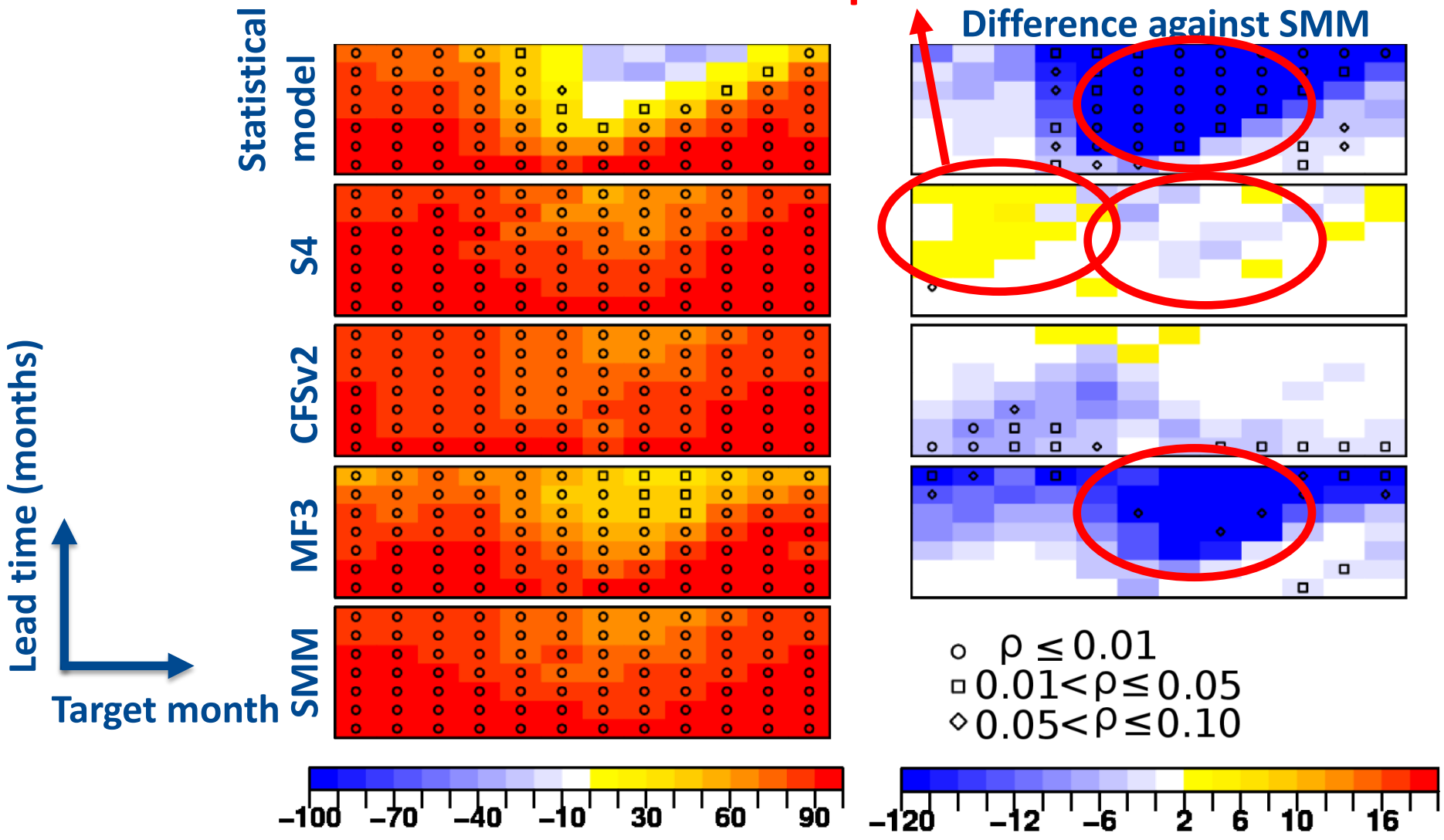
Downscaling **metadata** project,
CIM Control Vocabulary (CV) and schemas.
https://www.earthsystemcog.org/projects/downscalingmetadata/



Multi-model

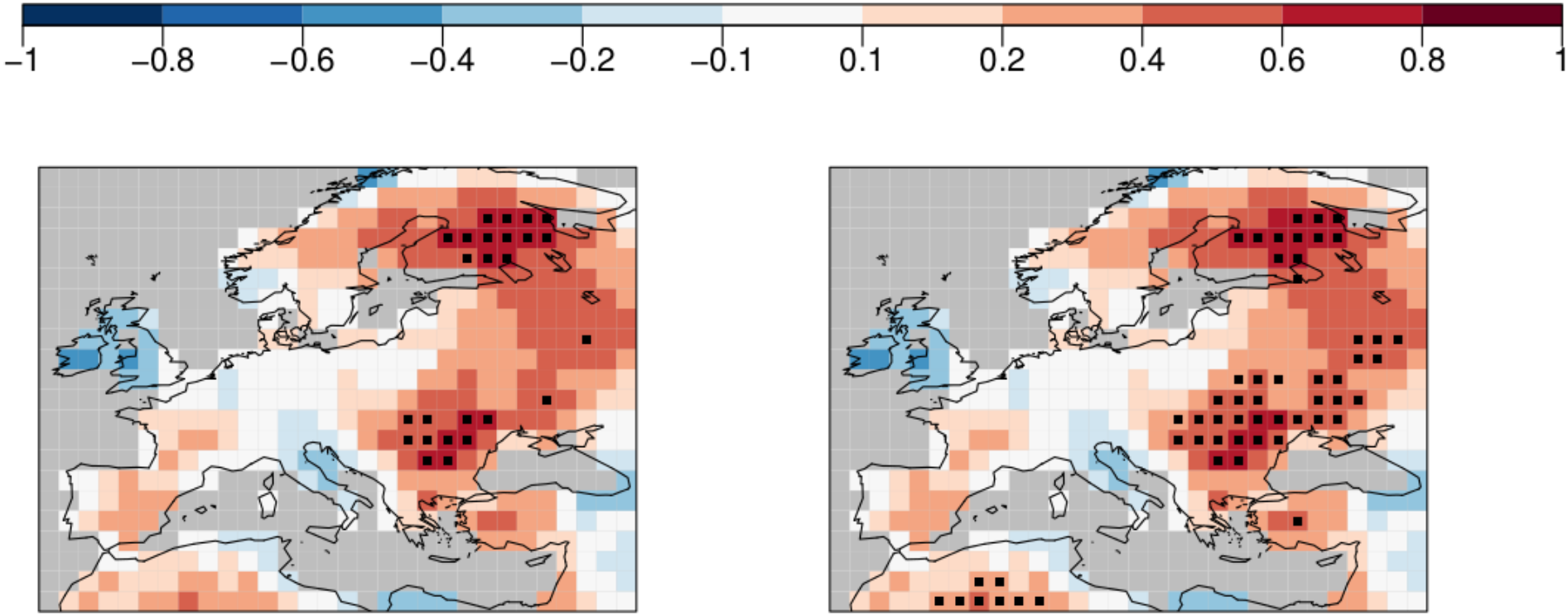
Correlation for the Niño3.4 SST over 1981-2010.

Rodrigues et al. (2014, Clim. Dyn.)



Uncertainty in the QA

Difference in correlation of EC-Earth3 (T511/ORCA025) seasonal one-month lead predictions started every May over 1993-2009 with ERAInt and GLORYS2v1 ics, and internal sea-ice reconstruction with realistic and climatological land-surface initial conditions: left for traditional Fisher test and right for Steiger test (with increased power).

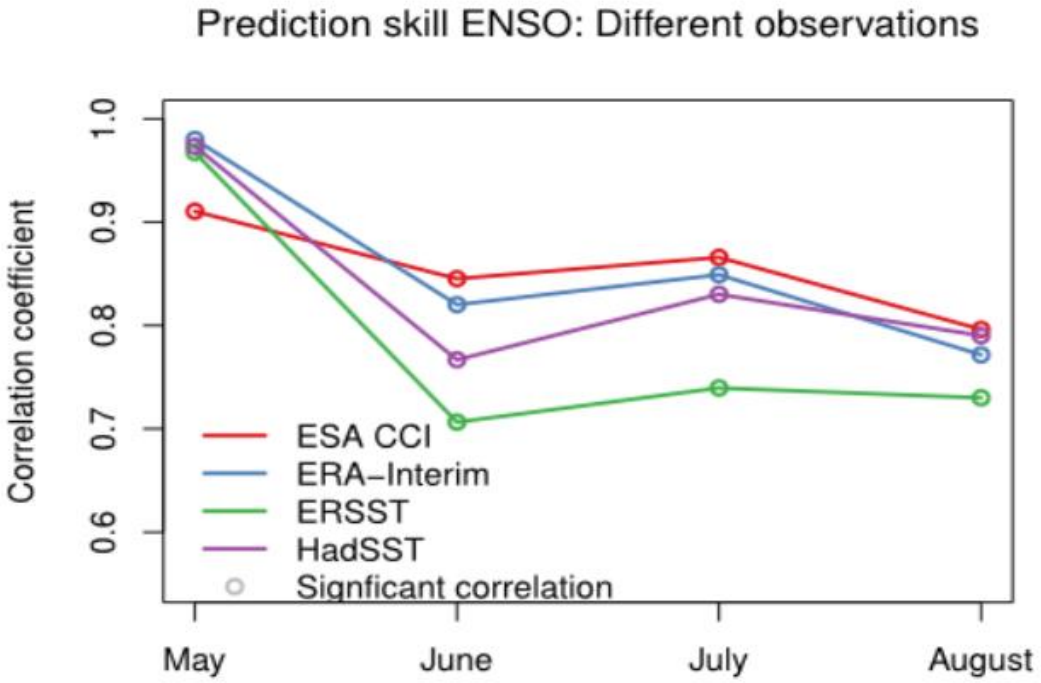


Siegert et al. (Submitted to MWR)

Observational uncertainty

Need to take into account the large observational uncertainty both in the initialisation and the forecast quality. The Global Climate Observing System (GCOS) is a key actor in this task.

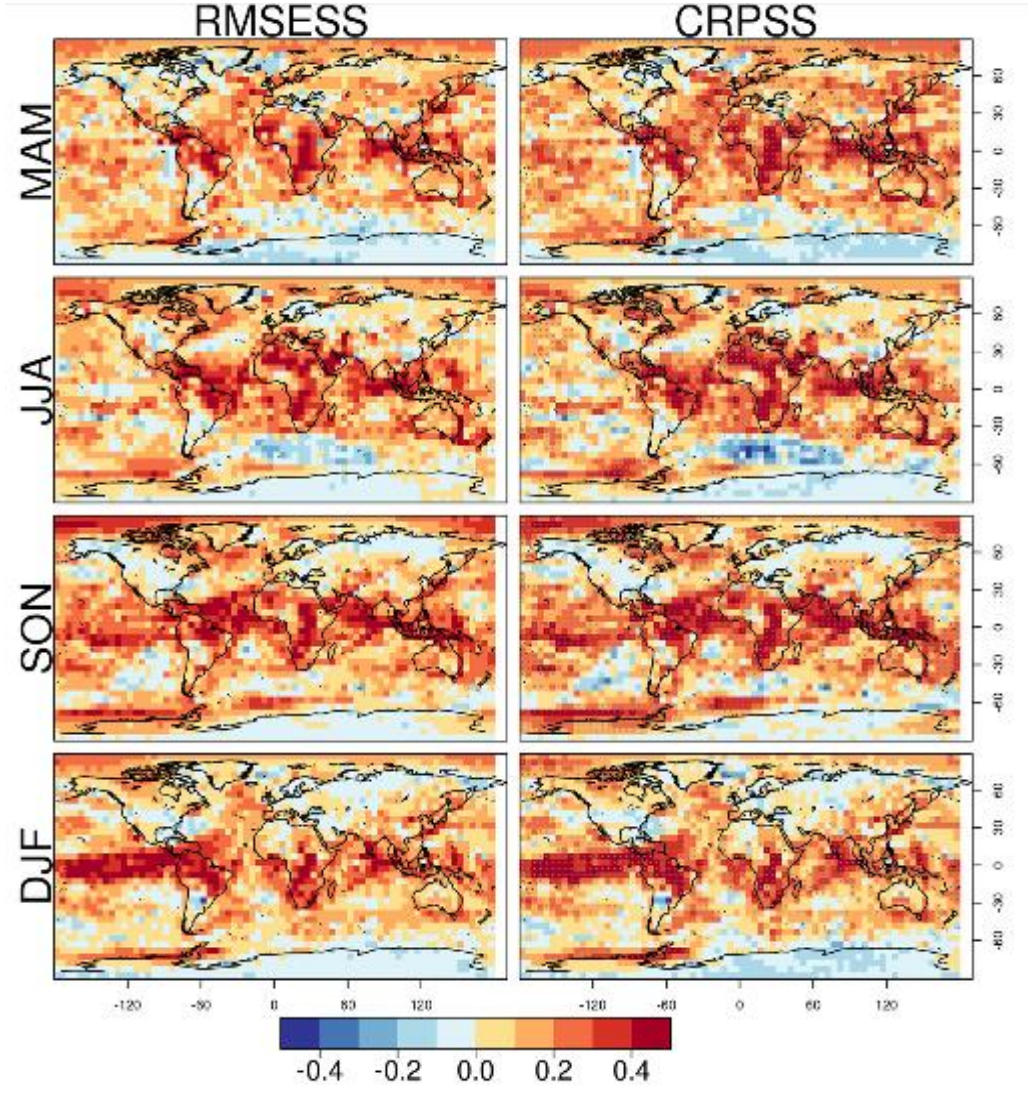
EC-Earth3 (T511/ORCA025) predictions started every May over 1993-2009 with ERAInt and GLORYS2v1 ics, and internal sea-ice reconstruction.



Bellprat et al. (2015, IC3 Tech. Note)

Generalised empirical forecasts

Empirical forecasts of one-month lead temperature using a wide range of observed predictors. A benchmarking opportunity.



Eden et al. (2015, GMD)

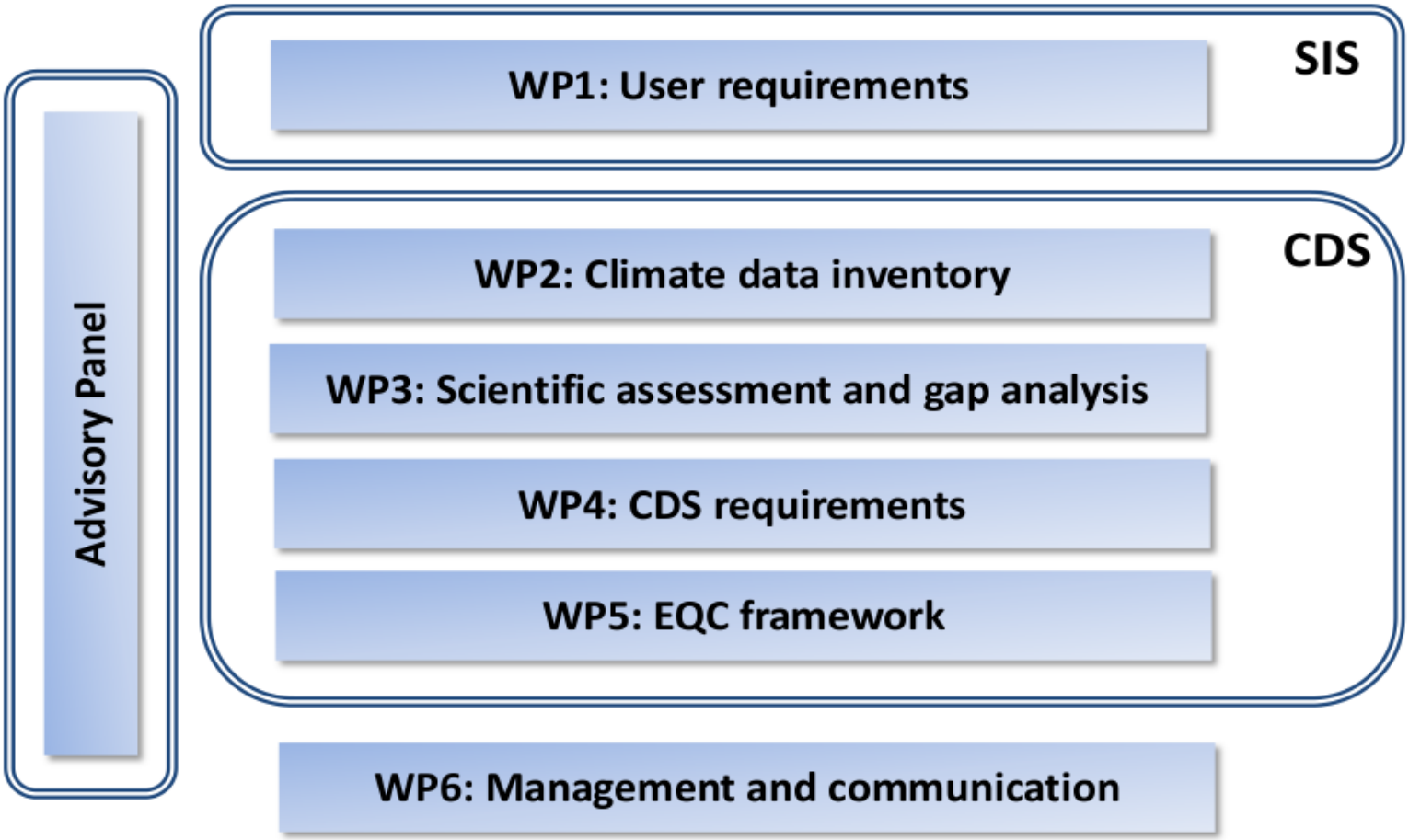
Consortium

Six partners and 10 in-kind participants.

Partner	Nature	Role	Effort (PM)
BSC-CNS	Main contractor	Coordination, data inventory and EQC framework and prototype	88
Univ. Leeds	Subcontractor	Assess user requirements	14
Meteoswiss-MCH	Subcontractor	Scientific quality assessment and gap analysis	23
Predictia	Subcontractor	CDS requirements and development of the prototype	28
Univ. Exeter	Subcontractor	Expert statistical advice	2
IFCA-CSIC	Subcontractor	Downscaling	14

Management

QA4Seas will have a) a steering committee composed of one representative of the BSC-CNS and of each subcontractor, and b) an expert advisory panel.



WP1

WP1: Assessment of users' requirements.

- Assess user requirements for the **evaluation and quality control of data** describing the physical climate system in the context of climate services **provision from climate predictions**, and in connection with other applications involving climate information development relevant to C3S.
- Methods used: data collection, surveys, and interviews.
- Workflow:
  - **Strategic framework:** collaboration and coordination with other C3S 51 lots and proof-of-concept projects regarding the users
  - **Collating data** on user requirements from other projects
  - **Survey** for assessing requirements of users involved in the project
  - **Interviews** for assessing requirements of specific users (JRC, EEA)
  - **Survey** for assessing needs of SIS user group and advisory panel
  - **Sharing findings** with other WPs e.g. WP3 scientific assessment and gap analysis and WP5 and EQC framework

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

An example from EUPORIAS.

Interviews	Survey	Workshops
• In depth interviews with EUPORIAS stakeholders and other European users	• Multi-lingual survey sent to European organisations across a range of sectors	• Workshop with European climate services providers • Workshop with developers of seasonal forecasts
80 interviews across 16 EU countries	450+ responses in 37 EU countries	Around 30 people in each workshop

UNIVERSITY OF LEEDS

WP1

List of tasks:

1. Develop a strategic framework for collaboration and coordination with other lots and POC projects (**Univ. Leeds**, BSC-CNS).
2. Collate user requirements from other European projects and initiatives (**Univ. Leeds,** BSC-CNS, MCH, Predictia, IFCA-CSIC).
3. Develop and implement a survey for assessing user requirements (**Univ. Leeds,** BSC-CNS, MCH, Predictia).
4. Develop an interview protocol based on the first survey (**Univ. Leeds,** BSC-CNS, MCH, Predictia).

WP1

List of tasks:

5.Analysis of the results from the survey and interviews (**Univ. Leeds,** BSC-CNS, MCH, Predictia, IFCA-CSIC).

6.Share findings from initial survey and interviews with other WPs (**Univ. Leeds,** BSC-CNS, MCH, Predictia, IFCA-CSIC).

7.Second round of surveys (**Univ. Leeds,** BSC-CNS, MCH, Predictia, IFCA-CSIC).

8.Share findings from second round of surveys with partners (**Univ. Leeds,** BSC-CNS, MCH, Predictia, IFCA-CSIC).

WP2

WP2: Climate data inventory.

- Develop **an inventory of existing climate data sets** to be considered for inclusion in the CDS to address the user requirements identified in WP1. The data requirements of a set of **user-relevant indices** will also be considered to ensure that the necessary ECVs are available.
- Tasks:

1. List of available ECVs to be used in the WP1 survey process (**BSC**, Univ. Leeds).
2. Inventory of additional ECVs and indices from the set of user requirements (**BSC**, Univ. Leeds, IFCA-CSIC, MCH).
3. Coordination with other lots of C3S 51 and with the seasonal prediction data providers of C3S 433 (**BSC**).

WP3

WP3: Scientific assessments and gap analysis.

- Provide scientific quality assessments for the European multi-model seasonal forecasts serving as an overview of the quality of the pre-operational and POC systems available in the CDS, including an uncertainty assessment of the quality estimates. Carry out a gap analysis of the current quality information available to users, in graphical, text and numerical form, and refer it to the user requirement analysis performed in WP1. Produce recommendations to inform the European research agenda.

- Tasks:

1. Evaluation of forecast quality of pre-operational and POC multi-model seasonal forecasts (**MCH**, BSC-CNS, Predictia, IFCA-CSIC).
2. Assessment of limitations and value added by downscaling of seasonal forecasts (**IFCA-CSIC**, MCH).
3. Gap analysis of information available to users (**MCH**, BSC-CNS).

WP4: CDS requirements.

- Produce recommendations for further CDS development based on the user requirements (WP1), the data availability (WP2), and the gap analysis and technical and scientific assessments of the current data sets (WP3) including additional data sets for the CDS, strategies for products and communicate uncertainty, suggestions of user guidance and visualisation requirements, and aspects linked to an efficient data access like metadata standards for the quality information, code development and numerical performance.
- Tasks:

1. Strategies for communicating uncertainty (**Predictia**, Univ. Leeds).
2. Specification of EQC products and standards to be included in the CDS (**Predictia**, BSC-CNS, MCH).
3. Recommendations for computational efficiency and code development guidance (**Predictia**, Univ. Exeter, BSC-CNS, MCH).

WP5

WP5: EQC framework.

•Produce recommendations for the development of the EQC function for multi-model seasonal forecasts and develop a framework for quality assessment, including a list of key performance indicators for the CDS illustrated with a functional prototype.

•Tasks:

1.EQC framework (**BSC-CNS**, Univ. Exeter, MCH).

2.EQC prototype (**BSC-CNS**, MCH, Predictia, Univ. Leeds).

WP6

WP6: Management and communication.

- Ensure the a) management of the project, including reporting to ECMWF, links to ECMWF and other relevant agencies (EEA, JRC), the partners and other relevant C3S contractors and b) the dissemination of the main public outcomes, mainly among the SIS user group.
- Tasks:

1. Description of the service (**BSC-CNS**).
2. Project meetings (**BSC-CNS**).
3. Periodic (quarterly and annual) and final reports (**BSC-CNS**).
4. Link to C3S and coordination with other C3S activities (**BSC-CNS**, Univ. Leeds).
5. Workshop (**BSC-CNS,** Univ. Leeds).
6. Web site and dissemination (**BSC-CNS**).

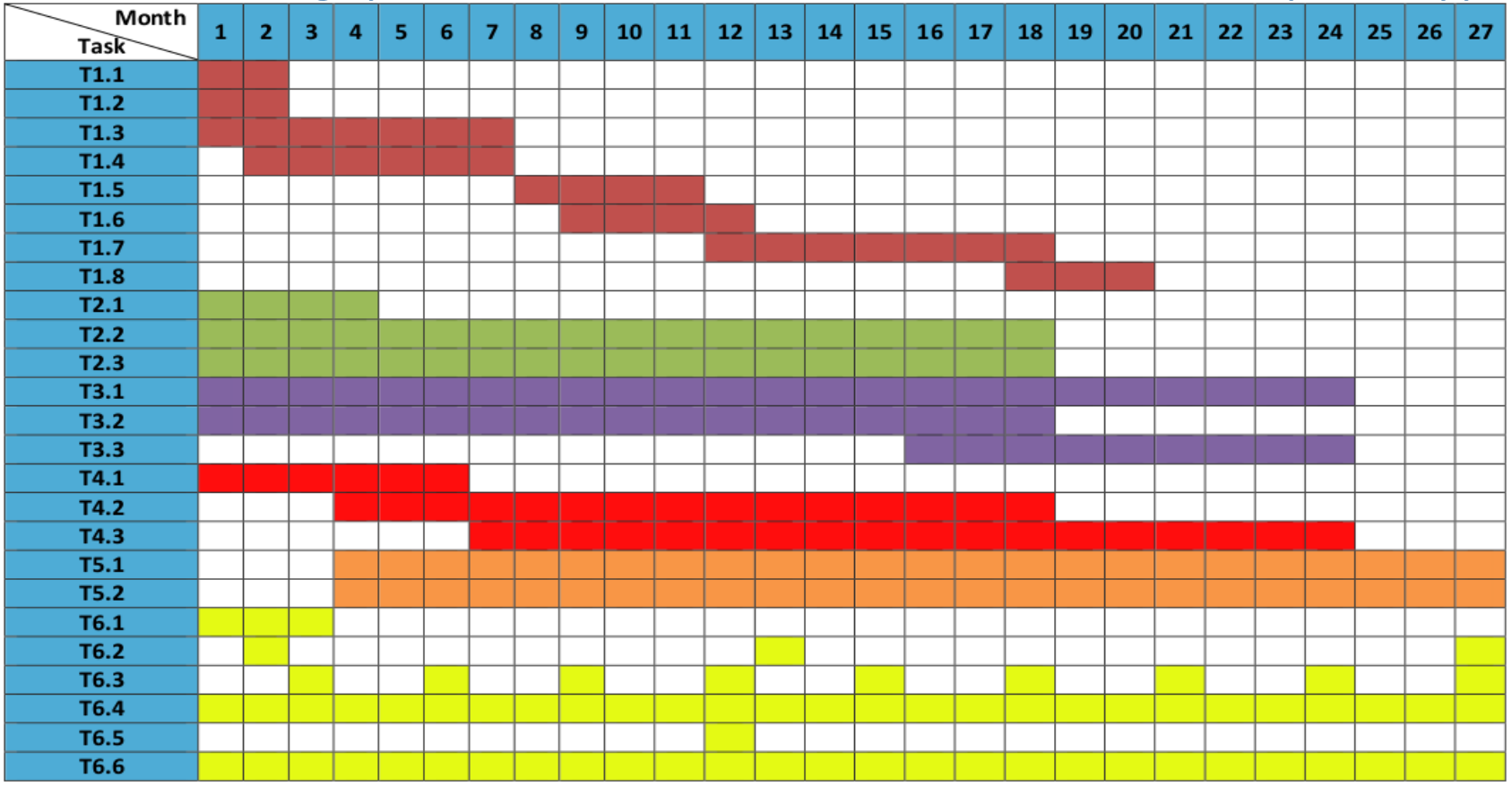
Management

- Coordination ensured by BSC-CNS.
- Eugene Griffiths and Mar Rodríguez project managers.
- Isadora Jiménez will coordinate the communication activities.
- Partners represented in the steering committee.
- Agile methodology for task management:
  - Monthly teleconferences. Some teleconferences may include the advisory panel and, if considered relevant, members of the SIS user group. Three main meetings will also take place, gathering subcontractors and users.
  - A first version of each quarterly activity report will be asked to each subcontractor two weeks before deadline of the quarterly activity report to be delivered to ECMWF.
  - The deadline for the completion of the deliverables will be 1 month before the delivery to ECMWF, allowing an internal review by all participants (including the advisory panel).
  - Use of version control (GIT).

Project development

The project will be developed in two main phases:

- Phase 1 (M1-18): First users' consultation, assessment of user requirements and preliminary scientific assessment
- Phase 2 (M7-27): Formulation of requirements, 2nd stage of users' consultation, gap assessment, and EQC framework and prototype



Deliverables and milestones

Deliverable/Milestone	Due	Responsible	Effort (PM)
D1.1: Survey of user requirements	3	Univ. Leeds	7
D1.2: Findings from survey and interviews	12	Univ. Leeds	5
D1.3: Findings from survey with subset of users	20	Univ. Leeds	4.15
M1.1: Framework for collaboration and coordination with other C3S 51 lots and proof-of-concept projects	2	Univ. Leeds	
D2.1: User-oriented inventory of ECVs	9	BSC-CNS	2
D2.2: Inventory of additional ECVs and indices for seasonal forecast evaluation	18	BSC-CNS	13
M2.1: User-oriented table of available ECVs relevant to seasonal forecast evaluation	1	BSC-CNS	
M2.2: List of user-relevant indices	8	IFCA-CSIC	
M2.3: Summary of coordination of observational needs with other lots of C3S 51	12	BSC-CNS	

Deliverables and milestones

Deliverable/Milestone	Due	Responsible	Effort (PM)
D3.1: Assessment of forecast quality of illustrative bias-corrected/adjusted seasonal forecasts	18	IFCA-CSIC	10
D3.2: Assessment of forecast quality of multi-model seasonal forecasts available in the CDS	24	MCH	16
D3.3: Gap analysis and recommendations to mitigate gaps	24	MCH	7
M3.1: Assessment of limitations and value added by downscaling and bias correction/adjustment	9	IFCA-CSIC	
M3.2: Preliminary evaluation of the multi-model seasonal forecast system available in the CDS	9	MCH	
M3.3: Preliminary gap analysis	18	BSC-CNS	

Deliverables and milestones

Deliverable/Milestone	Due	Responsible	Effort (PM)
D4.1: Recommendations for data (EQC products including uncertainty) and metadata formats for EQC products	18	Predictia	16
D4.2: Recommendations for computational efficiency and integration strategies with the CDS	24	Predictia	12.5
M4.1: Preliminary list of intermediate and final EQC product types and standards	9	Predictia	
M4.2: Estimation of computational efficiency of existing validation packages	12	Predictia	
M4.3: Strategies for storing and communicating uncertainty in EQC products	16	Predictia	
D5.1: EQC framework	27	BSC-CNS	16.3
D5.2: EQC prototype	27	BSC-CNS	20
M5.1: First version of EQC prototype	21	BSC-CNS	
M5.2: Draft version of the EQC framework	22	BSC-CNS	

Deliverables and milestones

Deliverable/Milestone	Due	Responsible	Effort (PM)
D6.1: Project meeting summary (3)	2,14,27	BSC-CNS	3
D6.2: Periodic and final reports	Quartely, annual, final	BSC-CNS	6
D6.3: Fully documented web site	27	BSC-CNS	32.5
M6.1: QA4Seas web site including description of service	3	BSC-CNS	
M6.2: Quality assessment workshop for CDS/SIS	12	BSC-CNS	

Points for discussion

- *Details of the plans for the assessment of uncertainties in the estimation of scores and metrics proposed for evaluation of the forecasting systems.*
- *The proposed work on calibration and multi-system combinations, to minimise the risk of overlap between activities of this lot and ongoing activities.*
- *The proposed interactions with other C3S activities (CDS development and toolbox, SIS). The main points will be:*
  - *Any critical dependencies (of this project on the others).*
  - *Ensuring there is no duplication and that any interaction with the other C3S activities happen in coordination with and in the context of ECMWF's activities on these topics.*