



Call: H2020-SC5-2014-two-stage

Topic: SC5-01-2014

**PRIMAVERA**

Grant Agreement 641727



**PRocess-based climate sIMulation: AdVances in high resolution modelling and  
European climate Risk Assessment**

## **Deliverable D1.2**

***Tools for the Process-Based Assessment  
of the PRIMAVERA Climate Experiments***

Deliverable Title	Tools for the Process-Based Assessment of the PRIMAVERA Climate Experiments		
Brief Description	This Deliverable documents a list of process-based model evaluation tools developed in PRIMAVERA WP1, particularly aiming at assessing processes affecting European climate.		
WP number	1		
Lead Beneficiary	CMCC		
Contributors	Francisco Doblas-Reyes - BSC Louis-Philippe Caron - BSC Javier Vegas - BSC Eduardo Moreno - BSC Eleftheria Exarchou - BSC Julien Boe - CERFACS Laurent Terray - CERFACS Panos Athanasiadis - CMCC Federico Fabiano - CNR Susanna Corti - CNR Fuentes Franco Ramon - SMHI Zimmermann Klaus - SMHI David Docquier - UCLouvain François Massonet - UCLouvain Thierry Fichefet - UCLouvain Malcolm Roberts - UKMO Alex Baker - UREAD Reinhard Schiemann - UREAD		
Creation Date		10-02-2019	
Version Number		v4	
Version Date		04-03-2019	
Deliverable Due Date		28-02-2019	
Actual Delivery Date		04-03-2019	
Nature of the Deliverable	R	R - Report	
		P - Prototype	
		D - Demonstrator	
		O - Other	
Dissemination Level/ Audience	PU	PU - Public	
		PP - Restricted to other programme participants, including the Commission services	
		RE - Restricted to a group specified by the consortium, including the Commission services	
		CO - Confidential, only for members of the consortium, including the Commission services	

Version	Date	Modified by	Comments
1	20/02/2019	P. Athanasiadis (CMCC)	Few contributions missing.
2	26/02/2019	P. Athanasiadis (CMCC)	All contributions included.

3	28/02/2019	P. Athanasiadis (CMCC)	Corrections made by partners.
4	04/03/2019	P. Athanasiadis (CMCC)	Feedback addressed.

## Table of Contents

1. Executive Summary.....	4
2. Project Objectives.....	6
3. Detailed Report.....	7
3.1 Introduction to PRIMAVERA metrics, data and methods.....	8
3.2 Processed-based metrics for single and coupled components.....	8
3.2.1 Implementation of CVDP on PRIMAVERA models (SMHI)	
3.2.2 Sea-ice drift - strength metric (UCLouvain)	
3.2.3 Ice formation efficiency diagnostic (UCLouvain)	
3.2.4 Northern hemisphere blocking (UREAD, CMCC)	
3.2.5 Position and strength of the North Atlantic eddy-driven jet (CMCC, UREAD)	
3.2.6 Tropical cyclone tracker (BSC)	
3.2.7 Ocean Heat Content (BSC)	
3.2.8 K-mean clustering of sea-ice (BSC)	
3.2.9 Wintertime Euro-Atlantic Weather Regimes (CNR)	
3.2.10 Solid and liquid freshwater volumes and exports from the Arctic (SMHI)	
3.2.11 The representation of extreme European precipitation (UREAD)	
3.2.12 Multi-algorithm analysis of tropical cyclones (UKMO)	
3.2.13 sea-ice - ocean heat transport metric (UCLouvain)	
3.2.14 Local coupled feedbacks over Europe in summer (CERFACS)	
4. Lessons Learned.....	42
5. Links Built.....	42
6. References.....	43

## 1. Executive Summary

A powerful assortment of processed-based metrics have been developed in WP1 meeting the aims of Task1.1 and Task1.2 as quoted below from the Description of Work:

*T1.1: Metrics will be developed to quantify the representation of processes that involve mainly one component of the climate system when considering their impact on European climate. These will include both aspects of the basic mean state (e.g. mid-latitude jet, storm track, ocean heat content) together with weather-related variability and extremes (storm clustering, atmospheric moisture transport, sea ice variability). Metrics will be combined to enable assessment of present day climate, and then used in an attempt to understand and constrain future projections.*

*T1.2: Metrics will be developed to quantify the representation of coupled processes and those processes that involve more than one component of the climate system and their impact on European climate. These will include coupled modes of variability and teleconnections (North Atlantic Oscillation (NAO), Atlantic Meridional Overturning Circulation (AMOC), troposphere-stratosphere coupling), local coupled feedbacks and indices of climate extremes. Metrics will be combined to enable assessment of overall present-day performance, and be used to understand and constrain future projections.*

Indeed, as presented below, emphasis has been given on assessing processes dominantly affecting the European climate, directly or indirectly. Each of the developed metrics is well-based on existing literature, while a series of new publications are made and other are in progress documenting the background, the significance and the correct interpretation of each of these new metrics.

A wide variety of tools (metrics) have been implemented on our data analysis platform, and used to assess the multi-model, multi-resolution performance of the PRIMAVERA-HighResMIP model simulations. Such summary metrics give insight into the model performance, as well as enabling all project members to gain a quick understanding of each models' strengths and weaknesses.

Specifically, there are metrics focussing:

1. on atmospheric processes and phenomena, such as: teleconnection patterns (§ 3.2.1), blocking (§ 3.2.4), jet variability (§ 3.2.5), tropical cyclones (§ 3.2.6, § 3.2.12), weather regimes (§ 3.2.9) and extremes in precipitation over Europe (§ 3.2.11),
2. on oceanic and processes and the sea-ice, such as: sea-ice drift and its relation to sea-ice thickness and concentration (§ 3.2.2), sea-ice formation efficiency as a function of existing sea-ice volume (§ 3.2.3), ocean heat content and its changes under climate change (§ 3.2.7), variability patterns of sea-ice thickness and concentration derived through clustering analysis (§ 3.2.8), freshwater exports in

solid and liquid phase from certain basins (§ 3.2.10) and the relationship between ocean heat transports and sea-ice changes (§ 3.2.13),

3. finally, there are also metrics assessing coupled processes, in particular evapotranspiration over land (§ 3.2.14).

It is important to note that seven (7) out of the total fourteen (14) metrics / assessment tools presented here have been particularly developed for PRIMAVERA (referring to: § 3.2.2, § 3.2.3, § 3.2.5, § 3.2.9, § 3.2.10, § 3.2.12, § 3.2.13), while for most of the remaining ones, significant modifications were needed in order to render the respective codes fully functional on JASMIN —so that they can be smoothly implemented into the ESMValTool— and to increase the quality of the final output (graphics).

In addition, there are a few more coupled and single-component processes that are being assessed in PRIMAVERA simulations, yet respective metrics are in development, or may need to be recoded in Python. Although this is work that will be done, there is no commitment that the associated metrics will be finally incorporated into the ESMValTool. Such metrics will assess, for example, the North Atlantic stormtrack activity and variability with methods different than storm-tracking (Athanasiadis et al., 2010) and certain aspects of the air–sea interaction along SST fronts, in particular the Gulf Stream Extension area (Bishop et al., 2017; Putrasahan et al., 2013).

## 2. Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (DOA, Part B Section 1.1) WP numbers are in brackets:

No.	Objective	Yes	No
A	To develop a new generation of global high-resolution climate models. (3, 4, 6)		X
B	To develop new strategies and tools for evaluating global high-resolution climate models at a process level, and for quantifying the uncertainties in the predictions of regional climate. (1, 2, 5, 9, 10)	X	
C	To provide new high-resolution protocols and flagship simulations for the World Climate Research Programme (WCRP)'s Coupled Model Intercomparison Project (CMIP6) project, to inform the Intergovernmental Panel on Climate Change (IPCC) assessments and in support of emerging Climate Services. (4, 6, 9)		X
D	To explore the scientific and technological frontiers of capability in global climate modelling to provide guidance for the development of future generations of prediction systems, global climate and Earth System models (informing post-CMIP6 and beyond). (3, 4)		X
E	To advance understanding of past and future, natural and anthropogenic, drivers of variability and changes in European climate, including high impact events, by exploiting new capabilities in high-resolution global climate modelling. (1, 2, 5)	X	
F	To produce new, more robust and trustworthy projections of European climate for the next few decades based on improved global models and advances in process understanding. (2, 3, 5, 6, 10)		X
G	To engage with targeted end-user groups in key European economic sectors to strengthen their competitiveness, growth, resilience and ability by exploiting new scientific progress. (10, 11)		X
H	To establish cooperation between science and policy actions at European and international level, to support the development of effective climate change policies, optimize public decision making and increase capability to manage climate risks. (5, 8, 10)		X

### 3. Detailed Report

The work documented in this Deliverable aimed to contribute to the following two primary PRIMAVERA objectives (quoted from Section 2):

- *To develop new strategies and tools for evaluating global high-resolution climate models at a process level, and for quantifying the uncertainties in the predictions of regional climate.*
- *To advance understanding of past and future, natural and anthropogenic, drivers of variability and changes in European climate, including high impact events, by exploiting new capabilities in high-resolution global climate modelling.*

WP1 has, indeed, provided a spectrum of process-based evaluation tools for climate models, optimized to aid the assessment of high-resolution climate models, as well as the understanding of drivers of climate variability and climate changes, particularly relevant to European climate. These tools are documented in the remaining of this Section.

Also, a number of articles have been submitted, or are in preparation and will be submitted in the next few months (2019). These include the following:

- Athanasiadis, P., Baker, A. and co-authors ([in preparation](#)): Assessing the representation of the North Atlantic eddy-driven jet in PRIMAVERA historical simulations.
- Bellucci, A. and co-workers ([in preparation](#)): Air-sea interactions over the Gulf Stream extension in HighResMIP models.
- Docquier, D., J. P. Grist, M. J. Roberts, C. D. Roberts, T. Semmler, L. Ponsoni, F. Massonnet, D. Sidorenko, D. Sein, D. Iovino, A. Bellucci, T. Fichefet ([in review](#)). Impact of model resolution on Arctic sea ice and North Atlantic Ocean heat transport.
- Docquier, D. and co-authors ([in preparation](#)): The relationships between sea-ice and ocean heat content variability in Barents and Bering Seas.
- Moreno-Chamarro, E., Ortega, P. and Massonnet, F. ([in preparation](#)): Impact of the ice thickness distribution on sea ice interannual variability in the NEMO3.6-LIM3 global ocean-sea ice model.
- Roberts, M. J., and co-authors ([in preparation](#)): A comparison of three popular tropical cyclone tracking algorithms using CMIP6 HighResMIP model simulations.
- Schiemann, R., Athanasiadis, P. and co-authors ([in preparation](#)): Northern Hemisphere blocking in PRIMAVERA historical simulations: benefits from increasing model resolution.

- Strommen, K., I. Mavilia, S. Corti, M. Matsueda, P. Davini, J. von Hadenberg, P-L. Vidale, R. Mizuta ([in review](#)). The Sensitivity of Euro-Atlantic Regimes to Model Horizontal Resolution.

### 3.1 PRIMAVERA process-based model assessment tools

The developed WP1 metrics are to be implemented in a brand-new version of the Earth System Model eValuation Tool (ESMValTool) that is aimed at handling efficiently large data volumes as produced by current (and beyond current) state-of-the-art high-resolution climate models. This implementation is ongoing. In fact, as documented in MS3, the implementation is already completed and tested for a small group of metrics.

Particular emphasis has been given in creating flexible and user-tuned tools so that the sensitivity of the metrics themselves to various parameters and methods (as per case) can be assessed. This is key for judging the robustness of the results and estimating uncertainties. Moreover, different observational datasets (or reanalyses) can be used for model evaluation, as well as different realisations of the same model runs (climate experiments) as will arise from Stream-2 simulations. This will allow a better evaluation of the uncertainties in estimating future climate change and in drawing conclusions on the effect of increasing model resolution. The implemented metrics will be delivered in D1.3 with the completion of task T1.3, as described in the Description of Work:

*T1.3 This task aims at providing a framework to integrate all the diagnostics developed to address the metrics described in Tasks 1.1 and 1.2 and to develop a long-term solution for the process-based analysis of the PRIMAVERA and CMIP6 European experiments (including HighResMIP). The solution will be based on existing national and international initiatives to develop metrics and will encompass the main developments in metrics performed in PRIMAVERA. In particular this task entails the conversion of AutoAssess to Python, the analysis of the common functionalities of the EMBRACE and WGNE/WGCM diagnostic packages including the necessary adaptations for a special focus on Europe and the design of a strategy for common metrics and diagnostics for the analysis of the CMIP6 experiments performed by European climate modelling institutions.*

### 3.2 Processed-based metrics for single and coupled components

#### 3.2.1 Implementation of CVDP on PRIMAVERA models [SMHI]

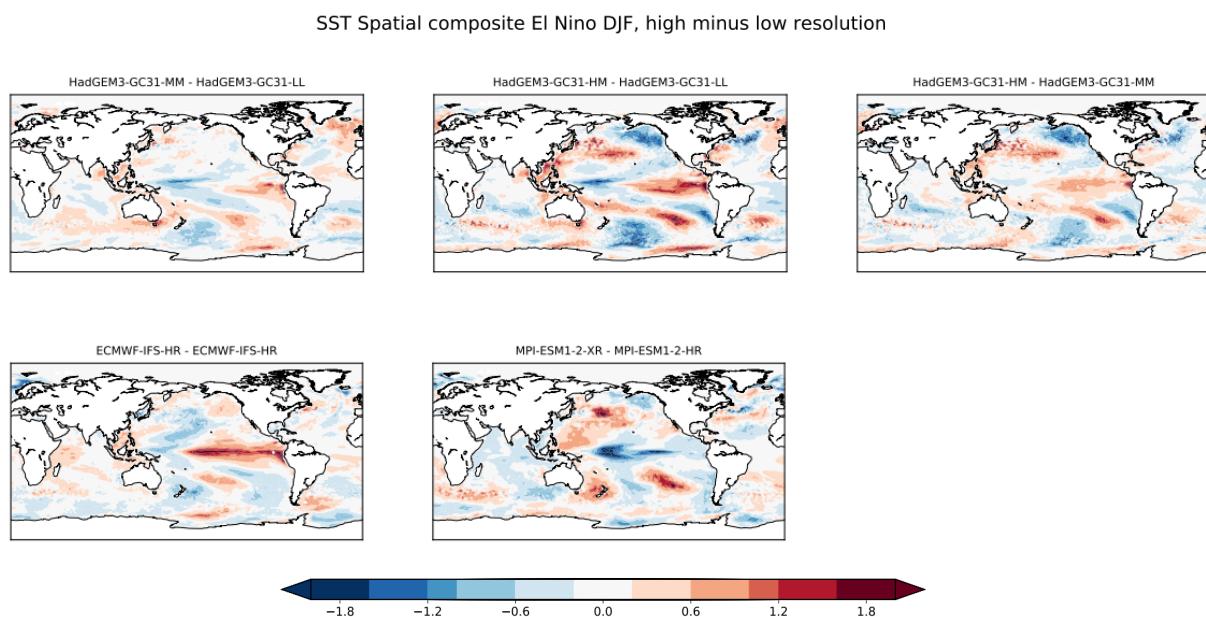
We have applied the Climate Variability Diagnostics Package (CVDP) developed by NCAR's Climate Analysis Section to the analysis of stream-1 simulations. The CVDP is an analysis tool that produces files with major modes of climate variability in models and observations. Among the modes included in this package are ENSO, the Pacific Decadal Oscillation (PDO), the Atlantic Multi-decadal Oscillation (AMO), Northern and

Southern Annular Modes (NAM and SAM), North Atlantic Oscillation (NAO), Pacific North and South American teleconnection patterns (PNA and PSA).

The CVDP calculates and saves as NetCDF time series of the different modes of variability, their spatial patterns and power spectra. CVDP also creates a web interface to be able to analyse the output in a web page. Example output for some PRIMAVERA models can be found in the following link:

<http://exporter.nsc.liu.se/2d2d9a3d8ad84338868530edf1937f73>

The CVDP also computes climatological fields, standard deviation and trend maps of mean sea level pressure, precipitation, surface air temperature and sea surface temperature. We have also implemented regridding programs to assess changes in spatial patterns on different variability modes due to increased model resolution.



**Figure 3.2.1.1:** Changes in SST anomalies related to ENSO due to changes in model resolution for HadGEM3-GC3.1, ECMWF-IFS and MPI-ESM1-2. Units are degrees K.

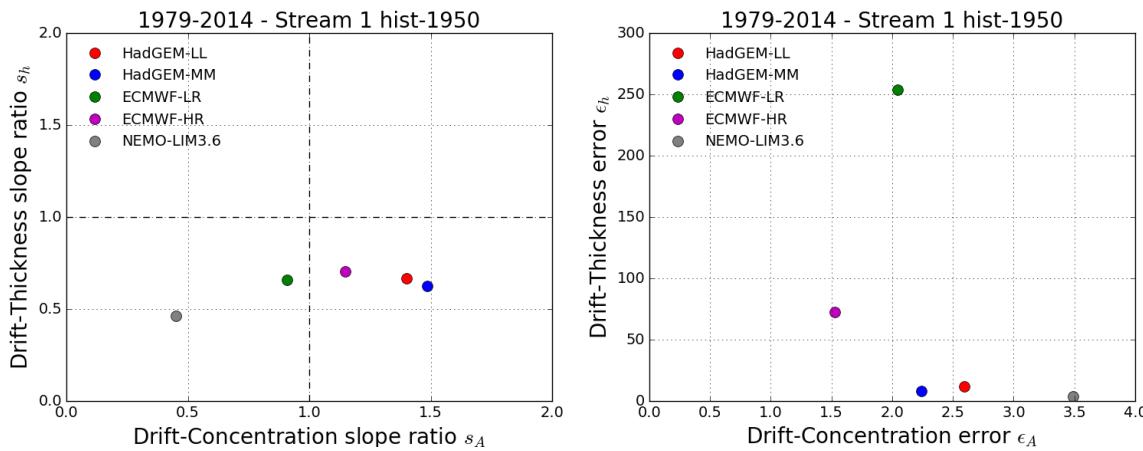
### 3.2.2 Sea-ice drift - strength metric [UCLouvain]

Both seasonal and recent long-term changes in sea-ice drift speed are primarily correlated to changes in sea-ice strength, more precisely sea-ice concentration and thickness (Olason and Notz, 2014). On seasonal time scales, when sea-ice concentration is low (from June to November), drift speed increases with decreasing concentration, while for high concentration (from December to March), drift speed changes are largely driven by changes in thickness (higher drift speed with lower thickness). We have developed the sea-ice drift - strength metric that allows quantifying these relationships. In particular, two different sub-metrics have been developed for the two drift-concentration and drift-thickness relationships: (i) a slope ratio, which is the

ratio between the modelled and observed drift - concentration and drift - thickness slopes (the closer the slope ratio to 1, the better the agreement between the modelled and observed relationships), (ii) an error, which is the mean normalized distance between modelled and observed points (the lower the distance, the closer the model to observations). A detailed description of these sub-metrics using the NEMO-LIM3.6 ocean - sea-ice model (forced by atmospheric reanalysis) and observational / reanalysis datasets is provided in Docquier et al. (2017).

The codes of this metric are written in Python programming language and are available on JASMIN as well as on the [PRIMAVERA svn repository](#) so all PRIMAVERA members can use it. [A web page](#) explaining this metric is available on the PRIMAVERA wiki. The metric has been refactored ([URL](#)) and is now close to be fully implemented into ESMValTool. It constitutes the main example of Deliverable D1.1.

The sea-ice drift - strength metric has been applied to PRIMAVERA Stream-1 coupled hist-1950 simulations, in particular four different model configurations, i.e. HadGEM3-LL, HadGEM3-MM, ECMWF-LR and ECMWF-HR. Figure 3.2.2.1 below shows the sea-ice drift - thickness slope ratio as a function of drift - concentration slope ratio (left panel), as well as the drift - thickness error as a function of drift - concentration error (right panel), for the four different model configurations and NEMO-LIM3.6. No clear sensitivity to model resolution is found related to this metric, but inclusion of more model configurations in this study would be needed to increase the robustness of our findings. In terms of slope ratios (Figure 3.2.2.1, left panel), ECMWF-LR and ECMWF-HR are the closest model configurations to observations (they are close to 1:1). In terms of errors (right panel), HadGEM3-MM and NEMO-LIM3.6 have the lowest drift - thickness errors, while ECMWF-HR has the lowest drift - concentration error, compared to observations. Note the unrealistically high error of ECMWF-LR (~250%), mostly due to high sea-ice thickness compared to observations and reanalysis.

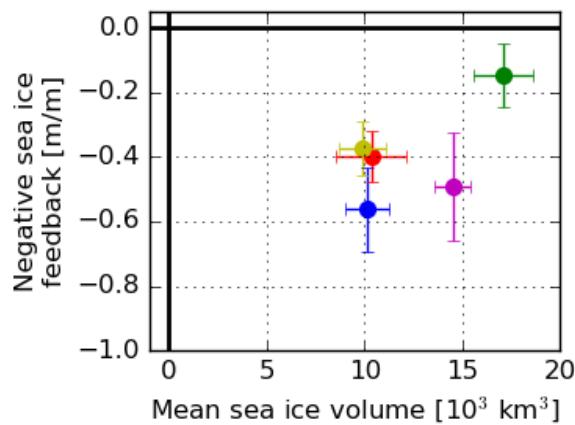


**Figure 3.2.2.1:** (left) Sea-ice drift-thickness slope ratio as a function of sea-ice drift-concentration slope ratio. (right) Sea-ice drift – thickness error as a function of sea-ice drift-concentration error (in %) for 4 PRIMAVERA model configurations (hist-1950) and NEMO-LIM3.6. Drift – concentration relationships are built using OSI SAF satellite observations, while drift – thickness relationships are based on PIOMAS reanalysis. All results are averaged over the period 1979-2014.

### 3.2.3 Ice formation efficiency diagnostic [UCLouvain]

A well-known feature of the Arctic climate is the existence of a powerful positive ice-albedo feedback, in which sea-ice plays a central role (Screen and Simmonds, 2010). Less known is the existence of a set of strong negative feedbacks (Notz and Bitz, 2017) that are mitigating the sea-ice loss in response to external forcing. The simplest of these feedbacks is the ‘negative ice growth - ice thickness feedback’: since heat conduction fluxes are inversely proportional to sea-ice thickness (all other things being equal), thin ice has a tendency to grow faster than thick ice. That is, late-summer negative anomalies of sea-ice thickness will be partly recovered during ice growth. We constructed a diagnostic named the ‘Ice Formation Efficiency’ (IFE) to measure this feedback indirectly. The IFE is estimated as the regression between minimum sea-ice volume for one year and the volume gained until the following winter maximum. It has been applied to the CMIP5 ensemble in a recent publication (Massonnet et al., 2018). The main finding was that (1) the IFE, along with its corresponding diagnostic for the melting season, are good predictors for the interannual to multi-decadal sea-ice volume changes in the Arctic and (2) that the IFE is a strong function of the background mean sea-ice state, and less so of the type of sea-ice model used.

The diagnostic is available from the ESMValTool private repository (<https://github.com/ESMValGroup/ESMValTool-private>) and from F. Massonnet’s Github Page (<https://github.com/fmassonn/paper-arctic-processes>) as a Python function. The function returns an estimate of the IFE along with error statistics and requires as input the monthly time series of Arctic sea-ice volume north of 80°N.



**Figure 3.2.3.1:** Ice Formation Efficiency (IFE) as a function of mean sea-ice volume for HadGEM3-LL (red), HadGEM3-MM (blue), HadGEM3-HM (yellow), ECMWF-LR (green), ECMWF-HR (pink).

The IFE diagnostic was applied to 5 model configurations from PRIMAVERA (HadGEM3-LL, HadGEM3-MM, HadGEM3-HM, ECMWF-LR, ECMWF-HR; Figure 3.2.3.1). The dependence of the IFE to the background mean sea-ice state (defined as the annual-mean sea-ice volume north of 80°N), initially found in Massonnet et al. (2018) with CMIP5 models, was

confirmed with the PRIMAVERA data. Since the high-resolution versions of the models tend to have lower climatological sea-ice volumes because of increased poleward oceanic heat transport, they also tend to simulate higher absolute values of the IFE. This corresponds to the notion that the negative ice growth - ice thickness feedback is stronger for thin ice (Bitz and Roe, 2004). Therefore, we conclude that across the range of resolutions studied, the efficiency of vertical sea-ice formation is sensitive to horizontal resolution, but this link is indirect: enhanced sea-ice formation efficiency is a consequence of lower baseline sea-ice thickness, which itself is a consequence of enhanced poleward heat flux that is caused by the resolution increase.

### 3.2.4 Northern hemisphere blocking [UREAD, CMCC]

Introduction: Multiple studies have shown that an increase in atmospheric resolution generally benefits the representation of blocking in climate models, though blocking simulation is also sensitive to a number of other factors including vertical resolution, sea surface temperature, the representation of orography, physical parametrisations, and the dynamical core numerical scheme (Woollings et al., 2018). This continues to hold true as the resolution in atmosphere-only simulations is increased from O(100km) to O(20km), but several models continue to exhibit sizeable biases even at about 20km resolution (Schiemann et al., 2017), and improvements seen in blocking have been shown to be due to compensating biases in the representation of eddies in one high-resolution model (Davini et al., 2017). Here, we evaluate the representation of blocking in the PRIMAVERA Stream-1 multi-model ensemble focussing in particular on whether benefits seen at higher resolution in atmosphere-only simulations are also seen in coupled atmosphere-ocean simulations.

Blocking indices and metrics: We use a one-dimensional and a two-dimensional blocking index. The one-dimensional index detects instantaneous blocking along the central blocking latitude (CBL) as in Athanasiadis et al., 2014. For each calendar season, the CBL is defined by ERA-Interim. The presented zonal profiles (Figures 3.2.4.1 and 3.2.4.2) have been smoothed with a 3-point running average. For the one-dimensional blocking analysis, daily Z500 data from each model have been interpolated from the respective native grid to a coarser regular grid ( $2.5^\circ \times 2.5^\circ$ ). The one-dimensional index is applied both without and with mean-bias correction following Scaife et al., 2010. At each grid point, mean bias correction has been accomplished by subtracting the model daily climatology (smoothed by 30-day running average) to compute Z500 daily anomalies and subsequently adding the respective daily observed (reanalysis) climatology.

The two-dimensional blocking index is the Absolute Geopotential Height (AGP) index as used by Scherrer et al., 2006. For a grid box to be identified as blocked, three criteria need to be fulfilled by the Z500 geopotential height field; (i) the climatologically negative equator-to-pole gradient must be reversed to the south of the grid box, (ii) there must be westerlies to the north of the grid box, and (iii) conditions (i) and (ii) must hold for at least 5 consecutive days.

Data: The two-dimensional blocking analysis has been applied to four (4) PRIMAVERA models (EC-Earth3, ECMWF-IFS, MPI-ESM-1-2, HadGEM3-GC31) and the respective atmosphere-only (highresSST-present) and coupled (hist-1950) historical Stream-1 simulations, as listed in Table 3.2.4.1. The simulated blocking is compared against that seen

in reanalyses data (NCEP-NCAR, ERA40, ERA-Interim). For the one-dimensional blocking, the results shown here are for (3) PRIMAVERA models (ECMWF-IFS, MPI-ESM-1-2, CMCC-CM2) with other details as above, that is for both the atmosphere-only (highresSST-present) and coupled (hist-1950) historical Stream-1 simulations.

**Table 3.2.4.1:** Ensemble of Stream-1 simulations the 2D blocking index has been applied to.

Experiment (coupling)	center/Model	Resolution atmosphere	Resolution ocean	Notional resolution
highresSST-present (forced)	MOHC HadGEM3-GC31	N96	(0.25° daily HadISST2)	L
		N216		M
		N512		H
	ECMWF IFS	Tco199		L
		Tco399		H
	EC-Earth-Consortium EC-Earth3	TL255		M
		TL511		H
	MPI-M MPI-ESM1-2	T127		L
		T255		M
hist-1950 (coupled)	MOHC HadGEM3-GC31	N96	1° (1/3° tropics)	LL
		N216	1/4°	MM
		N512	1/4°	HM
		N512	1/12°	HH
	ECMWF IFS	Tco199	1°	LL
		Tco399	1/4°	HM
	EC-Earth-Consortium EC-Earth3	TL255	1°	ML
		TL511	1/4°	HM
	MPI-M MPI-ESM1-2	T127	TP04	LM
		T255	TP04	MM

Selected results: Figure 3.2.4.1 shows results for the one-dimensional blocking obtained with and without mean-bias correction for the atmosphere-only simulations. In agreement with previous studies, models tend to underestimate Euro/Atlantic blocking considerably and this result is not found to be sensitive to model resolution. For all models, large part of the blocking bias disappears after mean-bias correction indicating that the former is reflected in (due to) the mean geopotential height bias of each model. Considering the results without mean-bias correction, the blocking biases are generally larger in the respective coupled

simulations (Figure 3.2.4.2) as opposed to the atmosphere-only simulations (Figure 3.2.4.1), forced with observed SSTs, arguably due to the resulting SST biases in the former. There is work in progress aiming at attributing biases in temperature and precipitation (as well as in the frequency and severity of weather extremes) to the biases found in blocking.

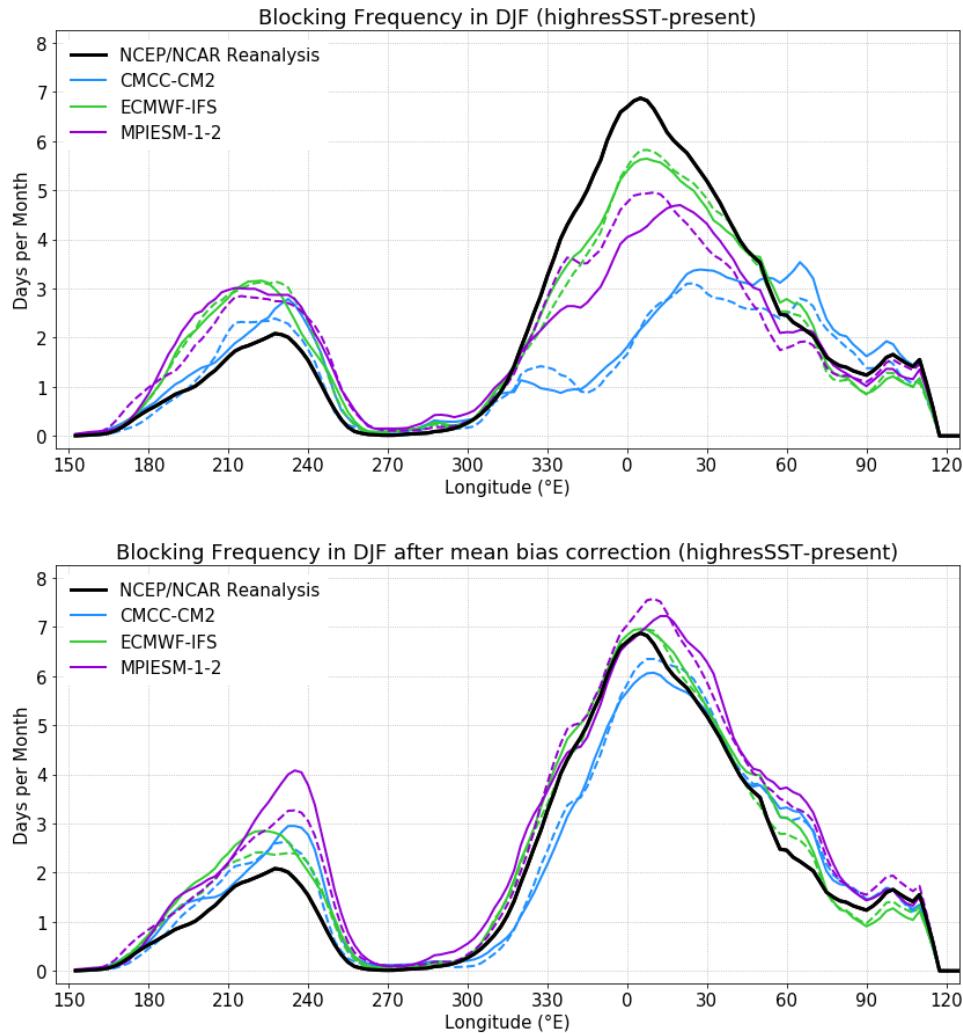
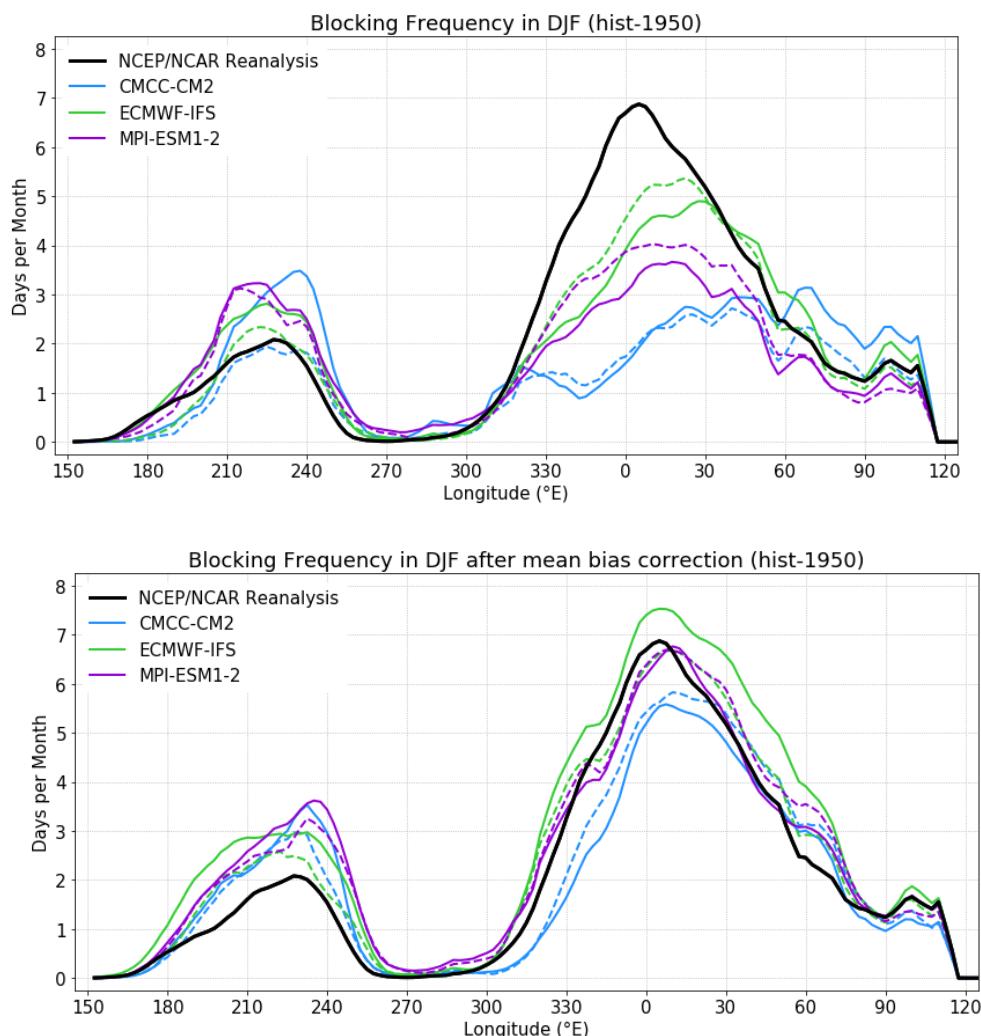


Figure 3.2.4.1: 1D blocking index for a number of atmosphere-only simulations (highresSST-present) at high-resolution (solid), low-resolution (dashed), and for NCEP/NCAR reanalysis (black). The bottom panel shows the blocking index after correcting the model mean state to the reanalysis mean state, and the top panel shows the blocking index for the uncorrected model data.

We further assess the simulated blocking frequency and geographical pattern of blocking distribution by means of the two-dimensional index. The frequency of blocked days over European domains for winter and summer is shown in Figure 3.2.4.3. During winter, the PRIMAVERA Stream-1 simulations underestimate blocking frequency, which is a long-standing bias in climate models, yet some models attain about 80% of the observed blocking frequency. There appears to be a small improvement with resolution for the four coupled models, but not in the forced models. Three of the four models simulate higher blocking frequencies than seen in the CMIP5 multi-model mean, yet this is not seen when considering the CMIP5 mean from the same modelling centers only. During summer, the models

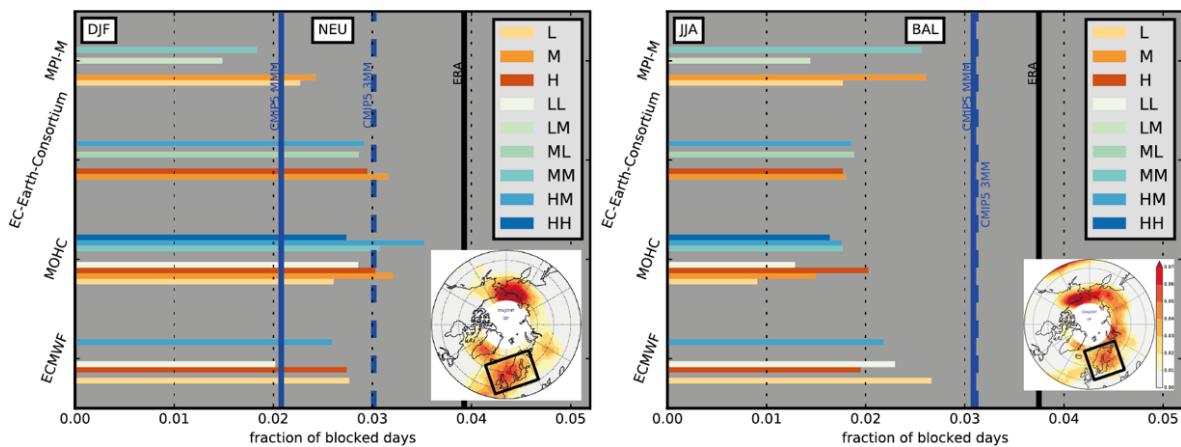
underestimate European blocking by about 50%, more so than in winter and also more than seen in the CMIP5 models. There is no systematic resolution sensitivity in this metric in summer.

We proceed by evaluating the geographical pattern of blocking occurrence (Figure 3.2.4.4). This figure shows scatter plots of the root-mean-square error and the spatial correlation with respect to the reanalysis climatology so that the better the agreement between a model simulation and the reanalysis, the further will the corresponding entry be located in the lower right of the scatter plot. During winter, an improvement in the simulated pattern can be seen in three out of the four forced simulations, and in the coupled simulations the improvement is larger and can be seen for all four models. It can also be seen that the high-resolution PRIMAVERA Stream-1 models show an improvement with respect to the CMIP5 multi-model mean, even if only models from the same modelling centers are considered. Similar conclusions can be drawn for summer. Across the four models considered here, a model's ability to capture the pattern of blocking occurrence in winter is no indication of how well it will perform in summer, as seen, for example, by the comparatively close agreement with reanalysis of the MOHC HadGEM3-GC31 model in winter and the comparatively large disagreement with reanalysis of the same model in summer.

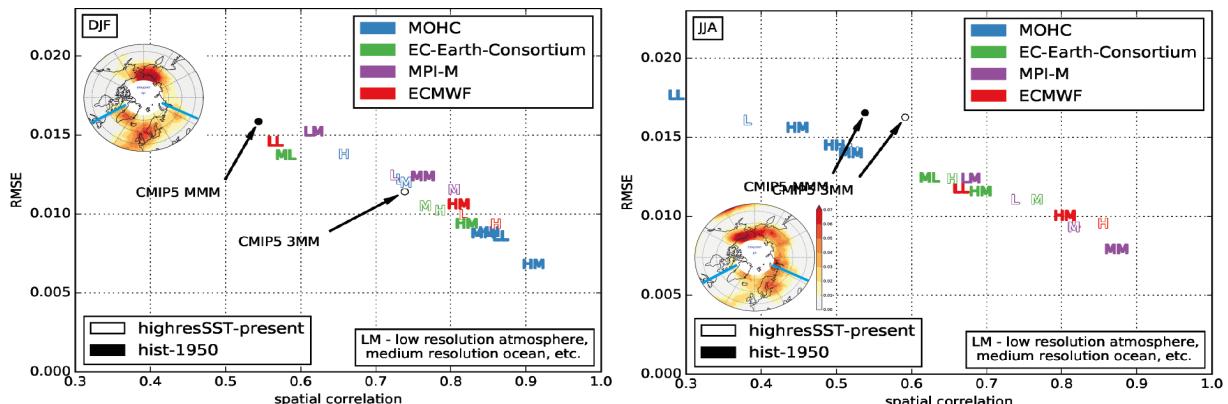


**Figure 3.2.4.2:** As in Fig. 3.2.4.1 but for the respective coupled Stream-1 simulations (hist-1950).

Eventually, the respective metrics as implemented in the ESMValTool, will allow the user to produce results for a specific area of interest by defining the respective Lat-Lon box.



**Figure 3.2.4.3:** Domain-mean blocking frequency for (left) DJF and a Northern European domain (reanalysis climatology and domain in inset) and (right) JJA and a Baltic domain. Forced and coupled models are shown in terms of their notional resolutions (Table 1). CMIP5 MMM denotes the CMIP5 multi-model mean and CMIP5 3MM denotes the mean over the models from three centers only (EC-Earth Consortium, MOHC, MPI-M). ERA denotes the reanalysis mean.



**Figure 3.2.4.4:** Blocking frequency root-mean square error and spatial correlation with respect to reanalysis climatology over the Atlantic European sector (insets show reanalysis climatology and domain) for (left) DJF and (right) JJA and for forced and coupled simulations as in Table 1. CMIP5 MMM denotes the CMIP5 multi-model mean and CMIP5 3MM denotes the mean over the models from three centers only (EC-Earth Consortium, MOHC, MPI-M).

### 3.2.5 Position and strength of the North Atlantic eddy-driven jet [CMCC, UREAD]

**Introduction:** Despite improvements over previous generations of climate models (Hannachi et al., 2013; Iqbal et al., 2018), an accurate representation of the pulses and the latitudinal shifts of the North Atlantic eddy-driven jet remains a challenge, upon which depends the representation of important aspects of European climate, including weather extremes and the frequency of severe prolonged anomalies, such as cold spells, dry spells and heat waves

(e.g. Woollings, 2010; Buehler et al., 2011). Thus, assessing the representation of the jet climatology and variability in the current generation of climate models is fundamental for understanding the strengths and the weaknesses of these models for European applications.

In PRIMAVERA WP1 we developed a powerful and concise metric for this purpose.

Methodology: To identify the eddy-driven jet, daily mean u-wind field at 850 hPa was extracted from each highresSST-present simulations and interpolated from native model grids to a  $2.5^\circ \times 2.5^\circ$  regular grid. Following Woollings et al. (2010) and Woollings et al. (2018) and applying an additional orography mask (to account for the 850 hPa isobaric level being underground over most of Greenland). Jet latitude and jet speed are defined over the domain 0–60W and the respective bivariate distributions were determined for each simulation and the NCEP/NCAR reanalysis, binned at  $2.5^\circ$  latitude and  $1.0 \text{ m s}^{-1}$  speed and smoothed by a PDF kernel (Silverman, 1986). The results are largely insensitive to the kernel estimation method, the practical effect of which is smoothing.

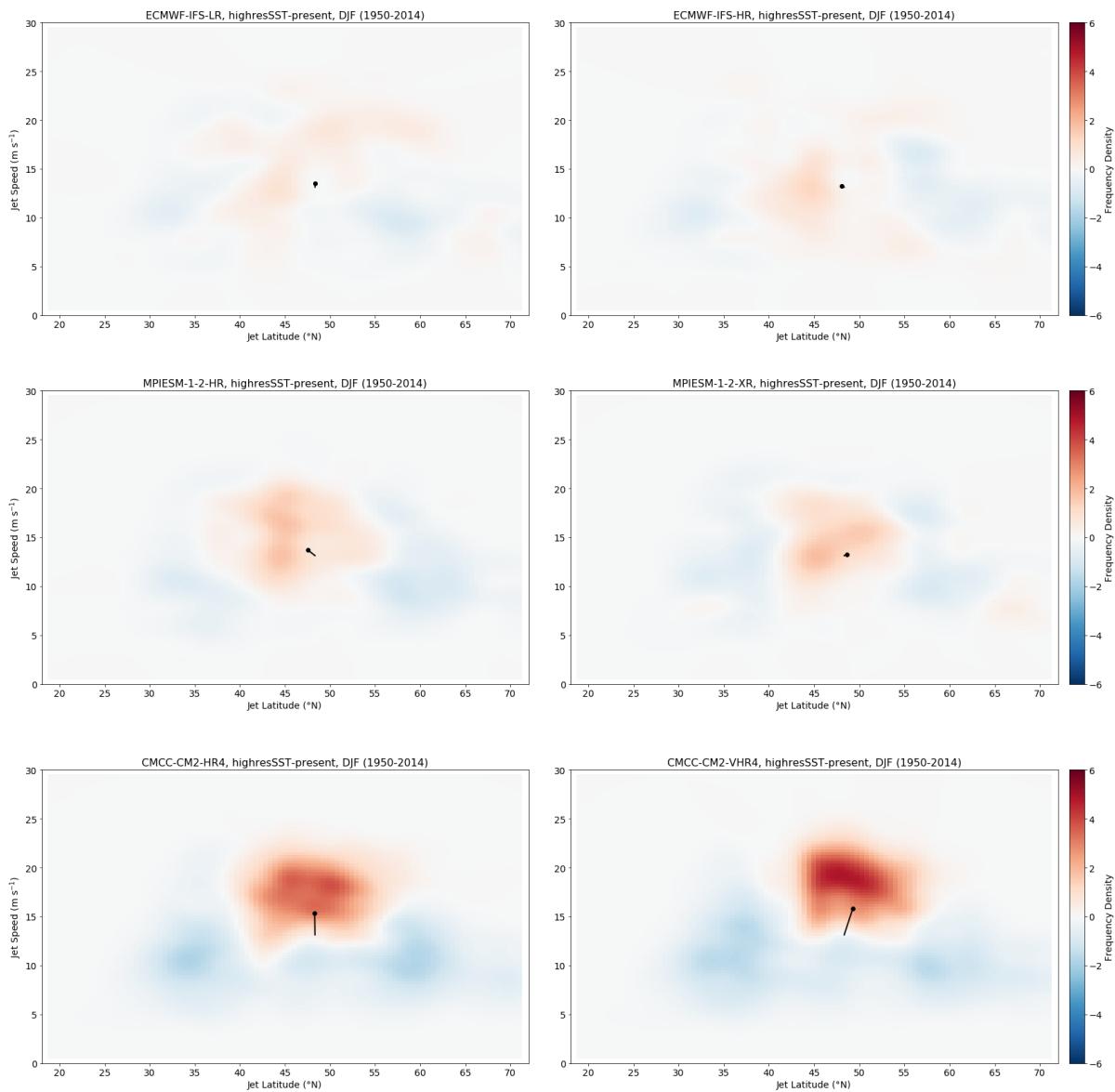
Results: Here we present results for three (3) PRIMAVERA models (ECMWF-IFS, MPI-ESM-1-2 and CMCC-CM2) and for both the atmosphere-only (highresSST-present) and coupled (hist-1950) historical Stream-1 simulations conducted with these models at two different resolutions (low-resolution: LR, high-resolution: HR). Nevertheless, the scope here is not to assess the role of increasing model resolution but to document the metric and its use.

The model biases displayed in Fig. 3.2.5.1 (LR left, HR right) make evident that there are significant differences across models in the representation of the North Atlantic jet. While the ECMWF-IFS model exhibits small biases in the bivariate distribution of the jet, capturing well the trimodality of the jet latitude distribution, the CMCC-CM2 model has large biases in the distribution of both the jet speed and jet latitude, indicating also a misrepresentation of the Euro-Atlantic low-frequency variability patterns, referring to blocking (§ 3.2.4) and circulation regimes (§ 3.2.9). More specifically, the CMCC-CM2 model, which is used here for demonstrating the power and the interpretation of the metric, exhibits an overly strong jet that also has a narrower meridional distribution compared to the observed jet (Fig. 3.2.5.1, bottom row). Given that the jet moves to the south during blocking over Greenland and to the north when blocking occurs at the eastern part of the North Atlantic, the above-described jet biases signify, also, large biases in the climatological blocking distribution.

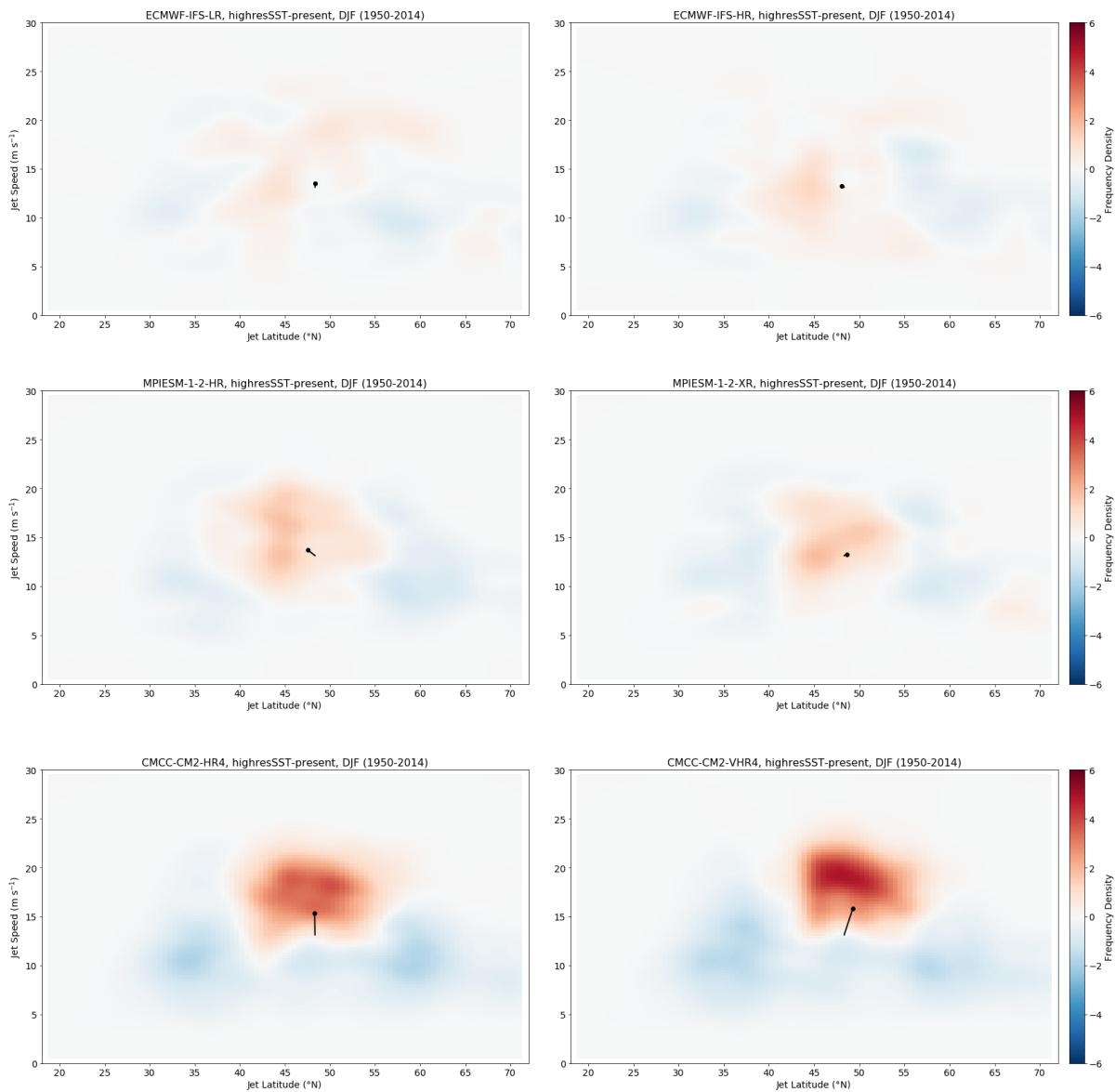
It is fair to ask whether a single realisation/member from a model configuration is sufficient for assessing its jet representation. Fortunately, results from a recent study (Kwon et al., 2018) using large ensembles, referring to the Large Ensemble Simulations (LENS) run with the Community Earth System Model (CESM) v.1, provide evidence that the jet statistics over a similar historical period (1951–2005) exhibit little intra-ensemble spread. Stream-2 PRIMAVERA simulations provide us the possibility to re-assess the intra-ensemble spread and use the latter to evaluate the respective uncertainty.

This gives us confidence that for certain diagnostics single-member analyses are a viable option.

Forthcoming research Eddy-driven jet tilt diagnostics. Extension of these analyses to coupled simulations upon delivery by PRIMAVERA partners.



**Figure 3.2.5.1:** Model biases in the bivariate distribution of the North Atlantic eddy-driven jet for the low- and high-resolution version of three PRIMAVERA models. Here results from the atmosphere-only (highresSST-present) simulations are shown. The jet latitude (x-axis) and jet speed (y-axis) correspond to the respective indices defined daily as described in the text. Biases are computed in respect to the NCEP/NCAR reanalysis for the same period (1950–2014). The black line segment at the center of each panel connects the mean position and strength of the jet in the reanalysis and in the model (round marker). Units: relative frequency density multiplied by  $10^3$ . Distributions are estimated by a PDF kernel.



**Figure 3.2.5.2:** As in Fig. 3.2.5.1 but for the respective coupled Stream-1 simulations (hist-1950).

### 3.2.6 Tropical cyclone tracker [BSC]

Regarding tropical cyclone tracking, the reader is kindly pointed also to Section 3.2.12.

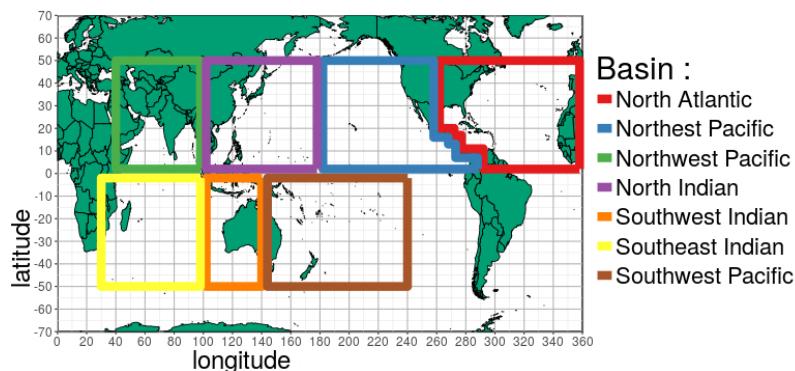
This tool detects the formation and propagation of tropical cyclones in climate simulations in order to provide an estimate of global tropical cyclone activity. The core tracking algorithm is derived from the [GFDL Vortex Tracker V3.5b](#), which was modified to read PRIMAVERA data and complemented with some post-processing tools (statistics at the global, basin and individual storm level and plotting tools). The tracker has also been modified to run with bsub on JASMIN in order to take advantage of the multi-core environment. Maximum or minimum values of different atmospheric fields (3-hourly / 6-hourly) are used to track the position of the hurricane center:

- mean sea level pressure (mandatory)
- vorticity or wind velocity at 850 and 700 hPa,
- geopotential height at 850 and 700 hPa,
- wind speed at 10m.

Wind velocities in the mid-troposphere are also mandatory to estimate the subsequent position of the storm and to construct the track. Winds at 850 or 700 hPa can be used, but winds at 500 hPa give the best result. The BSC Cyclone Tracker is stored in the [PRIMAVERA repository](#) under svn. It can be viewed on a web browser at this [URL](#).

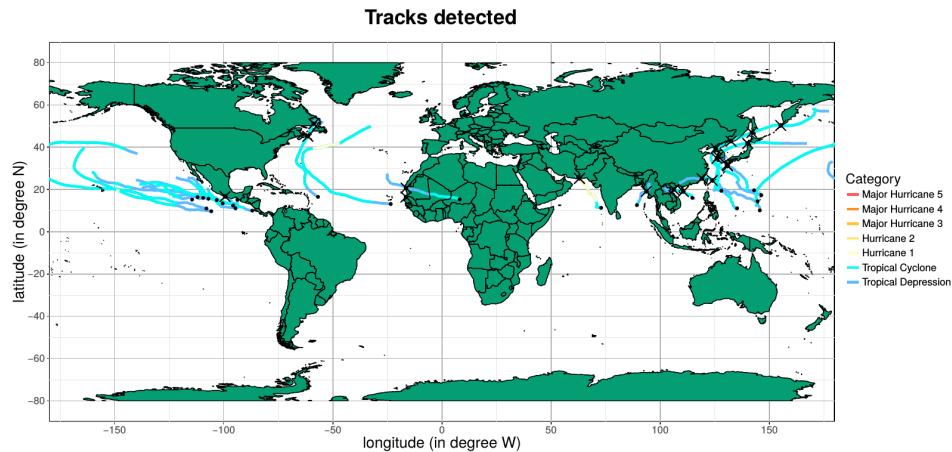
The full documentation on how to run the tracker can be found [online](#), and more information about the original GFDL Vortex tracker can be found on the DTC (Development Testbed Center) [Users' guide](#).

The tracker provides an estimate of the cyclone center position (latitude and longitude) along with metrics for intensity and structure at each time-step, such as the minimum surface pressure in the center of the storm, the maximum surface winds near the center of the storm and the radii of maximum wind speed in the four different quadrants. The algorithm can also construct the cyclone phase diagrams for each detected cyclone if the necessary fields are present. The tracker also includes some post-processing tools analysing the cyclones that have been detected, such as the integrated kinetic energy, the storm lifetime, the full yearly activity sorted by basins (see Figure 3.2.6.1) and the number of storms making landfalls.



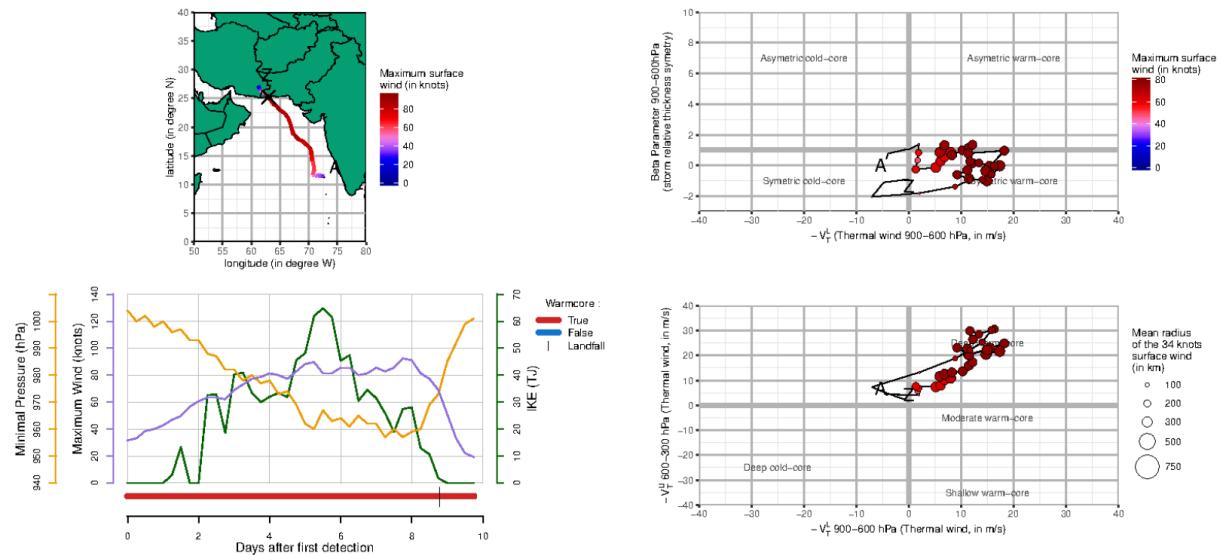
**Figure 3.2.6.1:** The different basins of tropical cyclone activity considered by the tropical cyclone tracker.

The figure below (Figure 3.2.6.2) shows the tracks identified in a 6-month simulation performed with EC-Earth3. We note that both the general level of simulated activity as well as the intensity of the storms are much lower than in the real world, but that is standard for climate models. We note however that tropical cyclone activity is detected in all the major basins of the northern hemisphere: eastern North Pacific, western North Pacific and North Atlantic.



**Figure 3.2.6.2:** Tropical cyclones detected during 6 months of simulation performed with EC-Earth3.

The figure below (Figure 3.2.6.3) gives an idea of the post-processing tools that have been added to the tracker. It shows the track of a particular cyclone (top left), the Hart diagrams (right column), and the evolution of the intensity of the storm, as measured by the maximum wind speed, minimum surface pressure and integrated kinetic energy (bottom left).



**Figure 3.2.6.3:** Example of information provided by the tracker on the evolution of a given tropical cyclone. Top left: the track and it's intensity (in knots). Bottom left: the maximum surface winds (purple), the minimum surface pressure (yellow) and the integrated kinetic energy (green). Right column: Hart diagrams.

### 3.2.7 Ocean Heat Content [BSC]

The ocean has the largest heat capacity of the Earth system and dominates the Earth's heat balance. Most of the total warming caused by climate change is manifested in increased OHC, which in turn causes thermal expansion and sea level rise. Estimating the OHC, with

the OHC tool, in simulations of the pre-industrial climate, is fundamental for understanding the energy balance occurring in the absence of external forcing and under natural variability, and how the oceanic changes relate to the surface climate variability. OHC estimates in simulations of future climate are important for assessing the oceanic response to the surface forcing, the heat uptake by the ocean and how they impact surface climate change.

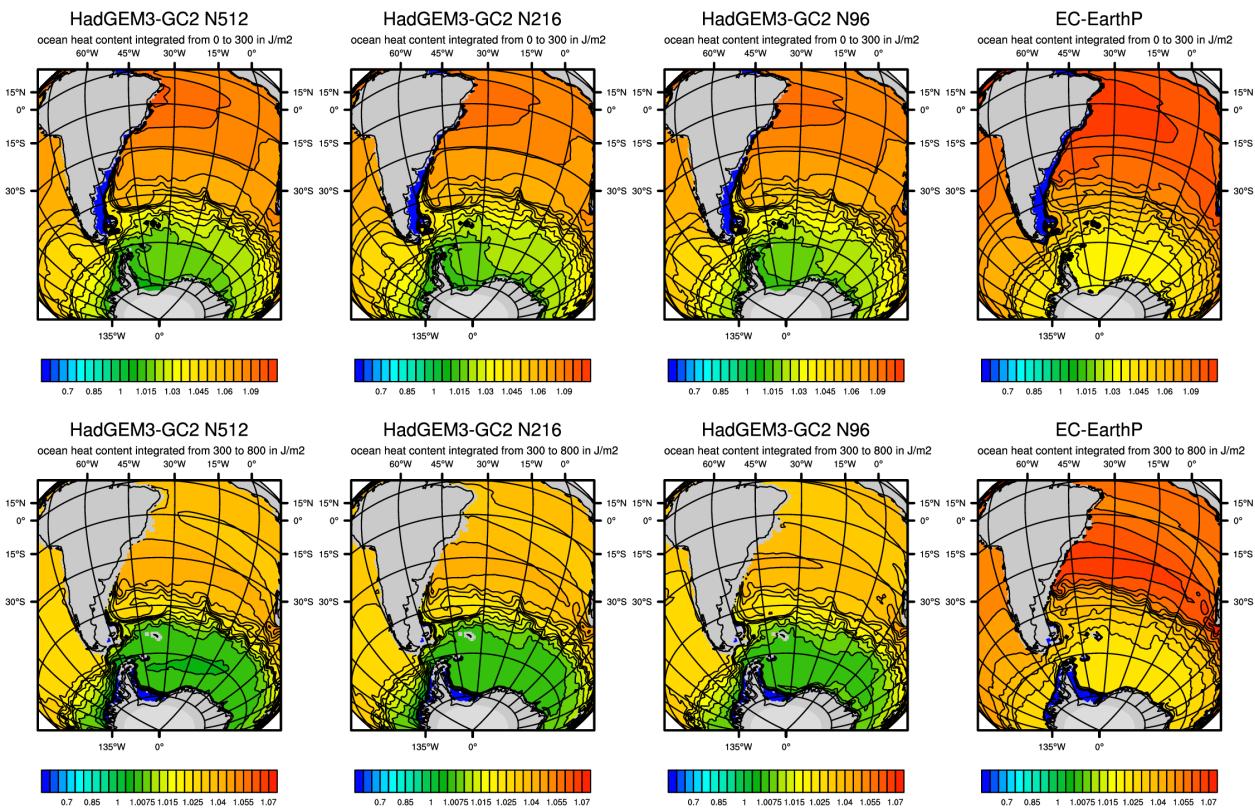
The Ocean Heat Content (OHC) tool is a python-based script which calculates the heat content of the global ocean, or of a given oceanic basin. It needs as input a time series of a 3D ocean temperature field, the land-sea mask, the grid information (the grid box dimensions dx, dy and dz) and a basin mask. The output can be a 2D field of the depth-integrated OHC, or a timeseries of the volume integrated OHC, depending on user request. The basin(s) and depth level(s) are specified by the user. The function requires a certain number of python libraries (for cdo, nco operations) which can be installed in a user-specific environment. It is available to run on JASMIN as a standalone script in:

```
/gws/nopw/j04/primavera1/tools/WP2/Topic1-6_NA/Ocean_heat_content
```

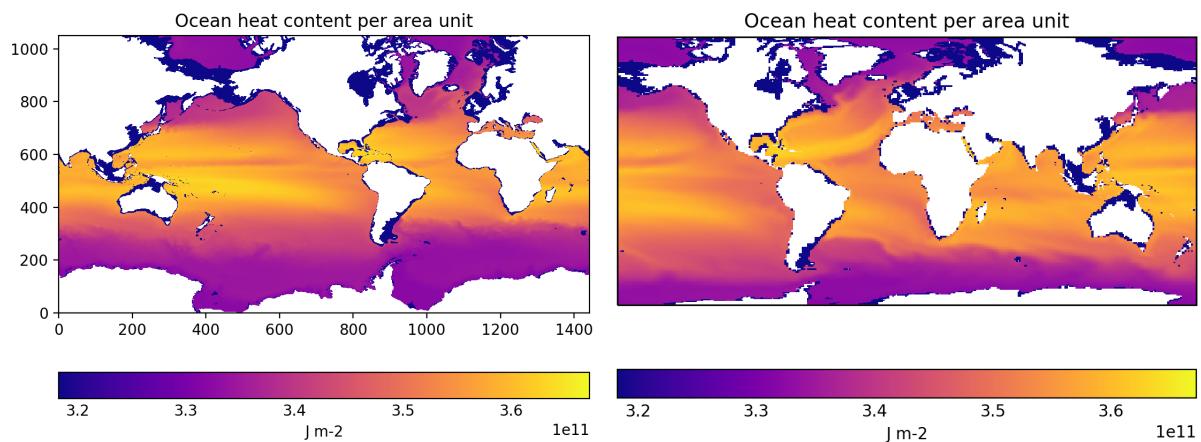
There is also a fully functional version of this tool in the Primavera\_ohc branch of the ESMValTool's GitHub repository (<https://github.com/ESMValGroup/ESMValTool>). This version is fully functional, but the plots are not as polished as they could be and some options (i.e. computing volume integration for multiple basins) are still missing. Also, since ocean cell area information (stored in areacello) is not available in the cmor files for PRIMAVERA runs, it is difficult and time consuming to produce results across different models, as it requires specific code for each model. Support for running this tool is made available through Gitlab at [URL](#).

The Ocean Heat Content tool aims to assess the heat content of the ocean, either integrated over depth (in units of  $J/m^2$ ) or integrated over volume (in units of  $J$ ). It is used to assess anomalies of OHC, which in turn can be used in studies to track oceanic heat uptake and redistribution.

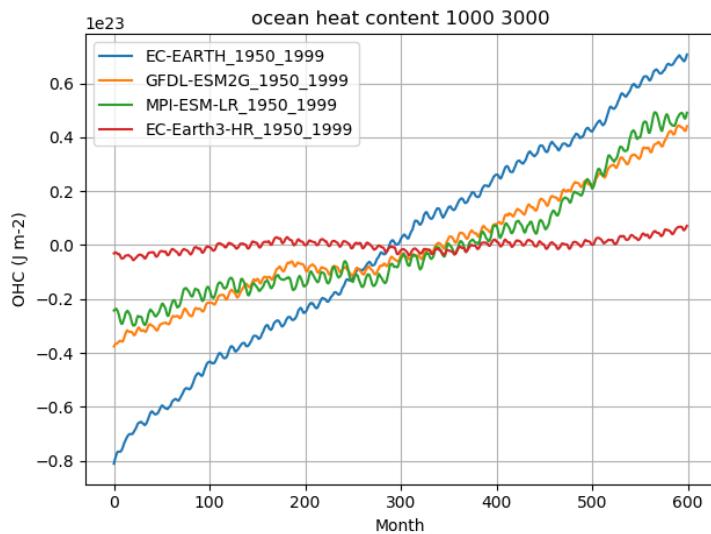
The tool allows comparison of ocean heat content between different PRIMAVERA simulations. This is demonstrated in Figure 3.2.7.1 using pre-PRIMAVERA simulations, while the metric is being applied to Stream-1/2 PRIMAVERA simulations.



**Figure 3.2.7.1:** Top row: Ocean heat content from the surface to 300m (in  $J/m^2$ ). Bottom row: Ocean heat content between 300-800m (in  $J/m^2$ ). The panels in the three leftmost columns show the OHC for the HadGEM3-GC2 model, which all share the same ocean model resolution (ORCA025 grid), but are coupled to atmospheric models with different resolutions, namely N512, N216 and N96. The panels in the rightmost column are from the EC-Earth3P-HR configuration, which uses a ORCA025 grid in the ocean and a T511 grid in the atmosphere.



**Figure 3.2.7.2:** Average November climatology of ocean heat content for the 0-300m layer for the PRIMAVERA EC-Earth3 (left) and the CMIP5 GFDL-ESM2G (right) historical simulations. The climatology is computed for the period 1950–1999.



**Figure 3.2.7.3:** Global ocean heat content for the 1000–3000m layer for the PRIMAVERA EC-Earth3P-HR simulations and 3 different CMIP5 models.

A great advantage of the latest version of the tool is that it has the ability to compare PRIMAVERA against other experiments. Such a comparison is shown below. It shows the PRIMAVERA EC-Earth3-HR control-1950 run along with three CMIP5 historical runs for the period 1950–1999. Adding extra CMIP5 runs to an analysis is just a matter of defining the appropriate simulations in the ESMValTool recipe. This feature will also be available with CMIP6 simulations once CMIP6 simulations become available.

### 3.2.8 K-mean clustering of sea-ice [BSC]

The K-means clustering method partitions data into groups or clusters based on the distance that separates them. The method aims to simultaneously minimize the distance between the members of a given cluster/mode and maximize the distance between the centers of the different clusters. The variable required to run this sea-ice clustering algorithm is either monthly mean of sea-ice thickness or monthly mean of sea-ice concentration. The number of cluster K, found to be three in this case, is determined using a set of clustering validity indices (Fučkar et al.,2016).

K-means clustering is available through the [s2d verification R-package](#). An installation is available on JASMIN through a conda environment. To use it, one needs to execute the following commands:

```
export PATH=/home/users/jvegas/miniconda3/bin:$PATH
source activate s2dv
```

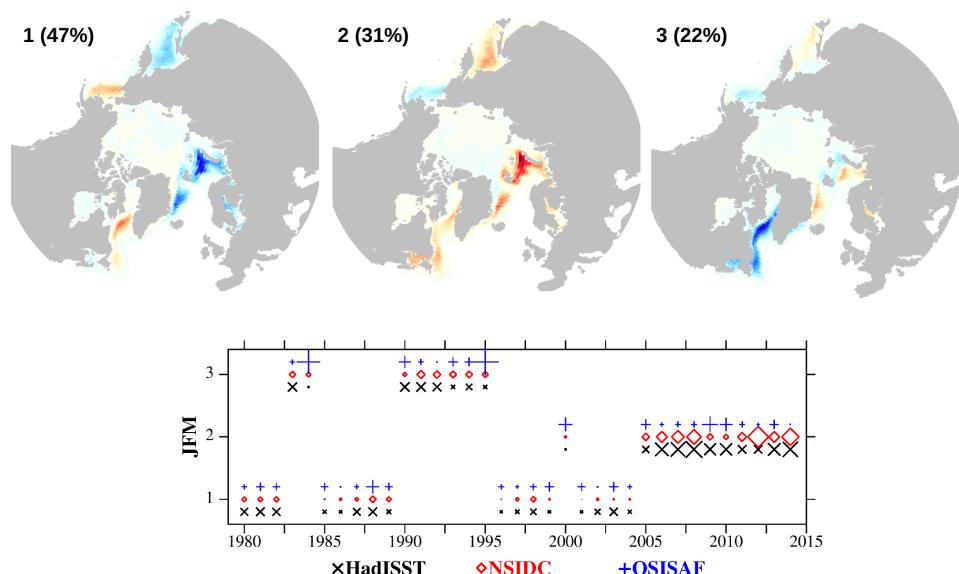
It is installed in the home filesystem instead of the group workspaces for better performance, as the new Jasmin4 group workspaces are very inefficient when dealing with small files.

This branch includes the relevant documentation (see [SealceModes](#)). The branch also contains a vignette explaining the Weather Regimes function. The example in that vignette

can be used as a starting point to run the cluster analysis for sea-ice by modifying the input data and replacing the call to [WeatherRegimes](#) with a call to [SealceModes](#).

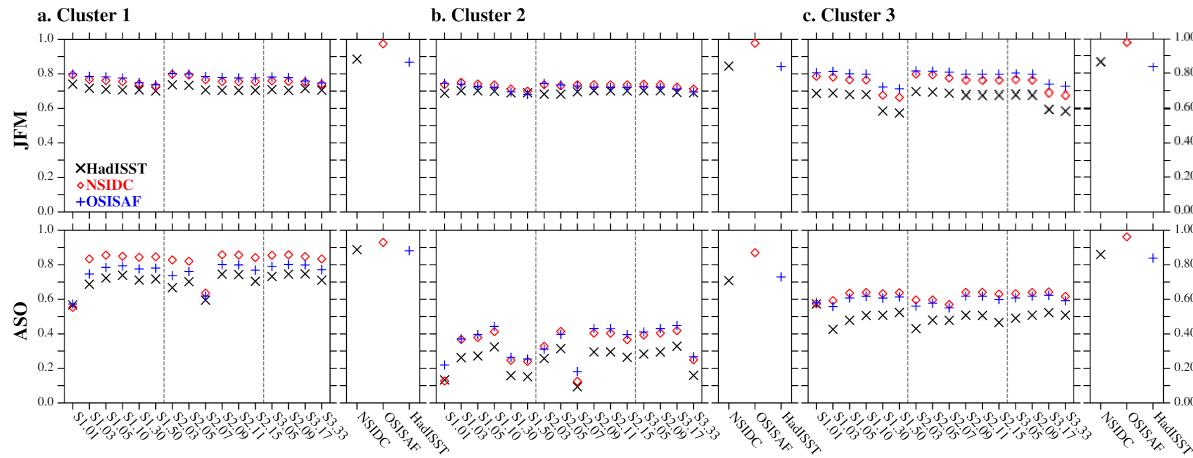
This cluster analysis represents the large-scale climate variability as variations in the occurrence of a limited number of representative spatial patterns (clusters). In other words, a K-means mode, or cluster, is characterized by a spatial pattern and discrete time series of its occurrence. While the EOF analysis has the advantage of parsimony in the context of explaining a maximum amount of data variance, the K-means cluster analysis has the advantage of finding patterns that maximize similarity between clusters and the fields that belong to those clusters, without orthogonality or linearity constraints.

The K-mean clustering technique can be used to extract robust modes of Arctic interannual sea-ice variability and to investigate potential mechanisms leading to these modes. As an example, we show below (Figure 3.2.8.1) the modes of variability of Arctic sea-ice concentration, after removing the long-term trend using a spatially varying second order polynomial fit for the January-March season.



**Figure 3.2.8.1:** Top: Winter (JFM) cluster patterns of anomalous sea-ice concentration in OSISAF (EUMETSAT SAF, 2016), with their respective percentage of occurrence over the period 1979–2015. Clusters in HadISST (Titchner and Rayner, 2014) and NSIDC (Cavalieri et al., 1996) are very similar and therefore are not shown. Bottom: the associated time series of cluster occurrences and the Euclidean distance (RMS difference) between a pattern in a year and the associated cluster, with a larger symbols indicating a larger distance.

This particular metric was used in WP3 to explore the impact of different configurations of the sea-ice thickness distribution (ITD) on the variability of the sea-ice concentration. To evaluate the impact of the ITD configuration on the sea-ice concentration, we compared the spatial correlation of the observed and simulated clusters (Figure 3.2.8.2). A similar analysis is planned by BSC once the melt pond scheme is made available in EC-Earth3P.



**Figure 3.2.8.2:** Spatial correlation coefficients between the observed and the simulated clusters in winter (JFM; top) and summer (ASO; bottom). Each different column corresponds to a different experimental set up. The spatial correlation coefficients between the different observational products for each cluster are also shown.

In winter the impact of the different ITD configuration is small and most simulations capture well the observed cluster of variability; nonetheless, there is a slight drop in the spatial correlation coefficient in the third clusters for a very high number of categories, which appears related to a too-high refinement of the thinnest ice. In summer, the ITD configuration has a slightly bigger impact, especially for the second cluster.

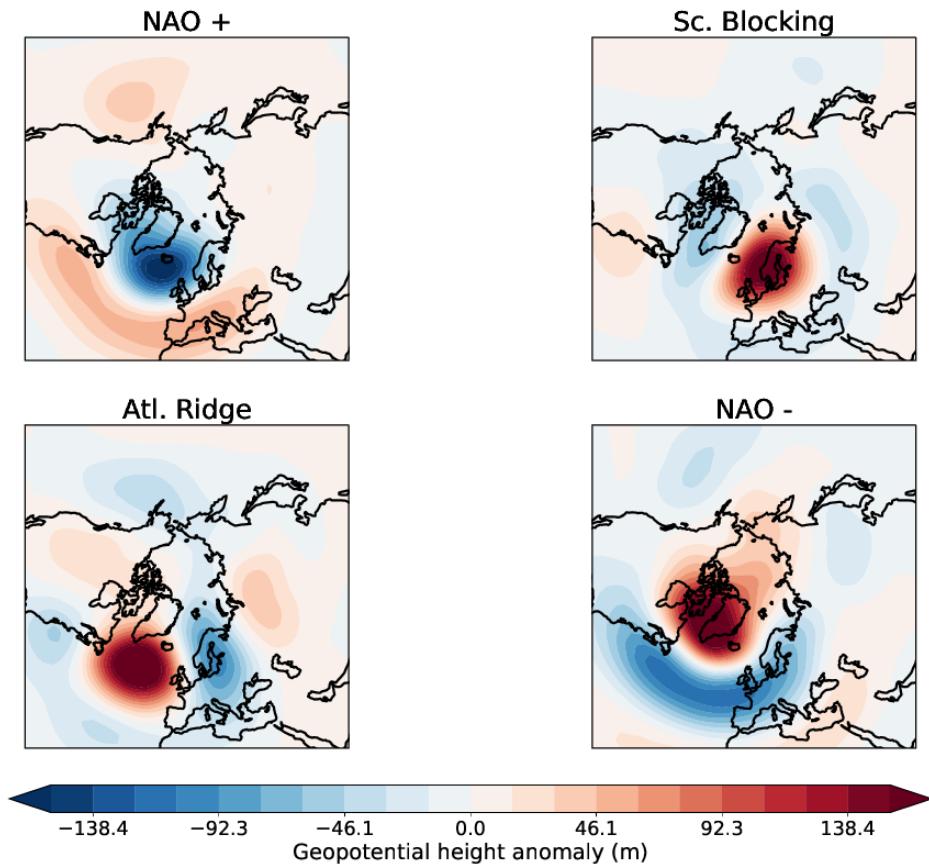
In this exercise, it is difficult to draw any particular conclusions beyond that one category tends to perform the worst, or that a particular setup is inappropriate.

### 3.2.9 Wintertime Euro-Atlantic Weather Regimes [CNR]

Over the last couple of decades evidence has begun to increase that indicates the geometry of the atmosphere's climate attractor exhibits a specific local structure characterised by quasi-persistent weather regimes (Kimoto and Ghil 1993; Cheng and Wallace 1993; Straus et al., 2007; Hannachi et al., 2017).

Weather Regimes (WRs) are persistent dynamical patterns that can last from a few days to two or three weeks. (Straus et al., 2007; Dawson et al., 2012; Hannachi et al., 2017). They have been mostly studied in the wintertime extra-tropical atmospheric circulation for the Pacific-North American and Euro-Atlantic sectors (where the non-normality in the PDF of large scale atmospheric circulation has been confirmed to a high significance) using daily 500 hPa geopotential fields. However, similar analysis have also been successfully applied to different atmospheric fields in other regions and seasons to characterize important climate processes (e.g. Molteni et al., 2003).

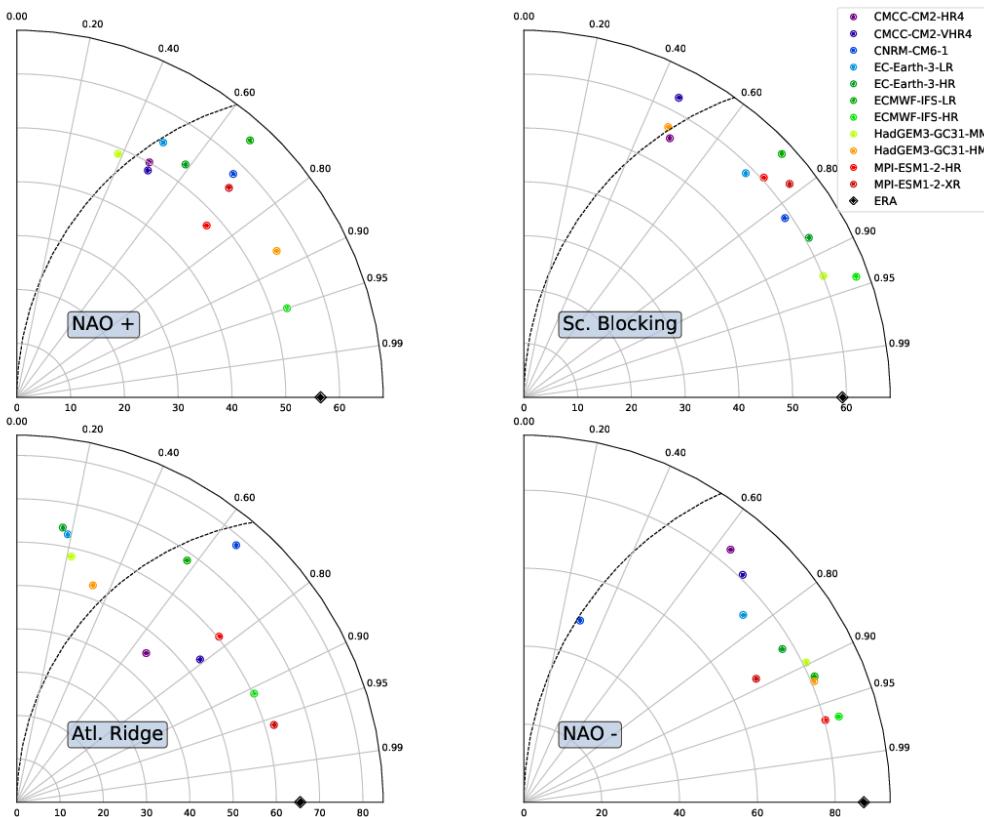
In recent years, there has been increasing interest in studying WRs and how well climate models reproduce them, due to their importance in influencing regional weather patterns and possibly future regional changes in the climate state (e.g. Corti et al., 1999; Cattiaux et al., 2010; Matsueda et al., 2018).



**Figure 3.2.9.1:** Patterns of observed wintertime Weather Regimes in the Euro-Atlantic sector as seen from ERA-Interim reanalysis in the period 1979–2010.

In the following we describe the functioning of WRtool, a tool for the assessment of Weather Regimes in climate models. After a technical description of the metric itself, we will show its application to the PRIMAVERA Stream-1 coupled simulations and briefly discuss the main results.

The analysis is performed on daily 500 hPa geopotential height fields. First of all, the daily anomaly field is computed subtracting the model mean daily climatology and optionally detrending it. The detrending is necessary for transient runs. The user defines a range of years to be considered, the season and the geographical area to be analyzed. An Empirical Orthogonal Function (EOF) decomposition is then applied to the anomaly fields. The number of Principal Components (PCs) to be retained are chosen by the user either selecting the number itself, or setting the total variance to be explained by the PCs. WRs are finally identified by applying a K-means clustering algorithm to the PCs in the phase space spanned by the said EOFs. The number of clusters to be computed can be defined by the user as well. The tool first calculates the WRs for the reanalysis (ERA-Interim or NCEP) and then it performs the WRs computation for each model considered. Each model's clusters are automatically matched with the ones obtained from reanalysis data, minimizing the relative errors for all possible permutations.



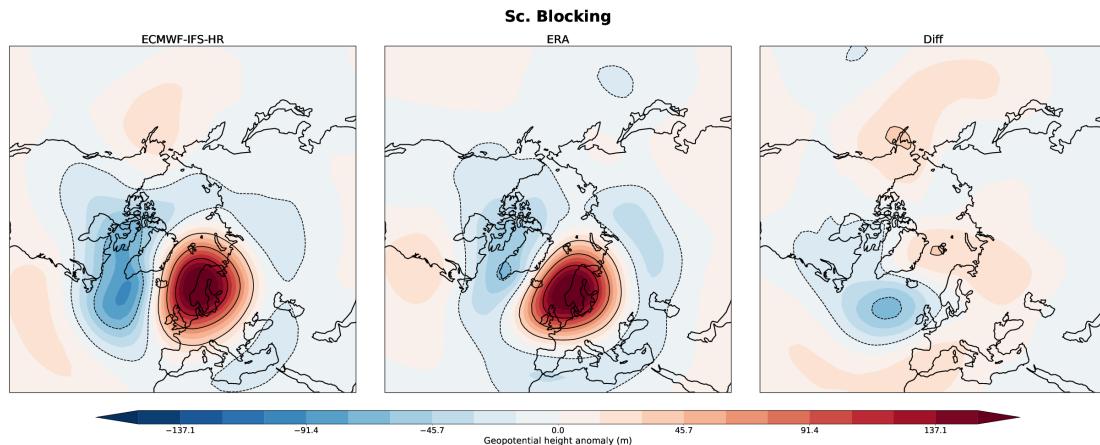
**Figure 3.2.9.2:** Taylor diagrams measuring the performance of each model in simulating the wintertime WRs in the Euro-Atlantic sector.

Here we show, as an example, the application of WRstool to the 500 hPa daily winter fields (DJF) in the Euro-Atlantic sector (EAT) for the ERA-Interim reanalysis in the period 1979–2010. A significant part of the observed variability is obtained with 4 clusters (Figure 3.2.9.1), representing the following WRs: NAO+, Scandinavian Blocking, NAO- and Atlantic Ridge.

The comparison between the observed and simulated WRs is carried out considering different features related to the quality of the model simulations, which lead to the definition of specific metrics. The first metric measures how well the simulated regime patterns resemble the observed ones. A measure of this is given by a Taylor diagram (Taylor, 2001) for each WR, which summarizes the performance of all models with respect to the reanalysis. The radial axis of the Taylor plot measures the internal standard deviation of the pattern, and the polar axis measures the pattern correlation. The difference between the corresponding observed and model clusters provides a measure of the quality of the WR representation in that model. The closer the two corresponding clusters are, the better that given WR pattern is represented.

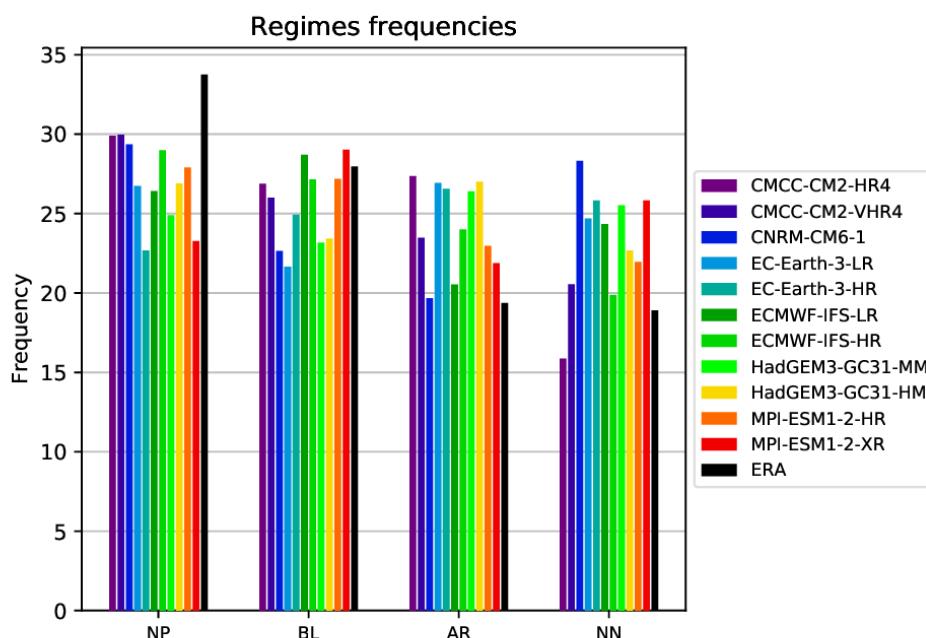
The result of this first metric on the PRIMAVERA Stream-1 coupled simulations is shown in Figure 3.2.9.2. The period considered for the simulations is 1979–2010, as for ERA-Interim. The performance of the coupled models is generally worse than those of the AMIP simulations (shown in the D2.2 deliverable), due to the missing coherent SST forcing. Nevertheless, some models perform very well for some WRs, although very few of them are able to reach a pattern correlation larger than 80% for all WRs. Figure 3.2.9.3 shows the

comparison with ERA-Interim for the ECMWF-HR Scandinavian Blocking regime, which is the closest to the observed one.

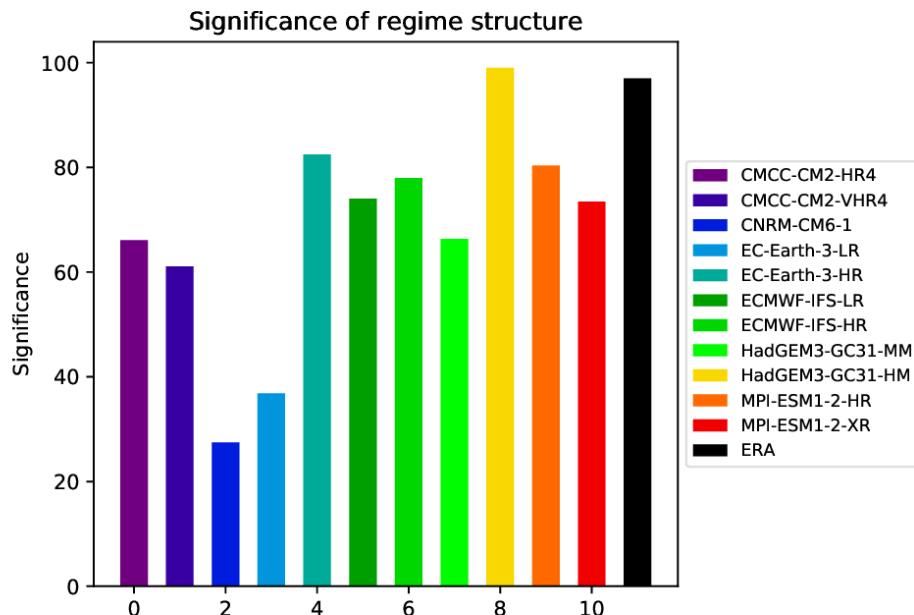


**Figure 3.2.9.3:** Comparison between the observed Scandinavian Blocking pattern and the one simulated by ECMWF-HR.

The second metric quantifies the model performance in reproducing the observed frequency of each WR. This is yet a difficult task for coupled models, which show large differences with respect to the observed frequencies, as shown in Figure 3.2.9.4. Models tend to underestimate the occurrence of the NAO+ regime, while they overestimate the occurrence of NAO- and Atlantic Ridge regimes. For the Sc. Blocking regime some models get very close to the observed frequency, while others tend to underestimate it.



**Figure 3.2.9.4:** Frequencies of occurrence for each regime and each model considered, in comparison with the observed frequencies in the period 1979–2010. The x-axis labels signify the four WRs: "NP" for NAO+, "BL" for Scandinavian Blocking, "AR" for Atlantic Ridge and "NN" for NAO-.



**Figure 3.2.9.5:** Significance of the cluster partition for each model considered with respect to ERA-Interim.

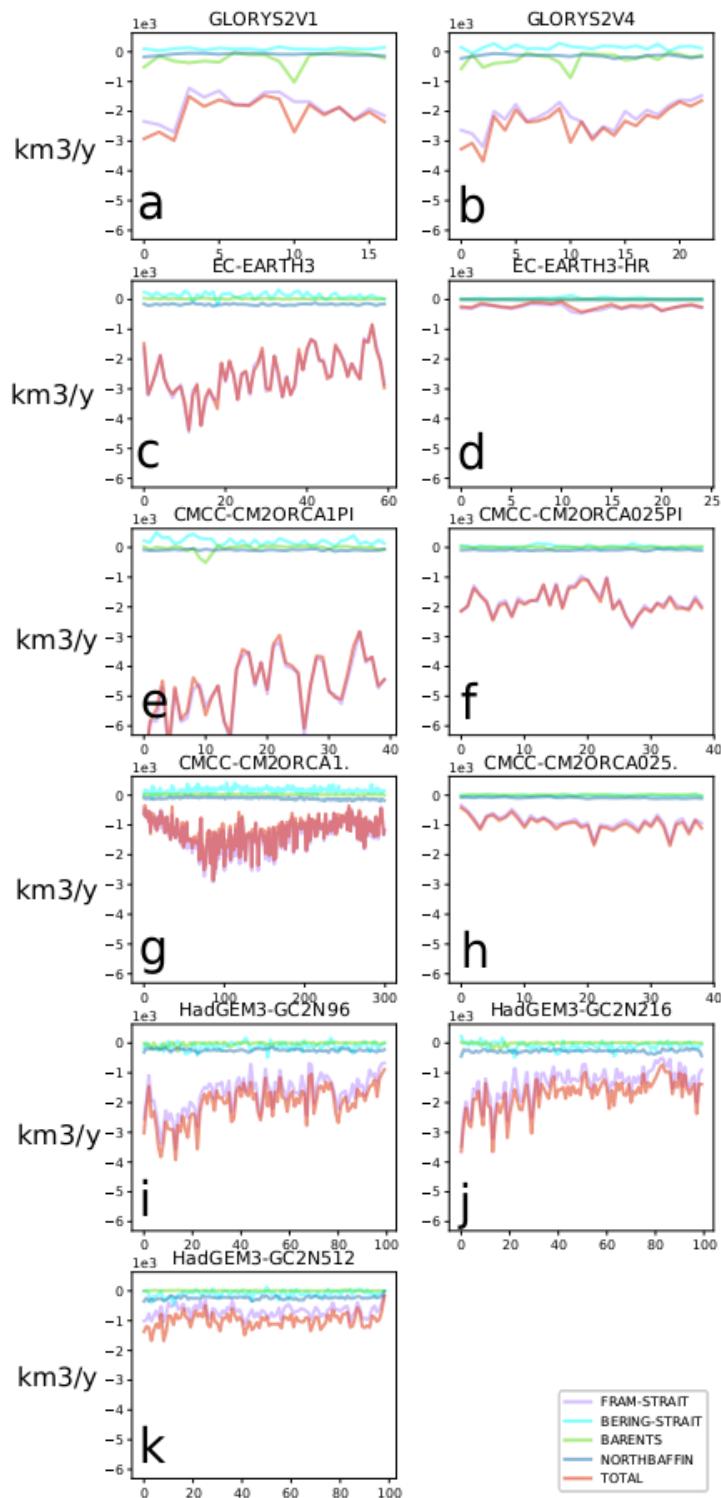
Finally, it is important to assess whether the simulated fields show the same tendency to cluster around the WRs attractors as the observed fields do. This is measured through the calculation of the significance of the cluster partition as described in Straus et al. (2007) and Strommen et al. (2019). The results for the PRIMAVERA coupled simulations are shown in Figure 3.2.9.5: some models reach a significance somehow comparable to that found in ERA-Interim, while many others are far below the observations.

Concluding, the coupled models show some difficulties in correctly reproducing the observed Weather Regimes over the Euro-Atlantic sector during winter. Nevertheless, for some models and depending on the regime considered, the simulated WR patterns show a satisfactory pattern correlation with the observed ones. More systematic differences are seen in the WRs frequencies of occurrence.

### 3.2.10 Solid and liquid freshwater volumes and exports from the Arctic [SMHI]

We have implemented python programs to calculate the Arctic liquid and solid freshwater volumes, and their transports to/from the Arctic, for models participating in the EU H2020 PRIMAVERA project, we have applied this tool to three different coupled global climate models (GCMs): EC-Earth3.1, HadGEM3-GC2 and CMCC-CM2.

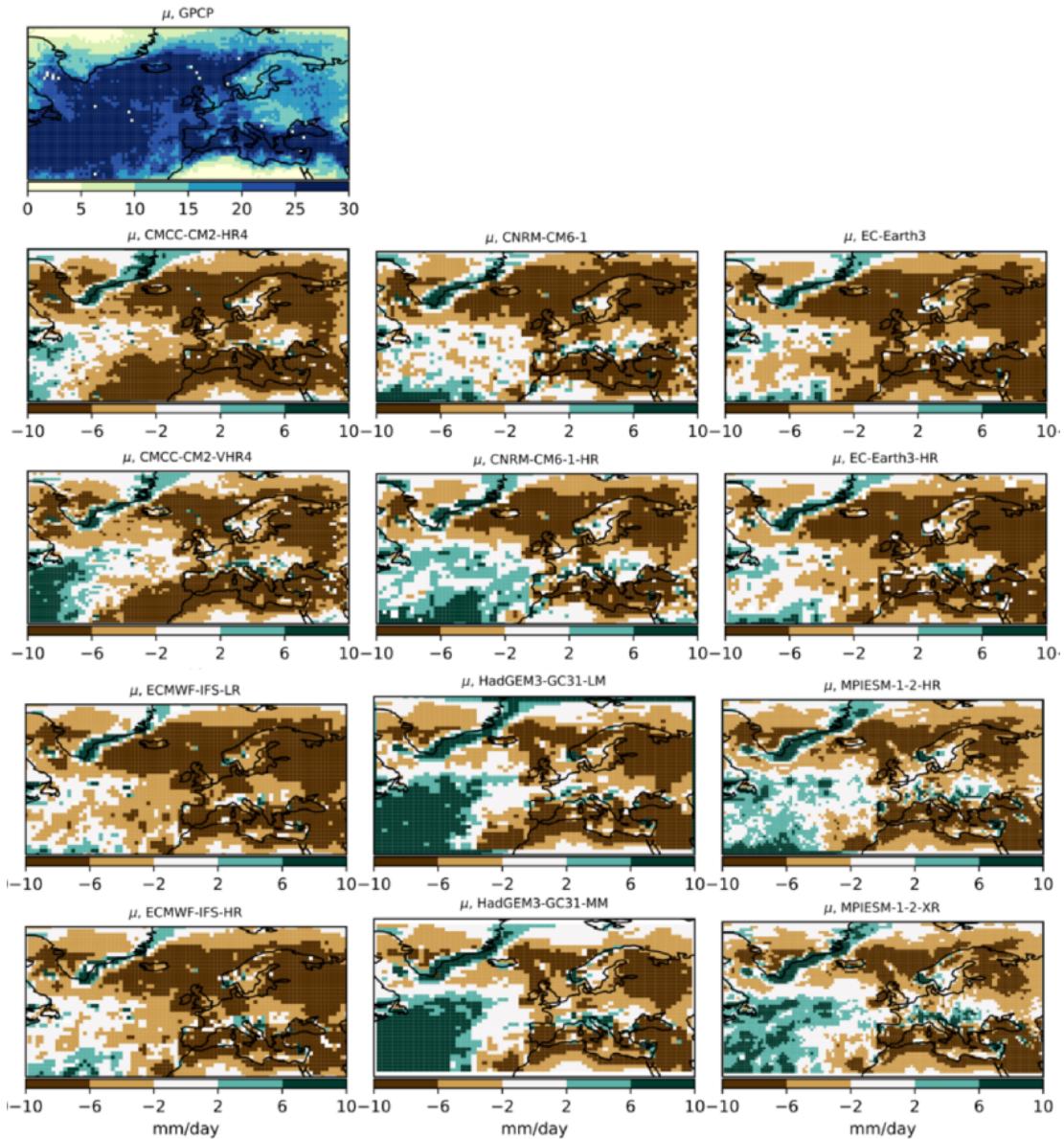
The set of programs calculate the liquid and solid freshwater volume and the transports across different straits that define the limit between the Arctic and the Atlantic or the Pacific oceans. They produce new NetCDF files that can be used after for further analysis. An example of solid freshwater exports output can be found in the figure below (Figure 3.2.10.1). This set of programs will be applied to all Stream-1 simulations.



**Figure 3.2.10.1:** Ice transport towards the Arctic across different straits in a) GLORYS2V1 reanalysis, b) GLORYS2V4, c) EC-Earth, d) EC-Earth-HR, e) CMCC-CM2 ORCA1 PI, f) CMCC-CM2 ORCA025 PI, g) CMCC-CM2 ORCA1 PD, h) CMCC-CM2 ORCA025 PD, i) HadGEM3-GC2 N96, j) HadGEM3-GC2 N216 and k) HadGEM3-GM2 N512. The units are  $\text{km}^3 \text{ year}^{-1}$ . Negative numbers mean ice export from the Arctic. Total values (sum of ice transport across all straits) are shown in red. GLORYS2V1 includes 1993–2009 period and GLORYS2V4 includes the period 1993–2015.

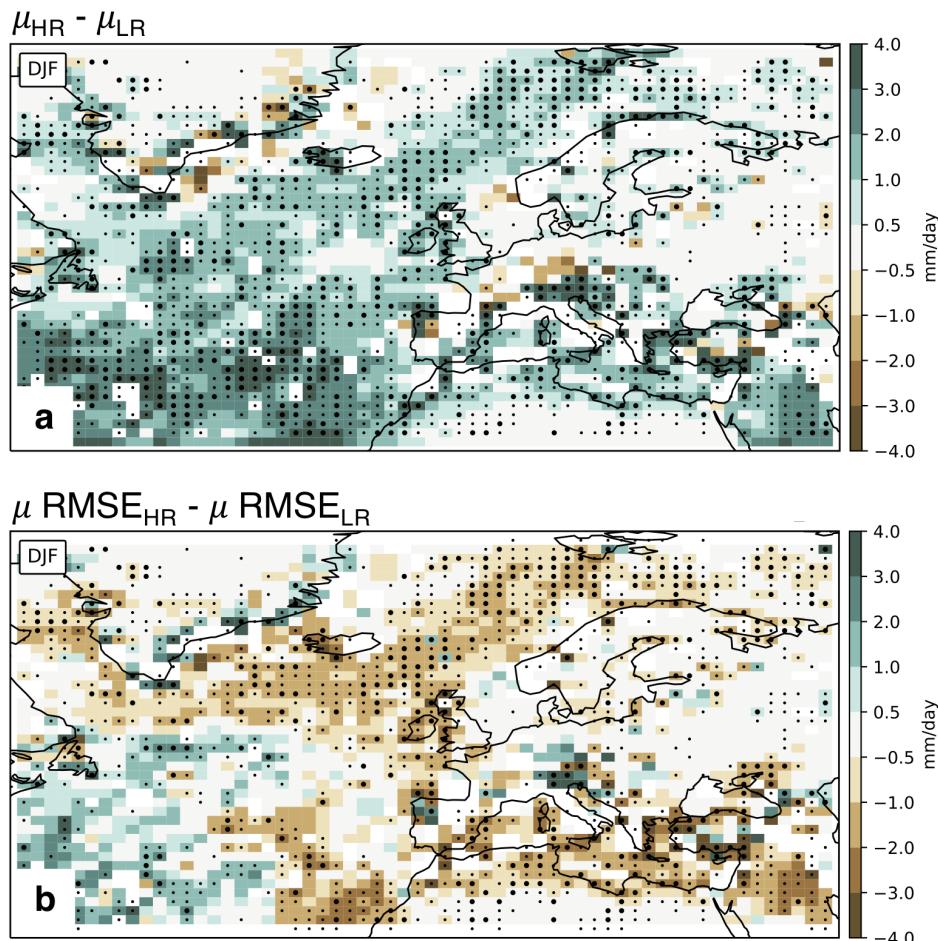
### 3.2.11 The representation of extreme European precipitation [UREAD]

Introduction: We evaluated the impact of increased horizontal atmospheric resolution on extreme daily precipitation across the Stream-1 ensemble of atmosphere-and-land-only and fully coupled simulations. Here, we show the added value of increased atmospheric resolution for winter (DJF) precipitation over Europe and the North Atlantic.



**Figure 3.2.11.1:** Wintertime mean bias (model-observation) in  $\mu$  for lowest (2<sup>nd</sup> and 4<sup>th</sup> rows) and highest (3<sup>rd</sup> and 5<sup>th</sup> rows) available resolution. Biases are computed versus GPCP daily, gridded ( $1^\circ$ ) precipitation data, available for 1996–2013 (Huffman et al., 2001).

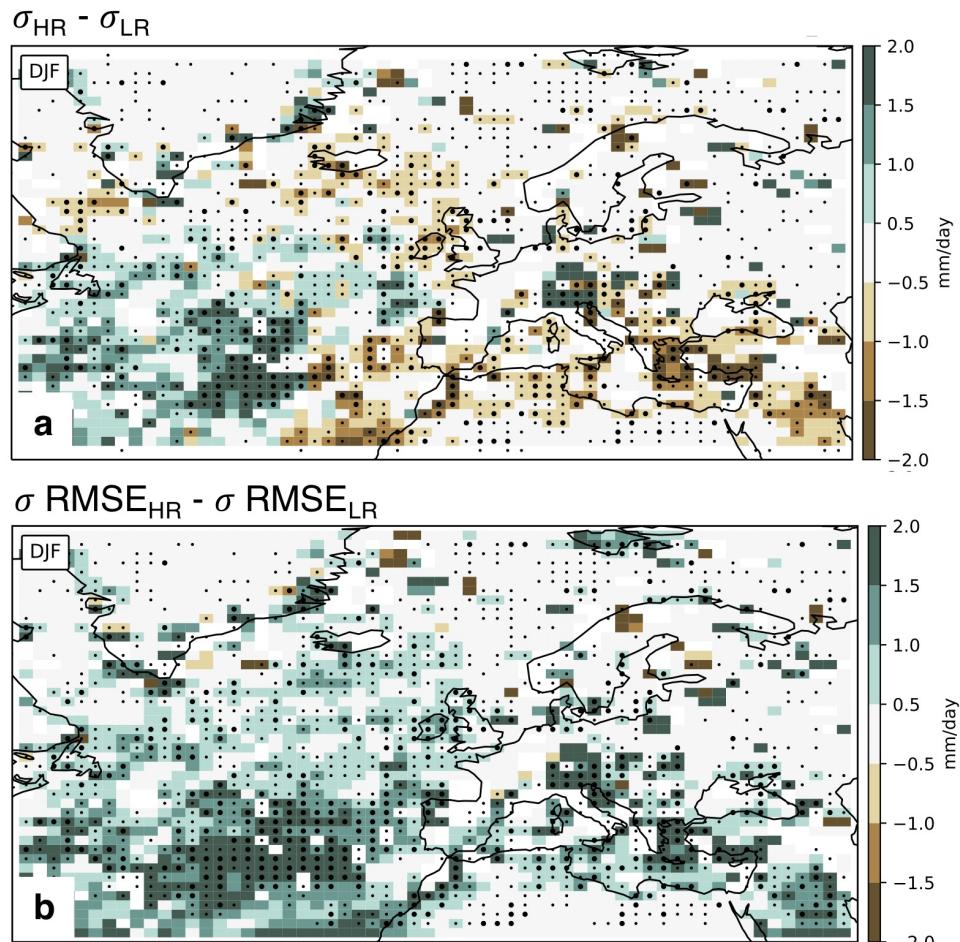
Methodology: We employed generalised extreme value (GEV) analysis and applied the parametric block maxima method globally. At each model grid point, globally, 1-day precipitation maxima were computed for each canonical season. GEV distributions were fitted to these seasonal precipitation block maxima time series, described by the location ( $\mu$ ), scale ( $\sigma$ ) and shape ( $\xi$ ) parameters, which determine the change in return value as a function of return period. Here, we focus on two quantities:  $\mu$  determines the vertical position of the GEV curve and thereby ‘typical’ return values and  $\sigma$  determines the slope of the GEV curve and thereby the year-to-year variability in extremes. We show results for highresSST-present simulations as delivery of coupled runs from PRIMAVERA partners is ongoing. An example of the application of GEV analysis to global climate model integrations is given in Schiemann et al. (2018).



**Figure 3.2.11.2:** Multi-model wintertime mean difference in (a)  $\mu$  (high-resolution minus low-resolution) and (b)  $\mu$  root-mean-square error (RMSE). Positive (negative) values indicate increased (decreased)  $\mu$  or  $\mu$  RMSE at high-resolution. Large (small) stippling indicates that all six (five out of six) models agree on the sign of  $\mu$  change with resolution increase. RMSE is computed versus GPCP data.

Key results: Models underestimate  $\mu$  over land, the Mediterranean, and north-east North Atlantic (Figure 3.2.11.1). Increasing resolution increases  $\mu$  across the mid-latitudes in all models (Figure 3.2.11.2a). Increased extremes are simulated over much of the North Atlantic, particularly the storm track region in winter (and the equinoctial

seasons – not shown). Simulated  $\mu$  is closer to observational Global Precipitation Climatology Project data (Huffman et al., 2001) over this region (Figure 3.2.11.3b). However, simulated  $\sigma$  is farther from observational estimates, indicating that typical return values are better-simulated in high-resolution forced simulations. Importantly, increased extreme precipitation is coterminous with reduced error over the north-eastern North Atlantic, Mediterranean and European orographic regions, exhibiting the added value of high-resolution integrations across much of the Euro-Atlantic domain of immediate interest to PRIMAVERA partners and stakeholders.



**Figure 3.2.11.3:** As Figure 3.2.11.2 but for  $\sigma$ .

Forthcoming research: (i) Assess observational uncertainty over European land. (ii) Extend analysis to remaining coupled simulations upon delivery. (iii) Link GEV evaluation to analyses of ETC activity and associated precipitation as well as North Atlantic eddy-driven jet variability (CMCC collaboration). (iv) Link evaluation of extremes to post-tropical cyclone analyses (KNMI collaboration).

### 3.2.12 Multi-algorithm analysis of tropical cyclones [UKMO]

Regarding tropical cyclone tracking, the reader is kindly pointed also to Section 3.2.6.

It is well known that different algorithms used to track particular structures in models and reanalyses (such as tropical and extra-tropical cyclones, polar lows, Medicanes etc.) have different strengths and weaknesses. These may relate to both how well the required features are found within these large datasets, as well as the efficiency and the focus of the algorithms themselves. Interpreting the output from such algorithms, and drawing conclusions from it (for example understanding model biases, or the impact of forcing, or storm structure) may not be robust if the algorithm itself has weaknesses.

Hence using more than one tracking algorithm can help to address some of these issues. There are several questions that arise:

1. does each algorithm find the same structures in a given model?
2. do results from different algorithms help us to better understand the models?
3. what are the strengths and weaknesses of the different algorithms?

In this work we use TRACK (Hodges et al., 2017) and TempestExtremes (Ullrich and Zarzycki, 2017; Zarzycki and Ullrich, 2017) for tropical cyclones tracking. The respective codes are available on JASMIN. We use a variety of input datasets, including reanalyses and multi-member ensembles of model simulations at different resolutions.

The basic differences between the algorithms are:

TRACK uses 850 hPa vorticity, filtered onto a common T63 spectral grid (regardless of the native model resolution) to find promising features, and then uses the vertical gradient of vorticity (850 to 250 hPa) to ensure a warm core, together with other filters.

TempestExtremes uses mean sea level pressure as the basic feature variable, and then the difference in geopotential height between 250 and 500 hPa to check for a warm core. It then stitches together individual points if they pass various criteria to make continuous tracks.

Here we focus on ensemble simulations using HadGEM3-GC31 on latitude-longitude grid (85 levels) at different horizontal resolutions:

LM = N96 (130km mid-latitude), 192x144 points, 13 members

MM = N216 (60km), 432 x 324 points, 13 members

HM = N512 (25km), 1024 x 768 points, 5 members

We use years 1979–2014 with the HighResMIP highresSST-present simulations (using HadISST2.2  $\frac{1}{4}$  degree daily SST and sea-ice forcing, MACv2-SP aerosol (not natural aerosol variability; Haarsma et al., 2016). We mainly use ACE (Accumulated Cyclone Energy) as the metric for tropical cyclone activity, as it is more robust than simple storm frequency (Roberts et al., in prep.).

Algorithm cost: Proviso – it is not simple to exactly compare the cost of the algorithms due to preprocessing (and currently re-using TRACK inputs to set up TempestExtremes inputs). In addition there are known optimisations of TRACK in later versions of the code which is not used here. However, of the main different processing costs for a 25km global lat-lon grid, for 1 year with 6 hourly data:

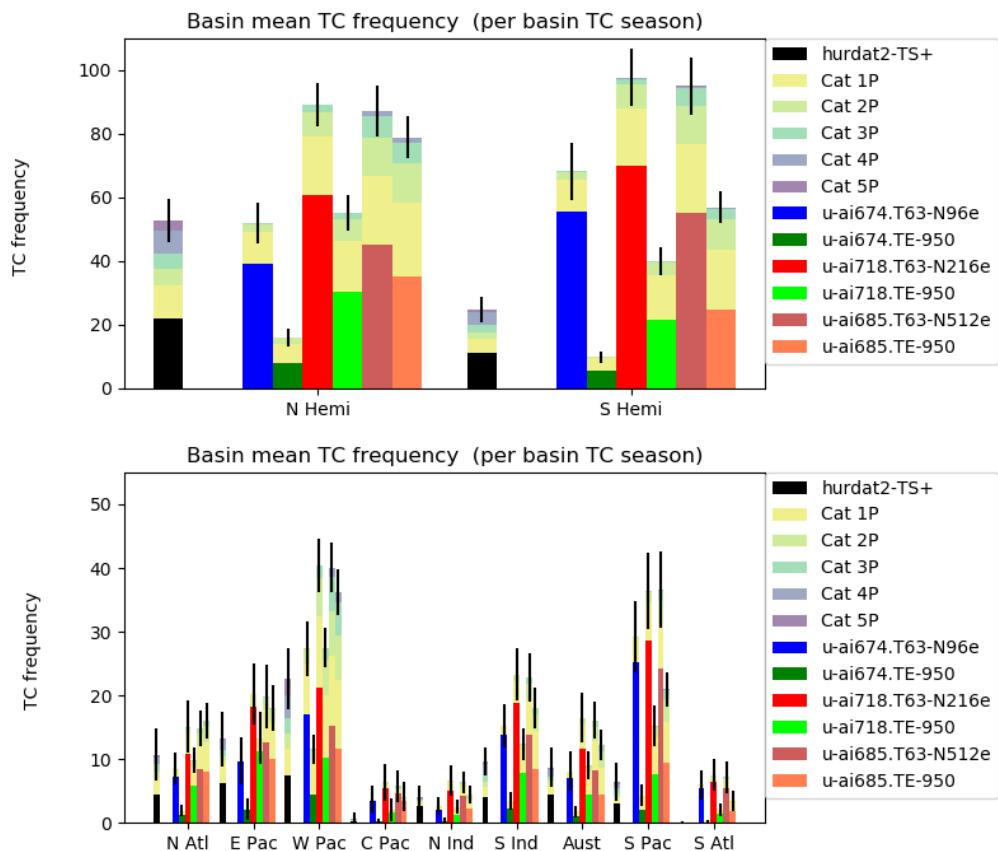
#### TRACK:

Spatial filtering 5 pressure levels to T63 (run concurrently): >=4 hours

Tracking: 1 hour per hemisphere, serially = 2 hours

Total: ~6 hours per model year

TempestExtremes: Identify + stitch for both hemispheres: 8 minutes.



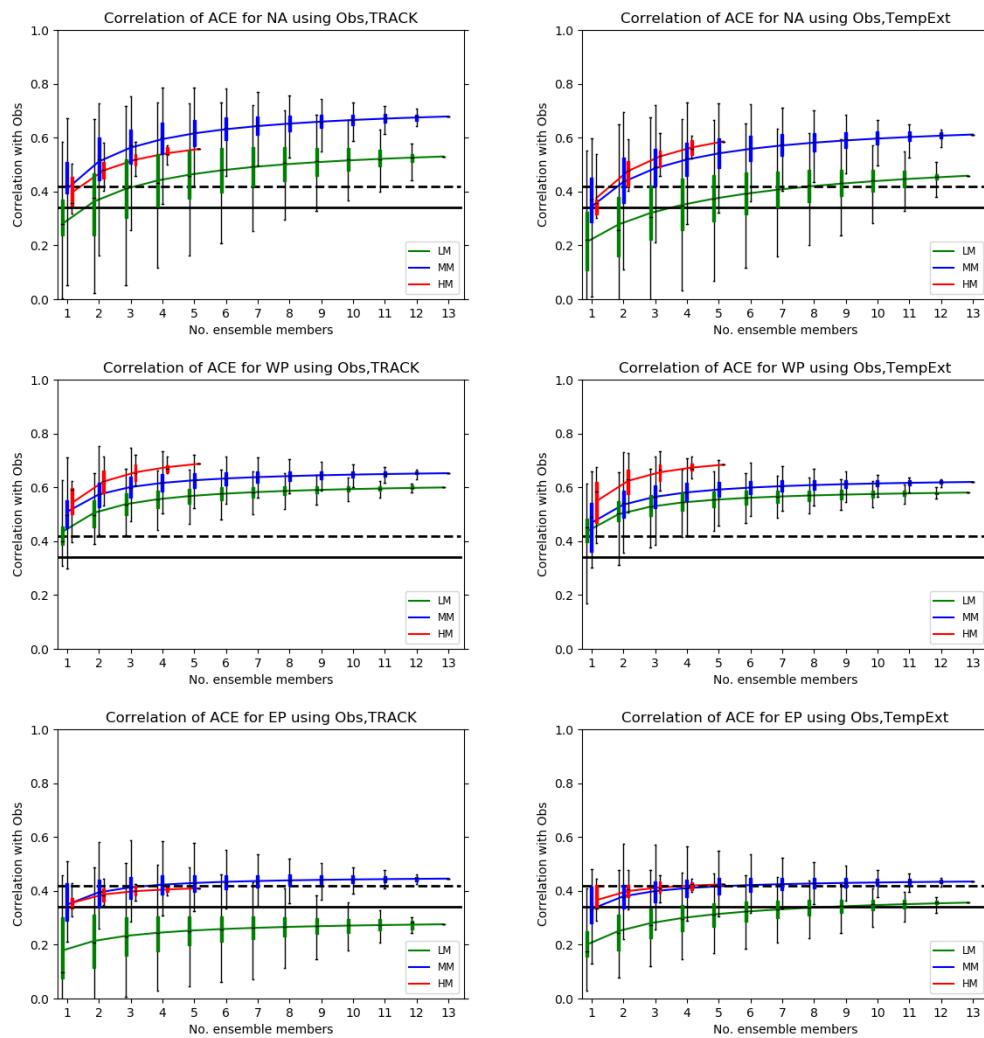
**Figure 3.2.12.1:** Mean tropical cyclone activity (frequency) in (top) Northern and Southern Hemispheres and (bottom) each ocean basin over the period 1979–2014. The colours in each bar show the proportion of the storms related to particular storm strengths. The initial colour indicates the model and corresponds to the weakest storms. In the legend, T63 indicates tracking with TRACK, while TE indicates tracking performed with TempestExtremes. The Cat xP shadings indicate how strong the tropical cyclone was (in terms of minimum sea level pressure), from weak (Cat 1P) to strong (Cat 5P). Models: u-ai674 = LM (130km); u-ai718 = MM (60km); u-ai685 = HM (25km). Observations are in black.

**Results:** We track all the ensemble members of each model resolution globally and assess any differences in the results between the different algorithms.

Figure 3.2.12.1 shows the mean number of tropical cyclones in each basin, at the three different model resolutions and for each algorithm. There are several differences to note:

1. The lower resolution is more sensitive to the algorithm in terms of storm counts, as expected;
2. The number of storms in the Southern Hemisphere is considerably reduced using TempestExtremes.

We then use the ensemble members to generate ensemble mean values at each year, and compare the ensemble mean interannual variability with that observed. This should act to average out the weather variability in each year (from the models), and allow us to assess whether the skill of the models (as assessed by the correlation of interannual variability with observations) is different between the two algorithms.



**Figure 3.2.12.2:** Number of ensemble members vs ACE interannual variability correlation for (top) North Atlantic and (middle) NW Pacific and (bottom) Eastern Pacific. The left column uses TRACK and the right column uses TempExtExt algorithms.

Figure 3.2.12.2 shows the correlation of interannual ACE variability with observations for a given ensemble size, for the North Atlantic (NA), North West Pacific (WP) and East Pacific (EP) for the two algorithms (left column: TRACK, right column: TempestExtremes). We see that for each basin, for the higher resolutions, the asymptote for the correlation looks similar for the two algorithms, but is generally lower in TempestExtremes for the lower resolutions. For the WP (and possibly the NA), there is a hint that the 25km model has higher skill, even with fewer ensemble members than the other resolutions.

Summary: We have implemented two different tracking algorithms to use for all the HighResMIP simulations, in order to have more confidence in our conclusions for tropical cyclone analysis. This has illustrated that an apparent bias in all models for too many tropical cyclones in the Southern Hemisphere may partly be due to the TRACK algorithm, as with TempestExtremes there are far fewer storms found here.

We have also shown that a reasonable number of ensemble members (perhaps five or more) is needed for a robust assessment of model interannual variability skill, and with enough members it is clear that the higher resolution models have more skill.

Further work: In the future we can attempt to exactly match the tracks from the different algorithms and hence attempt a deeper understanding of why one method finds particular tracks but not another.

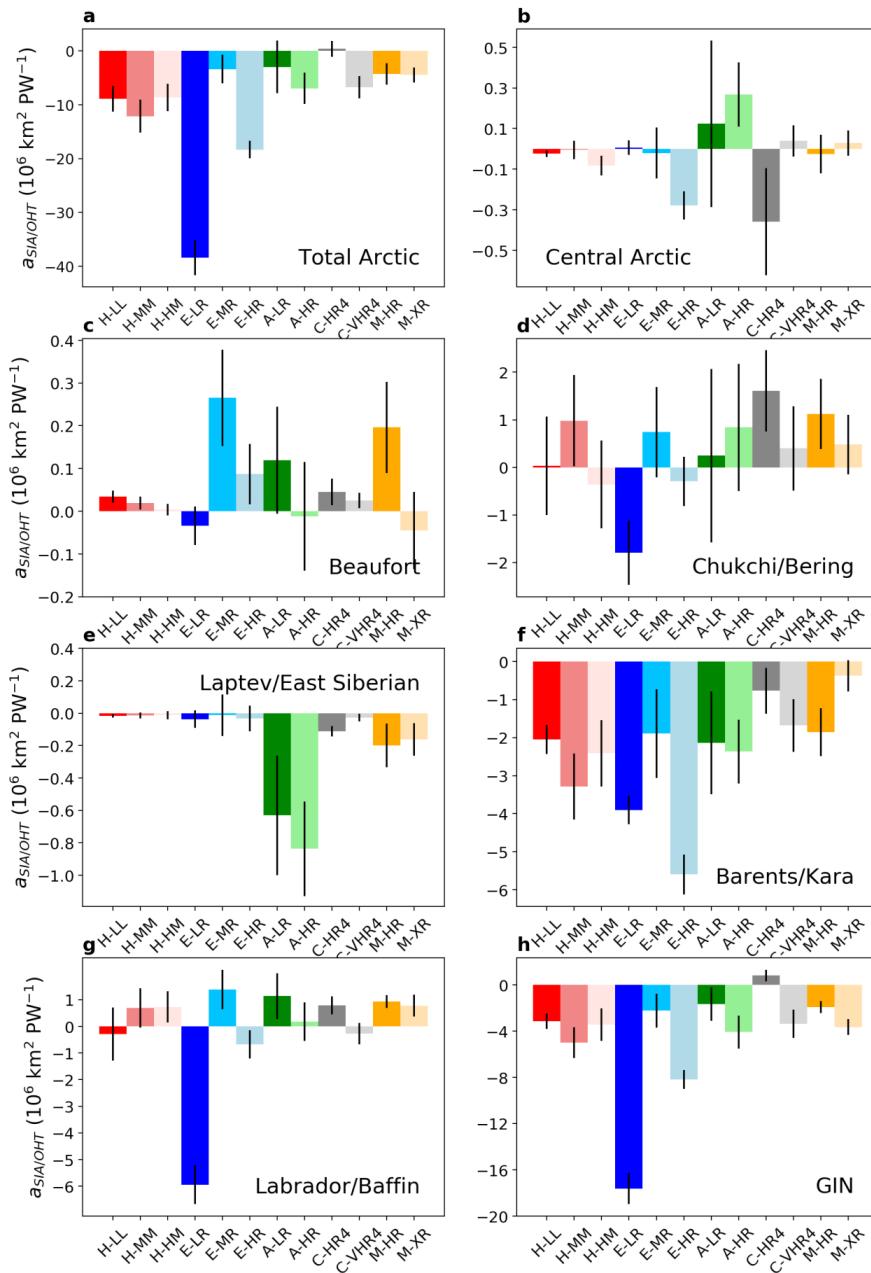
An overview of the results from the different methods may also have potential implications for understanding variability and climate change, in particular if the coupled model simulations produce significantly different results with different algorithms.

### 3.2.13 Sea-ice - ocean heat transport (OHT) metric [UCLouvain]

The recent observed sea-ice reduction in the Barents Sea occurred concurrently to an increase in Atlantic OHT due to both strengthening and warming of the oceanic inflow (Arthun et al., 2012). Other regions of the Arctic Ocean have also recently experienced an increased influence from the Atlantic Water on sea-ice (Polyakov et al., 2017). We have developed the sea-ice – ocean heat transport (sea-ice – OHT) metric, which assesses the relationship between Arctic sea-ice area/volume and poleward Atlantic OHT. We compute the regression slope between detrended monthly mean sea-ice area and detrended annual mean OHT for every year of a given period. Thus, this metric gives the amount of sea-ice loss (or gain) per PW of OHT flowing to the Arctic. Sea-ice area and volume for the whole Arctic Ocean and 7 specific Arctic seas (following Koenigk et al., 2016) are computed based on sea-ice concentration, equivalent sea-ice thickness and grid-cell area. Annual mean OHT is computed at 50N, 60N and 70N over the whole latitudinal band crossing the Atlantic Ocean. A detailed description of this metric is provided in Docquier et al. (in review) using outputs from 12 PRIMAVERA Stream-1 coupled hist-1950 model configurations.

The code of this metric is written in Python and is available upon request. It is planned in the near future to clean the code and make it available to PRIMAVERA users through JASMIN and PRIMAVERA svn repository. No observational reference is available at the moment due

to the lack of OHT observations for a whole Atlantic latitudinal band. However, an ongoing PRIMAVERA study focuses on the sea-ice – OHT relationship in specific Arctic seas where OHT observations are available (e.g. Barents and Bering Seas) (Docquier et al., in prep.).



**Figure 3.2.13.1:** Regression slopes between detrended March Arctic sea-ice area and detrended annual mean Atlantic OHT at 70N (computed over 1950–2014) for (a) the whole Arctic and (b–h) 7 specific Arctic regions (GIN stands for Greenland-Iceland-Norwegian). The X axis shows the 12 model configurations used, with the first letter indicating the model (H: HadGEM3; E: ECMWF-IFS; A: AWI-CM; C: CMCC-CM2; M: MPI-ESM). The black line on top of each bar indicates the standard deviation of these slopes. Credit: Figure 14 of Docquier et al. (in review).

The sea-ice - OHT metric has been applied to PRIMAVERA Stream-1 coupled hist-1950 (and control-1950) simulations, in particular 12 different model configurations, i.e.

HadGEM3-LL, HadGEM3-MM, HadGEM3-HM, ECMWF-LR, ECMWF-MR, ECMWF-HR, AWI-LR, AWI-HR, CMCC-HR4, CMCC-VHR4, MPI-HR, MPI-XR. Figure 3.2.13.1 below shows the application of this metric over all model configurations and the whole Arctic Ocean, as well as the 7 specific Arctic seas. In particular, Figure 3.2.13.1 shows that the reduced sea-ice area in March is clearly correlated to the increased OHT at 70N for the total Arctic Ocean (Figure 3.2.13.1.a; except CMCC-HR4 and AWI-LR), Barents/Kara Seas (Figure 3.2.13.1.f; except MPI-XR) and GIN Seas (Figure 3.2.13.1.h; except CMCC-HR4). The higher the latitude to compute OHT, the stronger the anticorrelation between Arctic sea-ice area and OHT. The specific Arctic regions that are more directly influenced by Atlantic OHT are in the Atlantic sector of the Arctic Ocean, i.e. Barents/Kara Seas and GIN Seas, which first receive the warm Atlantic water inflow. As HighResMIP is still ongoing at the time of writing, more model outputs and analyses will be provided, which will give insights into our understanding of the impact of resolution.

### 3.2.14 Local coupled feedbacks over Europe in summer [CERFACS]

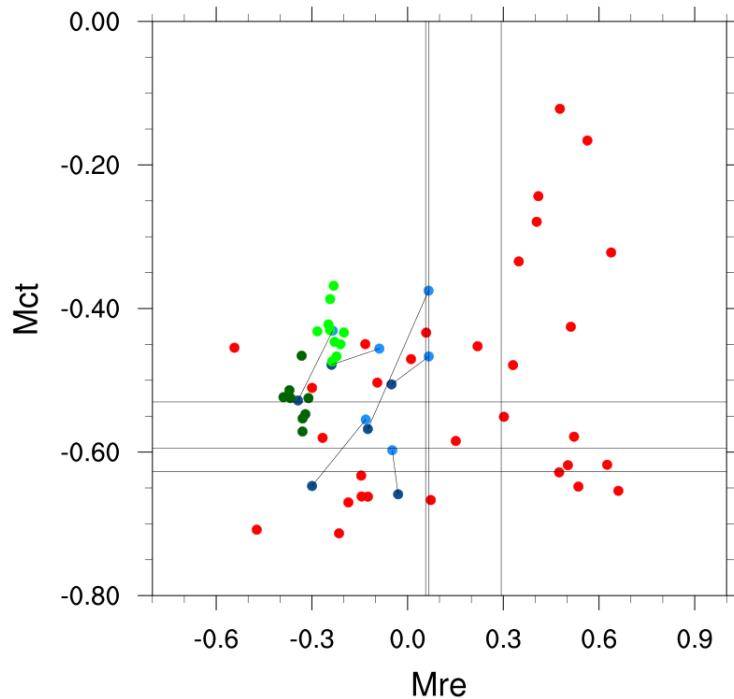
Different metrics of local coupled feedbacks have been computed. Two of them are analysed here:

- (i) The present-day interannual correlation in summer between evapotranspiration and the sum of incoming solar and longwave radiation at surface (Mre).
- (ii) The present-day interannual correlation in summer between surface temperature and cloud cover (Mct).

These metrics are calculated here on the 1979–2004 period and for land points only.

The first metric, Mre, aims to characterize the local coupling between the soil and the atmosphere through evapotranspiration. Where the evapotranspiration is predominantly limited by the energy available at surface (e.g. over Scandinavia), the correlation is positive. Where the evapotranspiration is predominantly limited by soil moisture (e.g. over southern Europe) the correlation is negative: namely, less energy at surface (more clouds and more precipitation) is associated with a stronger evapotranspiration because the latter, in this case, is controlled by soil moisture rather than the available radiative energy. Boé and Terray (2008, 2014) have shown that previous climate models often disagree on the present-day predominant control of evapotranspiration over an intermediate region of Europe, between Scandinavia and southern Europe, with important potential impacts on future climate changes, through the modulation of evapotranspiration changes.

The second metric, Mct, aims to assess the local cloud feedback. During summer over most of Europe, a smaller cloud cover is generally associated with higher surface temperature, but large uncertainties exist regarding the magnitude of this effect, especially over western Europe, with a potential impact on simulated future temperature changes (Boé and Terray, 2014).



**Figure 3.2.14.1:** Values of the  $M_{re}$  and  $M_{ct}$  metrics (no units) in summer averaged over western Europe ( $40^{\circ}\text{N}$ — $60^{\circ}\text{N}$ ,  $-10^{\circ}\text{E}$ — $15^{\circ}\text{E}$ ) on the 1979–2004 period. LR (HR) PRIMAVERA forced atmospheric simulations (highresSST-present experiment) are shown with a light (dark) blue dot. A line connects the value of HR and LR simulations for a given model. For CNRM-CM6, the ensemble mean is shown with the same colour code and the individual members (8 HR and 10 LR members) are shown with light (for LR members) and dark green (for HR members) dots. The metrics for a large ensemble of CMIP5 models (historical simulations) are also shown with red dots. Three observational estimates are shown for each metric. Three land surface analyses are used for  $M_{re}$ : ERAland (Balsamo et al., 2015), MERRALand (Reichle et al., 2011), and GLDAS (Rodell et al., 2014). Regarding  $M_{ct}$ , one estimate is based on cloud cover and temperature from the ERA-Interim atmospheric reanalysis (Dee et al., 2011), another is based on CRU TS (Harris et al., 2014) and a third one is based on the surface temperature dataset HadCRUT4 (Morice et al., 2012) and the clouds satellite dataset ISCCP (Rossow et al., 1999).

The inter-model spread for the two metrics is much smaller for the PRIMAVERA models than for the CMIP5 coupled models, very likely because observed sea surface temperatures are imposed as forcing in the highresSST-present PRIMAVERA simulations shown here (Figure 3.2.14.1). The observational uncertainties are quite large for both metrics. The impact of internal variability as estimated with CNRM-CM6 is large for  $M_{ct}$ , even in the forced atmospheric framework used here. Given the impact of internal variability and the large observational uncertainties, it is not possible to reject the consistency of PRIMAVERA HR simulations with the observations regarding  $M_{ct}$  for all the models. Some LR simulations are likely not compatible with the observational estimates, but the lack of members, except for CNRM-CM6, does not allow to reach a strong conclusion. Most PRIMAVERA simulations are likely not consistent with observations for  $M_{re}$ , with generally a stronger limitation of evapotranspiration by soil moisture.

Increasing model resolution may have a measurable impact using these metrics. This is visible for CNRM-CM6 for both metrics. For the other models, given the impact of internal variability, it is difficult to draw a conclusion based on the single member used here. It is still interesting to note that in all the models,  $M_{ct}$  is more negative in the higher resolution simulations. This is also true for  $M_{re}$  in all the PRIMAVERA models except one. Note that more negative present-day  $M_{re}$  and  $M_{ct}$  tend to be associated with larger future summer warming over Europe (Boé and Terray, 2014). It will be interesting to assess whether this holds true for the PRIMAVERA future climate projections.

The previous analysis shows the importance of taking into account the impact of internal variability and the observational uncertainties for the evaluation of the models, at least regarding the metrics used here, based on variables not necessarily well-observed, and, in some cases not observed at all. Oceanic biases are likely responsible for the too positive values of  $M_{re}$  and  $M_{ct}$  seen in many CMIP5 models but not in the PRIMAVERA highresSST-present simulations. Finally, the model resolution likely impacts the metrics analysed here. It will be interesting to try to confirm whether this result is robust through the analysis of the involved mechanisms.

## 4. Lessons Learned

The use of multiple metrics, and multiple algorithms calculating similar metrics, is important to gain different insights into the strengths and weaknesses of model simulations. Algorithms also have their own weaknesses, and hence it can be dangerous to rely on just one when making conclusions about model fidelity.

## 5. Links Built

Regarding the implementation of metrics into the ESMValTool, priority will be given to those metrics that are needed for evaluations related to WP10. Coordination with WP10 will be fostered in the forthcoming General Assembly (GA4).

## 6. References

- Athanasiadis, P. J., Bellucci, A., Hermanson, L., Scaife, A. A., MacLachlan, C., Arribas, A., et al. (2014). The representation of atmospheric blocking and the associated low-frequency variability in two seasonal prediction systems. *Journal of Climate*, 27(24), 9082–9100. <https://doi.org/10.1175/JCLI-D-14-00291.1>
- Arthun, M., T. Eldevik, L. H. Smedsrud, O. Skagseth, R. B. Ingvaldsen (2012). Quantifying the influence of Atlantic heat on Barents Sea ice variability and retreat. *Journal of Climate*, doi: 10.1175/JCLI-D-11-00466.1.
- Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., ... Vitart, F. (2015). ERA-Interim/Land: A global land surface reanalysis data set. *Hydrology and Earth System Sciences*. <https://doi.org/10.5194/hess-19-389-2015>.
- Bishop, S.P., R.J. Small, F.O. Bryan, and R.A. Tomas, 2017: Scale Dependence of Midlatitude Air–Sea Interaction. *J. Climate*, 30, 8207–8221, <https://doi.org/10.1175/JCLI-D-17-0159.1>.
- Bitz, C. M., G. H. Roe (2004). A mechanism for the high rate of sea ice thinning in the Arctic Ocean. *Journal of Climate*, doi: 10.1175/1520-0442(2004)017<3623:amfthr>2.0.co;2.
- Boé, J., & Terray, L. (2008). Uncertainties in summer evapotranspiration changes over Europe and implications for regional climate change. *Geophysical Research Letters*, 35(5), L05702. <https://doi.org/10.1029/2007GL032417>.
- Boé, J., & Terray, L. (2014). Land-sea contrast, soil-atmosphere and cloud-temperature interactions: Interplays and roles in future summer European climate change. *Climate Dynamics*, 42(3–4), 683–699. <https://doi.org/10.1007/s00382-013-1868-8>.
- Buehler, Raible & Stocker (2011): The relationship of winter season North Atlantic blocking frequencies to extreme cold or dry spells in the ERA-40. *Tellus*, doi:10.1111/j.1600-0870.2010.00492.x.
- Cattiaux, J., B. Quesada, A. Arakelian, F. Codron, R. Vautard, and P. Yiou (2013), North-Atlantic dynamics and European temperature extremes in the IPSL model: Sensitivity to atmospheric resolution, *Clim. Dyn.*, 40, 2293–2310.
- Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally. 1996, updated yearly. Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center.
- Corti S., Palmer T., Molteni F. (1999). Signature of climate change in frequencies of natural atmospheric circulation regimes. *Nature*, doi: 10.1038/19745.
- Davini, P., Corti, S., D'Andrea, F., Rivière, G., & von Hardenberg, J. (2017). Improved Winter European Atmospheric Blocking Frequencies in High-Resolution Global Climate Simulations. *Journal of Advances in Modeling Earth Systems*, 9(7), 2615–2634. <https://doi.org/10.1002/2017MS001082>.

Dawson A., Palmer T., Corti S. (2012). Simulating regime structures in weather and climate prediction models. *Geophysical Research Letters*, doi:10.1029/2012GL053284.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*. <https://doi.org/10.1002/qj.828>

Docquier, D., F. Massonnet, A. Barthélemy, N. F. Tandon, O. Lecomte, T. Fichefet (2017). Relationships between Arctic sea ice drift and strength modelled by NEMO-LIM3.6, *The Cryosphere*, doi: 10.5194/tc-11-2829-2017.

Docquier, D., J. P. Grist, M. J. Roberts, C. D. Roberts, T. Semmler, L. Ponsoni, F. Massonnet, D. Sidorenko, D. Sein, D. Iovino, A. Bellucci, T. Fichefet (in review). Impact of model resolution on Arctic sea ice and North Atlantic Ocean heat transport.

EUMETSAT SAF on Ocean and Sea Ice (2016). Global Sea Ice Concentration climate data record release 1.1 (period 1978-2009) - DMSP. OSI SAF. DOI: 10.15770/EUM\_SAF\_OSI\_0001.

Franzke C., Woolings T., Martius O. (2011). Persistent circulation regimes and preferred regime transitions in the North Atlantic. *Journal of the Atmospheric Sciences*, doi:10.1175/JAS-D-11-046.1.

Fučkar, N.S., V. Guemas, N.C. Johnson, F. Massonnet and F.J. Doblas-Reyes (2016). Clusters of interannual sea ice variability in the northern hemisphere. *Climate Dynamics*, 47, 1527-1543, doi:10.1007/s00382-015-2917-2.

Hannachi A. et al. (2013). Behaviour of the winter North Atlantic eddy-driven jet stream in the CMIP3 integrations. *Clim. Dyn.* 41: 995–1007.

Hannachi A., Straus D., Franzke C., Corti S., Woollings T. (2017). Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere, *Rev. Geophys.*, doi:10.1002/2015RG000509.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*. <https://doi.org/10.1002/joc.3711>.

Hodges, K., Cobb, A., & Vidale, P. L. (2017). How well are tropical cyclones represented in reanalysis datasets? *Journal of Climate*, 30(14), 5243–5264. <https://doi.org/10.1175/JCLI-D-16-0557.1>.

Huffman, G. J. et al., 2001: Global Precipitation at One-Degree Daily Resolution from Multisatellite Observations. *J. Hydrometeorol.* 2, 36-50.

Iqbal W. et al. (2018). Analysis of the variability of the North Atlantic eddy-driven jet stream in CMIP5. *Clim. Dyn.* 51: 235.

Koenigk T., M. Caian, G. Nikulin, S. Schimanke S (2016). Regional Arctic sea ice variations as predictor for winter climate conditions. *Climate Dynamics*, doi: 10.1007/s00382-015-2586-1.

Kwon, YO. et al. (2018). North Atlantic winter eddy-driven jet and atmospheric blocking variability in the Community Earth System Model version 1 Large Ensemble simulations. *Clim. Dyn.* 51: 3275.

Massonnet, F., M. Vancoppenolle, H. Goosse, D. Docquier, T. Fichefet, E. Blanchard-Wrigglesworth (2018). Arctic sea-ice change tied to its mean state through thermodynamic processes. *Nature Climate Change*, doi: 10.1038/s41558-018-0204-z.

Matsueda M., Palmer T. (2018). Estimates of the flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3265.

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research Atmospheres*. <https://doi.org/10.1029/2011JD017187>.

Notz, D., C. M. Bitz. Sea ice in Earth system models. In: *Sea Ice* (ed. Thomas, D. N.) (John Wiley & Sons, Chichester, 2017).

Olason, E., D. Notz (2014). Drivers of variability in Arctic sea ice drift speed. *Journal of Geophysical Research*, doi: 10.1002/2014JC009897.

Polyakov I. V., A. V. Pnyushkov, M. B. Alkire, I. M. Ashik, T. M. Baumann, E. C. Carmack, I. Gosczko, J. Guthrie, V. V. Ivanov, T. Kanzow, R. Krishfield, R. Kwok, A. Sundfjord, J. Morison, R. Rember, A. Yulin A (2017). Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. *Science*, doi: 10.1126/science.aai8204.

Putrasahan, DA, Miller AJ, Seo H. 2013. Isolating mesoscale coupled ocean-atmosphere interactions in the Kuroshio Extension region. *Dynamics of Atmospheres and Oceans*. 63:60-78. [10.1016/j.dynatmoce.2013.04.001](https://doi.org/10.1016/j.dynatmoce.2013.04.001).

Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., & Toure, A. (2011). Assessment and enhancement of MERRA land surface hydrology estimates. *Journal of Climate*. <https://doi.org/10.1175/JCLI-D-10-05033.1>.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., ... Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*. <https://doi.org/10.1175/BAMS-85-3-381>.

Rossow, W. B., & Schiffer, R. A. (1999). Advances in Understanding Clouds from ISCCP. *Bulletin of the American Meteorological Society*. [https://doi.org/10.1175/1520-0477\(1999\)080<2261:AIUCFI>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2)

Scaife, A. a., Woollings, T., Knight, J., Martin, G., & Hinton, T. (2010). Atmospheric Blocking and Mean Biases in Climate Models. *Journal of Climate*, 23(23), 6143–6152. <https://doi.org/10.1175/2010JCLI3728.1>

Scherrer, S. C., Croci-Maspoli, M., Schwierz, C., & Appenzeller, C. (2006). Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *International Journal of Climatology*, 26(2), 233–249. <https://doi.org/10.1002/joc.1250>

Schiemann, R., Demory, M.-E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., et al. (2017). The Resolution Sensitivity of Northern Hemisphere Blocking in Four 25-km Atmospheric Global Circulation Models. *Journal of Climate*, 30(1), 337–358. <https://doi.org/10.1175/JCLI-D-16-0100.1>

Schiemann, R. et al., 2018: Mean and extreme precipitation over European river basins better simulated in a 25km AGCM. *Hydrol. Earth Syst. Sci.* **22**, 3933-3950.

Screen, J. A., I. Simmonds (2010). The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, doi: 10.1038/nature09051.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, Vol. 26, Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1986.

Straus D. M., Corti S., Molteni F. (2007). Circulation regimes: chaotic variability versus SST-forced predictability. *Journal of Climate*, doi: 10.1775/JCLI4070.1

Strommen K., I. Mavilia, S. Corti, M. Matsueda, P. Davini, J. von Hardenberg, P-L. Vidale, R. Mizuta (2019). The Sensitivity of Euro-Atlantic Regimes to Model Horizontal Resolution. Submitted to GRL.

Taylor K. (2001). Summarizing multiple aspects of model performance in a single diagram. *JGR*, 106, D7 7183-7192.

Titchner, H. A., & Rayner, N. A. (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, 119(6), 2864-2889.

Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, 10(3), 1069–1090. <https://doi.org/10.5194/gmd-10-1069-2017>.

Woollings, T., Barriopedro, D., Methven, J., Son, S.-W., Martius, O., Harvey, B., et al. (2018). Blocking and its Response to Climate Change. *Current Climate Change Reports*, 287–300. <https://doi.org/10.1007/s40641-018-0108-z>.