Original

Behr, L.; Luther, N.; Josey, S.A.; Luterbacher, J.; Wagner, S.; Xoplaki, E.:

On the Representation of Mediterranean Overflow Waters in Global Climate Models.

In: Journal of Physical Oceanography. Vol. 52 (2022) 7, 1397 - 1413.

First published online by AMS: 15.06.2022

https://dx.doi.org/10.1175/JPO-D-21-0082.1

On the Representation of Mediterranean Overflow Waters in Global Climate Models

LORINE BEHR,[a] NIKLAS LUTHER,[a] SIMON A. JOSEY,[b] JÜRG LUTERBACHER,[a,c,d] SEBASTIAN WAGNER,[e] AND ELENA XOPLAKI[a,c]

[a] *Centre for International Development and Environmental Research, Justus Liebig University, Giessen, Germany*
[b] *National Oceanography Centre, Southampton, United Kingdom*
[c] *Climatology, Climate Dynamics and Climate Change, Department of Geography, Justus Liebig University, Giessen, Germany*
[d] *Science and Innovation Department, World Meteorological Organization, Geneva, Switzerland*
[e] *Institute of Coastal Systems—Analysis and Modeling, Helmholtz-Zentrum Hereon, Geesthacht, Germany*

(Manuscript received 20 March 2021, in final form 1 April 2022)

ABSTRACT: Accurate representation of the Atlantic–Mediterranean exchange in climate models is important for a reliable simulation of the circulation in the North Atlantic Ocean. We evaluate the performance of 10 global climate models in representing Mediterranean Overflow Water (MOW) over the recent period 1986–2005 by using various performance metrics. The metrics are based on the representation of the climatological mean state and the spatiotemporal variability of temperature, salinity, and volume transports. On the basis of analyses and observations, we perform a model ranking by calculating absolute, relative, and total relative errors E_j over each performance metric and model. The majority of models simulate at least six metrics well. The equilibrium depth of the MOW, the mean Atlantic–Mediterranean exchange flow, and the dominant pattern of the MOW are represented reasonably well by most of the models. Of those models considered, MPI-ESM-MR, MPI-ESM-LR, CSIRO Mk3.6.0, and MRI-CGCM3 provide the best MOW representation (E_j = 0.14, 0.19, 0.19, and 0.25, respectively). They are thus likely to be the most suitable choices for studies of MOW-dependent processes. However, the models experience salinity, temperature, and transport biases and do not represent temporal variability accurately. The implications of our results for future model analysis of the Mediterranean Sea overflow are discussed.

KEYWORDS: Coupled models; General circulation models; Empirical orthogonal functions; Mediterranean Sea; Model comparison; Model evaluation/performance; Ranking methods; Salinity; Seasonal variability; Temperature; Transport; Water masses/storage

1. Introduction

Evaporation losses over the Mediterranean Sea generate a horizontal density gradient with respect to the North Atlantic Ocean that is regulated by a two-layer exchange flow along the Strait of Gibraltar (SoG; Tsimplis et al. 2006; Soto-Navarro et al. 2015). Specifically, the mean eastward, warmer, and fresher Atlantic inflow in the surface layer overlies the westward, cooler, and saltier Mediterranean outflow in the intermediate layer (Ambar and Howe 1979; Bryden et al. 1994; Baringer and Price 1997; Hopkins 1999; Tsimplis et al. 2006; Soto-Navarro et al. 2010; García-Lafuente et al. 2011; Soto-Navarro et al. 2015). The resulting Mediterranean Overflow Water (MOW), and especially its salinity, play a key role in the stability of the convective cells in the eastern North Atlantic and hence the meridional overturning circulation (Artale et al. 2006; Koltermann et al. 2011; Rogerson et al. 2012). Temperature and salinity of western Mediterranean waters and MOW have been increasing significantly in observations and climate change scenarios, suggesting that the MOW has been and will continue to be an important mid-depth heat and salt source in the North Atlantic (Thorpe and Bigg 2000; Potter and Lozier 2004; Millot et al. 2006; Marcos and Tsimplis 2008; Schroeder et al. 2010). Therefore, the SoG—as a gateway to the Atlantic—is a crucial area for assessing the abilities of current-generation climate models.

MOW is the result of modified Eastern North Atlantic Central Water (ENACW) flowing into the Mediterranean Sea and then being transformed into Levantine Intermediate Water (LIW) as well as Eastern and Western Mediterranean Deep Water (WMDW), which in turn form the MOW (Ruti et al. 2016). MOW is the saltiest and warmest water mass in the intermediate layer of the eastern North Atlantic (36.3–36.5 and 11.7°–12.9°C, respectively) (Zenk 1970; Ambar and Howe 1979; Baringer 1993; Price et al. 1993; Price and O'Neil Baringer 1994; Baringer and Price 1997; Álvarez et al. 2004; Wu et al. 2007; Carracedo et al. 2016). It reaches neutral buoyancy and geostrophic balance at 7°W and 800–1200-m depth in the Gulf of Cádiz (GoC; Fig. 1, Ochoa and Bray 1991; Price et al. 1993; Sánchez-Leal et al. 2017; de Pascual-Collar et al. 2019). After exiting the narrow SoG (14.5 km wide and approximately 280 m deep), bathymetry has a strong impact on the MOW's composition and pathways and determines its integration with other Atlantic water masses (Sánchez-Leal et al. 2017).

The general circulation and variability of MOW have been explored using both observations and climate models (Ambar and Howe 1979; Bryden et al. 1994; Baringer and Price 1997; Hopkins 1999; Soto-Navarro et al. 2010; García-Lafuente et al. 2011; Soto-Navarro et al. 2015). Furthermore, thermohaline properties and volume transports at the SoG have been

Corresponding author: Lorine Behr, Lorine.Behr@zeu.uni-giessen.de

DOI: 10.1175/JPO-D-21-0082.1

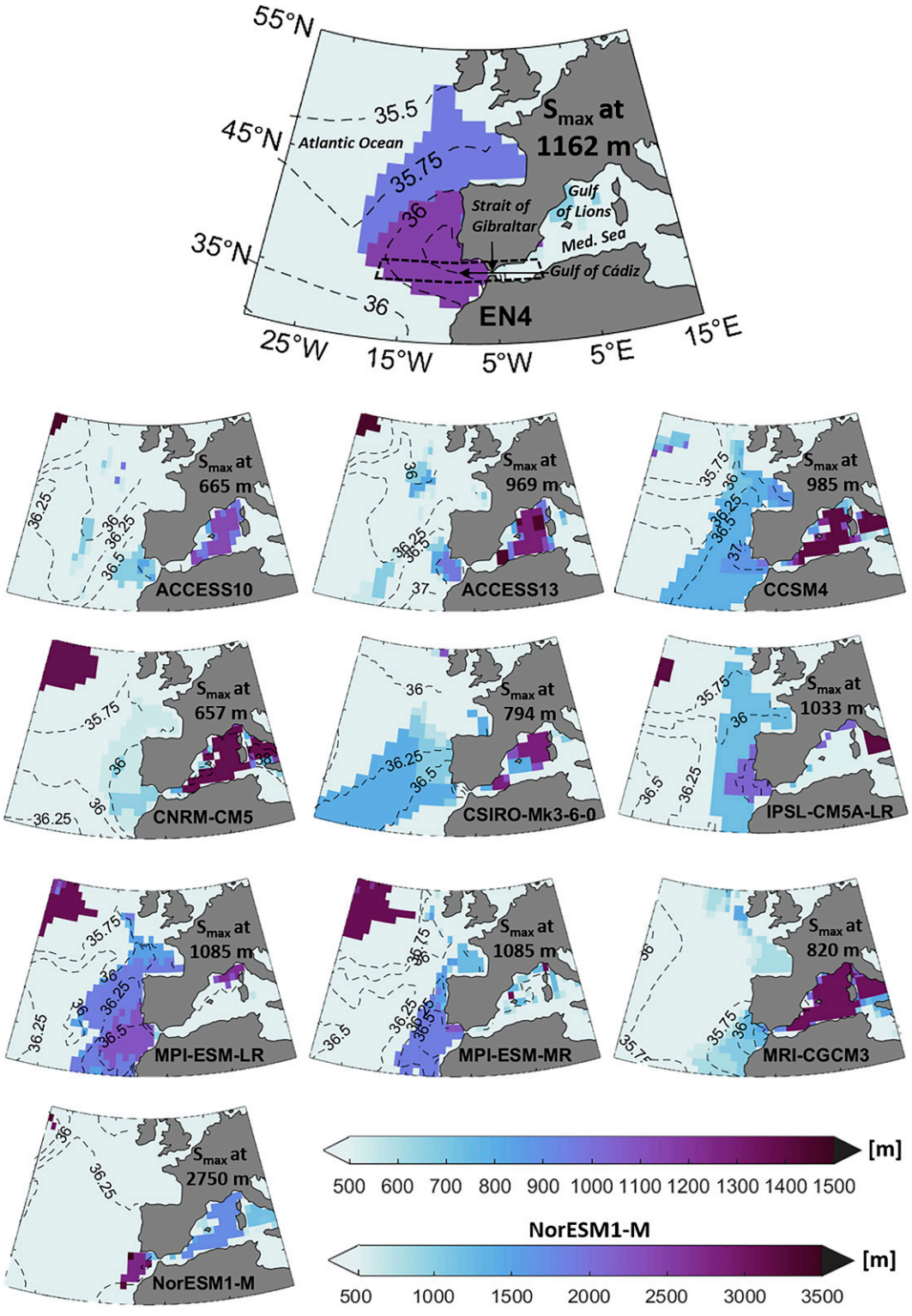

FIG 1. Twenty-year mean (1986–2005) salinity maxima (contours) and their corresponding depth levels (shading) between 500 (light blue) and 1500 (purple) m. The depth indication shows the depth of the salinity maximum S_{max} in the GoC in each model; the dash-outlined rectangle in the top image shows the main study area from 18°W (left side) to 0° (right side) and from 35°N (bottom side) to 37°N (top side) and the most important locations. Note the different color bar for NorESM1-M.

studied through hindcast runs (Béranger et al. 2005; Tonani et al. 2008; Oddo et al. 2009; Sannino et al. 2009), case studies (Vlasenko et al. 2009; Sánchez-Garrido et al. 2011), and regional climate change scenarios (Thorpe and Bigg 2000; Somot et al. 2006; Soto-Navarro et al. 2020). In addition, Soto-Navarro et al. (2015) investigated the climate variability of long-term hindcast simulations in comparison with long-term in situ observations. However, as yet, no comparable model intercomparison assessment of MOW exists for coupled general circulation models (CGCMs), including objective performance measures. This is likely due in part to the coarse horizontal resolution (typically ~1°) of CGCMs, which limits their ability to realistically represent bathymetry and small-scale processes like eddies, saltwater plumes, and overflows. In particular, the Mediterranean overflow involves several different physical processes that occur below the resolution of most climate models such as bottom boundary layer processes, entrainment, and rotating hydraulics of the exchange flow, including hydraulic control at the Spartel and Camerinal Sills, and possibly at the Tarifa Narrows. Overall control of the exchange may be maximal or submaximal (Armi and Farmer 1986; Farmer and Armi 1988), the two possibilities having important ramifications for mixing and exchange in the strait. Furthermore, the choice of vertical coordinates is particularly important as it is well known that coarse-resolution z-coordinate models suffer from spurious entrainment [for further discussion see Legg et al. (2006, 2009) and Fox-Kemper et al. (2019)]. In response to these problems, hydraulic control models (Armi and Farmer 1986; Farmer and Armi 1988; Siddall et al. 2002; Timmermans and Pratt 2005) and various parameterizations have been developed for use in climate models such as bottom boundary parameterizations (Tang and Roberts 2005), entrainment parameterizations, or overflow parameterizations (Price and Yang 1998; Wu et al. 2007; Legg et al. 2009). As an example, the NCAR CCSM3 overflow parameterization of Wu et al. (2007) [based on Price and Yang (1998)] involves hydraulic control theory to compute the overflow transport and entrainment. This reduces excessive numerical diffusion in coarse resolution z models [for further discussion see Legg et al. (2006, 2009)]. To the best of our knowledge, CCSM4 is the only model of those considered in this study that uses such an overflow parameterization, but unfortunately not at the SoG (Danabasoglu et al. 2012). Legg et al. (2009) propose the use of a "partially open barrier" algorithm that limits the width of the SoG to its actual width of 14.5 km and decreases the transports to realistic levels. However, in most coarse-resolution models, the strait is set to the size of the model grid and is thus much wider than in the real world (Wu et al. 2007). Without these approximations, models cannot hope to faithfully represent the physical conditions at the SoG. Nevertheless, because MOW is a key climate factor, it is important to determine to what extent it is represented in models. Here, we carry out the first such assessment focusing on CGCMs within phase 5 of the Coupled Model Intercomparison Project (CMIP5). Our paper addresses ocean and climate scientists who have used models participating in CMIP5 directly and/or have drawn conclusions about the behavior of the North Atlantic from their analyses. We aim to inform this community of the shortcomings with regard to their representation of MOW that have not previously been investigated. This information will prove valuable as studies now seek to understand differences between the North Atlantic properties of models participating in CMIP6 from those in CMIP5.

In particular, we consider the representation of MOW in 10 models participating in CMIP5. To do so, we first evaluate various performance metrics (13 in total) based on the representation of the climatological mean state and the temporal/spatial variability of seawater potential temperature and salinity (hereinafter simply temperature and salinity), and volume transports. Second, we perform a model ranking by calculating absolute ($A_{i,j}$), relative ($R_{i,j}$), and total relative (E_j) errors over each performance metric and model. Third, we identify and discuss the best performing models for representation of MOW properties as well as noting those models that have severe limitations. Last, we compare the model biases with those of two other water masses. This allows us to assess whether these biases can be explained by general model behavior or whether the models have difficulties in simulating the small-scale processes of the Atlantic–Mediterranean exchange.

This work is structured as follows: section 2 gives a brief description of the EN4 analyses (defined in section 2b), models, and outputs, and section 3 describes the calculation of volume transports, multivariate empirical orthogonal functions (MEOF), performance metrics, and the method employed for model ranking. Section 4 presents results and discussions and is split into five parts that focus on the MOW representation in models and EN4 analyses/observations as well as its seasonal and spatiotemporal variability, followed by a consideration of the overall model performance. In section 5 we compare the model performances with the works of Heuzé et al. (2013, 2015) and Heuzé (2017), who studied the water properties, transports, and formation processes of North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW) in models participating in CMIP5. Section 6 contains conclusions and a future outlook.

2. Data

a. *Models*

CMIP5 historical experiments (1850–2005) are widely used to assess the performances of CGCMs in comparison with other models or observations and have contributed to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Taylor et al. 2012). In this work we use monthly outputs of salinity, temperature, and zonal velocity to investigate how the MOW is simulated in each model. Salinity is represented on the practical salinity scale and hence has no unit. We define the MOW to be water in the GoC within the depth range 800–1200 m with a maximum salinity between 36.2 and 36.6 and a mean temperature between 10.5° and 12.9°C. We have retained the original horizontal and vertical discretizations because interpolation deforms the SoG in some models.

Of the 31 models in CMIP5 only 19 have a representation of the SoG and simulate an exchange flow between the Mediterranean Sea and the Atlantic Ocean. We chose 10 models

TABLE 1. Model names, model name abbreviations for MEOFs, local resolution of the study area (x is longitude, y is latitude, and L is the number of depth levels), type of horizontal Arakawa grid (Arakawa and Lamb 1977) and vertical grid in the ocean (z is geopotential, σ is terrain following, ρ is isopycnic, and H denotes a hybrid grid), ocean components, SoG width (km), and references.

Model name	Model acronym	Local resolution ($x/y/L$)	Arakawa/ vertical grid	Ocean component	SoG width (km)	Reference
ACCESS1.0	ACCESS10	1°/1°/50	B/z	MOM v.4p1	222	Bi et al. (2013)
ACCESS1.3	ACCESS13	1°/1°/50	B/z	MOM v.4p1	222	Bi et al. (2013)
CCSM4	CCSM	1.1°/0.5°/60	B/z	POP2	56	Danabasoglu et al. (2012)
CNRM-CM5A	CNRM	1°/0.8°/42	C/z	NEMO v3.2	178	Voldoire et al. (2013)
CSIRO Mk3.6.0	CSIRO	1.9°/0.9°/31	B/z	MOM v.2.2	100	Gordon et al. (2010)
IPSL-CM5A-LR	IPSL	1.4°/1.3°/31	C/z	NEMO v.3.2	144	Dufresne et al. (2013)
MPI-ESM-LR	MPI-LR	1°/0.8°/40	C/z	MPIOM	89	Jungclaus et al. (2013)
MPI-ESM-MR	MPI-MR	0.5°/0.5°/40	C/z	MPIOM	56	Jungclaus et al. (2013)
MRI-CGCM3	MRI	1°/0.5°/51	B/H σ–z	MRI.COM3	56	Tsujino et al. (2011)
NorESM1-M	NorESM	1.1°/0.5°/70	C/H z–ρ	MICOM	56	Bentsen et al. (2013)

from these 19 based on the requirement of realistic volume transports, considering a variety of horizontal resolutions and vertical coordinate systems (Table 1). In the horizontal plane, all models are formulated on Arakawa grids, with five using B grids and the other five using C grids. The majority of models solve the primitive equations with the Boussinesq approximation and conserve volume rather than mass. NorESM1-M employs an alternative non-Boussinesq mass-conserving formulation (Bentsen et al. 2013). Moreover, we use one ensemble member per model (rli1p1), since it is the only one that was available for all models as of September 2020. Eight of the models run on a vertical z-level grid. The other two, MRI-CGCM3 and NorESM1-M, run on hybrid coordinate systems.

In all models but CCSM4, the bottom topography is represented by a partial-step formulation. At the SoG, CCSM4 uses a cliff topography instead of its usual overflow parameterization to prevent excessive entrainment related to a staircase topography (Danabasoglu et al. 2012; Wu et al. 2007). For the local horizontal resolution, MPI-ESM-MR has the highest (0.5° × 0.5°) and IPSL-CM5A-LR the coarsest (1.4° × 1.3°) of the models under consideration (Table 1). The vertical resolution ranges from 31 levels in IPSL-CM5A-LR to 70 levels in NorESM1-M (Table 1). In particular, the sill depth of the strait is accurately represented as shallower than 300 m in all models, except for MRI-CGCM3 (~400-m depth). For our analysis, we use the 20-yr period from January 1986 to December 2005. Our study area ranges from 18°W to 0° and from 35° to 37°N (Fig. 1), and 1986–2005 is taken to be the reference period.

b. EN4 subsurface temperature and salinity dataset

Version 4 of the Met Office Hadley Centre "EN" series of datasets provides monthly gridded global quality-controlled temperature–salinity profiles as well as spatially complete analyses and covers the period from 1900 to present (Good et al. 2013). In this work we use these analyses (hereinafter EN4 analyses), which consist of a combination of profiles, a climatology, and an objective analysis scheme. EN4 has a horizontal resolution of 1° on a Cartesian grid with 42 levels from 5- to 5350-m water depth (Good et al. 2013). We use the 20 years of monthly salinity and temperature from January 1986 to December 2005. The EN4 dataset is used to assess the models' performance of the present-day overflow characteristics at the SoG.

3. Methods

a. Calculation of volume transports

Since streamfunctions or transports through the SoG are not directly available as outputs (as of September 2020), we use horizontal velocities to calculate the total in- and outflowing volume transports. We therefore integrate the zonal velocity from Morocco to Spain (35°–37°N, model dependent) at the longitude (between 5° and 7°W, model dependent) where the exchange flow can be measured in each model. We then use this result to calculate the Atlantic inflow in the surface layer and the Mediterranean outflow in the intermediate layer by first, filtering only positive (inflow) and negative (outflow) velocities and finally, integrating them over depth from the sill depth at approximately 300 m to the surface. Thereby, we define the Atlantic inflow as the eastward transport (positive values) and the Mediterranean outflow as the westward transport (negative values).

b. Multivariate empirical orthogonal function analysis

The MEOF method is a variant of the typical EOF method and has been widely used to study large-scale atmospheric and oceanic joint variability structures due to its ability to integrate different variables with their combined variances

(Xue et al. 2000; Sparnocchia et al. 2003; Xoplaki et al. 2003, 2004; Wheeler and Hendon 2004; Alvera-Azcárate et al. 2007). To represent the dominant spatiotemporal pattern of the MOW in the GoC, we perform MEOF analyses on monthly anomalies of salinity and temperature. Since the overflow originates from the influence of both variables, we prefer this variant to the general EOF, as it allows us to investigate their spatiotemporal covariability. Furthermore, MEOFs consider the joint connection and dependence of salinity and temperature, and they are able to combine the same physical mechanisms. The calculation is performed with two simultaneous fields, using time series of temperature and salinity. We first detrend the monthly anomaly data to remove possible first-order effects of anthropogenic climate change. Thereafter, we base the MEOFs on the correlation matrix because it standardizes both variables and thus removes their different units (Wilks 2011). Last, we project the resulting eigenvectors, where each eigenvector now consists of two spatial fields. The prevailing mode of the MOW is expected to be among the first few MEOFs because it dominates most of the variability. The selection of the simulated MEOFs was done qualitatively by choosing the mode closest to the spatial structure depicted in the EN4 analyses.

c. *Performance metrics and model ranking*

In this study, 13 metrics have been implemented to evaluate the performance of CGCMs (Table 2). Thereby, we focus on metrics that reflect the climatological mean state as well as the temporal and spatial variability of temperature, salinity, and volume transports at the SoG and in the GoC. The performance of the climatological mean state as well as its spread is measured by 20-yr climatological means and standard deviations based on monthly means. Additionally, we consider the performance of the exchange flow by evaluating its net flow as well as the amplitude of the annual cycle at the strait. Temporal variability is quantified by standard deviations as well as annual amplitudes of the exchange flow. To measure how well the models represent in MEOFs the spatial variability of MOW, we use the skill score S_{score} after Perkins et al. (2007), This skill score, calculated as

$$S_{\text{score}} = \sum_{1}^{n} \min(Z_m, Z_o),$$

measures the similarity between two probability density functions (PDFs) and allows a comparison across all statistical moments (Perkins et al. 2007). The EN4 and model data are binned around centers defined by the anomaly range of the models' MEOFs. Bin centers range from −1 to 1 with a bin size of 0.1 for all models and EN4 analyses. We quantify the common area between the PDFs of EN4 analyses and each model at 36°N, 18°W–0° and from the surface down to 1500-m depth (for NorESM1-M, down to 2500 m) by calculating the cumulative minimum value of each bin, where n is the number of bins (19) used for the PDF calculation of the spatial pattern of the MOW. Moreover, Z_m and Z_o are the frequencies of values in a given bin from each model and the EN4 analyses, respectively. The skill score can take on values between 0 and 1. It is equal to 1 if the total sum of the probability is reached at each bin center. A model then perfectly represents the spatial pattern of the MEOF of the EN4 analyses.

TABLE 2. List and acronyms of the considered performance metrics.

Performance metric	Acronym
Depth of Mediterranean Overflow (m)	Depth MOW
Mean salinity	Mean S
Salinity std dev	STD S
Mean temperature (°C)	Mean T
Temperature std dev (°C)	STD T
Mean Atlantic inflow (Sv)	Mean A_{in}
Atlantic inflow std dev (Sv)	STD A_{in}
Mean Mediterranean outflow (Sv)	Mean M_{out}
Mediterranean outflow std dev (Sv)	STD M_{out}
Ratio of exchange flow (Sv) (inflow ≥ outflow)	Net flow
Annual amplitude Atlantic inflow (Sv)	Amplitude A_{in}
Annual amplitude Mediterranean outflow (Sv)	Amplitude M_{out}
Skill score/error of MEOF analysis	Skill score/error

To be able to select and identify the best performing models, we have applied a ranking method that assigns equal weights to each of the 13 performance metrics (Rupp et al. 2013; Kamworapan and Surussavadee 2019). For a given metric i and climate model j, we first define an absolute error $A_{i,j}$,

$$A_{i,j} = |O_i - S_{i,j}|,$$

where O_i and $S_{i,j}$ are the analyzed/observed and simulated performance metrics, respectively. Since the metrics have different magnitude scales, we additionally calculate a relative error $R_{i,j}$,

$$R_{i,j} = \frac{A_{i,j} - A_{i,\min}}{A_{i,\max} - A_{i,\min}},$$

for each performance metric i and climate model j, where $A_{i,\min}$ and $A_{i,\max}$ are the minimum and maximum absolute errors for each performance metric, respectively. Since the skill score is already a scaled measure, we calculate a skill error SE_j instead of a relative error by subtracting the skill score from 1. Last, we sum the relative errors and the skill error of each model across all n metrics and divide them by n to get the total relative error E_j,

$$E_j = \frac{\sum_{i=1}^{n} R_{i,j} + \text{SE}_j}{n}.$$

The ranking of the models is then determined by ordering them according to their respective E_j, with the lowest E_j representing the most accurate model in terms of evaluations on MOW.

4. Results and discussion

This section is split into five parts: section 4a concentrates on the representation of the MOW in observations and

models and on respective absolute errors of the overflow and thermohaline characteristics. Section 4b assesses and discusses absolute errors of volume transports, section 4c focuses on the MOW's seasonal cycle, and section 4d considers the spatiotemporal variability by using MEOF analyses. Within these sections we follow a consistent structure: we first provide the results of the EN4 analyses or observed transports and thereafter show the model results, compare them first with analyses/observations and then among each other, and then we discuss the results. Section 4e assesses the ranking of the models considered and provides recommendations to the climate and modeling community.

a. *Representation and absolute errors of thermohaline properties*

First, it has to be mentioned that temperature and salinity in EN4 and the models participating in CMIP5 have been initialized with different datasets: the EN4 dataset is built on Argo (Argo 2000), Arctic Synoptic Basinwide Oceanography (ASBO), Global Temperature and Salinity Profile Program (GTSPP), and World Ocean Database 2013 (WOD13). The models are based on HadISST1.1, HadCRUT4, ERA40, the Global Ocean Heat Content data, and *World Ocean Atlas 2009* (Rayner et al. 2003; Uppala et al. 2005; Levitus et al. 2009; Antonov et al. 2010; Jones et al. 2012). Furthermore, EN4 and models are built on different methods for temperature bias correction (EN4: Gouretski and Reseghetti 2010; CMIP5: Levitus et al. 2009). Therefore, temperature and salinity may differ between EN4 and the model initialization datasets. For depths above 1000 m, the EN4 analyses are provided with observation weight analyses and uncertainty estimates (Good et al. 2013). Observation weights indicate how much a specific grid box in a specific time has been influenced by observations and how much by climatology, where high values imply good observational input. Uncertainty estimates provide the error standard deviations of the analyses, where low values imply a low standard error. The 20-yr mean standard errors for salinity and temperature within the depths of the MOW are not higher than 0.1 and 0.35°C, respectively (not shown). The observation weights lie between 40% and 70% at about 1000-m depth, increasing with time (not shown).

In EN4 analyses and models, we define the MOW according to the salinity maximum below 400-m depth, which indicates the equilibrium depth of the overflow (between 800 and 1200 m). In the EN4 analyses, the salinity maximum—and thus the prevailing westward salt tongue of the overflow—can be found around 1160-m depth, spreading from the GoC toward the central and northern North Atlantic (Fig. 1, top). For the modeled equilibrium depth, the absolute error ranges from 6 to 1550 m (Table 3). Six of 10 models simulate this metric reasonably well between 800- and 1200-m depth (Fig. 1): MPI-ESM-LR and MPI-ESM-MR as well as IPSL-CM5A-LR perform best (1033–1085 m), and ACCESS1.0, CNRM-CM5A, and CSIRO Mk3.6.0 have a shallow bias (657–794 m; Fig. 1). In NorESM1-M the MOW follows the isopycnals and reaches neutral buoyancy at two prevailing depths with the salinity maximum around 2750 m (Fig. 1). A reason for this might be the hybrid z–isopycnic grid of NorESM1-M, since the simulation of overflows is especially sensitive to the type of vertical coordinate scheme (Legg et al. 2009). In the coarse-resolution models considered here, the entrainment at shallower depths can be affected by spurious mixing and the ambient water properties as well as the indirect impact of topography; note that more-realistic topography can result in unrealistic entrainment. In addition, terrain-following and isopycnal models are relatively insensitive to resolution (Ezer and Mellor 2004; Ezer 2005). Low-resolution z-coordinate models are prone to excessive entrainment and spurious mixing in overflow regions (Legg et al. 2006, 2009).

Considering the EN4 20-yr mean vertical section in Fig. 2, the mean salinity and temperature between 7° and 11°W within the depth of the MOW plume are 36.2 and 10.5°C, respectively (Fig. 2, contours and shading). For the upper 300 m, the Atlantic inflow is characterized by salinities of 36.2 and temperatures of 15°C and the Mediterranean outflow is characterized by 38.4 and 13.5°C, respectively. In EN4, the standard deviations for salinity and temperature within the overflow are 0.08 and 0.19°C, respectively, which indicates that the variability of heat and salt in the overflow is low. Of the 10 models, 6 have a salty bias and 8 have a warm bias at the depth of the MOW plume (model dependent; Fig. 2): the absolute error, with regard to the EN4 analyses, reaches up to 0.7 and 3.5°C, respectively (Table 3). CNRM-CM5A, IPSL-CM5A-LR, MRI-CGCM3, and NorESM1-M simulate salinity accurately; ACCESS1.3, ACCESS1.0, and CCSM4 represent the saltiest models with 37.3, 37.1 and 37.1, respectively (contours in Fig. 2). MPI-ESM-LR and MPI-ESM-MR show an absolute error of 0.2 and 0.1 at the plume depth off the strait but simulate the plume correctly west of 12° (Fig. 2). Maximum salinity biases are reached in the western Mediterranean Sea by CCSM4, which can be up to +4 in the whole water column. IPSL-CM5A-LR and MRI-CGCM3 simulate the temperature properties within their overflow closest to observations (11.9 and 12.9°C, respectively; shading in Fig. 2), ACCESS1.0 and ACCESS1.3 show the strongest biases (16.4 and 16.0°C, respectively).

The T–S profiles from the GoC (box mean at 36°N, 15°–7°W) shown in Fig. 3 support some of the above results but also show that the total salinity profile in particular is not adequately represented by most models: in the EN4 analyses, salinity decreases constantly from the surface to a depth of about 500 m, then increases until the maximum is reached at ~1200 m, and then decreases again (Fig. 3a). One-half of the models are far too salty but are able to simulate a MOW typical profile. However, they show a weaker decrease in the surface to intermediate layer and a too-strong increase in the deeper layer before salinity declines again (Fig. 3a). The remaining five models have a fresh bias and hardly show the salinity decline in the surface layer but represent salinity at their depth of the MOW plume closest to EN4 analyses (Fig. 3a). For the temperature profile, the majority of models have strong negative biases in the surface layers and strong positive biases in the deeper layers relative to the EN4 analyses (Fig. 3b). The maximum temperature bias in the overflow plume

TABLE 3. Evaluation of 13 performance metrics showing the total relative error E_j, absolute values and errors (AV/AE) of each model, and EN4 analyses and observations over 1986–2005. For the skill score, a skill error (SE) has been calculated. See Table 2 for a list of the acronyms of the considered performance metrics and their units.

		Depth MOW		Mean S		STD S		Mean T		STD T		Mean A_{in}		STD A_{in}		Mean M_{out}	STD M_{out}		Net flow		Amplitude A_{in}		Amplitude M_{out}		Skill score	
Metrics/models	E_j	AV	AE	AV	AE	AV	AE	AV	AE	AV	AE	AV	AE	AV	AE	AV	AV	AE	AE	AE	AV	AE	AV	AE	AV	SE
EN4/Obs	—	800–1200	—	36.2–36.6	—	0.08	—	10.5–12.9	—	0.19	—	0.7–1.3	—	0.05	—	From −0.7 to −1.0	0.06	—	0.1	—	0.1	—	0.03–0.14	—	1	—
MPI-ESM-MR	0.14	1085	0	36.8	0.2	0.06	0.02	14.1	1.2	0.16	0.03	0.8	0.0	0.13	0.08	−0.7	0.16	0.10	0.1	0.0	0.23	0.13	0.36	0.22	0.73	0.27
MPI-ESM-LR	0.19	1085	0	36.7	0.1	0.04	0.04	14.0	1.1	0.07	0.12	1.0	0.0	0.12	0.07	−0.9	0.14	0.08	0.1	0.0	0.21	0.11	0.31	0.17	0.72	0.28
CSIRO Mk3.6.0	0.19	794	6	36.8	0.2	0.06	0.02	15.0	2.1	0.25	0.06	0.9	0.0	0.10	0.05	−1.0	0.12	0.06	−0.1	0.1	0.29	0.19	0.32	0.18	0.76	0.24
MRI-CGCM3	0.25	820	0	36.3	0.0	0.03	0.05	12.9	0.0	0.13	0.06	1.6	0.3	0.21	0.16	−1.0	0.13	0.07	0.6	0.5	0.42	0.32	0.32	0.18	0.72	0.28
ACCESS 1.3	0.34	969	0	37.3	0.7	0.04	0.04	16.0	3.1	0.20	0.01	0.6	0.4	0.15	0.10	−0.5	0.13	0.07	0.1	0.0	0.4	0.3	0.33	0.19	0.62	0.38
NorESM 1-M	0.36	2750	1550	36.4	0.0	0.03	0.05	8.5	2.0	0.12	0.07	0.9	0.0	0.10	0.05	−1.3	0.18	0.12	−0.4	0.3	0.25	0.15	0.52	0.38	0.62	0.38
ACCESS 1.0	0.38	665	135	37.1	0.5	0.14	0.06	16.4	3.5	0.32	0.13	0.5	0.5	0.13	0.08	−0.4	0.10	0.04	0.1	0.0	0.31	0.21	0.25	0.11	0.7	0.30
CNRM-CM5A	0.38	657	143	36.2	0.0	0.04	0.04	14.5	1.6	0.17	0.02	1.5	0.2	0.24	0.19	−1.9	0.32	0.26	−0.4	0.4	0.31	0.21	0.54	0.40	0.68	0.32
CCSM4	0.40	985	0	37.1	0.5	0.09	0.01	15.2	2.3	0.30	0.11	1.4	0.1	0.19	0.14	−1.4	0.27	0.21	0.0	0.0	0.4	0.3	0.74	0.60	0.68	0.32
IPSL-CM5A-LR	0.63	1033	0	36.5	0.0	0.01	0.07	11.9	0.0	0.06	0.13	3.6	2.3	0.30	0.25	−3.6	0.43	0.37	0.0	0.0	0.4	0.3	1.00	0.86	0.71	0.29

reaches up to ~5°C in ACCESS1.0 and ACCESS1.3 (Fig. 3b). Most models agree well with the EN4 analyses below 2000-m depth (Figs. 3a,b). NorESM1-M simulates a totally different T–S profile: T–S constantly decrease until 1100-m depth, then increase to a first maximum at ~1500 m, and finally decrease and increase again until the absolute maximum at ~2750-m depth (Figs. 3a,b).

The salinity and temperature standard deviations within the overflow are extremely variable from model to model. The absolute errors range from 0.01 to 0.07 and from 0.01° to 0.13°C, respectively (Table 3). The majority of models underestimate the salinity standard deviation, and ACCESS1.0 overestimates it (σ = 0.14). CCSM4, CSIRO Mk3.6.0, and MPI-ESM-MR simulate well the spread around the mean (σ = 0.09, 0.06, and 0.06, respectively; Table 3). The temperature standard deviation is most accurately represented by ACCESS1.3 and MPI-ESM-MR (σ = 0.20° and 0.16°C). Six of 10 models underestimate the metric and 4 overestimate it, indicating the disagreement among the models. ACCESS1.0 and IPSL-CM5A-LR represent the MOW temperature standard deviation worst (σ = 0.32° and 0.06°C).

There are several potential reasons for the high T–S biases in almost all models. Soto-Navarro et al. (2015) suggest that some models simulate the composition of MOW incorrectly because they underestimate the proportion of WMDW or overestimate the fraction of the much warmer and saltier LIW or the much warmer and fresher ENACW. Thus, this discrepancy might affect the T–S characteristics of the resulting mixed waters that exit the strait toward the North Atlantic. Furthermore, the dominant driver of the aspiration of WMDW is tides, which are not included in the models (Soto-Navarro et al. 2015). Thus, an overestimation of LIW is not unexpected. In addition, temperature and salinity may differ between EN4 and the models since they are built on different methods for bias correction (see section 4a). A further potential factor is variation in entrainment between different models as this depends on the ambient properties of the nearby Atlantic basin.

b. *Representation and absolute errors of volume transports*

Long-term time series of volume transports are not available at the SoG. Therefore, we refer to various studies that have examined observations of the exchange flow between the Atlantic and the Mediterranean Sea. The mean eastward flowing Atlantic inflow in the surface layer is 0.7–1.3 Sv (1 Sv ≡ 10^6 m^3 s^{-1}) and thus surpasses the westward-flowing Mediterranean outflow from −0.7 to −1.0 Sv in the intermediate layer (Ambar and Howe 1979; Bryden et al. 1994; Baringer and Price 1997; Hopkins 1999; Tsimplis and Bryden 2000; Soto-Navarro et al. 2010; García-Lafuente et al. 2011; Soto-Navarro et al. 2015). The mean standard deviations of the in- and outflow are 0.05 and 0.06 Sv, respectively, indicating a low variability of the exchange flow strength (Soto-Navarro et al. 2010).

The absolute errors of the mean in- and outflow are up to 2.3 and 2.6 Sv, respectively (Table 3). Four models simulate the flows well within the observed range. MPI-ESM-LR and

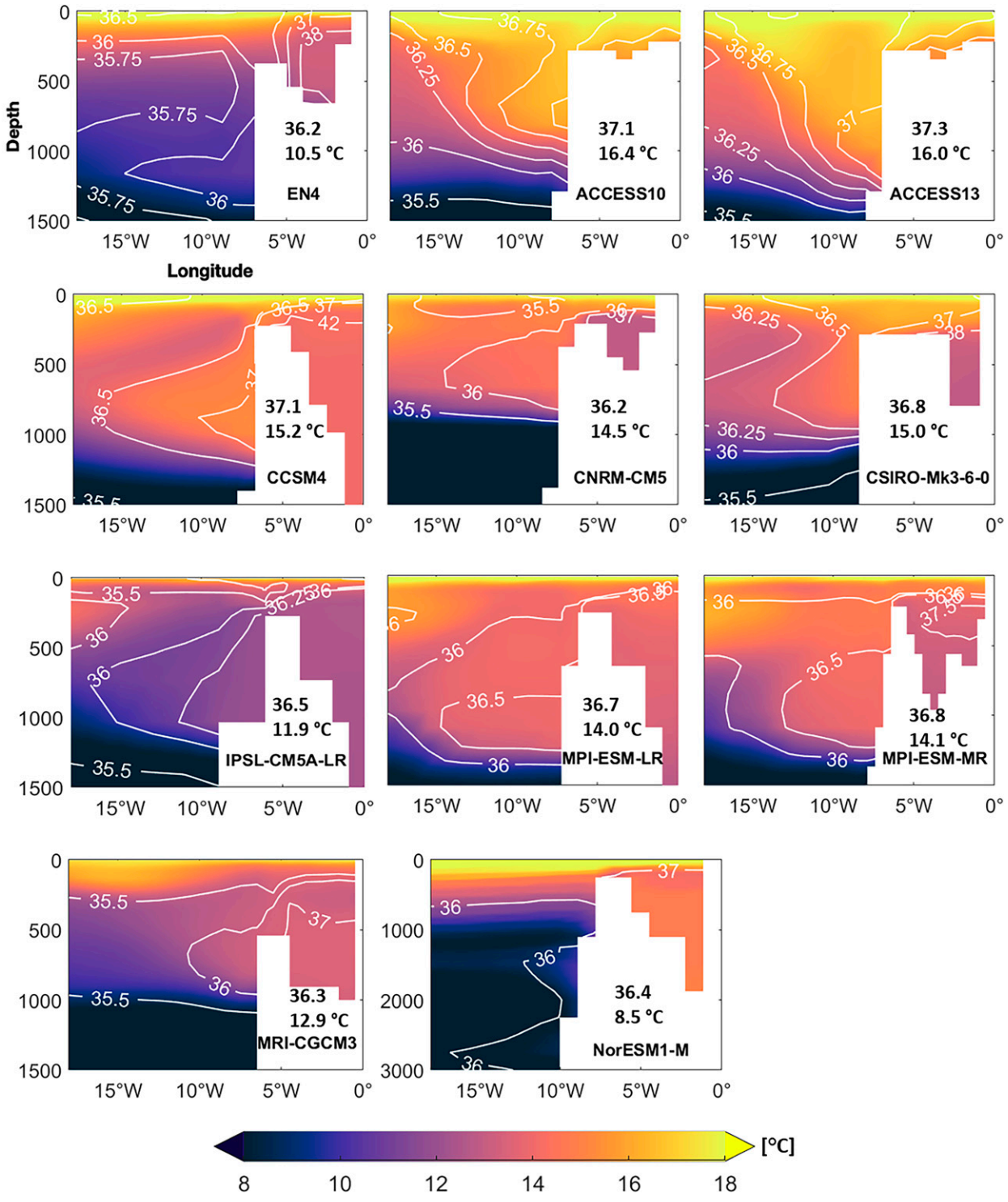

FIG 2. Twenty-year mean (1986–2005) salinity (contours) and temperature (°C) (shading) vertical sections at the SoG/GoC at 36°N and from 18°W (left side) to 0° (right side) as a function of depth (*y* axis; m). Shown are analyses and historical model output; the boldface values show the mean temperature and salinity at the depth of the salinity maximum. Note the different *y* axis for NorESM1-M.

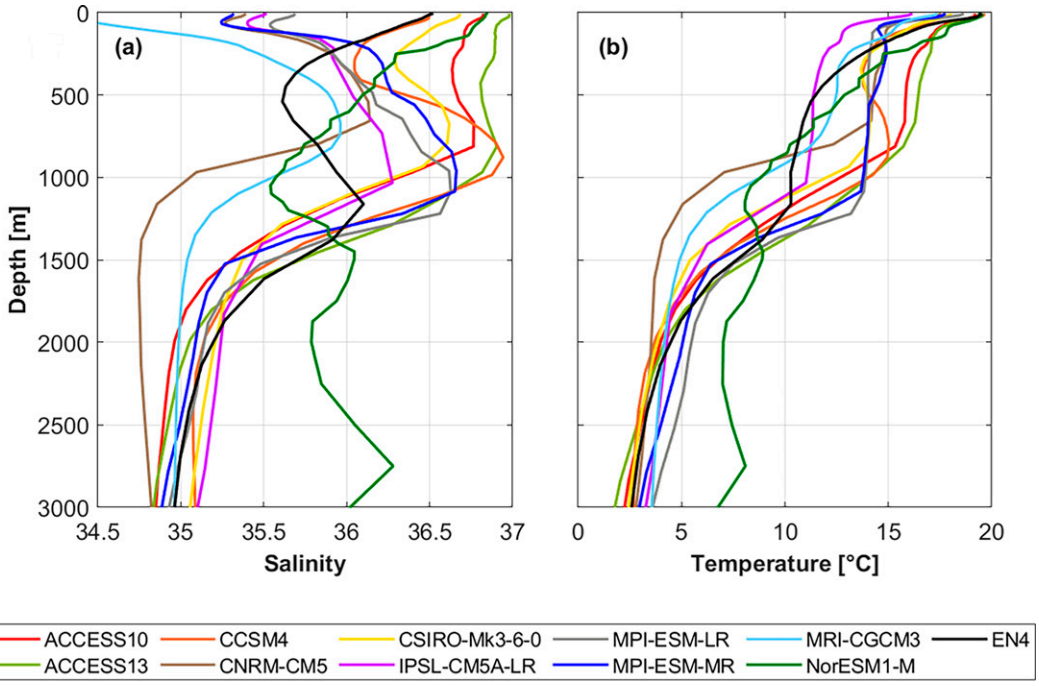

FIG 3. Twenty-year mean (1986–2005) (a) salinity and (b) temperature profiles along the GoC (box average at 36°N, 15°–7°W) as a function of depth (y axis; m) to detect and compare present MOW among the 10 models.

MPI-ESM-MR most accurately represent the observations with average Atlantic inflows of 1.0 and 0.8 Sv and Mediterranean outflows of −0.9 and −0.7 Sv (Table 3). ACCESS1.0 and ACCESS1.3 underestimate the flows, and especially IPSL-CM5A-LR overestimates them (±3.6 Sv). In CCSM4 and NorESM1-M, the in- and outflow are not simulated correctly, which might have implications for the volume transport strengths. In CCSM4 the exchange flow is not simulated at a common latitude. The inflow can be found at 35.7°N from the surface to the sill depth, and the outflow can be found at 36.2°N from the surface to approximately 150-m depth (not shown). In NorESM1-M, a westward flow off the strait partially blocks the Atlantic inflow and therefore also inhibits an accurate simulation of the exchange (not shown).

In nearly all models, the standard deviation is simulated at least 2 times as strongly as in observations (Table 3), with IPSL-CM5A-LR showing the highest spread (inflow: 0.30 Sv; outflow: 0.37 Sv, Table 3). The absolute errors of the in- and outflow standard deviation range from 0.05 to 0.25 Sv and from 0.04 to 0.37 Sv. The simulated variability of the Mediterranean outflow is stronger than that of the Atlantic inflow (Table 3). In addition, we have examined whether the models accurately reproduce the ratio of the exchange flow. Accuracy is achieved when the inflow is at least equal to or greater than the outflow. The majority of models simulate the ratio of the exchange flow correctly (Table 3). However, in CSIRO Mk3.6.0, CNRM-CM5, and NorESM1-M the outflow surpasses the inflow by 0.1, 0.4, and 0.4 Sv, respectively. Volume transports like these can potentially lead to a loss of water mass in the Mediterranean Sea—as more water flows out than in—and are therefore unreliable. Further complications can arise from excessive runoff or precipitation into the basin as a whole. An additional indicator for an accurate representation of the exchange flow is a high correlation between the Atlantic inflow and Mediterranean outflow. The majority of models considered show a correlation higher than 0.9 (statistically significant at the 99th confidence interval) pointing to a strong connection with a higher/lower inflow that is compensated by a higher/lower outflow and vice versa, consistent with observations (not shown).

Although the horizontal grid type (Arakawa B or C) is one of the key factors for an accurate simulation of the exchange flow, we have not found any relation between the type of grid and the performance of transports at the SoG. The horizontal grid has a crucial impact on how the models can represent the width of the SoG and where their horizontal velocity points are located. ACCESS1.0, ACCESS1.3, CCSM4, CSIRO Mk3.6.0, and MRI-CGCM3 are formulated on an Arakawa B grid, where the velocities are located at the same position, that is, in the center of the grid. CNRM-CM5A, IPSL-CM5A-LR, MPI-ESM-LR, MPI-ESM-MR, and NorESM1-M come with an Arakawa C grid, where the velocities are placed at their corresponding side centers (Arakawa and Lamb 1977; Griffies 2004; Collins et al. 2013). However, the width of the SoG seems to play a crucial role: Legg et al. (2009) state that the wider the strait in coarse-resolution models (typically the size of the model grid), the stronger the exchange through the SoG and vice versa. We work with models that predominantly have a horizontal resolution of approximately 1° and whose SoG is mostly set to the size of the model grid (Table 1). Thus, it is not surprising that most of them simulate the exchange flow

too high. Moreover, none of the models considered has an overflow parameterization at the SoG. This presents challenges, such as moving the fluid down the slope without excessive mixing, simulating hydraulic control at the strait, and achieving the right amount of entrainment (Legg et al. 2009). Furthermore, the negative net flow in NorESM1-M may arise because the model does not conserve volume. In volume conserving models, the in- and outflow almost balance each other, as can be observed in our results. The same applies to the standard deviations.

c. *Seasonal variability in models and observations*

In this section, we investigate whether the models are able to represent the seasonal cycle of the exchange flow, focusing on volume transports and the Mediterranean outflow temperatures (at ~300-m depth). In observational data, the maximum inflow is reached in late summer, being a result of the barotropic forcing of the evaporative cycle (Soto-Navarro et al. 2010). The outflow peaks in late winter–early spring and coincides with the minimum in temperature (García-Lafuente et al. 2007; Soto-Navarro et al. 2010; Sammartino et al. 2015). Thereby, the outflow is likely linked with the winter deep-water formation in the Gulf of Lion (García-Lafuente et al. 2007; Sánchez-Román et al. 2009).

From model to model, the seasonal exchange of the Atlantic inflow and Mediterranean outflow is highly variable. Figure 4 shows that the majority of models simulate too-weak Atlantic inflows (positive values) over the entire year relative to the maximum inflow in observations (dashed lines in Fig. 4a). In contrast, CNRM-CM5A, CCSM4, MRI-CGCM3, and IPSL-CM5A-LR have a too pronounced inflow. CNRM-CM5A, CCSM4, and NorESM1-M simulate the Mediterranean outflow (negative values) too strongly (Fig. 4a). A comparison of the monthly multimodel maximum outflow (average over all models) and the monthly multimodel mean temperature within the outflow (thick black lines in Figs. 4a,b) shows that the majority of models simulate a stronger and cooler outflow in winter (maximum in January) and a weaker and warmer outflow in summer (minimum in August), which is almost consistent with observations (Candela 2001; García-Lafuente et al. 2007).

As an additional measure of variability, the annual amplitude of the in- and outflow has been included (Table 3). In observations, the amplitude indicates a low annual variability of the exchange flow and does not exceed 0.1 Sv in the inflow and 0.03–0.14 Sv in the outflow (Table 3; Bryden et al. 1994; Candela 2001; García-Lafuente et al. 2007). The absolute model error of the annual amplitude ranges from 0.11 to 0.32 Sv in the inflow and from 0.11 to 0.86 Sv in the outflow (Table 3). MPI-ESM-LR and MPI-ESM-MR most accurately represent the variability of the inflow (0.21 and 0.23 Sv); ACCESS1.0 and MPI-ESM-LR represent that of the outflow (0.25 and 0.31 Sv). Our evaluation reveals that the simulated variability of the exchange flow (standard deviation and annual amplitude) is too high among all CMIP5 models relative to observations.

Moreover, the forcing of the Atlantic inflow is not represented correctly. In observations, the inflow is the result of a barotropic forcing and reaches its maximum at the same time as evaporation reaches its maximum in summer (Candela 2001; Soto-Navarro et al. 2010). In the models, the Atlantic inflow behaves almost like the outflow and reaches its trough in summer. We suggest that the evaporative cycle is not represented accurately by the low-resolution models over the Mediterranean basin, which therefore leads to an underestimation of the inflow in five models. Further work is needed to evaluate the evaporation/precipitation and runoff fluxes in CMIP5 (and CMIP6) models. One issue limiting such a study is that observation-based estimates of evaporation/precipitation fluxes still contain relatively large uncertainties when integrated over the Mediterranean Sea (Jordà et al. 2017). Nevertheless, it may still be possible to discriminate between the models at a useful level using the available observation-based datasets and we encourage research in this area.

In contrast to observations, temperatures within the outflow reach their minimum already in February and not in April. In addition, the maximum outflow is shifted to January instead of February (Fig. 4a). This implies that either the winter deep-water formation in the Gulf of Lion—and thus the cooling of intermediate water masses—is temporally shifted in the models or that the variability of temperature is based on a different process such as the annual cycle of air temperature. In any case, the Mediterranean outflow is stronger when the sea surface and mixed layer temperatures are cool, and it is weaker when temperatures reach their maximum in summer (Figs. 4a,b). Models that match this winter maximum/summer minimum (except CSIRO Mk3.6.0) show strong and significant negative correlations (99th confidence level) between the strength of the Mediterranean outflow and temperature within the strait (~−0.6; not shown). ACCESS1.0 and ACCESS1.3 constitute an exception: both models simulate minimum transports from February to April and maximum values in summer when the other models reach their maxima and minima, respectively (Fig. 4a). Therefore, both models show weak and/or positive correlation coefficients in comparison with the other models (ACCESS1.0: 0.6; ACCESS1.3: −0.1, not significant).

d. *Spatiotemporal variability in models and observations*

Performing the MEOF analysis is a key element of this work and allows a first insight into how models simulate the spatiotemporal variability of MOW on a multivariate field. To the best of our knowledge, there is no comparable study on MOW, making our work the first such assessment and a valuable approach for future research, for example, with the next generation CMIP6 group of models.

In this section, we present MEOFs of salinity and temperature for the EN4 analyses and each model on separate vertical sections (Fig. 5). In EN4, we find the pattern of the MOW on the first leading MEOF, explaining 19% of the joint (standardized) variance of salinity and temperature over the period 1986–2005 (Fig. 5, top). Positive salinity anomalies (green) are associated with positive temperature anomalies (red) in the prevailing depth of the overflow plume (800–1200 m; Fig. 5). The *T*–*S* patterns compensate for the density in the positive (salty and warm) and negative (fresh and cold) phases, and

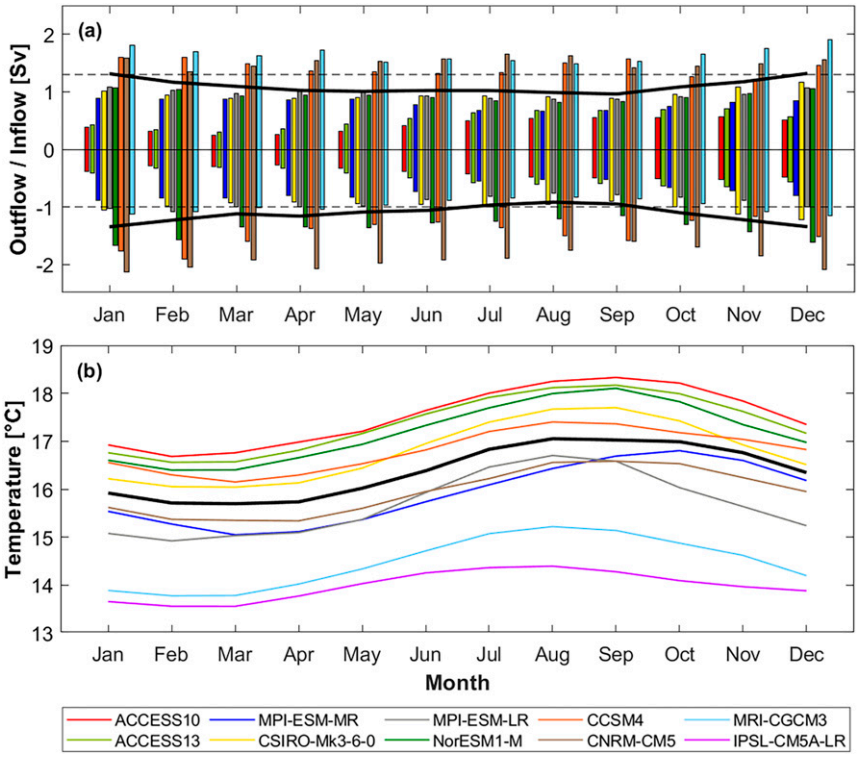

FIG 4. Seasonal variability of Atlantic inflow and Mediterranean outflow over 1986–2005. (a) Monthly mean volume transports of the inflow (positive values) and outflow (negative values); thick black lines show the multimodel maxima and dashed lines show maximum volume transports from observations as in Table 3. (b) Monthly mean temperature within the outflow; the thick black line shows the multimodel mean. For better readability, the overestimated volume transports of IPSL-CM5A-LR (±3.6 Sv) have been excluded in (a).

hence the prevailing depth of the overflow plume does not vary strongly from phase to phase.

Six of 10 models simulate the dominant pattern of the Mediterranean overflow on their first MEOF, and four models simulate it on their second MEOF, with the total explained variance ranging from 16% to 32% in the first MEOF mode and from 15% to 20% in the second MEOF mode (Table 4). Thereby, we select the first or second MEOF qualitatively by choosing the mode closest to the spatial structure depicted in the EN4 analyses. However, the orthogonality constraint, implicitly involved in the MEOF approach, complicates the comparison, since the leading pattern already controls at least some of the spatial pattern of the higher indexed MEOFs. In the overall pattern, and most likely owing to the coarse resolution of all models used in the analysis, profound differences to the observationally based EN4 dataset are evident (Fig. 5).

However, all models simulate positive salinity anomalies associated with positive temperature anomalies during a positive phase, consistent with EN4 (Fig. 5). The location of the anomalies varies from model to model and often corresponds to the MOW's equilibrium depth (cf. Figs. 2, 3, and 5). Differences between MPI-ESM-LR and MPI-ESM-MR may be related to the changes in horizontal resolution by using a bipolar and a tripolar grid, respectively. Both models share the same ocean model (MPIOM) as well as the number of vertical levels (Table 1), which might be an important characteristic, especially in the context of the SoG. To measure the performance of spatial variability, we consider the skill score of Perkins et al. (2007): this ranges from 0.62 in ACCESS1.3 to 0.76 in CSIRO Mk3.6.0, where a skill score of 1 denotes perfect agreement of the PDFs of model and analyses. The majority of models achieve skill scores of 0.7. Thus, their spatial variability shows a good correspondence to the EN4 dataset. However, the considerably different MEOF structure in comparison with EN4 also indicates that most models still lack a realistic simulation of the in- and outflow variance modes of the MOW.

e. *Model ranking*

In this subsection, we perform a model ranking by evaluating and interpreting the relative ($R_{i,j}$) and total relative (E_j) errors of the 10 models considered. In addition, the multimodel mean relative errors of the individual metrics are used to identify not-well-simulated fields across all models. The majority of models simulate at least 7 of 13 performance metrics well ($R_{i,j} < 0.2$) and 1 of 13 metrics poorly ($R_{i,j} > 0.8$; Table 3). MPI-ESM-MR provides the best MOW representation and

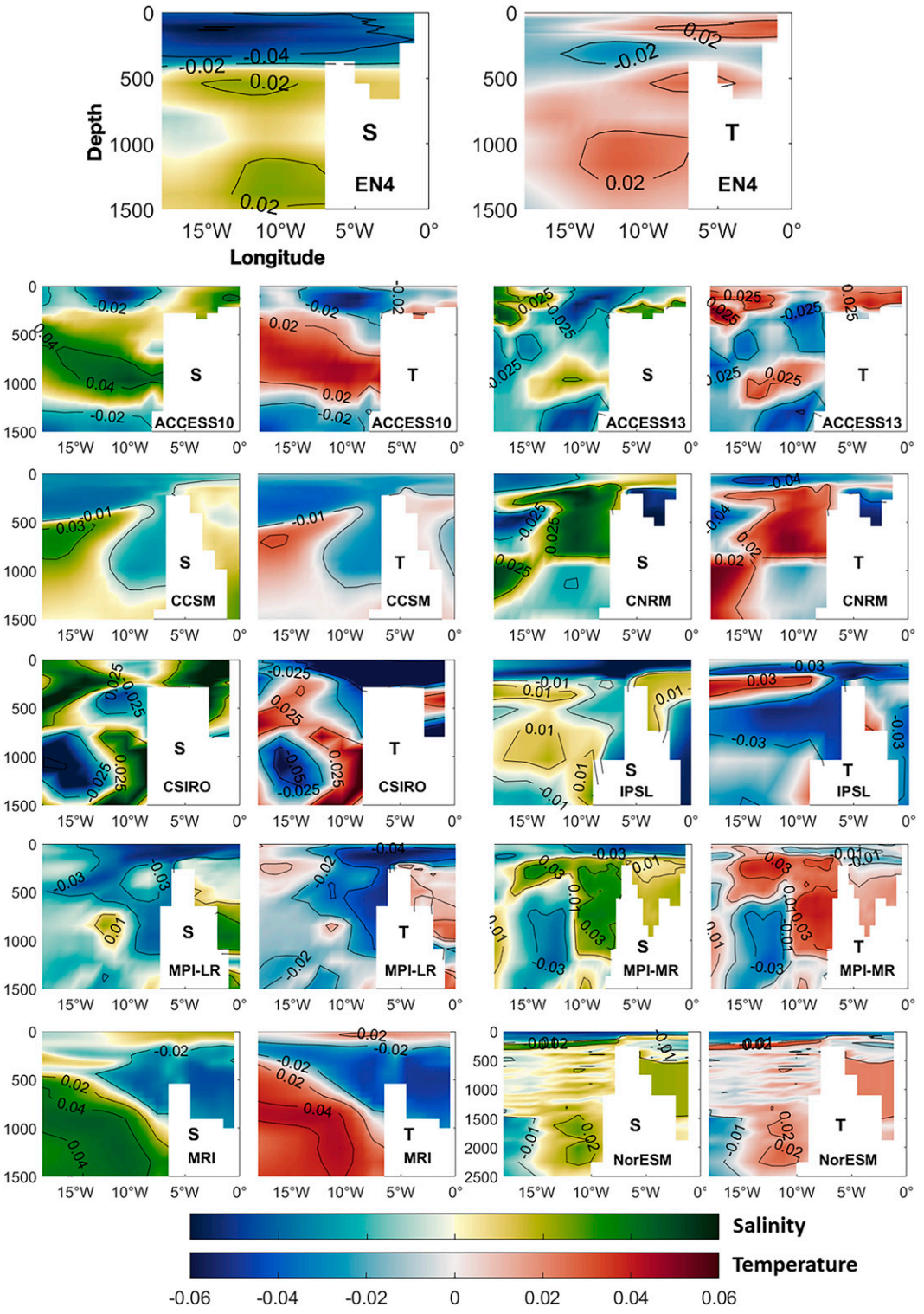

FIG 5. Linear detrended MEOFs of monthly salinity (*S*; shades from blue to green) and temperature (*T*; shades from blue to red), showing the prevailing pattern of the Mediterranean overflow along the SoG/GoC at 36°N and from 18°W (left side) to 0° (right side) as a function of depth (*y* axis; m) over 1986–2005. Shown are EN4 analyses and historical model output. Note the different *y* axis for NorESM1-M.

TABLE 4. Total explained variance of the first and second MEOF over 1986–2005. The corresponding MEOF, representing the pattern of the Mediterranean overflow, is shown in boldface type.

Analyses/Models	MEOF 1	MEOF 2
EN4 analyses	**19%**	9%
ACCESS1.0	39%	**18%**
ACCESS1.3	46%	**20%**
CCSM4	**32%**	10%
CNRM-CM5	30%	**20%**
CSIRO Mk3.6.0	43%	**15%**
IPSL-CM5A-LR	**29%**	15%
MPI-ESM-LR	**22%**	21%
MPI-ESM-MR	**16%**	12%
MRI-CGCM3	**29%**	14%
NorESM1-M	**17%**	13%

simulates well 10 of 13 metrics and none poorly, whereas IPSL-CM5A-LR only represents 4 metrics well and 8 poorly. These models also show the lowest (E_j = 0.14) and highest (E_j = 0.63) total relative error, respectively (Fig. 6a). We therefore suggest that MPI-ESM-MR is the best choice from this CMIP5 set for studies of MOW properties and its potential wider impacts. The next best models are MPI-ESM-LR, CSIRO Mk3.6.0, and MRI-CGCM3, reaching total relative errors of 0.19, 0.19, and 0.25, respectively (Fig. 6a). The three models simulate well 9, 8, and 8 of 13 metrics and do not perform well in 1, 0, and 1 metric, respectively (Table 3). All other models have weaknesses for studies on MOW because of the following characteristics: ACCESS1.0 and ACCESS1.3 have strong salinity and temperature biases and simulate the exchange flow too weakly (Figs. 2 and 3; Table 3); CNRM-CM5 and NorESM1-M do not accurately represent the net flow (outflow > inflow) and reach neutral buoyancy in too-shallow (657 m) and too-deep ocean layers (2750 m), respectively (Figs. 1 and 4; Table 3). This implies that the MOW might take wrong pathways in the North Atlantic because it interacts with different water masses. CCSM4 also shows excessive salinity and temperature biases that negatively affect most performance metrics (Figs. 2 and 3; Table 3). We point out that the MOW is only one of several potential criteria to qualify a model. Some models may show a good performance for biased physical reasons; for example, a salty and warm bias compensates for density and thus shows an accurate Mediterranean salt tongue. This could also be the case for less well performing models that are influenced by processes that are different in the real world, for example, a velocity bias due to a bad resolution of bathymetry.

Certain metrics have been poorly simulated by almost every model. In particular, the majority experience problems with the correct representation of temporal variability such as the standard deviations of the 20-yr salinity and temperature, as well as the annual amplitude of the Atlantic inflow (Fig. 6b). Thereby, the multimodel mean relative errors are 0.50, 0.53, and 0.53, respectively. The simulated spread around the mean is too weak for salinity and temperature and too high for the annual amplitude in comparison with EN4 and observations

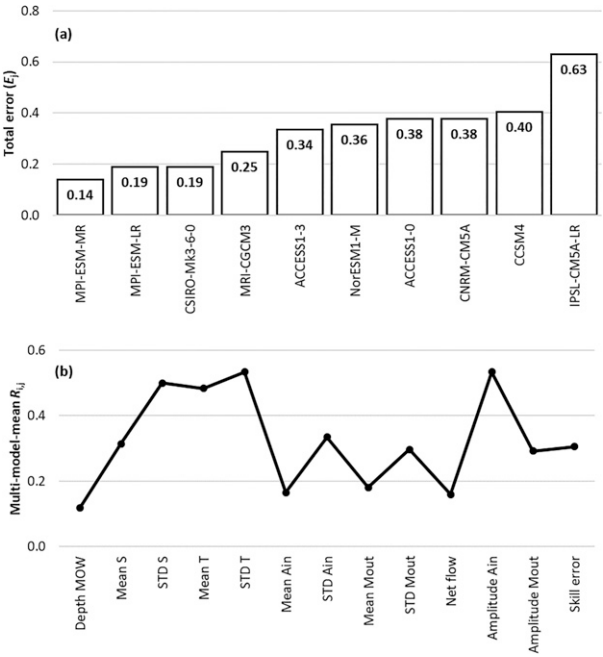

FIG 6. (a) Total relative errors E_j of all 10 CGCMs and (b) multimodel mean relative errors $R_{i,j}$ and skill error of all metrics over 1986–2005.

(Table 3). Furthermore, the multimodel mean relative error for the 20-yr mean temperature is 0.48, which reinforces the warm bias across most models. We suggest that this is an overall problem of the models that has to be considered for model development by the climate model community. Further, the equilibrium depth of the MOW, the 20-yr mean exchange flow, the net flow as well as the spatial variability (MEOF skill score) are represented reasonably well by almost all models under consideration (Table 3). However, it must be considered that we have initially made a preselection of models with regard to the most realistic representation of transports. Therefore, this positive result cannot be generalized to all models participating in CMIP5.

5. Representation of other water masses in CGCMs

Beyond our results, we are interested in whether similar *T*–*S* and volume biases can be found for other water masses that are not necessarily connected to the MOW and thus do not pass on any of their own biases. We intend to match the models' *T*–*S* biases of MOW with those of AABW and NADW and match volume biases with those of the Atlantic meridional overturning circulation (AMOC), using the results of Heuzé et al. (2013, 2015) and Heuzé (2017), who cover the same time period (1986–2005) as we do. For the water-mass properties of AABW, the CNRM-CM5, MPI-ESM-LR, and MRI-CGCM3 demonstrate a warm bias of 1 and CSIRO Mk3.6.0, IPSL-CM5-LR, and NorESM1-M demonstrate a warm bias of below 0.5°C (Heuzé et al. 2013). For NADW in the subpolar gyre and the Greenland–Iceland–Norwegian Seas, most models respectively have a warm bias (0.2°–3°C;

0.8°–3°C) and a salty bias (0.2–0.6; 0.1–0.6) at 750–1250-m water depth (Heuzé 2017) as is the case for MOW in this study. The saltiest models include CSIRO Mk3.6.0 and MPI-ESM-MR, the freshest include CNRM-CM5A and IPSL-CM5A-LR, which is also consistent with our findings. In terms of transports, the majority simulates the AMOC within the range of observations (17.4 ± 4.8 Sv; Heuzé et al. 2015). The best-performing models are ACCESS1.0, CCSM4, and MPI-ESM-LR (18–19 Sv). CNRM-CM5A-LR, IPSL-CM5A-LR, and MPI-ESM-MR show weaker transports (10–12 Sv), and CSIRO Mk3.6.0 and NorESM1-M show stronger transports (20 and 32 Sv; Heuzé et al. 2015; Heuzé 2017). In this work, ACCESS1.0 simulates the Atlantic–Mediterranean exchange flow with the weakest transports and IPSL-CM5A-LR simulates with the strongest transports. In both works, MPI-ESM-LR represents the transports most accurately and CSIRO Mk3.6.0 and NorESM1-M are within the models showing stronger transports. In summary, we cannot draw any significant conclusion about volume transports from the above results to our study. However, the warm and salty biases are consistent with this work. We therefore suppose that their origin is not only related to the poor representation of the Mediterranean, but that a general model behavior may be also assumed.

6. Conclusions and outlook

In this work, we evaluate the performance of 10 CGCMs, which participated in CMIP5, in representing recent (1986–2005) MOW properties along the SoG and the GoC. We assess the spatiotemporal variability and quantify 13 performance metrics based on the representation of the climatological mean state and the spatiotemporal variability of temperature, salinity, and volume transports. Using the EN4 dataset and observations, we performed a model ranking by calculating absolute, relative, and total relative errors over each performance metric and model.

The majority of models represent the equilibrium depth of the MOW, the 20-yr mean exchange flow, the net flow, and the spatial variability reasonably well. In addition, the seasonal cycle of the Mediterranean outflow is accurately simulated by most models. Almost all models experience strong temperature biases, and six models have strong salinity biases. We suggest that these stem from one or more of an incorrect simulation of the composition of MOW, the absence of tide simulations that drive the aspiration of Western Mediterranean Deep Waters, the use of different initialization datasets and methods for bias correction in models and EN4, or the lack of entrainment and overflow parameterizations at the SoG in every model. The majority of the models considered here experience problems in correctly representing temporal variability, including standard deviations and the annual amplitude of the Atlantic inflow. The simulation of the barotropic forcing of the Atlantic inflow is not correctly represented in the models because the inflow has a minimum rather than the expected maximum during summer. This may be due to inaccuracy in the representation of the Mediterranean basin evaporative cycle by the coarse-resolution models. In evaluating the total relative errors across all models, we find that MPI-ESM-MR, MPI-ESM-LR, CSIRO Mk3.6.0, and MRI-CGCM3 provide the best representation of the MOW properties. Their total errors range from 0.14 to 0.25, and they simulate 8–10 performance metrics well (0–1 metric poorly). All of the other models considered exhibit biases that negatively affect many performance metrics, for instance, an incorrect representation of the exchange flow. As possible reasons, we again suggest the lack of entrainment and overflow parameterizations and poorly resolved topographies near the sill. However, we note that the MOW is only one of various criteria used for model evaluation. A comparison with other water masses showed consistent warm and salty biases across the models considered; thus, a general model behavior may be also assumed.

For future model development and guidance of subsequent studies of the CMIP6 ensemble, we note that better representation of MOW properties in coarse-resolution models is likely to require common entrainment and overflow parameterizations, higher horizontal resolution, and more accurate SoG bathymetry. In addition, vertical resolution and the choice of vertical coordinate system can be critical in MOW studies: where terrain-following and isopycnal models are relatively insensitive to resolution, low-resolution z-coordinate models are prone to excessive entrainment and spurious mixing in overflow regions. Since most model communities decide to use z coordinates, we recommend increasing their resolution and/or implementing schemes configured to move dense water downslope more efficiently. Moreover, we suggest the establishment of regionally coupled atmosphere–ocean models, or models with unstructured grids allowing a considerably higher resolution over certain target areas like the narrow SoG. Future studies that build on our results, in particular an extension to the next generation CMIP6 models, would prove valuable and are needed before the wider impacts of MOW on the North Atlantic climate system can be reliably determined.

Acknowledgments. Lorine Behr and Elena Xoplaki acknowledge support by the German Federal Ministry of Education and Research (BMBF) project ClimXtreme and the German Academic Exchange Service (DAAD) project EM-MHeatwaves. Xoplaki also acknowledges support by the German Federal Ministry of Education and Research (BMBF) project NUKLEUS, the Academy of Athens and the Greek "National Research Network on Climate Change and its Impact" (project code 200/937). Simon Josey acknowledges support by the U.K. Natural Environment Research Council under the North Atlantic Climate System Integrated Study program (NE/N018044/1). Special thanks are given to Céline Heuzé for downloading and providing one-half of the model output.

Data availability statement. All model outputs are available online from the Earth System Grid Federation (ESGF). The EN4 observational dataset (EN.4.2.1) was obtained online (https://www.metoffice.gov.uk/hadobs/en4/) and is provided under a Non-Commercial Government License (http://www.nationalarchives.gov.uk/doc/non-commercial-government-licence/version/2/).

REFERENCES

Álvarez, M., F. F. Pérez, H. Bryden, and A. F. Ríos, 2004: Physical and biogeochemical transports structure in the North Atlantic subpolar gyre. *J. Geophys. Res.*, **109**, C03027, https://doi.org/10.1029/2003JC002015.

Alvera-Azcárate, A., A. Barth, J.-M. Beckers, and R. H. Weisberg, 2007: Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. *J. Geophys. Res.*, **112**, C03008, https://doi.org/10.1029/2006JC003660.

Ambar, I., and M. R. Howe, 1979: Observations of the Mediterranean outflow—I mixing in the Mediterranean outflow. *Deep-Sea Res.*, **26A**, 535–554, https://doi.org/10.1016/0198-0149(79)90095-5.

Antonov, J. I., and Coauthors, 2010: *Salinity*. Vol. 2, *World Ocean Atlas 2009*, NOAA Atlas NESDIS 69, 184 pp.

Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *General Circulation Models of the Atmosphere*, J. Chang, Ed., Methods in Computational Physics: Advances in Research and Applications, Vol. 17, Elsevier, 173–265, https://doi.org/10.1016/B978-0-12-460817-7.50009-4.

Argo, 2000: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). Sea Scientific Open Data Publication, http://doi.org/10.17882/42182.

Armi, L., and D. M. Farmer, 1986: Maximal two-layer exchange through a contraction with barotropic net flow. *J. Fluid Mech.*, **164**, 27–51, https://doi.org/10.1017/S0022112086002458.

Artale, V., S. Calmanti, P. Malanotte-Rizzoli, G. Pisacane, V. Rupolo, and M. Tsimplis, 2006: The Atlantic and Mediterranean Sea as connected systems. *Mediterranean Climate Variability*, P. Lionello, P. Malanotte-Rizzoli, and R. Boscolo, Eds., Developments in Earth and Environmental Sciences, Vol. 4, Elsevier, 283–323.

Baringer, M. O., 1993: Mixing and dynamics of the Mediterranean outflow. Ph.D. thesis, Massachusetts Institute of Technology, 240 pp.

——, and J. F. Price, 1997: Mixing and spreading of the Mediterranean outflow. *J. Phys. Oceanogr.*, **27**, 1654–1677, https://doi.org/10.1175/1520-0485(1997)027<1654:MASOTM>2.0.CO;2.

Bentsen, M., and Coauthors, 2013: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate. *Geosci. Model Dev.*, **6**, 687–720, https://doi.org/10.5194/gmd-6-687-2013.

Béranger, K., L. Mortier, and M. Crépon, 2005: Seasonal variability of water transport through the Straits of Gibraltar, Sicily and Corsica, derived from a high-resolution model of the Mediterranean circulation. *Prog. Oceanogr.*, **66**, 341–364, https://doi.org/10.1016/j.pocean.2004.07.013.

Bi, D., and Coauthors, 2013: The ACCESS coupled model: Description, control climate and evaluation. *Aust. Meteor. Oceanogr. J.*, **63**, 41–64, https://doi.org/10.22499/2.6301.004.

Bryden, H. L., J. Candela, and T. H. Kinder, 1994: Exchange through the Strait of Gibraltar. *Prog. Oceanogr.*, **33**, 201–248, https://doi.org/10.1016/0079-6611(94)90028-0.

Candela, J., 2001: Mediterranean water and global circulation. *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, International Geophysics Series, Vol. 77, Elsevier, 419–429, https://doi.org/10.1016/S0074-6142(01)80132-7.

Carracedo, L. I., P. C. Pardo, S. Flecha, and F. F. Pérez, 2016: On the Mediterranean water composition. *J. Phys. Oceanogr.*, **46**, 1339–1358, https://doi.org/10.1175/JPO-D-15-0095.1.

Collins, S. N., R. S. James, P. Ray, K. Chen, A. Lassman, and J. Brownlee, 2013: Grids in numerical weather and climate models. *Climate Change and Regional/Local Responses*, P. Ray and Y. Zhang, Eds., InTech, https://doi.org/10.5772/49933.

Danabasoglu, G., S. C. Bates, B. P. Briegleb, S. R. Jayne, M. Jochum, W. G. Large, S. Peacock, and S. G. Yeager, 2012: The CCSM4 ocean component. *J. Climate*, **25**, 1361–1389, https://doi.org/10.1175/JCLI-D-11-00091.1.

de Pascual-Collar, Á., M. G. Sotillo, B. Levier, R. Aznar, P. Lorente, A. Amo-Baladrón, and E. Álvarez-Fanjul, 2019: Regional circulation patterns of Mediterranean Outflow Water near the Iberian and African continental slopes. *Ocean Sci.*, **15**, 565–582, https://doi.org/10.5194/os-15-565-2019.

Dufresne, J.-L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth System Model: From $CMIP_3$ to $CMIP_5$. *Climate Dyn.*, **40**, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1.

Ezer, T., 2005: Entrainment, diapycnal mixing and transport in three-dimensional bottom gravity current simulations using the Mellor–Yamada turbulence scheme. *Ocean Modell.*, **9**, 151–168, https://doi.org/10.1016/j.ocemod.2004.06.001.

——, and G. L. Mellor, 2004: A generalized coordinate ocean model and a comparison of the bottom boundary layer dynamics in terrain-following and in z-level grids. *Ocean Modell.*, **6**, 379–403, https://doi.org/10.1016/S1463-5003(03)00026-X.

Farmer, D. M., and L. Armi, 1988: The flow of Atlantic water through the Strait of Gibraltar. The flow of Mediterranean water through the Strait of Gibraltar: Undefined. *Prog. Oceanogr.*, **21**, 1–103, https://doi.org/10.1016/0079-6611(88)90055-9

Fox-Kemper, B., and Coauthors, 2019: Challenges and prospects in ocean circulation models. *Front. Mar. Sci.*, **6**, 65, https://doi.org/10.3389/fmars.2019.00065.

García-Lafuente, J., A. Sánchez Román, G. Díaz del Río, G. Sannino, and J. C. Sánchez Garrido, 2007: Recent observations of seasonal variability of the Mediterranean outflow in the Strait of Gibraltar. *J. Geophys. Res.*, **112**, C10005, https://doi.org/10.1029/2006JC003992.

——, A. Sánchez-Román, C. Naranjo, and J. C. Sánchez-Garrido, 2011: The very first transformation of the Mediterranean outflow in the Strait of Gibraltar. *J. Geophys. Res.*, **116**, C07010, https://doi.org/10.1029/2011JC006967.

Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans*, **118**, 6704–6716, https://doi.org/10.1002/2013JC009067.

Gordon, H. B., S. P. O'Farrell, M. A. Collier, M. R. Dix, L. D. Rotstayn, E. A. Kowalczyk, A. C. Hirst, and I. G. Watterson, 2010: The CSIRO Mk3.5 climate model. CAWCR Tech. Rep. 021, 74 pp., http://hdl.handle.net/102.100.100/108810?index=1.

Gouretski, V., and F. Reseghetti, 2010: On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep-Sea Res. I*, **57**, 812–833, https://doi.org/10.1016/j.dsr.2010.03.011.

Griffies, S. M., 2004: *Fundamentals of Ocean Climate Models*. Princeton University Press, 518 pp.

Heuzé, C., 2017: North Atlantic deep water formation and AMOC in CMIP5 models. *Ocean Sci.*, **13**, 609–622, https://doi.org/10.5194/os-13-609-2017.

——, K. J. Heywood, D. P. Stevens, and J. K. Ridley, 2013: Southern Ocean bottom water characteristics in CMIP5 models. *Geophys. Res. Lett.*, **40**, 1409–1414, https://doi.org/10.1002/grl.50287.

——, ——, ——, and ——, 2015: Changes in global ocean bottom properties and volume transports in CMIP5 models under climate change scenarios. *J. Climate*, **28**, 2917–2944, https://doi.org/10.1175/JCLI-D-14-00381.1.

Hopkins, T. S., 1999: The thermohaline forcing of the Gibraltar exchange. *J. Mar. Syst.*, **20**, 1–31, https://doi.org/10.1016/S0924-7963(98)00068-2.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res.*, **117**, D05127, https://doi.org/10.1029/2011JD017139.

Jordà, G., and Coauthors, 2017: The Mediterranean Sea heat and mass budgets: Estimates, uncertainties and perspectives. *Prog. Oceanogr.*, **156**, 174–208, https://doi.org/10.1016/j.pocean.2017.07.001.

Jungclaus, J. H., and Coauthors, 2013: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *J. Adv. Model. Earth Syst.*, **5**, 422–446, https://doi.org/10.1002/jame.20023.

Kamworapan, S., and C. Surussavadee, 2019: Evaluation of CMIP5 global climate models for simulating climatological temperature and precipitation for Southeast Asia. *Adv. Meteor.*, **2019**, 1067365, https://doi.org/10.1155/2019/1067365.

Koltermann, K. P., V. Gouretski, and K. Jancke, 2011: Atlantic Ocean. Vol. 3, Hydrographic Atlas of the World Ocean Circulation Experiment (WOCE), International WOCE Project Office, https://doi.org/10.21976/C6RP4Z.

Legg, S., R. W. Hallberg, and J. B. Girton, 2006: Comparison of entrainment in overflows simulated by z-coordinate, isopycnal and non-hydrostatic models. *Ocean Modell.*, **11**, 69–97, https://doi.org/10.1016/j.ocemod.2004.11.006.

——, and Coauthors, 2009: Improving oceanic overflow representation in climate models: The Gravity Current Entrainment Climate Process team. *Bull. Amer. Meteor. Soc.*, **90**, 657–670, https://doi.org/10.1175/2008BAMS2667.1.

Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov, 2009: Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.*, **36**, L07608, https://doi.org/10.1029/2008GL037155.

Marcos, M., and M. N. Tsimplis, 2008: Comparison of results of AOGCMs in the Mediterranean Sea during the 21st century. *J. Geophys. Res.*, **113**, C12028, https://doi.org/10.1029/2008JC004820.

Millot, C., J. Candela, J.-L. Fuda, and Y. Tber, 2006: Large warming and salinification of the Mediterranean outflow due to changes in its composition. *Deep-Sea Res. I*, **53**, 656–666, https://doi.org/10.1016/j.dsr.2005.12.017.

Ochoa, J., and N. A. Bray, 1991: Water mass exchange in the Gulf of Cadiz. *Deep-Sea Res.*, **38A**, S465–S503, https://doi.org/10.1016/S0198-0149(12)80021-5.

Oddo, P., M. Adani, N. Pinardi, C. Fratianni, M. Tonani, and D. Pettenuzzo, 2009: A nested Atlantic-Mediterranean Sea general circulation model for operational forecasting. *Ocean Sci.*, **5**, 461–473, https://doi.org/10.5194/os-5-461-2009.

Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate*, **20**, 4356–4376, https://doi.org/10.1175/JCLI4253.1.

Potter, R. A., and M. S. Lozier, 2004: On the warming and salinification of the Mediterranean outflow waters in the North Atlantic. *Geophys. Res. Lett.*, **31**, L01202, https://doi.org/10.1029/2003GL018161.

Price, J. F., and M. O'Neil Baringer, 1994: Outflows and deep water production by marginal seas. *Prog. Oceanogr.*, **33**, 161–200, https://doi.org/10.1016/0079-6611(94)90027-2.

——, and J. Yang, 1998: Marginal sea overflows for climate simulations. *Ocean Modeling and Parameterization*, E. P. Chassignet and J. Verron, Eds., Springer, 155–170.

——, and Coauthors, 1993: Mediterranean outflow mixing and dynamics. *Science*, **259**, 1277–1282, https://doi.org/10.1126/science.259.5099.1277.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670

Rogerson, M., G. R. Bigg, E. J. Rohling, and J. Ramirez, 2012: Vertical density gradient in the eastern North Atlantic during the last 30,000 years. *Climate Dyn.*, **39**, 589–598, https://doi.org/10.1007/s00382-011-1148-4.

Rupp, D. E., J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote, 2013: Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *J. Geophys. Res. Atmos.*, **118**, 10 884–10 906, https://doi.org/10.1002/jgrd.50843.

Ruti, P. M., and Coauthors, 2016: Med-CORDEX initiative for Mediterranean climate studies. *Bull. Amer. Meteor. Soc.*, **97**, 1187–1208, https://doi.org/10.1175/BAMS-D-14-00176.1.

Sammartino, S., J. G. Lafuente, C. Naranjo, J. C. S. Garrido, R. S. Leal, and A. S. Román, 2015: Ten years of marine current measurements in Espartel Sill, Strait of Gibraltar. *J. Geophys. Res. Oceans*, **120**, 6309–6328, https://doi.org/10.1002/2014JC010674.

Sánchez-Garrido, J. C., G. Sannino, L. Liberti, J. García Lafuente, and L. Pratt, 2011: Numerical modeling of three-dimensional stratified tidal flow over Camarinal Sill, Strait of Gibraltar. *J. Geophys. Res.*, **116**, C12026, https://doi.org/10.1029/2011JC007093.

Sánchez-Leal, R. F., and Coauthors, 2017: The Mediterranean overflow in the Gulf of Cadiz: A rugged journey. *Sci. Adv.*, **3**, eaao0609, https://doi.org/10.1126/sciadv.aao0609.

Sánchez-Román, A., G. Sannino, J. García-Lafuente, A. Carillo, and F. Criado-Aldeanueva, 2009: Transport estimates at the western section of the Strait of Gibraltar: A combined experimental and numerical modeling study. *J. Geophys. Res.*, **114**, C06002, https://doi.org/10.1029/2008JC005023.

Sannino, G., L. Pratt, and A. Carillo, 2009: Hydraulic criticality of the exchange flow through the Strait of Gibraltar. *J. Phys. Oceanogr.*, **39**, 2779–2799, https://doi.org/10.1175/2009JPO4075.1.

Schroeder, K., S. A. Josey, M. Herrmann, L. Grignon, G. P. Gasparini, and H. L. Bryden, 2010: Abrupt warming and salting of the Western Mediterranean Deep Water after 2005: Atmospheric forcings and lateral advection. *J. Geophys. Res.*, **115**, C08029, https://doi.org/10.1029/2009JC005749

Siddall, M., D. A. Smeed, S. Matthiesen, and E. J. Rohling, 2002: Modelling the seasonal cycle of the exchange flow in Bab El Mandab (Red Sea). *Deep-Sea Res. I*, **49**, 1551–1569, https://doi.org/10.1016/S0967-0637(02)00043-2.

Somot, S., F. Sevault, and M. Déqué, 2006: Transient climate change scenario simulation of the Mediterranean Sea for the twenty-first century using a high-resolution ocean circulation

model. *Climate Dyn.*, **27**, 851–879, https://doi.org/10.1007/s00382-006-0167-z.

Soto-Navarro, J., F. Criado-Aldeanueva, J. García-Lafuente, and A. Sánchez-Román, 2010: Estimation of the Atlantic inflow through the Strait of Gibraltar from climatological and in situ data. *J. Geophys. Res.*, **115**, C10023, https://doi.org/10.1029/2010JC006302.

——, S. Somot, F. Sevault, J. Beuvier, F. Criado-Aldeanueva, J. García-Lafuente, and K. Béranger, 2015: Evaluation of regional ocean circulation models for the Mediterranean Sea at the Strait of Gibraltar: Volume transport and thermohaline properties of the outflow. *Climate Dyn.*, **44**, 1277–1292, https://doi.org/10.1007/s00382-014-2179-4.

——, and Coauthors, 2020: Evolution of Mediterranean Sea water properties under climate change scenarios in the Med-CORDEX ensemble. *Climate Dyn.*, **54**, 2135–2165, https://doi.org/10.1007/s00382-019-05105-4.

Sparnocchia, S., N. Pinardi, and E. Demirov, 2003: Multivariate Empirical Orthogonal Function analysis of the upper thermocline structure of the Mediterranean Sea from observations and model simulations. *Ann. Geophys.*, **21**, 167–187, https://doi.org/10.5194/angeo-21-167-2003.

Tang, Y. M., and M. J. Roberts, 2005: The impact of a bottom boundary layer scheme on the North Atlantic Ocean in a global coupled climate model. *J. Phys. Oceanogr.*, **35**, 202–217, https://doi.org/10.1175/JPO-2671.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Thorpe, R. B., and G. R. Bigg, 2000: Modelling the sensitivity of Mediterranean outflow to anthropogenically forced climate change. *Climate Dyn.*, **16**, 355–368, https://doi.org/10.1007/s003820050333.

Timmermans, M.-L. E., and L. J. Pratt, 2005: Two-layer rotating exchange flow between two deep basins: Theory and application to the Strait of Gibraltar. *J. Phys. Oceanogr.*, **35**, 1568–1592, https://doi.org/10.1175/JPO2775.1.

Tonani, M., N. Pinardi, S. Dobricic, I. Pujol, and C. Fratianni, 2008: A high-resolution free-surface model of the Mediterranean Sea. *Ocean Sci.*, **4**, 1–14, https://doi.org/10.5194/os-4-1-2008.

Tsimplis, M. N., and H. L. Bryden, 2000: Estimation of the transports through the Strait of Gibraltar. *Deep-Sea Res. I*, **47**, 2219–2242, https://doi.org/10.1016/S0967-0637(00)00024-8.

——, and Coauthors, 2006: Changes in the oceanography of the Mediterranean Sea and their link to climate variability. *Mediterranean Climate Variability*, P. Lionello, P. Malanotte-Rizzoli, and R. Boscolo, Eds., Developments in Earth and Environmental Sciences, Vol. 4, Elsevier, 227–282.

Tsujino, H., M. Hirabara, H. Nakano, T. Yasuda, T. Motoi, and G. Yamanaka, 2011: Simulating present climate of the global ocean–ice system using the Meteorological Research Institute Community Ocean Model (MRI.COM): Simulation characteristics and variability in the Pacific sector. *J. Oceanogr.*, **67**, 449–479, https://doi.org/10.1007/s10872-011-0050-3.

Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, https://doi.org/10.1256/qj.04.176.

Vlasenko, V., J. C. Sanchez Garrido, N. Stashchuk, J. Garcia Lafuente, and M. Losada, 2009: Three-dimensional evolution of large-amplitude internal waves in the Strait of Gibraltar. *J. Phys. Oceanogr.*, **39**, 2230–2246, https://doi.org/10.1175/2009JPO4007.1.

Voldoire, A., and Coauthors, 2013: The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dyn.*, **40**, 2091–2121, https://doi.org/10.1007/s00382-011-1259-y.

Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.

Wu, W., G. Danabasoglu, and W. G. Large, 2007: On the effects of parameterized Mediterranean overflow on North Atlantic Ocean circulation and climate. *Ocean Modell.*, **19**, 31–52, https://doi.org/10.1016/j.ocemod.2007.06.003.

Xoplaki, E., J. F. González-Rouco, J. Luterbacher, and H. Wanner, 2003: Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs. *Climate Dyn.*, **20**, 723–739, https://doi.org/10.1007/s00382-003-0304-x.

——, ——, ——, and ——, 2004: Wet season Mediterranean precipitation variability: Influence of large-scale dynamics and trends. *Climate Dyn.*, **23**, 63–78, https://doi.org/10.1007/s00382-004-0422-0.

Xue, Y., A. Leetmaa, and M. Ji, 2000: ENSO prediction with Markov models: The impact of sea level. *J. Climate*, **13**, 849–871, https://doi.org/10.1175/1520-0442(2000)013<0849:EPWMMT>2.0.CO;2.

Zenk, W., 1970: On the temperature and salinity structure of the Mediterranean water in the northeast Atlantic. *Deep-Sea Res. Oceanogr. Abstr.*, **17**, 627–631, https://doi.org/10.1016/0011-7471(70)90072-0.