Validation of an ensemble modelling system for climate projections for the northwest European shelf seas

Jonathan Tinker [a,*], Jason Lowe [a], Jason Holt [b], Anne Pardaens [a], Andy Wiltshire [a]

[a] Met Office Hadley Centre, FitzRoy Rd, Exeter, Devon EX1 3PB, UK
[b] National Oceanography Centre, Joseph Proudman Building, 6 Brownlow Street, Liverpool L3 5DA, UK

ARTICLE INFO

ABSTRACT

The aim of this study was to evaluate the performance of a modelling system used to represent the northwest European shelf seas. Variants of the coupled atmosphere–ocean global climate model, HadCM3, were run under conditions of historically varying concentrations of greenhouse gases and other radiatively active constituents. The atmospheric simulation for the shelf sea region and its surrounds was downscaled to finer spatial scales using a regional climate model (HadRM3); these simulations were then used to drive a river routing scheme (TRIP). Together, these provide the atmospheric, oceanic and riverine boundary conditions to drive the shelf seas model POLCOMS. Additionally, a shelf seas simulation was driven by the ERA-40 reanalysis in place of HadCM3. We compared the modelling systems output against a sea surface temperature satellite analysis product, a quality controlled ocean profile dataset and values of volume transport through particular ocean sections from the literature.

In addition to assessing model drift with a pre-industrial control simulation the modelling system was evaluated against observations and the reanalysis driven simulation. We concluded that the modelling system provided an excellent (good) representation of the spatial patterns of temperature (salinity). It provided a good representation of the mean temperature climate, and a sufficient representation of the mean salinity and water column structure climate. The representation of the interannual variability was sufficient, while the overall shelf-wide circulation was qualitatively good. From this wide range of metrics we judged the modelling system fit for the purpose of providing centennial climate projections for the northwest European shelf seas.

1. Introduction

The world's shelf seas are very important to the world economy, ultimately supporting 95% of global fish catches, and they play an important role in the global marine ecosystems. The UK's marine industry (predominantly located on the continental shelf) includes fisheries, oil and gas, shipping, renewable energy, and aggregate extraction (Pugh, 2008). Despite the importance of this region, during the UK's first Climate Change Risk Assessment (CCRA, a statutory requirement of the UK's Climate Change Act, 2008) there was found to be too little known about potential climate change impacts on UK's shelf seas (Pinnegar et al., 2012). Lack of regional scale projections for temperature, salinity and stratification have all been identified as important research areas (MCCIP, 2012).

The effects of climate trends and variability have already been observed on the North West European (NWE) shelf for a wide range of parameters (e.g. Cannaby and Hüsrevoğlu, 2009). There

has been a rapid increase in the north east Atlantic Sea Surface Temperature (SST) around the UK and Ireland between 1983 and 2012 (HadISST; Rayner et al., 2003; Holt et al., 2012b) with the largest increase evident in the southern North Sea and the eastern English Channel, at a rate of 0.4–0.5 °C/decade (Dye et al., 2013b). In addition to these observed trends there is also substantial year-to-year variability. SST variability is greatest in the eastern North Sea, while Near-Bed Temperature (NBT) variability is greatest in the eastern and southern North Sea (Holt et al., 2012b). Salinity on the shelf responds to changes in precipitation, evaporation and river inflow, in addition to effects of the link to open ocean salinity variations. Salinity records are characterised by their long-term variability. The present period is the most saline since the 1950, and has remained stable since 2003 (Dye et al., 2013a).

There is a clear requirement for marine climate projections for the NWE shelf seas (e.g. Pinnegar et al., 2012). Despite the useful skill in global climate model simulation of aspects of large-scale ocean behaviour (e.g. Shuckburgh, 2012), the shelf seas are poorly represented in Atmosphere–Ocean Coupled General Circulation

Models (AOGCM, or GCM). In addition to the poor spatial representation of the geography of shelf seas (e.g. GCMs often omit the English Channel and the Irish Sea) important shelf seas processes are excluded. Perhaps the most important of these is tides, which are often the dominant source of mixing energy on the shelf. One solution is to "dynamically downscale" the ocean component of the GCM, as is routinely done for the atmosphere. A higher resolution, regional shelf sea model is driven by forcings derived from a global model simulation. This approach has been used for a number of studies of the NWE shelf. There has been considerable modelling of the NWE shelf seas (e.g. Skogen et al., 1995; Holt and James, 2001; Proctor et al., 2003; Schrum et al., 2003; Holt and Proctor, 2008; O'Dea et al., 2012; Wakelin et al., 2012; Mathis et al., 2013), with several projections of the future climate (Meier, 2006; Ådlandsvik and Bentsen, 2007; Ådlandsvik, 2008; Holt et al., 2010, 2012a; Friocourt et al., 2012; Olbert et al., 2012; Gröger et al., 2013; Mathis et al., 2013). Most climate modelling studies have focused on individual basin, such as the Baltic Sea (Meier, 2006), the North Sea (Ådlandsvik and Bentsen, 2007; Ådlandsvik, 2008; Friocourt et al., 2012; Mathis et al., 2013) or the Irish Sea (Olbert et al., 2012), whereas here we look at the wider shelf seas region, including the North Sea, Celtic Sea, Irish Sea, English Channel.

In this study we present and evaluate a new ensemble modelling system which will be used in subsequent work to provide climate projections for the NWE shelf seas. The modelling system uses a shelf seas model (POLCOMS) to dynamically downscale the HadCM3 model. The ocean lateral forcings are taken directly from the ocean component of HadCM3 (the ocean boundary is beyond the shelf break giving the POLCOMS control over exchange with the open ocean (e.g. Holt et al., 2010)). The HadCM3 atmosphere over the European region is downscaled with the physically consistent regional climate model, HadRM3, which provides the surface forcing to POLCOMS. The HadRM3 run-off fields provide the riverine forcings via the river routing model TRIP. Thus POLCOMS can receive information about changes in the climate mean and variability from HadCM3 via the ocean, atmosphere or rivers.

We use this modelling system to downscale a Perturbed Physics Ensemble (PPE) designed to allow us to consider uncertainty in shelf sea projections arising from uncertainty and limitations in HadCM3. The coarseness of climate models (such as HadCM3) necessitates the use of (imperfect) parameterisations, which may have some poorly constrained parameters, to represent sub-grid scale processes. The ensemble is designed to sample the range of uncertainty associated with the parameters of the HadCM3 atmosphere. Other potential sources of uncertainty, which are outside the scope of this study, include: the forcing model or shelf sea model structure and grid size; the driving methodology; the effect of unforced climate variability on initial conditions and the future emissions of greenhouse gas. In addition, the regional atmosphere model is not coupled to the shelf seas model, introducing an additional source of uncertainty.

This paper focuses on evaluation of the present day NWE shelf seas, as given by our modelling system. Climate models are generally not initialised from, or constrained by, observations, and so the phases of the natural variability in a historical simulation are not expected to be as observed in the real world. This does not invalidate their use in looking at many of the statistics of natural variability or long-term climate trends. We typically compare 30-year modelled and observed mean (and variance) fields, so as to reduce the effect of natural year-to-year climate variability. Low frequency natural variability, however, will still influence the comparison against observations. To make an assessment of the role of low frequency variability, we run a shelf sea simulation using forcings from a long HadCM3 "control" climate simulation, which has fixed, pre-industrial radiative forcings.

2. Method

Our ensemble of shelf seas simulations is produced by POLCOMS (Proudman Oceanographic Laboratory Coastal Ocean Modelling System), which is driven by consistent atmosphere, ocean and riverine forcings derived from an 11-member PPE. The methodology is based on, but further developed from, that of Holt et al. (2010). Each shelf seas simulation is run as a transient experiment over the 1952–2098 period (here we focus on the historical period). The overall chain of models in our modelling system is as follows (Fig. 1): HadCM3 is run with historically time-varying concentrations of greenhouse gas and other radiatively active constituents; the ocean component of the HadCM3 provides the oceanic boundary forcing; the HadCM3 atmosphere over the NWE shelf sea region is dynamically downscaled with an ensemble of physically consistent variants of the (atmosphere-only) regional climate model HadRM3 (Jones et al., 2004) providing the surface forcing for the shelf seas model; The river inflow to the shelf seas is provided by passing the HadRM3 run-off through the river routing model TRIP (Total Runoff Integrating Pathways; Oki and Sud, 1998; Oki et al., 1999). This chain of models is run for each member of the PPE to span the range of uncertainty of the PPE. This hierarchy of models provides a set of self-consistent forcing with which to simulate the shelf seas, albeit with the limitation of not allowing the shelf seas to feedback to local atmosphere of the wider climate.

2.1. Climate forcings

The HadCM3 model (Gordon et al., 2000; Pope et al., 2000), on which the PPE is based, has been used extensively, both for climate projections, including in the IPCC (Intergovernmental Panel on Climate Change), Third Assessment Report (TAR), and Fourth Assessment Report (AR4) (IPCC, 2001, 2007), and also for investigations of climate variability (e.g. Gregory et al., 2004; Knight et al., 2005). The atmosphere has a horizontal resolution of $2.5° \times 3.75°$ with 19 vertical levels. The ocean has a resolution of $1.25° \times 1.25°$, with 20 depth levels. The PPE that we use was developed by the Quantifying Uncertainty in Model Projections (QUMP) project, and is described in detail in Collins et al. (2011) and Harris et al. (2013). Here we give a brief overview.

The PPE basis model differs from the CMIP3 (third phase of the Coupled Model Intercomparison Project) version of HadCM3 by using flux adjustment. Flux adjustment allows a wide range of parameter perturbations to be explored, while limiting model climate drift under fixed greenhouse gas conditions. Around 30 parameters of the atmosphere component (given in Table 1 in Rougier et al., 2009) were perturbed within expert-specified ranges. A wide range of single and multiple parameter perturbation simulations ($\sim$280) were run to equilibrium, using slab-ocean-model versions and under both fixed-present-day and doubled-CO_2 conditions. The resulting estimates of equilibrium climate sensitivity (a measure of how much the global mean temperature eventually rises with a doubling of CO_2) were divided into 16 equally probable bins and the parameter-set that best validated against observations in each bin was noted. This gave an ensemble of 17 multiple perturbed parameter settings, with inclusion of the unperturbed or standard ensemble member, with a range of equilibrium climate sensitivity ranging from (2.26–5.46 °C, Harris Pers. Comm.). The perturbations from this set were then applied to the fully coupled HadCM3 version for transient policy-relevant
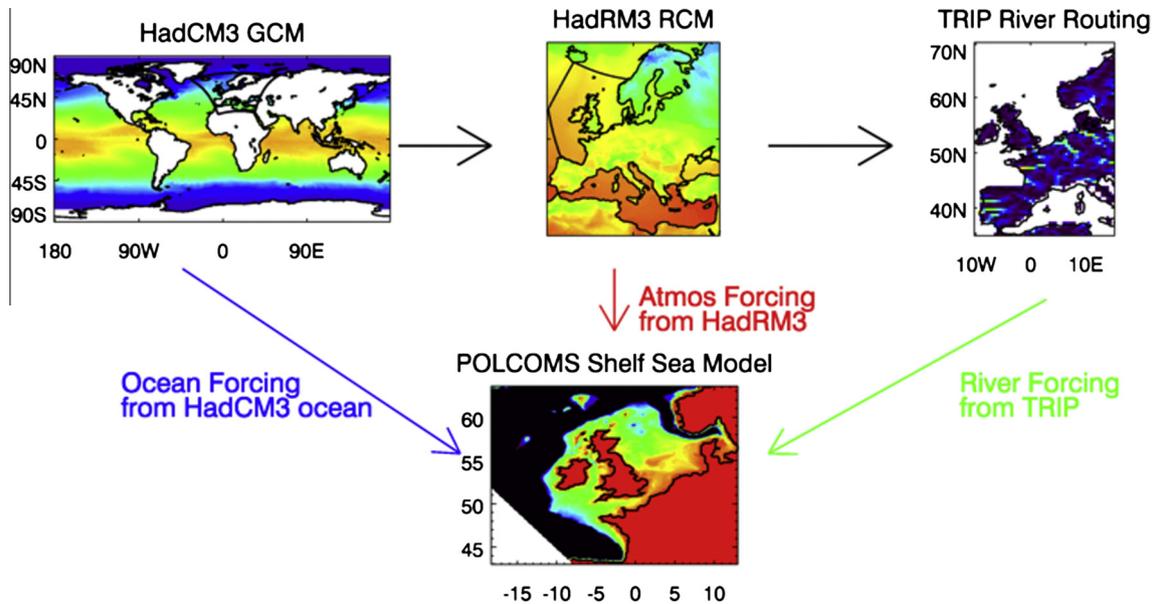
Fig. 1. Chain of models. The global climate model HadCM3 drives the regional climate model HadRM3, which provides forcings for the river routing model TRIP (black arrows). The shelf seas model POLCOMS gets its oceanic forcings from HadCM3 (blue arrow), atmospheric forcings from HadRM3 (red arrow) and riverine forcings from TRIP (green arrow).

historic and future forcings experiments. This ensemble was evaluated by Collins et al. (2011; labelled AO-PPE-A), who found that the ensemble mean was within the observational uncertainty for a number of variables, including temperature (both land-surface air temperature and SST), precipitation and radiation (net top-of-atmosphere flux, outgoing longwave and shortwave fluxes). Overall, the ensemble was found to perform well when compared to a CMIP3 multi-model ensemble.

Although climate sensitivity is an important index to describe of the behaviour of a global climate model, it may not be the most important index for the response of the NWE shelf seas to climate change. We will test this in following studies. Selecting parameters to span the distribution of a different index (e.g. NAO) would lead to a different ensemble of models, which may affect the uncertainty range of the resultant ensemble. For this reason the study should be considered as a preliminary investigation of uncertainty in simulated shelf sea behaviour due to the uncertainty in the driving global and regional climate models. Considerable further work is needed in future to understand the optimum approach to sampling the uncertainty associated with the driving climate models. Our contribution is to use a single sampling approach based on available GCM simulations. Production of further large ensembles with alternative sampling approaches will be computationally expensive and may best be organised as a community activity.

HadRM3, which downscales the PPE-simulated European atmosphere, is a high-resolution version of the atmospheric component of HadCM3, with a horizontal resolution of 0.22° × 0.22° (~25 km), and 19 vertical levels. The domain has been set up on a rotated pole to give near square grid boxes over Europe. It is forced at its lateral and surface boundary by the global PPE simulations. HadRM3 has been widely used for the European region (e.g. Déqué et al., 2007; Murphy et al., 2009) and other regions around the world (e.g. Bhaskaran et al., 2011). The atmosphere of each member of the PPE was dynamically downscaled with HadRM3 using appropriate parameter settings (which may be adapted for grid size). Six ensemble members with a common parameter setting did not validate well against atmospheric and surface observations and were excluded, leaving 11 members available for surface forcing. The rejected ensemble members are not thought to bias the

regional PPE, as they were distributed fairly evenly through the range of climate sensitivities. We consider each ensemble member to be equally likely, although future analysis may lead to weightings for the ensemble members. A number of studies have evaluated and used aspects of this 11-member ensemble, including extreme temperature (Brown et al., 2014), precipitation (Sanderson et al., 2012) and large-scale European atmospheric circulation (Clark and Brown, 2013).

The river routing TRIP model (Oki and Sud, 1998; Oki et al., 1999) post-processes the surface and subsurface runoff output from HadRM3 and, using a pre-defined half degree river network (Simulated Topological Network at 30-min spatial resolution (STN-30 version 6.01) (Vorosmarty et al., 2000a, 2000b; Fekete et al., 2001), advects moisture downstream. This approach ensures the river flow forcing is consistent with the atmospheric forcing. Previous work, which assessed UK river runoff from the same HadRM3 ensemble (not routed through TRIP), found that this was in good agreement with observations, capturing the magnitude and monthly variations in flow (Sanderson et al., 2012).

2.2. Shelf seas model: POLCOMS

POLCOMS (Holt and James, 2001) is a primitive equation, three-dimensional, baroclinic, finite-difference coastal ocean model. The configuration used here has a horizontal resolution of 1/9° latitude by 1/6° longitude (~12 km) with 34 vertical levels. The domain (Fig. 2) extends from 43°N to 63°33′20″N and 18°20′W to 13°E, however, the south west corner (from 51°53′20″N, 18°20W to 43°26′40″N, 8°20′W) is outside the domain of the regional atmospheric model (HadRM3 on a rotated grid) and so is excluded (e.g. Fig. 1).

The POLCOMS implementation used was based on that of Holt et al. (2010), but was adapted for our purposes in a number of ways. The oceanic, open boundary conditions were taken directly from the ocean component of the PPE. This has a number of advantages over climatological forcing. The notable differences between the behaviour of the coarse resolution ocean component of HadCM3 and POLCOMS, however, led to some complications. The ocean boundary forcings and initial conditions were interpolated
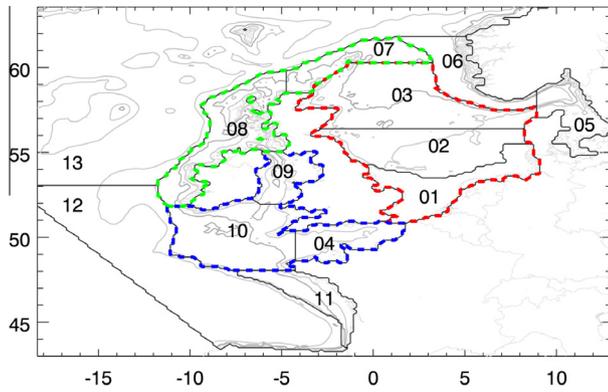
Fig. 2. The complete model domain showing validation regions (red dotted line: North Sea; green dotted line: Outer Shelf region; blue dotted line: Celtic Seas region), which are further broken down into smaller analysis regions (dark grey lines): 01 Southern North Sea; 02 Central North Sea; 03 Northern North Sea; 04 English Channel; 05 Skagerrak/Kattegat; 06 Norwegian Trench; 07 Shetland Shelf; 08 Irish Shelf; 09 Irish Sea; 10 Celtic Sea; 11 Armorican Shelf; 12 NE Atlantic (S); 13 NE Atlantic (N).

with a Gaussian interpolation to the POLCOMS horizontal grid and then linearly interpolated from the HadCM3 depth levels to the POLCOMS s-levels. The method used to relax the boundary forcings into the model domain was altered (both the width and weightings of the relaxation zone) to reduce the numerical instabilities introduced by sharp transitions between the HadCM3 and POLCOMS density fields.

The baroclinic ocean forcings of HadCM3 introduced a significant unrealistic, persistent eddy within the truncated Bay of Biscay. This was reduced by introducing a secondary 'sponge layer' (south of 48°33′20″N, and within the Porcupine Seabight) where the horizontal diffusion (of both the tracers and velocities) is doubled (it is increased by an order of magnitude in the main sponge layer). This was shown to have little effect on the circulation outside the region.

The model outputs were recorded monthly, rather than daily, and were de-tided with a Doodson tidal filter (e.g. Pugh, 1987) rather than with a simple 25 h mean. Fifteen tidal constituents, taken from a North Atlantic tidal model (Flather, 1976) were also used as boundary conditions. The atmospheric forcings from the unperturbed ensemble member were similar to those used in the study of Holt et al. (2010).

A region mask for POLCOMS (Holt et al., 2012a) is used to aid analysis (Fig. 2). The mask is further simplified into 3 regions: the 'North Sea' (combining the northern, central and southern North Sea regions from Fig. 2); the 'Celtic Seas' (combining the Celtic Sea, the Irish Sea and the English Channel); the 'Outer Shelf' region (combining the Irish Shelf and the Shetland Shelf). These regions are combined into the 'shelf' region for the purpose of validation (this excludes the Kattegat, Skagerrak and the Norwegian Trench).

2.3. Experimental design

The full 11-member regional downscaled atmospheric PPE ensemble was used to drive POLCOMS, together with associated oceanic boundary conditions and river inflow. Each shelf sea simulation was downscaled as a transient run for the 1952–2098 period by following the modelling chain described above (Fig. 1). The 1952–1960 period is considered spin-up. The unperturbed model simulation is referred to as ens_00, the perturbed members as ens_xx where xx range from 01 to 10. To provide an estimate of unforced climate variability of the NWE shelf seas in the absence

of historical changes in radiative constituents, a 146-year pre-industrial control run of a single HadCM3 PPE variant (the unperturbed member) was downscaled with POLCOMS (ens_ctrl). However this was run with atmospheric and (TRIP routed) riverine forcings from HadCM3. The use of lower spatial resolution surface fluxes to drive POLCOMS is common (Holt et al., 2012a, 2014) and sensitivity test repeating ens_00 with global forcings gave comparable results justifying this approach in the absence of a downscaled atmosphere.

In order to validate the model set-up, a re-analysis forced simulation was run. ERA-40 reanalysis data (Uppala et al., 2005) was used to force the regional climate model HadRM3, which then provided POLCOMS and TRIP atmospheric forcings. As data assimilation was not used there was a risk that HadRM3 will deviate from the ERA40 reanalysis, although this was minimised by carefully selecting the regional model domain (Jones et al., 1995, 1997). Observationally constrained oceanic forcings for this simulation were taken from Holt et al. (2010) and consist of a 30-year climatology created by driving a large area POLCOMS domain with the Forecasting Ocean Assimilation Model (FOAM; Bell et al., 2000). Refer to Holt et al. (2010) for further details. This model simulation is referred to as ERA_sim.

2.4. Observational datasets and analysis techniques

The observational OSTIA (Operational Sea-Surface Temperature and Sea-Ice Analysis) SST analysis product (Roberts-Jones et al., 2012) provides a predominantly satellite based dataset. This analysis combines different bias corrected satellite products to reduce the bias of the overall product. Monthly maps are used between 1986 and 2006, allowing the climatological monthly mean and standard deviation to be calculated directly. The OSTIA data was of a slightly lower resolution than the POLCOMS model grid and so was bi-linearly interpolated onto the model grid. The data was also slightly smoother than the resolution suggests, due to the correlation length-scales used in the OSTIA reanalysis system.

The quality-controlled *in-situ* EN3 temperature and salinity profile dataset (Ingleby and Huddleston, 2007) (between 1960 and 2005) was also used, to validate the simulated monthly mean fields of Sea-Surface and Near-Bed Temperature and Salinity (SST, SSS, NBT, NBS), as well as integrated water-column structure variables, including Potential Energy Anomaly (PEA) and Mixed Layer Depth (MLD). The EN3 data set was insufficient to assess the temporal performance of the model. The data were averaged over the years to produce a mean monthly climatology, although this was still affected by the limitations in temporal and spatial sampling, so it can only be used as a guide. Each observed profile within the model domain was assigned to the nearest grid box, with these then averaged for a given month and year. The dataset was then averaged over years to give a monthly mean, a standard deviation and number of profiles (in the each grid-box). For a given month, grid boxes with profiles from less than 4 years were excluded from the multi-annual statistics used in the later evaluation. This reduced the number of grid-boxes with EN3 observations to 11,500, approximately 12% of the available data (as shown in Fig. 5).

The Mixed Layer Depth (MLD), which typically is the depth of the pycnocline, was calculated as the depth of water where the density matches that of the equivalent density of the reference depth (6 m) if the temperature reduced by 0.5 °C (Wakelin et al., 2009).

$$MLD = z(\rho(T,S) = \rho(T_{6m} - 0.5\,°C, S))$$

In order to quantify stratification, we used the Potential Energy Anomaly (PEA, Simpson and Bowers, 1981) which is equivalent to the amount of energy required to fully mix a stratified water column. This is defined as:

$$PEA = -\frac{g}{h} \int_{z=-h}^{0} z(\rho(T,S) - \rho(\overline{T},\overline{S}))dz$$

where h is the depth of water (limited to 400 m), g is gravity, z the vertical coordinate (positive upwards), and an over bar represents depth average.

In the absence of spatially coherent, gridded current observations, we turned to observed volume transport cross-sections to assess the shelf-wide circulation. To quantify these current strengths, post-processed detided net-volume transport rates were calculated through the cross-sections presented in Lowe et al. (2009) (validating Holt et al. (2010); see Fig. 10), located close to observed estimates (Svendsen et al., 1991; Turrell et al., 1992; Prandle et al., 1996; Danielssen et al., 1997; Brown et al., 1999, 2003; Holt et al., 2001; Fernand et al., 2006). This however is a semi-qualitative test, as there are very few observations and those available are for particular years, but is useful as an illustrative tool to validate the model in the context of the broader scale circulation.

3. Statistical evaluation framework

The climate system has a wide range of temporal variability, including externally forced variability (e.g. the climatic response to changes in the external forcings such as solar or volcanism), and internally generated (unforced) variability.

The climate system has many components that respond on different timescales, with non-linear interactions, and so the system is rarely in equilibrium (IPCC, 2001) – this leads to unforced Natural Variability (NV) such as the El Niño-Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO). Additionally the climate is subject to both natural external forcings such as volcanic eruptions and changes in solar forcings, and anthropogenic forcings from greenhouse gases (GHGs) emissions. Thus the climate expresses variability over a wide range of frequencies, from high frequency events, such as individual storms, to low frequency events, such as the NAO and Atlantic Multidecadal Oscillation (AMO). When looking at a period (e.g. 30 years) of the climate (a sample), low-frequency variability (e.g. with a period greater than 60 years) can contaminate the sample mean, introducing an apparent bias compared to the underlying true (population) mean.

Climate models are designed to capture the behaviour of this climate variability: by representing their underlying mechanisms, they simulate the amplitude, period and spatial patterns of many modes of variability. However, except when set-up for seasonal or decadal forecasts (which is not the case here), they do not aim to reproduce the observed phase of the variability. Therefore directly comparing a climate model to observations year by year is not meaningful, as any apparent model-observation differences will very likely be contaminated by differences between the modelled and observed phase of the high- (and low-) frequency climate variability. A common approach is therefore to compare the model and observed climate statistics over a period (often 30 years). However, this is still contaminated by the low-frequency variability.

A pre-industrial climate simulation makes it possible to assess the distribution of 30-year climate mean due to unforced NV. This can give us bounds to how much a modelled 30-year mean can vary simply due to the phase of the NV. The real world is a single realisation from a theoretical distribution of real-world 30-year climate means associated with NV. However, as the climate is not stationary, we cannot use the recent past to estimate the spread of this distribution, and we have no grounds to assume that it has the same width as the modelled distribution or will remain the same in the future. Furthermore, we do not know whether the observed, single realisation is above or below the mean of the distribution,

or whether it is near the centre, or in the tails of the distribution. Thus, assuming the single realisation of the observations are not from a tail of the observations distribution, if the single realisation of the observations fit within the modelled distribution, we know that the modelled and observed distributions overlap. If the single realisation of the observations does not fit within the modelled distribution, we know that the two distributions are not identical, but we do not know if the distributions overlap. Therefore apparent discrepancies between the modelled distribution and the observation sample could be due to the relative phases of the NV, or due to the model-observation bias. We therefore do not ask how well the modelling system validates against the observations. Instead, we ask whether the observations are consistent with the modelled ensemble. We have insufficient information to separate any inconsistencies into model-observation bias from differences in the phase of NV. However, as ERA_sim has the same phase of variability as the observations, and perhaps similar downscaling model bias as ens_xx, it provides additional information to the problem.

We consider the modelled-, real- and reanalysis-world to be the combination of a low- and high-frequency variability components and the climate mean state, and calculate each component separately. For the ensemble, we are able to estimate each of these components. Two estimates of the low frequency variability can be made from the pre-industrial control simulation, and from the ensemble spread, each with its own complications. The modelled mean climate can be estimated from the ensemble of 30-year means, and the modelled high-frequency variability can be estimated from the near present day period of the simulations. The observed and reanalysis high-frequency variability can also be estimated from the respective near-present day periods, but no estimate can be made of the distribution of the low-frequency variability or the mean climate state.

Therefore we use the following statistical framework.

1. Compare the two estimates of the modelled low-frequency variability. The control simulation is relatively short (containing 5 samples) and is forced by a non-downscaled atmosphere component. The ensemble spread contains both NV and model parameter uncertainty, which may be important in the present day. These two variance estimates will be compared with an f-test.
2. Assess whether the single observed 30-year mean is consistent with the modelled distribution of 30-year means. We fit a Gaussian distribution to 30-year-means of the 11 ensemble members and see where the observations (and ERA_sim) fit within the associated cumulative probability distribution. Differences may be due to model (or observational) bias, phase of NV, or the differing forcings lead to different mean statistics.
3. Assess whether the single reanalysis (ERA_sim) 30-year mean is consistent with the modelled distribution of 30-year means. Differences are due to model-observation bias and differences in the modelled and reanalysis phase of low-frequency, however, as the reanalysis and observed NV should have the same phase, differences between test 2 and 3 give information to help differentiate between the model bias and NV.
4. Compare the modelled and observed high-frequency variance (inter-annual variability) with an f-test. The ensemble interannual high-frequency variability is calculated by removing the 30-year mean from present-day of each respective ensemble member, and then concatenating the resultant anomalies along the time dimension.
5. Compare the modelled and reanalysis high-frequency variance (inter-annual variability) with an f-test.

In addition to assessing test 2 and 3 individually in a spatial context (looking at maps of the position within the modelled
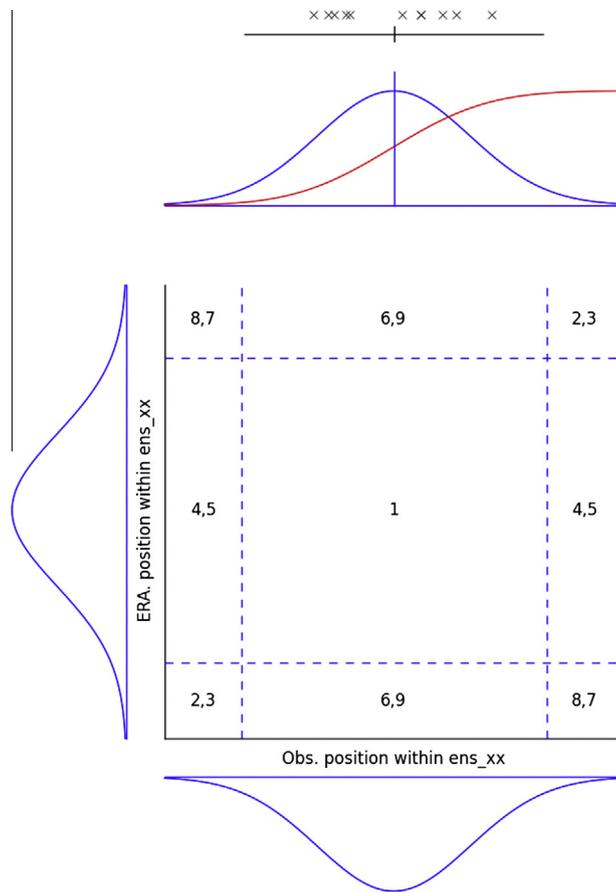
Fig. 3. Implications of the position of the observations and ERA_sim estimates within the modelled distribution of a particular variable. The x- and y-axes represents the position of the observations and ERA_sim within the modelled distribution respectively, with the three columns/rows representing the left tail ($p < 0.05$), middle of the distribution ($0.05 < p < 0.95$) and the right tail ($p > 0.95$). When the observations fall in the left tail of the modelled distribution, we look to the left column, and when the ERA_sim realisation falls within the left tail of the modelled distribution, we look to the bottom row. In such a case both the observations and ERA_sim realisation is below the modelled distribution, and possible explanations for this is given by points 2 and 3 in Table 1: There is a difference in NV between the observations and the modelled simulation, or there is a bias in the climatological forcings.

distribution), they are interpreted in concert (plotting the percentile values of ERA against observations for each co-located point). This shows how much agreement there is in the spatial patterns of how the observations and ERA_sim realisation fit within the modelled distribution. For example, this shows how much data there are where both the observations and ERA_sim are both greater than the modelled distribution, compared to the observations being greater, and ERA_sim being less, than the modelled. There are a number of causes (Table 1) which would lead to such scenarios, and these are illuminated in Fig. 3.

By assessing the variables, seasons and regions where the observations and ERA_sim fall within the modelled distribution, we can start draw some conclusions on the role of model/observation/reanalysis bias and NV. We, like most climate studies (e.g. Flato et al., 2013), have insufficient information to thoroughly validate the system in the way that one might validate a near-term forecasting system; however, we have adopted this approach to make as thorough an assessment as possible.

We use this approach to assess the ensemble against the OSTIA SST, EN3 NBT, SSS, NBS, MLD and PEA. We assess all months, but only show February and July as exemplar winter and summer months with good data coverage (only July for MLD and PEA). This approach is not used to assess the volume transport through the cross-sections. These currents are often topographically constrained and so climate variability is likely to be of secondary importance. Instead we assess their seasonal cycle against the observations.

4. Results

Here we describe the results of the model evaluation. We summarise the key findings in Table 5.

4.1. Assessment of model drift and low-frequency variability from a pre-industrial control simulation

The dynamically-downscaled NWE shelf in ens_ctrl allows us to both assess drift that may arise from the driving climate model, and also to assess the modelled unforced internal climate variability. This is one of the first published pre-industrial control runs downscaled for the NWE shelf, and so provides an important dataset. Previous studies have assessed the variability of the shelf seas from observations and from models driven by atmospheric reanalyses (Holt et al., 2012b). However, these studies are comparatively short (∼40 years) and implicitly include recent climate change signals. As we have noted, this pre-industrial shelf seas simulation is forced directly by relatively coarse resolution atmospheric forcing data, rather than being downscaled prior to use. It should be noted that this might be expected to affect, and possibly limit the shelf seas variability represented.

For most of the shelf regions analysed (Fig. 2), the annual mean SST and NBT were found to have statistically insignificant absolute linear trends of less than 0.01 °C/decade (all regions, including the open ocean had absolute linear trends less than 0.025 °C/decade), when tested with a regression analysis to the 5% level, and modelling the errors as an AR(1) process. SSS and NBS also have statistically insignificant (at the 5% level) absolute linear trends of less than 0.02 psu/decade for all shelf regions except the English Channel which has a statistically significant SSS increases at 0.0196 psu/decade (±1.96 S.E.).

The annual mean range of values on the shelf for ens_ctrl is given in Table 2). The 50th percentile SST for the whole shelf is 10.96 °C but ranges from 10.43 to 11.56 °C from the 5th to the 95th percentile. The NBT show a similar range, with 90% of the data

Table 1
Possible interpretations of relative observation and ERA_sim 30-year mean position within the ensemble modelled distribution.

1. Observations and ERA_sim are consistent with the modelled distribution
2. There is a different phase of NV between the reality and the modelled ensemble
3. There is a bias in the climate model forcings
4. There is a model bias associated with the shelf seas downscaling technique (common to both ens_xx and ERA_sim)
5. There is an observations bias
6. There is a bias in the ERA-40 forcings
There are also combinations between the biases
7. POLCOMS model bias (+ve/−ve) and ERA bias (+ve/−ve) respectively
8. Observation bias (+ve/−ve) and ERA bias (−ve/+ve) respectively
9. Model bias (+ve/−ve) and observation bias (+ve/−ve)

Table 2
Distribution of annual- and shelf-mean surface and bed, temperature and salinity from the downscaled pre-industrial control simulation, ens_ctrl, as represented by the 5%, 20%, 50%, 80% and 95% percentile values.

Annual shelf mean	5% percentile	20% percentile	50% percentile	80% percentile	95% percentile
SST:	10.43	10.68	10.96	11.36	11.58
SSS:	34.29	34.42	34.55	34.68	34.81
NBT:	8.79	9.03	9.34	9.64	9.89
NBS:	34.41	34.53	34.64	34.76	34.88

lying within 1.1 °C. Surface and bed salinity have a range of ~0.5 psu.

In a stable climate, as represented by a control run with little drift, the longer-term warming approaches zero. Shorter periods of warming (and cooling), however, are expected under the influence of unforced variability. Understanding the range of these short-period trends puts any analysis of the influence of greenhouse gas increases on similarly short-period observed or projected trends into context. The distribution of linear trends over n years (where $2 > n > 70$ yrs), was calculated (Fig. 4; Table 3; note
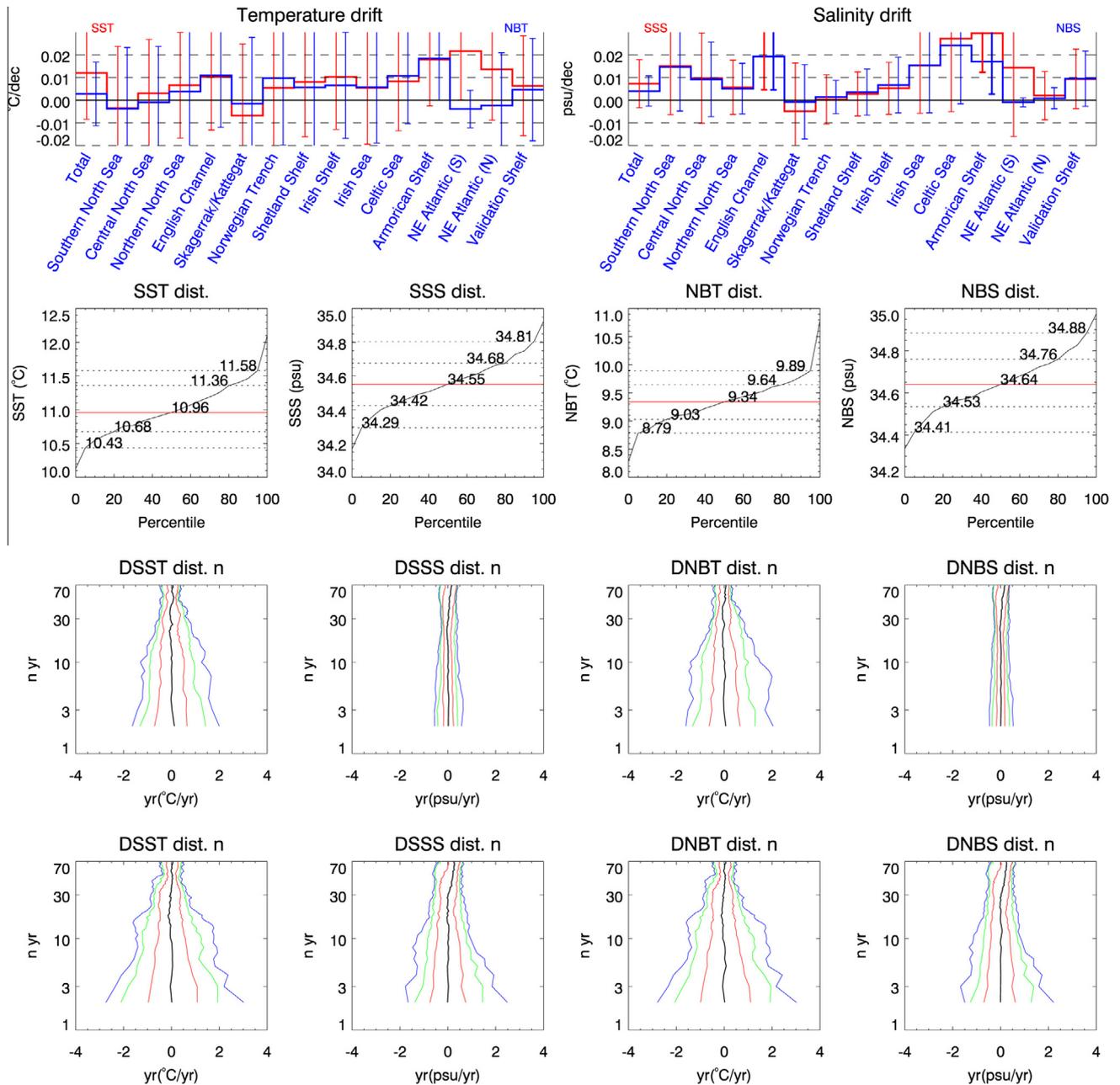

Fig. 4. Drift and variability from the pre-industrial control run. The four columns represent the Sea-Surface Temperature (SST), Sea-Surface Salinity (SSS), Near-Bed Temperature (NBT), Near-Bed Salinity (NBS) respectively. Upper row shows the area-averaged, linear drift in the annual mean temperature (left) and salinity (right) over the full 146-year control run, with error bars showing ±1.96 SE (bold when the trend is statistically significant at the 5% level). The second row shows the distribution of the annual mean SST (and SSS, NBT, SSS), with horizontal lines highlighting the 5th, 20th, 50th, 80th and 95th percentile values. The 3rd row shows the scales of variability over the entire shelf. The linear trend for n years (y axis) were calculated with a moving window through the 146 years, and the distribution calculated. This was repeated for $2 \geqslant n > 70$. Here we present the 1st, 5th, 20th, 50th, 80th, 95th and 99th percentile trend multiplied by n (as the blue, green, red, black, red, green and blue lines respectively) against n, on the y axis. This is repeated in the fourth row for the exemplar southern North Sea region, to illustrate the variability of the most variable regions.

Table 3
Range of sustained linear warming of *n* years in the pre-industrial control simulation. The 5th and 95th percentile values for the southern North Sea and the shelf validation region (excluding the Norwegian Trench and the Kattegat/Skagerrak). The southern North Sea is included as an exemplar region as it shows the greatest variability over the 4 variables than other regions.

Region	*n*	dSST/yr	dSSS/yr	dNBT/yr	dNBS/yr
Southern North Sea	2	−1.05: 0.97	−0.69: 0.72	−1.04: 0.96	−0.63: 0.64
Southern North Sea	5	−0.29: 0.27	−0.19: 0.25	−0.28: 0.27	−0.17: 0.22
Southern North Sea	10	−0.12: 0.13	−0.08: 0.08	−0.12: 0.13	−0.07: 0.07
Southern North Sea	30	−0.02: 0.02	−0.02: 0.02	−0.02: 0.02	−0.02: 0.02
Validation Shelf	2	−0.66: 0.72	−0.21: 0.21	−0.66: 0.64	−0.18: 0.18
Validation Shelf	5	−0.19: 0.20	−0.07: 0.07	−0.20: 0.21	−0.06: 0.06
Validation Shelf	10	−0.09: 0.09	−0.03: 0.03	−0.09: 0.10	−0.03: 0.03
Validation Shelf	30	−0.02: 0.02	−0.01: 0.01	−0.01: 0.02	−0.01: 0.01

that the number of samples decreases with increasing value of *n*). The analysis suggest that a warming or cooling of the entire shelf of 0.35 °C/yr over a 2 year period is not uncommon (the 80th (95th) percentile at 2 years is 0.35 °C/yr (0.72 °C/yr)), whereas over even a 5 year period, a sustained warming rate greater than 0.2 °C/yr is rarer (the 95 percentile is 0.2 °C/yr). When looking over longer periods, the likely rates of sustained change rapidly decrease, so the 95th percentiles for shelf wide SST warming for 10 and 30 years are 0.09 °C/yr and 0.02 °C/yr respectively.

Certain regions are more variable than others, while larger regions tend to be less variable, as they are able to average the small scale variability out. The southern North Sea is identified as a particularly variable region for the four fields considered here (surface and bed, temperature and salinity), and so is included in Fig. 4 and Table 4 to compare to the shelf wide value. For example, while the 95th percentile of 5-year linear trend of SSS change for the entire shelf is 0.07 psu/yr, for the southern North Sea it is 0.25 psu/yr.

4.2. Qualitative assessment of ERA_sim, ens_00 with observations

Overall, both ERA_sim and ens_00 capture the main features of the mean state of NWE shelf seas (Fig. 5a). In the winter, the observed (OSTIA) and modelled SST field clearly show the shelf break current. The warmest temperatures are in the south-west of the domain (in the Bay of Biscay), with the coolest temperatures in the Skagerrak/Kattegat and the eastern North Sea. In the summer, the modelled SST field reproduce (OSTIA) observed features such as the warm temperatures along the southern North Sea and relatively cool regions around north-east Scotland and Brittany. The ERA_sim SST field tends to be too warm (relative to observations), in the summer, while the ens_00 has a tendency to be too cool in the winter, and too warm in the summer. The modelled winter NBT field is spatially similar to the EN3 observations, although the ens_00 tends to be too cold. The summer NBT field (June–August) tends too be warm in the mixed regions, and too cool in the stratified regions, as a result of the modelled MLD begin too shallow. In the summer, when much of the shelf is stratified, the pattern of the NBT is dominated by the stratified/mixed regions. The model reproduce many of the observed features, including the warm (mixed) regions in the southern North Sea, Dogger Bank, English Channel and even east of the Isle of Man, although this is greater in the model. The EN3 NBT field shows a cooler plume emanating from the region of the Elbe river mouth, which is reproduced by the ens_00, but is absent in ERA_sim.

The salinity field modelled by ERA_sim and ens_00 also captures most of the key features of the EN3 observations. The models pick up the observed lower salinities in the coastal regions, with the higher salinity in the centre of the North Sea. The ERA_sim tends to be slightly more saline than the observations, while ens_00 tends to be fresher. The surface salinity observations in

the Norwegian Trench show a large region of low salinity water, associated with the Baltic outflow, which is much weaker in the modelled simulations. The EN3 observed salinities are also very low in the very eastern North Sea, from the German Bight (extending up the Norwegian Coast, including the Skagerrak/Kattegat) and this is reflected in the ens_00 and, to a lesser extent, in ERA_sim (which has higher salinities in these regions). The models tend to reproduce the observed (both SSS and NBS) haline intrusion from the English Channel into the southern North Sea (absent in the ens_00 summer). The NBS tend to be similar to the SSS, with the exception of the Norwegian Trench which is salinity stratified throughout the year, with NBS influenced by salty North Atlantic water, rather than fresh Baltic outflow – the models better represent the Norwegian Trench NBS than SSS.

4.3. Evaluation of modelled spatial patterns

Here we compare the modelled spatial patterns to the observations. These spatial patterns tend to be driven by internal processes (include geography, bathymetry and tides) and so are similar across the ensemble. We therefore compare the ERA_sim, and an exemplar ensemble member (ens_00) to the observations using Taylor diagrams, before considering the wider ensemble.

Overall the Taylor diagrams show that the ERA_sim and the ens_xx ensemble represent the spatial patterns of the observed temperature very well. The salinity patterns are also well represented by ERA_sim and the ens_xx ensemble, although the amplitude of the ERA_sim salinity pattern is weaker than in the observations. These results suggest the modelling systems is capturing the underlying processes and dynamics of the system, supporting their use as the basis of a set of climate projections.

The first Taylor diagram (Fig. 6a) shows that the spatial patterns of the shelf-wide ERA_sim SST tend to validate particularly well. There is an observed large-scale temperature pattern, with higher temperatures to the south, and cooler temperatures to the north (Fig. 5a), which the model captures (*r* = 0.99 for OSTIA annual mean SST). Validation regions with dominant large-scale temperature patterns (e.g. the north–south aligned Outer Shelf region, and the Celtic Seas region with a strong temperature gradient between the Celtic Sea in the SW and the Irish Sea in the NE) perform well (OSTIA *r* = 0.97, Celtic Seas and Outer Shelf). The North Sea is slightly disadvantaged compared to the other regions as the large-scale temperature pattern is weaker. Most of the region (75%) is of a very similar temperature (within 0.8 °C), and so the spatial pattern associated with the bathymetry, circulation and stratification is more important (Fig. 5a). These factors are reflected in the slightly lower correlation (North Sea annual mean SST: OSTIA *r* = 0.93). The model tends to validate better against OSTIA than the EN3 dataset, largely due to paucity of data in the EN3 observations.

Near-bed temperatures (Fig. 5b) tend to validate fairly well, with *r* > 0.8 for all regions. During the stratified months the NBT

develops a complex spatial pattern (Fig. 5b) with fronts delineating the mixed regions, where NBT matches the SST, from the stratified regions, where the summer NBT vary little from their winter values. This pattern, reflected in the annual means, is fairly well captured by the model, and so leads to strong correlations, $r > 0.8$, for all regions. The model has a tendency to express slightly too much variability, particularly in the Outer Shelf region, and in the summer.

The large-scale salinity pattern is very different from the temperature pattern (Fig. 5a) and is largely controlled by internal dynamics, and the location and strength of the freshwater inputs (rivers and the Baltic) rather than large-scale atmospheric or even oceanic patterns. Thus the main pattern is effectively a function of the distance from, and strength of, the river and Baltic outflows, and the shelf circulation. Therefore matching the model and observed salinity patterns is much more difficult, as reflected by

the lower correlations (r) and increased relative root mean square ($rrms$). The modelled Dooley Current in the North Sea separates the higher salinity oceanic water from the lower salinity shelf water, but this is less clear in the observations. Furthermore, the model does not capture the low Norwegian Trench SSS (Fig. 5a, mainly outside the validation region, but also along the eastern edge of the North Sea EN3 observations). This leads to a different large-scale pattern of surface salinity for the North Sea compared to that observed by EN3. This difference is not only reflected in the North Sea correlation ($r = 0.65$), but also in the shelf wide correlation ($r = 0.75$). Both the Outer Shelf region and the Celtic Seas region show good correlation reflecting the good spatial agreement ($r = 0.76$, 0.86 respectively). The spatial pattern of the model has a smaller amplitude than the observations, as reflected by the relative standard deviation ($0.4 < rsd < 0.65$). This is particularly apparent in the coastal freshening along the western coast of northern
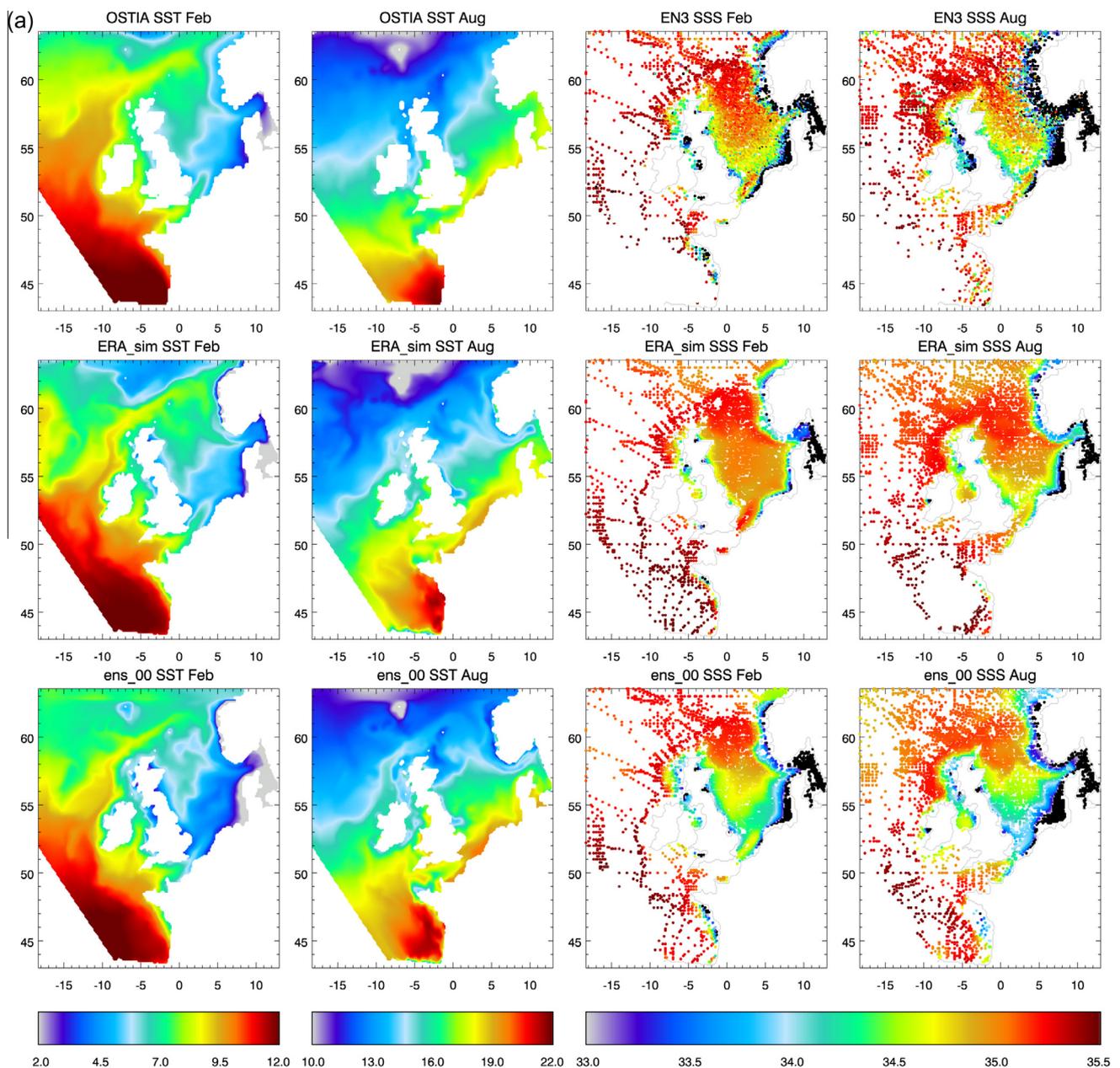

Fig. 5. (a) Present day mean climate. February and August SST and SSS is compared across observations (OSTIA SST 1986–2001, EN3 SSS 1960–2004, upper row), ERA_sim (2nd row), ens_00 (3rd row) and (b) same as (a) but for EN3 NBT and EN3 NBS.
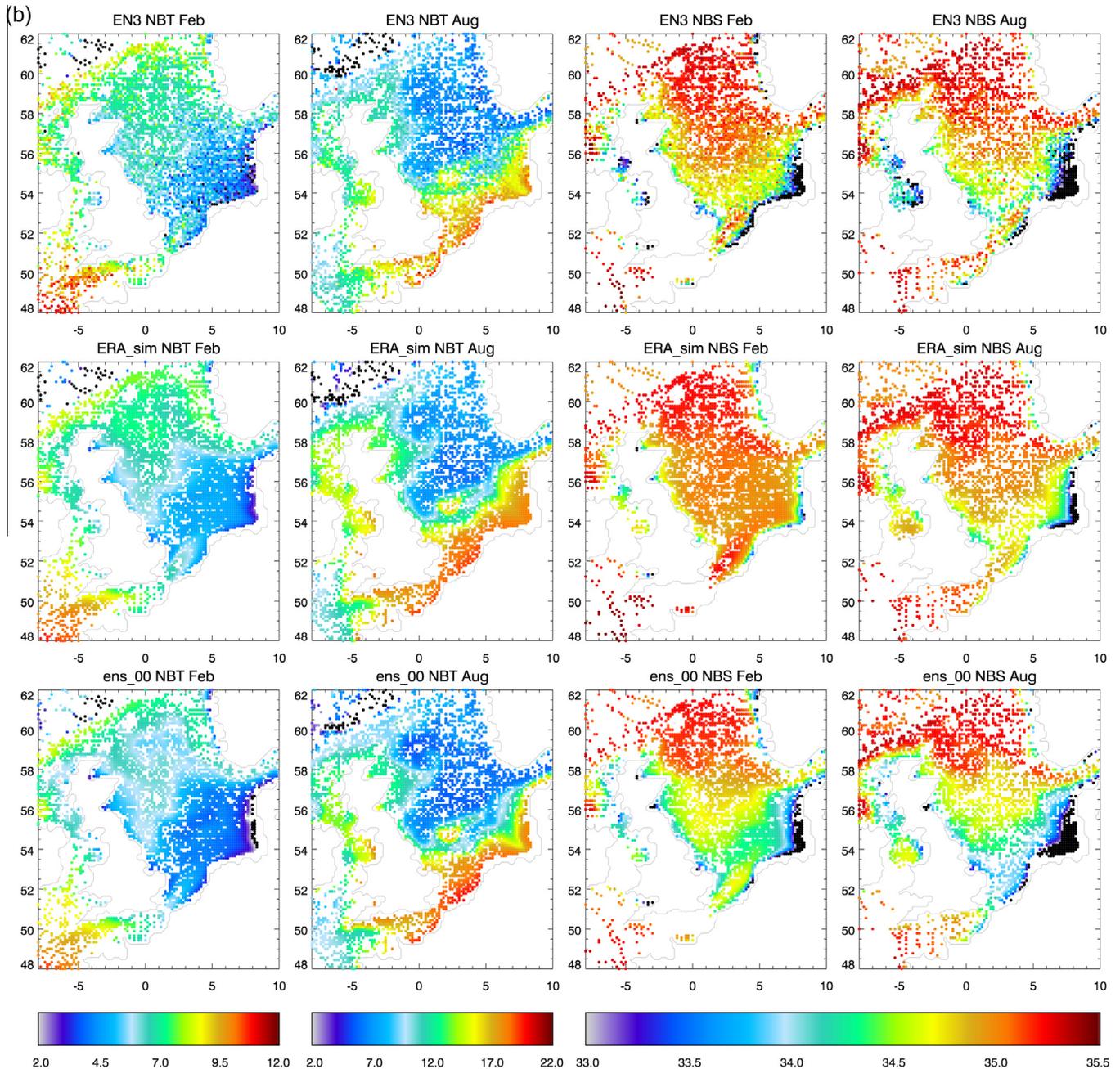
Fig. 5 (continued)

Britain, and the German Bight. The model performed better for the NBS (Fig. 5b) than the SSS (Fig. 5a) in all regions (Fig. 6a). The large fresh bias in the Norwegian Trench is a surface feature, and so the NBS correlation in the North Sea improved (SSS $r = 0.64$, NBS $r = 0.88$) as did the wider shelf (SSS $r = 0.75$, NBS $r = 0.85$).

Fig. 6b shows the Taylor diagram for the annual mean fields of ens_00 broken down into validation regions (analogous to Fig. 6a), but with lines showing how the positions on the Taylor diagram have changed from ERA_sim. Overall there is an increase in relative spatial variability when moving from ERA_sim to ens_00. This leads to a slight increase in *rrms* of the temperature fields, mainly due to the *rsd* increasing from the typical value of ~1 to ~1.1. The difference in the salinity fields is much greater. While the *r* and *rrms* remain relatively similar, the *rsd* typically increase from *rsd* 0.5 to ~1.1, although the Celtic Seas salinities *rsd* remained at *rsd* ≈ 0.60.

In order to understand these relatively large changes in the salinity on the Taylor diagram, we look at data used to calculate the Taylor statistics. The river inflows used in ens_00 are greater than those used in ERA_sim and the oceanic inflow to the North Sea less (Table 4, Shetland-Norwegian Trench, Fair Isle). Hence, the coastal regions tend to be fresher (Fig. 5a) in ens_00. It can be seen that the distribution of the salinity values over the annual mean SSS for the whole shelf for ens_00 is in better agreement with the observations than the ERA forced run. This is largely due to the pattern of the fresher coastal waters in the North Sea (along the east coast of the Britain, and the German Bight) which is in better agreement with observation in ens_00 compared to ERA_sim. Overall, their was little change in the *rrms* for the annual mean surface and bed salinity for most regions when moving from the ERA_sim to ens_00.

When looking across the ensemble, all members tend to sit within a similar region of parameter space on the Taylor diagram

(Fig. 6c). The parameter space spread across the ensemble for a particular variable is generally smaller than the distance between variables. The temperature values are more tightly grouped than salinity values. For each variable, the ensemble tends to be spread radially rather than in the polar direction (*rsd* varies more than *r* across the ensemble). The suppression of the polar spread reflects the control that internal dynamics have over the spatial pattern (common tides, bathymetry and a general north–south temperature gradient lead to a common general pattern between ensemble members). The radial position reflects the amplitude of the pattern (measured by the *rsd*) and so the ensemble's radial spread suggests that the strength of spatial pattern is more controlled by the varying lateral and surface boundary conditions (which vary between ensemble members) than the internal dynamics (which do not). For example, the summer NBT pattern (Fig. 5b) is dominated by the pattern of the stratified and mixed regions. As this pattern is largely controlled by bathymetry and tides (common to all ensemble members) the pattern varies little across the ensemble – thus all ensemble members have similar *r* values. However the temperature difference between the mixed and stratified regions varies between the ensemble members, and so the *rsd* is more variable across the ensemble.

The ERA_sim OSTIA SST has a slightly lower *rrms* error than any of the ensemble member. The ensemble correlations for shelf wide, annual mean SST (for both EN3 and OSTIA) range from $0.92 < r < 0.99$, with *rsd* ranging from $0.93 < rsd < 1.18$, with ens_00 towards the middle of the ensemble. Conversely, the ERA_sim NBT is slightly worse than any of the ensemble members. The ensemble NBT correlations range from $0.93 < r < 0.95$, and relative standard deviations from $1.03 < r < 1.12$, with the ERA_sim point having the worst correlation and error.

When looking at the SSS, the ensemble has correlations ranging from $0.72 < r < 0.80$ with *rsd* from $0.92 < rsd < 1.28$. The ERA_sim is quite distinct from the ensemble with an $rsd = 0.51$, although there is a similar correlation, $r = 0.75$. Again ens_00 is not the best performing ensemble member. The results for the NBS are similar to those of the SSS, with ensemble correlations ranging from $0.84 < r < 0.87$ with *rsd* from $1.0 < rsd < 1.30$, ERA_sim being a distinct outlier in terms of *rsd* ($rsd = 0.57$), the ens_00 performance being somewhere in the middle of the ensemble. The difference of the ERA and ensemble relative performance between SST and salinity is likely due to the riverine forcings. While all simulations are forced with TRIP-routed rivers, the ERA_sim is slightly drier than the ens_xx ensemble (in terms of precipitation minus evaporation), which leads to a systematic difference in riverine forcings (the total riverine fresh water enter domain (averaged over 1960–1989 and over the ensemble) is 0.029 Sv for ens_xx, and 0.021 Sv for ERA_sim). Thus ERA_sim has more saline coastal regions than the entire ens_xx ensemble, which reduces its spatial relative standard deviation compared to observations, and so impairing its skill compared to the ens_xx ensemble.

4.4. Evaluating the modelled mean climate and climate variability

4.4.1. Low frequency variability associated with natural variability and model parameter uncertainty

We have two estimates of the modelled low-frequency variability for each variable, from the pre-industrial control run (the variance of the five 28-year means from ens_ctrl), and from the PPE spread (the variance of the 11 30-year means). The first is possibly an underestimate, as it is driven with a global model atmosphere (which may reduce the variability) and is calculated from only 5
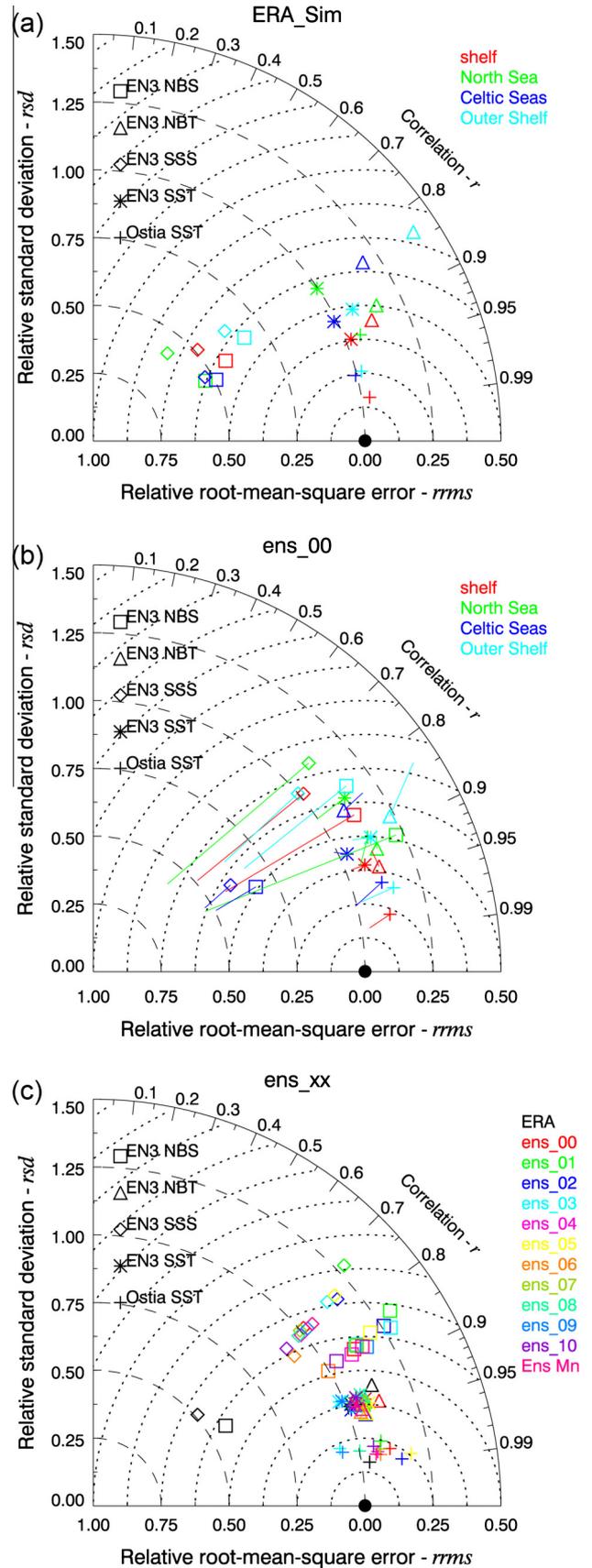

Fig. 6. Taylor diagrams. (a) annual mean ERA_sim against validation regions, (b) ens_00, with lines showing change from ERA_sim and (c) shelf-mean ens_xx.

samples, while the second may be an overestimate, as it also contains any signal of the perturbed physics in the present day.

We compare these two estimates with an *f*-test. For temperature, the low-frequency variability estimate from the ensemble spread is generally significantly greater than that of the pre-industrial control run (Fig. 7a). This is consistent with the different parameter settings in the different ensemble member (leading to different climate sensitivities and model biases) being greater than the low-frequency variability also sampled by the ensemble. This is not the case for salinity (Fig. 7a) where there tends to be an agreement between the two methods of estimating the low frequency variability. As the climate response of salinity in the near future (before the 2050s) is limited (not shown in this study), it is not surprising that the ensemble spread is similar to that NV estimate from the control run.

The PPE estimate of PEA low-frequency variability is greater than that from the control simulation, although this is insignificant in most of the North Sea. By contrast, the MLD low-frequency variability from the PPE is significantly less than that of the control simulation (Fig. 7b).

As the low-frequency NV estimate from the ensemble spread is generally greater or equal to the estimate from the pre-industrial control run, we are able to use the modelled distribution directly in the following sections.

4.4.2. Mean climate evaluation

It is not possible to directly assess biases in the modelled simulation of the mean climate with the observation or the ERA_sim simulation, due to the possible difference in phase of NV. Instead we consider where the observations and ERA_sim 30-year mean estimate fits within the modelled distribution from the ensemble. This informs our later interpretation on whether any differences from the observations are likely to be due to NV or model bias (see Fig. 3, Table 1).

Here we fit a normal distribution to the 11 30-year means of the ensemble data (for each grid box) and create a cumulative distribution function (CDF). We then identify where the multi-year mean observations and the ERA_sim fit within this CDF. Interpreting these two positions simultaneously give an insight into possible reasons for any disagreement with the modelled distribution.

Generally the observations and ERA_sim fit within the modelled temperature distributions, but are usually outside the salinity distribution. The water column structure variables (MLD and PEA) also tend to be outside the modelled distribution, particularly in the vicinity of the Norwegian Trench.

The OSTIA SST observations tend to be consistent with the modelled distribution between May and November, with small areas that tend to be to cooler than the distribution (Fig. 8a). During the cooler months (December to March/April), the observations are warmer than the modelled distribution in the Celtic Sea, English Channel, northern part of the North Sea (Fig. 8a).

In winter, the regions where ERA_sim and OSTIA SST fields fall outside the modelled distribution tend to agree (Fig. 8a). During the summer months, ERA_sim tends to be consistent with the modelled distribution, but in the warmer half of the modelled distribution rather than in the cooler half like OSTIA.

As the shelf tends to be fully mixed in the winter, in both the model and the observations, the conclusions drawn from SST fields tend to hold for NBT (Fig. 8b). In the stratified regions during the summer, there is a complex pattern of regions that are warmer and cooler than the modelled distribution. The northern North Sea EN3 NBT tends to be warmer that the modelled distribution. Often the regions where the NBT is warmer than the modelled distributions have EN3 SST that is cooler than the modelled distribution. This is consistent with the modelled stratification being too strong, which is considered with the PEA.

In the winter there is a general agreement between the ERA and EN3 in terms of regions that are within, above and below the modelled NBT distribution. In the summer, there is much more disagreement between ERA and EN3. The central North Sea ERA_sim NBT tends to fall within the modelled distribution, while the southern North Sea NBT field is often above the modelled distribution. In the vicinity of fronts, the EN3 NBT is often below the modelled distribution while the ERA NBT is above it, or vice versa.

The Norwegian Trench EN3 SSS are always fresher than the modelled distribution (Fig. 8c). The SSS in this region is strongly influenced by the Baltic outflow, which is a climatology boundary condition in the modelled distribution – a limitation of this study. Outside the Norwegian Trench, most of the observations are saltier than the modelled distribution, although the German Bight is often within the ensemble spread (e.g. February Fig. 8c). The ERA_sim SSS tends to be more haline than the modelled distribution throughout the shelf, including the Norwegian Trench. For most of the shelf, outside the Norwegian Trench, the EN3 and ERA_sim both have SSS greater than the modelled distribution. The ERA_sim SSS in the vicinity of the German Bight are more haline than the model distribution, despite the agreement between the EN3 SSS and the modelled distribution.

Most of the modelled shelf is fully mixed with respect to salinity, and there are inadequate data to examine a vertical salinity structure in observations, and so the conclusions drawn from the NBS are very similar to those from the SSS (Fig. 8d). The Norwegian Trench is the exception, being salinity stratified throughout the year. Furthermore, the Norwegian Trench bed and surface waters have very different origins – the low salinity Baltic outflow tends to flow off the shelf in the surface waters of the Norwegian Trench, while higher salinity Atlantic water flows into the Norwegian Trench at depth. This is reflected in the EN3 NBS in the Norwegian Trench (as with the rest of the domain) being more haline than the modelled distribution, despite the EN3 SSS being fresher than the modelled distribution.

Most of the NWE shelf seas are fully mixed in the winter, with the exception of the Norwegian Trench and Skagerrak/Kattegat. Many parts of the NWE shelf seas seasonally stratify in the summer, while other parts remain fully mixed throughout the year. Due to the outflow of low salinity Baltic water, the Norwegian Trench and Skagerrak/Kattegat are salinity stratified throughout the year.

We limit our model system evaluation of the water structure to the summer months and to regions that stratify. We assess the water column structure, through the depth of the Mixed Layer (MLD), and the strength of the stratification with the Potential Energy Anomaly (PEA). The EN3 observations tend to be fairly sparse outside the North Sea, Norwegian Trench/Skagerrak/Kattegat region, and so we focus our evaluation of water column structure in these regions.

Outside the Norwegian Trench, all the observations of MLD on the shelf are greater (deeper) than the modelled distribution, while in the Norwegian Trench the observed MLD are less (shallower) than the modelled distribution. The ERA_sim MLD field tend to agree with the modelled distribution for most of the central and western North Sea (Fig. 8e). In some other regions, ERA_sim MLD field is greater (deeper) than the modelled MLD distribution (e.g. Norwegian Trench and the southern/eastern North Sea), although the absolute values are similar.

The Norwegian Trench EN3 PEA is always significantly greater than the modelled distribution (Fig. 8e) while the rest of the North Sea tends to be slightly lower, although EN3 PEA values at the edge of the Norwegian Trench sometime agree with the modelled distribution (e.g. July central northern North Sea). The ERA_sim North Sea PEA values tend to agree with the modelled distribution, while the ERA_sim Norwegian Trench PEA is lower than the modelled distribution.
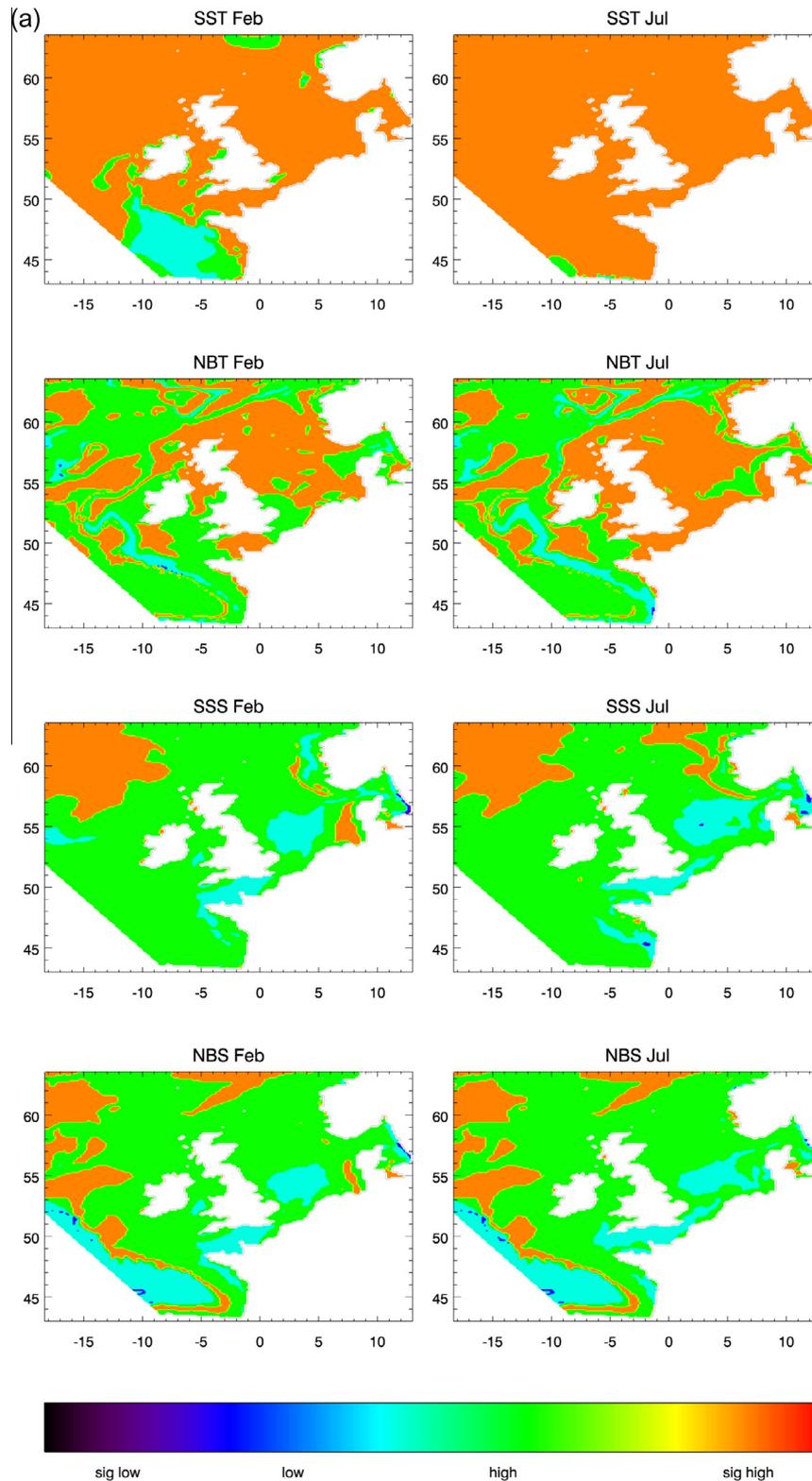
Fig. 7. Comparison of low frequency variance as estimated from the ensemble spread, and the pre-industrial control simulation, for February and July, for (a) SST, NBT, SSS and NBS, and (b) July MLD and PEA. The two estimates of variance have been tested with an *f*-test, and the results are presented as being ensemble spread variance estimate being significantly (or insignificantly) larger (orange, green respectively) or smaller (dark or light blue respectively) than the estimate from the pre-industrial control simulation.

4.4.3. Interpretation

Simultaneous interpretation of the position of the observations and ERA_sim within the modelled distribution provides additional insights underlying causes of any discrepancy – the possible role of model or climate forcing bias or NV.

In the summer, ERA_sim and OSTIA SST tends to agree with the modelled SST distribution. In the winter the ERA_sim and OSTIA SST fields tend to agree on which regions fall outside the modelled distribution. As observations and ERA have the same phase of natural variability, this agreement does not discount the possibility
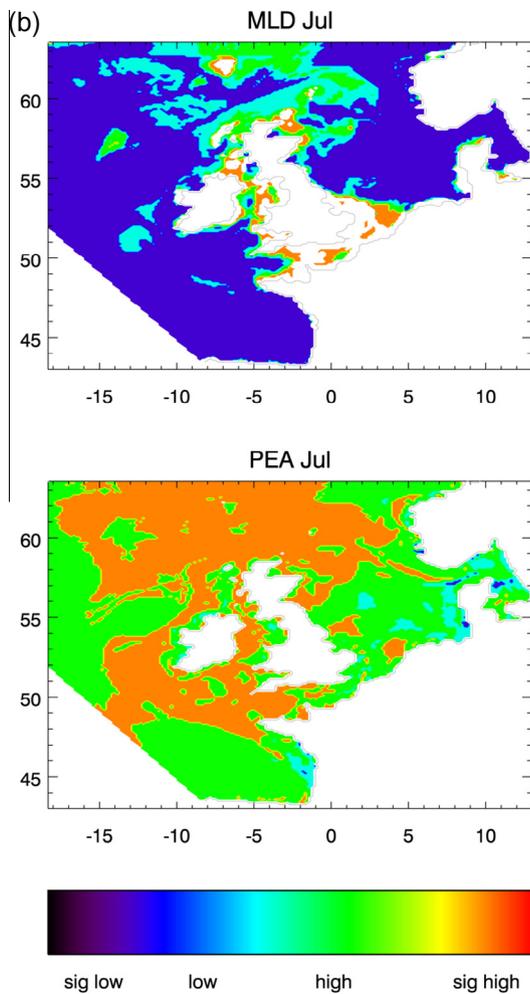
(b)

Fig. 7 (*continued*)

that the apparent bias is due to the phase of the low-frequency variability. The matching patterns are also consistent with the climate model forcings being biased. When considering the EN3 SST observations (not shown), there are regions where the EN3 observations are lower than the modelled distribution, while the ERA_sim values are consistent with the modelled distribution – this could be due to an observations bias or a downscaling model bias (common to both the modelled distribution and ERA_sim). However as this occurs in a location where the OSTIA observations tend to sit within the modelled distribution, tends to suggest an observational bias.

There tends to be agreement within ERA_sim and EN3 about which regions have NBT that are above and below the modelled distributions in winter. This could also be due to the phase of the low-frequency variability or model bias. In summer, there is less agreement, particularly in the regions of fronts, which is less likely to be due to the differing phases of NV, and more likely due to model biases.

The EN3 and ERA_sim SSS in most regions outside the Norwegian Trench are more haline than the modelled distribution, which is consistent with a fresh bias associated with the climate forcings, but is also consistent with low-frequency variability. The disagreement between the EN3 and ERA_sim SSS in the Norwegian Trench (EN3 begin fresher than the modelled distribution, while ERA_sim being more saline) is consistent with a bias in the downscaling methodology (i.e. the Baltic forcings), combined with a bias in the ERA_sim simulation forcings. The localised

disagreement between EN3 and ERA_sim in the German Bight perhaps suggests the ERA_sim forcings (e.g. river forcings) in this region are more biased than the ensemble forcings.

The ERA_sim (summer) MLD field tends to be in agreement with the modelled distribution, while the EN3 MLD tends to be shallower than the modelled distribution in the Norwegian Trench, and too deep elsewhere. In the Norwegian Trench, this is consistent with a bias in the downscaling technique, likely to be associated with the Baltic forcings, common to the ens_xx and ERA_sim. Outside this region it is consistent with a bias in either the MLD calculated from the EN3, or a bias associated with the downscaling model.

Within the Norwegian Trench, the EN3 PEA is greater than the modelled distribution, while the ERA_sim PEA is less than the modelled distribution. This is consistent with a concurrent shelf seas model bias and ERA_sim bias – the imperfect shelf seas model Baltic forcings leading to incorrect haline stratification in the Norwegian Trench relative to the observations, while the ERA_sim river forcing are less than the ensemble forcings, and so the ERA_sim Norwegian Trench is even less haline stratified than the ensemble modelled distribution. Outside this region (generally the western part of the North Sea), the both the EN3 and ERA_sim PEA tends to be generally in within the modelled ens_xx distribution (although some EN3 points are below).

These results suggest the modelling system generally reproduces the mean temperature climate well, although the winter SST field may be influenced by the relative phase of the modelled and observed NV or the biases in the climate forcings. The modelled salinity has a widespread statistically significant bias, associated with biases in the climate forcings or NV. Care should be taken interpreting the absolute values of the modelled salinity output. In the Norwegian Trench, the salinity bias is likely associated with the downscaling technique, and care should be taken when considering modelled SSS in this region. Outside the Norwegian Trench, the modelled water structure is considered sufficient for the purpose of climate projections, although care should be taken when using absolute values of MLD, as there appears to be a systematic bias associated with the shelf seas model. Care should be taken with the Norwegian Trench the water column structure results.

4.4.4. High-frequency (interannual) variability evaluation

We now compare the interannual variability estimates from ens_xx with that from the OSTIA and EN3 observations and ERA_sim.

Overall we find that the ens_xx summer high-frequency temperature variability is fairly consistent with the observations, while the ens_xx winter temperatures are less variable than the observations on the shelf. Outside the Norwegian Trench and adjacent waters, the ens_xx salinity interannual variability is consistent with the observations. The summer observed (and ERA_sim) stratified water column interannual variability is consistently more variable than in ens_xx. We consider that these results support the use of this model system for climate projections.

The OSTIA SST interannual variability off the shelf (to the north west of the domain) is significantly less than that of ens_xx throughout the year (Fig. 9a). In ERA_sim, the interannual variability of the entire boundary is significantly lower than that of ens_xx, due to the use of climatological boundary conditions.

On the shelf, the OSTIA variability is significantly greater than ens_xx interannual variability from February–May (Fig. 9a). From June–November, there is generally an agreement between the ensemble and OSTIA variability, although the Skagerrak and Norwegian Trench tend to have too low interannual variability in ens_xx (Fig. 9a).
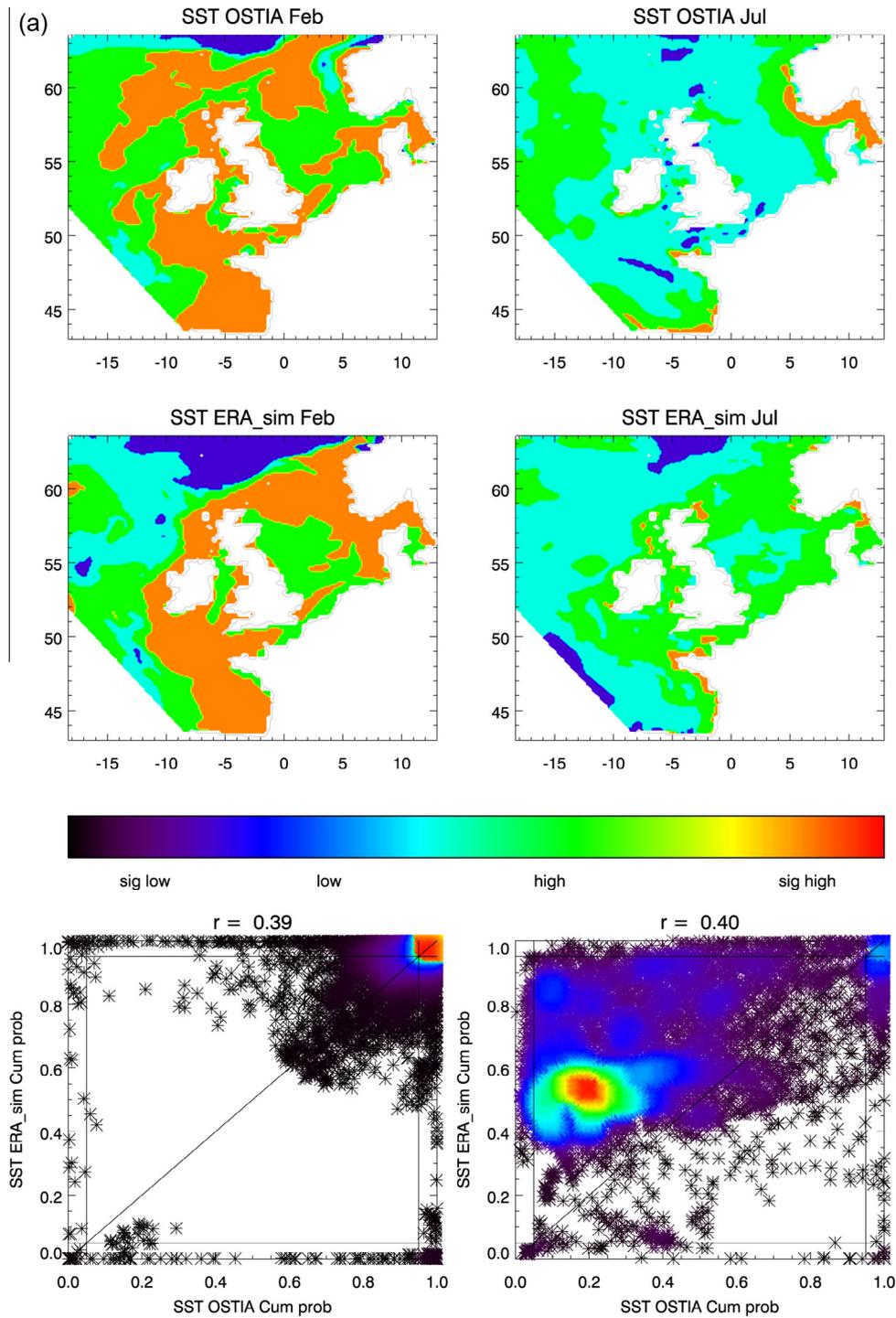
Fig. 8. Consistency of 30 year mean estimates of the observations and ERA_sim with the modelled distribution from ens_xx, for February and July (a) OSTIA SST, (b) EN3 NBT, (c) EN3 SSS, (d) EN3 NBS and (e) July EN3 MLD and PEA. The upper (middle) panel shows spatial maps showing where the 30 year mean observations (ERA_sim) fit within the ens_xx ensemble: in the left tail ($p < 0.05$, dark blue) or right tail (orange, $p > 0.95$) of the distribution. The lower panel compares position of the observations within the modelled distribution to the position of the ERA_sim within the modelled distribution, akin to Fig. 3. The colouring is indicative of data density.

The ERA_sim shelf SST interannual variability is significantly greater than that of ens_xx from November–May, although this is not significant in the northern and central North Sea for November, December and April (not shown). There is a general agreement between June and October, although there are still regions where the ERA_sim interannual variability is greater than that of ens_xx (Fig. 9a).

In the winter, the NBT interannual variability is similar to the SST interannual variability. In stratified regions in the summer (on the shelf) the EN3 NBT interannual variability tends agree with that of ens_xx, while the ERA_sim interannual variability is greater than that of ens_xx. In the sparsely sampled mixed southern North Sea and English Channel, the ERA_sim interannual variability tends to agree with that of ens_xx (not shown).
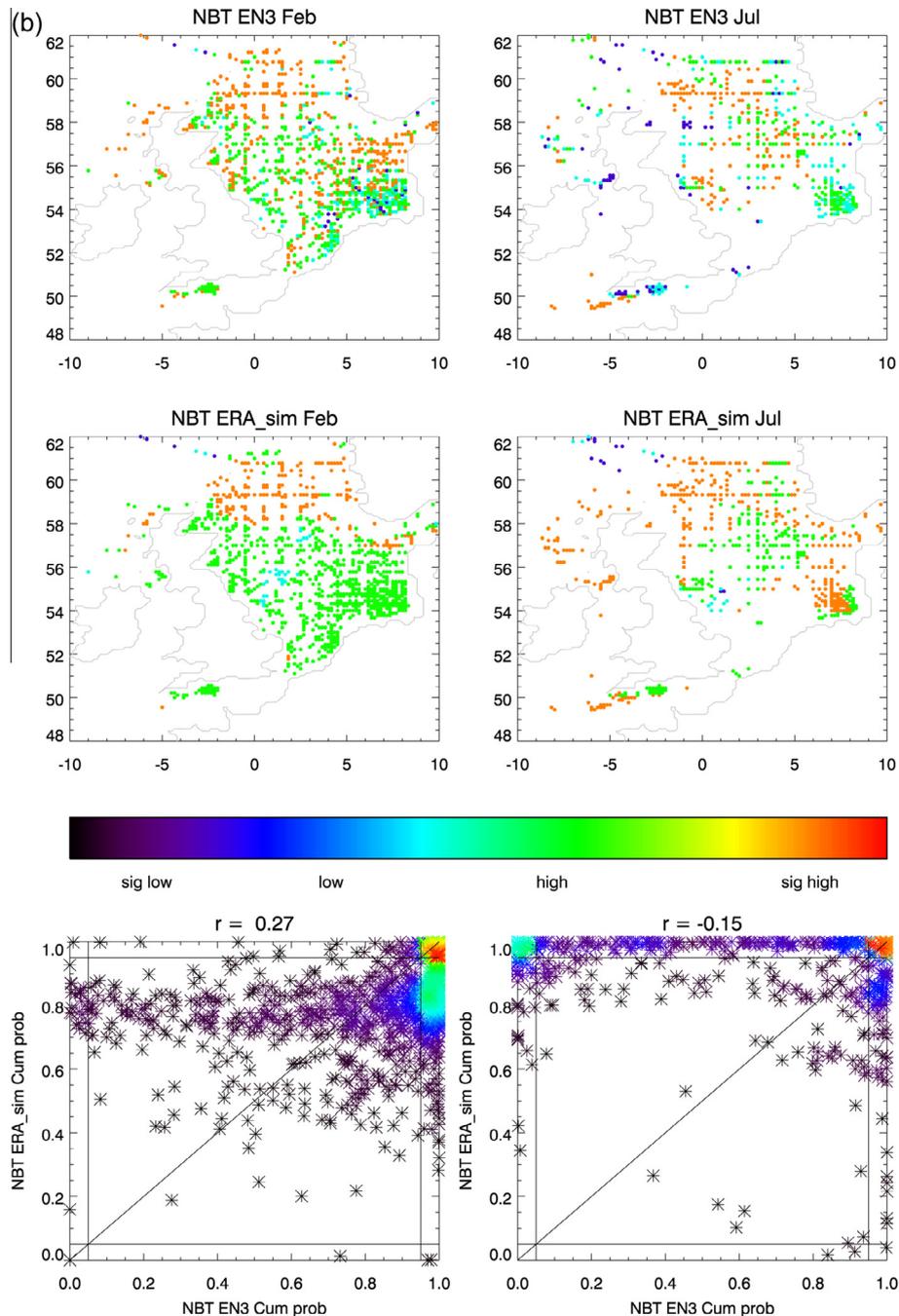
Fig. 8 (*continued*)

The EN3 SSS interannual variability is significantly greater than ens_xx in the Skagerrak and in the vicinity of the Norwegian trench, due to the use of climatology Baltic boundary conditions. This extends to towards the centre of the North Sea in June and July (Fig. 9b). The EN3 SSS interannual variability in the south-western North Sea and German Bight also has a tendency to be higher than ens_xx. Otherwise there is a generally agreement between interannual variance of the ensemble interannual variability and the EN3 SSS variability.

Off the shelf, the ensemble SSS variability is much greater than the ERA_sim SSS variability, and this tends to penetrate quite far onto the shelf (into the north and central North Sea, the Celtic Sea, western part of the English Channel and the southern part of

the Irish Sea). These patterns are consistent with the ERA_sim SSS variability being set to zero at the boundaries (climatology lateral boundary conditions). In the southern North Sea winter salinity, the ERA_sim interannual variability is greater than ens_xx in the west and weaker in the east (the German Bight). The western region reflects the greater variability of the current through the Dover Straits in ERA_sim (compared to the ens_xx). The weaker ERA_sim SSS variability in the German Bight is due to the variability of the width of the region of low salinity water adjacent European coast – it is much thinner in ERA_sim, and less variable.

For most regions away from the Norwegian Trench, the NBS interannual variability results mirror those of SSS (Fig. 9b). Due to the strong salinity stratification in the Norwegian Trench, the
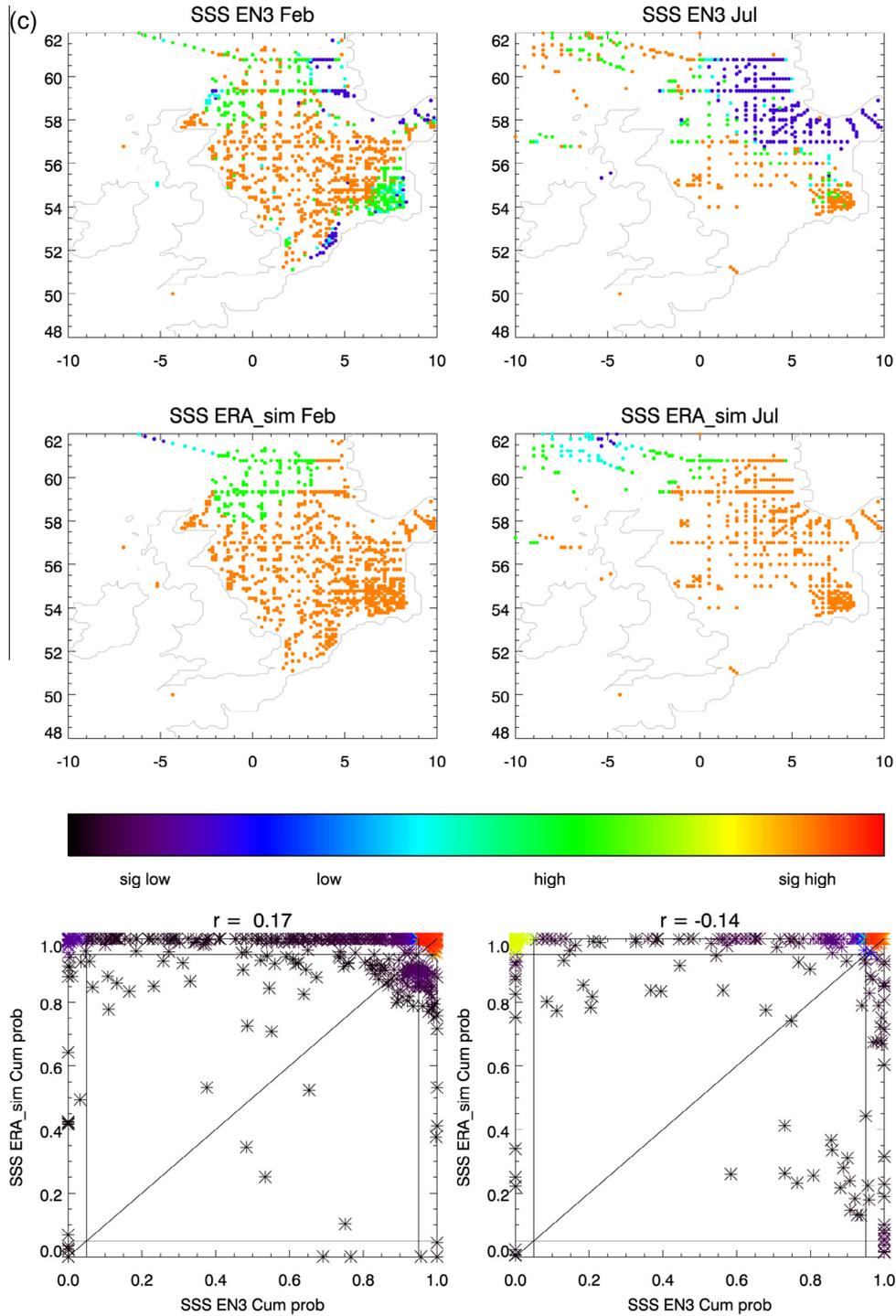
Fig. 8 (continued)

NBS is decoupled from the SSS, and as water in this region has a different origin (North Atlantic rather than Baltic) it has different interannual variability. In this region, the EN3 NBS interannual variability tends to be consistent with that of ens_xx, while the ERA_sim interannual variability is more consistent, or even lower than that of ens_xx, consistent with the use of climatology boundary conditions.

On the shelf, the EN3 and ERA_sim MLD and PEA interannual variability is consistently higher than that of ens_xx (Fig. 9c).

4.5. Circulation and volume transport through observed cross-sections

It is important that the model broadly reproduces the configuration of the shelf circulation. As we do not have detailed altimetry products for the shelf seas, we instead make use of limited observed volume transport cross-sections. Here we use the results of the quantitative current tests to make qualitative assessments of the overall circulation pattern.

These model-observed volume transport comparisons suggest that the modelled circulation configuration is qualitatively correct.

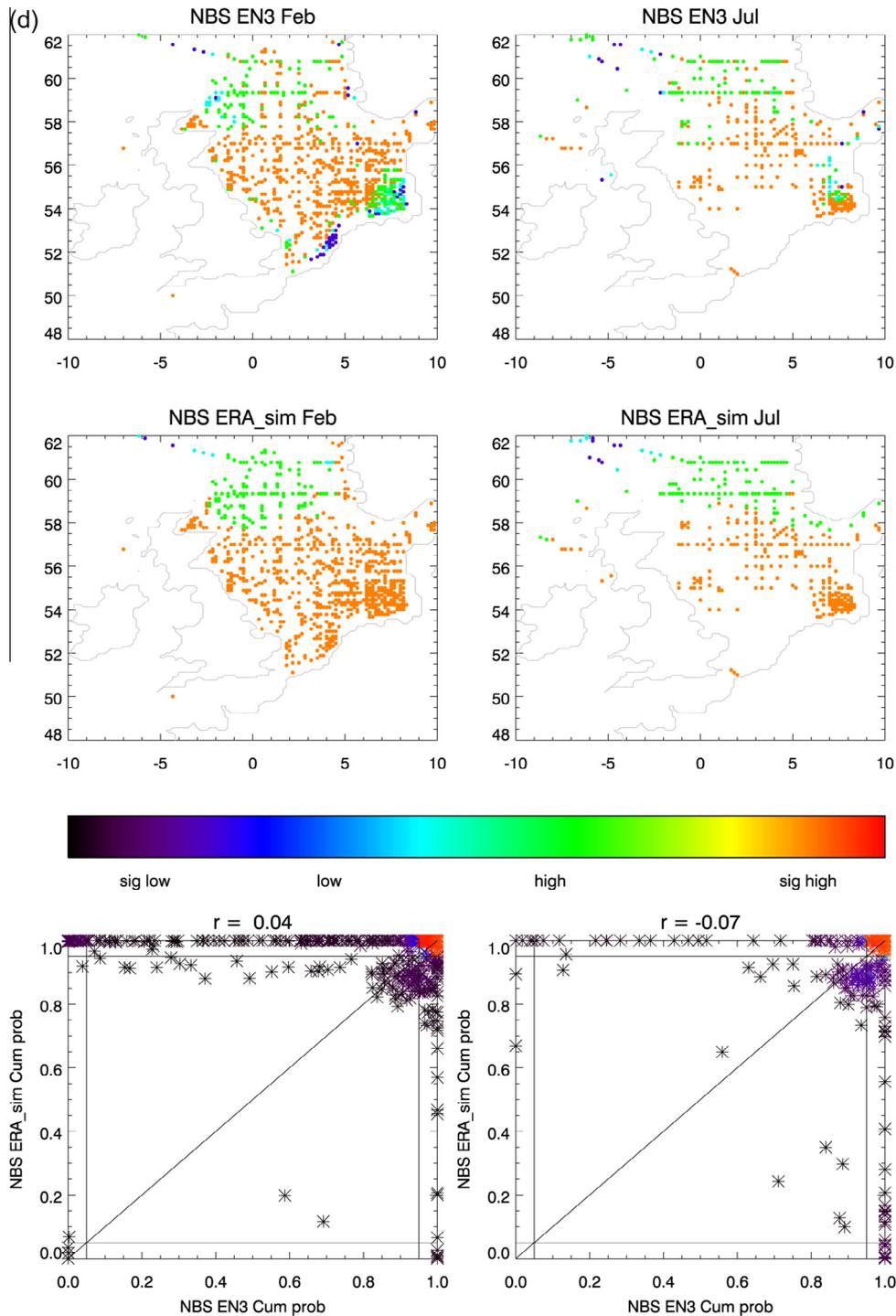sig low   low   high   sig high

Fig. 8 (continued)

This supports the overall modelling approach, by suggesting the advective processes are generally correct. However, differences in current magnitudes suggest that the care must be taken when using the modelled currents results directly.

The ERA_sim mean transport through most of the 12 observed cross-sections (Svendsen et al., 1991; Turrell et al., 1992; Prandle et al., 1996; Danielssen et al., 1997; Brown et al., 1999, 2003; Holt et al., 2001; Fernand et al., 2006) presented in Lowe et al. (2009) are similar to the observed transport rates (Fig. 10, Table 4), considering the coarse resolution of POLCOMS. The cross-sections north of the Dogger Bank, the Dooley Current, between Shetland and the Norwegian Trench, across the

Norwegian Trench and across the Dover Straits (Fig. 10, sections 02, 03, 04, 05 and 10 respectively) are in relatively good agreement with the observations (with the observations being within the modelled seasonal cycle, or the model within the observational range if present). The modelled currents in ERA_sim are too strong in the Fair Isle, Hebrides Shelf, Irish Shelf and shelf break current at 56°N cross-sections (Fig. 10, sections 06, 07, 08 and 11).

Both the ERA_sim St George's Inflow and the shelf break current at the Faroe Shetland Trench are too weak (Fig. 10, sections 09, 12). The observed transport cross-section for the Skagerrak (Fig. 10, sections 01) is only for the eastward component of the current,
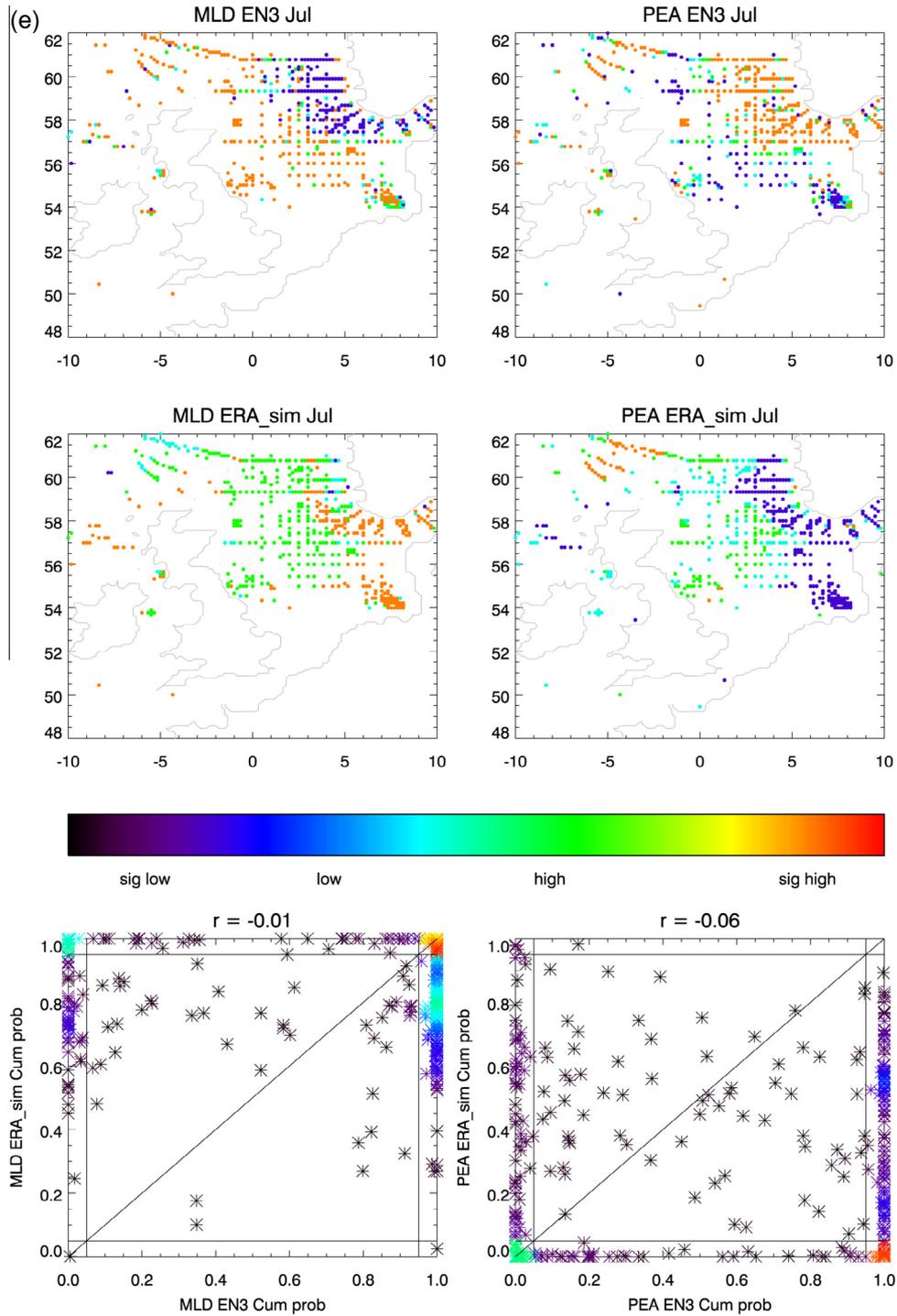
Fig. 8 (*continued*)

and so the present day modelled section excludes westward out-flow. When this is compared to the observations, the modelled sea-sonal cycle overlaps with the observational errors.

Overall, the ens_00 modelled current though most of the obser-vation cross-sections, particularly those in the interior of the shelf, are consistent when considering the constraints mentioned above. The current cross-sections with the largest differences between the observed and modelled currents tend to be towards the edge of the shelf, under the influence of the oceanic conditions. The currents from ens_00 are similar to those of ERA_sim, but there are some

differences. The ens_00 currents tend to be a little weaker, and often the seasonal cycle is depressed (Fig. 10, Table 4).

The current section into the Skagerrak (Fig. 10, section 01) has a very similar seasonal cycle to ERA_sim and the same annual mean. Others show similar annual mean values, but depressed seasonal cycles (e.g. north of Dogger Bank, Fig. 10, section 02), or similar seasonal cycles but depressed annual mean values (e.g. Shetland to Norway trench; Norway – Norwegian Trench; Fair Isle; Hebrides Shelf; Dover Strait (Fig. 10, sections 04, 05, 06, 07 and 10 respectively)). Cross sections that show both a depressed sea-sonal cycle and annual means include: Irish Shelf; Shelf current
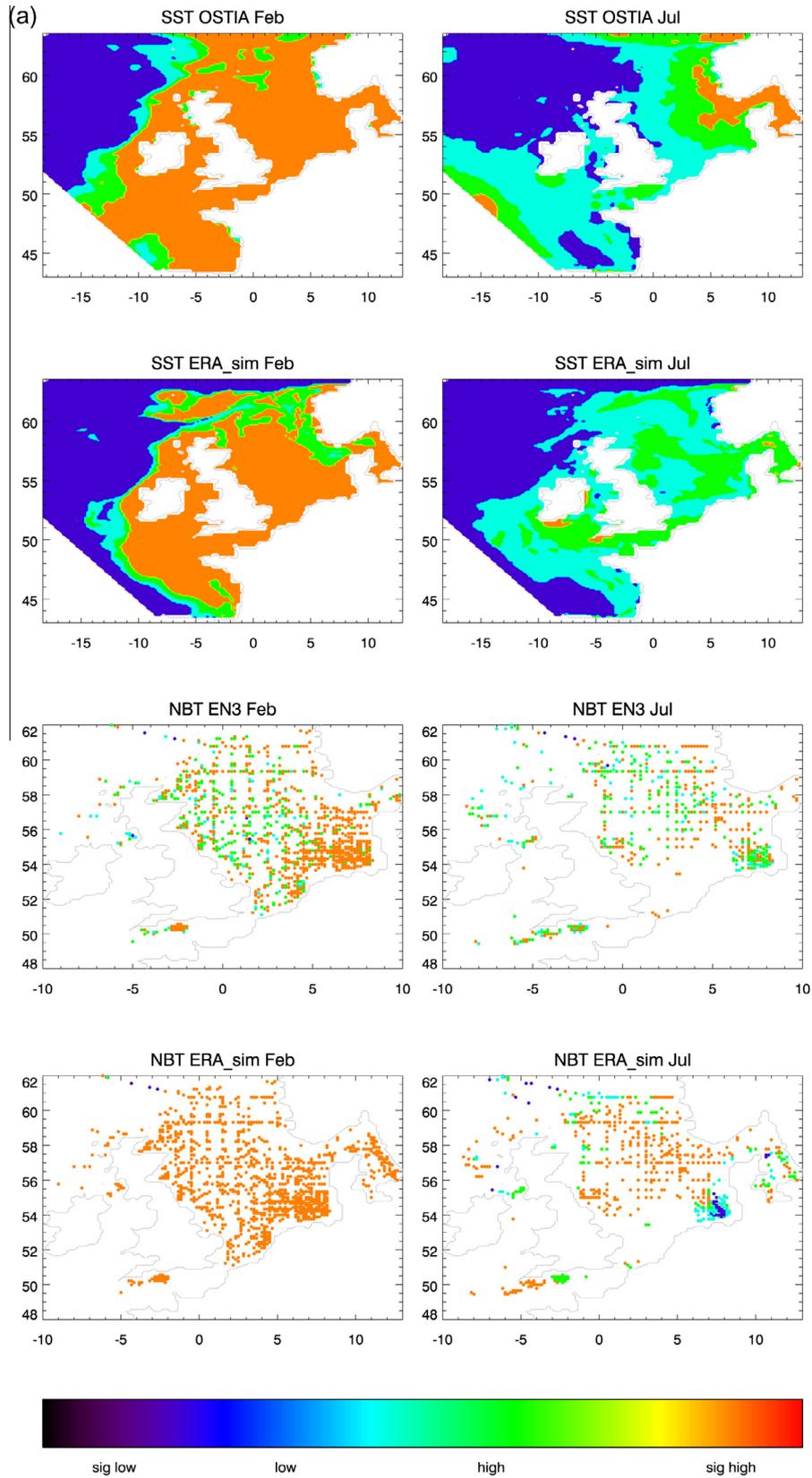
Fig. 9. Comparison of high frequency (interannual) variability between the observations (ERA_sim) and the ensemble, as tested with an *f*-test. The results are presented showing where the observations (ERA_sim) high frequency (interannual) variability is significantly (or insignificantly) larger (orange, green respectively) or smaller (dark or light blue respectively) than the estimate from ens_xx. Comparison with observations and ERA_sim are in the upper and lower rows for each variable. February and July (a) OSTIA SST and EN3 NBT, (b) EN3 SSS and EN3 NBS, (c) July EN3 MLD and EN3 PEA.
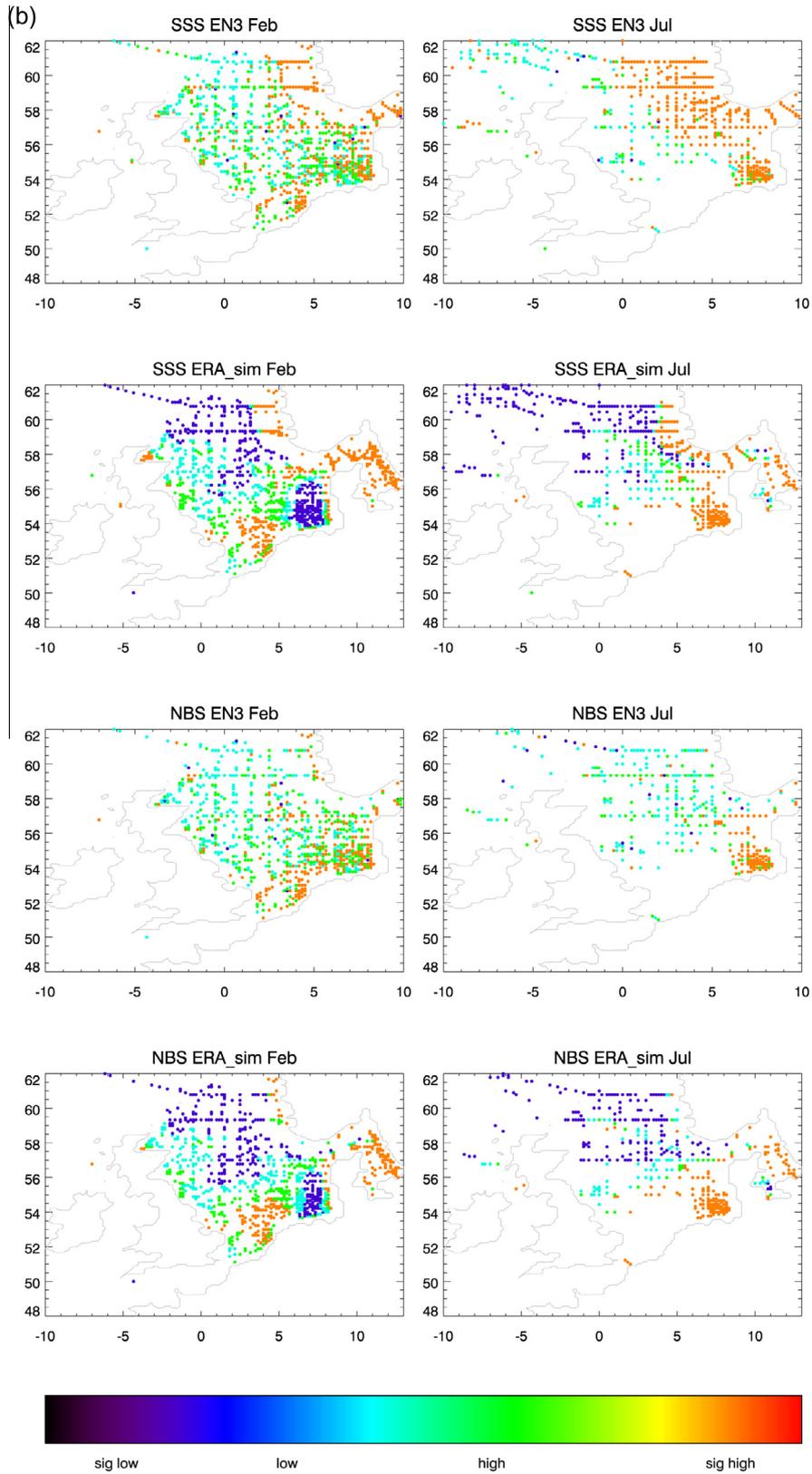
Fig. 9 (*continued*)

at 56°N; and Shelf current at FS current Strait (Fig. 10, sections 08, 11 and 12 respectively). The ens_00 Dooley Current (Fig. 10, section 03) is similar to the ERA_sim in the summer (the annual minimum), but the stronger winter values lead to greater annual mean and increases the positive current strength anomaly. The annual

mean St Georges inflow (Fig. 10, section 09) is stronger in ens_00, and the seasonal cycle is stronger.

Despite the differences between the ERA_sim and ens_00 simulations, the climate simulation does generally perform well. The Fair Isle, Hebrides Shelf, Irish Shelf, Shelf current at 56°N current
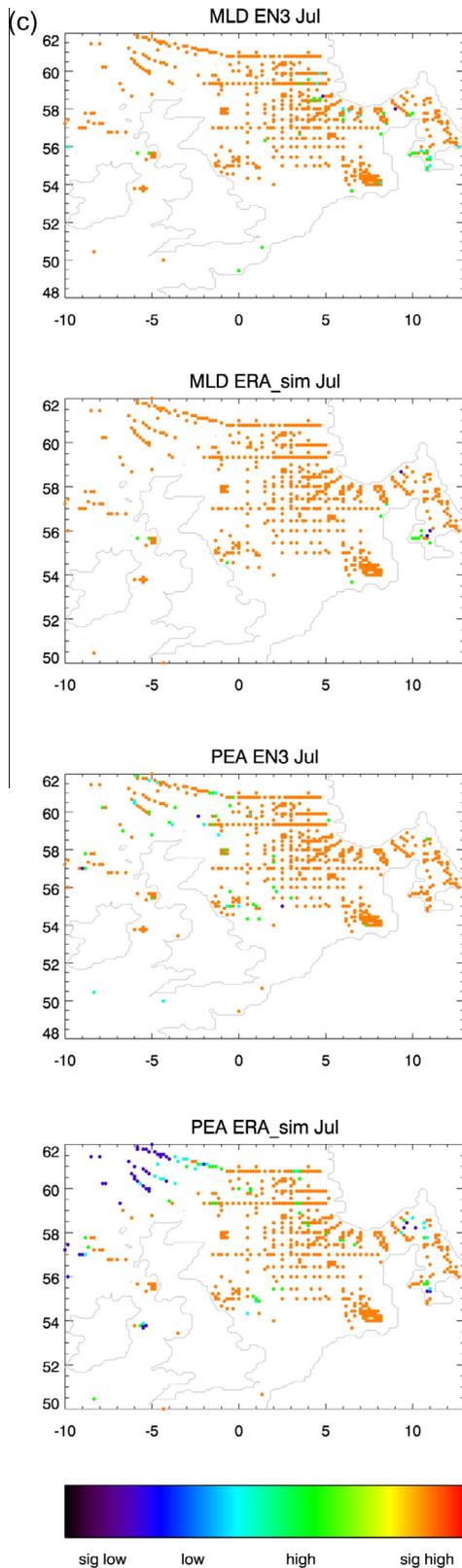
(c)

Fig. 9 (*continued*)

ensemble, similar to the ensemble mean (Fig. 10). The two exceptions to this are the Irish Shelf, and the Shelf current at 56°N, where ens_00 is notable stronger (Fig. 10, sections 08, 11).

We find there is relatively little ensemble spread. For most cross-sections, the annual mean ensemble range (max ensemble member − min ensemble member) is generally much smaller than the mean values (for most cross-sections, the spread is less than half the mean value). Exceptions include the Irish Shelf, and the Dover Strait (81% and 151% respectively, Table 4, sections 08, 10). Both these cross-sections tend to show greater spread in the winter month.

5. Discussion

We have developed a modelling system to downscale a global climate model (HadCM3) for the NWE shelf seas with a chain of existing models, in such a way that the climate variability and change can enter the shelf seas from the atmosphere, ocean or rivers. We have used this system to downscale an ensemble of transient simulations, from 1952 to 2098. The Perturbed Physics Ensemble has been designed to span the range of climate sensitivities associated with uncertainties in the model parameters of the HadCM3 atmosphere. In this paper we focus on describing the model set-up, experimental design, and the evaluation of the modelling system.

We believe our comprehensive evaluation methodology is more appropriate than those of other studies, given the available data. Some studies have been satisfied with a qualitative validation for certain variables, and have simply compared the observations and model output visually (whether surface maps or point time-series). For example, this approach is taken by Ådlandsvik and Bentsen (2007) to assess the current field. Olbert et al. (2012) compare their climate model forced present day simulation to a hindcast (of the same modelling system forced by NCEP reanalysis), rather than directly to observations. A more quantitative approach is more common, where model-observations differences are quantified (e.g. temperature and salinity time series of Ådlandsvik and Bentsen (2007)), and is often accompanying by a plausible explanation of the biases. Few climate projections of the NWE shelf have compared the model biases to the interannual variability (with the exception of Holt et al., 2010).

We have noted that low-frequency variability can lead to a distribution of 30-year means, of which the observed, and reanalysis climate are a single realisation. The use of an ensemble modelling approach allows the modelled distribution of these 30-year means to be calculated, and this then be compared to the observations. This is a more appropriate method to validate a climate model forced simulation against observations, and a fairer assessment of the model, given the available data.

When considering climate trends it is equally important to consider the Natural Variability (NV) of the system. There are a number of methods that can be used to quantify and analyse NV: time-series analysis of observations or reanalysis products (assimilative modelling systems); or coupled-climate models with stationary radiative forcings (such as pre-industrial climate control simulations). Both methods have advantages and disadvantages; the former is limited by the length of the available timeseries and includes climate trends (and so is not stationary); the latter is a limited by the model's ability to represent NV. In the shelf seas, the number of the long-term observations is very limited compared in the terrestrial system; despite the NWE shelf seas being some of the most observed regional seas in the world, there are very few timeseries extending over 60 years. There have been a number of studies using observations/reanalysis simulations to investigating NWE shelf seas NV (e.g. Holt et al., 2012b), however there is always the complication due to the non-stationarity of

cross-sections are closer to observations than in ERA_sim (Fig. 10, sections 06, 07, 08, 11), while ens_00 Dooley Current and Dover Strait are further from observations than ERA_sim.

When we compare the volume transport though the current cross-sections (Fig. 10), and compare the full ensemble and with ens_00 we find that ens_00 is generally fairly central in the
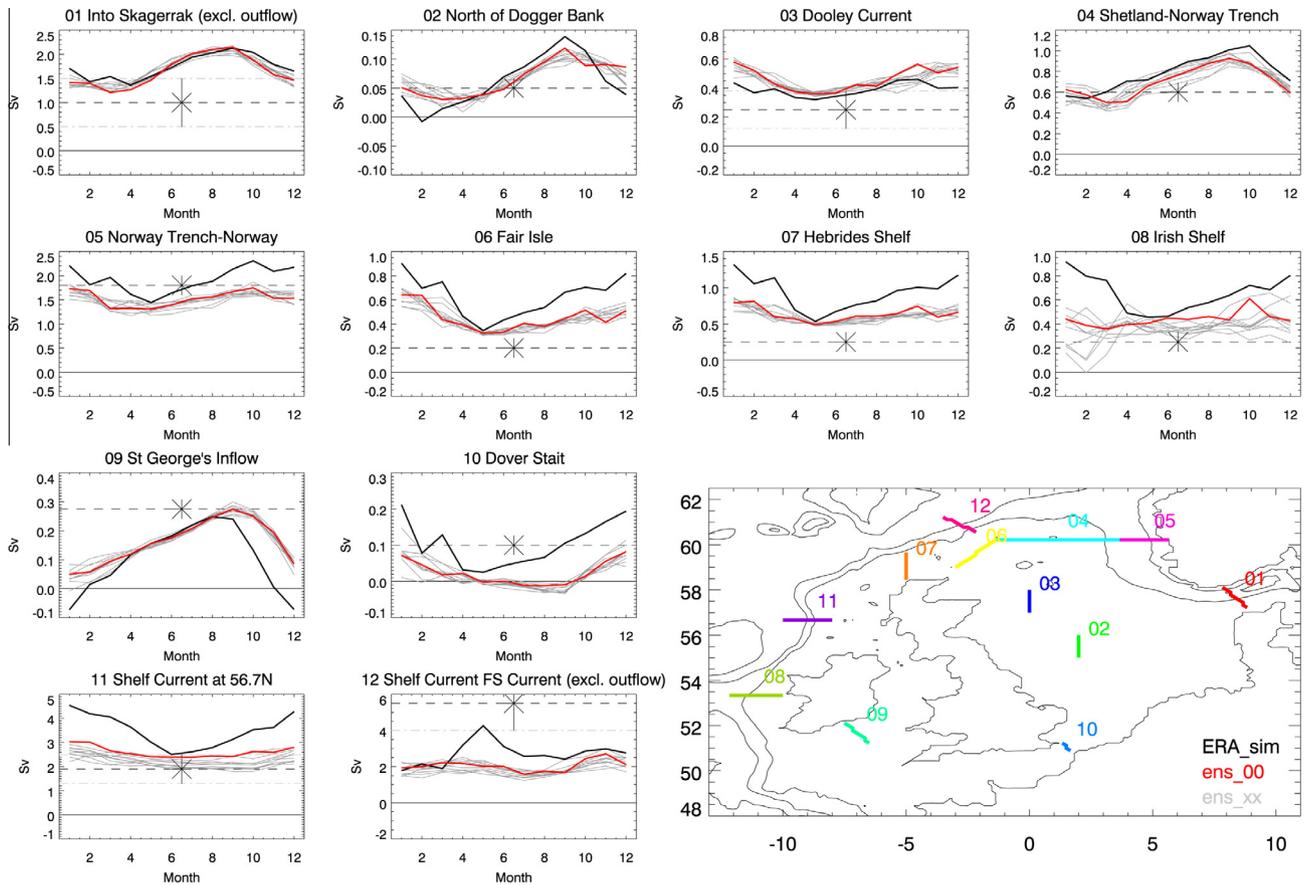
Fig. 10. Mean seasonal cycle of volume transport though observed cross-sections (Svendsen et al., 1991; Turrell et al., 1992; Prandle et al., 1996; Danielssen et al., 1997; Brown et al., 1999, 2003; Holt et al., 2001; Fernand et al., 2006). Star = observations (with/without error bars). Black line = ERA_sim, red = ens_00, grey = ens_xx (other ensemble members).

Table 4

Observed and modelled volume transport through cross-sections given in Fig. 10 (Svendsen et al., 1991; Turrell et al., 1992; Prandle et al., 1996; Danielssen et al., 1997; Brown et al., 1999, 2003; Holt et al., 2001; Fernand et al., 2006).

Cross-section	Cross-section number	Observed annual mean (Sv)	(Observed range (Sv))	ERA_sim annual mean (Sv)	(ERA_sim seasonal min/max (Sv))	ens_00 annual mean (Sv)	(ens_00 seasonal min/max (Sv))	ens annual mean range (Sv)	(ens_00 seasonal min/max range (Sv))	ens spread/ ens_00 (%)
Into Skagerrak (excluding outflow)	01	1	(0.50–1.50)	1.74	(1.36–2.13)	1.64	(1.21–2.15)	0.08	(0.16–0.21)	4.95
North of Dogger Bank	02	0.05		0.06	(−0.01 to 0.14)	0.07	(0.03–0.12)	0.01	(0.02–0.02)	14.53
Dooley Current	03	0.25	(0.12–0.38)	0.38	(0.32–0.46)	0.43	(0.36–0.58)	0.04	(0.03–0.09)	9.26
Shetland-Norway Trench	04	0.6		0.77	(0.54–1.05)	0.61	(0.50–0.93)	0.29	(0.17–0.13)	48.08
Norway Trench-Norway	05	1.8		1.87	(1.44–2.31)	1.36	(1.30–1.75)	0.37	(0.16–0.28)	27.01
Fair Isle	06	0.2		0.6	(0.35–0.90)	0.42	(0.32–0.64)	0.05	(0.03–0.16)	11.88
Hebrides Shelf	07	0.25		0.9	(0.53–1.32)	0.6	(0.49–0.81)	0.04	(0.04–0.19)	6.44
Irish Shelf	08	0.25		0.64	(0.46–0.92)	0.36	(0.36–0.61)	0.29	(0.36–0.30)	81.13
St George's Inflow	09	0.28		0.1	(−0.07 to 0.25)	0.17	(0.05–0.27)	0.02	(0.09–0.05)	12.58
Dover Strait	10	0.1		0.1	(0.03–0.21)	0.01	(−0.01 to 0.08)	0.02	(0.02–0.09)	151.28
Shelf current at 56.7N	11	1.9	(1.30–2.50)	3.42	(2.51–4.55)	2.01	(2.39–3.04)	0.94	(0.59–0.63)	46.72

such time-series (the inclusion of climate trend signal). We have downscaled a HadCM3 pre-industrial control with POLCOMS in order to investigate the NV of the NWE shelf seas (and the model drift). Our control run was limited to 146 years, and was forced by the coarse resolution HadCM3 atmosphere, which is a limitation. HadCM3 is considered suitable to investigate climate variability, and has been used as the basis of a number of variability studies (e.g. Knight et al., 2005).

Through our analysis of this downscaled control run we have been able to consider whether the NWE shelf seas are able to

sustain prolonged periods of warming and cooling under a stationary climate. We have shown that rapid periods of warming are possible for short periods, while more sustained periods of warming are possible at more modest rates. For example, with little model drift a sustained SST warming of 0.72 °C/yr is possible (95th percentile) for a 2-year sustained period, compared to 0.08 °C/yr for a 10-year sustained period. Furthermore, the pre-industrial control simulation also allows us to put the observed shelf warming into context of the background climate variability. For example, the observed warming of 0.4–0.5 °C/decade in the Southern North

Table 5
Summary of model system validation conclusions.

	Temperature	Salinity	Water column structure	Circulation
Model drift in downscaled pre-industrial control simulation, ens_ctrl	**Excellent:** SST and NBT trends statistically insignificant in all shelf regions Absolute trends <0.02 °C/decade in all shelf regions	**Good:** SSS and NBS trends statistically insignificant in all shelf regions except English Channel where there is a trend of 0.0196 psu/decade Absolute trends <0.02 psu/decade in all shelf regions except Celtic Sea	N/A	N/A
Qualitative assessment of spatial pattern of mean state from ens_00	**Excellent:** ens_00 captures the main observed temperature features of the NWE shelf seas including: ● the large-scale temperature patterns ● the shelf break current in winter SST ● mixed and stratified regions in NBT	**Good:** ens_00 captures the main observed salinity features of the NWE shelf seas including: ● the large-scale salinity patterns ● low salinity coastal regions ● higher salinity in the centre of the North Sea There are disagreements in the amount of freshening in the southern North Sea and Norwegian Trench	N/A	**Qualitatively Good:** We conclude that the configuration of the NWE shelf seas is qualitatively correct
Quantitative assessment of spatial pattern of mean state with Taylor diagrams	**Excellent:** SST and NBT: $r > 0.9$ for all members of ens_xx $rrms < 0.5$ for all members of ens_xx	**Good:** SSS: $r > 0.7$ for all members of ens_xx $rrms < 0.9$ for all members of ens_xx NBS: $r > 0.8$ for all members of ens_xx $rrms < 0.75$ for all members of ens_xx	N/A	N/A
Low frequency variability	Low frequency variability estimated from ens_xx spread is generally significantly greater than that from ens_ctrl	Low frequency variability estimated from ens_xx spread is generally consistent with that from ens_ctrl	PEA: Low frequency variability estimated from ens_xx spread is consistent with, or greater than, that from ens_ctrl MLD: Low frequency variability estimated from ens_xx spread is generally significantly less than that from ens_ctrl	N/A
Mean climate	**SST: Good** Summer: OSTIA and ERA_sim tend to agree with ens_xx Winter: differences between OSTIA/ERA_sim and ens_xx could be due to NV or climate forcing bias **NBT: Sufficient** Summer: differences between OSTIA/ERA_sim and ens_xx likely due to model bias Winter: differences between OSTIA/ERA_sim and ens_xx could be due to NV or climate forcing bias Care should be taken when using NBT in the vicinity of summer mixing fronts	**Excluding Norwegian Trench, Skagerrak/Kattegat: Sufficient** Differences between EN3/ERA_sim and ens_xx likely due to climate forcings, but also consistent with NV Care should be used when using he absolute values **Norwegian Trench, Skagerrak/Kattegat: Poor** SSS in this region are affected by a significant bias associated with the Baltic Outflow, and even the modelled relative changes should only be used with extreme care. NBS in this region are often of an Atlantic origin, and so less biased	**Excluding Norwegian Trench: Sufficient** Care must be taken when using absolute values of MLD as there is a systematic bias, associated with the downscaling model **Norwegian Trench, Skagerrak/Kattegat: Poor** MLD and PEA in this region are affected by a significant salinity bias, associated with the Baltic Outflow, and even the modelled relative changes should only be used with extreme care	**Inner shelf currents: Good** Cross-sections: 01; 02; 03; 04; 05; 09; 10 ens_xx seasonal cycle generally overlaps with the observed value for at least one month for one ensemble member **Outer shelf currents: Sufficient** Cross-sections: 06; 07; 08; 11; 12 There ens_xx transport is often biased for the entire ensemble. The transport is always of the correct direction **Overall shelf-wide circulation: Qualitatively Good**

High-frequency (interannual) variability	**Summer SST/NBT: Good** There is a general agreement between observed and ens_xx interannual variability on the shelf **Winter SST: Poor** OSTIA interannual variability is greater then ens_xx **Winter NBT: Sufficient** EN3 interannual variability is in agreement with ens_xx away from the south eastern North Sea	**Outside Norwegian Trench, Skagerrak/Kattegat: SSS: Sufficient** There is a general agreement between observed and ens_xx interannual variability on the shelf, although in summer the influence from the Baltic spreads further west from the Norwegian Trench **NBS: Good** There is a general agreement between observed and ens_xx interannual variability on the shelf	**Outside Norwegian Trench, Skagerrak/Kattegat: Poor** Interannual variability of EN3/ERA_sim consistently too high compared to ens_xx	N/A

Sea (MCCIP, 2013; HadISST1.1) from 1983 to 2012, is greater than the 95th percentile value (for a 30-year period) for the southern North Sea of 0.2°/decade.

By analysing the low-frequency variability in the downscaled pre-industrial control simulation, we are able to quantify the low-frequency unforced natural variability on the scale of 28-year mean periods. This was found to be generally small compared to the ensemble spread. However, as the ensemble was a perturbed parameter ensemble rather than an initial conditions ensemble, it includes this additional source of variability. Furthermore, the length of the simulation only allowed 5 samples from which to estimate this low-frequency variability, while the coarser resolution of the atmospheric forcings may reduce the simulated variability. A future study describing the range of variability from downscaled long (>500 year) pre-industrial control simulation (with high resolution atmospheric forcings) for the NWE shelf seas would be a very useful addition to the literature.

As the observed reality is a single realisation of an unknown possible distribution of real-world states, a model driven by un-initialised GCM model output, which is a random sample from such a distribution, cannot be expected to match the observations. However, as we have an estimate of this distribution from the modelled world, we are able to assess where the observations fit within the model distribution. Our simulation driven by ERA40 forcings (ERA_sim), which should have the same phase of NV as the observations, provides us with additional information. By comparing the position of the observations and ERA_sim within the ens_xx modelled distribution, we are able at make qualitative assessments as to whether model bias or NV is the likely cause of any differences with the modelled distribution. Although we are not able to make as solid statements of the model performance, we feel that this approach is more realistic summary of the available information.

By separating the modelled mean state, high- and low-frequency variability, we are able to consider how well the modelled distribution of each compares to the observations and to a reanalysis driven simulation.

The results suggest that the observations (and reanalysis driven simulation) and modelled mean are generally consistent, although there are discrepancies for some regions, seasons and variables – these tend to be attributed to differing phases in NV or a range of biases (relating to the climate forcings, downscaling technique, etc.). Generally the temperature results are consistent, while there is a widespread salinity bias particularly in the Norwegian Trench, which affects the water column structure variables in this region. This gives confidence in using the modelling system for projections of mean temperatures, but care should be taken when using the absolute projected salinity. Use of Norwegian Trench salinity results (and the related water column structure variables) should be avoided, as the present day bias combined with the limitation of using a climatology Baltic boundary condition reduce their credibility. Outside the Norwegian Trench the water column structure products are considered sufficient as the basis of climate projections, however care should be taken with absolute projections of MLD due to a bias associated with the shelf seas model.

The modelled and observed temperature and salinity high-frequency (inter-annual) variability is generally consistent on the shelf (excluding the Skagerrak/Kattegat and the Norwegian Trench). Care must be taken when using the modelled interannual variability for the temperature and salinity from the Skagerrak/Kattegat and the Norwegian Trench (due to the use of a climatology Baltic boundary condition) or the water column structure across the shelf.

The most important question that this study attempts to answer is whether the described modelling system is suitable as the basis of future climate projections. We have found that the

modelling system generally performs well, and we conclude that it is sufficient as the basis of future climate projections. However, we have identified certain circumstances (particular regions, variables, seasons, etc.) where care must be taken when using the model output. Generally, we have been able to ascribe possible reasons for poor model performance in such instances. For convenience, we tabulate the where the modelling system is sufficient, and where care is needed in using the model outputs (Table 5).

Tinker et al. (2015) uses this modelling system (and the simulations described within) to look at how differ the NWE shelf seas will be at the end of the 21st century. Later papers will investigate: the relationship between present day bias and rate of change; the temporal evolution of the NWE shelf seas; when the climate signal will emerge from the natural variability, and use this as the basis of near-future projections.

Acknowledgements

References

Ådlandsvik, B., 2008. Marine downscaling of a future climate scenario for the North Sea. Tellus A 60, 451–458.

Ådlandsvik, B., Bentsen, M., 2007. Downscaling a twentieth century global climate simulation to the North Sea. Ocean Dynamics 57, 453–466.

Bell, M.J., Forbes, R.M., Hines, A., 2000. Assessment of the FOAM global data assimilation system for real time operational ocean forecasting. Journal of Marine Systems 25, 1–22.

Bhaskaran, B., Ramachandran, A., Jones, R.G., Moufouma-Okia, W., 2011. Regional climate model applications on sub-regional scales over the Indian monsoon region: the role of domain size on downscaling uncertainty. Journal of Geophysical Research. http://dx.doi.org/10.1029/2012JD017956.

Brown, J., Carrillo, L., Fernand, L., Horsburgh, K.J., Hill, A.E., Young, E.F., Medler, K.J., 2003. Observations of the physical structure and seasonal jet-like circulation of the Celtic Sea and St. George's Channel of the Irish Sea. Continental Shelf Research 23, 533–561.

Brown, J., Hill, A.E., Fernand, L., Horsburgh, K.J., 1999. Observations of a seasonal jet-like circulation at the central North Sea cold pool margin. Estuarine, Coastal and Shelf Science 48, 343–355.

Brown, S.J., Murphy, J.M., Sexton, D.M.H., Harris, G.R., 2014. Climate projections of future extreme events accounting for modelling uncertainties and historical simulation biases. Climate Dynamics. http://dx.doi.org/10.1007/s00382-014-2080-1.

Cannaby, H., Hüsrevoğlu, Y.S., 2009. The influence of low-frequency variability and long-term trends in North Atlantic sea surface temperature on Irish waters. ICES Journal of Marine Science 66 (7), 1480–1489. http://dx.doi.org/10.1093/icesjms/fsp062.

Clark, R.T., Brown, S., 2013. Influences of circulation and climate change on european summer heat extremes. Journal of Climate 26, 9621–9632. http://dx.doi.org/10.1175/JCLI-D-12-00740.1.

Collins, M., Booth, B.B.B., Bhaskaran, B., Harris, G.R., Murphy, J.M., Sexton, D.M.H., Webb, M.J., 2011. Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. Climate Dynamics 36 (9–10), 1737–1766. http://dx.doi.org/10.1007/s00382-010-0808-0.

Danielssen, D.S., Edler, L., Fonselius, S., Hernroth, L., Ostrowski, M., Svendsen, E., Talpsepp, L., 1997. Oceanographic variability in the Skagerrak and Northern Kattegat, May–June, 1990. ICES Journal of Marine Science 54, 753–773.

Déqué, M., Rowell, D.P., Lüthi, D., Giorgi, F., Christensen, J.H., Rockel, B., Jacob, D., Kjellström, E., de Castro, M., van den Hurk, B., 2007. An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. Climatic Change 81 (Suppl. 1), 53–70. http://dx.doi.org/10.1007/s10584-006-9228-x.

Dye, S.R., Holliday, N.P., Hughes, S.L., Inall, M., Kennington, K., Smyth, T., Tinker, J., Andres, O., Beszczynska-Möller, A., 2013a. Climate change impacts on the waters around the UK and Ireland: salinity. MCCIP Science Review. http://dx.doi.org/10.14465/2013.arc07.060-066.

Dye, S.R., Hughes, S.L., Holliday, N.P., Kennedy, J., Berry, D.I., Kent, E.C., Inall, M., Kennington, K., Smyth, T., Nolan, G., Tinker, J., Andres, O., Beszczynska-Möller, A., 2013b. Climate change impacts on the waters around the UK and Ireland: temperature (air and sea). MCCIP Science Review. http://dx.doi.org/10.14465/2013.arc01.001-012.

Fekete, B.M., Vorosmarty, C.J., Lammers, R.B., 2001. Scaling gridded river networks for macroscale hydrology: development, analysis, and control of error. Water Resources Research 37 (7), 1955–1967.

Fernand, L., Nolan, G.D., Raine, R., Chambers, C.E., Dye, S.R., White, M., Brown, J., 2006. The Irish coastal current: a seasonal jet-like circulation. Continental Shelf Research 26, 1775–1793.

Flather, R.A., 1976. A tidal model of the northwest European continental shelf. Mémoires Société Royale des Sciences de Liège 10 (6), 141–164.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., Rummukainen, M., 2013. Evaluation of climate models. In: Stocker, T.F., Qin, D., Plattner, G.-K., et al. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Friocourt, Y.F., Skogen, M.D., Stolte, W., Albretsen, J., 2012. Marine downscaling of a future climate scenario in the North Sea and possible effects on dinoflagellate harmful algal blooms. Food Additives & Contaminants: Part A 29 (10), 1630–1646. http://dx.doi.org/10.1080/19440049.2012.714079.

Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F.B., Wood, R.A., 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics 16 (2–3), 147–168.

Gregory, J.M., Banks, H., Stott, P.A., Lowe, J.A., Palmer, J.R., 2004. Simulated and observed decadal variability in ocean heat content. Geophysical Research Letters 31, L15312. http://dx.doi.org/10.1029/2004GL020258.

Gröger, M., Maier-Reimer, E., Mikolajewicz, U., Moll, A., Sein, D., 2013. NW European shelf under climate warming: implications for open ocean – shelf exchange, primary production, and carbon absorption. Biogeosciences 10, 3767–3792. http://dx.doi.org/10.5194/bg-10-3767-2013.

Harris, G.R., Sexton, D.M.H., Booth, B.B.B., Collins, M., Murphy, B.H., 2013. Probabilistic projections of transient climate change. Climate Dynamics 40 (11–12), 2937–2972. http://dx.doi.org/10.1007/s00382-012-1647-y.

Holt, J., Butenschön, M., Wakelin, S., Artioli, Y., Allen, I., 2012a. Oceanic controls on the primary production of the northwest European continental shelf: model experiments under recent past conditions and a potential future scenario. Biogeosciences 9, 97–117.

Holt, J., Hughes, S., Hopkins, J., Wakelin, S., Holliday, N.P., Dye, S., Gonzalez, P., Hjøllo, S., Mork, K.A., Nolan, G., Proctor, R., Read, J., Shammon, T., Sherwin, T., Smyth, T., Tattersall, G., Ward, B., Wiltshire, K., 2012b. Multi-decadal variability and trends in the temperature of the northwest European continental shelf: a model-data synthesis. Progress in Oceanography 106, 96–117.

Holt, J., Proctor, R., 2008. The seasonal circulation and volume transport on the northwest European continental shelf: a fine-resolution model study. Journal of Geophysical Research – Oceans 113 (C6), C06021. http://dx.doi.org/10.1029/2006jc004034.

Holt, J., Schrum, C., Cannaby, H., Daewel, U., Allen, I., Artioli, Y., Bopp, L., Butenschön, M., Fach, B., Harle, J., Pushpadas, D., Salihoglu, B., Wakelin, S., 2014. Physical processes mediating climate change impacts on regional sea ecosystems. Biogeosciences Discussion 11, 1909–1975. http://dx.doi.org/10.5194/bgd-11-1909-2014.

Holt, J., Wakelin, S., Lowe, J.A., Tinker, J., 2010. The potential impacts of climate change on the hydrography of the northwest European continental shelf. Progress in Oceanography 86 (3–4), 361–379. http://dx.doi.org/10.1016/j.pocean.2010.05.003.

Holt, J.T., James, I.D., 2001. An s coordinate density evolving model of the northwest European continental shelf – 1. Model description and density structure. Journal of Geophysical Research – Oceans 106 (C7), 14015–14034.

Holt, J.T., James, I.D., Jones, J.E., 2001. An s coordinate density evolving model of the northwest European continental shelf 2. Seasonal currents and tides. Journal of Geophysical Research – Oceans 106 (C7), 14035–14053.

Ingleby, B., Huddleston, M., 2007. Quality control of ocean temperature and salinity profiles – historical and real-time data. Journal of Marine Systems 65, 158–175. http://dx.doi.org/10.1016/j.jmarsys.2005.11.019.

IPCC, 2001. Climate change 2001: the scientific basis. In: Griggs, D.J., Noguer, M. (Eds.), Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, p. 881.

IPCC, 2007. Climate change 2007: the physical basis. In: Solomons, S., Qin, D., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Contributions of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York, NY, USA, p. 996.

Jones, R.G., Murphy, J.M., Noguer, M., 1995. Simulation of climate change over Europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. Quarterly Journal of the Royal Meteorological Society 121, 1413–1449. http://dx.doi.org/10.1002/qj.49712152610.

Jones, R.G., Murphy, J.M., Noguer, M., Keen, A.B., 1997. Simulation of climate change over Europe using a nested regional-climate model. II: Comparison of driving and regional model responses to a doubling of carbon dioxide. Quarterly Journal of the Royal Meteorological Society 123, 265–292. http://dx.doi.org/10.1002/qj.49712353802.

Jones, R.G., Noguer, M., Hassell, D.C., Hudson, D., Wilson, S.S., Jenkins, G.J., Mitchell, J.F.B., 2004. Generating High Resolution Climate Change Scenarios Using PRECIS. Met Office Hadley Centre, Exeter, UK.

Knight, J.R., Allan, R.J., Folland, C.K., Vellinga, M., Mann, M.E., 2005. A signature of persistent natural thermohaline circulation cycles in observed climate. Geophysical Research Letters 32 (20). http://dx.doi.org/10.1029/2005GL024233.

Lowe, J.A., Howard, T.P., Pardaens, A., Tinker, J., Holt, J., Wakelin, S., Milne, G., Leake, J., Wolf, J., Horsburgh, K., Reeder, T., Jenkins, G., Ridley, J., Dye, S., Bradley, S., 2009. UK Climate Projections Science Report: Marine and Coastal Projections. Met Office Hadley Centre, Exeter, UK.

Mathis, M., Mayer, B., Pohlmann, T., 2013. An uncoupled dynamical downscaling for the North Sea: method and evaluation. Ocean Modelling 72, 153–166.

MCCIP, 2012. In: Buckley, P.J., Dye, S.D., Frost, M., Wallace, C.J. (Eds.), Marine Climate Change Impacts Knowledge Gaps, Summary Paper. Lowestoft, p. 8.

Meier, H.E.M., 2006. Baltic Sea climate in the late twenty-first century: a dynamical downscaling approach using two global models and two emission scenarios. Climate Dynamics 27, 39–68. http://dx.doi.org/10.1007/s00382-006-0124-x.

Murphy, J.M., Sexton, D.M.H., Jenkins, G.J., Boorman, P.M., Booth, B.B.B., Brown, C.C., Clark, R.T., Collins, M., Harris, G.R., Kendon, E.J., Betts, R.A., Brown, S.J., Howard, T.P., Humphrey, K.A., McCarthy, M.P., McDonald, R.E., Stephens, A., Wallace, C., Warren, R., Wilby, R., Wood, R.A., 2009. UK Climate Projections Science Report: Climate Change Projections. Met Office Hadley Centre, Exeter.

O'Dea, E., Arnold, A.K., Edwards, K.P., Furner, R., Hyder, P., Martin, M.J., Siddorn, J., Storkey, D., While, J., Holt, J., Lui, H., 2012. An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European North-West shelf. Journal of Operational Oceanography 5 (1), 3–17.

Oki, T., Nishimura, T., Dirmeyer, P., 1999. Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP). Journal of the Meteorological Society of Japan 77 (1B), 235–255.

Oki, T., Sud, U.C., 1998. Design of total runoff integrating pathways (TRIP) – a global river channel network. Earth Interactions 2 (1), 1–36.

Olbert, A.I., Dabrowski, T., Nash, S., Hartnett, M., 2012. Regional modelling of the 21st century climate changes in the Irish Sea. Continental Shelf Research 41, 48–60. http://dx.doi.org/10.1016/j.csr.2012.04.003.

Pinnegar, J., Watt, T., Kennedy, K., 2012. Climate Change Risk Assessment for the Marine and Fisheries Sector, Defra.

Pope, V.D., Gallani, P.R., Rowntree, P.R., Stratton, R.A., 2000. The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. Climate Dynamics 16 (2–3), 123–146. http://dx.doi.org/10.1007/s003820050009.

Prandle, D., Ballard, G., Flatt, D., Harrison, A.J., Jones, S.E., Knight, P.J., Loch, S., McManus, J.P., Player, R., Tappin, A., 1996. Combining modelling and monitoring to determine fluxes of water, dissolved and particulate metals through the Dover Strait. Continental Shelf Research 16, 237–257.

Proctor, R., Holt, J.T., Allen, J.I., Blackford, J., 2003. Nutrient fluxes and budgets for the North West European Shelf from a three-dimensional model. Science of the Total Environment 314, 769–785. http://dx.doi.org/10.1016/s0048-9697(03)00083-4.

Pugh, D., 1987. Tides, Surges and Mean Sea-Level. John Wiley & Sons, Chichester.

Pugh, D., 2008. Socio-economic Indicators of Marine-related Activities in the UK Economy. The Crown Estate, p. 68.

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C., Kaplan, A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late Nineteenth Century. Journal of Geophysical Research 108 (D14), 4407. http://dx.doi.org/10.1029/2002JD002670.

Roberts-Jones, J., Fieldler, E., Martin, M.J., 2012. Daily, global, high-resolution SST and sea ice reanalysis for 1985–2007 using the OSTIA system. Journal of Climate 25 (18), 6215–6232. http://dx.doi.org/10.1175/JCLI-D-11-00648.1.

Rougier, J., Sexton, D.M.H., Murphy, J.M., Stainforth, D., 2009. Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. Journal of Climate 22, 3540–3557.

Sanderson, M., Wiltshire, A.J., Betts, R.A., 2012. Projected changes in water availability in the United Kingdom. Water Resources Research 48 (8), W08512. http://dx.doi.org/10.1029/2012WR011881.

Schrum, C., Hübner, U., Jacob, D., Podzun, R., 2003. A coupled atmosphere/ice/ocean model for the North Sea and Baltic Sea. Climate Dynamics 21 (2), 131–151.

Shuckburgh, E.F., 2012. Oceanographers' contribution to climate modelling and prediction: progress to date and a future perspective. Philosophical Transactions of the Royal Society A 370 (1980), 5656–5681.

Simpson, J.H., Bowers, D., 1981. Models of stratification and frontal movement in shelf seas. Deep Sea Research Part A. Oceanographic Research Papers 28 (7), 727–738.

Skogen, M.D., Svendsen, E., Berntsen, J., Aksnes, D., Ulvestad, K.B., 1995. Modelling the primary production in the North Sea using a coupled three-dimensional physical–chemical–biological ocean model. Estuarine, Coastal and Shelf Science 41, 545–565.

Svendsen, E., Saetre, R., Mork, M., 1991. Features of the northern North Sea circulation. Continental Shelf Research 11, 493–508.

Tinker, J., Lowe, J., Holt, J., Pardaens, A., Barciela, R. (2015). The uncertainty in climate projections for the 21st century NW European shelf seas. Progress in Oceanography (submitted for publication).

Turrell, W.R., Henderson, E.W., Slesser, G., Payne, R., Adams, R.D., 1992. Seasonal changes in the circulation of the northern North Sea. Continental Shelf Research 12, 257–286.

Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V.D.C., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., Berg, L.V.D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., Mcnally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological Society 131, 2961–3012. http://dx.doi.org/10.1256/qj.04.176.

Vorosmarty, C.J., Fekete, B.M., Meybeck, M., Lammers, R.B., 2000a. Geomorphometric attributes of the global system of rivers at 30-minute spatial resolution. Journal of Hydrology 237 (1–2), 17–39.

Vorosmarty, C.J., Fekete, B.M., Meybeck, M., Lammers, R.B., 2000b. A simulated topological network representing the global system of rivers at 30-minute spatial resolution (STN-30). Global Biogeochemical Cycles 14, 599–621.

Wakelin, S., Holt, J., Blackford, J., Allen, I., Butenschön, M., Artioli, Y., 2012. Modeling the carbon fluxes of the northwest European continental shelf: validation and budgets. Journal of Geophysical Research – Oceans 117 (C5). http://dx.doi.org/10.1029/2011JC007402.

Wakelin, S.L., Holt, J.T., Proctor, R., 2009. The influence of initial conditions and open boundary conditions on shelf circulation in a 3D ocean-shelf model of the North East Atlantic. Ocean Dynamics 59 (1), 67–81. http://dx.doi.org/10.1007/s10236-008-0164-3.