EMI R&D PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2026
Project Title:	Enhancing regional and global ocean data assimilation
Computer Project Account:	spitstor
Principal Investigator(s):	Andrea Storto
Affiliation:	CNR ISMAR, Rome, Italy
Name of ECMWF scientist(s) collaborating to the project (if applicable)	NA
Start date of the project:	2025
Expected end date:	2027

Computer resources allocated/used for the current year and the previous one

(if applicable)
Please answer for all project resources

		Previous year		**Current year**	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	9M	9M	9M	9M
Data storage capacity	(Gbytes)				

Summary of project objectives (10 lines max)

The overarching goal of this project, which is in full continuity with the previous Special Project run by the P.I. (spitstor-2019, spitstor-2022) is to allow us to experiment several new ensemble generation and assimilation techniques, which are well aligned with the priorities outlined by the ocean data assimilation community and may also contribute to advancing the ocean data assimilation science at European level. The proposal is articulated in different tasks: 1. Enhancing and updating stochastic physics schemes for NEMO, extending the STOPACK package. 2. Coupled data assimilation experiments with the MESMAR system, formed by NEMO, WRF and HD models and covering the Mediterranean region. 3. Data assimilation and bias correction for historical reanalyses, focussing on use of hybrid covariances and climatological and flow-dependent large-scale mode bias correction methods.

Summary of problems encountered (10 lines max)

None

Summary of plans for the continuation of the project (10 lines max)

The continuation of the project will primarily focus on optimal methods for hybrid ensemble-variational data assimilation, with particular emphasis on using coarser-resolution ensemble simulations to compute background-error covariances for higher-resolution analysis and prediction systems.

List of publications/reports from the project with complete references

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
………………………………

Summary of results

Attribution of ocean reanalysis biases to atmospheric forcing

A key open question in ocean reanalysis development is how much of the residual model-observation bias originates from the prescribed atmospheric forcing rather than from ocean model physics or the data assimilation scheme itself. Two small ensemble reanalyses were run, differing only in atmospheric forcing, 20CRv3 (long-record, loosely constrained) versus ERA5 (modern, observationally rich), over 2001-2015, validated against EN4. These included two 4-member ensembles, both run for the ERA5-20CR overlapping period from 1991 to 2015 (earlier periods were considered too data-sparse and thus less relevant). Each of the two ensembles was generated by perturbing model physics, input observations, and bulk formulas, thus allowing for a statistically robust assessment of the net impact of the atmospheric forcing on the systematic model errors.

Several diagnostics were calculated and analysed: analysis-increment differences, bias/MAE against EN4, and the percentage of temperature bias explained by forcing as a function of latitude and depth. The latter is shown in Figure 1; the cross-section indicates that the forcing explains a large fraction of the reanalysis bias in the tropics, coherently extending into the upper thermocline along the equatorial waveguides and into polar regions where sea-ice/surface-flux processes dominate, while mid-latitudes are comparatively insensitive. This points to forcing data quality as the priority for tropical/polar bias reduction, and ocean-model/assimilation development as the priority elsewhere. This work was supported by the Copernicus Marine Service through the Service Evolution project ORACS.

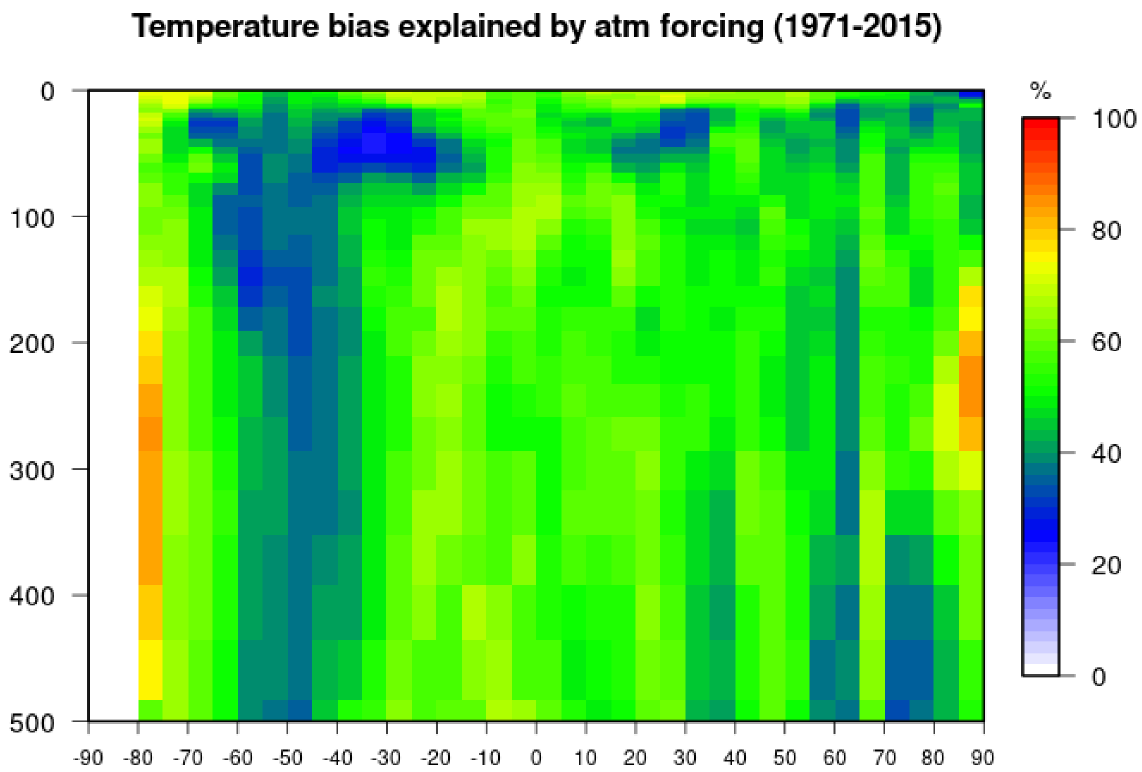

Figure 1. Percentage of bias explained by the atmospheric forcing, as a function of depth and latitude.

Impact of an LSTM-corrected river discharge dataset on NEMO ocean simulations

A new LSTM-corrected GloFAS discharge dataset (41 rivers, daily, gap-free) was developed by Magellium within the framework of the Copernicus Marine Service F3O project, and tested here in six free-running NEMO4.0.7/SI3 ORCA1 simulations (1991-2023, mimicking the CIGAR ocean reanalysis), isolating climatological vs. corrected/time-varying river forcing, using a purpose-built remapping tool for grid injection.
By validating SSS (sea surface salinity) against EN4 as a function of distance from the coast (Figure 2), the enhanced dataset (HYBRID/HYBRID-CLIM) gives the largest, most persistent RMSE reduction, remaining positive out to 2000-3000 km offshore, far beyond the immediate river-mouth influence, while raw GloFAS and JRA55 underperform. Most of the improvement comes from mitigating fresh biases present when using the widely adopted river runoff monthly climatology from Dai and Trenberth (2004). Impact of the interannual variability is generally negligible. The resulting denser surface ocean weakens stratification globally and modestly strengthens the Atlantic gyres, ACC, and AMOC (0.5-3 Sv, 0.8-1 Sv, respectively; not shown).

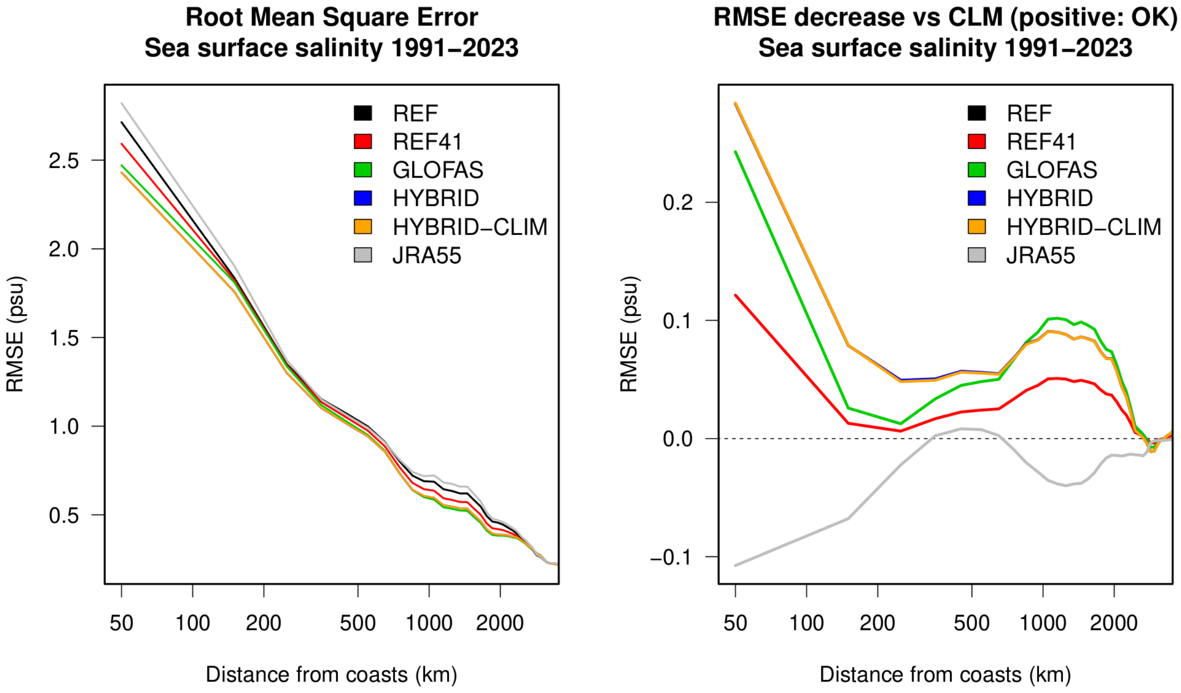

Figure 2. RMSE and RMSE decrease for several ocean simulation experiments with different river runoff datasets, as explained in the text, as a function of the distance from the shoreline.

Development of the ocean stochastic physics package (STOPACK)

Building on STOPACK (Storto and Andriopoulos, 2021), used to stochastically perturb sea-ice and ocean parameters within ECMWF's coupled IFS-NEMO4/SI3 system, three areas of methodological development were carried out.

Perlin noise: the random-field generator was upgraded from the iterated Shapiro-filter approach (cost scaling with the square of the number of filter passes) to multi-scale Perlin noise, controlled via base spatial scale, octaves, lacunarity and persistence, with an explicit relationship to grid spacing and e-folding correlation length derived for tuning. This made it computationally affordable to draw an independent random field per perturbed parameter rather than sharing one field across all of them, leading to more physically consistent perturbations, also aligned with the atmospheric SPP approach.

New sea-ice and liquid-ocean SPP parameters: the default nine perturbed sea-ice parameters (strength, ridging/rafting, drag, thermodynamics, salinity restoring, melt-pond flushing) were complemented with new (currently unused) albedo perturbation modes and a new-ice-thickness option. In parallel, an analogous liquid-ocean parameter set was implemented (lateral diffusivity/viscosity, Langmuir-cell TKE coefficient, SST/SSS surface relaxation, vertical mixing, 3D damping, solar penetration, air-sea exchange), plus new TKE-profile-shape and wave-ocean-coupling perturbation options.

SKEB: the Stochastic Kinetic Energy Backscatter scheme (re-injecting dissipated numerical/convective/eddy energy at larger scales) was brought to a stable, documented configuration for NEMO4/eORCA025, with a conservative reference namelist shown to run multi-year forced integrations without the previously encountered instabilities.

These STOPACK developments are presently being tested within ECMWF's coupled forecasting system, and part of them are documented on internal Confluence pages and partly supported by the HE ACCIBERG and ERC PREDDYCT projects.

May 2026