https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

pypfda v1.0: An open-source, model-agnostic particle-filter engine for online sequential paleo data assimilation in Earth system models

Bijan Fallah[1], Hugues Goosse[2], and Willem Huiskamp[1]

[1]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany
[2]Université catholique de Louvain, Earth and Life Institute, Place Louis Pasteur 3, B-1348 Louvain-la-Neuve, Belgium

Correspondence: Bijan Fallah (fallah@pik-potsdam.de)

Abstract. Data assimilation combines observations with the dynamics of a model to estimate the state of a system; in paleoclimatology it merges proxy records with climate models. Most reconstructions use the offline approach, in which the assimilated state does not persist, so slow processes such as the Atlantic overturning are not carried through time. We present `pypfda` v1.0, an open-source Python engine that runs an online particle filter inside a coupled Earth system model. At each analysis, ensemble members are weighted by their fit to the proxies, and the best-fitting members are copied and perturbed before the simulation continues, which preserves the ocean's multiyear memory without offline calibration or surrogate models. The engine connects to any model through four functions, so adding a model requires recompiling neither the model nor the filter. We tested it with the coupled model CM2Mc-BLING in twin experiments, in which 59 synthetic sea surface temperature proxies drawn from one model run constrain a 100-member ensemble that reconstructs the Atlantic Meridional Overturning Circulation (AMOC), a quantity the proxies do not observe directly. Skill appears only when the members start from diverse ocean states and when the analysis interval matches the ocean's timescales; a five-year interval works best. In a 300-year experiment, the reconstructed AMOC correlates with the truth at $r = +0.78$, while an ensemble without assimilation shows no correlation. Copying members at each analysis concentrates the ensemble's ancestry, and the 300-year run converges to a single lineage after about 155 years, while skill is maintained because each analysis realigns the ensemble with the proxies. Because the observation operator is a separate module, proxy system models can be added for reconstructions with real proxies.

1 Introduction

Two main methods reconstruct past climate, each with important drawbacks. The first uses scattered proxy records, such as corals, tree rings, and sediments. These records are unevenly distributed in space and time. They carry dating uncertainties and rest on the assumption that the relationship between proxy and climate variable stays constant. The second method uses climate model simulations, which range from atmosphere-only and coupled ocean–atmosphere general circulation models to Earth system models (ESMs) that also represent components such as ocean biogeochemistry. These simulations are based on physical laws but are affected by structural and parametric uncertainties. Data assimilation (DA) combines these two approaches to generate spatially complete and temporally continuous reconstructions of a climatological field of interest (Tierney et al., 2025).

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Offline data assimilation is used in most current paleoclimate reconstructions (Hakim et al., 2016; Steiger et al., 2018; Tierney et al., 2025). With this method, a previously computed set of model simulations is statistically combined with proxy observations at each analysis time. The assimilated state does not influence the model's later development (Hakim et al., 2016; Steiger et al., 2018), so different analysis states can arise from completely separate model trajectories, and no new information can be incorporated between assimilation windows. As shown by the Last Millennium Reanalysis (Hakim et al., 2016) and the Paleo Hydrodynamics Data Assimilation product (Steiger et al., 2018), offline data assimilation reconstructs global temperature well, particularly with proxy system models, the forward models that translate simulated climate into the quantity a proxy records, and a large proxy database.

Offline data assimilation removes the temporal continuity offered by coupled models, which in turn hinders reconstruction of long-term memory processes, such as ocean heat content, Southern Ocean upwelling (Goosse, 2017), and the Atlantic Meridional Overturning Circulation (AMOC). In contrast, online assimilation transfers information from long time-scale processes from one analysis interval to the next, as shown in coupled atmosphere-ocean online reconstructions (Perkins and Hakim, 2021). Each correction is then carried forward by the model's own dynamics, so the state stays physically consistent from one cycle to the next (Tierney et al., 2025). Online assimilation also lets observations of one component constrain the others through coupled physics of the model (Okazaki et al., 2021).

Despite these advantages, online data assimilation has had limited use in paleoclimatology (Goosse et al., 2012; Tierney et al., 2025), and the main obstacle is cost: every analysis requires the whole ensemble of coupled-model integrations to be run forward again to the next observation time, whereas an offline reconstruction reuses one existing ensemble at every time step. Online assimilation is carried out by a sequential filter (van Leeuwen et al., 2019). An ensemble of model simulations is run forward to the next observation time; each member is then corrected using the observations, and the corrected members are run forward again. Ensemble Kalman filters make this correction by shifting every member toward the observations, using error covariances estimated from the ensemble (Evensen, 1994). They assume that errors follow a Gaussian distribution, which can fail to capture the highly non-linear dynamics and possible state transitions of, for example, the AMOC. A particle filter instead weights each member, or particle, by how well it fits the observations and then resamples: the ensemble is replaced by copies of the well-fitting particles, and the others are discarded (Gordon et al., 1993). Because a weighted ensemble can represent any distribution, particle filters do not make this assumption. Resampling has a known cost, called degeneracy, which takes two forms (van Leeuwen et al., 2019). When the observations are many or precise relative to the ensemble size, nearly all the weight can fall on a single particle, so that the ensemble no longer represents the uncertainty. This is weight degeneracy. Even when the weights stay well spread at every analysis, repeated resampling removes lineages one by one until all members descend from a single ancestor. This slower form is path degeneracy.

Particle-filter-based online assimilation has been used in paleoclimatology for more than ten years (Goosse et al., 2012). Online ensemble Kalman filtering has also been applied to intermediate-complexity Earth system models (Masoum et al., 2024). Meng et al. (2025) created a reanalysis at seasonal resolution by carrying out online assimilation using a coupled linear inverse model. The first systematic comparison of online and offline assimilation in paleoclimatology showed that both methods have similar skill at continental and hemispheric scales (Matsikaris et al., 2015). Later, Acevedo et al. (2017) compared online

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

and offline ensemble Kalman filters using the SPEEDY atmospheric model with a slab ocean and found no benefit to the online method, consistent with the lack of ocean memory in slab-ocean setups. More recently, Sun et al. (2025) replaced the ESM forecast step with a deep learning emulator and showed that online cycling can outperform offline versions. Jebri and Khodri (2023) used a particle filter with a linear inverse model as an ESM emulator and showed that large particle ensembles, cheap to propagate through the emulator, can reconstruct skillfully without filter degeneracy.

The most recent studies (Sun et al., 2025; Jebri and Khodri, 2023) replace the coupled ESM altogether with a statistical or machine-learning surrogate. The closest fully online precedent is the pseudo-proxy experiment of Meng and Hakim (2024), which assimilates proxy-like sea surface temperature observations online into a deep-learning emulator of the tropical Pacific upper ocean.

None of the online studies above produced reusable software, because each implementation was tied to its model. The LOVECLIM particle filter (Dubinkina et al., 2011; Dubinkina and Goosse, 2013; Goosse et al., 2012) works entirely online within a coupled Earth system model of intermediate complexity (EMIC), but was developed solely for that host model, and the equivalent-weights particle filter has been cycled online in the coupled general circulation model (GCM) HadCM3 through the MPI-coupled EMPIRE framework (Browne and Wilson, 2015; Browne and van Leeuwen, 2015). Several general data-assimilation frameworks exist. The Parallel Data Assimilation Framework (PDAF) (Nerger and Hiller, 2013) and the Data Assimilation Research Testbed (DART) (Anderson et al., 2009) are compiled libraries that couple to a model either in memory, which requires the model to be built against the library, or through file exchange controlled by shell scripts. AWI-CM-PDAF (Nerger et al., 2020) is an online coupled ensemble system built on PDAF. Both libraries offer particle filters. PDAF includes a resampling particle filter with several resampling schemes and a nonlinear ensemble transform filter, and DART provides similar choices. PDAF has recently gained a Python interface, pyPDAF (Chen et al., 2025), which keeps the filter algorithms in compiled Fortran, cycles Python models in memory, and reaches compiled models through their restart files. DAPPER (Raanes et al., 2024) is a lightweight experimental package that compares particle-filter methods on idealized systems and is not linked to Earth system models. In paleoclimatology, applications of these frameworks have mainly used ensemble Kalman methods. ParticleDA.jl (Giles et al., 2024) and CrocO (Cluzet et al., 2021) are dedicated particle-filter packages with general model interfaces, the first written for distributed filtering on idealized and geophysical testbeds, the second for snowpack assimilation. The DASH toolbox (King et al., 2023) and the `cfr` package (Zhu et al., 2024) serve paleoclimatology directly and combine proxy forward models with mainly offline ensemble techniques; neither drives a running model between analyses. In short, the existing frameworks either require the model to be built against a compiled library, keep the filter in compiled code, or are offline by design. An engine that cycles an arbitrary coupled ESM online from Python through a minimal interface, without linking or recompiling the model, and that has been tested with a particle filter at paleoclimate cycle lengths, does not yet exist.

This study presents `pypfda` v1.0, an open-source online particle-filter engine that couples directly to Earth system models without offline calibration, surrogate models, or previously computed priors. The engine couples to any model through a minimal common interface (forecast, observe, get/set state, inflate), called from Python in the same process. Adding a model therefore requires recompiling neither the model nor the filter, and one filter serves models of any complexity. The package

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

also contains two fast–slow Lorenz testbeds that run in the same process; one is the two-level Lorenz-96 model (Lorenz, 1996; Lorenz and Emanuel, 1998), a standard benchmark for simulation methods (Grudzien et al., 2020; Haussaire and Bocquet, 2016), the other a coupled fast–slow pair of Lorenz-63 systems (Section S4 in the Supplement). A cluster-side implementation of the same interface drives the coupled GCM CM2Mc-BLING (Section 2.2).

The study also examines ocean-state inflation, that is, small perturbations added to the upper-ocean temperature and salinity of the copied members after each resampling, as a possible remedy for path degeneracy.

As a main example, we use CM2Mc-BLING with an ensemble of 100 members. We use a sequential importance resampling particle filter, the weight-and-resample scheme described above, in its global form, in which all proxies weight each member jointly and no covariance localization limits the reach of an observation (Section 2.3). Localized versions (Poterjoy, 2016) are left to future work (Section 4). We test the system with Observing System Simulation Experiments (OSSEs), a standard test for reconstruction methods (Smerdon, 2012). In such an OSSE, the synthetic truth comes from a known model trajectory, and pseudo-observations are drawn at 59 sea surface temperature (SST) proxy locations. Only SST is assimilated, which deliberately restricts the scope. The observations enter only through the `observe` operator, so the engine is the same whatever variable that operator returns (Section 2.2). This approach separates data assimilation from model error and the uncertainty associated with proxy calibration, setting an upper limit on the skill achievable with real proxies.

We focus on the AMOC because it is the slow, ocean-borne component whose multiyear memory online assimilation is designed to carry forward, and because surface proxies observe it only indirectly. Reconstructing the AMOC is especially difficult because its state at any moment depends heavily on its previous path, as it integrates surface forcing over many decades. The direct observational record is also too short for forced signals to stand out above the background internal variability (McCarthy et al., 2025). The surface-temperature fingerprints that are used as indicators of the AMOC are themselves non-stationary and depend on external forcing (McMonigal et al., 2025). The overturning measured in the subpolar North Atlantic also differs from that observed further south (Petit et al., 2025). These difficulties are why dynamically based reconstructions have to be tested in observing-system experiments before they are used with real proxies.

Four questions decide whether online assimilation is practical for paleoclimate reconstruction. The first is whether surface-ocean proxies can constrain the AMOC, given that the AMOC involves a complex network of surface and deep-ocean currents. The second question is which assimilation cycle length best reconstructs decadal AMOC variations without aliasing the signal. The third is whether the system retains predictive skill when ensemble members are initialized from independent ocean states, simulating real-world situations in which the initial ocean circulation is unknown. The fourth question, the main motivation of this study, is how much genealogical diversity the ensemble retains over a long reconstruction, and whether perturbing the upper-ocean state after each analysis can preserve that diversity while keeping the memory and the skill it provides. Resampling is what gives the online filter its memory: at each analysis it carries the overturning signal of the best-fitting members forward, and this is the mechanism by which online assimilation produces predictive skill for a slow variable such as the AMOC. The same step concentrates the ensemble's ancestry, because it keeps copies of some members and drops the others, so a long reconstruction gradually converges toward a single ancestral lineage. We call this balance the *diversity–memory trade-off*.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

2 Methods

2.1 The CM2Mc-BLING Earth System Model

The assimilation framework uses the CM2Mc Earth system model, a lower-resolution version of the Geophysical Fluid Dynamics Laboratory (GFDL) CM2, intended for computationally efficient simulations on millennial timescales (Galbraith et al., 2011). For the ocean component, we use Modular Ocean Model version 5 (MOM5), with a nominal horizontal resolution of $3°$; this increases to $0.6°$ at the equator to better represent tropical dynamics. The model uses a rescaled vertical pressure coordinate divided into 28 levels. The physical ocean is coupled to the Biogeochemistry with Light, Iron, Nutrients, and Gases (BLING) module (Galbraith et al., 2010), which in this configuration also carries radiocarbon as a tracer and will later allow radiocarbon proxies to be assimilated; $\delta^{13}\mathrm{C}$ is not simulated.

The atmospheric component is GFDL's AM2 model, which has a cubed-sphere finite-volume dynamical core at C48 resolution (about $2° \times 2°$) and 24 vertical levels running from the surface up to about 3 hPa. The atmospheric fields are stored on the model's history grid, a uniform 96×60 longitude–latitude grid ($3.75° \times 3°$).

In this study the AMOC index is the maximum of the annual-mean Atlantic meridional overturning stream function below 500 m at $26.5°$N, the latitude of the RAPID-MOCHA array (McCarthy et al., 2015; Smeed et al., 2018). The same index is used to score the reconstruction and to compare the synthetic truth with the RAPID observations (Section 2.6).

2.2 The Pluggable Assimilation Engine and the Models it Drives

The assimilation engine is independent of the forecast model and is available as a standalone, open-source Python package called `pypfda`. The model, called the *forward model* in the code, advances the state. The online particle filter (Section 2.3) needs the model to do only two things. It must advance the ensemble and return the model equivalent of each observation, for example, sea surface temperature at the proxy sites. The coupling goes through four model-specific methods: `forecast`, `observe`, `get/set_state`, and `inflate`. To integrate a new model, one must implement these four operations along with the member-setup method. All other components of assimilation, such as weighting, resampling, diagnostics, and genealogy tracking, are common to all cases and therefore do not need to be modified. Table 1 gives a full description of this interface, the engine's public entry points, and their default settings in version 1.0. Version 1.0 includes the filter, four resampling schemes, the weight cap, a Gaspari-Cohn localization module (which we do not use in this study), and in-process Lorenz adapters. The package also has subprocess adapters for two additional models, although we did not assess them in this study. It does not include the CM2Mc-BLING orchestrator's scheduling and restart handling.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1: The `pypfda` v1.0 interface. (a) The `ForwardModel` adapter contract through which a model is coupled: the four model-specific operations named in the text (`forecast`, `observe`, `get_state`/`set_state`, `inflate`), the member-setup method, and two further optional hooks. An integer index addresses members, and the engine never inspects model state. (b) The `ParticleFilter` constructor arguments and analysis entry points. (c) The `CycleDriver` arguments and the order of operations in one online cycle. Defaults match v1.0; the settings used for CM2Mc-BLING are given in Section 2.3, and the cluster-side orchestrator that drives CM2Mc-BLING implements the same contract against SLURM job arrays and restart files.

Name	Call and return	Contract
(a) `ForwardModel` adapter, implemented once per model		
`n_members`	attribute, `int`	number of ensemble members; set by the adapter before the driver runs
`initialize_member`	`(member_id, ic_spec)`	prepares a member from an opaque initial-condition specification (a restart path, a control-run year, or a perturbation seed); FREE and DA ensembles receive identical arguments so that assimilation is the only difference between them
`forecast`	`(member_id, window)`	advances the member by one assimilation window and leaves it ready for `observe` and `get_state`; the window mean of the observed field is accumulated here rather than an end-of-window snapshot
`observe`	`(member_id, window)` → array of shape `(n_obs,)`	the observation operator H applied at the fixed proxy network, in an ordering that is stable across members and cycles; a missing proxy is `NaN` and drops out of the likelihood
`get_state`	`(member_id)` → state handle	an independent, clonable snapshot of the member's full state (a copied restart for file-based models, deep-copied arrays in process) that stays valid after the source member is overwritten
`set_state`	`(member_id, state)`	overwrites the member's full state with a parent's snapshot
`inflate` (optional)	`(member_id, amplitude, seed)`	post-resampling perturbation of one member, called on every member with a distinct seed; default no-op, and a zero amplitude leaves the state untouched
`target_diagnostic` (optional)	`(member_id)` → `float`	a scalar recorded per member and cycle for evaluation only, never assimilated (e.g., an AMOC index); default `NaN`

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1, continued.

Name	Call and return	Contract
`finalize` (optional)	`()`	end-of-run cleanup; default no-op
(b) `ParticleFilter`: weights, effective sample size (ESS), resampling		
`ess_threshold`	`float`, default 0.5	resampling is triggered when ESS$/N$ falls below this fraction (τ_{ess}; 0.85 for CM2Mc-BLING)
`resampling`	`systematic` (default), `stratified`, `residual`, `multinomial`	resampling scheme applied when triggered
`max_weight`	`float` or `None` (default)	optional ceiling on the largest normalized weight, with the surplus redistributed among the remaining members; off by default
`rng`	`numpy.random.Generator`	random stream for resampling; created if omitted
`assimilate`	`(ensemble, ensemble_obs, observations, obs_err)` → `(ensemble, info)`	one analysis step under a diagonal Gaussian likelihood with the given observation-error standard deviations (the effective error σ_{eff} of Equation 4)
`assimilate_log_weights`	`(ensemble, log_weights)` → `(ensemble, info)`	the same step from user-supplied log-weights, for non-Gaussian or correlated proxy errors and proxy system models
`info`	`AssimilationInfo`	normalized weights, log-weights, ESS, ESS fraction, weight entropy, resampled flag, and the resampling indices
(c) `CycleDriver`: one online cycle		
`model`, `pf`	adapter and filter	the two objects above
`observations`	callable, `cycle` → `(y, obs_err)`	the observations assimilated at each cycle and their error standard deviation (scalar or per proxy); `NaN` allowed
`n_cycles`, `window`	`int`, `float`	number of cycles and the model-time length of each
`inflation_amplitude`	`float`, default 0	passed to `inflate` after each resampling
`backend`	`SerialBackend` (default) or `ThreadPoolBackend(max_workers)`	how the member forecasts of a cycle are executed
`base_seed`	`int`, default 2000	root of the per-member, per-cycle inflation seeds
`outdir`, `resume`	path (default `None`), `True`	a JSON checkpoint (genealogy, history, completed cycle) written each cycle atomically and resumed from if present

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1, continued.

Name	Call and return	Contract
order of operations	per cycle	forecast all members; observe; NaN-aware Gaussian log-weights; ESS check and resampling; snapshot of every parent before any clone is applied; `inflate` on every member; record ESS, resampling map, effective ancestor size and target diagnostics; checkpoint

The package installs with `pip install pypfda`; the version described here, and used for the idealized testbeds and the equivalence replay below, is v1.0, installable as `pip install "pypfda==1.0"`. The package includes working examples that run on a laptop: the two coupled fast–slow Lorenz benchmarks (two-scale Lorenz-96 and coupled Lorenz-63) described in Section S4 in the Supplement (see also the Code and data availability section). A typical assimilation cycle using the engine is only a few lines of Python (Listing 1).

Listing 1. Minimal use of the `pypfda` particle filter: one model-agnostic assimilation cycle. `forecast` and `observe` are the only model-specific methods (the common interface above); `assimilate` returns the resampled ensemble along with diagnostics (effective sample size, resampling flag). The full `CycleDriver` (not shown) additionally applies the post-resampling inflation kernel and records the ensemble genealogy.

```
from pypfda import ParticleFilter

pf = ParticleFilter(ess_threshold=0.85, resampling="systematic")
for cycle in range(n_cycles):
    ensemble = forecast(ensemble)          # advance all members one cycle
    obs_pred = observe(ensemble)           # predicted SST at the proxy sites
    ensemble, info = pf.assimilate(        # weights, ESS check, resample
        ensemble, obs_pred, obs, obs_err=4.0 * 0.3)  # sigma_eff = eta * sigma_obs (tempering)
```

The idealized testbeds call `pypfda` in process. CM2Mc-BLING runs as job arrays under the SLURM workload manager, so a cluster-side orchestrator drives its cycle. That orchestrator implements the same interface and the same weight, tempering, and resampling rules, rather than importing the package (Section 2.3 and the Code and data availability section). We verified that the two implementations agree to machine precision in their weight and effective-sample-size computations, and that both snapshot every parent before any clone is written (Section 2.3.1). Replaying the 120 archived per-cycle cost vectors of the diverse-IC annual and five-year experiments through `pypfda` reproduces the orchestrator's importance-weight and effective-sample-size computations to 1.7×10^{-16} and 9×10^{-14}, which is float64 rounding.

The two coupling modes cost very different amounts, and the difference comes from state exchange rather than the filter. In the in-process mode, a single Python session holds the ensemble bookkeeping and drives the model directly, so no scheduler enters the assimilation loop. In the cluster-side mode, each cycle is a job-array submission whose members exchange state through restart files on a shared filesystem, and that exchange is not free. Over the 300-year integration, the analysis step, comprising history archiving, likelihood evaluation, weighting, resampling, and restart preparation, averaged 14.9 minutes

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.



against a 50.5-minute ensemble forecast, so about 22% of each cycle went to the analysis rather than to the model. Two points matter for anyone using the engine. First, this cost is roughly fixed per cycle rather than per model year, so longer assimilation windows reduce it, while it would dominate a sub-annual cycle. Second, because it is restart I/O and not filter arithmetic, it scales with the model's restart size and with the number of members that change ancestry, not with the number of observations. A denser proxy network is nearly free while a finer-resolution model is not. Section S7 in the Supplement gives the full per-cycle decomposition.

The OSSE is a self-contained, within-model (perfect-model) test, meaning the truth, FREE, and DA experiments all use the same model. This within-model test differs from cross-model reconstruction, where one model's SST–AMOC fingerprint constrains another (see the Discussion).

Figure 1 illustrates the architecture: a single common interface connects the `pypfda` engine to a model.

2.3 Online Particle Filter Data Assimilation

Particle filters represent the posterior distribution of the state as a weighted ensemble of model trajectories, free of the symmetric-error assumption that constrains ensemble Kalman filters. That is the advantage for paleoclimate noted in the Introduction, because proxy–climate relationships are nonlinear and states such as the AMOC strength can be skewed or multimodal. The trade-off is that they require more members to represent that posterior reliably (van Leeuwen, 2009; Carrassi et al., 2018; Doucet et al., 2001).

2.3.1 The Filtering Recursion

We denote the model state at analysis time t by $\mathbf{x}_t \in \mathbb{R}^n$. The pseudo-observations are the annual-mean SST at the N_p proxy sites, written $\mathbf{y}_t \in \mathbb{R}^{N_p}$. The forecast model defines a Markov transition density $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$, which is one assimilation cycle of coupled integration. The observation operator defines a likelihood $p(\mathbf{y}_t \mid \mathbf{x}_t)$. The filtering distribution then follows the standard Bayesian recursion (Doucet et al., 2001; van Leeuwen, 2009),

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) \;=\; \frac{p(\mathbf{y}_t \mid \mathbf{x}_t) \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1})\, \mathrm{d}\mathbf{x}_{t-1}}{p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1})}, \tag{1}$$

where the integral propagates the previous cycle's posterior through the model (the *prediction* step) and multiplication by the likelihood performs the Bayesian *update*.

For nonlinear models and non-Gaussian posteriors, the integral in Equation 1 is intractable. Particle filters approximate the posterior by an N-member weighted empirical measure:

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) \;\approx\; \sum_{m=1}^{N} w_m^t\, \delta\big(\mathbf{x}_t - \mathbf{x}_m^f(t)\big)\,, \tag{2}$$

where the particles $\mathbf{x}_m^f(t)$ are the forecast states (the forecast step of Section 2.3.2) and δ is the Dirac measure.

A simple, widely used choice is the *bootstrap proposal* (Gordon et al., 1993), where each particle is sampled from the transition prior by integrating the model forward. This choice reduces the importance-weight recursion to the likelihood alone,

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

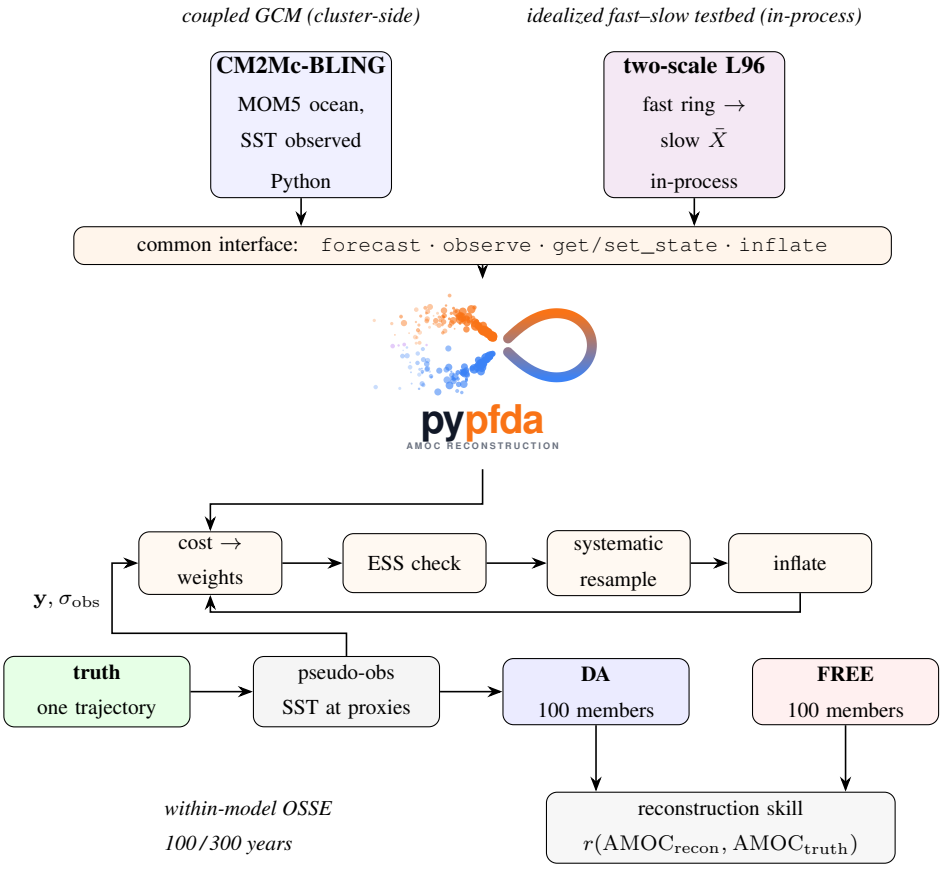


Figure 1. Architecture of the online particle-filter assimilation framework. A model-agnostic engine (`pypfda`) implements the sequential-importance-resampling cycle (cost → weights → effective-sample-size check → systematic resampling → inflation, which loops back to the cost/weight step at the start of the next assimilation cycle; Section 2.3) and couples to a model via a single common interface. The observations $\mathbf{y}$ and their error σ_{obs} enter the engine at the cost/weight step, where they are compared with the model equivalents $H(\mathbf{x})$ returned by `observe`. The coupled model is CM2Mc-BLING, whose ocean is MOM5 and whose sea surface temperature is the observed variable. The identical engine is additionally exercised on an idealized fast–slow chaotic testbed (two-scale Lorenz-96, shown in violet), which plugs into the same interface in-process and reconstructs a slow variable from fast observations as a minimal benchmark that runs on a laptop (Section S4 in the Supplement). The CM2Mc-BLING OSSEs are run over 100 years (the campaign) and, for the headline configuration, extended to 300 years for the significance test (Section 3.4).

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

$$w_m^t \propto w_m^{t-1}\, p(\mathbf{y}_t \mid \mathbf{x}_m^f(t)), \qquad \sum_{m=1}^{N} w_m^t = 1. \tag{3}$$

After each analysis we reset all weights to $1/N$ (resampling, Section S10 in the Supplement). The factor w_m^{t-1} in Equation 3 is therefore the same for every member, and the weight of a member is simply proportional to its observation likelihood, which we define as Gaussian in the cost and weight steps below (Equations 4–5). Weights thus act only within the cycle in which they are computed. When the effective sample size stays above the threshold, no resampling takes place: every member continues unchanged into the next cycle, and the observations of that cycle have no effect on the ensemble.

Both the `pypfda` driver and the CM2Mc-BLING orchestrator save every parent's restart state before any child is written, so the ensemble that runs is the systematic draw that was logged. Every genealogy diagnostic in this paper is therefore read directly from the per-cycle draw.

2.3.2 The Assimilation Cycle

Assimilation occurs in cycles of duration Δt, such as annual or five-year intervals. At the beginning of each cycle, the $N = 100$ ensemble members are advanced by Δt using the coupled climate model. This is the forecast step. Each member m progresses from time t to $t+\Delta t$ according to the model dynamics, producing forecast states $\mathbf{x}_m^f(t+\Delta t)$. The model runs on 28 processors per member and produces distributed tile output. During the archive step, we assemble these tiles into single NetCDF files for each member. From these files, we calculate the cycle-mean sea surface temperature at the 59 proxy locations, giving model-predicted observations $\mathrm{SST}_{\mathrm{model}}^{(m)}(p)$ for each member m and proxy p.

The mismatch between each ensemble member and the observations is computed as a weighted sum of squared residuals:

$$J_m = \sum_{p=1}^{N_p} \frac{\left(\mathrm{SST}_{\mathrm{model}}^{(m)}(p) - \mathrm{SST}_{\mathrm{obs}}(p)\right)^2}{\sigma_{\mathrm{eff}}^2} \tag{4}$$

Here, $N_p = 59$ denotes the number of proxy locations, $\mathrm{SST}_{\mathrm{obs}}(p)$ is the pseudo-observation, and $\sigma_{\mathrm{eff}} = \sigma_{\mathrm{obs}} \times \eta$ defines the effective observation error. This cost function aggregates all proxies into a single global misfit per member, unlike localized alternatives. It implements a global sequential-importance-resampling particle filter without covariance localization. The error inflation factor $\eta = 4.0$ widens the likelihood to delay weight collapse due to the curse of dimensionality (Snyder et al., 2008). In this perfect-model configuration, there is no structural model error or spatial representativeness mismatch. η is therefore a tuning parameter of the algorithm. Bracket tests showed that $\eta = 2$ resulted in an effective sample size (ESS) below 5, whereas $\eta = 8$ produced an ESS greater than 80 with minimal observational constraint.

The costs are then transformed into normalized importance weights using the Gaussian likelihood:

$$w_m = \frac{\exp(-J_m/2)}{\sum_{m'=1}^{N} \exp(-J_{m'}/2)} \tag{5}$$

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

The factor $1/2$ in the exponent results from the Gaussian likelihood $p(\mathbf{y}|\mathbf{x}) \propto \exp(-J_m/2)$.

The effective sample size quantifies the number of independent samples represented by the weighted ensemble:

$$\mathrm{ESS} = \frac{1}{\sum_{m=1}^{N} w_m^2} \tag{6}$$

ESS ranges from 1, where a single member dominates and the ensemble has collapsed completely, to N, where all members are equally weighted. When ESS falls below 85% of N (i.e., ESS < 85 for $N = 100$), Sequential Importance Resampling (SIR) is triggered using systematic (low-variance) resampling (Kitagawa, 1996; Carpenter et al., 1999). A single uniform deviate $u \sim \mathcal{U}(0, 1/N)$ defines N equally spaced pointers $u_k = u + (k-1)/N$ for $k = 1, \ldots, N$. The ancestor of each new member is the member whose cumulative weight $\sum_{m' \le m} w_{m'}$ first exceeds u_k. This low-variance scheme duplicates high-weight members and eliminates low-weight ones, and it is the "systematic resample" step shown in Figure 1. After resampling, we reset all weights to $1/N$. Finally, we perturb the resampled members to restore ensemble diversity (inflation; see Section 2.3.4).

2.3.3 Reconstruction-skill convention

Unless otherwise specified, we quantify reconstruction skill as the Pearson correlation between the *forecast* ensemble-mean AMOC and the truth AMOC at lag 0, calculated over the entire period in the standard OSSE forecast convention. The forecast ensemble mean is taken over each integration window, before the analysis and resampling that end it. The same convention is used for every experiment.

2.3.4 Post-Resampling Inflation

Resampling leaves several identical copies of the high-weight members, and without a perturbation the ensemble would collapse onto them. To maintain diversity, the engine invokes the model's `inflate` operator for each member, and the field it perturbs is a per-model choice (Table S4 in the Supplement). In CM2Mc-BLING, small perturbations are applied to the atmospheric temperature field of each member.

$$T'^{(m)}(i,j,k) = T^{(m)}(i,j,k) + \epsilon^{(m)}(i,j,k) \tag{7}$$

At grid point (i,j,k) the atmospheric temperature for member m is denoted by $T^{(m)}(i,j,k)$, and $\epsilon^{(m)}$ stands for a spatially-smoothed random perturbation field. This perturbation is generated as

$$\epsilon^{(m)}(i,j,k) = \sigma_T \cdot \left(G_{\sigma_h} *_h \xi^{(m)} \right)(i,j,k) \cdot M(k) \tag{8}$$

where $\xi^{(m)}(i,j,k) \sim \mathcal{N}(0,1)$ is a white noise field with a unique seed for each member and each resampling event; G_{σ_h} is a horizontal Gaussian smoothing kernel with $\sigma_h = 2.5$ grid cells, σ_T is the perturbation amplitude, and $M(k)$ is a vertical mask function:

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

$$M(k) = \begin{cases} 1 & \text{if } 6 \le k \le 18 \text{ (mid-troposphere)} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Perturbations are restricted to the mid-troposphere (approximately 850 to 200 hPa) to prevent disturbances in the boundary layer, where perturbations dissipate rapidly, and in the stratosphere, where they may generate gravity waves. During post-resampling inflation, perturbations are applied to temperature rather than atmospheric pressure thickness (DELP), because DELP is the prognostic variable for air mass in each model layer, and temperature perturbations are more numerically stable during adjustment after restart-file modification. In the CM2Mc-BLING orchestrator, the base amplitude $\sigma_T = 0.2$ K is scaled at each resampling by $1 + 10\,(1 - \mathrm{ESS}/N)$, so that the amplitude grows with ensemble degeneracy. This scaling produced amplitudes ranging from 0.5 to 2.1 K at the effective sample size (ESS) values observed in the experiments. Before scaling, we renormalize the field to unit variance, smooth it vertically over 1.5 levels, and apply it to every member. We would have applied an additional twofold increase if the sea surface temperature (SST) spread had dropped below 0.6 of its initial value, but this condition was never met. These amplitudes kept the model stable and gave the copies enough spread to diverge in the next forecast.

In the CM2Mc-BLING experiments, the engine uses three additional operators, each with fixed settings: the likelihood tempering behind $\sigma_{\mathrm{eff}} = \eta\,\sigma_{\mathrm{obs}}$, an optional cap on the largest weight, and the adaptive resampling rule that triggers on the effective sample size. Section S10 in the Supplement defines all three and reports the settings used here.

2.3.5 A Usable-Cycle Bound

The assimilation cycle length Δt is constrained by two intrinsic timescales. The lower bound is determined by the rate at which resampling and inflation can regenerate independent ensemble members, which cannot exceed the rate of chaotic divergence. This divergence rate is governed by the leading Lyapunov (error-doubling) time τ_L; assimilation error tends to concentrate in the flow's unstable and neutral subspace (Bocquet and Carrassi, 2017). If cycles are shorter than τ_L, resampling occurs faster than diversity can recover, and genealogical memory erodes. The ocean introduces an additional lower bound: the AMOC decorrelation (autocorrelation e-folding) time τ_{adj}, approximately one quarter of the dominant period, during which the surface temperature integrates an overturning signal suitable for assimilation. Although τ_L was not measured in this study, the atmospheric error-doubling time is on the order of days and therefore much shorter than τ_{adj}, making τ_{adj} the relevant lower bound. The upper bound is the Nyquist–Shannon sampling criterion (Shannon, 1949), which requires Δt to resolve the target variability. Sampling slower than half the dominant period $T^\star$ aliases the signal the filter is meant to track. Together, these constraints give

$$\underbrace{\max(\tau_L,\, \tau_{\mathrm{adj}})}_{\text{diversity / memory floor}} \lesssim \Delta t < \underbrace{\tfrac{1}{2}\,T^\star}_{\text{Nyquist ceiling}} . \tag{10}$$

If the interval defined by these bounds is not empty, the filter has a usable cycle length. If the interval is narrow or empty, no cycle length can recover skill. For CM2Mc-BLING, the interval is narrow because $\tau_{\mathrm{adj}} \sim 3$ yr is close to the Nyquist ceiling of

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

$\frac{1}{2}T^{\star} \approx 6.7$ yr, based on the approximately 13.3-year dominant AMOC period (see Figure S8 in the Supplement). The five-year cycle falls within this interval. In contrast, the annual cycle (over-resampled and short of diversity) and the ten-year cycle (aliased) fall outside it, which reproduces the hump-shaped skill curve found in Section 3.5. Although Equation 10 is a scaling argument rather than a strict theorem, it predicts the optimal cycle length from independently measurable timescales.

2.4 Spectral Analysis of AMOC Variability

We estimated the AMOC spectrum with Welch's method (Welch, 1967) to find the dominant periods that bound the assimilation cycle length. We analyzed the final 1500 years of the equilibrated control simulation, with a segment length of $L = 200$ years and 50% overlap. This configuration gives a frequency resolution of $\Delta f = 1/L = 0.005$ cycles per year, based on approximately 14 Hanning-tapered segments. This resolution is sufficient for decadal to multidecadal signals, and segment averaging reduces the variance of the estimate. The Nyquist limit on the cycle length is then half the dominant period (Equation 10).

2.5 Observing System Simulation Experiment (OSSE) Design

In an OSSE the observations are generated from a known model trajectory, so reconstruction skill can be measured without model error or proxy-calibration uncertainty. OSSEs are standard in weather forecasting (Atlas, 1997) and have been used in paleoclimate studies (Acevedo et al., 2017; Okazaki et al., 2021).

The OSSE setup consists of three parallel 100-year experiments. The first is the truth trajectory, a single deterministic model integration that is the synthetic reality; pseudo-observations are taken from it at the proxy locations, and the truth itself is never shown to the filter. The second experiment is the FREE ensemble, comprising 100 members integrated forward without assimilation and serving as the no-assimilation baseline for reconstruction skill. The third experiment is the DA ensemble, also with 100 members, in which the particle filter is applied whenever pseudo-observations become available.

2.5.1 Initial Condition Generation

For a fair comparison, the FREE and DA ensembles must begin from byte-identical initial conditions. We verify this requirement by calculating the root-mean-square difference in atmospheric pressure thickness (DELP) between corresponding DA and FREE ensemble members on the model grid. This difference is zero to machine precision (0.0 Pa).

All 100 members originate from a single equilibrated control state (year 5363 during the spin-up), by which time the AMOC strength and global sea surface temperatures had stabilized. Ensemble spread is introduced by perturbing the atmospheric pressure thickness field (DELP), the prognostic variable for the mass of air in each model layer, so the ensemble begins from a shared stable state with controlled spread.

The perturbation procedure consists of four steps (Section S8 in the Supplement). For each ensemble member m, a three-dimensional white noise field is generated, and smoothed with a Gaussian kernel so that it is spatially coherent. The smoothed noise field is scaled to 0.1% of the global mean DELP and added to the base state, resulting in perturbations with a root-mean-square value of approximately 10 Pa. The perturbations are small enough to keep the model stable and large enough to grow

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

into ensemble spread. This initial-spread perturbation is distinct from the post-resampling inflation described in Section 2.3.4, which instead perturbs atmospheric temperature.

For the truth trajectory, a perturbation amplitude ten times greater (1.0%) is applied, along with a different random seed (9999), so that it diverges from the ensemble members. Its initial ocean state is shared with the similar ocean initial-condition ensemble (Section 2.5.2). We use 100 members, as earlier particle-filter paleoclimate studies did (Dubinkina et al., 2011), and century-long integrations remain affordable at that size.

2.5.2 Two initial-condition strategies

The experiments evaluate two approaches for setting the initial ocean state (see Figure S11 in the Supplement), reflecting the uncertainty of the ocean state at the start of the assimilation period in real reconstructions. We call the previously described setup the *similar ocean initial condition* (IC) case, where all members share the same initial ocean state (AMOC ∼20.2 Sv). In contrast, the *diverse ocean IC* case assigns each member a different control year, with samples taken every 50 years over a 4950-year interval (years 4900 to 9850). In this configuration, no DELP seed is applied, as each member's coupled ocean-atmosphere state is already distinct by design. The perturbation procedure described above is therefore specific to the similar-IC configuration. This sampling gives 100 members whose ocean states span the model's internal AMOC variability (17–26 Sv). Given the dominant AMOC period of 13.3 years (Section 3.5), a 50-year separation corresponds to approximately 3.75 oscillation cycles, resulting in substantial but incomplete decorrelation among members. Both configurations use the same truth trajectory, pseudo-observation network, and DA parameters, so differences in skill reflect the effect of initial ocean-state diversity. Section 3.6 examines how this diversity interacts with the assimilation process.

2.5.3 The OSSE Experiment Campaign

Table 2 lists all experiments, over cycle length, both initial-condition strategies and the post-resampling regularization, including the unsuccessful one. The ten-year cycle aliases the AMOC spectral peak, and that failure marks where online particle-filter AMOC reconstruction stops working.

2.6 Realism of Simulated AMOC against RAPID

Because we evaluate OSSE skill against the model's own truth, the synthetic truth must be physically plausible. We therefore compared the truth run with the RAPID–MOCHA observational array (McCarthy et al., 2015; Smeed et al., 2018) (Table S7 in the Supplement). The mean overturning at 26.5°N is 1.9 Sv (11%) higher than the RAPID mean, and the run shows no collapse, reversal, or drift. Interannual variability is reduced relative to RAPID's monthly value but remains comparable to its annual value. The covariance analysis of the truth run (Section S5 in the Supplement) confirms that the subpolar-gyre SST–AMOC fingerprint used by the filter is positive. However, the sign and magnitude of this fingerprint are not consistent across all models (Section 4).

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 2. The online particle-filter OSSE campaign. Each experiment is a 100-member CM2Mc-BLING FREE/DA pair over 1801–1900 (100 yr). All experiments assimilate the global 59-site network except Diverse-5yr (Atlantic-34), which assimilates only its 34 Atlantic sites. "Reg." is the post-resampling regularization. Here atm denotes the ESS-scaled atmospheric-temperature inflation of Section 2.3.4 (base 0.2 K, applied 0.5–2.1 K), and +ocean denotes that inflation with additional ocean temperature and salinity perturbation (base 0.01°C / 0.003 psu, applied at its $3\times$ cap, i.e. 0.03°C / 0.009 psu, at every event; Section 3.8). Δr is the AMOC reconstruction-skill gain (DA − FREE), using the forecast-convention skill defined in Section 2.3; each Δr matches its entry in Table 3, and the Atlantic-34 experiment is described in Section 3.7. Confidence intervals are the moving-block bootstrap of Section 2.8. The outcome column summarizes each experiment.

Experiment	Ocean IC	Cycle	Reg.	Outcome
Systematic campaign (CM2Mc-BLING)				
Similar-1yr (pilot)	similar	1 yr	atm	first end-to-end online-DA run; no gain over FREE
Similar-1yr	similar	1 yr	atm	$\Delta r = +0.00$, 95% CI $[-0.23, +0.14]$; degeneracy prevention added
Similar-5yr	similar	5 yr	atm	$\Delta r = +0.17$, 95% CI $[+0.02, +0.26]$ (FREE already $r = +0.52$)
Similar-10yr	similar	10 yr	atm	$\Delta r = +0.01$, 95% CI $[-0.13, +0.12]$; CI includes zero
Diverse-1yr	diverse	1 yr	atm	$\Delta r = +0.65$; CI $[+0.35, +0.91]$; weight-degenerate (mean ESS 28%)
Diverse-5yr	**diverse**	**5 yr**	atm	$\Delta r = +0.68$ **(headline)**; CI $[+0.40, +0.94]$; $q = 0.04$
Diverse-10yr	diverse	10 yr	atm	$\Delta r = +0.35$; CI $[-0.08, +0.81]$ includes zero: Nyquist aliasing (failure mode)
Diverse-5yr (+ocean)	diverse	5 yr	+ocean	$\Delta r = +0.77$; largest gain and smallest q (0.009); CI $[+0.52, +1.02]$
Diverse-5yr (Atlantic-34)	diverse	5 yr	atm	$\Delta r = +0.66$; CI $[+0.44, +0.86]$; $q = 0.02$

2.7 Pseudo-Observation Network

To assess how strongly sparse surface observations constrain the deep overturning circulation, we constructed a global network of 59 pseudo-proxy locations. The network targets regions where the AMOC exerts a strong and detectable influence on sea surface temperature (Figure 2). It follows the observational evidence that AMOC variability leaves its clearest thermal signature in the subpolar North Atlantic (Caesar et al., 2018; Rahmstorf et al., 2015; Zhang et al., 2019). The final placement matches the truth run's maximum covariance analysis (MCA) fingerprint (Figure S10 in the Supplement), although we did not use the fingerprint as a selection criterion.

The network includes 10 sites in the subpolar gyre (SPG), where coupling between AMOC and SST is strongest and the "cold blob" pattern is most pronounced. An additional 19 sites span the Atlantic basin, from the subtropical to the tropical Atlantic, to capture the broader meridional heat-transport signal. Remote teleconnections may also influence the AMOC through atmospheric bridges or oceanic pathways; therefore, we include 12 Pacific sites and 9 Southern Ocean sites. We also include 5 Nordic Seas sites, where deep-water formation directly modulates overturning strength, and 4 Indian Ocean sites, which capture Indo-Pacific warm-pool variability. This uneven distribution mimics the uneven coverage of real proxies while still sampling the regions that matter most for the AMOC.

For each proxy location $p \in \{1, 2, \ldots, 59\}$ and each year $t \in \{1801, \ldots, 1900\}$, we extract the annual-mean sea surface temperature from the truth trajectory at the nearest ocean grid point. To simulate the combined effects of measurement error

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

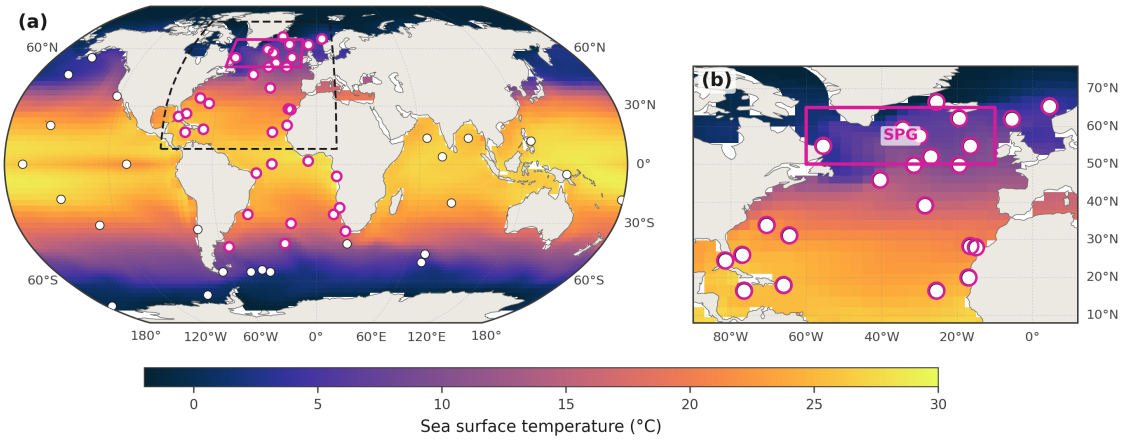


Figure 2. Pseudo-proxy network over the model's climatological sea surface temperature (°C). (a) Global distribution of the 59 proxy locations (white circles; subpolar and Atlantic AMOC-relevant sites ringed in magenta). These sites were placed *a priori* at named subpolar, deep-water-formation, and overflow locations of the North Atlantic, e.g. the Labrador and Irminger Seas, Denmark Strait, and Rockall Trough, rather than by a quantitative fingerprint-loading criterion. The black dashed rectangle marks the North-Atlantic region enlarged in panel (b), and the solid magenta rectangle marks the subpolar-gyre (SPG) averaging region. (b) North-Atlantic detail, with the subpolar-gyre averaging region outlined in magenta.

and proxy-temperature calibration uncertainty, we add independent Gaussian noise, giving

$$\mathrm{SST}_{\mathrm{obs}}(p,t) = \mathrm{SST}_{\mathrm{truth}}(p,t) + \varepsilon(p,t), \tag{11}$$

where $\varepsilon(p,t) \sim \mathcal{N}(0, \sigma^2_{\mathrm{obs}})$ with $\sigma_{\mathrm{obs}} = 0.3°\mathrm{C}$. This value is consistent with the error ranges reported for annually resolved tropical coral SST reconstructions (Tierney et al., 2015) and covers both instrumental and calibration uncertainties. For cycles longer than one year, the assimilated pseudo-observation is the truth's cycle-mean SST at each site plus a single draw from the same $\mathcal{N}(0, \sigma^2_{\mathrm{obs}})$ per site and cycle. Averaging annual values does not reduce the observation error. The same error is used at all 59 sites. This is an idealization, because the subpolar-gyre, Nordic Seas and Southern Ocean sites, which constrain the AMOC most, have no annually resolved SST proxies of this accuracy and would in reality depend on lower-resolution sediment records. The network is therefore optimistic and its skill an upper bound (Section 5). The nearest grid point for each proxy location is determined through minimizing the approximate spherical distance over all ocean grid points (Section S8 in the Supplement).

2.8 Significance Testing and Multiple-Comparison Correction

The 100-year records show strong decadal persistence, resulting in underpowered individual correlations (effective degrees of freedom $N_{\mathrm{eff}} \approx 20$, Section 4). Skill is therefore assessed with differences ($\Delta r = r_{\mathrm{DA}} - r_{\mathrm{FREE}}$) rather than individual correla-

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

tions, using two complementary non-parametric procedures (i, ii) and a multiple-comparison (false-discovery-rate) correction (iii).

(i) A moving-block bootstrap (20,000 resamples, fixed 10-year blocks to preserve autocorrelation) provides a 95% confidence interval (CI) for each Δr.

(ii) A surrogate test gives a counted one-sided p-value for each experiment, $p = (k+1)/(N_{\text{surr}}+1)$ following Phipson and Smyth (2010), where k is the number of surrogate realizations reaching the observed Δr. The minimum attainable p-value is determined by the number of *distinct* realizations the null hypothesis allows, not by the number of draws. This support is stated explicitly for each null. The first null shifts the truth trajectory circularly through all $n-1$ non-trivial lags, preserving the truth's autocorrelation exactly and yielding a minimum $p = 1/n$. The second null randomizes the Fourier phases of the truth while preserving its power spectrum and, thus, its full autocorrelation function, including the decadal (12–14-year) spectral peak. Its support is continuous, and we draw 100,000 surrogates. Both nulls disrupt the phase matching between the truth and the FREE/DA pair while maintaining each series' persistence. We score each surrogate against FREE and DA in pairs, allowing a paired comparison.

(iii) Across the eight quantitatively scored CM2Mc campaign experiments (Table 2), the false-discovery rate is controlled using the Benjamini-Hochberg (BH-FDR) procedure (Benjamini and Hochberg, 1995) at $q < 0.05$. Each experiment's phase-randomization p-value from (ii) is one test in the family, and this correction lowers the risk of selecting an apparently skillful experiment due to multiple comparisons. The initial end-to-end run, the Similar-1yr pilot, is reported qualitatively and is not included in the family. Development tiers preceding the campaign (32- and 133-site networks, with and without covariance localization) were used solely to complete the network and filter settings and are not scored here. The family includes all experiments quantitatively scored against the same truth, including the Atlantic-network experiment, with no exclusions based on outcome. Reported q-values are the Benjamini-Hochberg-adjusted p-values across these eight experiments, while the bootstrap in (i) provides confidence intervals only. After applying this correction, three experiments are significant at $q < 0.05$ (smallest $q = 0.009$; Section 4). Because the exhaustive circular-shift null cannot resolve below its own attainable bound of $p = 1/n$, we treat the campaign as a configuration ranking and the independent 300-year integration as the confirmation.

3 Results

The experiments address the four research questions outlined in Section 1. Sections 3.1–3.4 examine whether surface proxies can constrain the AMOC spatially, across components, and over three centuries. Section 3.5 examines the effect of cycle length. Section 3.6 evaluates the effect of initial ocean state diversity, while Section 3.7 considers the extent of the observing network required. Section 3.8 examines whether resampling erodes the ensemble's genealogical diversity and whether upper-ocean perturbations might preserve it.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

3.1 Spatial Reconstruction Skill

Assimilation improves reconstruction skill in the subpolar North Atlantic (Figure 3). This analysis uses the diverse ocean initial-condition ensemble (Section 2.5.2). The proxy network constrains this region, where the SST–AMOC fingerprint sits. Averaging 100 free members with diverse initial states removes their internal variability and leaves only a weak shared signal, a slow cooling common to all runs branched from the control. This pattern is consistent with the adjustment to the halved background diffusivity (Section 2.5.1). Incomplete decorrelation among members separated by approximately 3.75 oscillation cycles likewise contributes. The resampled DA ensemble, by contrast, draws on fewer independent states and does not benefit from far-field averaging. The gain is therefore regional and sits where the observations are. It is the spatial counterpart of the AMOC-index gain, not a global improvement.

3.1.1 Sensitivity to the evaluation window

The full-record skill reported is conservative. Because the diverse initial conditions are drawn from the control simulation, the free ensemble retains residual memory of those states during the first few decades, which lets it follow the truth for a while. Over years 31–100, excluding the initial 30-year ensemble spin-up ($n = 70$), the free-running subpolar-gyre SST correlation changes sign, while DA continues positive. This shift approximately doubles the proxy-region gain (Table S5 in the Supplement), as mapped in Figure 3. We keep the full 100-year window as the main metric all the same, for two reasons: it maximizes the number of independent samples for autocorrelation-corrected significance analysis (Section 2.8), and shorter late windows reduce the degrees of freedom for the free-run correlation. Although windowing affects magnitudes and, for near-zero surface air temperature (SAT), its sign, it does not alter the ranking of DA over FREE for the AMOC index, and the 300-year integration remains insensitive to this choice (Table S5 in the Supplement).

In the diverse ocean IC experiment (the Diverse-5yr headline experiment; Table 2), the particle filter draws the SST ensemble toward the truth despite the different initial ocean states. Subpolar-gyre SST ensemble trajectories for FREE and DA are shown in Figure S9 in the Supplement. The coupled SST–AMOC spatial patterns, their coupling intensity, and their causal direction are characterized in Section S5 in the Supplement.

3.2 Cross-Component Signal Propagation

Only SST is assimilated, so any change in SAT, the 2-meter air temperature, must come through the model's air–sea coupling. Relative to the FREE baseline with the same diverse initial conditions, SAT skill improves over the subpolar North Atlantic, where the SST is constrained, but not in the area-weighted global mean (Figure 3b), which matches the SST result. The cross-component constraint therefore reaches the atmosphere only above the observed ocean. Propagation into the unobserved ocean interior is examined in Sections 3.3 and 3.7.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

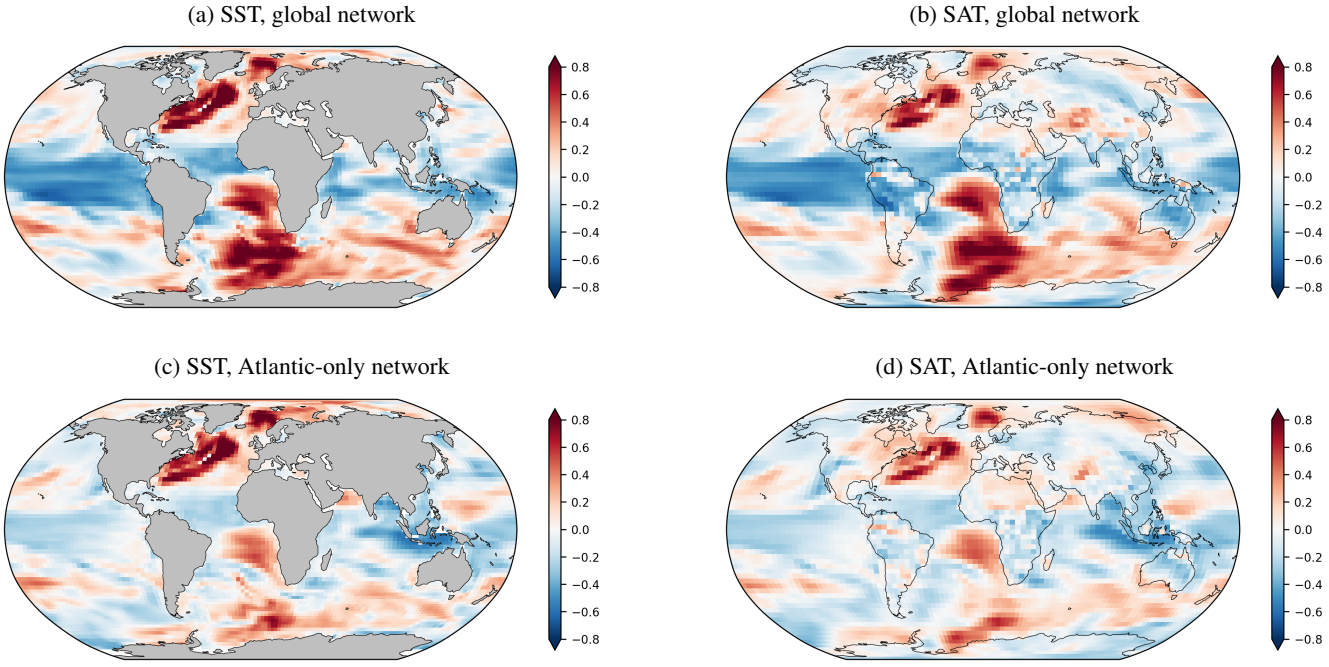


Figure 3. Per-grid-cell difference in time correlation with the truth, $\Delta r = r(\mathrm{DA}, \mathrm{truth}) - r(\mathrm{FREE}, \mathrm{truth})$, over years 31–100, the initial-condition-forgotten window of Section 3.1. Red: positive Δr (DA more correlated with the truth than FREE); blue: negative Δr. *Top row*, the global 59-proxy network: (a) sea surface temperature (land gray), (b) surface air temperature (SAT). *Bottom row*, the Atlantic-only 34-proxy network (Section 3.7): (c) SST, (d) SAT. The same comparison in the Atlantic overturning plane is given in Figure S15 in the Supplement. Area-weighted subpolar-gyre-box and deep-cell (0–60°N below 500 m) values of Δr are given in Sections 3.6 and 3.7.

3.3 Reconstruction of the AMOC

Under diverse initial conditions, every 100-year AMOC correlation except the ten-year one is individually significant after autocorrelation correction ($N_{\mathrm{eff}} \approx 20$, critical correlation $r_{\mathrm{crit}} \approx 0.44$; Table S1 in the Supplement and Section 4). The ten-year cycle is the one configuration that fails this test. Because the aim is to rank configurations, the pattern across experiments matters more than any single value.

The AMOC is a deep-ocean variable that is not directly observed by the surface proxy network. The analysis first considers the diverse-IC configuration, which is the more realistic scenario (Section 2.5.2). Table 3 summarizes the AMOC skill for each experiment, spanning cycle length and both IC strategies.

In the diverse-IC scenario, the FREE ensemble is uncorrelated with the truth ($r = -0.14$, consistent with zero at $N_{\mathrm{eff}} \approx 17$), as its members sample independent phases of the model's internal variability. Assimilation restores positive skill, with the gain increasing as the cycle length approaches the AMOC decorrelation time. The five-year cycle is the headline configuration, with $r = +0.54$ and $\Delta r = +0.68$ (Figure 4; Section 3.5). The annual cycle attains a comparable correlation but from a far more degenerate weight distribution (Table 3; Section 2.8). At the five-year cycle the assimilated ensemble is also statistically

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

reliable. It keeps 86% of the FREE spread, and its spread–skill ratio is 0.96 (a ratio of 1.00 is calibrated, below 1 overconfident, above 1 over-dispersed; the free ensemble, calibrated by construction here, scores 0.92). The truth falls outside the ensemble in 5% of years, against the 2% a flat rank histogram implies (Section S9 and Figure S13 in the Supplement). The annual cycles are the exception: they compress the spread by about half, to a spread–skill ratio of 0.44, and their rank histograms are U-shaped.

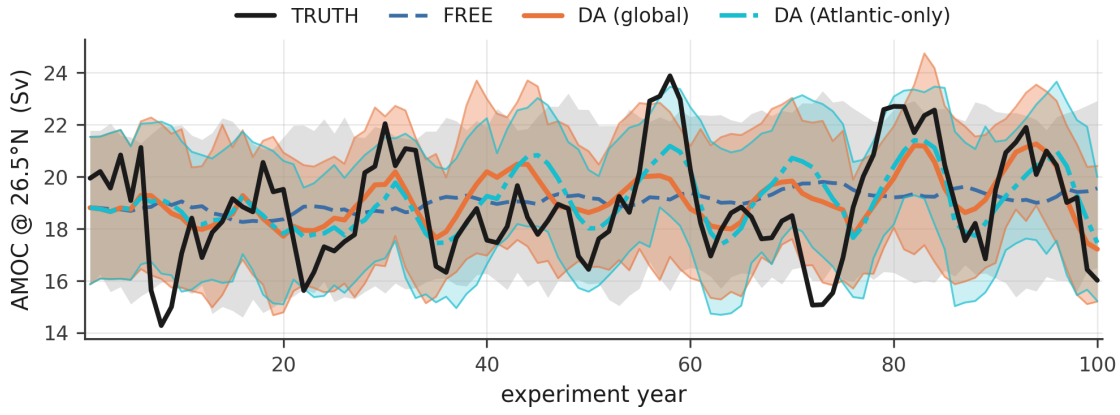


Figure 4. CM2Mc-BLING AMOC reconstruction (the Diverse-5yr experiment: 100-member ensemble, five-year assimilation cycles, diverse initial conditions, 100-year integration). Annual AMOC at 26.5°N (max below 500 m): the truth is black; the FREE ensemble mean is blue dashed with its 5–95% band in gray; and the DA ensemble mean, with its 5–95% band, is orange for the global 59-proxy network. The cyan dash-dotted line and its 5–95% band are the DA ensemble under the reduced 34-proxy Atlantic-only observation network (Section 3.7). Skill statistics (correlation, RMSE, and MSSS for each curve) are given in Sections 3.3 and 3.7.

The full-record AMOC gain is similarly conservative due to residual FREE initial-condition memory (Section 3.1). Evaluating years 31–100 further increases the diverse-IC five-year AMOC-index gain (Table S5 in the Supplement).

For comparison, in the similar-IC scenario, the FREE ensemble already tracks the truth due to shared initial ocean memory (Figure S12a in the Supplement), so it is not an independent baseline, and assimilation has little left to add. Only the five-year cycle produces a gain whose confidence interval excludes zero, and it is small. The five-year cycle, the best of the three, reaches $r = +0.69$ against the similar-IC FREE baseline of $+0.52$ ($\Delta r = +0.17$, 95% CI $[+0.02, +0.26]$). The annual and ten-year cycles match that baseline ($\Delta r = +0.00$ and $+0.01$; Table 3). The annual cycle narrows the spread by half, the same compression as with diverse initial conditions.

This extended free-running predictability is not an intrinsic property of the AMOC, but a consequence of shared initial ocean-state memory, as discussed for the diverse ensemble in Section 3.1. This memory diminishes as ensemble members diverge chaotically, while the diverse ensemble mean converges toward the model climatology (Figure S12b in the Supplement). Assimilation is thus largely redundant when the initial ocean state is known (similar IC) and essential when it is unknown (diverse IC), as in any real reconstruction.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

3.4 Summary of Results from a 300-Year Integration

Over 300 years, assimilation recovers the Atlantic Meridional Overturning Circulation (AMOC) at $r = +0.78$, whereas the free-running ensemble remains uncorrelated with the truth. This outcome supports the protocol selected in the 100-year campaign, which used diverse initial conditions, five-year cycles, and a global 59-proxy network. Each experiment in the 100-year campaign covers fewer than eight AMOC oscillation cycles, so the campaign is designed to rank configurations rather than to certify any one of them. This section tests the protocol selected from the cycle-length and initial-condition comparisons (Sections 3.5 and 3.6) on an independent integration. We repeated the selected protocol in a second 300-year OSSE, with 100-member FREE and DA ensembles re-branched from the same 100 control years as the campaign. The truth trajectory begins from the same control state but with a different perturbation ($r = +0.49$ with the 100-year truth over the first century). The reconstruction both reproduces and sharpens the 100-year result (Figure 5). Across the full record, the free-running ensemble mean remains uncorrelated with the truth ($r = +0.03$, 95% CI $[-0.12, +0.19]$), while assimilation recovers the overturning at $r = +0.78$ (95% CI $[+0.72, +0.83]$; $\Delta r = +0.74$ computed from unrounded correlations, 95% CI $[+0.59, +0.90]$; moving-block bootstrap, 20,000 resamples with 10-year blocks). Both the assimilated and true AMOC have strong serial correlation (lag-1 autocorrelation $\rho_1 = 0.88$ for the DA ensemble mean and $\rho_1' = 0.80$ for the truth). We therefore report significance with the two-series Bretherton et al. (1999) effective sample size, $N_{\text{eff}} = N_t\,(1 - \rho_1\rho_1')/(1 + \rho_1\rho_1') \approx 52$ over 300 years. Because both series share the 13-year oscillation, surrogate nulls from Section 2.8 are also considered. The DA skill is significant under both tests (phase randomization $p = 1 \times 10^{-5}$, the smallest value 100,000 surrogates can attain; exhaustive circular shift $p = 0.003$; $p = 1 \times 10^{-11}$ from the lag-1 N_{eff}), whereas the free-running correlation is indistinguishable from zero (surrogate $p = 0.39$). Excluding the first 30 years for spin-up does not alter the result ($r = +0.80$, $\Delta r = +0.80$), and for years 101–300, which do not share an initial state with the campaign, $r = +0.82$ ($\Delta r = +0.81$; surrogate $p = 0.005$ and 1×10^{-5}; $N_{\text{eff}} \approx 34$). With 23 rather than 8 oscillation cycles, the autocorrelation-aware test shows that assimilating surface temperature turns an uncorrelated free-running baseline into a reconstruction whose skill is significant on its own. In mean-square error, the reconstruction reaches RMSE $= 1.28$ Sv and MSSS $= +0.58$, compared to 1.99 Sv and -0.02 for FREE. The assimilated ensemble mean outperforms the truth's own climatology and the free-running mean does not, and that margin is resolved (moving-block bootstrap CI on the DA MSSS $[+0.51, +0.65]$, with no resample at or below zero). MSSS stays below the mean-square ceiling of $r = +0.78$, which is $r^2 = 0.61$, because the ensemble mean's amplitude is uncalibrated (Section S3 in the Supplement). The same 300-year integration also improves subpolar-gyre SST (Figure 5b; FREE $r = +0.36$, DA $r = +0.72$, $\Delta r = +0.36$, 95% CI $[+0.28, +0.52]$). Both the truth and the free ensemble mean show a gradual cooling of approximately 0.5°C over three centuries. After linear detrending, FREE falls to $r = +0.04$ and DA to $+0.68$ ($\Delta r = +0.64$). Figure 5c counts ensemble membership three ways in one unit, and the three counts answer different questions. Two of them are snapshots of a single analysis. The teal curve is the effective sample size written as a number of members, which measures how evenly the weights are shared before the draw; the purple curve is the number of separate members that the draw then selected as parents. The orange curve is a running total over every analysis so far: the number of members present at the start of the run that still have at least one descendant in the ensemble. Only the two snapshots say whether a given analysis was well conditioned, and both stay high for

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

all 300 years. The effective sample size averages 54.2 of the 100 members and never falls below 31.8, and the draw averages 65.5 distinct parents, never fewer than 50 and never more than 82, with no parent copied more than 10 times in any of the 60 draws. Weight degeneracy, in which one member takes almost all of the weight and the draw collapses onto a handful of parents, therefore never occurs here (Section 3.8). The running total falls all the same, and it falls for an arithmetic reason rather than a diagnostic one. Every analysis permanently discards the lineages of the members it does not select, and a discarded lineage cannot reappear; because that loss compounds over 60 successive analyses, the number of surviving founders is driven toward one however well conditioned each individual draw is. Member ancestry accordingly contracts from 100 founders to three after a century and to a single founder by year 155, remaining there for the rest of the run. This is path degeneracy, and neither curve can stand in for the other. A per-cycle count cannot detect it, because every draw looks well conditioned on its own. The founder count cannot be read as a measure of filter health, because it must decay in any scheme that resamples at all. The loss of founders does not propagate into the state. The ensemble's spread persists near 1.5 Sv throughout, about 77% of the FREE spread (Section S9 in the Supplement), because the coupled model regenerates diversity between analyses about as fast as selection removes it, so members that share a founder are not copies of one another. The genealogy, rather than the spread, collapses, and the reconstruction's skill over years 201–300 ($r = +0.83$) is measured long after the single-lineage point. Sustaining independent lineages over multi-century horizons remains the open path-degeneracy problem that ocean-state inflation seeks to address (Section 3.8), but it is not a precondition for skill at the horizons tested here.

The reconstruction is coherent across the whole overturning plane, not only at the index latitude (Figure 6). When the Atlantic streamfunction $\psi(\mathrm{depth}, \mathrm{latitude})$ is scored cell by cell against the truth over all 300 years, the assimilated ensemble mean has a lower error than the free-running one in 92.3% of the 1234 scoreable cells. It also beats the truth's own 300-year climatology in 96.2% of them, against 73.7% for FREE. Over the deep-cell box the Atlantic-network comparison already uses (0–60°N below 500 m), the area-weighted root-mean-square error falls from 1.196 to 0.887 Sv and the mean-square skill score rises from $+0.065$ to $+0.405$. The gain is largest between about 25 and 45°N at 0.8–3 km, which is where the North Atlantic Deep Water cell sits, and it is smallest in the wind-driven surface layer and in the abyss. That pattern is what a surface-temperature constraint should produce: the filter improves the density-driven circulation that surface temperature is linked to, and leaves the parts of the plane that carry no multiyear memory of it largely unchanged. A single index at one latitude could not show that the whole circulation is reconstructed coherently. The plane does.

3.4.1 Does assimilation add information beyond an offline linear reconstruction?

We compare the online filter with a linear reconstruction using the same noisy 59-site pseudo-observations that the filter processes. At analysis times, a regression based on a single predictor relating subpolar-gyre SST to AMOC from these observations gives a correlation of $r = +0.40$ (95% CI $[+0.18, +0.60]$). This regression is fitted on the truth itself at those 60 times, so it is the most favorable linear relationship the data allow. The full 59-site ridge regression performs less well ($r = +0.27$) since the extra-Atlantic sites add noise without adding any signal, and even a regression on the noise-free field, which no real reconstruction could use, reaches only $r = +0.53$. The DA ensemble mean reaches $r = +0.82$ at those same times, exceeding the best linear map by $\delta = +0.42$ (95% CI $[+0.24, +0.62]$; no bootstrap resample places the filter below the regression).

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

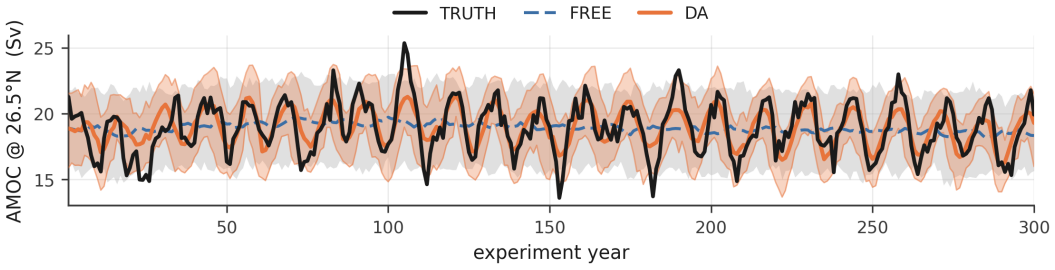
For the component between observations alone (each year's anomaly relative to the five-year block mean), the DA ensemble mean obtains $r = +0.72$ against the interpolated linear reconstruction's $r = +0.49$, a paired difference of $\delta = +0.23$ (95% CI $[+0.11, +0.36]$). We did not calculate an offline map based on the annual observations. Over the entire annual record, the median value among individual members is $r = +0.49$, which is already higher than the result from the single offline linear reconstruction. The ensemble mean ($r = +0.78$) exceeds even the best individual member ($r = +0.64$), which is consistent with the error cancellation expected in a coherent ensemble. The assimilated spread is mildly over-dispersed rather than over-confident (spread–skill ratio 1.20, 95% CI $[1.06, 1.38]$; Section S9 in the Supplement). The offline linear reconstruction yields $r = +0.33$ over the same full annual record (Figure 7). We did not carry out a full offline particle-filter benchmark, which would have assimilated the same observations into a fixed, non-propagated prior. The interpolated linear reconstruction stands in as the cheaper offline comparison. The filter also needs no calibration against the target, and it returns a complete, dynamically consistent ocean state and an ensemble rather than a single index. Unlike a fixed linear map, it can also handle the nonlinear, non-Gaussian observation operators that real proxies require, and in the idealized test below it does better than the linear reconstruction for some of them. We examine this in the idealized coupled Lorenz-63 twin OSSE, where a nonlinear operator can be introduced at negligible cost (Section S4 in the Supplement); in the coupled-model experiments, the identity operator of Equation 11 is used. Under a saturating operator, the most physically realistic proxy nonlinearity, the filter reaches $r = +0.81$ compared to a regression giving $r = +0.40$ on identical noise-free observations (Figure 8). Under the identity operator, the same pair of values is $+0.81$ versus $+0.53$ (Table S3 in the Supplement). The advantage depends on the operator. In the absence of observation noise, a smooth cubic map makes the two methods indistinguishable, and a non-invertible sign-folding map prevents the filter from resolving the sign ambiguity. Table S3 in the Supplement gives all results, including those from the noisy-observation case in which the static map fails. Both the DA method and the linear reconstruction use the same model-specific relationship between SST and AMOC, so this shows added value in the perfect-model scenario and does not cover cross-model applicability.

3.5 Sensitivity to Data Assimilation Cycle Length

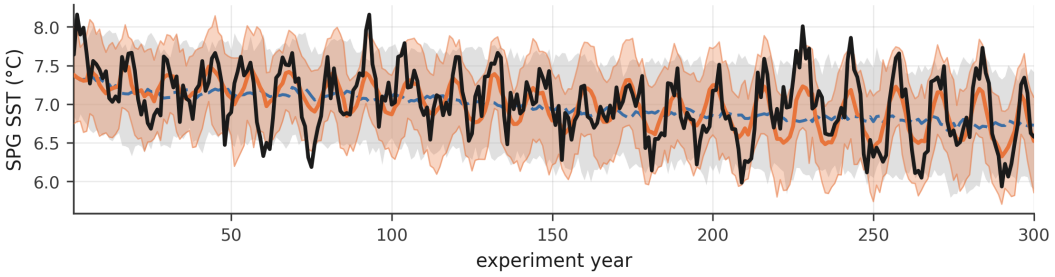
Of the three cycle lengths tested, the five-year cycle reconstructs the AMOC best and the ten-year cycle worst. The 1-, 5- and 10-year cycles bracket the usable-cycle bound of Equation 10.

The control AMOC spectrum, analyzed using Welch's method (Section 2.4; Figure S8 in the Supplement), shows a dominant period of approximately 13.3 years. This spectral peak exceeds the 95% confidence level of a first-order autoregressive (AR(1)) red-noise null model. The detected period also falls within the 12 to 28-year range reported for CMIP5/CMIP6 preindustrial simulations (Ma et al., 2023), where the timescale is determined by Rossby-wave propagation across the subpolar basin, influenced by the mean flow. A 3000-year control simulation run at twice the background diffusivity also yields a 13.3-year period, and the diverse ocean-IC ensemble, whose members differ by 50 model years, shows the same peak. The SST spectrum has the same peak too, so SST and AMOC reflect the same intrinsic coupled mode. Decadal to multidecadal AMOC variability is common in such models (Ferster et al., 2025).

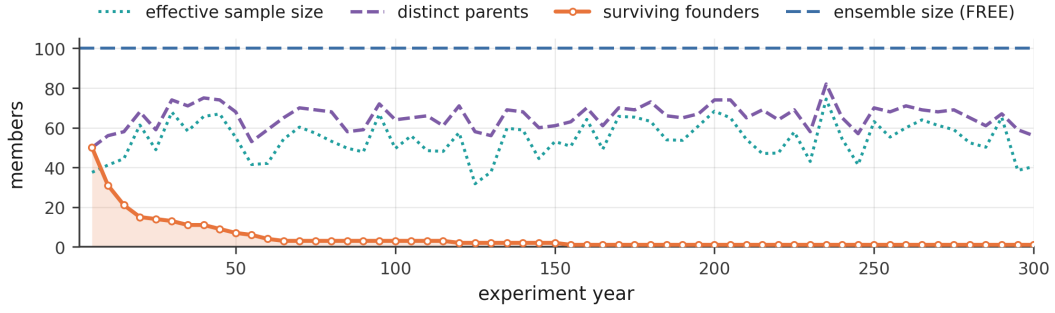
https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.



(a) AMOC



(b) Subpolar-gyre SST



(c) Ensemble diversity

Figure 5. CM2Mc-BLING reconstruction over the 300-year OSSE (diverse initial conditions, five-year cycles, global 59-proxy network; a second synthetic truth, branched from the same control state as the 100-year truth with a different perturbation). (a) Annual AMOC at 26.5°N (maximum below 500 m): truth (black), FREE ensemble mean (blue dashed) with its 5–95% band (gray), and DA ensemble mean (orange) with its 5–95% band. (b) Area-weighted subpolar-gyre SST, same color convention. (c) Three counts of ensemble membership, all in members out of the ensemble size. Teal dotted: the effective sample size at that analysis, written as an effective number of members. Purple dashed: the number of distinct members selected as parents in that cycle's resampling draw. Orange with circles: the number of members present at the start of the run that still have a descendant in the ensemble, traced through every draw back to the start (Section 2.3.1). The blue dashed line marks the 100-member ensemble size, which FREE keeps for the whole run because it never resamples. Skill statistics, genealogy timescales, and spread–reliability diagnostics are given in Section 3.4 and Section S9 in the Supplement.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

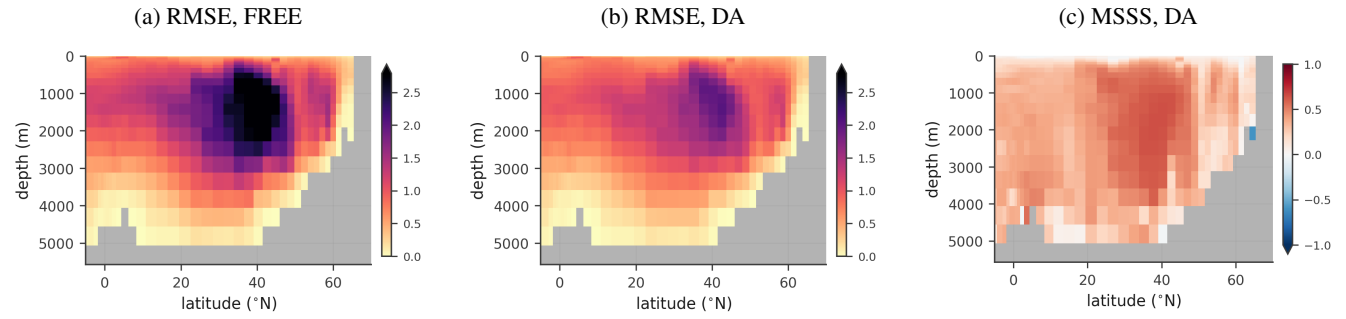


Figure 6. Error of the Atlantic meridional overturning streamfunction ψ in the depth–latitude plane over the 300-year OSSE (diverse initial conditions, five-year cycles, global 59-proxy network). Gray marks the seafloor and cells outside the Atlantic basin. (a, b) Root-mean-square error of the ensemble mean against the truth, for the free-running and the assimilated ensemble, on a common color scale. (c) Mean-square skill score of the assimilated ensemble mean, referenced to the truth's own 300-year climatology at each cell, so that positive values mark cells where the reconstruction improves on the long-term mean. Area-weighted values over the deep-cell box and the fraction of cells with positive skill are given in Section 3.4.

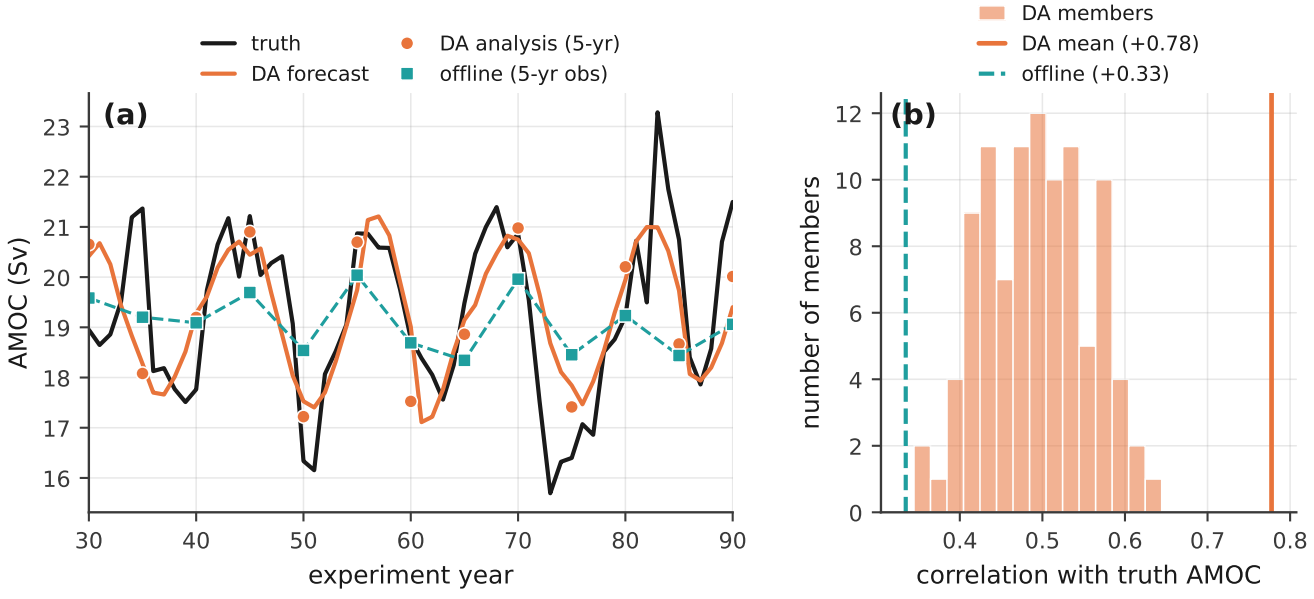


Figure 7. Online DA versus an offline linear reconstruction of the AMOC over the 300-year OSSE. The offline reconstruction is built from the same noisy 59-site pseudo-proxies the filter processes and is interpolated between the five-year observation times. (a) Years 30–90: truth AMOC (black), the DA forecast ensemble mean (orange) with the post-resampling DA *analysis* at each five-year cycle end (orange dots), and the offline linear reconstruction (teal dashed) with its values at the observation times (teal squares). (b) Histogram of the 100 DA members' full-record correlations with the truth AMOC, with the DA ensemble mean (solid) and the offline linear reconstruction (dashed) as reference lines. Skill statistics and their uncertainties are given in Section 3.4.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

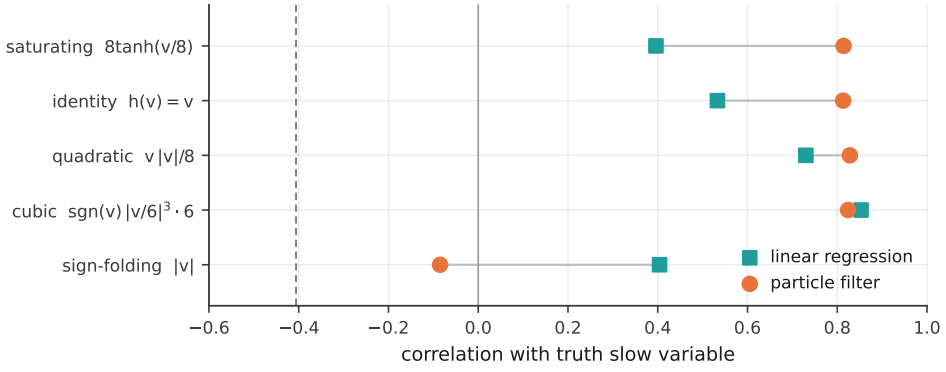


Figure 8. Reconstruction skill of the online particle filter (orange circles) and of a fair linear regression calibrated on the free-running ensemble alone (teal squares) on the coupled Lorenz-63 twin OSSE, under five observation operators h applied to the observed fast subsystem. Both estimators receive the identical noise-free standardized observations; values are means over three seeds, and the dashed line marks the free-running ensemble's own correlation with the target. Configuration, the noisy-observation counterpart, and the full sweep are given in the Supplement (Table S3); the results are discussed in Section 3.4.

The ten-year cycle performs worst under both initial-condition strategies, and the two strategies produce the same ranking of results (Table 3, Figure 9; autocorrelation-corrected p-values are provided in Section S3 in the Supplement). For similar initial conditions, performance is indistinguishable from its own FREE baseline ($\Delta r = +0.01$, 95% CI $[-0.13, +0.12]$), so assimilation at this cycle length adds nothing. Under diverse initial conditions, the ten-year cycle still improves ($\Delta r = +0.35$), but it is the smallest improvement among all cycles tested and the only diverse-IC interval whose confidence interval includes zero (Table 2).

At ten years the filter corrects itself too rarely. With at most 10 updates in 100 years, the ensemble runs free for most of an AMOC oscillation, the members diverge, and the filter loses the true phase. The ten-year reconstruction looks plausible, because the ensemble mean varies smoothly, but it does not follow the truth.

The five-year cycles outperform the annual cycles under both initial-condition strategies ($+0.69$ against $+0.52$ with similar ICs, $+0.54$ against $+0.50$ with diverse ones), and under diverse ICs the margin in correlation is small while the difference in the state of the filter is large. More frequent updates explain neither difference. The AMOC autocorrelation e-folding timescale in the control simulation is about 3 years, one quarter of the dominant period (Figure S3 in the Supplement). The annual cycle resamples and inflates before the ocean can respond to the previous correction (we did not measure the post-restart transient directly), and its weights are far more degenerate. Its mean effective sample size is 28% of N against 53% at five years, a single parent takes up to 48 of the 100 slots, and its ensemble collapses to one lineage by year 15 (Section 3.8). The five-year window spans this timescale and lets the coupled system equilibrate between updates.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 3. AMOC and SST reconstruction skill for all experiments, using the forecast-convention skill of Section 2.3. Correlations are evaluated over the full available record ($n = 100$ years for every experiment). Spread is the time-mean ensemble standard deviation of AMOC (Sv); "Spr. Red." is the percentage reduction in the DA AMOC spread relative to the FREE spread for the same IC strategy. Δr_{AMOC} is computed from the unrounded correlations, each experiment being differenced against its own FREE ensemble, and may therefore differ by ± 0.01 from the difference of the rounded r columns. The Atlantic-34 experiment assimilates the 34 Atlantic sites of the same 59-site network (Section 3.7).

IC Strategy	Cycle	AMOC r	SST r	Δr_{AMOC}	Spread (Sv)	Spr. Red.
Similar ocean IC	FREE	+0.52	+0.56	—	1.73	—
	1-yr	+0.52	+0.65	+0.00	0.86	50%
	5-yr	+0.69	+0.67	+0.17	1.40	19%
	10-yr	+0.53	+0.63	+0.01	1.66	4%
Diverse ocean IC	FREE	−0.14	+0.18	—	1.98	—
	1-yr	+0.50	+0.52	+0.65	0.92	54%
	5-yr	+0.54	+0.46	+0.68	1.70	14%
	10-yr	+0.21	+0.12	+0.35	1.91	4%
Diverse IC + ocean infl.	5-yr	+0.62	+0.51	+0.77	1.60	19%
Diverse IC, Atlantic-34 network	5-yr	+0.51	+0.45	+0.66	1.64	17%

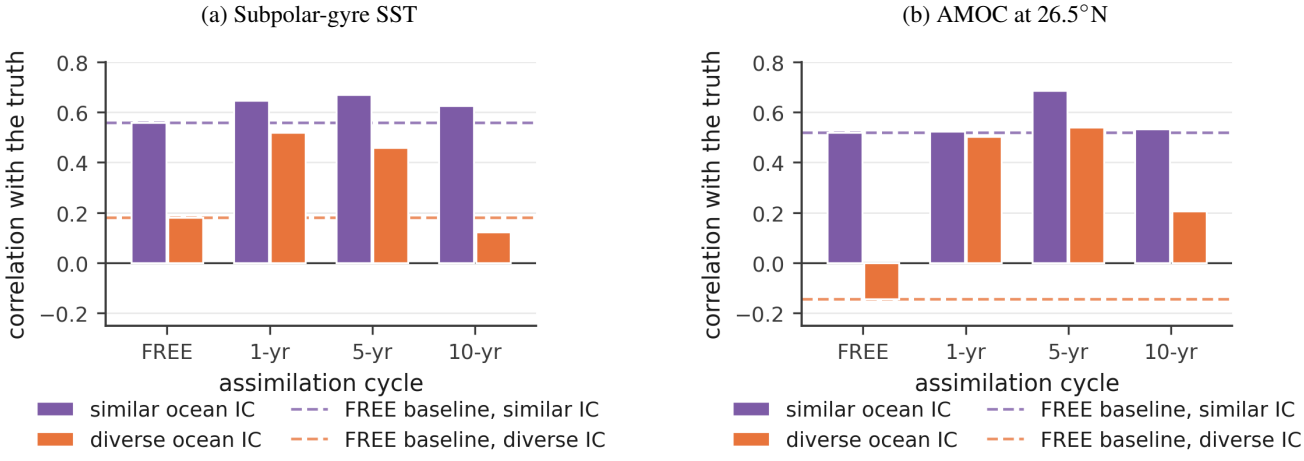


Figure 9. Correlation with the truth across assimilation cycle lengths for the similar ocean IC (purple) and diverse ocean IC (orange) ensembles: (a) subpolar-gyre SST and (b) AMOC at 26.5°N. Dashed lines mark the FREE baseline of each IC strategy. All correlations are evaluated over $n = 100$ years; the values are listed in Table 3.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

3.6 Sensitivity to Initial Condition Diversity

Assimilation improves skill only when the ensemble is initialized from diverse ocean states. The two strategies, similar and diverse ocean ICs, are defined in the Methods, along with their construction (Section 2.5.2, Figure S11 in the Supplement). Two subpolar-gyre SST metrics are relevant. The first is the correlation of the area-averaged SPG SST time series (the "SST r" column of Table 3). Diverse-IC DA reaches $r = +0.52$ at annual and $+0.46$ at five-year cycles, while similar-IC DA reaches $r = +0.67$. The second metric is the area-weighted mean of the per-gridcell correlation map over the SPG box. The corresponding DA minus FREE difference for this box, $\Delta r = +0.36$, is mapped over years 31–100 in Figure 3. This gridded metric is lower than the index correlation because it penalizes spatial-pattern errors, whereas index correlation requires only the area-mean phase to be correct.

The diverse-IC result is the relevant one, because the ocean state at the start of an actual reconstruction (for example, 850 CE) is unknown. The filter increases skill when initialized from independent ocean states. The absolute skill is lower, though. At the five-year cycle diverse-IC $r = +0.54$ stands against similar-IC $+0.69$, and at annual cycles $+0.50$ against $+0.52$ (Table 3). Subpolar-gyre SST skill is similar under both IC strategies, and surface variables adjust more rapidly than the overturning circulation. We recommend starting with a diverse prior and allowing a spin-up period. The AMOC index converges to the true value within the first two to three cycles (Figure 4), and the proxy-region gain approximately doubles when the first 30 years are excluded (Table S5 in the Supplement).

3.7 Restricting the Proxy Network to the Atlantic Basin

Assimilating only the 34 Atlantic proxies preserves the reconstruction skill of the full 59-site global network. The main experiments use the full network, but the physical constraint on the AMOC comes from the North Atlantic SST–AMOC fingerprint. To test whether restricting assimilation to this better-constrained basin changes the filter's performance, we repeated the diverse-IC, five-year-cycle experiment using only the 34 Atlantic pseudo-proxies (longitude $\in [-100°, 20°]$). Over the full 100-year record, the Atlantic-only DA reaches a forecast-convention AMOC correlation of $r = +0.51$ with the truth, and a root-mean-square error (RMSE) of 1.79 Sv (Figure 4, green line and band). This RMSE corresponds to a mean-square skill score relative to the truth's own climatology, $\mathrm{MSSS} = 1 - \mathrm{MSE}/\sigma^2_{\mathrm{truth}}$, of $+0.25$. For comparison, the global 59-proxy network gives $r = +0.54$ (RMSE 1.77 Sv, MSSS $+0.27$), while the unassimilated FREE ensemble produces $r = -0.14$ (RMSE 2.16 Sv, MSSS -0.09). Only the assimilated reconstructions outperform climatology.

The filter is also less degenerate. The importance weights are less concentrated, giving a mean effective sample size of 66% against 53% for the global network. Founders, defined as distinct ancestral trajectories traced to the start of the run, persist in greater number in the Atlantic-network ensemble. Seven remain after 100 years against three under the global network, and neither ensemble reaches a single lineage within the century (Section S9 in the Supplement). The heavier degeneracy under the global network likely arises from both physical and algorithmic factors. Physically, remote SST proxies from the Pacific, Indian, and Southern Oceans are dynamically decoupled from North Atlantic overturning, introducing non-informative

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

variance such as El Niño–Southern Oscillation (ENSO) or Southern Ocean modes into the likelihood. Algorithmically, the filter may discard particles with high AMOC accuracy in favor of those that better match unrelated remote SSTs.

The spatial pattern of the skill is nearly the same as with the global network (Figure 3c and d). The surface improvement stays confined to the subpolar North Atlantic, while the global-mean Δr remains negative. Over the subpolar-gyre box, the Atlantic-only network gives $\Delta r = +0.38$ for SST and $+0.27$ for SAT over years 31–100, against the global-network values of $+0.36$ and $+0.30$ (the global-network entries are in Table S5 in the Supplement; the Atlantic-network values are reported here only). In the overturning plane, the deep-cell gain matches the global network (area-weighted $\Delta r = +0.40$ versus $+0.41$ over 0–60°N below 500 m; Section S9 in the Supplement). Keeping only the observations in the basin that carries the SST–AMOC fingerprint gives the filter fewer observations to match, which eases weight degeneracy (Snyder et al., 2008), without losing skill. An Atlantic-only network is therefore a reasonable choice for the real-proxy reconstructions to follow.

3.8 Ocean-State Inflation and Ensemble Genealogy

The rate at which an SIR particle filter loses genealogical diversity depends on how often and how hard it resamples. Traced backward through the per-cycle draws, the ancestry of all 100 members contracts toward a single surviving lineage. Under annual cycling with diverse initial conditions that contraction completes within 15 years (within 41 years with similar ones); under a five-year cycle three lineages still survive after a century, and the 300-year integration reaches a single lineage near year 155 (Section 3.4). Annual cycling resamples five times as often and its weights are far more peaked: the effective sample size averages 28% of N (minimum 3%), against 53% for the five-year cycle (minimum 31%; Section S9 in the Supplement). This is path degeneracy rather than weight degeneracy, because the effective sample size stays sufficient at every analysis, averaging 54% over the 300-year run and never dropping below 32%. Even at the five-year cycle the filter loses most of its genealogical diversity within a few decades, while its skill is unaffected (Section 3.4). Whether that still holds over a millennium is untested, which is why we looked for a way to preserve diversity.

To assess whether perturbing the upper-ocean state after resampling could maintain genealogical diversity, we ran an ocean-inflation experiment (Diverse-5yr, +ocean) with 100 members spanning 100 years, using five-year cycles and the same diverse initial conditions, truth and proxy network as the headline experiment. After each resampling step, we perturbed ocean temperature and salinity over model levels 2 to 15 (15–514 m), excluding the surface level. The base perturbation amplitudes were 0.01°C and 0.003 psu, decaying exponentially over approximately eight levels (about 90 m) to zero below 684 m. We applied the same effective sample size (ESS) scaling as for the atmosphere, capped at three times the base amplitude. This cap was reached at every event, resulting in applied amplitudes of 0.03°C and 0.009 psu throughout.

Over its 100-year record the Diverse-5yr (+ocean) experiment ended with two surviving lineages, against three for the plain five-year filter on the same truth, so the perturbation did not measurably slow genealogical erosion at this amplitude. It did, however, raise reconstruction skill: the +ocean experiment reached an AMOC correlation of $r = +0.62$ (detrended $+0.60$) against $r = +0.54$ without ocean inflation, the largest gain over FREE in the campaign ($\Delta r = +0.77$, 95% CI $[+0.52, +1.02]$) and its smallest adjusted p-value ($q = 0.009$; Table 2 and Section 2.8). We did not test the $+0.08$ difference between the two experiments themselves, and it is well inside the width of either confidence interval. Given only one realization of each

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

experiment, this is suggestive rather than conclusive. Once a lineage is lost it cannot be recovered, so any resampling filter will contract over multi-century timescales, and whether upper-ocean perturbation can delay that collapse remains an open question at the single amplitude tested here. The perturbation left the assimilated spread at 1.60 Sv against 1.70 Sv without it (Table 3), both close to the ensemble-mean error and within a fifth of the 1.98 Sv FREE spread (Section S9 in the Supplement).

Only the $0.01°$C base amplitude was run at a length sufficient to resolve the AMOC correlation, so the dependence on perturbation amplitude remains untested.

4 Discussion

This study tests online particle-filter assimilation for AMOC reconstruction in a perfect-model setting. The filter recovers internal AMOC variability from sea surface temperature, and over three centuries its skill is significant on its own, while the free-running ensemble stays uncorrelated with the truth. At observation times, it exceeds the best linear reconstruction built from the same noisy proxies. Three caveats limit what these results say about assimilating real proxies. First, the OSSE design removes model–observation mismatch, as pseudo-observations are generated from the same model as the ensemble. The skill estimates are therefore upper bounds. Second, the 59-proxy network was optimized for SST coverage in AMOC-sensitive regions, including 10 sites in the subpolar gyre, which is the center of action for the leading coupled SST–AMOC pattern (Maximum Covariance Analysis Mode 1, accounting for 75% of the squared covariance). The actual Past Global Changes (PAGES) 2k temperature database (PAGES2k Consortium, 2017), by contrast, contains only a few high-resolution records in the high-latitude North Atlantic. Third, the limited number of annually resolved SST archives available in this region does not match the accuracy of the uniformly assumed $\sigma_{\text{obs}} = 0.3°$C. All three simplifications work in the assimilation system's favor, so skill with real proxies will be lower than reported here.

The second question is whether these results transfer to other models. The reconstruction skill found here depends on the specific SST–AMOC fingerprint of the CM2Mc model. In CM2Mc, stronger overturning goes with warming of the subpolar gyre on the basin mean (box-mean correlation $r = +0.45$, Figures S1–S2 in the Supplement). However, the leading maximum-covariance mode links overturning with a west–east dipole across the northwest Atlantic, rather than uniform subpolar warming (Figure S10 and Section S5 in the Supplement). The sign and magnitude of this fingerprint are not universal, reflecting a balance between northward heat transport, which warms the subpolar surface under stronger AMOC, and local air–sea heat loss, which cools water that remains at the surface; deep convection itself brings relatively warm water up from below rather than cooling the surface. The localized negative AMOC–SST correlation in the Labrador Sea in this simulation is consistent with the surface heat loss that drives Labrador Sea deep-water formation and precedes AMOC intensification (Zhang et al., 2019). In ocean models where convective and local processes dominate the subpolar surface response, the basin-scale fingerprint may weaken or, in some regimes, reverse sign relative to CM2Mc. The surface-temperature fingerprint of AMOC change is also non-stationary and depends on external forcing (McMonigal et al., 2025). Since an SST-only particle filter constrains AMOC solely through this fingerprint, a model with a fingerprint of opposite sign to the target system would be driven toward the wrong overturning state, and skill would fall. This sign sensitivity is a serious constraint on real-proxy reconstruction, because

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

the assimilating model's SST–AMOC relationship may not match that of the real climate. The within-model experiments presented here do not test whether one model's fingerprint can constrain another model's overturning. The engine makes that testable. Because a model is coupled through the four-method interface rather than compiled into the filter, the same proxy network, observation operator and filter settings can be run on several models without changing the assimilation code. We therefore recommend running real-proxy reconstructions with several models rather than one, and reporting the spread of the reconstructed overturning across models as fingerprint-dependency uncertainty alongside the within-model ensemble spread. Agreement across models with different SST–AMOC fingerprints would point to an observation-driven reconstruction, and disagreement would expose a model dependence that a single model cannot detect.

When initial conditions are similar, trajectories remain comparable over decadal timescales regardless of DA, whereas diverse initial conditions generate a spread that DA can selectively reduce. This agrees with Okazaki et al. (2021), who showed that online DA outperforms offline DA when practical predictability exceeds the temporal resolution of observations. In contrast, Acevedo et al. (2017) reported no advantage for online DA in a slab-ocean model lacking multi-decadal memory.

Path degeneracy is the main problem over long horizons. Since atmospheric perturbations do not penetrate the ocean interior, the deep state that determines the AMOC remains nearly identical to the surviving founder, causing the genealogy to contract toward a single lineage (Section 3.4).

Random perturbations are expected to violate model balances (van Leeuwen et al., 2019), which in the ocean interior degrades the temperature–salinity density structures that govern deep-water formation. In our experiment, a small structure-blind perturbation at the one amplitude tested improved skill over a century-long record without measurably slowing genealogical erosion (Section 3.8). A single realization of each experiment is insufficient to establish either effect. So we do not claim that ocean-state inflation consistently buys skill, nor that it consistently trades skill for diversity. Regardless of whether such a trade-off is confirmed at certain amplitudes, the structure-preserving alternatives reviewed by van Leeuwen et al. (2019) are the natural next step, because they guide particles toward the observations instead of adding noise to resampled copies. Equivalent-weights particle filters adjust the proposal density so that particles converge near the observations with approximately equal weights, while localized particle filters (Poterjoy, 2016) update state variables only within a defined localization radius of each observation. Dynamically balanced perturbations, such as bred vectors or singular-vector increments, are another option, because they keep the model's density structure instead of perturbing it blindly.

The AMOC autocorrelation e-folding time is approximately 3 years, which determines the assimilation cycle length (Section 3.5). Four decorrelation measures are discussed in this paper: the e-folding time (3 years, setting the cycle bound); the integral timescale of about 2 years; the two-series Bretherton equivalent of approximately 5 years underlying N_{eff}; and the single-series equivalent of about 10 years used for the rank-histogram degrees of freedom in Section S9. The timescale governing the degrees of freedom for statistical significance is longer, around 5 years, as derived from the first-order autoregressive (lag-1, AR(1)) two-series Bretherton et al. (1999) N_{eff} estimator (Section S3 in the Supplement). This persistence reduces the 100-year record to roughly 20 effective degrees of freedom. The critical correlation at $p = 0.05$ for $N_{\mathrm{eff}} \approx 20$ is $r_{\mathrm{crit}} \approx 0.44$. Under diverse initial conditions, the assimilated global-network correlation exceeds this threshold ($r = +0.54$), as do the annual ($r = +0.50$), Atlantic-network ($r = +0.51$) and ocean-inflation ($r = +0.62$) experiments, while only the ten-year cycle

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

does not. In contrast, under similar initial conditions, all experiments surpass this threshold, including the free-running ensemble ($r = +0.52$), which is why that configuration supplies no usable baseline. In the headline experiment, the observed shift $\Delta r = r_{\mathrm{DA,5yr}} - r_{\mathrm{FREE}} = +0.68$ is matched by 1 of the 99 exhaustive circular-shift surrogates ($p = 0.02$) and by 1.5% of 100,000 phase-randomized surrogates ($p = 0.015$; Section 2.8). Under the Benjamini–Hochberg false-discovery-rate correction across the eight quantitatively scored CM2Mc campaign experiments (Table 2), three remain significant at $q < 0.05$: the ocean-inflation experiment ($q = 0.009$), the Atlantic network ($q = 0.02$) and the headline diverse-initial-condition five-year design ($q = 0.04$). The conservative circular-shift null cannot resolve below its own attainable bound of $p = 1/n = 0.01$; on that null the ocean-inflation and Atlantic-network experiments reach $q = 0.040$ and the headline experiment $q = 0.053$, but those values sit at the null's resolution floor and should not be read as independent confirmation. We therefore continue to present the campaign as a configuration ranking, and the 300-year integration of Section 3.4 as the confirmation, using a second reference and a window with no shared initial state with the campaign.

The error inflation factor $\eta = 4.0$ and the 85% ESS resampling threshold are tuning parameters. The η bracketing tests that support the choice of $\eta = 4.0$ are detailed in Section 2.3. A full $\eta \times \tau_{\mathrm{ess}}$ sweep was not affordable in the coupled model, so we ran it on the cheap Lorenz-63 testbed (Table S2 in the Supplement). There, reconstruction skill varies by less than 0.2 in r across both parameters wherever the filter resamples, and retained genealogical diversity increases with η, so on that testbed $\eta = 4$ sits inside a flat region rather than at a sharp optimum. In the coupled model, we did not evaluate the skill of the $\eta = 2$ and $\eta = 8$ brackets. The analysis also identifies a region to avoid: strong tempering combined with a low resampling threshold, where the effective sample size never drops below the threshold, and the filter never resamples. The 85% resampling threshold is higher than the 50% commonly used in the particle-filter literature. Path degeneracy accumulates over multiple resampling steps rather than through weight collapse at a single step (Section 3.4). At the five-year cycle, where the mean ESS is approximately 53% (Section S9 in the Supplement), a lower threshold would result in less frequent resampling. That is a cheap way to reduce degeneracy, and we leave it to a future $\tau_{\mathrm{ess}} = 0.5$ experiment.

5 Best Practices and Caveats

Four recommendations follow from the experiments. *(1) Start from a diverse ensemble.* Online particle filtering improves skill only when ensemble members represent distinct phases of internal variability. In contrast, an ensemble with similar initial conditions already tracks the truth, leaving little for assimilation to recover (Section 3.6). Diversity is advantageous only if the model's sampled states include the true ocean state. This condition is guaranteed in a perfect-model OSSE but cannot be confirmed in a real reconstruction. *(2) Align the assimilation cycle with the AMOC decorrelation time.* Among the three tested cycle lengths, a five-year cycle performed best. We recommend selecting a cycle longer than the AMOC decorrelation time but shorter than half the target's dominant period. The lower bound prevents resampling before the ocean responds to the previous correction, while the upper bound avoids aliasing of variability (Section 3.5). *(3) Monitor genealogy in addition to spread.* Persistent ensemble spread may mask a collapsing genealogy. Report distinct-founder counts alongside the spread and the rank histogram, and verify that the member copies performed by a cluster orchestrator reproduce the draw that the filter

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

logged (Section 2.3.1; Section S9 in the Supplement for the diagnostics themselves). *(4) Perturb both the upper ocean and the atmosphere, and verify within each model whether genealogical diversity is maintained.* Atmospheric-only inflation does not influence the deep state that determines the AMOC. Post-resampling ocean-state perturbation is therefore worth testing in any new model; in ours it raised reconstruction skill but did not slow the loss of lineages (Section 3.8), and a single realization at one amplitude cannot settle either effect.

The principal limitations are summarized as follows. Five are discussed in detail above: the perfect-model upper bound on all skill estimates, the subpolar-gyre-dense network compared with the actual PAGES 2k temperature database (PAGES2k Consortium, 2017), genealogical collapse of the assimilated ensemble, model dependence of the SST–AMOC fingerprint, and the limited statistical power of 100-year records at $N_{\text{eff}} \approx 20$. The assimilated spread is calibrated at the five- and ten-year cycles but overconfident at the annual ones, where the weights are most degenerate (Section S9 in the Supplement). Four additional limitations are directly relevant to real-proxy applications. (i) *Initial-condition and background dependence.* Skill improves only with diverse initial conditions and a sufficient effective ensemble size, as the similar-IC experiments show. The initial ocean state was also drawn from a control equilibrated at a different interior diffusivity than the OSSE integrations (Section 2.5.1). (ii) *Observation operator.* The pseudo-observations are composed of true SST at the proxy sites plus independent Gaussian noise (Equation 11, $\sigma_{\text{obs}} = 0.3°\text{C}$), that is, an identity operator with diagonal error covariance. This configuration isolates the filter (weighting, resampling, inflation, genealogical collapse) from forward-model error. Real proxies, by contrast, require a proxy system model (PSM) with a calibrated, often nonlinear forward operator and correlated error (Fischer, 2024). A PSM is incorporated through the `observe` operator. Correlated or non-Gaussian proxy error is introduced via the engine's user-supplied log-weight hook (`assimilate_log_weights`), rather than through the diagonal Gaussian cost of Equation 4. The engine therefore accepts a PSM without modification, like existing paleoclimate assimilation tools (King et al., 2023; Zhu et al., 2024). We defer real-proxy AMOC reconstruction to a planned follow-up study.

(iii) *Unforced truth versus forced target.* The OSSE assesses reconstruction of *internal* AMOC variability exclusively, as both the truth and control are derived from a constant-forcing integration. The motivating last-millennium target is externally forced (volcanic, solar, greenhouse-gas), where forced variability may dominate. Therefore, the OSSE does not evaluate skill in reconstructing forced variability. (iv) *Real-proxy realism of timing and resolution.* The OSSE assumes perfect proxy timing on annual-to-five-year cycles, whereas real records have age-model uncertainties of decades to centuries, and subpolar North Atlantic marine sediments resolve only decadal to centennial variability. The real proxy network therefore supports neither the assumed chronology nor the cycle length.

The 100-year OSSE does not show that the 1000-year last-millennium reconstruction that motivates this study is feasible, because it covers fewer than eight AMOC oscillation cycles. Each 100-year experiment requires 10,000 member-years of coupled ESM integration, corresponding to approximately 3.6×10^4 core-hours on the PIK HPC cluster at a measured throughput of 3.6 core-hours per member-year (Section S7 in the Supplement). A 1000-year experiment would require ten times more resources, and the entire campaign, including spin-up, controls, and restarted integrations, would need a much larger allocation. Long-term genealogical stability and an ocean-state-inflation amplitude sweep of the same length are therefore left to future work. Future work should address six priorities: (1) an offline particle-filter benchmark using identical proxy networks; (2) a

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

skill benchmark against the online emulator-based pseudo-proxy experiment of Meng and Hakim (2024), the closest comparable experiment; (3) experiments with realistic proxy distributions from the PAGES 2k database; (4) a within-model robustness test in a second, structurally independent model, and extension to cross-model reconstruction, which is closest to real-proxy reconstruction; (5) implementation of localized (Poterjoy, 2016) or equivalent-weights (van Leeuwen et al., 2019) particle filters within the coupled ESM; and (6) real-proxy AMOC reconstruction using calibrated proxy system models supplied to the engine's `observe` operator (King et al., 2023; Zhu et al., 2024; Fischer, 2024).

6 Conclusions

`pypfda` v1.0 is an open-source, model-agnostic online particle-filter engine developed for sequential data assimilation in coupled Earth system models. Its four-method interface couples it to any model without recompiling either side, and because the filter runs online it needs no offline calibration, surrogate model, or pre-computed prior.

This study makes three main contributions. First, it connects a particle filter to a coupled Earth system model using a Python interface rather than compiled-in coupling, which enables the addition of a model without modifying or recompiling either the model or the filter. Second, the algorithm, executed through a cluster-side implementation of the interface, scales to 100-member ensembles over multiple centuries in a coupled Earth system model and treats ensemble genealogy as a primary diagnostic. Third, under a saturating observation operator the filter recovers skill that the linear reconstruction tested here does not. The filter also performs better under the identity operator on the testbed; however, this advantage is operator-dependent, disappearing for a smooth cubic map and reversing for a non-invertible sign-folding operator.

The central finding is that an online particle filter can recover AMOC variability from surface temperature alone, provided the ensemble starts from diverse ocean states and the cycle length matches the ocean's intrinsic timescales. Among the three cycle lengths evaluated, a five-year cycle had the highest skill. Shorter cycles resampled before the ocean could respond to previous corrections, while longer cycles lost synchronization with the decadal oscillation. Although the filter reconstructs the AMOC well, it loses genealogical diversity over a few decades, reaching a single lineage after about 155 years in the 300-year integration. Its spread, by contrast, is mildly over-dispersed rather than overconfident (spread–skill ratio 1.20 over that integration, against 1.00 for FREE). Perturbing the upper ocean after resampling gave the largest skill gain of the campaign at the only amplitude scored over a full century, without measurably slowing the loss of lineages. Whether any amplitude can defer genealogical collapse over centuries remains an open question.

These results are a perfect-model baseline for online particle-filter reconstruction of the AMOC from surface proxies, and they expose one limitation. Atmospheric-only inflation, the standard approach to mitigating ensemble collapse, does not affect the deep ocean state that determines the AMOC. Sustaining ensemble diversity over multi-century timescales will likely require shaping the proposal density, rather than solely inflating the posterior, as nudging-proposal particle filters in LOVECLIM have shown (Dubinkina and Goosse, 2013). The engine is built for that. `pypfda` already modifies importance weights through tempering and a maximum-weight cap, and exposes the perturbation kernel via the same model interface used for coupling models. This framework can implement a localized or equivalent-weights update by replacing the weighting and resampling

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

stage without altering model coupling, cost computation, or diagnostics. The per-cycle ancestry record lets any claimed gain in diversity be checked directly.

Code and data availability. **Engine.** `pypfda` v1.0 is archived on Zenodo at https://doi.org/10.5281/zenodo.21281526 (Fallah et al., 2026b) under the MIT License; that deposit, which includes the documentation and the test suite, is the version used in this study. Development continues at https://github.com/bijanf/pypfda (tag `v1.0`), and the user manual is at https://bijanf.github.io/pypfda/.

Experiment code and diagnostics. The cluster-side orchestrator, the cost module, the run configurations, the figure and table scripts, and the diagnostics behind every reported number are deposited on Zenodo at https://doi.org/10.5281/zenodo.22788425 (Fallah et al., 2026a) under the Creative Commons Attribution 4.0 License. The orchestrator is `run_online_da.py`, and the run configurations are the CM2Mc-BLING `input.nml` and `field_table`. The diagnostics comprise the truth, FREE and DA ensemble series of AMOC and subpolar-gyre SST, the effective-sample-size, weight and genealogy histories, the significance family, the pseudo-observation network, and the derived fingerprint-regression series; a README and scripts recompute the headline skill and significance results from the deposited arrays alone. The figure scripts document the processing steps and address cluster paths. Because the coupled model is not bit-for-bit reproducible across restarts, the deposited diagnostics are the definitive record of the results.

Model. CM2Mc-BLING (Galbraith et al., 2010, 2011) is not our software, and its developers do not distribute the source publicly; the version used is commit `e664daa1fff332e81b6adf044f50bdc20880f7dd`, recorded in the deposited run logs. As the GMD Code and Data Policy requires for restricted code, we state that restriction here, and we will give the editor and reviewers confidential access to that source tree through the handling editor. The model was run with its distributed input datasets, unmodified, with land cover, aerosols and ozone held at 1860 values.

Other data. The raw coupled-model output (∼16 TB) exceeds the capacity of any repository meeting the GMD archiving criteria; access for the editor and reviewers is arranged through the handling editor, as documented in the deposit's `AVAILABILITY.md`. The SODA 2.2.4 reanalysis (Carton and Giese, 2008; Giese and Ray, 2011) is distributed by its providers at https://soda.umd.edu, and the derived series used in Section S1 in the Supplement are in the deposit. The PAGES 2k database is available from PAGES2k Consortium (2017), and the RAPID–MOCHA data (McCarthy et al., 2015; Smeed et al., 2018) from the British Oceanographic Data Centre at https://doi.org/10.5285/48d0bf43-0598-ceb2-e063-7086abc062f1 (Moat et al., 2026).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/gmd-0-1-2026-supplement.

Author contributions. B. Fallah conceived the study, developed the `pypfda` engine, designed and ran the OSSEs, performed the analysis, and wrote the manuscript. W. Huiskamp contributed to the model configurations, diagnostics, and analysis. H. Goosse and W. Huiskamp supervised the work and contributed to the interpretation. All authors discussed the results and edited the manuscript.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. The authors thank the German Federal Ministry of Education and Research and the State of Brandenburg for supporting this project by providing resources on the high-performance computer system at the Potsdam Institute for Climate Impact Research. The authors thank Stefan Rahmstorf for discussing the results, and colleagues at PIK and the Earth and Life Institute, UCLouvain, for helpful discussions. The authors wrote the text; Grammarly checked grammar, spelling, and word choice. We used Claude (Anthropic) as an AI coding assistant to help lint, add docstrings to, and review the research software underlying this study (the `pypfda` engine and analysis scripts); AI generated no data, numerical results, or scientific interpretation. BF inspected and validated all AI-assisted code and assumes full responsibility for the software and scientific content.

Financial support. This research has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the AMOC-Millennia project (grant no. 550037087).

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Acevedo, W., Fallah, B., Reich, S., and Cubasch, U.: Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model, Climate of the Past, 13, 545–557, https://doi.org/10.5194/cp-13-545-2017, 2017.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., and Avellano, A.: The Data Assimilation Research Testbed: A Community Facility, Bulletin of the American Meteorological Society, 90, 1283–1296, https://doi.org/10.1175/2009BAMS2618.1, 2009.

Atlas, R.: Atmospheric Observations and Experiments to Assess Their Usefulness in Data Assimilation, Journal of the Meteorological Society of Japan, 75, 111–130, https://doi.org/10.2151/jmsj1965.75.1B_111, 1997.

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society: Series B (Methodological), 57, 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x, 1995.

Bocquet, M. and Carrassi, A.: Four-Dimensional Ensemble Variational Data Assimilation and the Unstable Subspace, Tellus A: Dynamic Meteorology and Oceanography, 69, 1304 504, https://doi.org/10.1080/16000870.2017.1304504, 2017.

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., and Bladé, I.: The effective number of spatial degrees of freedom of a time-varying field, Journal of Climate, 12, 1990–2009, https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2, 1999.

Browne, P. A. and van Leeuwen, P. J.: Twin experiments with the equivalent weights particle filter and HadCM3, Quarterly Journal of the Royal Meteorological Society, 141, 3399–3414, https://doi.org/10.1002/qj.2621, 2015.

Browne, P. A. and Wilson, S.: A simple method for integrating a complex model into an ensemble data assimilation system using MPI, Environmental Modelling & Software, 68, 122–128, https://doi.org/10.1016/j.envsoft.2015.02.003, 2015.

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V.: Observed Fingerprint of a Weakening Atlantic Ocean Overturning Circulation, Nature, 556, 191–196, https://doi.org/10.1038/s41586-018-0006-5, 2018.

Carpenter, J., Clifford, P., and Fearnhead, P.: Improved particle filter for nonlinear problems, IEE Proceedings – Radar, Sonar and Navigation, 146, 2–7, https://doi.org/10.1049/ip-rsn:19990255, 1999.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data Assimilation in the Geosciences: An Overview of Methods, Issues, and Perspectives, Wiley Interdisciplinary Reviews: Climate Change, 9, e535, https://doi.org/10.1002/wcc.535, 2018.

Carton, J. A. and Giese, B. S.: A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA), Monthly Weather Review, 136, 2999–3017, https://doi.org/10.1175/2007MWR1978.1, 2008.

Chen, Y., Nerger, L., and Lawless, A. S.: A Python interface to the Fortran-based Parallel Data Assimilation Framework: pyPDAF v1.0.2, Geoscientific Model Development, 18, 8235–8252, https://doi.org/10.5194/gmd-18-8235-2025, 2025.

Cluzet, B., Lafaysse, M., Cosme, E., Albergel, C., Meunier, L.-F., and Dumont, M.: CrocO_v1.0: a particle filter to assimilate snowpack observations in a spatialised framework, Geoscientific Model Development, 14, 1595–1614, https://doi.org/10.5194/gmd-14-1595-2021, 2021.

Doucet, A., de Freitas, N., and Gordon, N., eds.: Sequential Monte Carlo Methods in Practice, Statistics for Engineering and Information Science, Springer, New York, ISBN 978-0-387-95146-1, https://doi.org/10.1007/978-1-4757-3437-9, 2001.

Dubinkina, S. and Goosse, H.: An assessment of particle filtering methods and nudging for climate state reconstructions, Climate of the Past, 9, 1141–1152, https://doi.org/10.5194/cp-9-1141-2013, 2013.

Dubinkina, S., Goosse, H., Sallaz-Damaz, Y., Crespin, E., and Crucifix, M.: Testing a Particle Filter to Reconstruct Climate Changes over the Past Centuries, International Journal of Bifurcation and Chaos, 21, 3611–3618, https://doi.org/10.1142/S0218127411030763, 2011.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, https://doi.org/10.1029/94JC00572, 1994.

Fallah, B., Huiskamp, W., Goosse, H., and Rostami, M.: Online particle-filter OSSE diagnostics and experiment code for AMOC reconstruction [Dataset], Zenodo, https://doi.org/10.5281/zenodo.22788425, Creative Commons Attribution 4.0 License, 2026a.

Fallah, B., Rostami, M., Huiskamp, W., Goosse, H., and Rahmstorf, S.: pypfda: An Online Particle-Filter Data Assimilation Engine in Pure Python [Software], Zenodo, https://doi.org/10.5281/zenodo.21281526, Version v1.0; MIT License, 2026b.

Ferster, B. S., Fedorov, A. V., Mignot, J., and Guilyardi, E.: AMOC Variability in Climate Models and Its Dependence on the Mean State, Geophysical Research Letters, 52, https://doi.org/10.1029/2024GL110356, 2025.

Fischer, M. J.: The Measurement Error Proxy System Model: MEPSM v0.2, Geoscientific Model Development, 17, 6745–6760, https://doi.org/10.5194/gmd-17-6745-2024, 2024.

Galbraith, E. D., Gnanadesikan, A., Dunne, J. P., and Hiscock, M. R.: Regional impacts of iron-light colimitation in a global biogeochemical model, Biogeosciences, 7, 1043–1064, https://doi.org/10.5194/bg-7-1043-2010, 2010.

Galbraith, E. D., Kwon, E. Y., Gnanadesikan, A., Rodgers, K. B., Griffies, S. M., Bianchi, D., Sarmiento, J. L., Dunne, J. P., Simeon, J., Slater, R. D., Wittenberg, A. T., and Held, I. M.: Climate variability and radiocarbon in the CM2Mc Earth System Model, Journal of Climate, 24, 4230–4254, https://doi.org/10.1175/2011JCLI3919.1, 2011.

Giese, B. S. and Ray, S.: El Niño variability in simple ocean data assimilation (SODA), 1871–2008, Journal of Geophysical Research: Oceans, 116, C02 024, https://doi.org/10.1029/2010JC006695, 2011.

Giles, D., Graham, M. M., Giordano, M., Koskela, T., Beskos, A., and Guillas, S.: ParticleDA.jl v.1.0: a distributed particle-filtering data assimilation package, Geoscientific Model Development, 17, 2427–2445, https://doi.org/10.5194/gmd-17-2427-2024, 2024.

Goosse, H.: Reconstructed and simulated temperature asymmetry between continents in both hemispheres over the last centuries, Climate Dynamics, 48, 1483–1501, https://doi.org/10.1007/s00382-016-3154-z, 2017.

Goosse, H., Crespin, E., Dubinkina, S., Loutre, M.-F., Mann, M. E., Renssen, H., Sallaz-Damaz, Y., and Shindell, D.: The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly", Climate Dynamics, 39, 2847–2866, https://doi.org/10.1007/s00382-012-1297-0, 2012.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M.: Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation, IEE Proceedings F - Radar and Signal Processing, 140, 107–113, https://doi.org/10.1049/ip-f-2.1993.0015, 1993.

Grudzien, C., Bocquet, M., and Carrassi, A.: On the numerical integration of the Lorenz-96 model, with scalar additive noise, for benchmark twin experiments, Geoscientific Model Development, 13, 1903–1924, https://doi.org/10.5194/gmd-13-1903-2020, 2020.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The Last Millennium Climate Reanalysis Project: Framework and First Results, Journal of Geophysical Research: Atmospheres, 121, 6745–6764, https://doi.org/10.1002/2016JD024751, 2016.

Haussaire, J.-M. and Bocquet, M.: A low-order coupled chemistry meteorology model for testing online and offline data assimilation schemes: L95-GRS (v1.0), Geoscientific Model Development, 9, 393–412, https://doi.org/10.5194/gmd-9-393-2016, 2016.

Jebri, B. and Khodri, M.: Large ensemble particle filter for spatial climate reconstructions using a linear inverse model, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 094, https://doi.org/10.1029/2022MS003094, 2023.

King, J., Tierney, J., Osman, M., Judd, E. J., and Anchukaitis, K. J.: DASH: a MATLAB toolbox for paleoclimate data assimilation, Geoscientific Model Development, 16, 5653–5683, https://doi.org/10.5194/gmd-16-5653-2023, 2023.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.



Kitagawa, G.: Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, Journal of Computational and Graphical Statistics, 5, 1–25, https://doi.org/10.1080/10618600.1996.10474692, 1996.

Lorenz, E. N.: Predictability: A Problem Partly Solved, in: Proceedings of the Seminar on Predictability, vol. 1, pp. 1–18, ECMWF, Reading, UK, reprinted 2006 in: Predictability of Weather and Climate, edited by: Palmer, T. and Hagedorn, R., Cambridge University Press, 40–58, https://doi.org/10.1017/CBO9780511617652.004, 1996.

Lorenz, E. N. and Emanuel, K. A.: Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model, Journal of the Atmospheric Sciences, 55, 399–414, https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2, 1998.

Ma, X., Huang, G., Li, X., and Li, S.: The potential mechanisms of the dominant timescale of AMOC multidecadal variability in CMIP6/CMIP5 preindustrial simulations, Climate Dynamics, 60, 2131–2145, https://doi.org/10.1007/s00382-022-06440-9, 2023.

Masoum, A., Nerger, L., Willeit, M., Ganopolski, A., and Lohmann, G.: Paleoclimate data assimilation with CLIMBER-X: An ensemble Kalman filter for the last deglaciation, PLoS ONE, 19, e0300 138, https://doi.org/10.1371/journal.pone.0300138, 2024.

Matsikaris, A., Widmann, M., and Jungclaus, J.: On-line and off-line data assimilation in palaeoclimatology: a case study, Climate of the Past, 11, 81–93, https://doi.org/10.5194/cp-11-81-2015, 2015.

McCarthy, G. D., Smeed, D. A., Johns, W. E., Frajka-Williams, E., Moat, B. I., Rayner, D., Baringer, M. O., Meinen, C. S., Collins, J., and Bryden, H. L.: Measuring the Atlantic Meridional Overturning Circulation at 26°N, Progress in Oceanography, 130, 91–111, https://doi.org/10.1016/j.pocean.2014.10.006, 2015.

McCarthy, G. D., Hug, G., Smeed, D., Morris, K. J., and Moat, B.: Signal and Noise in the Atlantic Meridional Overturning Circulation at 26°N, Geophysical Research Letters, 52, e2025GL115 055, https://doi.org/10.1029/2025GL115055, 2025.

McMonigal, K., Larson, S. M., Gervais, M., Klavans, J. M., He, C., Cane, M. A., Corti, S., and Bellomo, K.: Fingerprints of AMOC Decline Are Sensitive to External and Mechanistic Forcing, Geophysical Research Letters, 52, https://doi.org/10.1029/2025GL116307, 2025.

Meng, Z. and Hakim, G. J.: Reconstructing the Tropical Pacific Upper Ocean Using Online Data Assimilation With a Deep Learning Model, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 422, https://doi.org/10.1029/2024MS004422, 2024.

Meng, Z., Hakim, G. J., and Steig, E. J.: Coupled seasonal data assimilation of sea ice, ocean, and atmospheric dynamics over the last millennium, Journal of Climate, 38, 7229–7247, https://doi.org/10.1175/JCLI-D-25-0048.1, 2025.

Moat, B. I., Smeed, D., Rayner, D., Johns, W. E., Smith, R. H., Volkov, D. L., Elipot, S., Petit, T., Kajtar, J. B., Baringer, M. O., and Collins, J.: Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2024 (v2024.1a) [Dataset], NERC EDS British Oceanographic Data Centre NOC, https://doi.org/10.5285/48d0bf43-0598-ceb2-e063-7086abc062f1, 2026.

Nerger, L. and Hiller, W.: Software for ensemble-based data assimilation systems—Implementation strategies and scalability, Computers & Geosciences, 55, 110–118, https://doi.org/10.1016/j.cageo.2012.03.026, 2013.

Nerger, L., Tang, Q., and Mu, L.: Efficient ensemble data assimilation for coupled models with the Parallel Data Assimilation Framework: example of AWI-CM (AWI-CM-PDAF 1.0), Geoscientific Model Development, 13, 4305–4321, https://doi.org/10.5194/gmd-13-4305-2020, 2020.

Okazaki, A., Miyoshi, T., Yoshimura, K., Greybush, S. J., and Zhang, F.: Revisiting Online and Offline Data Assimilation Comparison for Paleoclimate Reconstruction: An Idealized OSSE Study, Journal of Geophysical Research: Atmospheres, 126, e2020JD034 214, https://doi.org/10.1029/2020JD034214, 2021.

PAGES2k Consortium: A Global Multiproxy Database for Temperature Reconstructions of the Common Era, Scientific Data, 4, 170 088, https://doi.org/10.1038/sdata.2017.88, 2017.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.

Perkins, W. A. and Hakim, G. J.: Coupled atmosphere–ocean reconstruction of the last millennium using online data assimilation, Paleoceanography and Paleoclimatology, 36, e2020PA003 959, https://doi.org/10.1029/2020PA003959, 2021.

Petit, T., Robson, J., Ferreira, D., Yeager, S. G., and Evans, D. G.: Coherence of the AMOC Over the Subpolar North Atlantic on Interannual to Multiannual Time Scales, Geophysical Research Letters, 52, e2025GL115 171, https://doi.org/10.1029/2025GL115171, 2025.

Phipson, B. and Smyth, G. K.: Permutation P-values Should Never Be Zero: Calculating Exact P-values When Permutations Are Randomly Drawn, Statistical Applications in Genetics and Molecular Biology, 9, Article 39, https://doi.org/10.2202/1544-6115.1585, 2010.

Poterjoy, J.: A localized particle filter for high-dimensional nonlinear systems, Monthly Weather Review, 144, 59–76, https://doi.org/10.1175/MWR-D-15-0163.1, 2016.

Raanes, P. N., Chen, Y., and Grudzien, C.: DAPPER: Data Assimilation with Python: a Package for Experimental Research, Journal of Open Source Software, 9, 5150, https://doi.org/10.21105/joss.05150, 2024.

Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S., and Schaffernicht, E. J.: Exceptional Twentieth-Century Slowdown in Atlantic Ocean Overturning Circulation, Nature Climate Change, 5, 475–480, https://doi.org/10.1038/nclimate2554, 2015.

Shannon, C. E.: Communication in the Presence of Noise, Proceedings of the IRE, 37, 10–21, https://doi.org/10.1109/JRPROC.1949.232969, 1949.

Smeed, D. A., Josey, S. A., Beaulieu, C., Johns, W. E., Moat, B. I., Frajka-Williams, E., Rayner, D., Meinen, C. S., Baringer, M. O., Bryden, H. L., and McCarthy, G. D.: The North Atlantic Ocean is in a state of reduced overturning, Geophysical Research Letters, 45, 1527–1533, https://doi.org/10.1002/2017GL076350, 2018.

Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, WIREs Climate Change, 3, 63–77, https://doi.org/10.1002/wcc.149, 2012.

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J.: Obstacles to High-Dimensional Particle Filtering, Monthly Weather Review, 136, 4629–4640, https://doi.org/10.1175/2008MWR2529.1, 2008.

Steiger, N. J., Smerdon, J. E., Cook, E. R., and Cook, B. I.: A Reconstruction of Global Hydroclimate and Dynamical Variables over the Common Era, Scientific Data, 5, 180 086, https://doi.org/10.1038/sdata.2018.86, 2018.

Sun, H., Lei, L., Liu, Z., Ning, L., and Tan, Z.-M.: An online paleoclimate data assimilation with a deep learning-based network, Journal of Advances in Modeling Earth Systems, 17, e2024MS004 675, https://doi.org/10.1029/2024MS004675, 2025.

Tierney, J. E., Abram, N. J., Anchukaitis, K. J., Evans, M. N., Giry, C., Kilbourne, K. H., Saenger, C. P., Wu, H. C., and Zinke, J.: Tropical Sea Surface Temperatures for the Past Four Centuries Reconstructed from Coral Archives, Paleoceanography, 30, 226–252, https://doi.org/10.1002/2014PA002717, 2015.

Tierney, J. E., Judd, E. J., Osman, M. B., King, J. M., Truax, O. J., Steiger, N. J., Amrhein, D. E., and Anchukaitis, K. J.: Advances in Paleoclimate Data Assimilation, Annual Review of Earth and Planetary Sciences, 53, 625–650, https://doi.org/10.1146/annurev-earth-032320-064209, 2025.

van Leeuwen, P. J.: Particle Filtering in Geophysical Systems, Monthly Weather Review, 137, 4089–4114, https://doi.org/10.1175/2009MWR2835.1, 2009.

van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., and Reich, S.: Particle filters for high-dimensional geoscience applications: A review, Quarterly Journal of the Royal Meteorological Society, 145, 2335–2365, https://doi.org/10.1002/qj.3551, 2019.

Welch, P.: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics, 15, 70–73, https://doi.org/10.1109/TAU.1967.1161901, 1967.

https://doi.org/10.5194/egusphere-2026-5746
Preprint. Discussion started: 13 November 2026
© Author(s) 2026. CC BY 4.0 License.



Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G., Amrhein, D. E., and Little, C. M.: A Review of the Role of the Atlantic Meridional Overturning Circulation in Atlantic Multidecadal Variability and Associated Climate Impacts, Reviews of Geophysics, 57, 316–375, https://doi.org/10.1029/2019RG000644, 2019.

Zhu, F., Emile-Geay, J., Hakim, G. J., Guillot, D., Khider, D., Tardif, R., and Perkins, W. A.: cfr (v2024.1.26): a Python package for climate field reconstruction, Geoscientific Model Development, 17, 3409–3431, https://doi.org/10.5194/gmd-17-3409-2024, 2024.