Table 4. 4DVarNet performance on the GULFSTREAM and OSMOSIS domain compared to DUACS OI over the period from 22 October to 2 December 2012 (42 d).

Domain	Method	μ (RMSE)	σ (RMSE)	λx (degrees)	λt (days)	Train/test
	DUACS OI (one SWOT plus four nadirs)	0.92	0.01	1.22	11.31	–
GULFSTREAM	4DVarNet (one SWOT plus four nadirs)	0.96	0.01	0.62	5.29	GULFSTREAM/GULFSTREAM
	4DVarNet (one SWOT plus four nadirs)	0.95	0.01	0.86	5.67	GULFSTREAM2/GULFSTREAM
	4DVarNet (one SWOT plus four nadirs)	0.92	0.02	1.25	10.93	cNATL/GULFSTREAM
	DUACS OI (one SWOT plus four nadirs)	0.81	0.02	1.04	17.80	–
OSMOSIS	4DVarNet (one SWOT plus four nadirs)	0.89	0.02	0.35	6.84	OSMOSIS/OSMOSIS
	4DVarNet (one SWOT plus four nadirs)	0.88	0.02	0.41	8.05	OSMOSIS2/OSMOSIS
	4DVarNet (one SWOT plus four nadirs)	0.84	0.02	0.93	9.59	cNATL/OSMOSIS
	DUACS OI (one SWOT plus four nadirs)					–
cNATL	4DVarNet (one SWOT plus four nadirs)					cNATL/cNATL
	4DVarNet (one SWOT plus four nadirs)					cNATL/GULFSTREAM
	4DVarNet (one SWOT plus four nadirs)					cNATL/OSMOSIS

Table 5. 4DVarNet performance on the GULFSTREAM domain based on nine different training sessions with a random initialization of both Φ and Γ weights but similar training parameters (number of epochs, learning rates, optimizers, gradient steps, etc.) over the period from 22 October to 2 December 2012 (42 d).

Members	μ (RMSE)	σ (RMSE)	λx (degrees)	λt (days)
4DVarNet (no. 1)	0.96	0.01	0.68	5.16
4DVarNet (no. 2)	0.96	0.01	0.66	4.52
4DVarNet (no. 3)	0.96	0.01	0.62	4.66
4DVarNet (no. 4)	0.96	0.01	0.63	4.12
4DVarNet (no. 5)	0.96	0.01	0.87	4.92
4DVarNet (no. 6)	0.96	0.01	0.86	5.07
4DVarNet (no. 7)	0.96	0.01	0.68	5.18
4DVarNet (no. 8)	0.96	0.01	0.85	4.99
4DVarNet (no. 9)	0.96	0.01	0.62	5.29
4DVarNet (median)	0.96	0.01	0.67	4.62

dian field and the associated standard deviation. We report the performance metrics for the GULFSTREAM domain of the nine trained models and the median model in Table 5. It reveals the internal variability in the training process. Though it does not reach the best performance, the median model combines a resolved spatial scale below 0.7° and a resolved timescale below 5 d, which is only the case for 6 out of 9 of the trained models. Figure 12a further illustrates this aspect. Interestingly, the standard deviation of the ensemble of 4DVarNet-SSH schemes correlates to the interpolation error, with an R^2 coefficient of determination equal to 0.86. Even if the scales between the interpolation error and the training-related 4DVarNet internal variability differ (see Fig. 13), the latter can be regarded as an indicator of the interpolation error, usually with an appropriate localization of large errors. In future works, we plan to draw from traditional ensemble DA methods or ensemble Gaussian-based simulations to address all the components of the interpolation error related to the data assimilation scheme.

6 Conclusion and discussion

This paper introduced the 4DVarNet-SSH scheme, an end-to-end neural architecture for the space–time interpolation of SSH fields from nadir and wide-swath satellite altimetry data. The 4DVarNet-SSH scheme draws from recent methodological development to bridge data assimilation and deep learning with a view to training 4D-Var DA models and solvers from data. Numerical experiments within an OSSE setting support the relevance of the 4DVarNet-SSH scheme with respect to the state of the art.

We further discuss our main contributions according to three aspects, namely the added value of a deep learning scheme for satellite altimetry and operational oceanography, the exploitation of upcoming SWOT data, and the ability to scale up learning approaches from regional case studies to the global scale.

- *Deep learning for satellite altimetry and operational oceanography.* This study contributes to a growing research effort regarding the potential benefit of deep learning schemes for space and operational oceanography challenges (see, e.g., Ballarotta et al., 2020).