Comments

This manuscript develops a SwinTrans-ConvLSTM framework for 1–7-day SST forecasting and subsequent marine heatwave (MHW) identification in the East China Sea. The reported prediction accuracy is impressive, and the chronological train–validation–test split is appropriate. However, in its current version, I do not find the manuscript to provide sufficiently strong scientific or methodological novelty to justify publication. The claimed "physics-informed" contribution is substantially overstated, the physical interpretation of the learned processes is not adequately demonstrated, and several aspects of the experimental design and extreme-event verification raise important concerns. In particular, the study mainly represents an integration and optimization of existing deep-learning components rather than a major methodological advance, while the MHW analysis remains largely diagnostic rather than revealing new physical or physical–ecological mechanisms. I therefore recommend rejection.

Major Comments:

(1) The manuscript is presented as a physics-guided study of marine heatwave prediction, but the actual scientific contribution is primarily improved SST forecasting. MHWs are subsequently diagnosed from the predicted SST using a conventional percentile-based definition. No ecological variables, ecosystem responses, biological processes, or physical–ecological interactions are explicitly analyzed. More importantly, the manuscript does not identify a new physical mechanism controlling MHW development in the East China Sea. The discussions regarding the Western Pacific Subtropical High, air–sea heat exchange, Kuroshio intrusion, monsoon forcing, stratification, and coastal fronts mainly rely on mechanisms already established in previous studies. Therefore, the manuscript currently provides limited new process-level understanding of MHW dynamics.

(2) The manuscript repeatedly describes the framework and loss function as "physics-informed". However, the model does not explicitly impose governing physical equations, heat conservation, mixed-layer heat-budget constraints, advection–diffusion equations, or dynamically consistent momentum/thermodynamic balances. The wind components and air–sea temperature difference are simply added as predictor variables, while the gradient loss is essentially a spatial-structure regularization term rather than a physical conservation constraint. Therefore, the method would be more appropriately described as physics-guided or physics-aware feature engineering.

(3) Several sections infer that the model captures processes such as Ekman transport, wind-driven mixing, warm-water advection, Kuroshio intrusion, mixed-layer thermal memory, and air–sea heat exchange simply because wind vectors, ΔT, or a three-day moving average are included as inputs. These interpretations are plausible hypotheses, but they are not demonstrated by the experiments. For example, including u_10 and v_10 does not by itself prove that the network has learned Ekman transport or vertical mixing. Similarly, interpreting a three-day moving average as a direct analogue of mixed-layer thermal inertia is not physically validated. If process understanding is intended as a major contribution, additional diagnostics such as heat-budget analysis, feature attribution, perturbation/counterfactual experiments, or process-conditioned evaluation would be required. Otherwise, these mechanistic claims should be substantially weakened.

(4) There is a fundamental inconsistency between the forecasting setup and the physical interpretation of forecast-error growth. According to the experimental design, the model uses information from the preceding 14 days to predict SST for the subsequent 7 days. The manuscript therefore needs to state very clearly whether atmospheric variables during the forecast period are available to the model. If only past SST and past ERA5 atmospheric variables are used, then future atmospheric forcing over lead days 1–7 is not provided. In that case, statements attributing the degradation of long-lead forecasts to the "accumulation of uncertainty in atmospheric forcing and boundary conditions" are not logically consistent with the actual forecasting configuration. The more direct limitation would be the absence of future atmospheric forcing itself.

Conversely, if ERA5 atmospheric variables from the forecast period are used, then the experiment is not a true operational forecast because future reanalysis information would be available to the model, resulting in information leakage or an unrealistic hindcast advantage. The authors must explicitly clarify whether atmospheric predictors are past observations/reanalysis only or include future atmospheric fields. If future forcing is required, an operational experiment should instead use genuine numerical weather prediction forecasts. This issue directly affects the interpretation and practical validity of the claimed 7-day MHW forecasting capability.