**Ocean Science
Discussions**

Interactive comment on "NEMO on the shelf: assessment of the Iberia–Biscay–Ireland configuration" *by* C. Maraldi et al.

Anonymous Referee #2

This paper provides a broad brush comparison between a NEMO model simulation at an ocean margin and several diverse sources of observational information. While there are plenty of interesting comparisons here, there is a tendency for unquantified statements of model skill and unsubstantiated explanations for model deficiencies. For example, the explanation for the overestimation of the extent in stratified region in the Irish Sea and east UK coast is spurious on two accounts: first how this SST field can be used to accurately infer stratification in the first place and second the attribution of the deficiency to salinity seems very confused. The summary approach is a good way of giving the reader an overview of many aspects of the performance, but here is cluttered with lots of methodological detail which might be better in an appendix. While this paper could certainly be publishable, it suffers from some serious shortcomings, which need be address before it is acceptable for publication.

C125

1. The purpose of the paper is either unclear or not achieved. There is already an extensive literature on the methodology and application of model validation in shelf sea regions – this is not 'relatively uncharted territory', neither is the assessment of high frequency dynamics a 'new issue'. If the objective of this work is to propose a new set of metrics then these should be contrasted with what has been used before. However, I have some difficulty in seeing what is new in the methods of comparing model and observations here – most of the metrics seem common place, although the consideration of a comprehensive range of different components of the system is excellent, this needs to be better brought out from the results. What is perhaps missing is a clear statement of what/who these metrics are for: if they are to inform model development, then how and why these metrics, or if they are to inform the reliability of this information for users then which users, what decisions this will inform and again why these metrics? A list of the proposed metrics is needed and how the model performs against them, including a clear statement of what aspect and in which regions does the model perform well and what/where less so. For such a cross-comparison normalised metrics such as cost-functions and Taylor skill scores are a help e.g. in a summary table or diagram.

2. There is very little here to put this model simulation in context. With virtually no mention of other models in this region it is difficult to judge what the frequent, unquantified statements of 'this process is well modelled' actually means in the current state-of-the-art. The obvious, and straightforward, exercise would be to compare these results with the PSY2V3 model – asking does this finer resolution model 'add value' to the larger area simulation, if so in what aspect and by how much? Such a comparison would add substantially to the scientific interest in this paper. Beyond this a semi-qualitative comparison with other models in the literature (e.g. tides, SST RMS errors, fluxes across sections) would add substantial interest, e.g. to inform model choice/development.

3. This paper is far from ready for publication in terms of its presentation. It needs a very careful proof read for sense and meaning as it is very confused in some places,

some figures are missing (fig 18c and d), tables columns without explanation (Table 3 and 4) and missing units.

C127