Thank you for taking the time and care to provide valuable feedback and contributions to this manuscript. Please see our responses to the comments below, which we are ready to implement for a future revision.

Copy of review comments (RC) are given below, followed by the author comments (AC).

Reviewer 2:

RC40	Overview In this paper, the authors use complex network theory with outputs from a model simulation of the North-West European Shelf (NWES) to identify 1) spatial correlation length scales of biogeochemical variables, 2) geographical regions with strong spatial correlation within them and weak correlation between them and 3) correlations between biogeochemical variables. Point 1) is achieved by computing the Spearman's correlation coefficient between the time series of the different grid points. For point 2), for each variable, they build a spatial network with the previous coefficient, apply spectral graph clustering to gather grid-points and identify the boundaries of these clusters. Then, they define the regions base on the fraction of variables that have a boundary in each grid point. For point 3), they compute the Spearman's correlation coefficient between the spatial distributions of each variable, build a spatial network with that and use the spectral graph clustering to cluster biogeochemical variables. A first result of this work is to show that complex network theory can be used to identify biogeochemical regions based on spatial correlation or to identify correlation between biogeochemical variables. This is of interest for reducing the complexity of biogeochemical dynamics and for helping the analysis of simulations. The correlation length scales are of interest for data assimilation as it quantify the range of the influence between grid points. I very much appreciated to read the paper. It is clear and well written. The results are of interest and worth to be published. It presents an interesting way to analyse biogeochemical model outputs. The definition of biogeochemical provinces is particularly interesting as it can help the analysis of models. The methods are clearly explained. I do not have major comments on the paper, but rather a list of minor or specific comments that I think could further improve the paper. The comments that are more important are highlighted in red (see the pdf file attached for colored version).
	As a summary of my comments, here are my answers to the review criteria at Biogeosciences. I just selected the relevant questions:
RC41	1. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? Yes. Maybe a bit of comparison with the literature on correlation length scales could benefit the paper.
AC	We agree. We will include some comparison to addition literature on lengthscales. Some examples are given below: *Fowler, A.M., Skákala, J. and Ford, D., 2023. Validating and improving the uncertainty assumptions for the assimilation of ocean-colour-derived chlorophyll a into a marine biogeochemistry model of the Northwest European Shelf Seas. Quarterly Journal of the Royal Meteorological Society, 149(750), pp.300-324.* *Desroziers, G., Berre, L., Chapnik, B. and Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131(613), pp.3385-3396.* *Hollingsworth, A. and Lönnberg, P., 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. Tellus A, 38(2), pp.111-136.*

RC42	2. Does the abstract provide a concise and complete summary? Mostly. It could be improved by more clearly stating the results
AC	Thanks. The abstract will be amended according to the “minor and specific comments” to address this.

	Minor and specific comments
	Abstract
	I think the results should be more clearly/precisely stated in the abstract. It seemed a bit to vague to me. For example:
RC43	- l. 4: « to identify the functional types », which one are they exactly?
AC	Different expression to be used: “functional groups” when referring to a group of similarly behaving variables, to avoid confusion with PFTs used by ERSEM. We will also mention the key groups in brackets after (i.e. phytoplankton, detritus and heterotrophs & DOM).
RC44	- l. 6: « identifying the (geographically varying) connectivity lengthscales and the clusters of spatial locations that are connected. » What are the main findings concerning the length scales? What are the different clusters? For the length scales, results that seems particularly interesting is that spatial variability is quite similar between variables, requiring only to scale it using the mean length.
AC	We will update the abstract to give more detail on the lengthscale results. “We show that the spatial correlation lengthscales vary significantly between variables and are not directly transferable, however they are distinguished only by a constant scaling factor: the spatial distribution of lengthscales is similar for each variable.” We may update the abstract to include some of the specific regions (e.g defined by river input, or open-ocean to shelf-sea exchange), although we identify 13 different regions/clusters in our analysis, each only labelled with a letter A-M. Detailing all of them would seem unnecessarily specific in the case of the abstract.
RC45	- l. 9: « The results of this study help to understand how natural, or antrophogenic, perturbations propagate through the shelf-sea ecosystem », it is difficult to agree with that last sentence since the results where not clearly stated before. After finishing reading paper, I also do not think the results help to understand how perturbations propagate in the ecosystem. The results rather offer a analysis framework to do that.
AC	We agree that our method provides a framework to describe the propagation of information. We also agree that using the word “understand” in relation to how perturbations propagate was perhaps too strong and we will change this to “describe”. The use of “describe” can be justified since: • The horizontal lengthscales help to identify areas where the information is shared across space. • The regionalisation indicate that an area has some level of shared behaviour, meaning information from these regions are more likely to have a stronger influence within the region than outside of the regions. • The inter-variable analysis indicates how information of a particular variable can spread to other linked/clustered variables, as certain subsets of the state variables are shown to behave with a strong correlation. The existing statement will be rephrased as:

	"The results of this study describe how information is expected to propagate through the shelf-sea ecosystem on the time-scale of interest, and how it can be used in multiple future applications such as stochastic noise modelling, data assimilation, or machine learning."
RC46	- l. 9: « antrophogenic » -> anthropogenic
AC	*Agreed, thanks for pointing this out.*
	Introduction
RC47	l. 35: « an abstraction that will allow for smarter decision-making when considering data sampling and feature selection for ML. » Not that clear to me how and why abstraction can allow smarter decision-making.
AC	We agree that the word abstraction" is inappropriate and misleading here, we will replace it with the word "information". The key message of this sentence is that identifying connections across the NWES and variables, would indicate which variables and locations are unneeded as input features into ML algorithms.
RC48	l. 37-50: Very nice paragraph clearly stating the objective of the work.
AC	*Thanks.*
	Model and Data
RC49	Sec. 2.1: I think it will be nice to have a bit more details about the configuration. Things like: numerical schemes, diffusion, viscosity, equation of state, what forcings (wind, temperature?). How the simulations are run (spin-up procedure, initialisation...). The reference to the papers should be for further details. The reader should not need to read these papers to get a basic understanding of the configuration.
AC	We agree, and we will give more details about the configuration used.
	Methodology
RC50	Sec. 3.1: maybe a figure showing the raw and filtered time series in the supplementaries could be useful to illustrate what are the timescale filtered? Or maybe some periodogram? It should probably be stated before (introduction? Or somewhere in the methods?) what are the timescales of interest? And why? Out of curiosity have you tried your analysis with the seasonal signal?
AC	We agree and we will provide a Figure comparing the raw and filtered time-series in the Supplementary Information. The time-scales are limited by the resolution of outputs (1 day) on one end, and by the need to remove seasonality on the other end, since seasonality introduces artificial correlations. The links between variables are explored through multi-year simulations, since these we could computationally afford. We however believe that the links between variables and regions that we identified here could be applicable to longer time-scales than the time-scale of the simulation. Note, some tests were done using the seasonal signal, however as already said this signal results in a very large temporal correlation across the entire domain, obfuscating any detail within the region.
RC51	l. 154: « to a 21 km spatial resolution » make me wonder if the results are sensitive to the resolution of the model? Longer length scale because of eddy mixing? Or shorter one

	because of dynamical barrier created by filaments or eddies? This somehow questions also the isotropy assumption.
AC	As for the nature of the horizontal lengthscales identified in the manuscript: We expect that physics (eddy-mixing, filaments, eddies, which should be resolved by the 7km model) is important contribution to the length-connections (e.g. increase of Rossby radius in the open ocean), but one needs to keep in mind that for the biogeochemical variables these drivers are intertwined with biogeochemical drivers that will also be reflected in the output (e.g. river delta geography, interaction of sunlight with biology…etc). To clarify the 7km and 21km resolutions: the model is run at 7km resolution, and only the daily model outputs are upscaled to 21km, because the 21km is the highest resolution we could feasibly use in our complex network analysis for computational reasons. When it comes to the model resolution (7km), obviously using higher resolution than 7 km would improve the physics of the model (e.g. increase of the model resolution to 1.5km is an ongoing stream of work at multiple involved institutions), but for the spatial scales considered by this manuscript (I.e. the NWES-wide analysis) it is widely accepted that 7 km model resolution provides a good approximation to physics and biology. At the end, the 7 km model is used operationally at the UK Met Office and its outputs are supplied to CMEMS. To summarize, we expect that increasing spatial resolution would lead to only higher order corrections to the results presented in this work.
RC52	l. 162: I do not understand why the authors say : « As opposed to the biogeochemical lengthscales computed in Sect. 3.2.1 [...] here we manipulate the spatial networks to look at the spatial dependency of this length scale. » In section 3.2.1 you also have a map of the length scales that give you the spatial information (Fig. 2). I do not get the interest of these two definitions. Note that this also bring a bit of confusion about which are the length scales used for the different plots. For example in Fig. 4 which one is it? And for Fig. 5? I kind of got that Fig. 4 is the length scale define in sec. 3.2.1 and Fig. 5 the one in sec. 3.2.3 but it is not so clear.
AC	It is correct that 3.2.1 relates to Fig. 4, and Sec. 3.2.3 relates to Fig.5. We will add some text to Fig. 2's captions to make clearer – this is just a visualisation of what a length scale calculation looks like. The difference here, is that 3.2.1 calculates the average length-scale of each variable, while the networks used in section 3.2.3 aim to look at the way that these lengthscales vary spatially. Utilising the network structure (with pruning and totalling the number of connections) allows us to effectively normalise these lengthscales, so we can directly compare the spatial distribution of each variable to each other. We will make these need for these differences clearer in the text.
RC53	l. 167: « black » rather than « red »?
AC	Agreed. Will rephrase to: "shows a set of nodes (red) connected to the current target node (black)"
RC54	Sec. 3.3: This part is not easy to follow. Maybe a short description of the objective at the beginning could help the reader. What are the objects to be clustered, following which criteria? If I understood well, the goal is to clusters grid-points depending on their temporal correlation between each other for each variables so that grid-points with strong correlation are group together.

AC	Agreed. We will rephrase the opening sentence of the section to: "With the spatial networks, the graphs, from Sect.~3.2 at hand, we aimed to cluster geographical points (represented as nodes in each network), so that areas with similar temporal behaviour are grouped together."
	Results and Discussion
RC55	Sec. 4.1: As mentioned before, mentioning which length scale (the one from sec. 3.2.1 or sec. 3.2.3) the authors refer to would help the reader. Since two definition of length scale seems to be used, it feels natural to wonder how they compare?
AC	Agreed, we will open section 4.1 with: "Figure 4 shows the estimated correlation lengthscales for each model variable using three correlation thresholds (0.5, 0.6 and 0.7) as found from the analysis described in Sect. 3.2.1." This makes it clear which length scale is being spoken about.
RC56	l. 275-278: I think I got the general idea here: the spatial distribution of the length scale of a specific variable is the product between Fig. 5a and Fig. 4. However, as it seems that it is not the same definition of the length scale between Fig. 4 and Fig. 5a it is a bit confusing.
AC	Yes, we will make this clearer as to which definition is being used.
RC57	Sec. 4.1: I am not familiar with length scale, but it seems that there is some literature on length scales (just saying that based on a quick search on google scholar). Some comparison of the results and the methods with the literature is missing there. Are there other definition of length scale? How does the method used in this paper compare with other? Are the length scales similar to former estimations?
AC	In variational DA where we often parametrize the horizontal length-correlations, the lengthscales can be supplied as a free parameter fitting a specific function (e.g. Gaussian, Lorenzian, SOAR, Gaspari-Cohn) The length-correlation functions are identified either by ensemble runs, or diagnostic methods, such as by Desroziers et al. (2005). For example, the UK Met Office system on the NWES uses sum of two Gaussian functions and the length-correlation functions have been recently re-assessed through diagnostic methods by Fowler et al (2022). The ambition of this work is not to provide fitted functions for the length-correlation, but rather assess through single length-scale parameter how the length-correlations spatially vary across the NWES. This can then feed into future length-correlation analysis. We will compare the spatial length-scale maps from this manuscript with the only study (we are aware of) that analysed biogeochemistry length-correlations on the NWES in some detail, the Fowler et al (2022) paper, which is already cited in our manuscript. We will include discussion on this comparison in the upcoming revision of our manuscript. Please note that as stated in our response to reviewer's comment 2 (RC41), we will also add to the list of references other, more general, papers on the length-scale estimation.
RC58	Fig. 7: How is it done? I guess it is some kind of generalisation of Fig. 6 but it would be good to know more than « We used those robust boundaries to identify 13 regions representing areas of NWES connectivity. Results of this regionalisation are represented in Fig. 7. » (line 315)
AC	Yes, it is a generalisation of Fig 6. We will expand the caption for Fig. 7 to give more detail.

RC59	l. 350: « or build simpler models than ERSEM » I think this need to be say a bit differently. Complexity of models tends to increase to better (or hoping to better) represent the real world. NPZD models already exist with just one phytoplankton, one zooplankton... Here the issue is to simplify ERSEM while keeping an accurate representation. Maybe something like line 51 « simplified (yet realistic with respect to the objectives) ».
AC	Agreed, we will add the following to improve readability as suggested: “or build simplified (but realistic with respect to the objectives) models than ERSEM.”
RC60	l. 363-366: I do not see that in Fig. 8. The mean correlation between POM (yellow) and the Higher Trophic Levels + DOM (pink) is rather low. The authors should clarify.
AC	Current text, "Fig. 8 demonstrates two more clusters of variables grouped together: the group of particulate organic matter (POM)" We will remove "grouped together" to clarify the meaning. These words are not necessary, and they might imply that the clusters are linked in some way.
	Conclusions
RC61	l. 410-426: You are here a bit more specific about the results and this could be used for the abstract. E.g. « we can conclude that the biogeochemical lengthscales vary significantly between variables and are not directly transferable. » or « we have provided an approximation for the lengthscale of each variable, and each spatial location, that is informed by the high correlation in the spatial variability between lengthscales of each variable »...
AC	Agreed. We will add a sentence to the abstract that gives some more detail about the length-scale results.
RC62	l. 421-424: « Our analysis demonstrated that the chemical components (e.g., nitrogen, carbon, silicon. . . etc) of each pelagic variable (e.g., diatoms, nanophytoplankton, microzooplankton) are closely linked and a simpler version of the model can be built, by reducing these variables through parametrization. » I do not know ERSEM but I assume that as many models it started from a simple version and the complexity has been increased (e.g. addition of more phytoplankton types). I am wondering how the grouping compare with a former simpler version of ERSEM? I suppose it should be relatively similar (e.g. all types of phytoplankton in gather in only one) however it will be quite interesting if some grouping where different.
AC	Early version of ERSEM has been published in Baretta et al (1995). It is however only slightly simpler model than the current one, i.e the small and larger phytoplankton functional types are in the older model grouped together. This corresponds well with our clustering.
	Extra comments
RC63	« lengthscales »: After a quick search on google scholar, it seems that it is rather written « length scales » or « length-scales ».
AC	We agree that ‘lengthscale’ is a less common spelling of the word. To better align with other literature, we will replace instances of ‘lengthscale’ with ‘length-scale’.
RC64	The regions define in Fig. 7 could be used for sampling the domain to analyse the inter-variable interaction network. Maybe selecting grid points only within one region and to compare with the same done with another region. Are the interaction between variables different between two regions? Or sampling evenly between the regions to have a fair

	general representation? This point is mostly for curiosity as it seems natural to try to use these regions.
AC	This is interesting and would well worth considering in future work. However, it is out-of-scope for this particular work.
RC65	l. 367: Butenschon et al. (2015) and Butenschon et al. (2016) are similar paper (2015 is the discussion version of 2016). Better to keep only 2016.
AC	Agreed, we will change accordingly.

Let us thank again the reviewer for their important suggestions and we hope that after the suggested changes, addressing the reviewer comments, the manuscript will be in a good shape to be accepted for publication.

Best wishes,

Ieuan Higgs and the co-authors