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).

Responses to RC1:

RC00	This paper analyses output from a complex biogeochemical model, ERSEM, using network analysis. The analysis is used for several purposes: evaluating the spatial length scale of the variables, determining areas of coherent biogeochemical interactions and boundaries of low connectivity, and establishing which variables are highly connected with each other. This information is useful when setting up regional systems. and evaluating the interactions between model variables and weather the system can be approximated well by a simpler representation. The length scales are useful in data assimilation systems, when setting the area of influence of the observations. I think the paper provide new knowledge worth publishing, but before I would like the following points addressed:
RC01	- Only surface data is used, this is reasonable to reduce the amount of data, but it would require a discussion of the implications of such a choice. For example, in the resulting network from the analysis (Figure 9) the detritus is completely disconnected from the photo and zooplankton, but as that quickly sinks out it would not remain one on the surface and maybe using only surface data is the reason for this disconnect? There is also a question wether there are other methods to reduce the data size that would retain more information throughout the water-column that could have been used?
AC	We thank the reviewer for pointing this issue out and we will write a paragraph transparently discussing this limitation immediately after introducing the dataset we used throughout the manuscript (currently section 2, line 75). It should be noted that we expect the method of reducing the data to the surface, will only impact the variable cross-correlation matrix in Fig. 9. We believe that choosing the surface data is useful, as: (i) it is directly relevant to DA horizontal lengthscales near the surface, which is the most observed part of the water column on the NWES. (ii) it transparently captures ecosystem connections in the mixed layer, which is the most biologically active part of the ocean. (iii) in this mixed layer, we capture the lengthscales for probabilistic modelling (ensemble development). We do not know a better methodology that could retain advantages of points (i), (ii) and (iii) in a computationally affordable way.
RC02	- The longer time-scales are filtered out, so there could be biogeochemical feedback mechanisms that work on timescales >10 days that are filtered out. So what happens when resulting network is used to inform an emulator, and then applied in the context of climate as suggested by the authors? This also needs to be addressed in the discussion.
AC	While we could imagine a strong connection within sub-monthly timescales that gradually weakens (or disappears) on long (e.g. climate) timescales, in this case it is hard for us to think of a physical or biogeochemical process that would cause this, making it quite unlikely. We filter the seasonal timescales as they superficially increase the correlation thresholds, due to the seasonal harmonics common to many variables. We can think of this filter as a sort of "normalization" technique that makes the results more human understandable, rather than something that affects the analysis of connectivity (it will not affect which

	variables are more correlated and which less, but it will lower the overall correlation between all the variables). We cannot check what happens on climate timescales, but we argue that the connections captured on the short-time lengthscales (e.g. the differences between more and less connected variables and regions) of the high-pass filtered data are very likely the dominant factor in any (non-filtered) long-term connectivity analysis. While we fully agree that caution is required for any machine learning model trained on local data to be applicable on a climate scale, we would expect the simplifications and relationships learned from the complex network (this can be applied to e.g. select types of features for the climate emulator) to survive. To summarise, we believe that these connections could be used to design a climate emulator, but the emulator would need to be ideally trained on climate data. This is a long-term goal of some ongoing research (e.g. at PML), which is why we mention it in the manuscript.
RC03	- Applicability of results: Would this results of the analysis be valid other models? For example could the length scales obtained be used in data assimilation system using another BGC model than ERSEM? Would the length scales apply when assimilation observations deeper in the water column even if your results that are only based on surface model data?
AC	The methodology highlighted by this manuscript is naturally applicable to different models and datasets. We would expect the key characteristics derived here from ERSEM to be representative of ecosystem itself, and therefore of other trustworthy representations of the ecosystem (i.e. trustworthy models). Wherever the derived characteristics could be compared to the general knowledge, they compared very well. We note that there are severe constrains on validating our results with observations, due to the intermittency of satellite observations - and doing the same analysis with other models is beyond the scope of this project. This has been discussed in the manuscript, please see the line 70. We would expect the lengthscales to be broadly applicable in the mixed layer of the ocean, and to be not applicable beneath the mixed layer. The majority of biological growth happens in the mixed layer, and so it is of higher significance and relevance to investigate (we might think differently if we were instead investigating global or deep ocean systems). In addition to this, on the NWES there aren't that many sub-mixed-layer observations, at least compared to the number of surface observations, making our knowledge of horizontal lengthscales at these depths more challenging, and not our particular focus.
RC04	- The description of the methods could be improved for the benefit of the reader, I provide some suggestions for what needs to be clarified below.
AC	The suggestions are extremely helpful and go to strengthen the manuscript. Details of each improvement are given with the corresponding suggestion.

	Specific comments
RC05	Title: Could the title be improved but adding "Investigating" at the beginning?
AC	Yes agreed. We will change the title accordingly.
RC06	Abstract: The expression "functional types of variables" is used in the abstract and in the text, it is a bit unclear to me what this means. The expression becomes particularly confusing since the ERSEM itself also includes functional types of plankton. Consider either using a different expression or define it properly before using it.
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.
RC07	"Be also used" should be "also be used"
AC	Thanks for noticing this, we will change as suggested.
RC08	What is meant by "flow of information between degree of freedom"
AC	Will rewrite to say: "within which there is a large exchange of information within the ecosystem"
RC09	The first part of the last sentence is unclear to me: I don't see that it is demonstrated anywhere how these results can be used to understand how a perturbation propagate through the ecosystem.
AC	Will rephrase : "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." Some examples include: The horizontal lengthscales help to identify areas where the information is simultaneously 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.
RC10	Line 38: "…investigate three relevant questions related …" either formulate the three topic as questions or rewrite the sentence on line 38.
AC	We will rewrite the sentence on line 38 to read "…investigate three relevant *topics* related …"
RC11	Line 40: "based on" should be "apply".
AC	Thanks, and agreed. We will change accordingly.
RC12	Line 40: Is this length scale only useful when applying variational data assimilation, not other (ensemble) data assimilation techniques?
	Thanks, this is correct. The results are also highly relevant to ensemble data assimilation (EnDA) too. For instance, it can guide scale-aware localisation for ensemble-based error covariance. We will add text on line 40 to highlight this applicability.

RC13	Line 49: as mentioned before, the use of the expression the use of the expression ”functional type” is a bit confusing, please define it here.
AC	We will rephrase this expression in line with our earlier comment on “functional types” ->” functional groups”. Here we will also define a functional group (for additional clarity): “functional groups (i.e. a set of state variables that are generally highly correlated with each other).”
RC14	Line 51: The statement that these traditional biogeochemical models are unsuitable to address response to climate change, effectively writing off all CMIP simulations is quite severe, I would suggest to moderate the statement. However I do agree that lighter model systems are more suitable for ensemble simulations, but it they are trained on data from the present day, they may not be very good at representing future ecosystem response.
AC	As well as line 51, we believe the reviewer may also be referring to line 21, which states: "such as ecosystem’s response to climate change and anthropogenic pressures across large variety of scenarios". We agree and recognise that these large models can display a high degree of realism, but the computational burden of these simulations means there is also a need for reduced order models that can cover a “large variety of scenarios” (where these ESMs may only explore some subset of this range). The current text may have led to this misunderstanding, so we will tidy up the language to avoid this.
RC15	Line 88: Were the river nutrients also included and were they also annual?
AC	Yes, the rivers included nutrients, varying daily. We will provide this information in the manuscript.
RC16	Line 120: the transformation to the time-local standardised form is very well explained, but I wonder what happens in period when standar deviation is low or zero (for example I winter), does and stay finite?
AC	Thanks. The aim is to create a data set with a uniform standard deviation (i.e. unity). When sigma is small (denominator in (3)), the numerator is small also. The values remain finite.
RC17	Line 120: Would river input influence the network results, for example would there be a stronger connection between the biogeochemistry and salinity in a region of strong river influence. I.e. would the network presented in figure 9 differ from region from region to region?
AC	The river inputs influence the network structure, which is clear from Fig. 7 “Region G” – a region largely defined by the delta of River Elbe (see also the difference in nutrients in supplementary material). In the context of Fig. 8, we would expect certain connections to be amplified/dampened if we were to reduce the sample size to a specific region. The figures provided aim to show an overview for the system behaviour, acknowledging these effects with the coefficients of variation shown in Fig.8. While we acknowledge that there is a great variety of additional questions one can ask and explore, in this work, we don't have scope to address all of these questions and we have to be selective (e.g. we don't specifically address the difference between river delta areas and areas further from the coast).

RC18	Line 124: I did not see it specified anywhere that data were treated any differently, so could you just simply write that all dat were treated this way?
AC	Agreed. Will rewrite to simply state: "All data used in this study have been preprocessed using the procedure in Eqs…"
RC19	Sections 3.2.1: Biogeochemical length scale estimation: What did you do in regions close to land or the boundary? Did you not compute the length scale or only consider the ocean points? The same question applies to the method in 3.2.2
AC	In Sec. 3.2.1, we only considered the ocean points away from the boundary for this initial calculation of the average lengthscale of each variable. We will make this clear in the manuscript. In section 3.2.2, we upscale from a 7km -> 21km grid. We use the arithmetic mean of the relevant points to achieve this upscaling. To account for the boundaries, we consider a 21km grid point to be ocean only if more than half of the relevant 7km points are also ocean. In addition, this provided good results, and the lengthscales at the boundaries are discussed in section 4.1: "It is notable that another area of low-connectivity is the open (Atlantic) AMM7 domain boundary regions. This indicates that the boundary conditions of the regional model de-correlate from the rest of the domain…"
RC20	Difference between method in 3.2.1 and 3.2.2: Am I correct that the difference between 3.2.1 and 3.2.2 is that 3.2.1 is done on a finer grid and uses a different method to compute the length scale? The coarsening before computing the length scale is primarily used to reduce the amount of data given to the SGC? Is this correct or are there other resort to compute length-scales twice? This could be made clear in the manuscript.
AC	Thank you. Your conclusion is mostly correct, and we shall make this clearer in the manuscript. Some slight clarifications: Section 3.2.1 is used to determine the average lengthscale of each variable. Section 3.2.2 is used to determine the average lengthscale of each geographical point (independent of the specific variable). Since Sec. 3.2.1 shows that the average lengthscales are different, we account for this by using the different correlation coefficient thresholds on each variables network, such that each network has the same number of links. This effectively normalises the lengthscales for each variable, and allows us to calculate an average for each point. While we allude to this in the opening sentence of section 3.2.3: "As opposed to the biogeochemical lengthscales computed in Sect. 3.2.1, which refer to each variable and reflect their physical properties averaged on the domain, here we manipulate the spatial networks to look at the spatial dependency of this length-scale." We will make a statement earlier in the manuscript to make this clearer.
RC21	Line 153: How was the grid upscaled from 7 to 21 km?
AC	Will add text to specify that "arithmetic averaging" was used.
RC22	Line 154-160 Explanation of pruning: This is very hard to understand, please explain better how this was done.
AC	We will revise text used to explain the pruning will be revised to make it clearer. We will likely re-phrase the relevant text, and move it to a more appropriate section of the manuscript (Sec. 3.2.3) instead of in the section for initially generating the spatial networks (3.2.2).

RC23	LIne 170 from “We took …” and the next sentence mean exactly the same thing. Remove the first sentence (or last, up the author, but I preferred the last).
AC	We will remove the first sentence (and keep the last).
RC24	Before line 180: This is not easy to understand, could you please try to make this clearer: “This was done by taking the mean lengthscale at each grid point across all variables from the dynamically thresholded spatial networks. In order to assess whether this spatial variation could be well approximated by the mean of these lengthscales, we compared the spatial distribution of lengthscales between each different variable using Pearson’s correlation. Here, we would expect to see a high correlation if the structure of the spatially varying lengthscales is consistent. This set of spatially varying lengthscales was then represented as a ratio of the mean.”
AC	Agreed that this is not clear. We will rewrite this paragraph to make it much clearer.
RC25	Line 193: “a links … defined by the Spearman correlation.. ” at this point there has been introduces severe spearman correlation, the length scale of the correlation with itself on a 7 km grid, the length-scale om a 21 km grid and the correlations between the length scales of different variables, so which one does this refer to here?
AC	We will adjust the text to make specific reference to the relevant case: “defined in this case by the Spearman's correlation between each node on the 21km grid)”
RC26	Paragraph line 190-200: Please write out the equations on its own line (as on page 7) and give them numbers to benefit the reader.
AC	Agreed, we will give the equations their own lines to make it clearer for readers.
RC27	Line 220: This is difficult to follow: “In order to compare the regionalisation of each variable, we first projected the cluster labels of each node back onto the horizontal plane. Then, we applied an edge detection kernel to identify the boundaries between differently labelled regions, creating a boundary map for each variable (with value 1 at boundary grid points and 0 elsewhere).” Please refer back to the appropriate equation on the previous page (ref. my comment above).
AC	We will rewrite this to make it clearer, and refer to the relevant clustering equation: We identified ``robust regions'' as connected areas of ocean that rarely, or never, contain the boundaries from the clustering of any individual state variable. For the spatial network of each variable, we identified every node that is geographically adjacent to another node with a different cluster label (as found from Eq. (X)). These nodes represent the boundaries between different regions. Since each node in a spatial network will have a corresponding node in the spatial network of every other variable (i.e., they share the same geographical point), we could then calculate the frequency with which each geographical point occupies a boundary node, across all ERSEM state variables. These ``boundary frequency'' values are then plotted onto a grid, according to their geographic location, so that the robust regions can be identified visually.

RC28	Line 235: You calculate the mean adjacency matrix over 300 point randomly selected over the shelf <200 meter and then average that. Then later you say "the boundaries particularly seem to reflect shallower bathymetry (approx. 100 m) than the 200 m depth usually applied to delimit the margins of shelf-seas, including NWES." So why not samle within 100 meters?
AC	We chose to use the 200m bathymetry as it is the standard convention to define shelf-seas, including the NWES (Skakala et al (2022), Huthnance et al (2009), Borges et al (2006). This delimitation corresponds to the continental shelf. The 100m area highlights that a big part of this region is linked to the open North Atlantic, which is of interest. *Skákala, J., Bruggeman, J., Ford, D., Wakelin, S., Akpınar, A., Hull, T., Kaiser, J., Loveday, B.R., O'Dea, E., Williams, C.A. and Ciavatta, S., 2022. The impact of ocean biogeochemistry on physics and its consequences for modelling shelf seas. Ocean Modelling, 172, p.101976.* *Huthnance, J.M., Holt, J.T. and Wakelin, S.L., 2009. Deep ocean exchange with west-European shelf seas. Ocean Science, 5(4), pp.621-634.* *Borges, A.V., Schiettecatte, L.S., Abril, G., Delille, B. and Gazeau, F., 2006. Carbon dioxide in European coastal waters. Estuarine, Coastal and Shelf Science, 70(3), pp.375-387.*
RC29	Line 255: Be precise: inclusion of new types of observations *for assimilation* …
AC	Agreed. This change will be implemented.
RC30	Line 255: I suggest to remove "profound".
AC	Agreed. This change will be implemented.
RC31	Line 265: suggest: "oxygen have different lengthscales …"
AC	Agreed. This change will be implemented.
RC32	Line 379: "… we applied SGC…": did you also test different values of k here?
AC	Yes, we mention this on lines 217.
RC33	Line 390: "Ammonium dynamics are relatively more complex than the ones of nitrate." This sentence can be removed.
AC	Agreed. This will be done.
RC34	Figure 9: How was the lines connecting the different variables decided?
AC	Added text to Figure 9 caption: "The highest correlations (top 25\%) of all possible pairwise correlations between variables are shown (grey lines)."
RC35	Line 427: I suggest to use another word than "dismantling".
AC	We will change to "simplifying" instead.
RC36	Concerning the supporting information, this would be easier to understand if the variables plotted were given standard names and the y-axis were supplied with the units.
AC	We will change the plots in the supporting information, re-format them with standard names, and correct units on the y-axis.

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