

Appendix A AEM

Firstly, a connectivity matrix between the sampling sites is constructed. An element in this matrix represents the probability of reaching site B from site A. This matrix can be highly asymmetric (Figure (A.1)). This method is only a very rough first approximation. Instead of basing the connectivity matrix on global currents (Figure (A.2)), more precise biophysical models should be used for the dispersal probability (Xuereb et al., 2018).

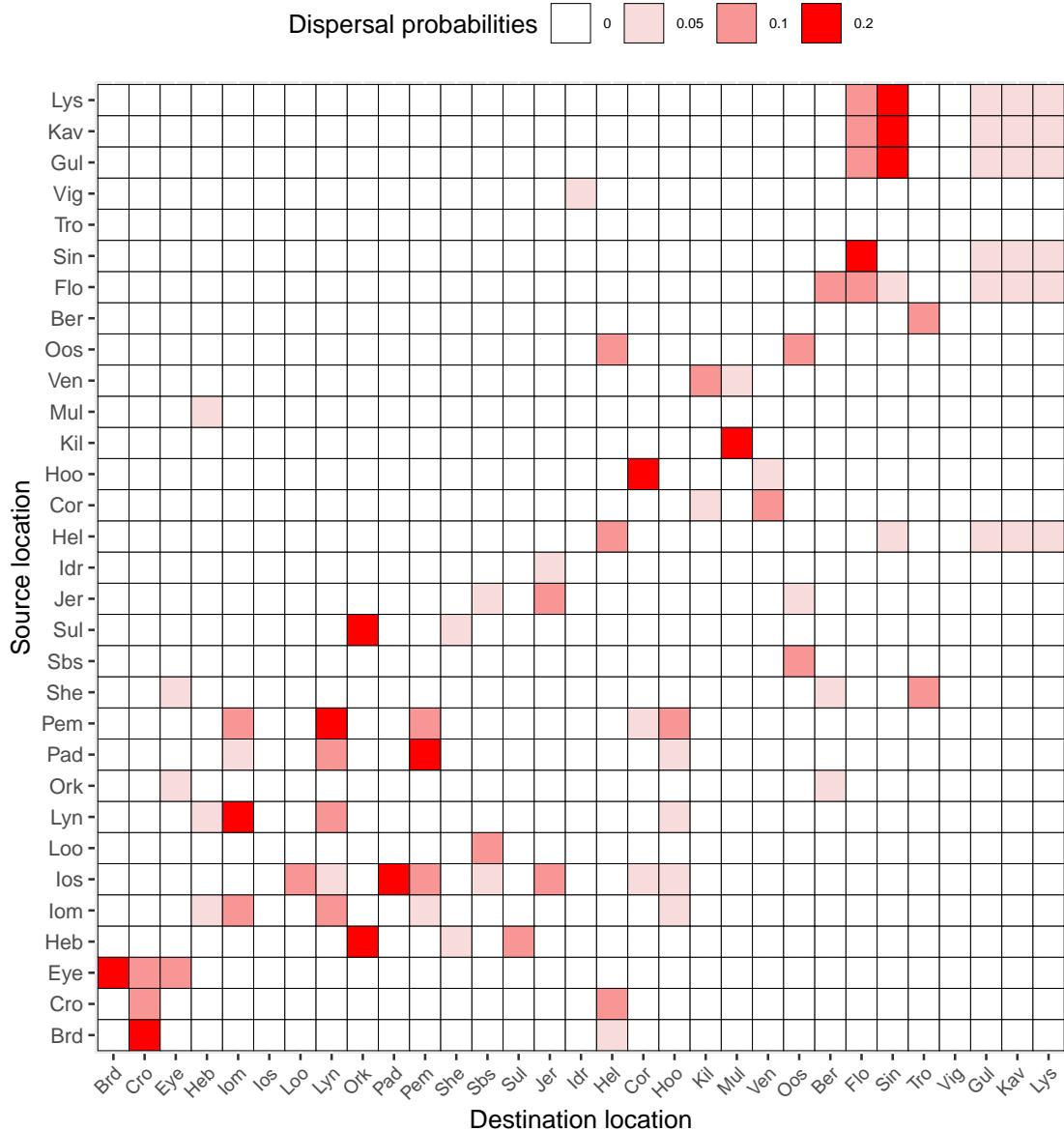


Figure A.1: Connectivity matrix based on dominant ocean currents

Secondly, a site-by-edge matrix is constructed from this connectivity matrix as described in Xuereb et al. (2018). Figure (A.3) provides a geographic representation of the site-by-edge matrix. Note the added node in the far south of the study area, which ensures directionality.

Finally, the asymmetric eigenvectors are calculated from the site-by-edge matrix. This matrix is first weighted by the dispersal probabilities from the connectivity matrix.

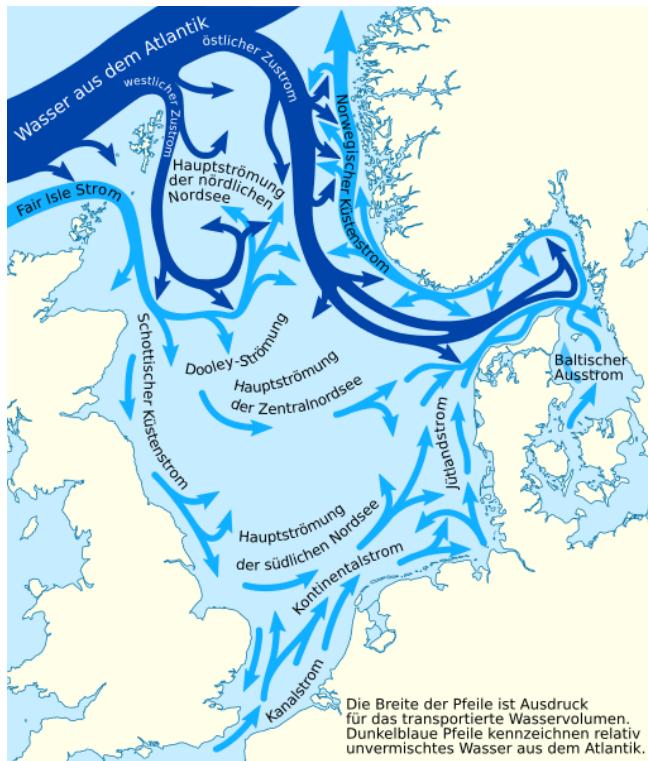


Figure A.2: Map of dominant ocean currents around the North Sea (Wikipedia)

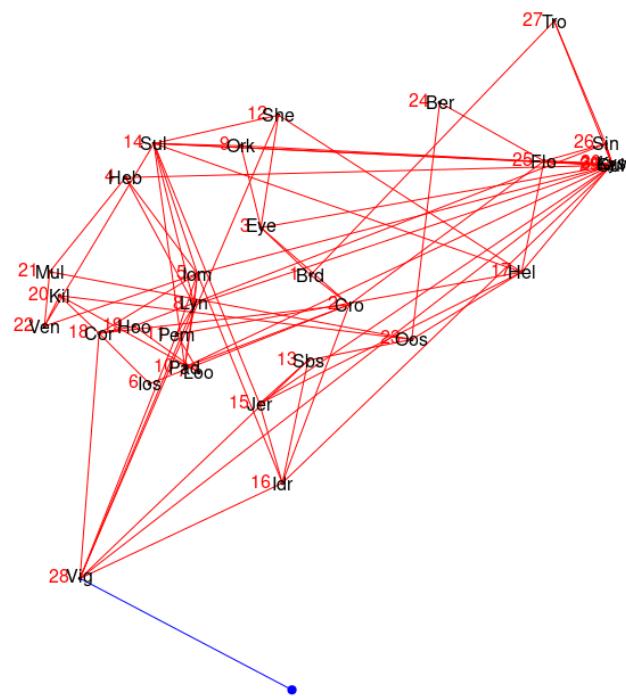


Figure A.3: Geographical representation of the binary site-by-edge matrix

Appendix B PCoA versus PCA

Figure (B.1) shows the comparison between the PCoA and PCA for the outlier- and neutral SNPs.

(A) and (B) show the analysis with the 8 outlier SNPs for all the sites and (C) and (D) for the Atlantic sites. All eight locus descriptors are shown.

(E) and (F) show the analysis with the 71 putatively neutral SNPs for all the sites and (G) and (H) for the Atlantic sites. In (F), only the descriptors that lie with at least 80 percent of their length along the first two axes are retained. In (H), this minimum is 70 percent.

All PCA biplots are shown in scaling 1.

In general, very similar patterns are visible in both ordinations. However, the PCA with allele frequencies consistently captures more variation in its first two axes than the PCoA with the F_{st} distances.

The first two axes also capture more variation when the 8 outlier SNPs are used. As the neutral SNPs represent more neutral processes, a higher randomness can be expected in the neutral variation.

The Oosterschelde again provides an interesting case. While its placement among the Scandinavian sites for the 8 outlier SNPs (A-D) is unexpected, the Dutch site falls more or less in the vicinity of the other sites. For the 71 neutral SNPs, however, the Oosterschelde is clearly different from the other sites. Consequently, the differentiation of the Oosterschelde population could possibly be attributed to neutral processes.

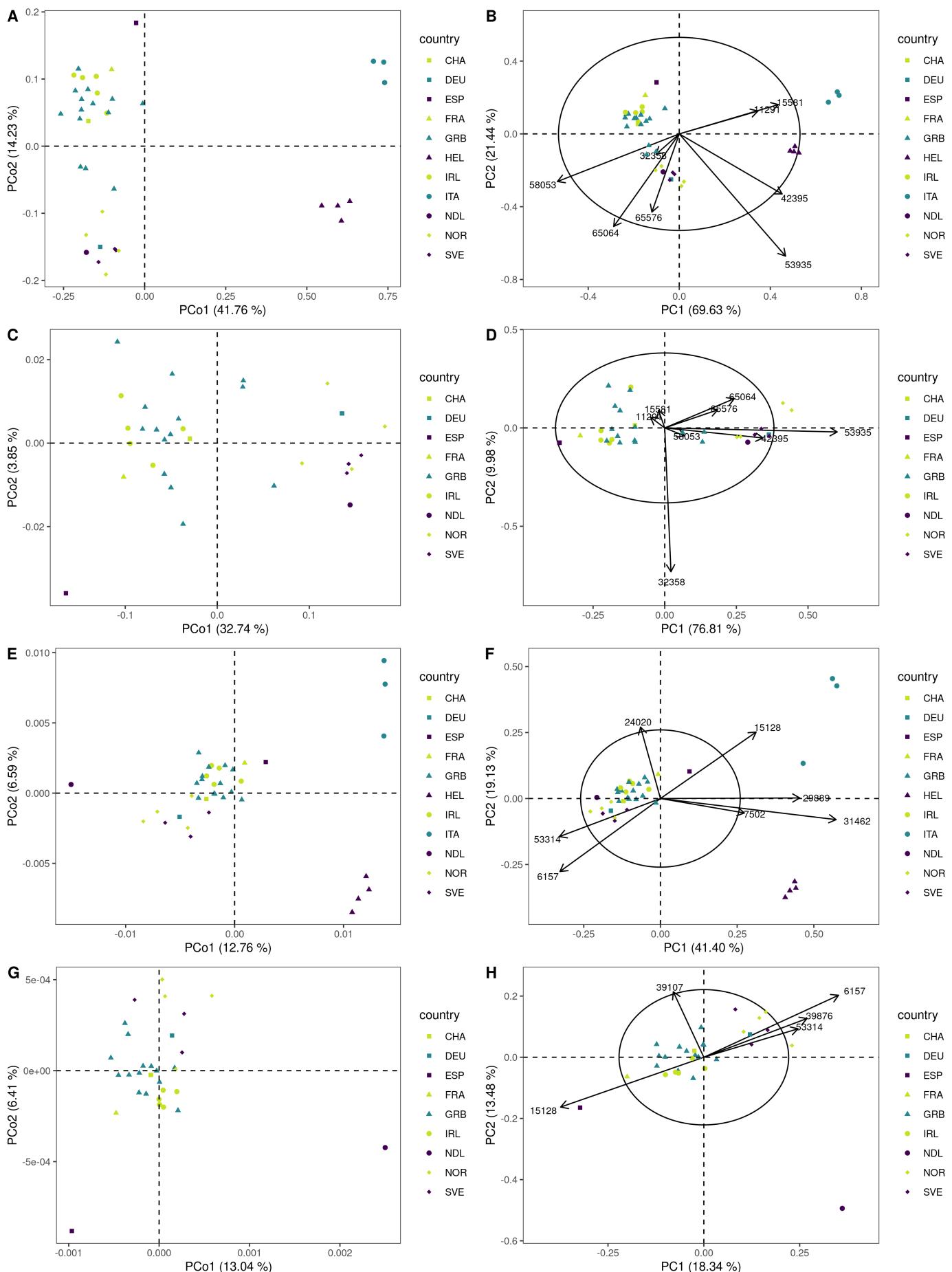


Figure B.1: PCoA versus PCA; (A) and (B): 8 SNPs - all sites; (C) and (D): 8 SNPs - Atlantic sites; (E) and (F): 71 SNPs - all sites; (G) and (H): 71 SNPs - Atlantic sites

Appendix C RDA for allele frequencies

The allele frequencies are used in the RDA. The matrix of allele frequencies now constitutes the matrix of response variables.

The list of variables that are retained after forward selection can be found in Table (C.2).

For the 71 neutral variables, the environmental variables are found to be non-significant after removing the PP variables ($[\text{env}] \mid [\text{lin}] + [\text{MEM}] + [\text{AEM}]$; $R_{\text{adj}}^2 < 0$; $p_F = 0.6019$). The full model with AEM variables is not significant ($p_F = 0.079$). For consistency, the AEM variables are retained in the further analysis.

For the 8 outlier SNPs, the selected variables are the same variables that were selected for the analysis with the F_{st} distances with the sole exception of the mean salinity. However, for the 71 neutral SNPs, the selected variables are quite different. Fewer AEMs are selected as well as more MEMs. Nonetheless, with the exception of MEM7, no new variables are selected as compared to previous analyses.

Figure (C.1) shows the variation partitioning of the linear regression and Table (C.1) contains the testable fractions.

For the 8 outlier SNPs, the results are very similar to the results with the F_{st} distances.

However, the dbMEMs explain a borderline significant fraction of the variation after the environmental variables and the linear coordinates are taken into account.

For the 71 neutral SNPs, however, the results deviate. The AEMs lose their dominant role and the dbMEMs are now the most informative group. Interestingly, no group of variables explains a significant fraction of the variation when the other two groups are taken into account.

Table C.1: Permutation test of the variation partitioning using the allele frequencies

8 outlier SNPs				71 neutral SNPs			
fraction	R_{adj}^2	dof	p_F	fraction	R_{adj}^2	dof	p_F
[\text{env}]	0.648	2(30)	< 0.0001	[\text{lin}]	0.107	2(30)	< 0.0001
[\text{lin}]	0.645	2(30)	< 0.0001	[\text{MEM}]	0.128	4(30)	< 0.0001
[\text{MEM}]	0.629	3(30)	< 0.0001	[\text{AEM}]	0.098	3(30)	< 0.0001
[\text{env}] + [\text{lin}]	0.688	4(30)	< 0.0001	[\text{lin}] + [\text{MEM}]	0.156	6(30)	< 0.0001
[\text{env}] + [\text{MEM}]	0.667	5(30)	< 0.0001	[\text{lin}] + [\text{AEM}]	0.137	5(30)	< 0.0001
[\text{lin}] + [\text{MEM}]	0.692	5(30)	< 0.0001	[\text{MEM}] + [\text{AEM}]	0.148	7(30)	< 0.0001
[\text{env}] + [\text{lin}] + [\text{MEM}]	0.715	7(30)	< 0.0001	[\text{lin}] + [\text{MEM}] + [\text{AEM}]	0.171	9(30)	< 0.0001
[\text{env}] \mid [\text{lin}] + [\text{MEM}]	0.023	2(25)	0.052	[\text{lin}] \mid [\text{MEM}] + [\text{AEM}]	0.023	2(23)	0.084
[\text{lin}] \mid [\text{env}] + [\text{MEM}]	0.047	2(25)	0.0049	[\text{MEM}] \mid [\text{lin}] + [\text{AEM}]	0.034	4(25)	0.055
[\text{MEM}] \mid [\text{env}] + [\text{lin}]	0.027	3(26)	0.047	[\text{AEM}] \mid [\text{lin}] + [\text{MEM}]	0.015	3(24)	0.20
[\text{env}] \mid [\text{lin}]	0.043	2(28)	0.0039	[\text{lin}] \mid [\text{MEM}]	0.028	2(26)	0.024
[\text{env}] \mid [\text{MEM}]	0.039	2(27)	0.019	[\text{lin}] \mid [\text{AEM}]	0.039	2(27)	0.0041
[\text{lin}] \mid [\text{env}]	0.041	2(28)	0.0067	[\text{MEM}] \mid [\text{lin}]	0.049	4(28)	0.011
[\text{lin}] \mid [\text{MEM}]	0.064	2(27)	0.0004	[\text{MEM}] \mid [\text{AEM}]	0.050	4(27)	0.0063
[\text{MEM}] \mid [\text{env}]	0.020	3(28)	0.11	[\text{AEM}] \mid [\text{lin}]	0.030	3(28)	0.047
[\text{MEM}] \mid [\text{lin}]	0.047	3(28)	0.0083	[\text{AEM}] \mid [\text{MEM}]	0.020	3(26)	0.12

Figure (C.2) shows the results of the RDA graphically. The patterns are generally very similar to the patterns in the RDA with the F_{st} distances. Information on the individual loci could be used to associate certain loci with certain explanatory variables, but no clear patterns are visible. All loci are shown in (A) and (B), while (C) and (D) only show loci that lie with more than 50 percent of their length along the first two axes.

Table (C.2) shows the results of the hierarchical variation partitioning. For the 8 outlier SNPs, the results are very similar to the results with the F_{st} distances. However, for the 71 neutral SNPs, the MEM variables take over the role of the AEM variables in the analysis with the F_{st} distances. None of the AEM variables explains a significant fraction of the variation.

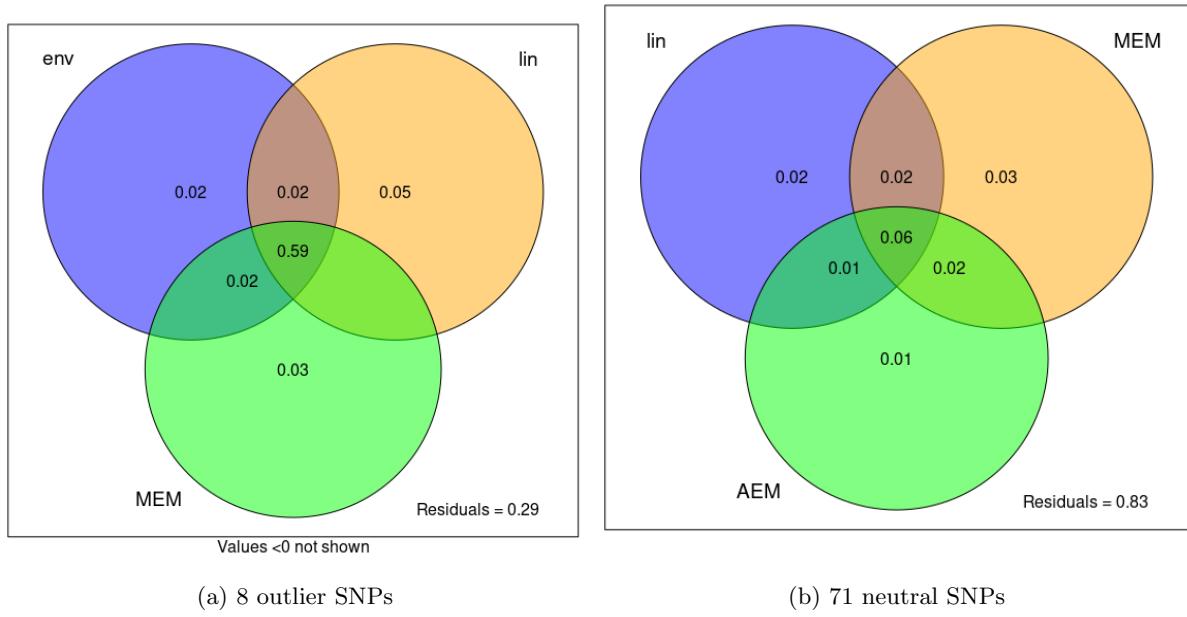


Figure C.1: Venn diagram of the variation partitioning of the linear regression using allele frequencies

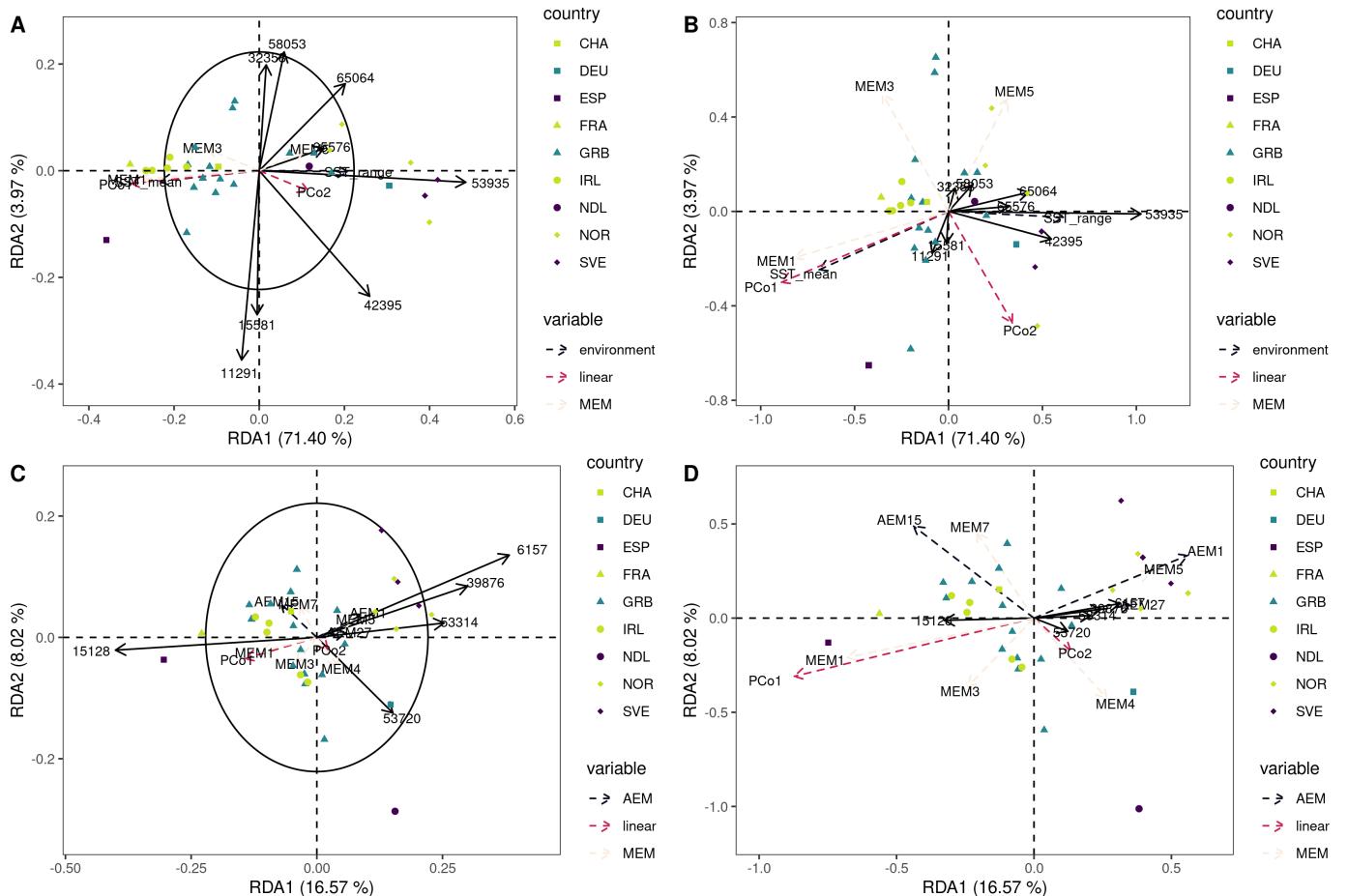


Figure C.2: RDA with the allele frequencies;
 (A) 8 outlier SNPs - scaling 1; (B) 8 outlier SNPs - scaling 2;
 (C) 71 neutral SNPs - scaling 1; (D) 71 neutral SNPs - scaling 2

Table C.2: Remaining explanatory variables after forward selection for allele frequencies;
 results of hierarchical variation partitioning;
 9999 permutations for the test with 8 outlier SNPs;
 9999 permutations for the test of the variable groups with the 71 neutral SNPs;
 999 permutations for the test of the individual variables with the 71 neutral SNPs

8 outlier SNPs			71 neutral SNPs		
variable	R^2_{adj}	p_R	variable	R^2_{adj}	p_R
<i>environment</i>	0.2371	0.0003	<i>linear</i>	0.0524	0.0007
SST mean	0.1397	0.0109	PCo1	0.0478	0.003
SST range	0.0964	0.0343	PCo2	0.0085	0.182
<i>linear</i>	0.2481	< 0.0001	<i>dbMEMs</i>	0.0879	0.0008
PCo1	0.1975	0.0034	MEM1	0.0282	0.019
PCo2	0.0658	0.0726	MEM3	0.0148	0.098
<i>dbMEMs</i>	0.2296	0.0042	MEM4	0.0219	0.039
MEM1	0.1571	0.0070	MEM5	0.0042	0.260
MEM3	0.0341	0.1357	MEM7	0.0159	0.094
MEM5	0.0243	0.1825	<i>AEMs</i>	0.0483	0.0121
			AEM1	0.0173	0.074
			AEM15	0.0197	0.067
			AEM27	0.0102	0.146

In conclusion, the linear regression and variation partitioning with the allele frequencies displays similar trends as the analysis with F_{st} distances. However, in particular for the neutral SNPs, clear differences exist. The analysis with F_{st} distances hints at asymmetric processes while the one with allele frequencies points more to isotropic patterns. This discrepancy deserves further attention.

Appendix D Mantel test

The package `ade4` (Dray & Dufour, 2007) is used to perform a Mantel test. The Mantel test is a simple tool for detecting *isolation by distance* (Mantel, 1967; Diniz-Filho et al., 2013).

Figure (D.1) shows the results of the Mantel test. The test for all the sites is highly significant. However, it is clear from Figure (D.1a) that two distinct meta-populations exist in the dataset: the Atlantic meta-population and the Mediterranean meta-population.

Figure (D.1b) shows the Mantel test for the Atlantic sites. Surprisingly, the correlation between F_{st} values and in-water distances is not significant. However, the culprit for this negative test can be easily identified: the Oosterschelde. Due to the high genetic differentiation of the Dutch site, every pair formed with the Oosterschelde shows a high F_{st} value for relatively low in-water distances. When the Oosterschelde is removed from the dataset, the correlation between the F_{st} values and the in-water distances becomes highly significant for the Atlantic sites as well (Figure (D.1c)).

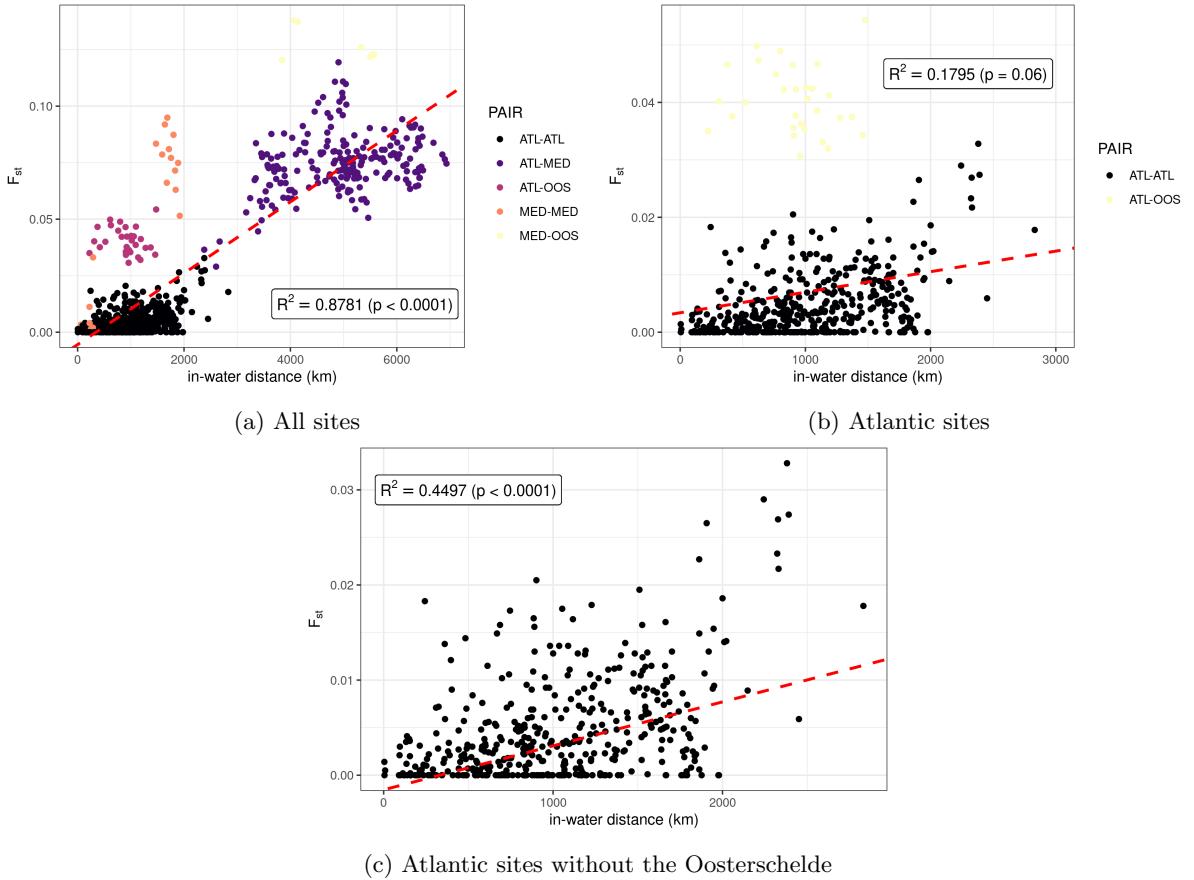


Figure D.1: Mantel test scatter plots

Appendix E Distribution of the environmental variables

Figure (E.1) shows the distributions of the normalised environmental variables. While the SST variables are more or less normally distributed, the other variables clearly are not. The salinity- and productivity variables show stretched tails; the bathymetry suffers from an outlier; and the current velocity appears to be more uniformly distributed.

Certain transformations could be used to remedy the non-normality of these variables. However, as most of the statistical tests in this exercise are quite robust against deviations from normality, the original variables are retained.

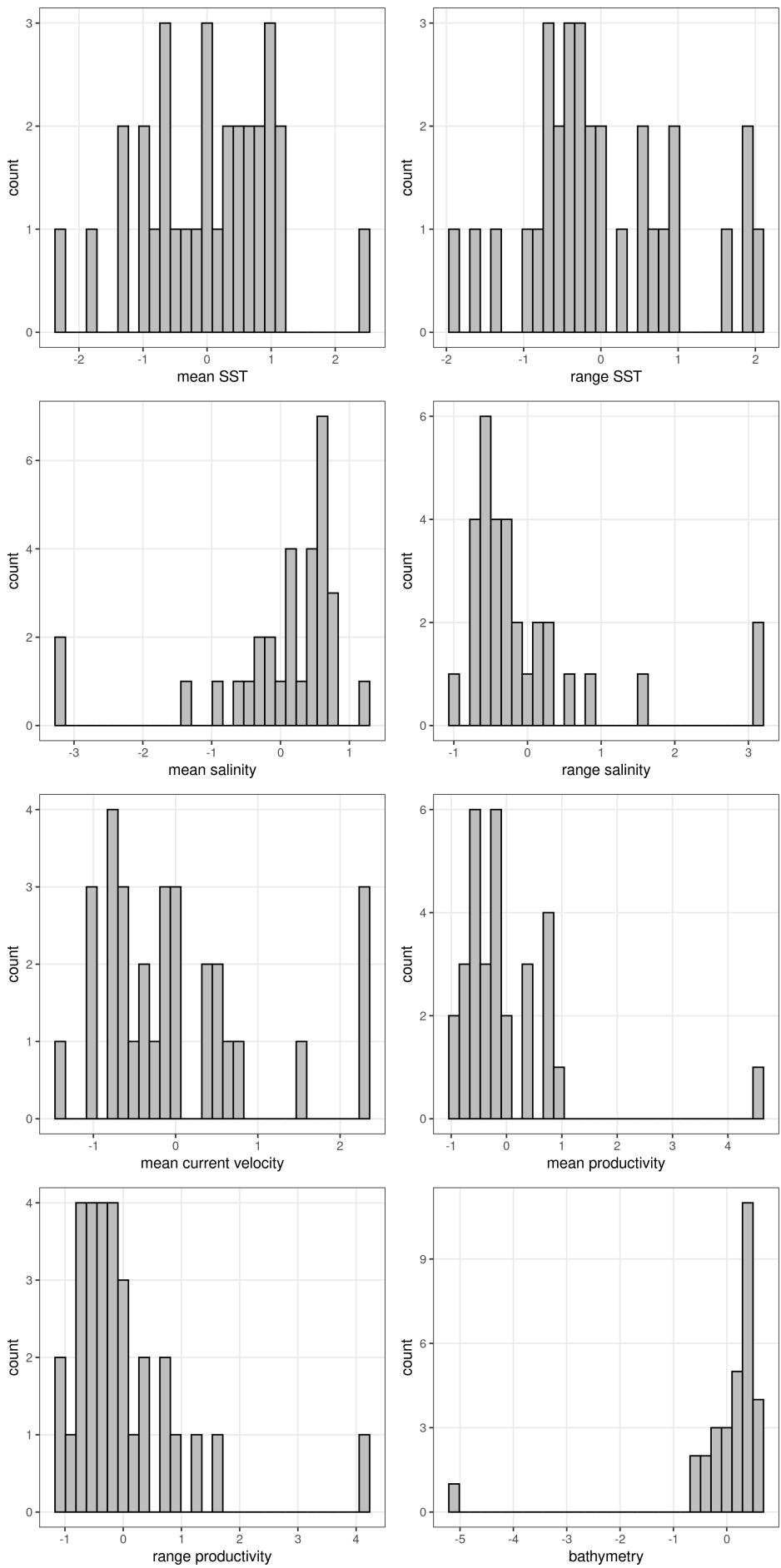


Figure E.1: Distributions of normalised environmental variables

Appendix F Maps

The packages `rnaturrearth` (South, 2017a), `rnaturrearthdata` (South, 2017b), `sf` (Pebesma, 2018), `ggspatial` (Dunnington, 2022) and `cowplot` (Wilke, 2020) are used to generate the maps in Figures (F.1-F.6).

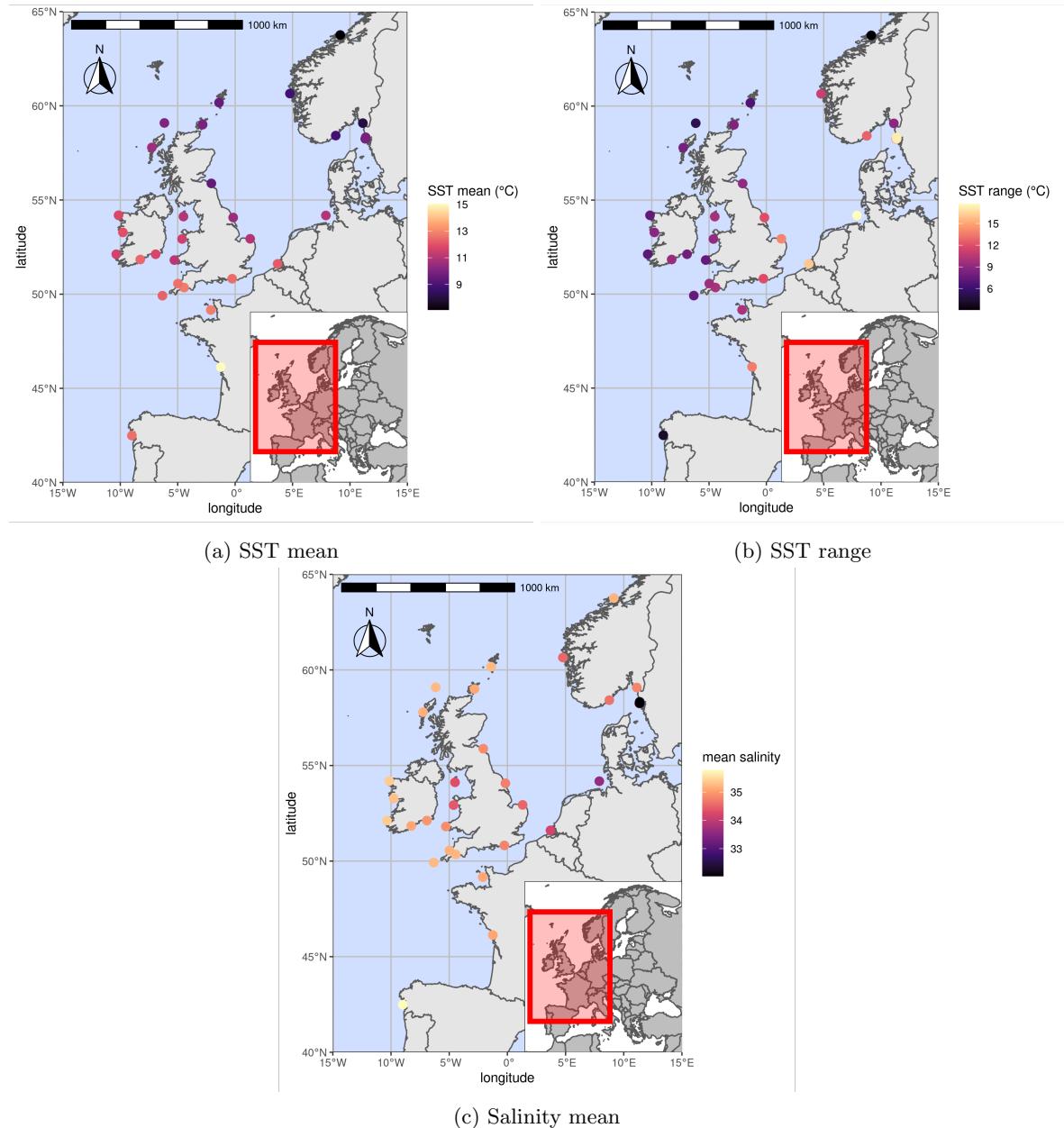


Figure F.1: Geographic maps with the environmental variables

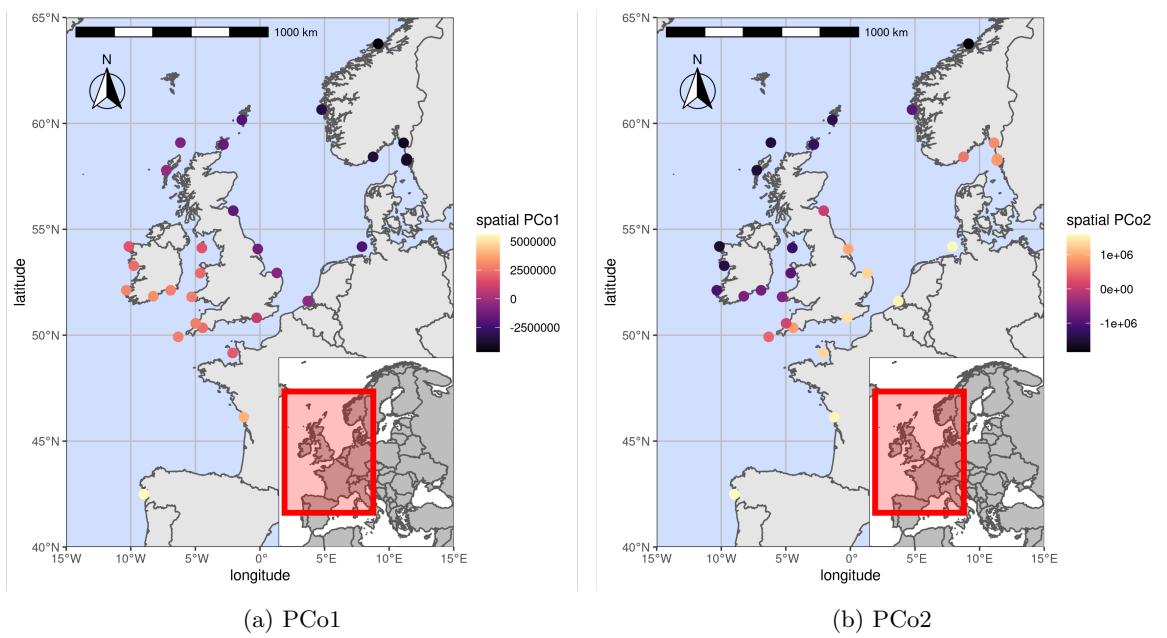


Figure F.2: Geographic maps with the linear coordinates

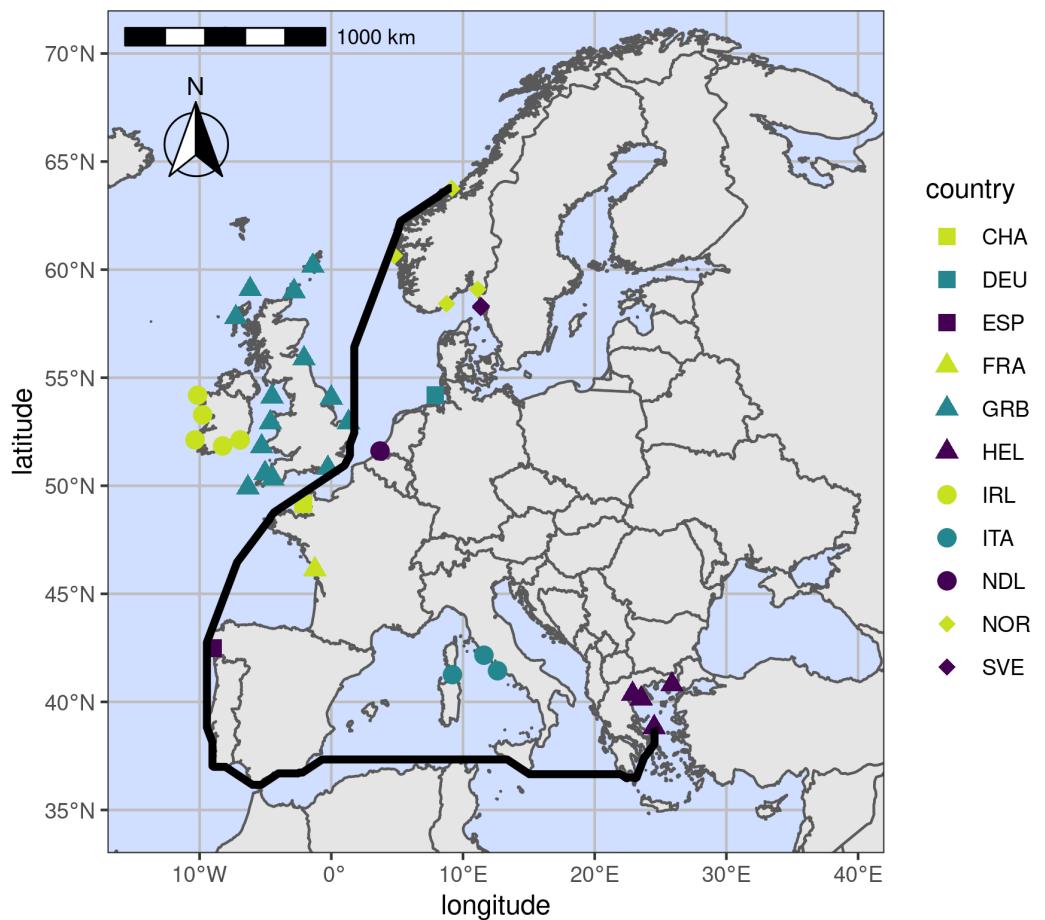


Figure F.3: In-water distance between Trondheim and Skyros

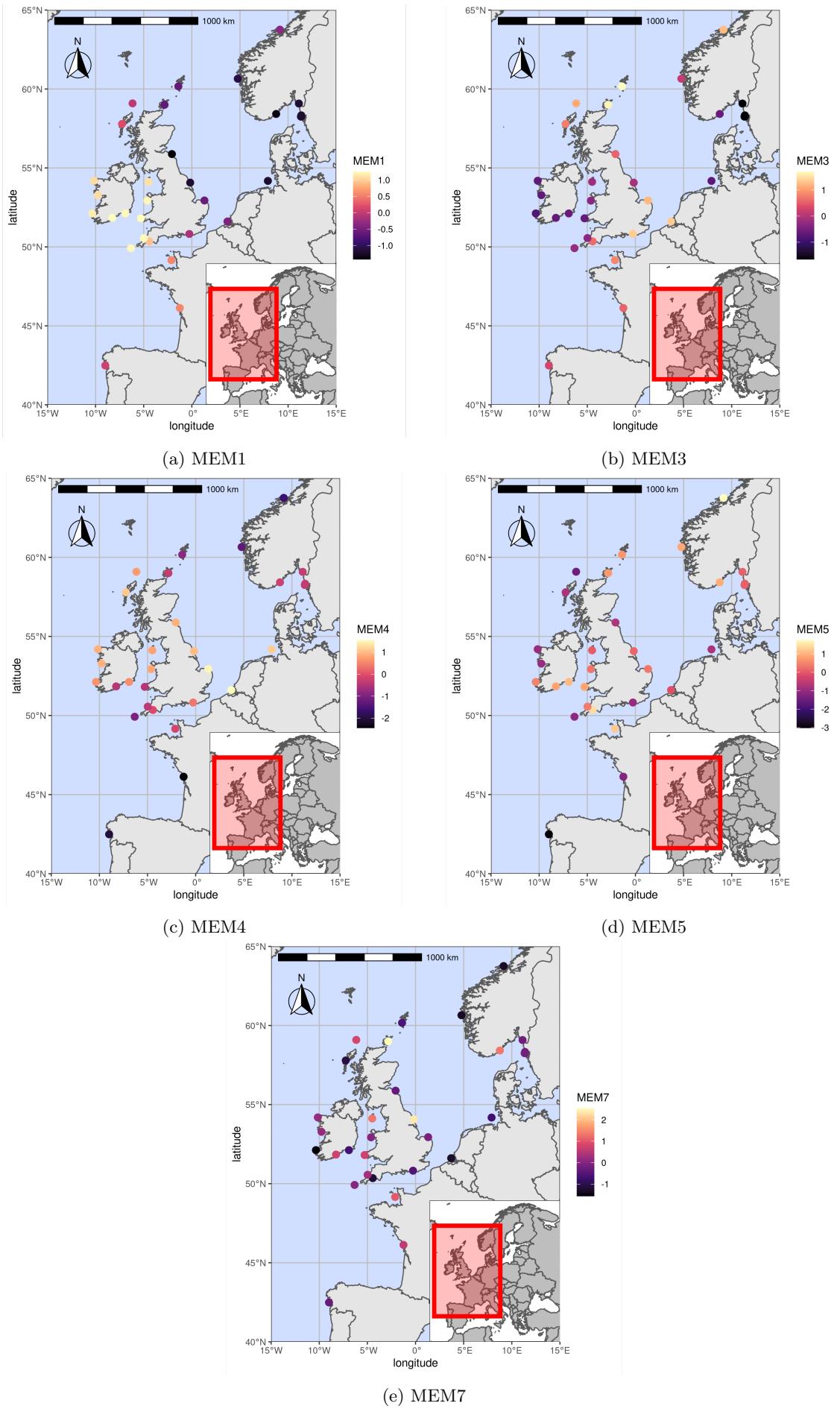


Figure F.4: Geographic maps with the dbMEMs

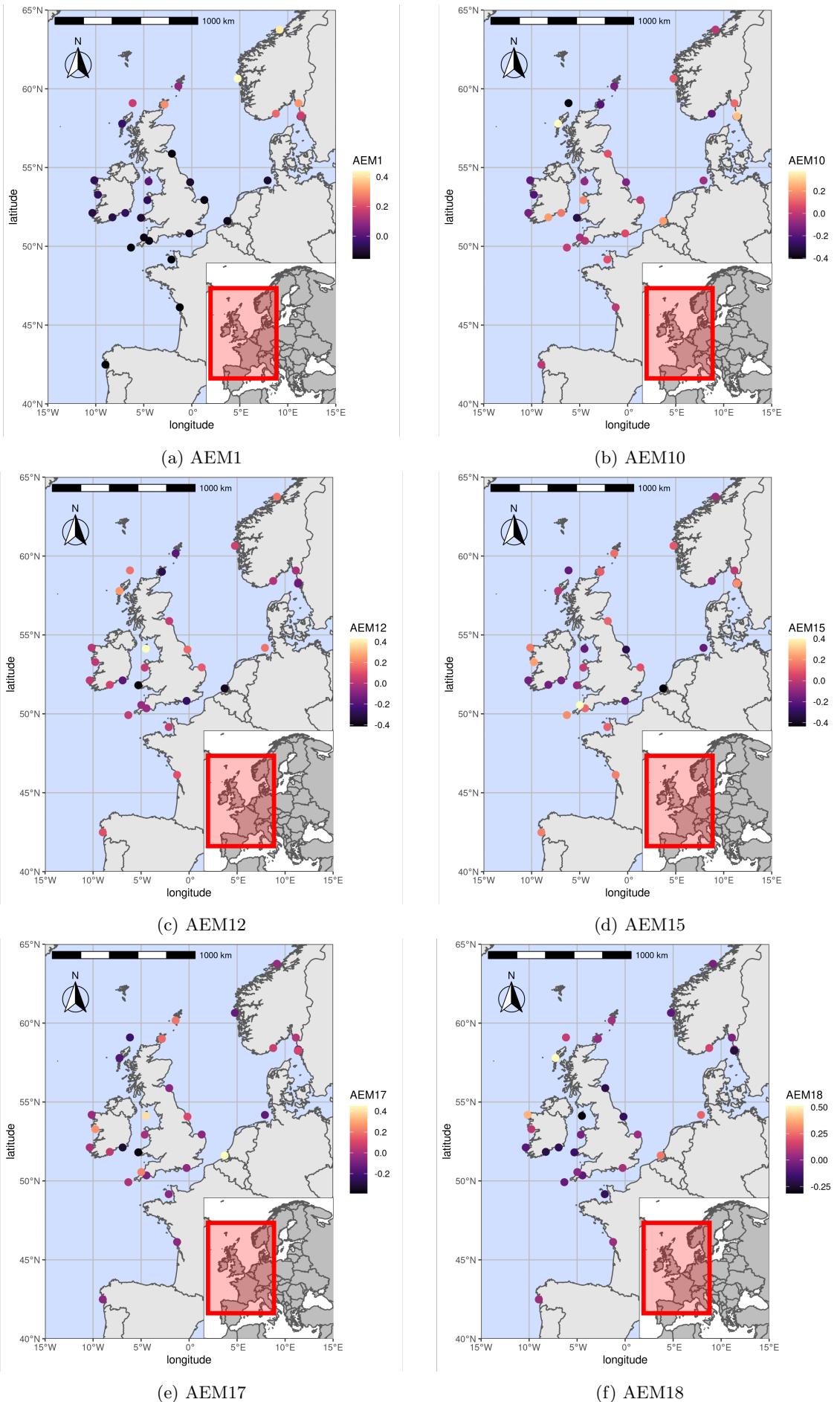


Figure F.5: Geographic maps with the AEMs

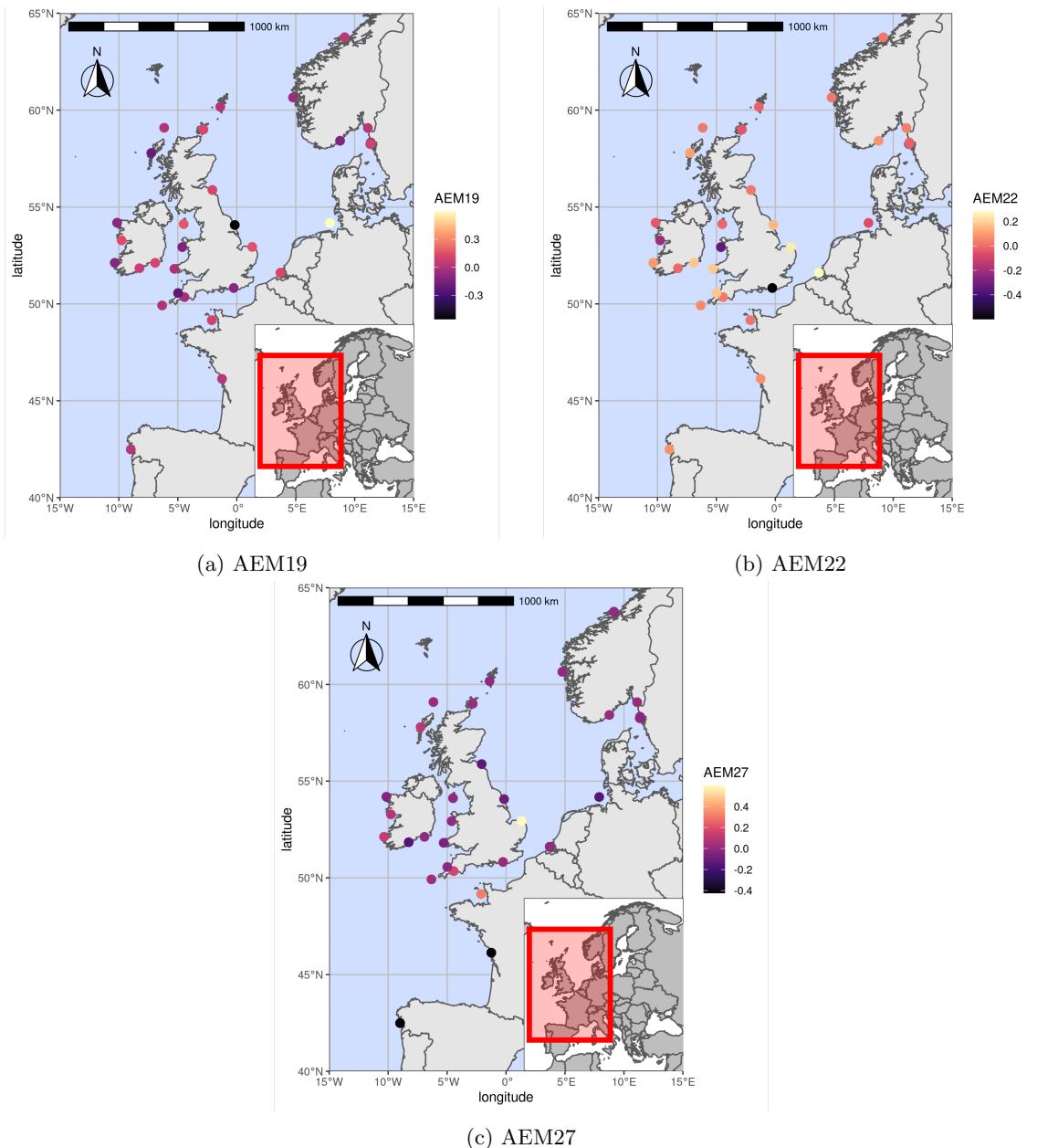


Figure F.6: Geographic maps with the AEMs - continuation

Appendix G Packages

package	reference
ade4	(Dray & Dufour, 2007)
adegenet	(Jombart & Ahmed, 2011)
adespatial	(Dray et al., 2022)
cowplot	(Wilke, 2020)
gdistance	(van Etten, 2017)
geosphere	(Hijmans, 2021)
ggforce	(Pedersen, 2022)
ggnewscale	(Campitelli, 2022)
ggpubr	(Kassambara, 2020)
ggspatial	(Dunnington, 2022)
gridExtra	(Auguie, 2017)
hierfstat	(Goudet & Jombart, 2022)
maptools	(Bivand & Lewin-Koh, 2022)
poppr	(Kamvar et al., 2015)
raster	(Hijmans, 2022)
rdacca.hp	(Lai et al., 2022)
rnaturalearth	(South, 2017a)
rnaturalearthdata	(South, 2017b)
sf	(Pebesma, 2018)
tidyverse	(Wickham et al., 2019)
vegan	(Oksanen et al., 2022)

Table G.1: List of used packages

References

- Auguie, B. (2017). *gridExtra: Miscellaneous functions for “grid” graphics* [R package version 2.3]. <https://CRAN.R-project.org/package=gridExtra>
- Bivand, R., & Lewin-Koh, N. (2022). *maptools: Tools for handling spatial objects* [R package version 1.1-4]. <https://CRAN.R-project.org/package=maptools>
- Campitelli, E. (2022). *ggnewscale: Multiple fill and colour scales in ‘ggplot2’* [R package version 0.4.8]. <https://CRAN.R-project.org/package=ggnewscale>
- Diniz-Filho, J. A. F., Soares, T. N., Lima, J. S., Dobrovolski, R., Landeiro, V. L., de Campos Telles, M. P., Rangel, T. F., & Bini, L. M. (2013). Mantel test in population genetics. *Genetics and Molecular Biology*, 36(4), 475–485. <https://doi.org/10.1590/S1415-47572013000400002>
- Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guénard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., & Wagner, H. H. (2022). *adespatial: Multivariate multiscale spatial analysis* [R package version 0.3-19]. <https://CRAN.R-project.org/package=adespatial>
- Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 1–20. <https://doi.org/10.18637/jss.v022.i04>
- Dunnington, D. (2022). *ggspatial: Spatial data framework for ggplot2* [R package version 1.1.6]. <https://CRAN.R-project.org/package=ggspatial>
- Goudet, J., & Jombart, T. (2022). *hierfstat: Estimation and tests of hierarchical F-statistics* [R package version 0.5-11]. <https://CRAN.R-project.org/package=hierfstat>
- Hijmans, R. J. (2021). *geosphere: Spherical trigonometry* [R package version 1.5-14]. <https://CRAN.R-project.org/package=geosphere>
- Hijmans, R. J. (2022). *raster: Geographic data analysis and modeling* [R package version 3.6-3]. <https://CRAN.R-project.org/package=raster>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6, 208. <https://doi.org/10.3389/fgene.2015.00208>
- Kassambara, A. (2020). *ggnpubr: ‘ggplot2’ based publication ready plots* [R package version 0.4.0]. <https://CRAN.R-project.org/package=ggnpubr>
- Lai, J., Zou, Y., Zhang, J., & Peres-Neto, P. (2022). Generalizing hierarchical and variation partitioning in multiple regression and canonical analysis using the rdacca.hp R package. *Methods in Ecology and Evolution*, 13(4), 782–788. <https://doi.org/10.1111/2041-210X.13800>
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2), 209–220. https://aacrjournals.org/cancerres/article/27/2_Part_1/209/476508/The-Detection-of-Disease-Clustering-and-a
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., ... Weedon, J. (2022). *vegan: Community Ecology Package* [R package version 2.6-4]. <https://CRAN.R-project.org/package=vegan>

- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Pedersen, T. L. (2022). *ggforce: Accelerating ‘ggplot2’* [R package version 0.4.1]. <https://CRAN.R-project.org/package=ggforce>
- South, A. (2017a). *rnatuarlearth: World map data from natural earth* [R package version 0.1.0]. <https://CRAN.R-project.org/package=rnatuarlearth>
- South, A. (2017b). *rnatuarlearthdata: World vector map data from natural earth used in ‘rnaturalearth’* [R package version 0.1.0]. <https://CRAN.R-project.org/package=rnatuarlearthdata>
- van Etten, J. (2017). R package *gdistance*: Distances and routes on geographical grids. *Journal of Statistical Software*, 76(13), 21. <https://doi.org/10.18637/jss.v076.i13>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wilke, C. O. (2020). *cowplot: Streamlined plot theme and plot annotations for ‘ggplot2’* [R package version 1.1.1]. <https://CRAN.R-project.org/package=cowplot>
- Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M.-J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology*, 27(10), 2347–2364. <https://doi.org/10.1111/mec.14589>