1. **Major comments**

- *[1] The paper suffers from wrong word choices and poor English. This obstructs the reading flow and sometimes it is unclear what is actually meant by a given sentence. The problems occur throughout the text and are too numerous to be listed. In the revision, the text should be thoroughly screened and improved (e.g. by a English).*

  After having been revised to match the reviewers advices, the manuscript will be reviewed by an external English adviser.

  ─────────────────

- *[2] The model description in section 2.1 is very, very brief, and it's not highly informative. This way, the model basically remains essentially a black box to the reader (which I personally don't like when reading modelling papers). This especially pertains to the biogeochemical model: it impossible to get a feel of how simple or complex this biogeochemical model formulation is. For example, I think the sediment domain only covers the shelf, and not the whole Black Sea. Similarly, no information at all is provided on the parameterization of the model. It would be highly coincidental that all parameters exactly the same as in Grégoire et al. [10]? For example, the resuspension description is not in in Grégoire et al. [10]. In principle, scientific work should be repeatable as to verify the conclusions. Based on the limited information, the model simulations presented here cannot be repeated.*

  The physical model is the GHER 3D hydrodynamical model already described in a lot of publications (e.g. Beckers [2], Delhez [5]) and in the particular case of the Black sea in Grégoire et al. [8, 9], Stanev and Beckers [14], Beckers et al. [3]). The implementation used in the frame of these simulations is described in Capet et al. [4]

  The biogeochemical model is described in details in Grégoire et al. [10]. Compared to Grégoire et al. [10], the model has been coupled (online) in 3D with the GHER 3D hydrodynamic model and extended with a dynamic representation of the benthic compartment (in Grégoire et al. [10], a reflective boundary condition was used for describing benthic degradation). This representation is based on the comparative analysis performed by Soetaert et al. [12] on the representation of benthic processes that is both reliable and tractable for long term simulations in a 3D ocean model. Hence, the benthic compartment is described by a semi-empirical model whose state variables are 2D variables (vertically integrated C and N content with two degrees of liability). From these 2D stock variables, the fluxes of solutes are estimated using bulk parameters (i.e. fraction of organic matter degraded by denitrification and anaerobic oxidation, oxygen consumption in nitrification). As in Soetaert et al. [12], these bulk parameters are estimated from Monte Carlo simulations performed with an existing 1D diagenetic model of the Black sea sediments developed and validated by Wijsman et al. [17]. Concerning the resuspension, we used a setting similar to that used in Stanev and Kandilarov [13].

1

The benthic module is used for the whole domain and is not restricted to the shelf area. We are on the point of finalizing a manuscript describing in details the mechanisms and results of the sediment model, as well as the validation of bentho-pelagic fluxes against in-situ estimates. However, the detailed description and validation of such a model would be too long for inclusion in this manuscript that we prefer to keep focused on the hypoxia issue and oxygen dynamics. Therefore, here, we propose to extend the description of the benthic model and we will refer to the work of Soetaert et al. [12] and to the submitted manuscript.

—————————

- *[3] The model output for oxygen is compared to point measurements available in the WOD and BSC databases. This is done in a systematic way, by first introducing the model skill metrics in section 2.3. and then doing the model data comparison in section 3 (I like the question and answer format of this section). However, the O2 is just one of the many state variables. How does a model data comparison work out for these other state variables? For example, the temperature of the water column in spring is critical (March SST features in simple model for hypoxia index H). How good is the model at representing March SST? No information is provided here, and neither is it discussed whether such a model data comparison has been done in any of the previous papers.*

The objective here was to restrain the focus of the manuscript to the issue of hypoxia and hence on the oxygen dynamics and the ability of the model to reproduce this dynamics. We therefore use all the observations on oxygen (the most available observation in the Black sea with temperature) to assess the ability of the model to represent the oxygen dynamics during the last decades over the whole shelf.

Oxygen dynamics on the shelf derives form hydrodynamic and physical factors, water column and benthic biochemical process.

Concerning the hydrodynamics, the interannual variability of the model hydrodynamics and physics has been presented in Capet et al. [4] and validated in details with satellite SST and Sea level elevation data. Among others this validation exercise demonstrates the ability of the model to correctly reproduce the interannual variability of the SST, and that of the mixed layer depth. The good performances of the model to simulate the physical fields suggest that it will be able to represent the ventilation process and oxygenation of waters.

The simulated benthic fluxes of Oxygen , Nitrate and Ammonia match the ranges and spatial variability revealed by the available estimates [6, 7, 1]. This comparison suggests that the contribution of the sediment to the oxygen dynamics is satisfactorily represented by the model.

Concerning the biogeochemical model, it has been validated for the deep sea using all the variables available during the eutrophication period Grégoire et al. [10]. For the shelf, the very good fit of simulated oxygen values with observations and the good representation by the model of the ventilation and benthic oxygen consumption implies a good simulation of the balance between photosynthesis and respiration. These points are detailed in the revised version.

—————————

- *[4] An important novel message in this paper is that accumulation of organic matter in the sediment during hypoxic periods may aggravate future hypoxia. This is illustrated in Figure 12. However, this conclusion has to be thoroughly checked, as the figure 12 is probably wrong. A typical accumulation of 10 mmol C m-2 (see figure 12) and a decay constant of 0.1 yr-1 for semi-labile organic matter provides an extra oxygen demand of 1 mmol C m-2 yr-1. This number is way too small to have an impact of future hypoxia. Typically, sediments have an oxygen demand of 2000 – 10000 mmol C m-2 yr-1, so the extra organic matter accumulation only contributes 0,025% or less to the O2 demand. The units of figure 12 are probably wrong.*

The message may have been misunderstood, which of course denote the need of clarifying the issue in the manuscript.

We indeed identified an accumulation of organic matter in the sediment that aggravates hypoxia in the years following high riverine discharge. However this accumulation is due to the high riverine discharge regime and to the low decay rate of the semi-labile sediment pool of organic matter, and not directly to the hypoxic conditions. In the model, the remineralization rate is not influenced by the bottom oxygen but is imposed as a fixed rate and modulated by temperature. This issue may be the object of long discussion but the lack of data allowing to relate the interannual variability of benthic remineralization to the intensity of seasonal hypoxia would prevent to calibrate this process in the model.

The units of figure 12 are effectively wrong, in the sense that a mention "10^3" has been omitted for the Y-axis. This figure is corrected in the revised manuscript. The sediment stock typically ranges spatially between 3000-14000 mmol C/m^2. In winter, this stock is mainly composed of semi-labile organic matter since the fresh labile component has been either remineralized during the summer or resuspended in early winter. We use a decay rate of 0.003 $d^{-1}(\sim 1$ $yr^{-1})$ for the semi-labile stock while a value of 0.0753 $d^{-1}(\sim 27.5$ $yr^{-1})$ is used for the labile part. The oxygen demand from sediments (~ 27.4 mmolC m-2/d) then effectively lies within the range of 2000 –10000 mmol C m-2 /yr . For comparison, in-situ measurements (benthic chambers estimates) range between 10-50 mmolO/m^2/d (Apr-Oct).

The important fact that the interannual accumulation of organic matter in the sediment intensifies the phenomenon of hypoxia has also been evidenced by Turner et al. [15], for the Gulf of Mexico. We clarify that point in the revised version of the manuscript.

- *5] Hypoxia index : In my view, the hypoxia index defined in Equation (1) can be defined in a much more meaningful way. Right now it has strange physical units (unit of area times unit of time) and is not directly interpretable. However if one would properly renormalize H for the integration period and define it as:*

$$H = \frac{1}{T} \int_0^T A(t)dt \tag{1}$$

If T= 1 year then H would denote yearly averaged hypoxic area. If one further normalizes for the total shelf area A_{shelf}

$$H = \frac{1}{T.A_{shelf}} \int_0^T A(t)dt \qquad (2)$$

This way, one obtains a hypoxic index between 0 and 1, defining the yearly averaged fraction of the shelf that becomes hypoxic.

The definition of the hypoxia index can indeed be discussed. Usually, the surface of hypoxia is computed from in-situ observations assuming synoptic data and is interpreted as the maximal extension of the hypoxic area. It is effectively difficult to gather from in-situ measurements the detailed dynamics of the surface of hypoxia. Examples of such dynamics are given in Fig. 1 to illustrate that these are not regular, which stresses the difficulty to establish a correspondence with field estimates. Here we have this dynamics at our disposal (weekly outputs) and therefore take the opportunity to exploit it.

Regarding ecosystem issues, both spatial and temporal extension import but these two values cannot be merged into one number without some loss of information. As detailed in the answer to the second referee, the damaging effect of hypoxia on the ecosystem is a complex matter that can not be directly addressed here. In fact indexes should be defined specifically for given species, given groups or given ecosystem functions, also considering the issue of the threshold of oxygen concentration used in the definition.

Here, our objective, for the sake of simplicity and accessibility, is to represent generally the environmental pressure of hypoxia through one single number, assuming the simplifications it may require. Integrating the hypoxic area over the year then appears as the simplest way to merge both spatial and temporal aspect.

As accurately underlined by the referee, it is true that units of area are more interpretable and friendly than unit of area times unit of time and we revised the index definition accordingly (see below). However, we think that the yearly averaged hypoxic area does not bear sense since we know it is a seasonal process. Eventually, such an H expressed in unit of area could be wrongly interpreted as an area of hypoxic area effectively encountered at a given time, which should be avoided

In order to produce an index expressed in area units, that can be compared from year to year and that corresponds to some meaningful value, we thus propose to use a typical temporal scale of hypoxia to normalize the index.

$$H = \frac{1}{\overline{D}} \int_{year} A(t)dt, \qquad (3)$$

where $\overline{D}$ denotes the average D over all the years.

If we define, for each year, the duration of hypoxia by

$$D = \frac{1}{\max A(t)} \int_{year} A(t)dt, \qquad (4)$$

4

which corresponds roughly to the time during which hypoxia occurs over half of the maximal area (see Fig. 1). With this definition of the duration of hypoxia, Eq. 3 implies that H is equivalent to the maximal extension if $D = \overline{D}$, i.e. if the duration is assumed to be equal to the average (see Fig. 1). Finally, this formulation allows an easy algebraic relationship between the index H and the duration and maximal area of hypoxia, which permits a graphical representation (Fig. 1).

As it appears that this issue deserves more attention, we develop Section 5.1 in the revised version of the manuscript and discuss the difficulties of defining H and D.

In particular we introduce the equivalence in terms of ratio of the shelf area for the sake of comparison with other sites suffering from seasonal hypoxia.
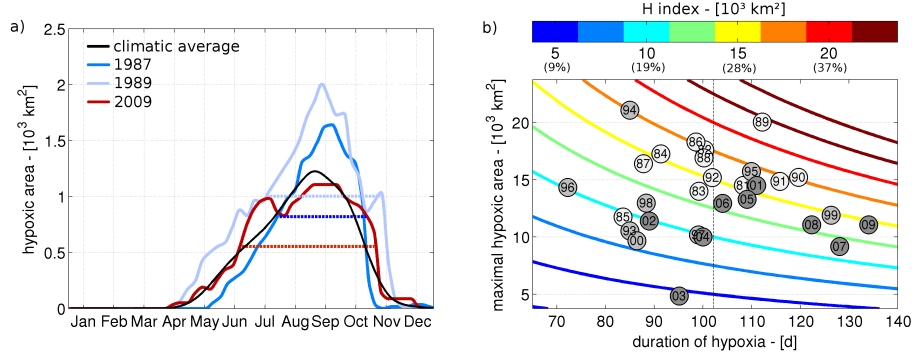

Figure 1: (a) Seasonal dynamics of the hypoxic area for selected years (color plain lines) and climatological average (black plain line). Dotted lines indicate the duration of hypoxia defined by Eq. 4. (b) The H index (colored lines) quantifies the intensity of the seasonal hypoxia event for a given year, and is computed (Eq. 3) as the integral of the hypoxic area over the year, normalized for the average duration of hypoxia (dotted line). The circles locates the hypoxic event diagnosed for each years (1981 -1991 : light gray, 1992-2000 : mid gray and 2001-2009: dark gray) which evidence the variabilities and the lack of relationship between the spatial and temporal extension of hypoxia. H is expressed in units area (if the duration is equal to the average duration, H equates the maximal extension of hypoxia). For inter-site comparability purposes the percentages given below the color bar express H in terms of ratio of the shelf area depicted in Fig. 1.

- *[6] P18415 It has been found. . . No details are given why the index H only depends on 4 variables and how this selection procedure has been performed.*

This issue is addressed in the preceding paragraph (last lines of P18414)

"Among a preselected set of potential predictors issued from the model (i.e. winter SST, March SST, mean winter organic carbon stock in the sediments, average summer winds: magnitude, zonal and meridional components, minimum summer average Brunt–Vaïsala frequency, river discharges in water, nitrate and phosphate) a stepwise regression procedure (backward elimination based on p = 0.05 significance threshold) is used to identify a subset of predictors for which no additional predictor will increase significantly the coefficient of determination of the regression (R , computed as the ratio of the variance of the response explained by the predictors Y and the true variance of Y) (Legendre and Legendre, 1998)."

Following the advice given in remark 8, we revised the set of potential predictors proposed to the stepwise procedure by considering more easily accessible atmospheric predictors. Doing so, the stepwise procedure identified that the "late summer SST", computed as the average SST over the months of Aug., Sept. and Oct. could advantageously take the place of the fourth predictor (duration of the stratification), while being close in terms of interpretation. Obviously, this modification has some implications on the formulations of the empirical equation and the manuscript has been revised accordingly.

In particular the preceding paragraph now writes :

"Among a preselected set of potential predictors issued from the model (i.e. winter SST, mean summer SST, late summer SST (Aug., Sept., Oct.) March SST, mean winter organic carbon stock in the sediments, average summer winds: magnitude, zonal and meridional components, minimum summer average Brunt–Vaïsala frequency, water and nutrients riverine discharge) a stepwise regression procedure (backward elimination based on p = 0.01 significance threshold) is used to identify an optimal subset of predictors such that no additional predictor would increase significantly the coefficient of determination of the regression (R , computed as the ratio of the variance of the response explained by the predictors Y and the true variance of Y) (Legendre and Legendre, 1998)."

- *[7] Why does fig 13 need a power law? A linear regression seems fine to me.*

We agree that for Nitrogen discharge above 25 Gmol/yr, (Fig. 13 of the manuscript), the non-linear behavior of H is not clearly evidenced suggesting that a linear law would be enough. However, we choose a non-linear law for two reasons:
1) a non-linear law has similar (very slightly better) fit statistics (considering the adjusted R^2 that accounts for the lower degree of freedom due to the additional parameter),
Linear model : $p1 * x + p2$,$R^2 : 0.9012, R^2_{adj} : 0.8976$, dfe: 27
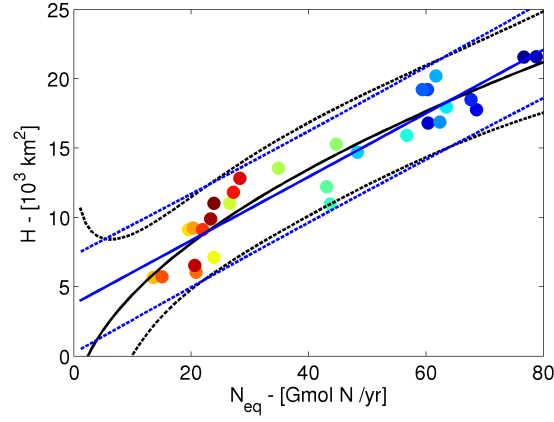Power law model : $a * x^b + c$,$R^2 : 0.076, R^2_{adj} : 0.0.9005$, dfe: 26

Figure 2: Comparison between linear and power law for Fig 13..

2) for values of N_eq below 25 Gmol/yr a non-linear fit would allow H=0 for low value of N, which bears more sense.

While the statistical model presented here is not validated for very low nitrate discharges (not occurring during the period 1980-2010), we nevertheless prefer that the proposed statistical law behaves as we may expect for low nitrate discharge. We can reasonably expect that when N_{eq} is approaching zero, the H will also tend to 0. There should be an "assimilation capacity" of the system which allows "absorbing" a certain level of nitrogen riverine discharge without the occurrence of hypoxia. This is not the case if we adopt a linear curve, which prognosticates hypoxia even with no nitrate river discharge (see both curves on Fig. 2)

The absence of hypoxia for very low nitrate discharges can be justified by the analysis of historical observations [16, 18] .

7

- *[8] A comment on the practical use of index of the H index for predicting future hypoxia. Two parameters are easy to measure or constrain (March SST, annual nitrate discharge), but the other two are not easy to determine (winter stock of sedimentary semi-labile carbon and duration of stratification). The latter hence impede the practical use of the H index for predicting future hypoxia.*

The first objective of using stepwise multi linear regression model was to identify and to appreciate the respective impacts of the main drivers of hypoxia. It is delicate to use such model for "prediction" purposes afterwards, for the main reason that these are only valid for the range of predictors and conditions used to derive the model, which finally strongly limits the "predictive" possibilities.

We acknowledge however the perspicacity of the comment and explored therefore more accessible potential drivers proposed to the stepwise procedure. Doing so, the late summer SST was identified as able to replace advantageously the predictor "duration of the stratification period", given the benefit of easier accessibility, while staying close in terms of interpretation.

The manuscript is revised accordingly.

The predictors C, the winter stock of sedimentary semi-labile carbon, may not be easily overpassed or replaced, as it introduces an important aspect of the interannual dynamics of hypoxia. The gain from its inclusion as a predictor in the regression is justified by the stepwise procedure.

Being aware of the practical aspect of a simplified relationship giving a H value for various values of the predictors, we rather proposed to use the influences evidenced for the drivers other than nitrate discharge to redraw the relationship between H and N, "filtered" from the influence of climatic factor and the non-equilibrium of the sediment pool. This is what has been done in Sect. 5.4.

This section is extended in the revised manuscript and provides now a relationship between H and the predictors (using absolute values), obtained by reinserting the linear impact of climatic predictors to the power law obtained in Section 5.4.

_____

2. Minor comments

- *Title: "Sensibilities" is a wrong term here, probably one means "sensitivity". I would suggest to change the title to: Interannual variability of seasonal hypoxia on the North-Western shelf of the Black Sea.*

We propose the following :

Drivers, mechanisms and long term variability of seasonal hypoxia in the Black Sea NWS. Is there any recovery after eutrophication?

_____

8

- **The hypoxia threshold of 2 mg L-1 corresponds to 62 mmol m-3. A few times, it is referred inconsistently equaled to 65 mmol m-3**

The value of 62 mmol m-3 has effectively been used in the analysis. The manuscript is corrected accordingly.

———————————

- **Table 1. What is the difference between symbols Z and z? Why do they differ?**

Z refers to the total depth of the water column, i.e. the bathymetry. This first criterion is introduced in order to exclude "hyper coastal" measurements that may be influenced by small scale processes unresolved by the model. The particular value of 17 m is used to filter out the coastal grid point, as the minimal bathymetry used in the model is 15m. This point has been clarified in the revised manuscript.

z refer to the depth of the observation. This criterion restricts the data selection to below the mixed layer depth, where hypoxia takes place.

———————————

- **P18398 L3 "to compensate" : wrong word choice.**

"Compensate" is replaced by "counterbalance"

———————————

- **Repartition: wrong term used many times throughout the ms : change to "distribution"**

The term "repartition" is replaced accordingly.

———————————

- **P18400 L15 "modifies the sedimentary geochemical cycling through the removal of bioturbating infauna"**

The manuscript is modified accordingly.

———————————

- **P18404 L 11 I don't understand : "horizontal variables" ?**

We meant that the sediment layer is represented by 2D variables, i.e. vertically integrated. This section has been rewritten and extended in order to answer to both reviewers comments.

9

- *P18404 L 25 represents 3Gmol yr-1 of what?*

Total nitrogen deposition (inorganic and organic), [11], introduced in the model in the form NOx. The text has been changed to *"3GmolN yr-1"*

Bibliography

[1] Akoumianaki, I.: , in prep, 2010.

[2] Beckers, J.-M.: Application of the GHER 3D general circulation model to the Western Mediterranean, J. Marine Syst., 1, 315–332, doi:DOI:10.1016/0924-7963(91) 90001-B, URL http://www.sciencedirect.com/science/article/B6VF5-48B0D4F-2H/2/3638e58a3b0406bfbe4ba4e771dbfad2, 1991.

[3] Beckers, J.-M., Grégoire, M., Nihoul, J. C., Stanev, E., Staneva, J., and Lancelot, C.: Modelling the Danube-influenced North-western Continental Shelf of the Black Sea. I: Hydrodynamical Processes Simulated by 3-D and Box Models, Estuar. Coast. Shelf S., 54, 453–472, doi:DOI:10.1006/ecss.2000.0658, URL http://www.sciencedirect.com/science/article/pii/S0272771400906589/, 2002.

[4] Capet, A., Barth, A., Beckers, J.-M., and Grégoire, M.: Interannual variability of Black Sea's hydrodynamics and connection to atmospheric patterns, Deep-Sea Res. Pt. II, 77–80, 128–142, doi:10.1016/j.dsr2.2012.04.010, URL http://www.sciencedirect.com/science/article/pii/S0967064512000586, satellite Oceanography and Climate Change, 2012.

[5] Delhez, E.: Reconnaissance of the general circulation of the North-Western European Continental Shelf by means of a three-dimensional turbulent closure model, Earth Sci. Rev., 41, 3–29, doi:DOI:10.1016/0012-8252(96)00012-8, URL http://www.sciencedirect.com/science/article/pii/0012825296000128, 1996.

[6] Friedl, G., Dinkel, C., and Wehrli, B.: Benthic fluxes of nutrients in the northwestern Black Sea, Mar. Chem., 62, 77–88, 1998.

[7] Friedrich, J., Dinkel, C., Friedl, G., Pimenov, N., Wijsman, J., Gomoiu, M.-T., Cociasu, A., Popa, L., and Wehrli, B.: Benthic nutrient cycling and diagenetic pathways in the north-western Black Sea, Estuar. Coast. Shelf S., 54, 369–383, 2002.

[8] Grégoire, M., Beckers, J.-M., Nihoul, J. C., and Stanev, E.: Reconnaissance of the main Black Sea's ecohydrodynamics by means of a 3D interdisciplinary model, J. Marine Syst., 16, 85–105, doi:10.1016/S0924-7963(97)00101-2, URL http://www.sciencedirect.com/science/article/pii/S0924796397001012, 1998.

[9] Grégoire, M., Soetaert, K., Nezlin, N., and Kostianoy, A. G.: Modeling the nitrogen cycling and plankton productivity in the Black Sea using a three-dimensional interdisciplinary model, Journal of Geophysical Research (Oceans), 109, C05 007, doi:10.1029/2001JC001014, 2004.

[10] Grégoire, M., Raick, C., and Soetaert, K.: Numerical modeling of the central Black Sea ecosystem functioning during the eutrophication phase, Progr. Oceanogr., 76, 286–333, doi: 10.1016/j.pocean.2008.01.002, 2008.

[11] Kanakidou, M., Duce, R. A., Prospero, J. M., Baker, A. R., Benitez-Nelson, C., Dentener, F. J., Hunter, K. A., Liss, P. S., Mahowald, N., Okin, G. S., Sarin, M., Tsigaridis, K., Uematsu, M., Zamora, L. M., and Zhu, T.: Atmospheric fluxes of organic N and P to the global ocean, Global Biogeochem. Cycles, 26, GB3026, doi:http://dx.doi.org/10.1029/2011GB004277, 2012.

[12] Soetaert, K., Middelburg, J. J., Herman, P. M., and Buis, K.: On the coupling of benthic and pelagic biogeochemical models, Earth Sci. Rev., 51, 173–201, doi:DOI: 10.1016/S0012-8252(00)00004-0, URL http://www.sciencedirect.com/science/article/pii/S0012825200000040, 2000.

[13] Stanev, E. and Kandilarov, R.: Sediment dynamics in the Black Sea: numerical modelling and remote sensing observations, OD, 62, 533–553, doi:10.1007/s10236-012-0520-1, URL http://dx.doi.org/10.1007/s10236-012-0520-1, 10.1007/s10236-012-0520-1, 2012.

[14] Stanev, E. V. and Beckers, J.-M.: Numerical simulations of seasonal and interannual variability of the Black Sea thermohaline circulation, Journal of marine systems, 22, 241–267, 1999.

[15] Turner, R. E., Rabalais, N. N., and Justic, D.: Gulf of Mexico hypoxia: Alternate states and a legacy, Environmental science & technology, 42, 2323–2327, 2008.

[16] UkrSCES: State of the Black Sea Environment, National report of Ukraine, 1996-2000, Astroprint, Odessa, 2002.

[17] Wijsman, J. W., Herman, P. M., and Gomoiu, M.-T.: Spatial distribution in sediment characteristics and benthic activity on the northwestern Black Sea shelf, Marine Ecology Progress Series, 181, 25–39, 1999.

[18] Zaitsev, Y.: Marine biological diversity in the Black Sea : A study of change and decline, vol. 3 of Black Sea environmental series, United Nation Publications, New York, 1997.