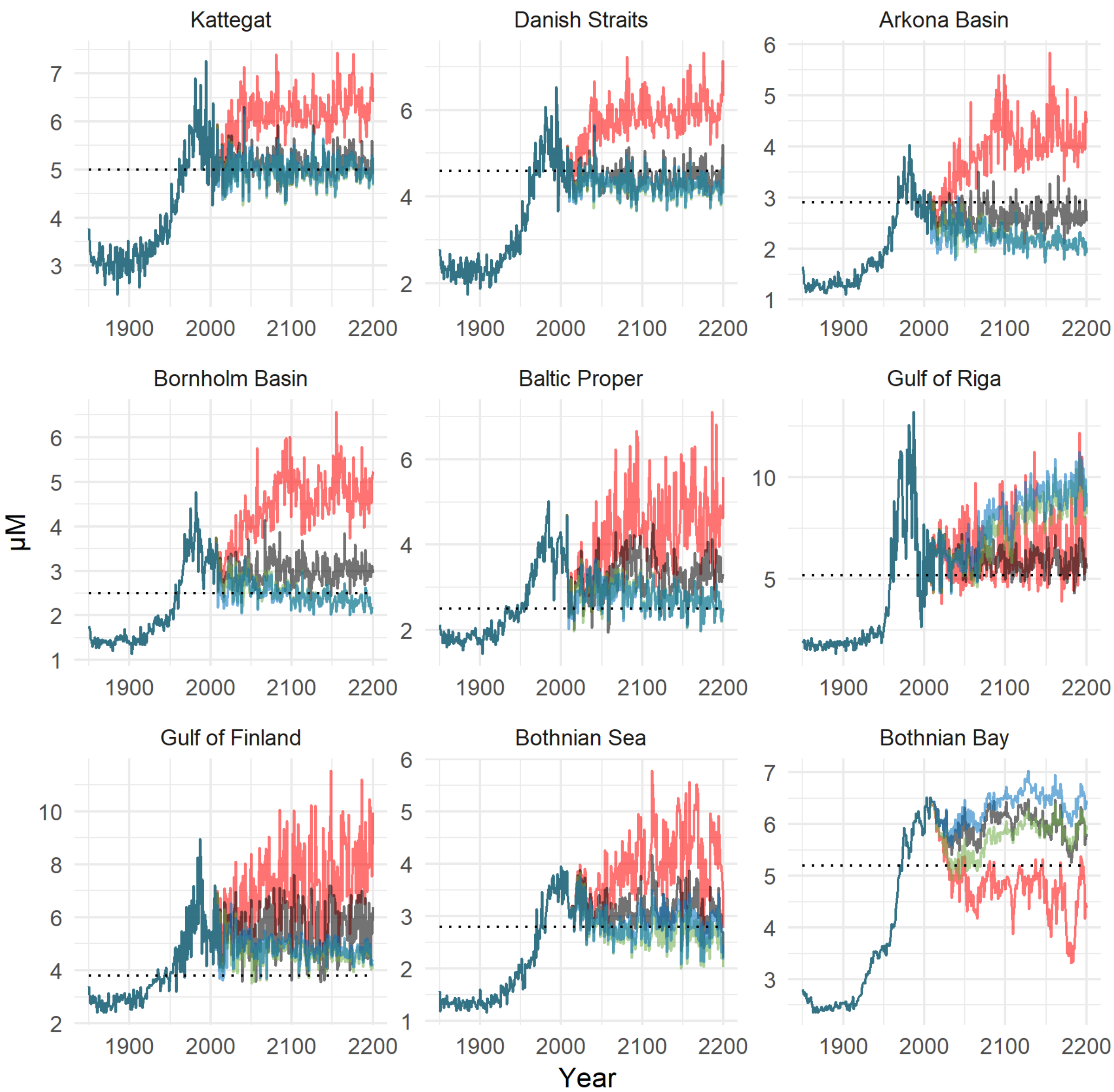
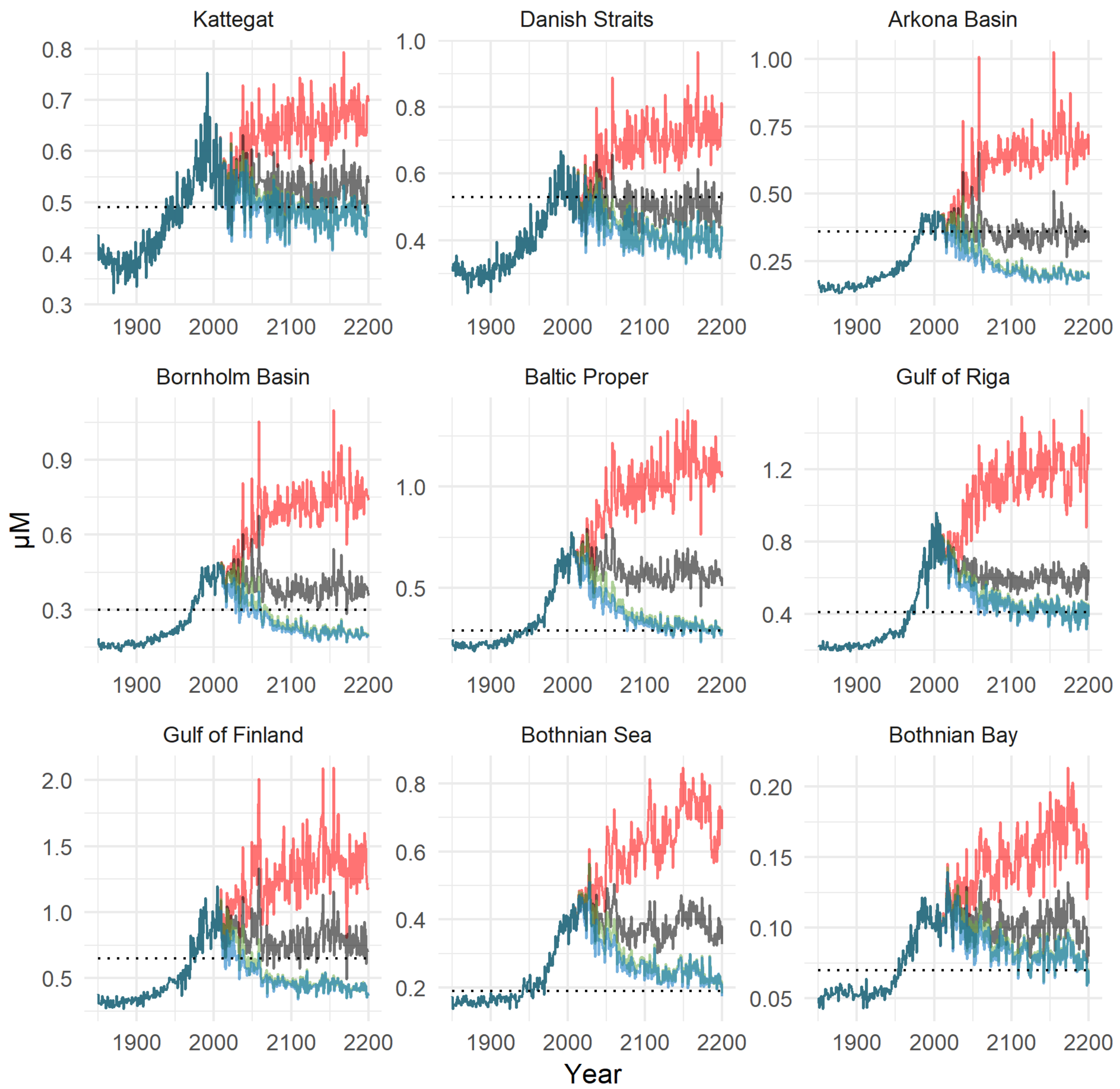
Supplementary Material

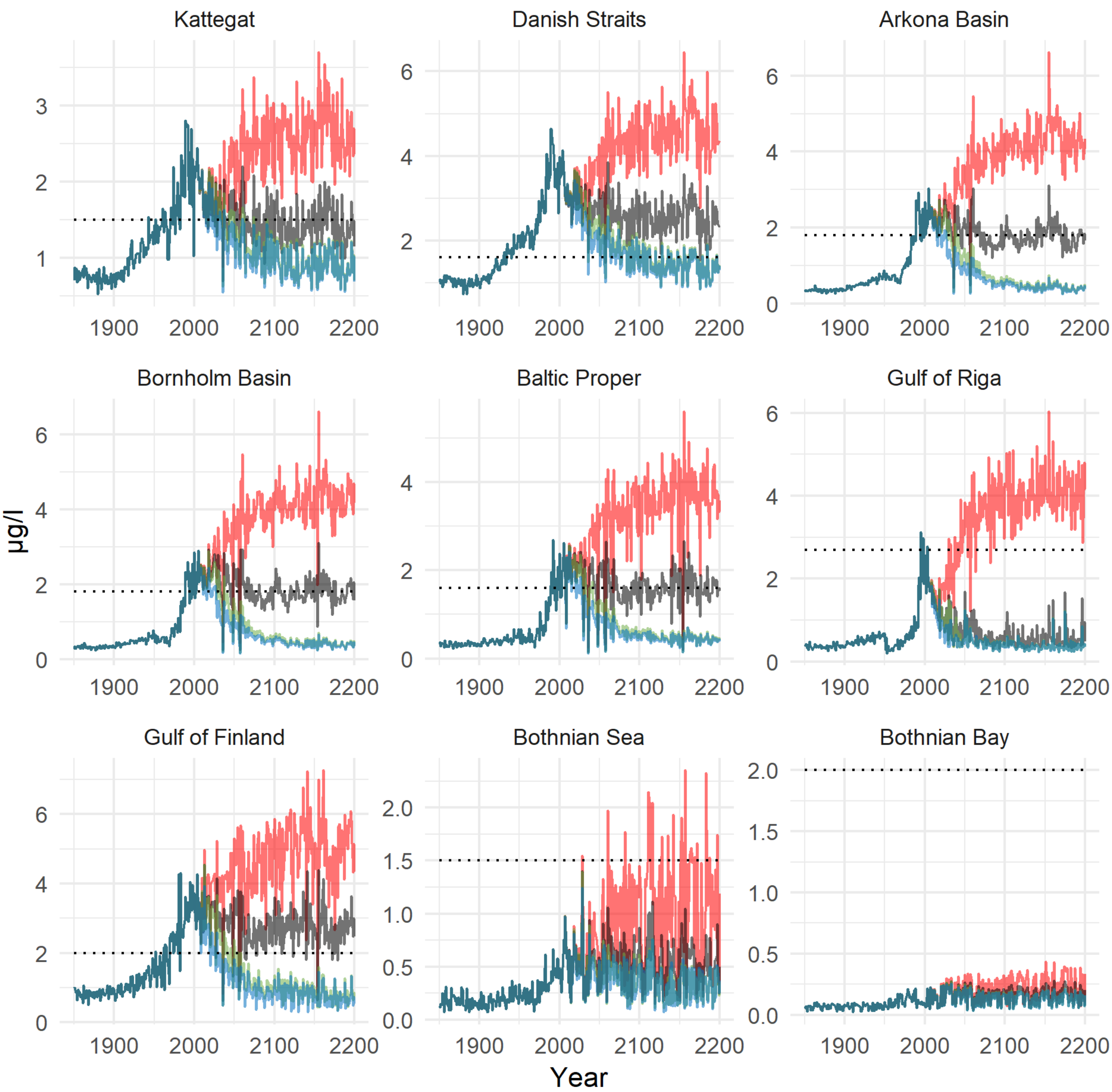
1 Supplementary Figures



Supplementary Figure S1. Winter DIN. BALTSEM results by basin from 1850 to 2200 for 4 load scenarios (blue: BSAP0, green: BSAP30, black: PLC5.5, red: BAU30). The dotted line indicates the basin-specific target value used in the HEAT assessment calculations.

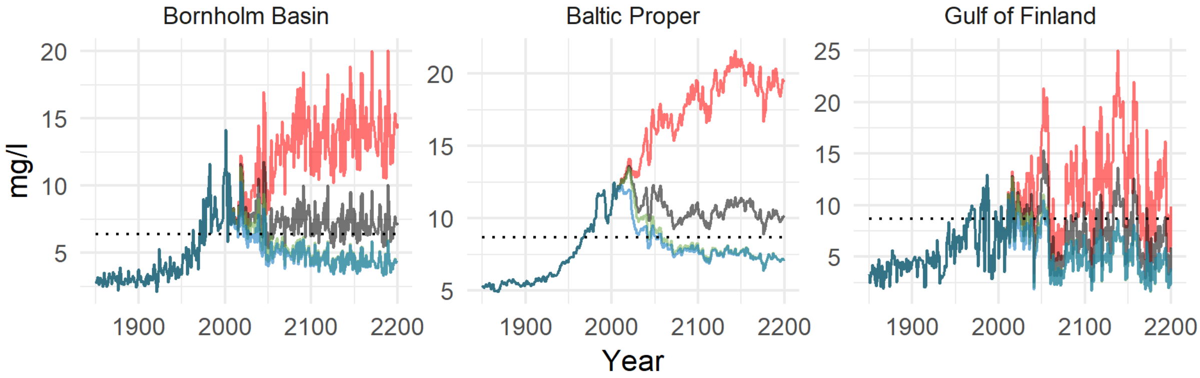


Supplementary Figure S2. Winter DIP. BALTSEM results by basin from 1850 to 2200 for 4 load scenarios (blue: BSAP0, green: BSAP30, black: PLC5.5, red: BAU30). The dotted line indicates the basin-specific target value used in the HEAT assessment calculations.

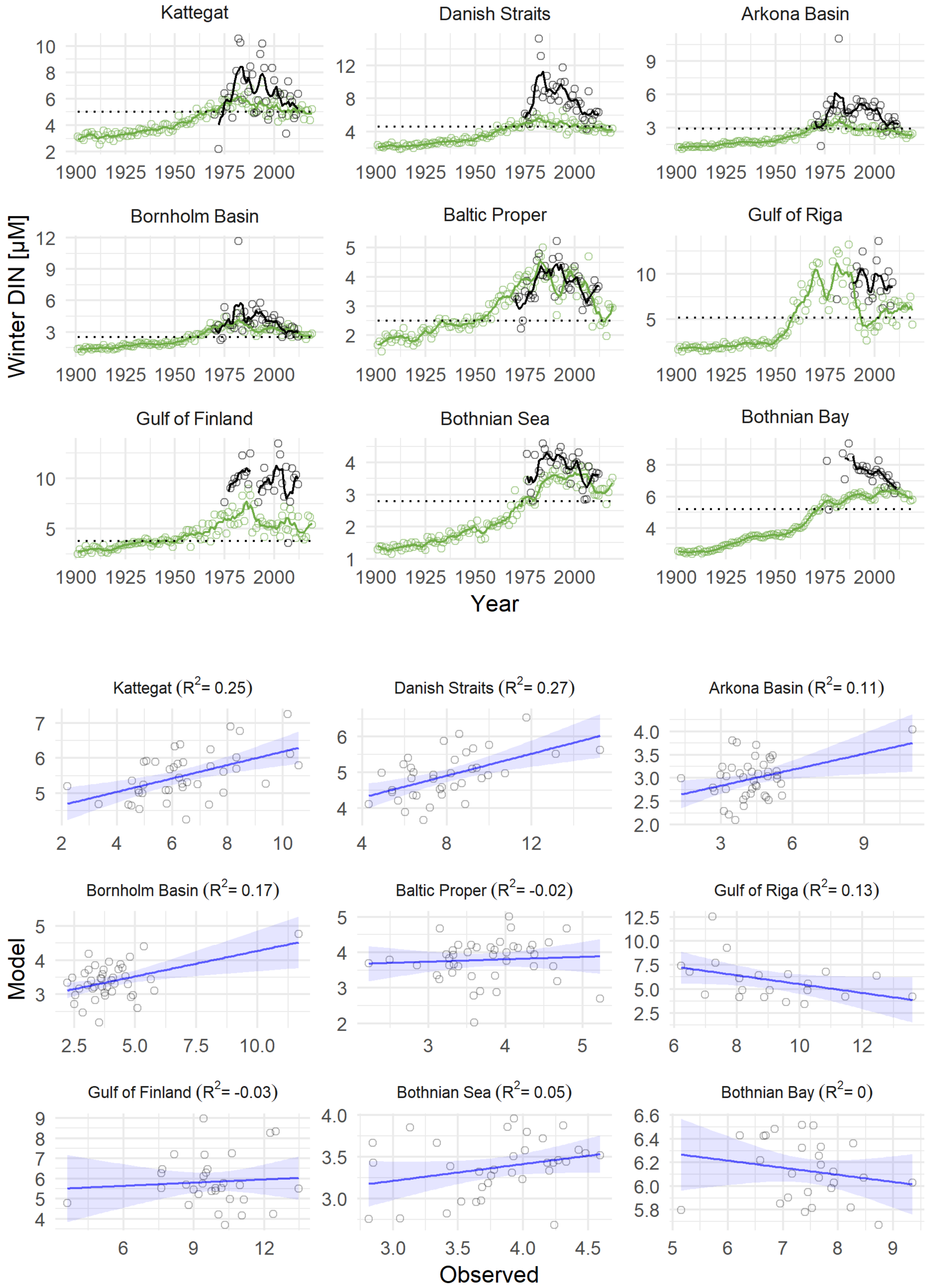


Supplementary Figure S3. Summer Chl a. BALTSEM results by basin from 1850 to 2200 for 4 load scenarios (blue: BSAP0, green: BSAP30, black: PLC5.5, red: BAU30). The dotted line indicates the basin-specific target value used in the HEAT assessment calculations.

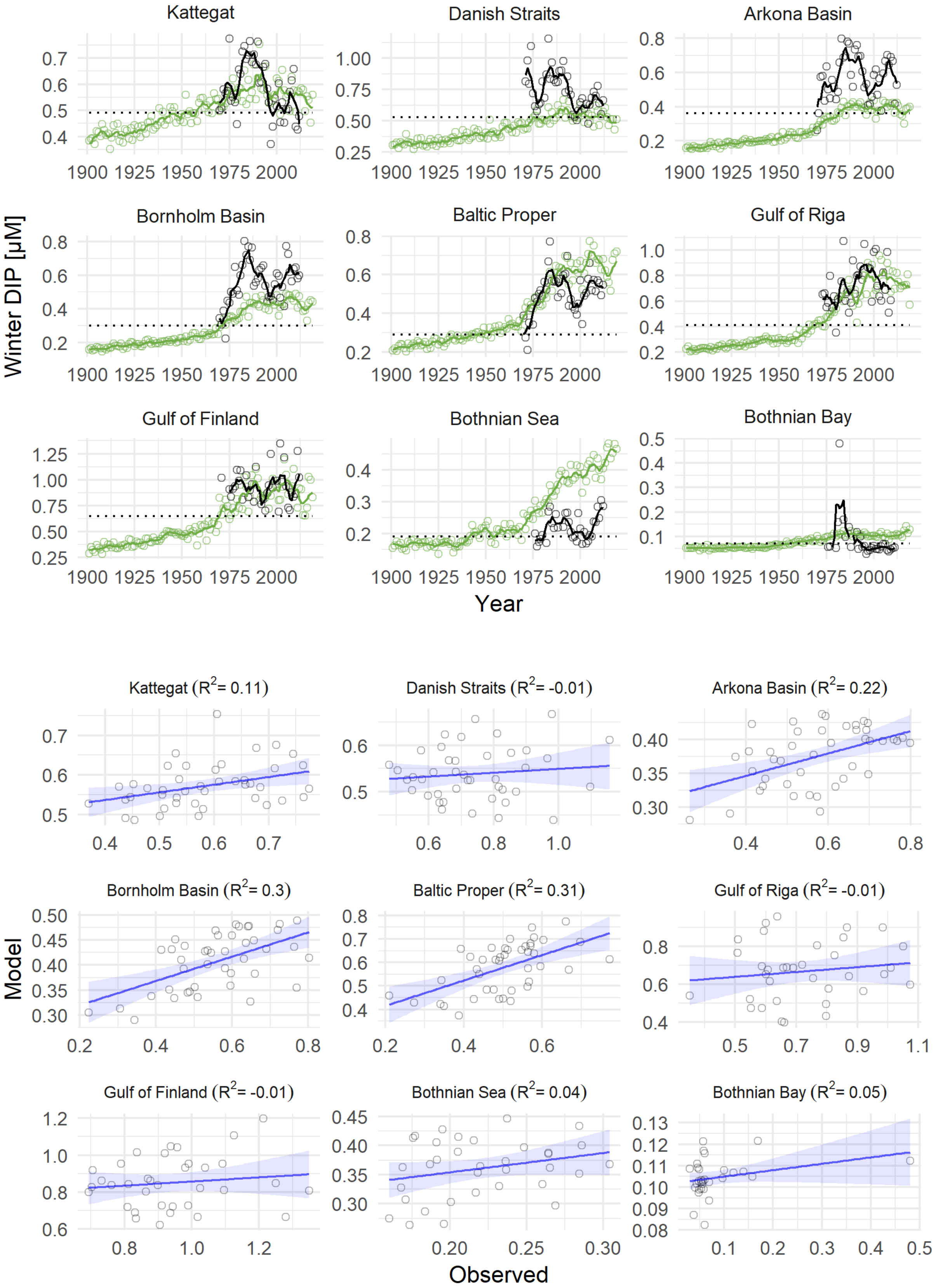
Supplementary Figure S4. Secchi depth. BALTSEM results by basin from 1850 to 2200 for 4 load scenarios (blue: BSAP0, green: BSAP30, black: PLC5.5, red: BAU30). The dotted line indicates the basin-specific target value used in the HEAT assessment calculations.



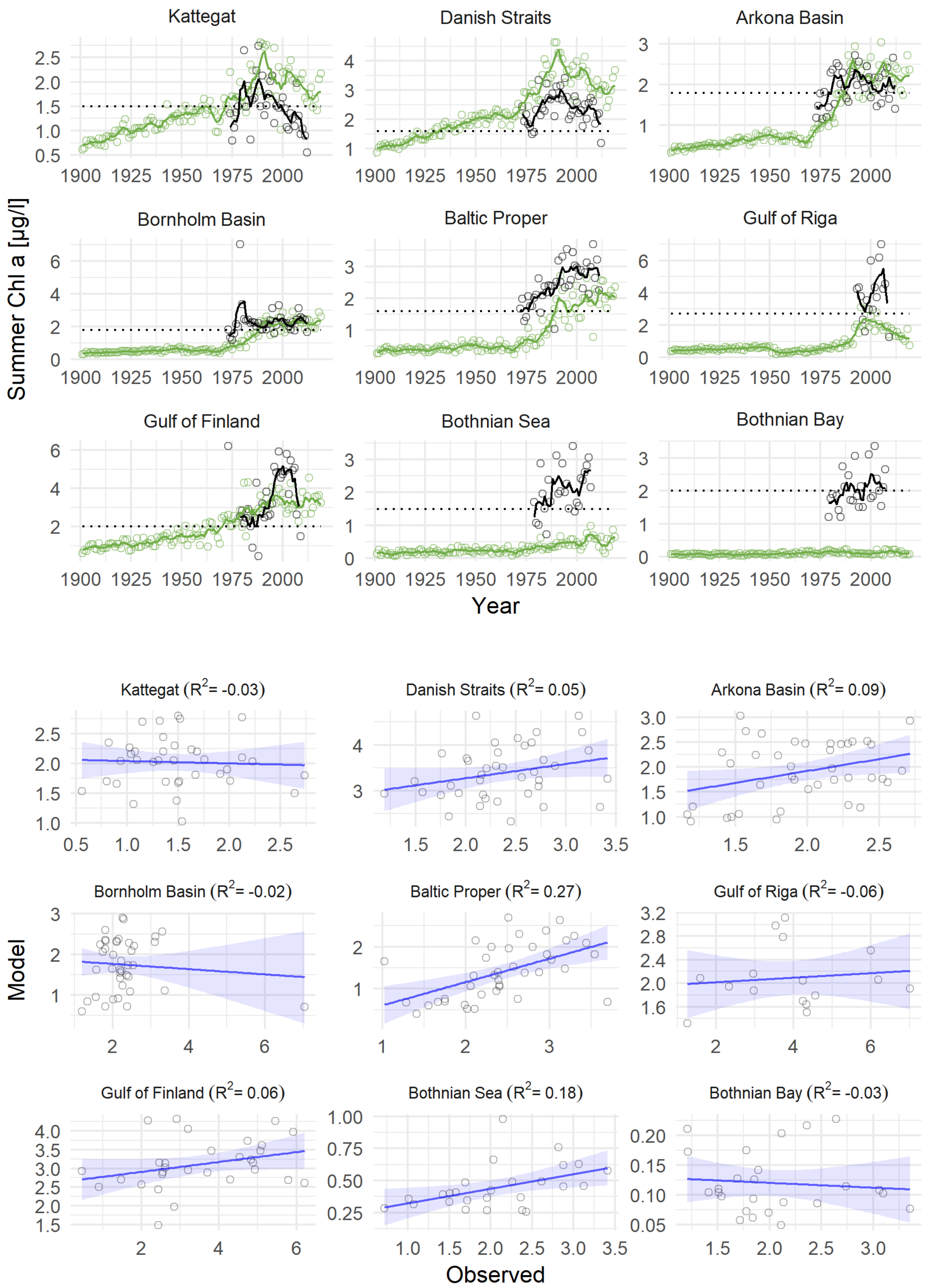
Supplementary Figure S5. Oxygen debt. BALTSEM results by basin from 1850 to 2200 for 4 load scenarios (blue: BSAP0, green: BSAP30, black: PLC5.5, red: BAU30). The dotted line indicates the basin-specific target value used in the HEAT assessment calculations.



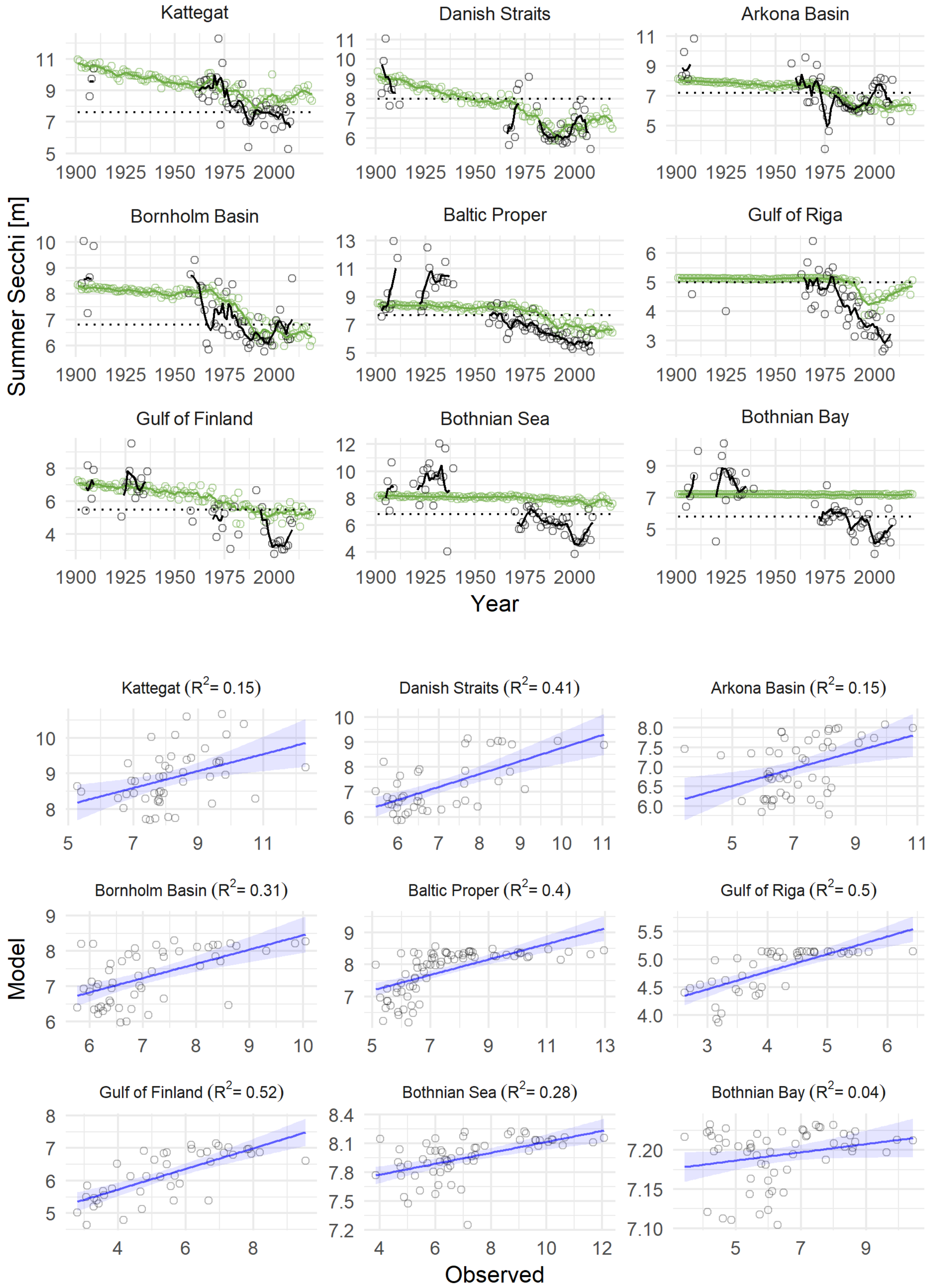
Supplementary Figure S6. Winter DIN. Upper panel: Timeseries BALTSEM (BSAP30) vs. observations (Andersen et al. 2017). Lower panel: Regression of BALTSEM (BSAP30) vs. observations.



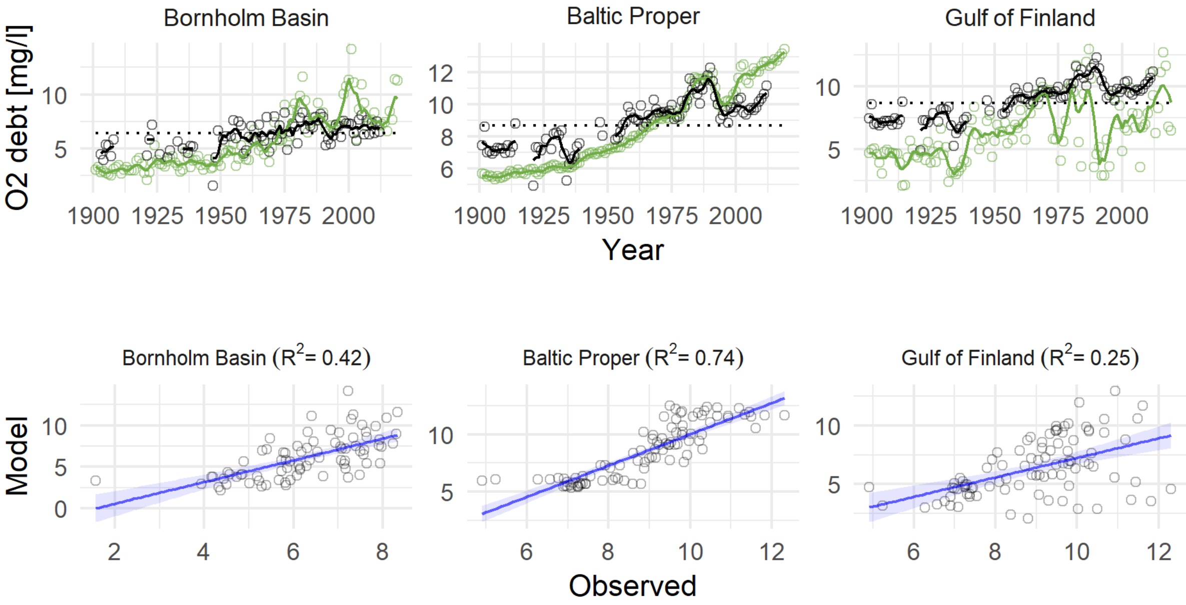
Supplementary Figure S7. Winter DIP. Upper panel: Timeseries BALTSEM (BSAP) vs. observations (Andersen et al. 2017). Lower panel: Regression of BALTSEM (BSAP) vs. observations.



Supplementary Figure S8. Summer Chl a. Upper panel: Timeseries BALTSEM (BSAP) vs. observations (Andersen et al. 2017). Lower panel: Regression of BALTSEM (BSAP) vs. observations.



Supplementary Figure S9. Secchi depth. Upper panel: Timeseries BALTSEM (BSAP) vs. observations (Andersen et al. 2017). Lower panel: Regression of BALTSEM (BSAP) vs. observations.



Supplementary Figure S10. Oxygen debt. Upper panel: Timeseries BALTSEM (BSAP) vs. observations (Andersen et al. 2017). Lower panel: Regression of BALTSEM (BSAP) vs. observations.

2 Description of HEAT calculation method

This is the method described by Andersen et al. (2017). At the core of the method is the comparison of indicator observations with target values. Briefly, the ratios of observed and target values are then averaged within 3 categories and the category having the worst result determines the overall status.

In more detail, the steps for HEAT assessment are as follows:

Step 1: Targets (Eutrophication Quality Target or ET)

The target value for an indicator is defined as being within an *acceptable deviation* from the value which defines *reference conditions.*

$$ET = RefCon \times (1 - AcDev) \quad (1)$$

where *RefCon* is reference conditions and *AcDev* is acceptable deviation from RefCon.

In this study the target values were not re-calculated but were taken directly from Andersen *et al.* (2017) and calculations started with Step 2.

Step 2: Calculation of Eutrophication Ratio (ER)

For concentrations of nutrient and chlorophyll *a*, as well as oxygen debt, the indicators have a numerically positive (+ve) response. i.e. their value increases with worsening eutrophication status. For these indicators, a Eutrophication Ratio (ER) is calculated as:

$$ER = ES/ET \text{ (+ve response)} \quad (2)$$

where *ES* is eutrophication state, i.e. the observed (or modelled) value for a given indicator in a given year.

Secchi depth shows a numerically negative (−ve) response to increasing (worsening) eutrophication. In this case, the ER is calculated as follows:

$$ER = ET/ES \text{ (-ve response)} \quad (3)$$

In this way, the value of each indicator is translated into a eutrophication ratio (ER) where a value above 1.00 indicates eutrophic status. ER values for different indicators are subsequently combined (see steps 3 and 4).

Step 3: Grouping of indicators

The indicators are grouped in three categories: C1 = nutrient levels (DIN and DIP), C2 = direct effects of eutrophication (chlorophyll-*a* and Secchi), and C3 = indirect effects of eutrophication (oxygen

debt). The average eutrophication ratio (ER) is calculated for each group. In Andersen et al (2017), C3 also included a benthic invertebrate index. Here, obviously, the ER for C3 is equal to the ER for oxygen debt.

Step 4: Classification of status

The ER for an indicator category (C1, C2 or C3) determines the status classification. Where ER is less than 1.0 the status is determined to be ”unaffected by eutrophication” and where ER is greater than, or equal to, 1.0 then the status is ”affected by eutrophication”

These two classes are further divided into five sub-classes, corresponding to the EU Water Framework Directive status classes, according to the following table:

Supplementary Table S1. Eutrophication Ratio (ER) intervals and corresponding eutrophication status, eutrophication classes from Andersen et al (2017)

ER	**Status**	**Class**
$0.0 \leq ER < 0.5$	Unaffected by eutrophication	High
$0.5 \leq ER < 1.0$		Good
$1.0 \leq ER < 1.5$	Affected by eutrophication	Moderate
$1.5 \leq ER < 2.0$		Poor
$ER \geq 2.0$		Bad

Step 5: Integrated Assessment

Using a ”one-out all-out” principle the indicator category having the worst (highest) ER value determines the overall assessment status for a basin. The HEAT score is thus the ER value of the worst category and the overall class (Bad, Poor, Moderate or Good) is the same as that of the worst category.

Reference:

Andersen, J.H., J. Carstensen, D.J. Conley, K. Dromph, V. Fleming-Lehtinen, B. Gustafsson, A. Josefson, A. Norkko, A. Villnäs and Murray, C. (2017). Long-term temporal and spatial trends in eutrophication status of the Baltic Sea. Biological Reviews 92, 135-149.