ANNEX 5

ICES DATA FOR AN HOLISTIC ANALYSIS OF FISH DISEASE PREVALENCE

1 INTRODUCTION

At its 1998 meeting, the ICES Working Group on Pathology and Diseases of Marine Organisms (WGPDMO) reviewed the results of a statistical analysis of the ICES fish disease data; a written report was presented providing information on location-specific temporal trends in the prevalence of lymphocystis, epidermal hyperplasia/papilloma, and skin ulcers in dab (*Limanda limanda*) and lymphocystis and skin ulcers in flounder (*Platichthys flesus*), as well as a comparison of trends found at different geographical locations (ICES, 1999).

The results of the analysis revealed marked spatial differences with respect to both the absolute levels and the temporal changes of the disease prevalences. Furthermore, areas characterized either by decreasing, increasing or stable temporal trends over the past years were identified. In some areas, temporal changes in the prevalence of two or more diseases were similar, thereby indicating the presence of common underlying ecological factors affecting the disease prevalence. However, the results of the analysis did not provide any information on possible causes of the observed trends, since potential explanatory factors known or suspected to be involved in disease aetiology and pathogenesis were not included in the analysis at that time.

WGPDMO emphasized that the integration of the ICES fish disease data and other types of data (e.g., contaminant, oceanographic, and fisheries-related data) would constitute a subsequent next step in the attempt to analyse the ICES fish disease data in a more holistic way, with the aim of obtaining better insight into cause-effect relationships between diseases and environmental factors. Since ICES has established different data banks with relevant data, it was recommended that available data should be assessed with respect to their usefulness for such a holistic approach.

When reviewing the 1998 WGPDMO report, the ACME endorsed the view that a more holistic statistical analysis is desirable and that the various ICES data banks could serve as a suitable data pool from which the information required could be extracted. The ACME emphasized that the first step taken, before any further actions are decided upon, should be to obtain a detailed overview of the data availability and compatibility. This should be done in close collaboration between selected WGPDMO members and the ICES Secretariat. A second step, a pilot study, could be subsequently started, using a selected subset of suitable data extracted from the ICES databases, in order to assess the practicality and perspectives of a future holistic data analysis.

The results of intersessional activities carried out according to the above recommendations, prior to the 1999 WGPDMO meeting, are presented below. The first part of the report describes the outcome of the assessment of the ICES data banks with respect to the types and amounts of data available, with particular emphasis on spatial and temporal data coverage and overlap. In the second part, preliminary results of a multivariate statistical analysis of a selected subset of data are presented. The third part provides conclusions and perspectives, e.g., by focusing on data limitations identified and on ways to accomplish a more comprehensive analysis.

2 OVERVIEW OF AVAILABLE DATA

Information on the ICES data banks from which data can be extracted for a holistic analysis and on strategies applied to obtain an overview of available data is described below. In addition, some of the data are presented in order to demonstrate temporal trends.

2.1 ICES Data Banks

ICES data considered relevant for a holistic analysis are available from the following data banks:

1) ICES Environmental Data Centre;

2) ICES Oceanographic Data Centre;

3) ICES Fishery Data Banks.

ICES provides a detailed overview (partly interactive) of the data included on its website http://www.ICES.dk. Most of the information on the data banks detailed below was extracted from this website.

A ICES Environmental Data Centre

The ICES Environmental Data Centre contains data on:

- contaminants in marine invertebrates, fish, birds, and mammals (approximately 275 000 records);
- contaminants in sea water (approximately 280 000 records);
- contaminants in sediments (approximately 80 000 records);
- biological effects of contaminants: EROD, oyster embryo bioassay (approximately 4000 records);
- fish disease prevalences (approximately 110 000 records);

- supporting data: nutrients, oxygen, temperature, salinity;
- quality assurance (QA) information.

B ICES Oceanographic Data Centre

The ICES Oceanographic Data Centre maintains two data banks in the ICES Secretariat:

- the ROSCOP data bank (information on cruise activities);
- the hydrochemical data bank (temperature, salinity, nutrients, oxygen, etc.).

In addition to these holdings, there is access to a number of project data sets (including oceanographic data from the International Young Fish/Bottom Trawl Surveys).

C ICES Fishery Data Banks

The ICES Fishery Data Banks include five fisheries-related data banks:

- STATLANT 27A (official statistics on nominal catches of fish and shellfish);
- ICES Fisheries Assessment Package (for use by approximately twenty working groups for ICES stock assessments—includes catches in tonnes, fishing effort, catch in number at age, and relevant biological data);
- International Bottom Trawl Survey (IBTS) (results from an international survey conducted each year in the North Sea since 1970 which provides an annual index of abundance by ICES Statistical Rectangle; additional data on temperature, salinity, nutrients);
- North Sea data bank (contains details of catches and fishing effort originally set up by the EC);
- North Sea multispecies data bank (contains stomach content data for each of the main predatory fish species in the North Sea—for use in multispecies models).

2.2 Strategies Applied to Obtain an Overview of ICES Data

Since the ICES data that are available and potentially relevant for a holistic analysis are overwhelming in terms of parameters measured and results of measurements, it was considered impossible to present a full overview on their spatial and temporal distribution patterns. It was, therefore, decided to extract some of the data by using *a priori* selection criteria that were mainly based on the availability of ICES disease prevalence data for common dab (*Limanda limanda*).

2.1.1 Selection criteria

Sites

Three North Sea areas (extended German Bight, Dogger Bank, Firth of Forth) were selected for which a considerable amount of disease data are available and which differ both in the absolute prevalence levels and the temporal trends recorded over the past years. As shown in Figure A5.1, the areas were relatively large and consisted of four to nine ICES Statistical Rectangles in order to obtain sufficient data for the subsequent statistical analysis.

Time span

Since the fish disease data date back to 1981, this was the year used as the starting point for the temporal overview.

Parameters

From the ICES Fishery Data Banks, data on catch per unit effort (CPUE) for dab (all specimens, independently of size) derived from the International Young Fish/Bottom Trawl Survey (IYFS/IBTS) were selected. Originally, it was also planned to incorporate fishing effort data (STECF data, EU Scientific Technical and Economic Committee for Fisheries). However, since they are not structured in an easily accessible way (e.g., effort data are only available for a short time period and separately for sixty different fishing fleets), they were excluded. From the ICES Oceanographic Data Centre, information on water temperature, salinity, dissolved oxygen, total phosphorus, phosphate, ammonium, nitrite, nitrate, silicate, and chlorophyll were considered, partly derived from the International Young Fish/Bottom Trawl Survey. From the ICES Environmental Data Centre, data on contaminants (Pb, Hg, Cd, HCH, HCB, CB118, CB153, *o,p* DDT) in water, sediments, dab muscle and liver, and blue mussel (*Mytilus edulis*) were selected.

After the selections were made, the ICES Secretariat was contacted and the data were requested in electronic form. According to the ICES data policy, raw data are not available as this would require permission from the data originators. Instead, aggregated data, consisting of, e.g., calculated mean values, were provided by the Secretariat.

2.3 Brief Overview

Figures A5.2–A5.4 provide an overview of the data available according to the selection criteria described above. In addition to the parameters mentioned, ICES disease prevalence data for female dab, size group 20–24 cm, are also included. The figures clearly show that most data are available for Area 1 (extended German Bight), followed by Area 2 (extended Dogger Bank) and, finally, Area 3 (extended Firth of Forth). Apart from the disease data, there is relatively good temporal coverage of data on CPUE, water temperature, salinity, and nutrients for all three areas for almost the whole period, but only few data are available on contaminants in water (with the exception of Area 1, where data are available from 1985 onwards), sediments (with the exception of data from 1990–1992), and dab (with the exception of Areas 1 and 2 from 1990–1996).

In order to provide an overview of temporal trends, Figures A5.5a–A5.5f show the data for selected parameters measured in Area 1, the extended German Bight. Both the observed values and the values derived from interpolation in preparation of the subsequent analysis (see below) are shown, with their corresponding confidence intervals. For contaminants in dab, only data for female dab, size group ≥ 20 cm, are included since they correspond to the standard disease prevalence data used in previous statistical analyses (females, 20–24 cm; see ICES, 1999). In the ICES Environmental Data Centre, contaminant and disease data for other size groups and for males are also available.

3 CASE STUDY (STATISTICAL ANALYSIS)

As requested by ACME, a case study was performed for which only the data set from Area 1 (extended German Bight) was used since it constituted the largest set with the best temporal coverage of parameters.

3.1 Material and Methods

Figures A5.2–A5.4 reveal that not all data to be compared were recorded at the same time and, therefore, some values had to be interpolated prior to analysis. Since the intention was to relate the data to the disease prevalence data, interpolations were calculated for those time points (days) for which disease data were available. A kernel smoother with Gaussian kernel was used to interpolate and to remove random fluctuation. The smoother bandwidth was selected by generalized cross-validation. No extrapolation outside the time range covered by real observations was done. Pointwise confidence intervals (upper and lower lines in Figures A5.5a–A5.5f) for interpolated values were derived as two standard deviations around smoothed values.

As a first approach to identify relationships between the disease prevalence and the parameters listed in Section 2.2, univariate logistic models involving observed and interpolated values were fitted (see Table A5.1).

As a second approach, a multivariate logistic model was fitted for each of three scenarios comprising different sets of parameters (see Table A5.2). In the long-term model, parameters with observations for nearly the entire range from 1981–1997 were included. The medium and short-term models contained additional parameters that covered approximately one half and one third, respectively, of this range. A stepwise selection was used to identify parameters with the highest explanatory values. A multivariate analysis comprising all parameters simultaneously was considered inappropriate at this stage as it would have required either massive extrapolations or would have been restricted to a very narrow time span.

3.2 Results and Discussion

The results of the univariate analysis are given in Table A5.1. A comparison of the ratio of the number of observations (n_o) and the number of interpolated data points (n_i) used in the analysis for each parameter provides not only an overview of the data availability but also a rough indicator of the reliability of the results of the analysis; if the balance is towards the number of observations, interpolated data can be considered more valid than if the balance is towards the interpolated values.

For each parameter and disease tested, the table includes information on the direction of the relationship and the significance levels. In total, a significant relationship was identified in 31 of 66 cases, which could be taken as an indicator of the multifactorial aetiology of the diseases considered. Water temperature was the parameter with the strongest impact and the prevalences of all three diseases were positively and highly significant when related to temperature. CPUE was also significantly correlated to the prevalence of the diseases. However, a positive relationship was only found for lymphocystis and epidermal hyperplasia/papilloma, while a negative relationship was found for acute/healing skin ulcerations. Information on the significance of other parameters can be taken directly from the table.

Table A5.2 provides information on the parameters found significantly correlated to disease prevalence by using a stepwise multivariate analysis. Assuming an 'ideal', e.g., consistent, relationship between disease prevalence and certain parameters, one would have expected that terms with a highly significant impact in the long-term model would also appear in the medium-term and short-term models for the respective diseases. However, this was only the case for water nitrate and epidermal hyperplasia/papilloma. Other significant terms occurring at least twice were salinity and cadmium in *M. edulis* (lymphocystis), CB153 in *M. edulis* and cadmium in unfiltered sea water (epidermal hyperplasia/-papilloma), and cadmium in unfiltered sea water and CPUE (acute/healing skin ulcerations).

The inconsistency between parts of the univariate and the multivariate analyses with regard to the parameters identified as significant (e.g., for water temperature) is at least partially due to the fact that these analyses referred to different time ranges. Furthermore, some of the parameters were found to be highly correlated (e.g., water temperature with CPUE, nitrate, and phosphate), which means that, in the multivariate models, certain parameters might replace others in order to explain variation in prevalence.

For interpretation of the results of the analyses, it has to be taken into account that a considerable part of the data used was derived from data interpolation, which is always problematic since one does not know how the parameters have behaved in between two sampling dates. With increasing distance between sampling dates, this problem becomes more serious and it is self-evident that the reliability of results from any statistical analysis suffers from this. Furthermore, automatic interpolation without adaptation to the type of data used might lead to biased data, e.g., if data series characterized by pronounced seasonal variation are not consistent in terms of temporal coverage of data points. An example from the existing ICES data set is the measurement of water temperature (Figure A5.5a). For the period 1981–1991, only winter data are available and interpolated values used for the analysis were generally low. From 1991, however, spring/summer values are also available and, therefore, interpolated values were considerably higher although winter values remained at about the same level as in previous years. This example clearly demonstrates that, for further analyses, it will be more appropriate to employ specific parameter-oriented interpolation techniques.

4 CONCLUSIONS AND PERSPECTIVES

First of all, with regard to the activities of the WGPDMO during previous years, the data overview reveals that the fish disease data constitute one of the most comprehensive and consistent data sets in the ICES Environmental Data Centre, in terms of both spatial and temporal coverage.

In contrast, there is a striking lack of other data, in particular for contaminants in biota and sediments, creating major problems in the initiated holistic data analysis. However, there are undoubtedly more data around held by national data banks. Therefore, ICES Member Countries should be encouraged to submit these data to the ICES Environmental Data Centre by using standard procedures already developed by ICES and, therefore, applicable without major effort.

The availability of additional data would improve the spatial and temporal data coverage and would, therefore, possibly facilitate an analysis based on smaller geographical areas than those used in the present case study. Large areas create problems since conditions normally are not the same over the entire area. For instance, Area 1 (extended German Bight) includes coastal, estuarine, and offshore areas and a comparison of, e.g., contaminant levels in mussels from estuarine areas with disease prevalences of dab collected offshore can only be regarded as a rough approach to identify possible relationships. Furthermore, additional data could minimize the need for temporal interpolation and related problems (see above).

Despite the shortcomings identified, the results of the analysis seem to be promising since, for a number of parameters included, a close relationship with variation in disease prevalence could be identified. It is, therefore, concluded that further activities with respect to enhancing the data basis and improving the models applied and the methods for statistical analysis are desirable and should be initiated.

5 ACKNOWLEDGEMENT

This report was prepared by Dr T. Lang, Bundesforschungsanstalt für Fischerei, Institut für Fischereiökologie, Cuxhaven, Germany, and W. Wosniok, Universität Bremen, Institut für Statistik, Bremen, Germany.

6 REFERENCES

ICES. 1999. Statistical analysis of fish disease prevalence data from the ICES Environmental Data Centre. *In* Report of the ICES Advisory Committee on the Marine Environment, 1998. ICES Cooperative Research Report, 233: 297–327.

Abbreviations used in the figures and tables, if not further specified

CPUE	catch per unit effort (number of fish per hour of trawling)
DIS OXY	dissolved oxygen
F1	females, size group 10–14 cm
F2	females, size group 15–19 cm
F3	females, size group 20–24 cm
F4	females, size group ≥ 25 cm
F20	fraction < 20 μm
F63	fraction < 63 μm
LI	liver tissue
LIMA	*Limanda limanda*
M1	males, size group 10–14 cm
M2	males, size group 15–19 cm
M3	males, size group 20–24 cm
M4	males, size group ≥ 25 cm
MU	muscle tissue
MYTI	*Mytilus edulis*
P TOTAL	total phosphorus
SB	soft body tissue
SED	sediment
U00	unfractionated
WAT	water
WAT A	filtered sea water
WAT B	unfiltered sea water

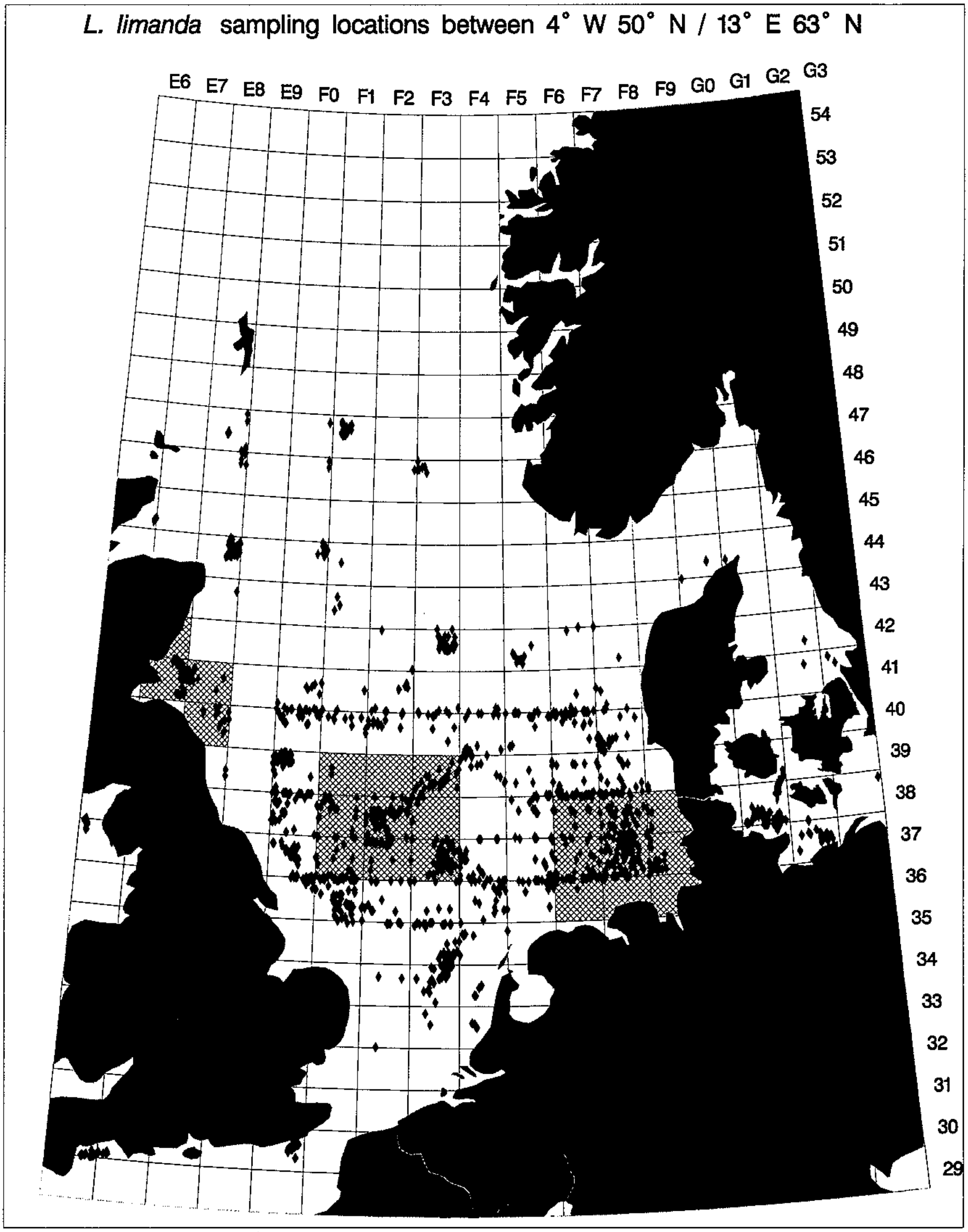
Table A5.1. Area 1 (extended German Bight): case study on the relationship between the prevalence of dab (*L. limanda*) diseases and parameters available from the ICES data banks, results of a univariate analysis (if $p < 0.05$, a significant relationship existed). n_o: number of observations; n_i: number of interpolations; dir: direction of relationship; p: significance level.

Parameter	n_o	n_i	Lymphocystis		Epidermal hyperplasia/ papilloma		Acute/healing ulcerations	
			dir	p	dir	p	dir	p
W TEMPERATURE	261	98	+	< 0.0001	+	< 0.0001	+	< 0.0001
W SALINITY	239	98	+	< 0.0001	+	< 0.0001	-	0.8911
W PHOSPHATE	48	71	+	0.1688	-	0.0638	-	0.0418
W DIS OXY	20	29	+	0.0013	+	0.1358	-	0.0733
W NITRATE	48	71	+	0.0362	-	0.7437	-	0.0832
WAT B CB153	120	43	+	0.2149	-	0.7491	+	0.9132
WAT B HCB	114	43	-	0.6210	-	0.7535	-	< 0.0001
WAT B Cd	185	52	-	0.5397	+	0.2417	-	< 0.0001
WAT B Hg	186	52	-	0.0066	-	0.0051	-	0.0555
SED F63 CB153	5	42	+	0.5098	-	0.0098	+	0.2030
SED F63 HCB	3	27	-	0.0006	-	0.0534	+	0.0317
SED F63 Cd	7	34	+	0.2290	+	0.7014	+	0.9162
SED F63 Hg	8	42	+	0.0114	-	0.4048	+	0.8843
MYTI SB CB153	36	96	+	< 0.0001	+	< 0.0001	+	0.4567
MYTI SB HCB	33	39	-	0.0183	-	0.0544	-	0.0128
MYTI SB Cd	50	96	+	0.1833	+	0.1119	-	0.0238
MYTI SB Hg	51	96	+	< 0.0001	+	< 0.0001	-	0.1370
LIMA LI CB153	4	13	+	0.0192	-	0.4267	+	0.9582
LIMA LI HCB	3	5	+	0.1218	-	0.7319	+	0.6986
LIMA LI Cd	5	28	+	0.0235	+	0.8689	-	0.0115
LIMA MU Hg	5	28	+	0.0019	-	0.8375	-	0.0064
CPUE	295	97	+	< 0.0001	+	< 0.0001	-	0.0060

Table A5.2. Area 1 (extended German Bight): case study on the relationship between the prevalence of dab (*L. limanda*) diseases and parameters available from the ICES data banks, results of a multivariate analysis. n_i= number of interpolations (data values) remaining for the analysis; dir = direction of relationship; p = significance level.

Long-term model	**Parameters included in the model**		
	W TEMPERATURE, W SALINITY, W PHOSPHATE , W NITRATE, WAT B CB153, WAT B HCB, WAT B Cd, WAT B Hg, MYTI SB CB153, MYTI SB Cd, MYTI SB Hg, CPUE		
n_i	43		
Disease	**Significant terms**	**dir**	**p**
Lymphocystis	W SALINITY MYTI SB Cd	- -	< 0.0001 < 0.0001
Epidermal hyperplasia/ papilloma	W NITRATE WAT B Cd MYTI SB CB153 MYTI SB Cd	+ + - -	0.0003 < 0.0001 < 0.0001 < 0.0001
Acute/healing ulcerations	WAT B Cd CPUE	- -	0.0008 0.0163
Medium-term model	**Parameters included in the model**		
	W TEMPERATURE, W SALINITY, W PHOSPHATE , W NITRATE, WAT B CB153, WAT B HCB, WAT B Cd, WAT B Hg, SED F63 CB153, SED F63 Hg, MYTI SB CB153, MYTI SB Cd, MYTI SB Hg, CPUE		
n_i	33		
Disease	**Significant terms**	**dir**	**p**
Lymphocystis	W SALINITY MYTI SB Cd CPUE	- - -	< 0.0001 < 0.0001 0.0213
Epidermal hyperplasia/ papilloma	W NITRATE WAT B CB153 SED F63 CB153 MYTI SB CB153	+ - - +	< 0.0001 0.0089 < 0.0001 0.0004
Acute/healing ulcerations	CPUE	-	0.0016
Short-term model	**Parameters included in the model**		
	W TEMPERATURE, W SALINITY, W PHOSPHATE , W NITRATE, WAT B CB153, WAT B HCB, WAT B Cd, WAT B Hg, SED F63 CB153, SED F63 Hg, MYTI SB CB153, MYTI SB HCB, MYTI SB Cd, MYTI SB Hg, LIMA LI Cd, LIMA MU Hg, CPUE		
n_i	23		
Disease	**Significant terms**	**dir**	**p**
Lymphocystis	W NITRATE MYTI SB CB153 LIMA LI Cd	+ + -	0.0093 < 0.0001 0.0070
Epidermal hyperplasia/ papilloma	W NITRATE WAT B Cd	+ +	0.0005 0.0288
Acute/healing ulcerations	WAT B Cd	-	0.0070

Figure A5.1. Location of the three North Sea areas for which an overview of available ICES data is presented (single marks indicate positions for which prevalence data of dab (*L. limanda*) diseases are available).



Figre A5.2. Area 1 (extended German Bight): overview of ICES data available for a holistic analysis of fish disease data (extract).

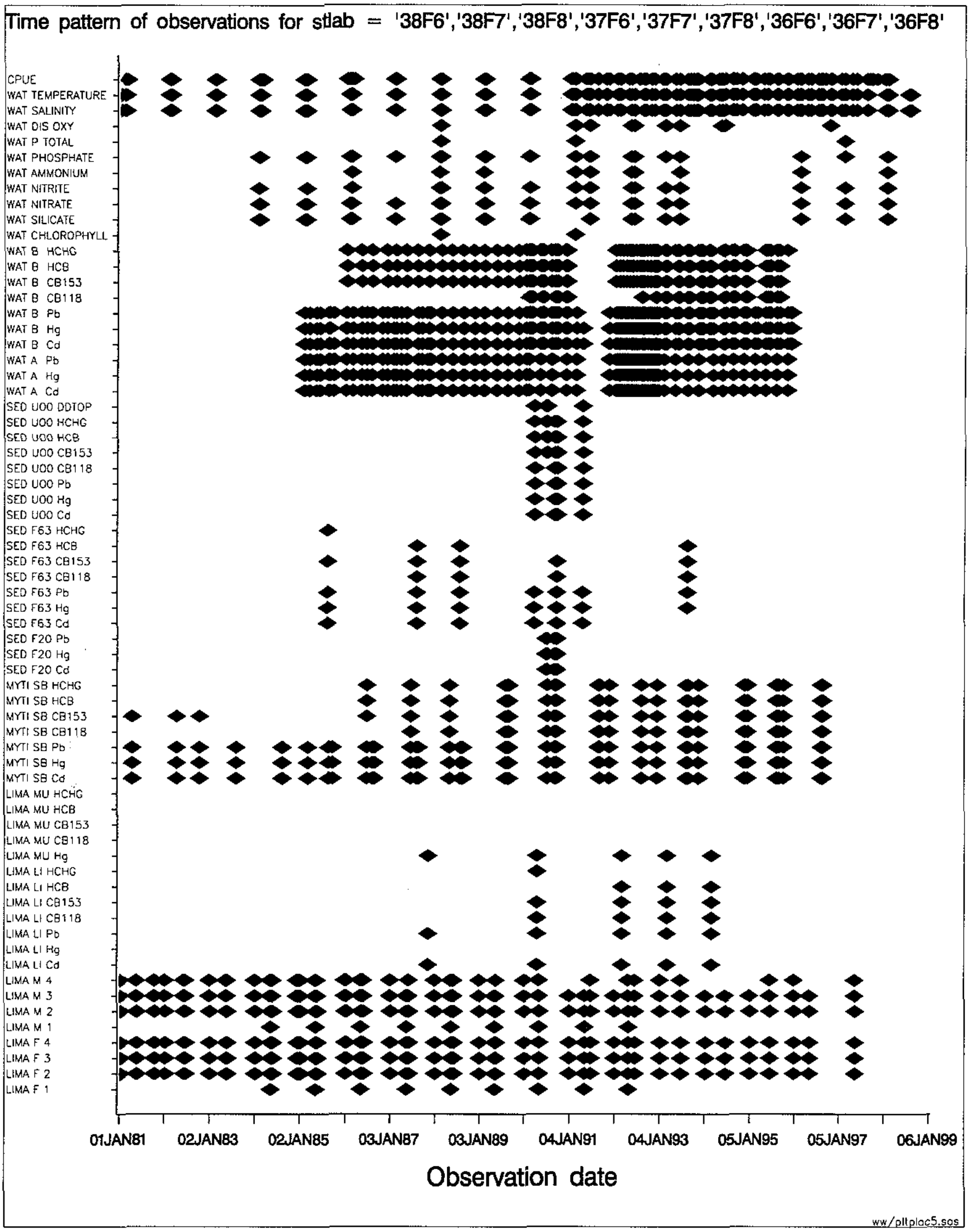


Figure A5.3. Area 2 (extended Dogger Bank): overview of ICES data available for a holistic analysis of fish disease data (extract).

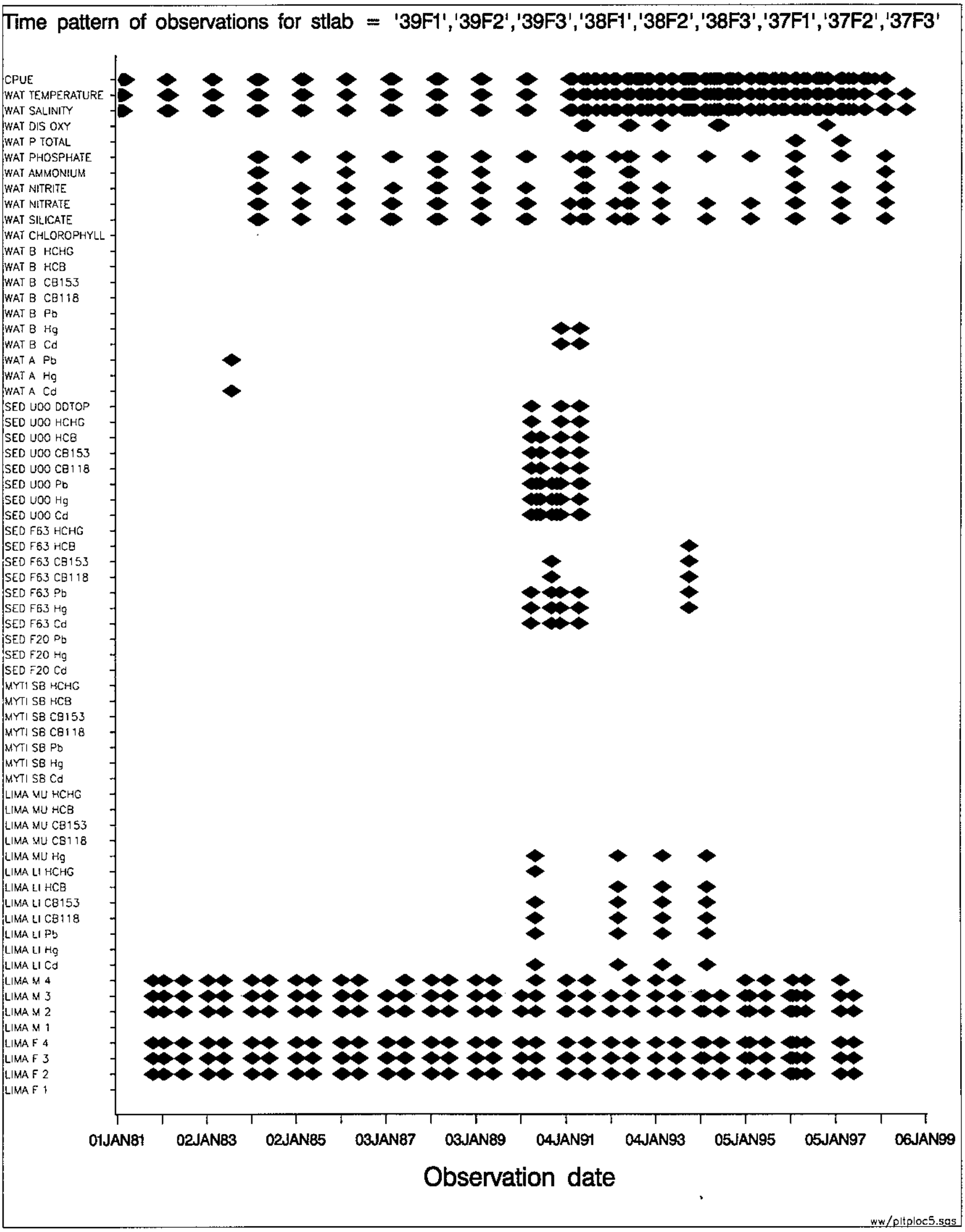


Figure A5.4. Area 3 (extended Firth of Forth): overview of ICES data available for a holistic analysis of fish disease data (extract).

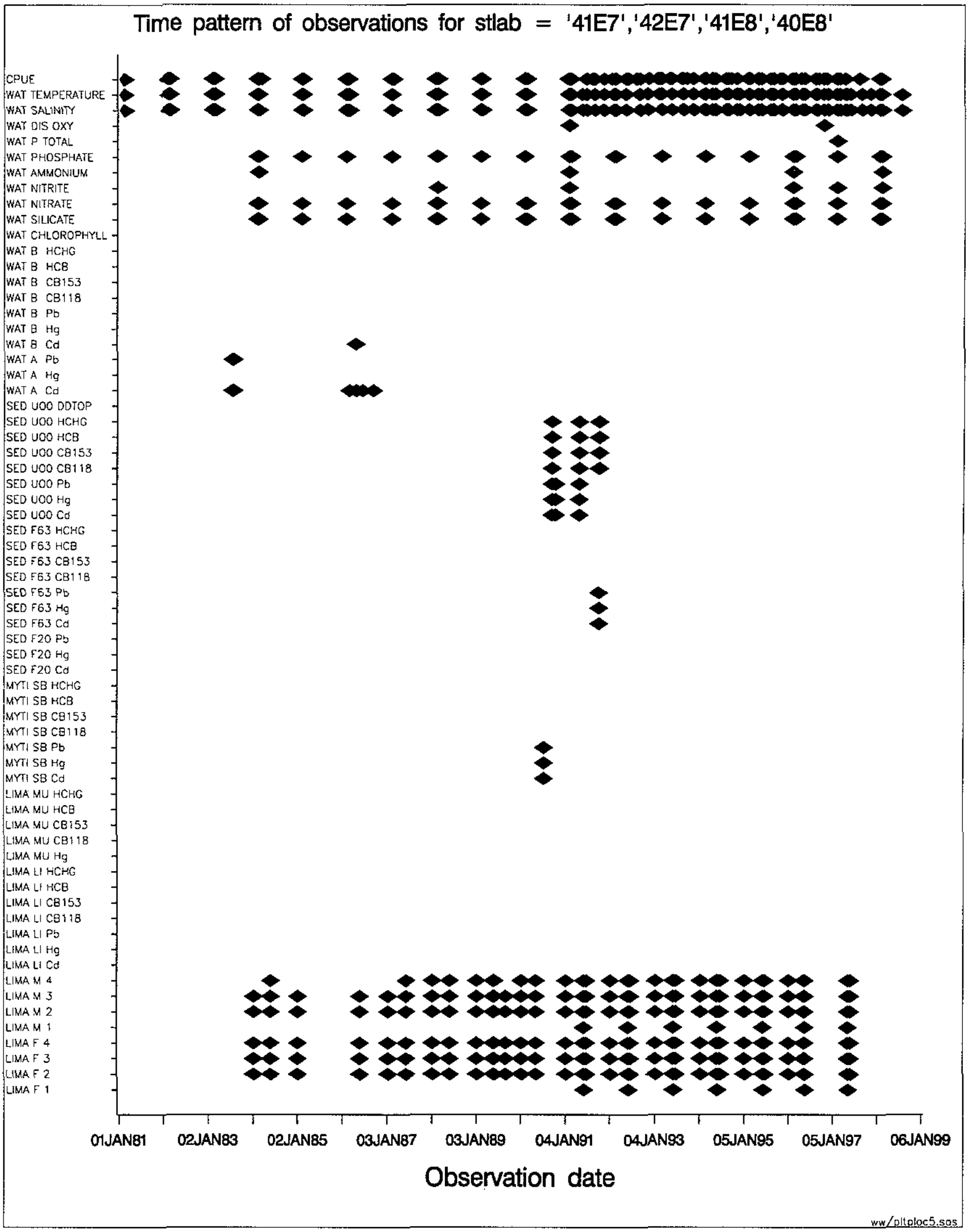


Figure A5.5.a. Area 1 (German Bight, extended): water data from the ICES Oceanographic Data Centre (extract) (filled circles: empirical data, empty circles: interpolated values).

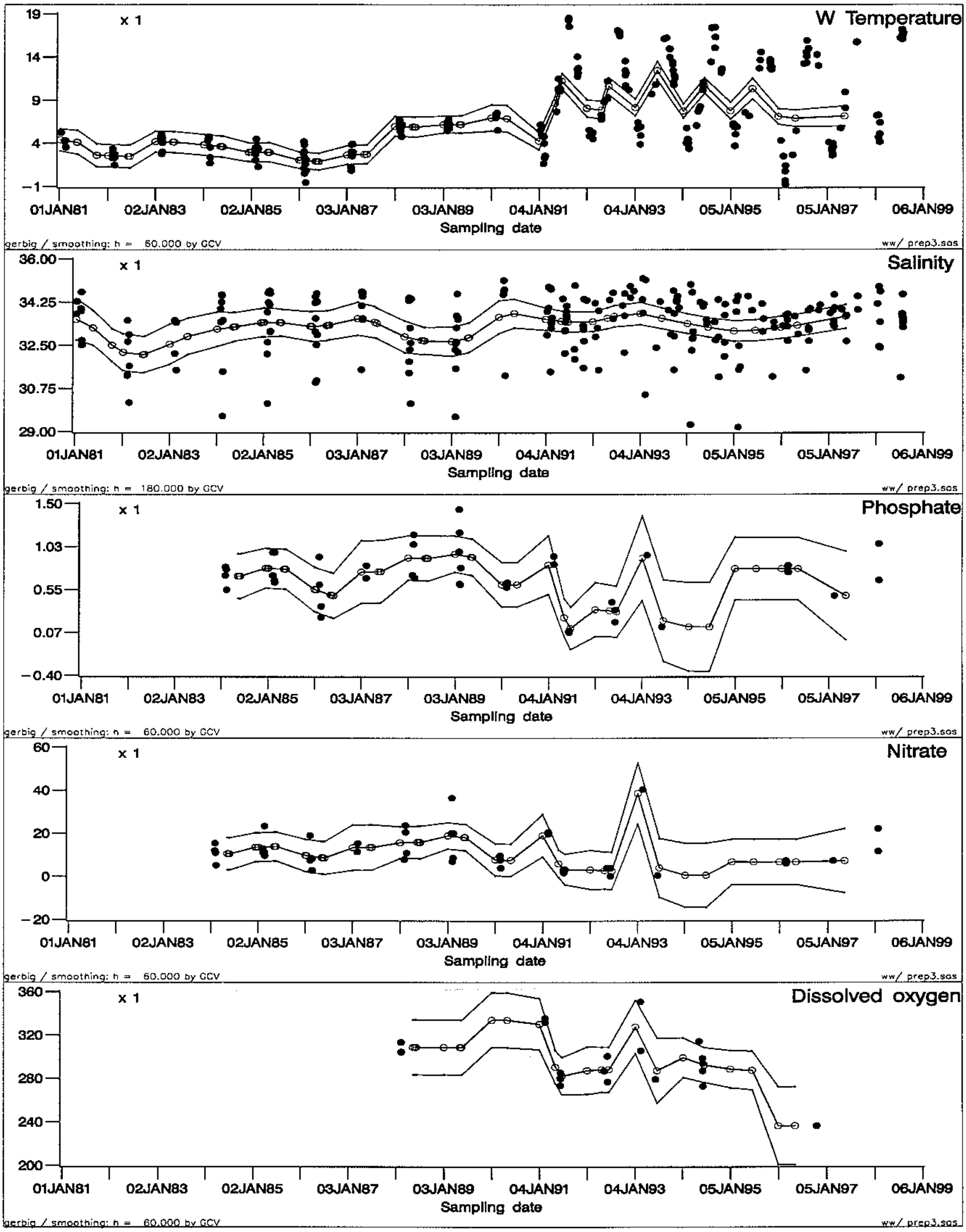


Figure A5.5.b. Area 1 (extended German Bight): data on contaminants in sea water (unfiltered) from the ICES Environmental Data Centre (extract) (filled circles: empirical data; empty circles: interpolated values).

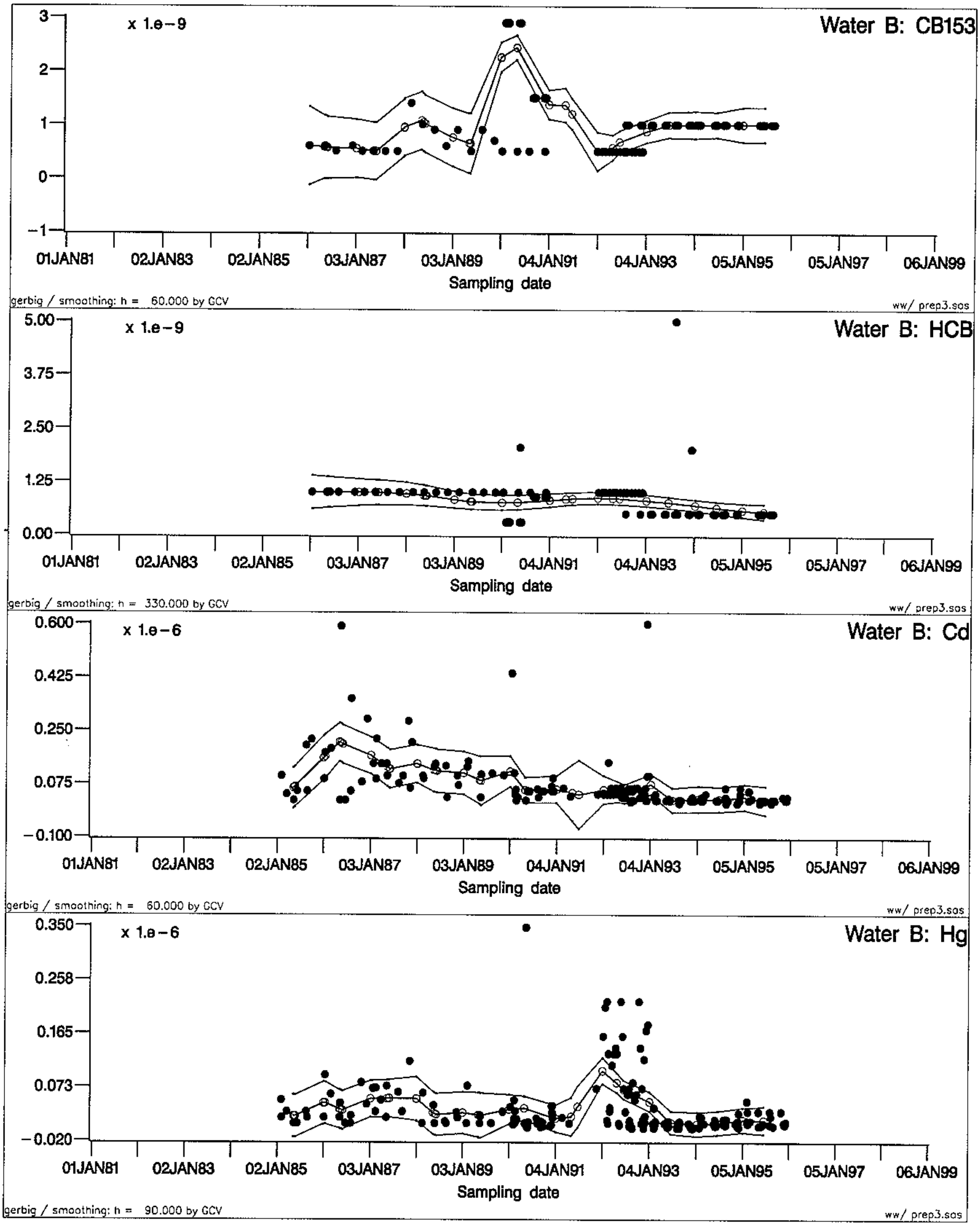


Figure A5.5.c. Area 1 (extended German Bight): data on contaminants in sediments (fraction < 63µm) from the ICES Environmental Data Centre (extract) (filled circles: empirical data; empty circles: interpolated values).

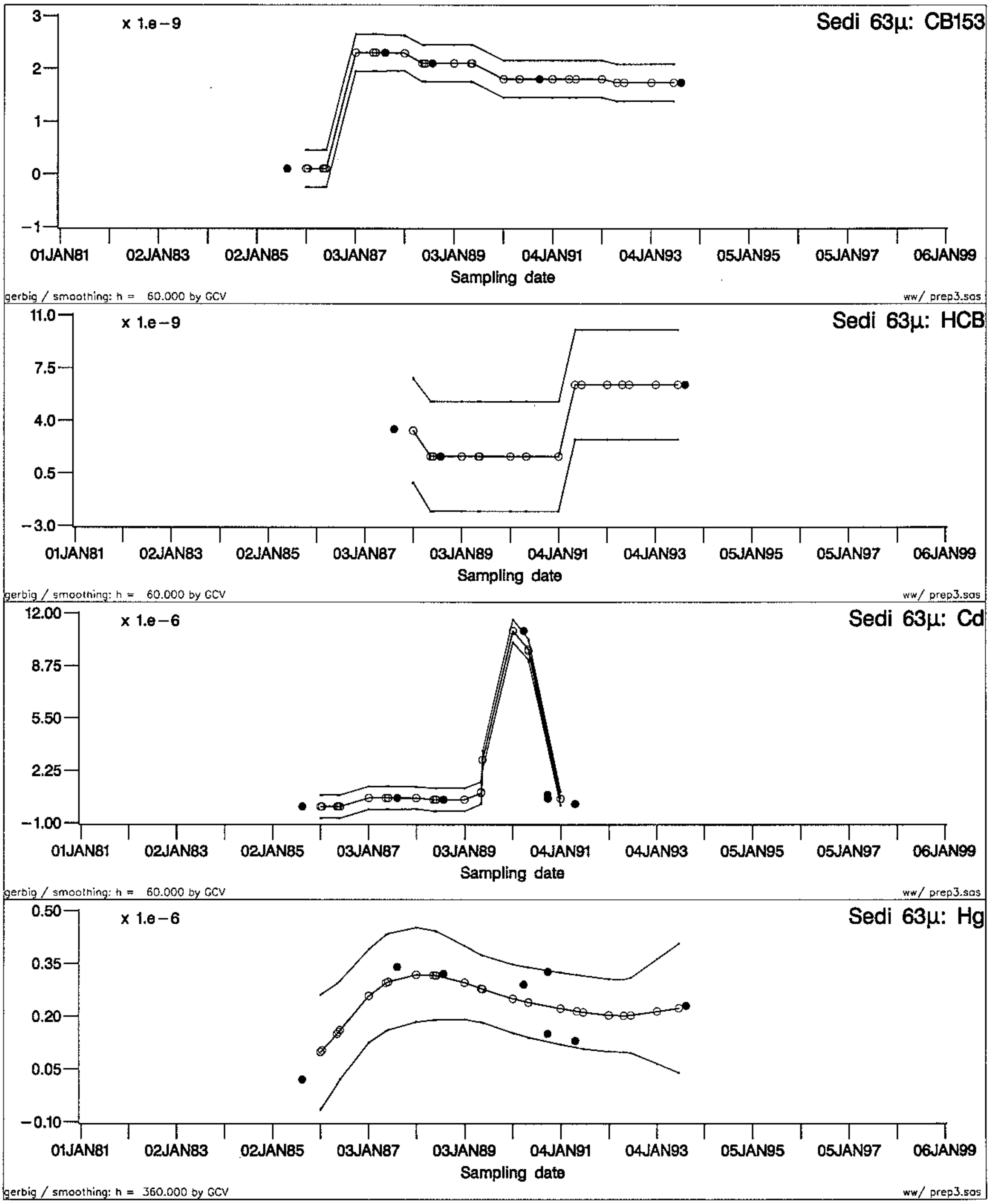


Figure A5.5.d. Area 1 (extended German Bight): data on contaminants in *Mytilus edulis* (soft body tissue) from the ICES Environmental Data Centre (extract) (filled circles: empirical data; empty circles: interpolated values).

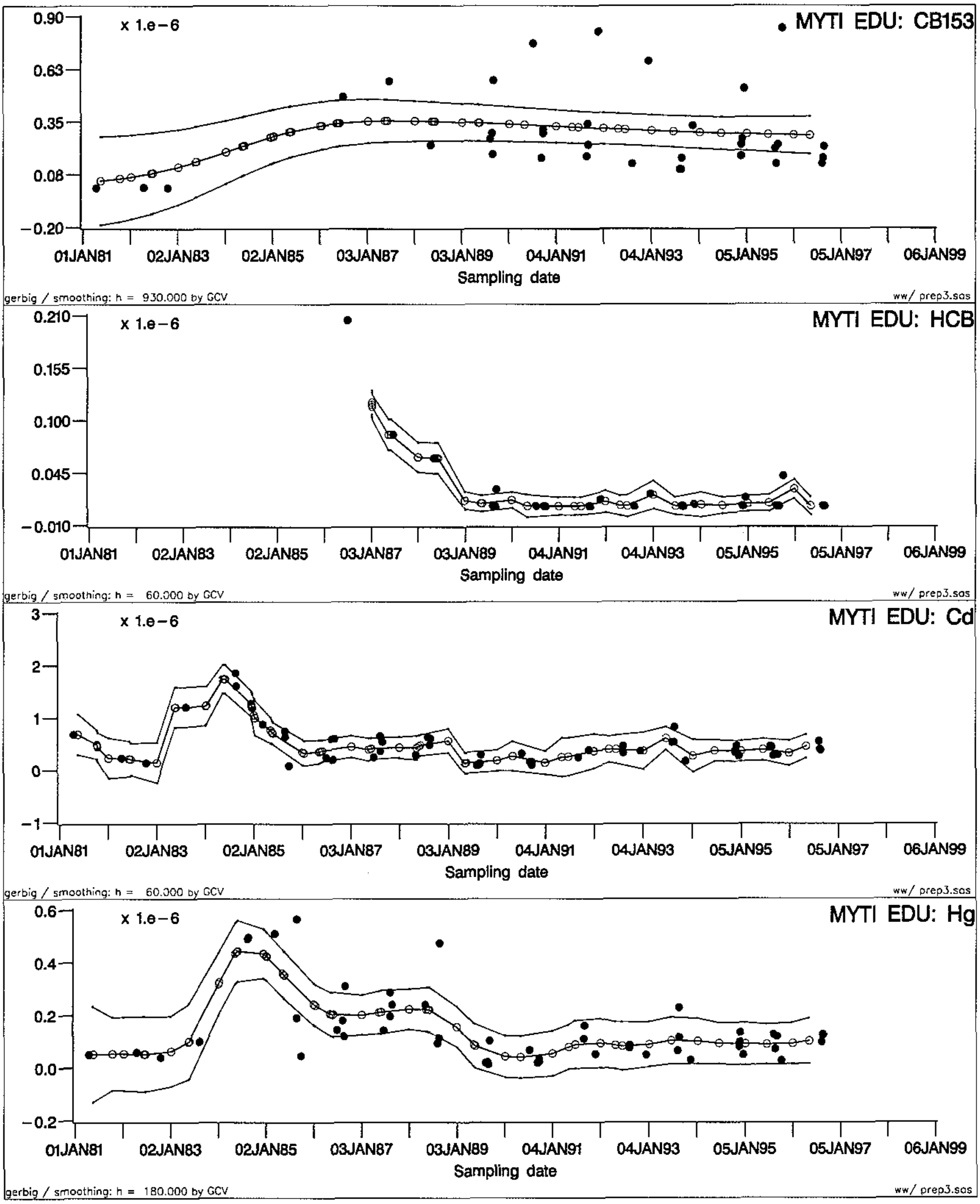


Figure A5.5.e. Area 1 (extended German Bight): data on contaminants in liver and muscle of dab (*Limanda limanda*) from the ICES Environmental Data Centre (extract) (filled circles: empirical data; empty circles: interpolated values).

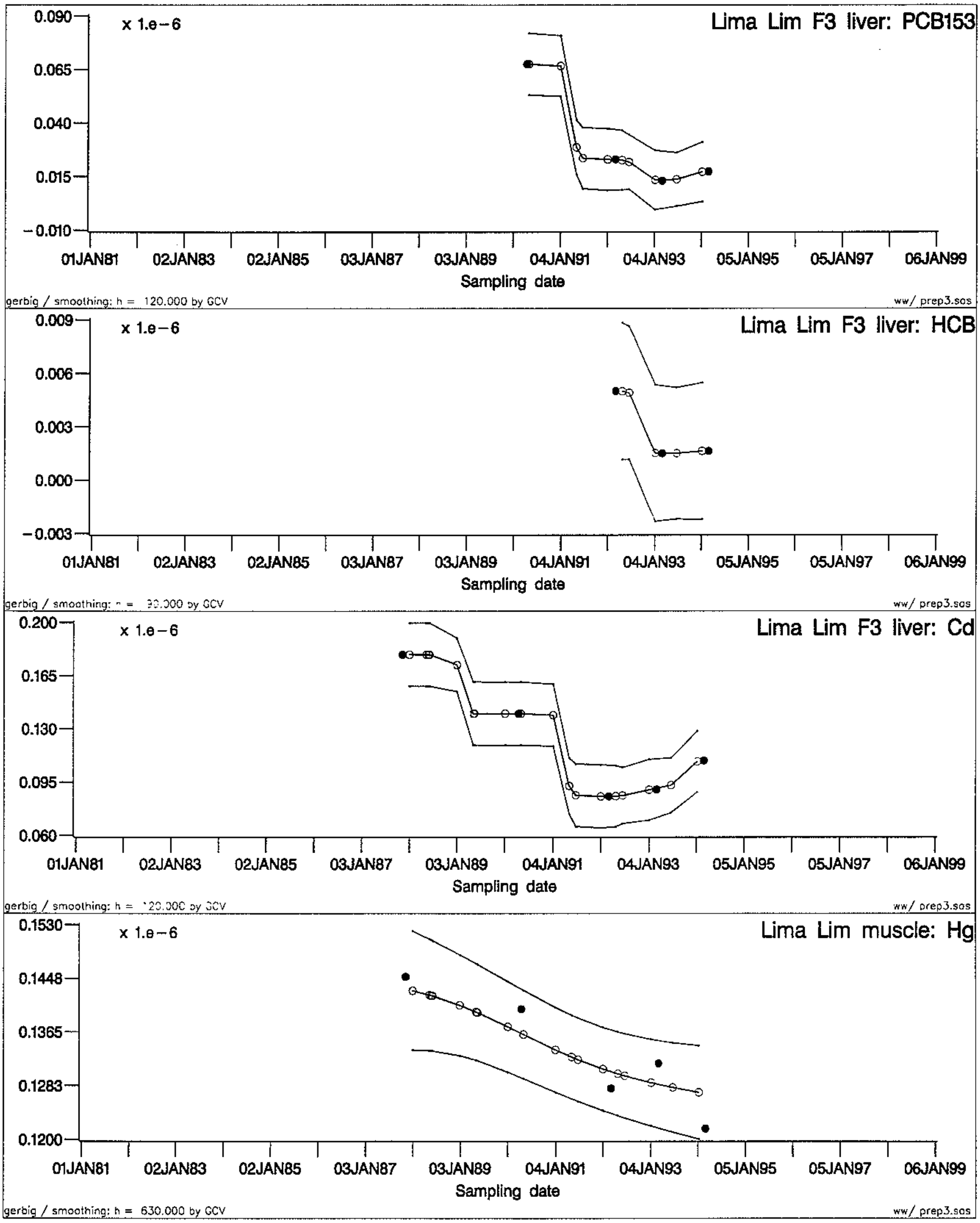


Figure A5.5.f. Area 1 (extended German Bight): *Limanda limanda*, data on catch per unit effort (CPUE) (all dab) from the ICES Fishery Data Bank and on diseases (females, 20–24 cm) from the ICES Environmental Data Centre (extract) (filled circles: empirical data; empty circles: interpolated values).