ANNEX 8

STATISTICAL ANALYSIS OF FISH DISEASE PREVALENCE DATA FROM THE ICES ENVIRONMENTAL DATA CENTRE

1 INTRODUCTION

Data on disease prevalence in wild marine fish stocks form a major component of the ICES Environmental Data Centre. ICES Member Countries conducting fish disease surveys as part of their national environmental monitoring programmes submit the data to ICES on a regular basis. The surveys are carried out according to standardized ICES methodologies established through the work of the ICES Working Group on Pathology and Diseases of Marine Organisms (WGPDMO) (ICES, 1989; Bucke *et al.*, 1996).

The ICES fish disease data comprise information from studies on the occurrence of externally visible diseases and macroscopic liver nodules/tumours in the common dab (*Limanda limanda*) and the European flounder (*Platichthys flesus*) from the North Sea and adjacent waters (including the Baltic Sea, Irish Sea, and the English Channel). In total, data on the length, sex, and health status of more than 420,000 individual specimens, partly dating back to 1981, have been submitted to ICES so far, as well as information on sampling characteristics (dates, locations, gear types, etc.). Therefore, the ICES fish disease data bank can be considered a unique set of data providing information on biological responses of marine organisms to, *inter alia*, environmental change.

In 1992, the Sub-Group on Statistical Analysis of Fish Disease Data in Marine Stocks (SGFDDS) was established. The Sub-Group existed until 1996 and was then replaced by a Study Group with the same name, which met once in 1997. The major tasks of the Sub-Group/Study Group were:

- to establish standardized procedures for the submission of fish disease data to the ICES Environmental Data Centre (the former ICES Environmental Databank) by ICES Member Countries; and
- to develop methods for and carry out a statistical analysis of the fish disease prevalence data submitted to the ICES Environmental Data Centre.

Disease prevalence is defined as the ratio between the number of diseased fish and the number of examined fish in a sample. This ratio and hence the prevalence is an empirical probability, which is thought to be a realization of an underlying stochastic process, whose probabilistic properties, including its relation to explaining quantities, must be estimated from observed data. Such an analysis needs specific statistical methods, which account for the particular distributional properties of prevalences.

At the 1997 meeting of WGPDMO, a progress report was presented on the activities of the SGFDDS, and its tasks were considered fulfilled. Through the work of the SGFDDS carried out in collaboration with the ICES Secretariat, the submission of fish disease data to ICES has been standardized via the implementation of the ICES Fish Disease Data Reporting Format (part of the ICES Environmental Data Reporting Formats), the ICES Fish Disease Data Entry Program, and procedures for validation of the data submitted. Statistical methods for analysing spatial and temporal trends in the disease prevalence have been elaborated and applied successfully using the data available prior to the 1997 meeting of the SGFDDS.

Recognizing that the data were still incomplete at that time, it was recommended that Member Countries submit additional fish disease data (historic and current data) to the ICES Environmental Data Centre and that a more comprehensive statistical analysis be carried out intersessionally prior to the 1998 WGPDMO meeting by selected experts using the complete data set. All data that were included in this comprehensive analysis had been collected following the ICES standard methodologies and were subjected to the same validation process before being entered into the ICES data bank.

The present report contains the main results of this statistical analysis. It highlights the methodologies applied and provides information on spatial and temporal trends with respect to the prevalence of the major diseases of dab and flounder in the North Sea and Baltic Sea.

The value of the ICES fish disease data as well as of the results of the analysis will increase in the future, since studies on externally visible diseases and liver nodules/tumours of dab are among the techniques designated for the biological effects component of the new OSPAR Joint Assessment and Monitoring Programme (JAMP) and since ICES serves as the data centre for the OSPAR Commission's monitoring data and will provide environmental data to be incorporated in the OSPAR Quality Status Report 2000. Furthermore, the status of the completed fish disease data bank and the establishment of methodologies for data submission, validation, and analysis will facilitate a more holistic approach for the future analysis of environmental data, combining fish disease data with other types of biological effects data, oceanographic data, and fisheries data held in the ICES data banks.

2 DATA AVAILABLE FOR THE STATISTICAL ANALYSIS

The present analysis was carried out on data from the ICES fish disease data bank provided by the ICES Secretariat on 27 January 1998. At that time, data on a total of 424,998 specimens of fish were available, most of them on dab (*Limanda limanda*, n = 399,262) and some on flounder (*Platichthys flesus*, n = 25,736). The data had been submitted to ICES by various laboratories (see Table A8.1) and covered the time period from 1981 to 1997. The data were not uniformly distributed over time, neither over years nor over months or seasons within years. Table A8.2 shows the distribution of dab data over calendar years and months within the year. Table A8.3 contains the corresponding figures for flounder. Data on flounder refer in most cases to samples taken in the last quarter (October to December) of a year, while for dab usually more than one quarter of a year is covered by sampling.

Table A8.1. Number of dab (*Limanda limanda*) and flounder (*Platichthys flesus*) for which disease data were reported, by laboratory.

Reporting laboratory	Number of fish reported		
	Dab	Flounder	Total
ALUK	28,476	0	28,476
BFCG	274,951	7,465	282,416
DFHU	62,911	0	62,911
DGWN	10,631	0	10,631
DOUK	16,290	0	16,290
RIVO	6,003	18,271	24,274
Total	399,262	25,736	424,998

Reporting laboratories:

ALUK = Fisheries Research Services, the Marine Laboratory, Aberdeen, UK

BFCG = Federal Research Centre for Fishery, Institute of Fishery Ecology, Cuxhaven, Germany

DFHU = Danish Institute for Fisheries Research, Charlottenlund, Denmark

DGWN = National Institute for Coastal and Marine Management, Ecotoxicology Section, Middelburg, The Netherlands

DOUK = The Centre for Environment, Fisheries and Aquaculture Science, Fish Disease Laboratory, Weymouth, UK

RIVO = Netherlands Institute for Fishery Investigation, IJmuiden, The Netherlands

The geographical spread of sampling locations, as well as the frequency with which locations were visited, vary each year. Tables A8.4 and A8.5 summarize for dab and flounder, respectively, the geographical and temporal spread of sampling in terms of the number of visits per ICES statistical rectangle and year. Figure A8.1 provides an overview of rectangles for which disease reports for dab and flounder are available in the ICES Environmental Data Centre. The number of rectangles which have been visited at least once is relatively high (123 for dab, 29 for flounder, five area designations were excluded because of implausible geographical details). However, not all rectangles were visited so frequently that the resulting time series of disease prevalence seems informative with respect to the investigation of temporal trends.

The data bank is organized in records, each of which refers in principle to an individual fish. Observations referring to fish specimens with the same host-specific attributes (species, gender, size in length class), the same location of observation (expressed as ICES rectangle), the same observation time (expressed by the sampling date), the same haul and the same reporting laboratory are summarized in one record, stating the number of specimens to which this record refers ('number examined').

The disease information in one record comprises, among others, the prevalence of lymphocystis, epidermal hyperplasia/papilloma, acute or healing skin ulcerations, skeletal deformities, and liver nodules/tumours. Most records also contain, depending on the reporting laboratory, information about additional diseases which were examined only occasionally. Prevalence information is expressed as the number of specimens examined (one variable per record) in relation to the number of fish infected. A missing value code for the number infected is used to express that the fish was not examined for the respective disease, while a zero indicates that no fish was affected. It is important that this distinction be maintained in future data reporting and administration within the ICES data bank.

Tables A8.6 and A8.7 contain an overview of the number of fish examined and affected; however, the number examined shown is the sum taken over all records referring to the respective ICES rectangle. As not all fish were inspected for all diseases, it is generally not possible to calculate the prevalence from these tables, though for the main diseases (lymphocystis, epidermal hyperplasia/papilloma, ulcerations), which were considered in nearly all samples, the error would be negligible.

3 AIM OF ANALYSIS

The aim of the present statistical analysis is the identification of location-specific temporal trends in disease prevalence and, as far as such exist, the comparison of such trends found at different geographical locations.

A temporal trend is generally understood to be a long-term change in a target quantity (here, the disease prevalence). Baggelaar *et al.* define a trend to be a '(semi-)permanent change in the location (mean or median) of a process over at least several years. It does

Table A8.2. Number of dab (*Limanda limanda*) examined, by year and month of sampling.

Year	January	February	March	April	May	June	July	September	October	December	Total
1981	4,809	0	0	0	7,401	0	0	0	13,380	0	25,590
1982	18,040	0	0	0	0	18,886	0	0	0	1,385	38,311
1983	5,579	0	0	0	15,716	0	0	0	0	10	21,305
1984	3,731	0	348	0	19,499	0	0	0	0	4,172	27,750
1985	6,408	0	0	0	12,100	0	0	0	0	144	18,652
1986	10,429	0	924	0	18,804	3,119	0	0	0	1,501	34,777
1987	8,183	0	1,321	0	4,888	7,607	0	0	0	119	22,118
1988	9,066	0	669	0	8,356	3,366	0	0	0	534	21,991
1989	4,145	0	0	959	9,379	2,043	0	83	0	0	16,609
1990	6,689	0	0	1,322	11,204	0	0	0	0	520	19,735
1991	9,408	837	0	1,145	6,584	8,770	1,383	0	387	673	29,187
1992	8,551	0	1,234	7,297	2,375	11,164	1,972	0	0	0	32,593
1993	5,088	879	1,057	0	4,911	8,161	1,943	0	0	1,237	23,276
1994	6,453	1,051	720	877	1,353	3,454	1,731	0	0	0	15,639
1995	3,595	5,399	0	0	0	7,084	1,665	0	0	729	18,472
1996	6,548	4,083	0	0	9,840	0	0	0	0	529	21,000
1997	0	4,322	0	0	5,889	2,046	0	0	0	0	12,257
Total	116,722	16,571	6,273	11,600	138,299	75,700	8,694	83	13,767	11,553	399,262

Table A8.3. Number of flounder (*Platichthys flesus*) examined, by year and month of sampling.

Year	Sep	Oct	Dec	Total
1983	1,336	0	121	1,457
1984	2,574	0	0	2,574
1985	1,089	0	97	1,186
1986	1,570	0	150	1,720
1987	1,188	0	615	1,803
1988	692	0	590	1,282
1989	1,265	0	0	1,265
1990	0	0	295	295
1991	0	1,517	1,365	2,882
1992	0	1,318	0	1,318
1993	0	1,305	1,568	2,873
1994	0	1,604	0	1,604
1995	0	1,402	1,472	2,874
1996	0	1,411	1,192	2,603
Total	9,714	8,557	7,465	25,736

not comprise changes related to seasonal cycles, or sudden and short-lived changes, caused by calamities' (ICES, 1997a). Though these authors discuss trend detection methods mainly for concentration or load measurements, not for prevalence, their definition can be used here as well. The analysis methods must, however, be adapted for the present problem.

4 DATA SUBSET FOR ANALYSIS

From the definition of a trend, it follows that only time series of a certain minimum length can be used for trend identification. In an earlier analysis, the Study Group on Statistical Analysis of Fish Disease Data in Marine Stocks (ICES, 1997b) decided, for dab data, to consider for analysis only those data series which contained on average at least one observation (sampling) within every two years over the reporting period (1981–1996) present in the 1997 database. Using a corresponding criterion for the selection of usable time series from the 1998 database leads to the requirement that a time series within one ICES rectangle should contain at least nine samplings to be used for statistical analysis. Table A8.4 shows that this criterion is fulfilled for the prevalence series in 42 ICES rectangles. However, the series for two of these rectangles (30F0 and 37F8) extend over only six years, which was considered as too short in view of the requirements of the trend detection technique (see below), so that these rectangles were excluded from further analysis. Furthermore, locations with fewer than four samples from the 20–24 cm size group were also excluded from the analysis in order to ensure a sufficiently accurate trend estimation. Combining all criteria resulted in 32 ICES rectangles for trend analysis. In Tables A8.4 and A8.5, samplings were counted on a

Figure A8.1. Map showing the ICES statistical rectangles from which fish samples were taken for disease investigation. Disease reports for dab (*Limanda limanda*) are indicated by circles; disease reports for flounder (*Platichthys flesus*) are denoted by plus signs.

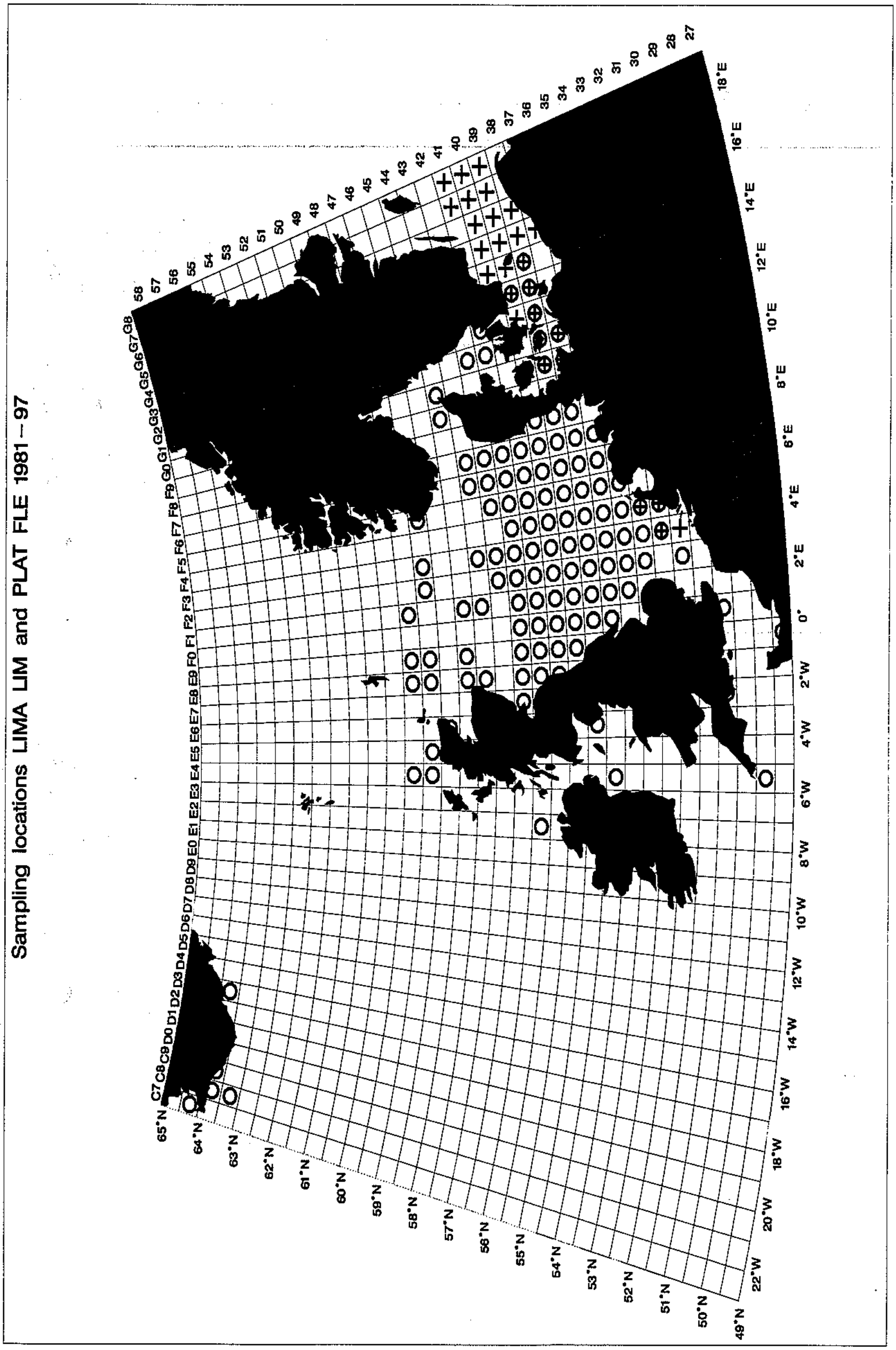


Table A8.4. Number of reports per ICES rectangle for dab (*Limanda limanda*). Hauls are combined; reports for different days are counted individually.

No.	ICES rect.	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	Total
1	37F7	12	5	3	8	4	5	3	6	4	4	3	3	2	2	2	2	1	69
2	38F7	6	3	2	7	5	3	4	3	1	1	1	1						37
3	40F7				2	2	3	3	3	4	4	2	4	2	2	2	2	1	36
4	41E7									4	2	4	5	5	3	4	4	3	34
5	38F2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	33
6	36F1			1	3	2	2	3	2			2	1	1	1	4	3	2	27
7	39F7	1	2	2	5	1	3	4	3	1	1	1	1						25
8	37F2	2	3			1	1		1	1	2	2	2	2	2	2	2	1	24
9	37F6	4	4	2	4	2	2	2	1		1	1	1						24
10	39F6	1	3	1	6	2	3	2	2	1	1	1	1						24
11	35F3		1		1		3	1	1	2	2	1	1		2	3	3	2	23
12	37F0	1	1	1	2	3		3	2			2		1	1	2	1	1	21
13	34F3	1	2				2	2	1	1	1	1	1	1	2	2	2	1	20
14	36F4				1		4	2	2	1	1	1	1	1		3	3		20
15	37F3					1	1	2	2	3	2	1	2	1	2	1	2		20
16	41F7				2	2	2	4	2	2	2	2	1	1					20
17	39E9				1		2		1	2	2	2	2	2		2	2	1	19
18	41E8							1	1	2	2	2	2	2	1	2	2	1	18
19	37F4						2	2	2	3	1	1	2	1	1		2		17
20	37F5	4	2	2	1	2	1					1	1			1	2		17
21	38G0		3	1		2	3	1	1		1	1		2		1	1		17
22	39F3		2	2	3	2	2	2	1			1	1				1		17
23	44E8									2	2	2	2	2	2	2	2	1	17
24	37F1	2	4	2	2	2	2	1	1										16
25	41F6		1	1	1	2	2	3	1	1	1	1	1	1					16
26	38F6	2	2	2	2	1	2	1			1								13
27	40F6			2	2	2	2	2	1					1	1				13
28	44F9				1	2	1	1	2		2	1	1	1					12
29	36F2		2	2	2	1	1	2	1										11
30	39F0		1	2	1	1			1	1	1	1				1	1		11
31	39F5	1	4	1	1	1	2	1											11
32	37F8	4	1	1	2	1	1												10
33	38F0	2		1		1	2	2	1			1							10
34	38F1	3	3				1					1	1					1	10
35	41G1				1	1	1	1	1	1	1	1	1	1					10
36	42F3									1	1	2	2	1	1		1	1	10
37	42G1				1	1	1	1	1	1	1	1	1	1					10
38	30F0												1	1	2	1	2	2	9
39	40F4		1		1	1	2	1			1	1	1						9
40	41G2					1	1	1	1	1	1	1	1	1					9
41	44G0				1	1	1	1	1		1	1	1	1					9
42	47E8									2	1		1	1	1	1	1	1	9
43	38F5	1	1	2	1		2				1								8
44	40F5				3	2	3												8
45	42F7					1	1	1	1	1	1	1	1						8
46	47F1								1	1	2		1	2			1		8
47	33F3						2	1	1	1	1	1							7
48	36F0		1	2	1		1					1						1	7
49	37G1		1					1	1			1		1		1	1		7
50	40E8				2	1	1	2	1										7
51	36F5			1			1	1				1	1		1				6
52	39F4		1			1	1				1	1	1						6
53	40E9				2	1		2	1										6
54	40F1		1	3	1		1												6
55	40F3		1	1	1		2	1											6
56	41F0			2			1	2		1									6
57	38F3		2		1	1		1											5
58	40F0		1	2	2														5
59	41F2			2	1			1					1						5
60	41F5		1	2	1	1													5
61	45E6												1	1	1	1	1		5
62	30E6														1	1	1	1	4
63	35F2		1	1	1		1												4
64	35F4						1	1								1		1	4
65	36F3		1	1					1						1				4

Table A8.4. Continued.

No.	ICES rect.	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	Total
66	36F7		1	1	1		1												4
67	37G0		1			1	1	1											4
68	40E2														1	2	1		4
69	41F1						1	2			1								4
70	41F3			2	1		1												4
71	41F4			1			2	1											4
72	44E9								1	1					1	1			4
73	30E7														1		1	1	3
74	34E3														1	1	1		3
75	37E6														1	1	1		3
76	38F8	1	1		1														3
77	38G3											1		1		1			3
78	39G3		1											1		1			3
79	40F2				1					1	1								3
80	42F5		1						1							1			3
81	46E4														1	1	1		3
82	46E5														1	1	1		3
83	46F3									1	1			1					3
84	33F4																1	1	2
85	34F2		1										1						2
86	34F4															1	1		2
87	36E4															1	1		2
88	36F6	1						1											2
89	38E9				1		1												2
90	38F4						2												2
91	38G2							1	1										2
92	38G4						1							1					2
93	39F8				1			1											2
94	41E9						1	1											2
95	42F2							1					1						2
96	46E8									1	1								2
97	56D3												1	1					2
98	27E9																	1	1
99	28E4															1			1
100	30F2														1				1
101	32F2												1						1
102	35F1			1															1
103	35F5														1				1
104	38G1							1											1
105	39E8											1							1
106	39F1																	1	1
107	39F2		1																1
108	42F6		1																1
109	43E8													1					1
110	43F1																1		1
111	43F3							1											1
112	43F6							1											1
113	43F7							1											1
114	44F1															1			1
115	46E9										1								1
116	46F2									1									1
117	46F5															1			1
118	47E4										1								1
119	47E9									1									1
120	55C8												1						1
121	56C8												1						1
122	56C9												1						1
123	57C7												1						1
Total		51	72	57	89	61	97	88	61	54	58	57	63	48	41	57	57	28	1,039

Table A8.5. Number of reports per ICES rectangle for flounder (*Platichthys flesus*). Hauls are combined; reports for different days are counted individually.

No.	ICES rect.	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	Total
1	38G4			3		1	2	2	2		1	2		2		1	1		17
2	38G5			2		1	1	3	2		1	1		2		1	1		15
3	32F3			1	1	1	1	1	1	1		1	1	1	1	1	1		13
4	38G3			1			1	2	1		1	2		1		2	2		13
5	33F4					1	1	1	1	1		1	1	1	1	1	1		11
6	39G6						2	1	2		1	1		2		1	1		11
7	38G0			1		1	3		1					1		1	1		9
8	39G8			1				1	2		1	1		1		1	1		9
9	39G3							1	1		1	1		1		1	1		7
10	31F3											1	1	1	1	1	1		6
11	34F4											1	1	1	1	1	1		6
12	35F6											1	1	1	1	1	1		6
13	37G1			1				1				1		1		1	1		6
14	40G7						1	2	3										6
15	39G7						1	2	2										5
16	40G6			1			1		1							1	1		5
17	38G2			1		1		1	1										4
18	40G5							1				1		1		1			4
19	37G5			1				1				1							3
20	38G6						1	1	1										3
21	40G8							1	1					1					3
22	33F3			1	1														2
23	39G4			1		1													2
24	41G7						1	1											2
25	41G8											1					1		2
26	37G0						1												1
27	39G2			1															1
28	39G5													1					1
29	40G4													1					1
Total		0	0	16	2	7	17	23	22	2	6	17	5	20	5	16	16	0	174

daily basis: all hauls reported in one rectangle for one day were combined, but hauls from different days were counted as different samples. In previous studies, samples were combined within a whole month, which led to a smaller number of samples per ICES rectangle and year.

Carrying over the selection criteria for dab data to the flounder data would have left the time series from only five ICES rectangles for further analysis (32F3, 33F4, 38G3, 38G4, 38G5). In order to provide at least a partial impression about temporal developments in flounder disease prevalence, the original criterion was relaxed by performing an exploratory analysis also for time series with six to eight sampling dates. This resulted in the performance of an analysis for a total of twelve rectangles.

5 METHOD OF STATISTICAL ANALYSIS

For all diseases under study, the target quantity is a prevalence. This makes it natural to use a logistic model for the analysis of relationships between disease prevalence and (potentially) explaining variables. The standard form of a (linear) logistic model is

Table A8.6. Number of dab (*Limanda limanda*) examined and numbers found with specific diseases. Only rectangles with at least nine reports are shown. Note: While all fish were examined for the presence of lymphocystis, epidermal hyperplasia/papilloma, and skin ulcers, only some were examined for the presence of skeletal deformities and liver nodules.

Area	Number of specimens examined	Lymphocystis	Epidermal hyperplasia/ papilloma	Skin ulcers	Skeletal deformities	Liver nodules
30F0	3,773	122	63	110	44	16
34F3	6,744	332	203	28	13	25
35F3	7,164	607	302	166	9	15
3651	13,265	1,474	584	354	68	34
36F2	3,484	274	62	92	6	.
36F4	8,603	372	201	50	6	49
37F0	7,337	677	218	196	5	132
37F1	7,701	985	190	148	9	.
37F2	6,586	822	283	391	18	.
37F3	8,074	1,340	298	78	8	1
37F4	6,810	860	256	54	12	4
37F5	6,368	357	83	81	3	2
37F6	9,247	505	126	101	7	.
37F7	33,097	3,207	1,625	485	36	.
37F8	3,700	91	14	48	2	.
38F0	2,939	391	67	18	1	4
38F1	7,292	862	118	173	6	58
38F2	19,072	2,356	706	1,643	37	.
38F6	3,769	256	72	20	1	.
38F7	17,525	1,332	563	164	4	.
38G0	5,188	304	10	65	21	.
39E9	8,822	2,096	257	63	28	.
39F0	3,848	548	46	39	3	.
39F3	8,150	1,360	227	335	10	5
39F5	3,245	323	31	42	3	.
39F6	7,868	921	174	59	8	.
39F7	8,493	778	288	89	7	.
40F4	1,609	375	58	34	1	.
40F6	4,420	605	104	53	8	.
40F7	18,366	2,124	791	214	24	.
41E7	24,991	3,816	644	928	178	15
41E8	4,812	1,413	244	351	114	.
41F6	8,010	846	174	69	3	.
41F7	6,757	576	200	62	1	.
41G1	4,224	329	64	11	.	.
41G2	3,725	264	65	3	.	.
42F3	6,878	1,786	118	212	15	.
42G1	3,395	206	60	7	.	.
44E8	6,643	1,579	316	152	172	.
44F9	5,817	57	11	45	.	.
44G0	2,196	10	3	33	.	.
47E8	3,987	349	62	45	0	2
Total	333,994	37,887	9,981	7,311	891	362

Table A8.7. Number of flounder (*Platichthys flesus*) examined and numbers found with specific diseases. Only rectangles with at least nine reports are shown. Note: While all fish were examined for the presence of lymphocystis, epidermal hyperplasia/papilloma, and skin ulcers, only some were examined for the presence of skeletal deformities and liver nodules.

Area	Number of specimens examined	Lymphocystis	Epidermal hyperplasia/ papilloma	Skin ulcers	Skeletal deformities	Liver nodules
31F3	1,774	14	.	13	.	7
32F3	5,300	177	.	60	.	11
33F4	5,339	760	.	137	.	55
34F4	1,800	64	.	134	.	6
35F6	1,320	22	.	14	.	1
37G1	369	76	.	5	.	.
38G0	50	13	.	0	.	.
38G3	2,512	571	.	58	.	.
38G4	1,421	261	.	25	.	.
38G5	1,192	249	.	31	.	.
39G3	508	185	.	11	.	.
38G6	408	103	.	16	.	.
39G8	250	30	.	14	.	.
40G7	34	1	.	0	.	.
Total	22,277	2,526	.	518	.	80

$$\text{Prob(disease present)} = f(a_0 + a_1x_1 + a_2x_2 + \ldots + a_qx_q) \quad (1)$$

where $x_1, x_2, \ldots x_q$ denote the explaining variables, $a_0, a_1, \ldots, a_q$ are unknown coefficients that must be estimated from empirical data, and $f(\cdot)$ is the logistic function

$$f(z) = 1 / (1 + \exp(-z)). \quad (2)$$

A logistic model (1) is a member of the class of Generalized Linear Models (GLMs; cf. McCullagh and Nelder, 1989).

The explaining variables (the x_i) in the present problem are:

- gender, technically a factor to be coded by a dummy variable for female/male;
- size (length), a factor with groupings ≤ 14 cm, 15–19 cm, 20–24 cm, ≥ 25 cm for dab, and with groupings 15–19 cm, 20–24 cm, 25–29 cm, ≥ 30 cm for flounder (these size group definitions allow the use of early data which were reported only for size groups);
- quarter of the year, a factor expressing seasonal variation and biological phase, with levels: Q1 (for Jan, Feb, Mar), Q2 (for Apr, May, Jun), Q3 (for Jul, Aug, Sep), Q4 (for Oct, Nov, Dec);
- calendar time (to describe the temporal trend).

While gender, size, and quarter can easily be modelled within the framework of the linear logistic model (1), this is not true for the temporal trend, the quantity of central interest. There is no reason to assume that a trend is linear over a long time or could be represented by simple elementary functions such as low-order polynomials, logarithms, or exponentials of time. Instead, earlier analyses suggest that trends of a more general non-linear shape are present (ICES, 1997b). Consequently, a generalization of (1) is needed to allow the incorporation (and estimation) of general non-linear trend functions. Hence, an extended model of the form

$$\text{Prob(disease present at time } t) = f(a_0 + a_1x_1 + a_2x_2 + \ldots + a_qx_q + s_1(t) + s_2(t) + \ldots + s_p(t)) \quad (3)$$

was used, which contains smooth functions $s_j(t)$ to describe temporal trends. These functions are estimated as a whole, i.e., also their shape is estimated by a fitting procedure. The only restriction is that the shape is to exhibit a certain degree of smoothness. This is not a real restriction here, as a certain smoothness is by definition a required property of a trend. The degree of smoothness can be expressed in terms of 'equivalent degrees of freedom' (df), which must be specified by the investigator as a prerequisite to use the Generalized Additive Model (GAM, cf. Hastie and Tibshirani, 1990) in (3). High df values indicate high variability or, equivalently, non-smooth curves.

For the dab data, the optimal smoothing parameter was determined for the longest time series (reported for rectangle 37F7) by cross validation within the GAM model. The optimal number of degrees of freedom there was found to be df = 7. This value was then used uniformly for all dab time series. For the flounder data, a similar calculation on the basis of the data from rectangle 32F3 (13 samples in 13 years, values in quarters Q3 and Q4) recommended a smoothing with 9 df, while a cross validation for the short-term series for rectangle 31F3 (6 samples in six years, only values in Q4) led to a value of df = 2. Though this difference seems enormous at first sight, it does not for the present problem go along with a severe change in the prevalence that is predicted by (3). The reason seems to be that a major component of the temporal variation is the seasonal fluctuation, which, with a very smooth trend $s_i(t)$, is accounted for by large coefficients of the seasonal components, while, in the case of a non-smooth trend (high df), the seasonal changes are incorporated in the trend, not in the model term for season. This shift of weights within the mathematical model could generate various interpretations of the model components, but leads to nearly identical estimates for Prob(disease present). As a result of these considerations, a smoothing with 2 df was generally used for all flounder data in order to separate trend and seasonal fluctuations clearly also for the shorter data series.

It should be noted that the choice of a smoothing parameter in general is still a field of ongoing research. The procedure applied here follows general recommendations as given by Hastie and Tibshirani (1990). It seems to cover the actual practical needs and is at the same time driven by some economic considerations, as the choice of an optimal smoothing parameter is a computer- and time-consuming exercise, for which, at present, execution times must be measured in hours. But before establishing a general protocol for the application of a GAM approach to fish disease data, one might want to review the strategy for choosing a smoothing parameter, particularly if future data collection requires assessing longer time series.

For each combination of gender and size class, an individual trend $s_i(t)$ was estimated. This is equivalent to introducing a threefold interaction term for gender*size* time. All of these trends were estimated with the same degree of smoothness.

All model fitting was done individually for each ICES rectangle. Confidence intervals for the predicted prevalence (the trends) were determined by a parametric bootstrap simulation.

All calculations within the GAM approach were carried out using the function gam() of the S-Plus software (MathSoft, 1997).

6 RESULTS

The primary result of fitting a Generalized Additive Model is a set of fitted coefficients $a_0, a_1, \ldots, a_q$ and the fitted trends $s_1(t), s_2(t), \ldots, s_p(t)$. Approximate confidence intervals for all of these terms are also supplied directly by the software, however, not for the estimated prevalence, for which confidence intervals were estimated by simulation. The aim of the analysis, namely the identification of temporal trends, requires the presentation of estimated prevalence trends together with confidence bounds. These trends are shown for each of the ICES rectangles with sufficiently long time series, where the operational definition of 'sufficiently long' was described above. Only the trends for female specimens with a length of 20–24 cm for dab and a length of 25–29 cm for flounder are shown. These subgroups were selected because they are covered with the highest frequency. Members of these length groups are relatively homogeneous in length, due to the grouping, but can also be assumed to be relatively homogeneous with respect to age, as individuals in these length classes are neither extremely young nor likely to be extremely old, but instead somewhere in the middle of their growth phase.

Trends are shown in Figures A8.2 to A8.6 with full detail, i.e., including seasonal variation (for the locations of the ICES rectangles covered, see Figures A8.7 and A8.8). Solid black lines in the figures refer to trends in fish disease prevalence and dotted lines to the 90 % confidence bounds for the time series at the observed points. These can lie in different quarters of a year and can thus be influenced by seasonal variation. Frequently a data series contains varying quarters of years, which causes the (erroneous) impression of erratic fluctuation over the years. To remove this problem and to facilitate the interpretation of the temporal trends, an additional season-adjusted trend estimation is added to the trend at observed time points. For dab, this additional trend shows the prevalence estimate valid for 15 February of each year, for the time period that is covered by the observed series. For flounder, the estimate valid for 15 November is shown. It refers as before to females in the length class 20–24 cm (dab) or 25–29 cm (flounder) and is shown as a dashed line. As this trend is calculated only for dates in one quarter of a year (dab: Q1; flounder: Q4), no seasonal effects can show up, so this curve is easier to use for an interpretation.

The season-adjusted trend can also be used for a first assessment of recent temporal developments. Also for this trend, a 90 % confidence bound was estimated by simulation. By comparing the prevalence estimates for 1992 with those for 1997 (or 1996, if a series does not extend to 1997), a rough statement about the general temporal development of the trend (constant, downward, upward) can be derived. The comparison can be performed by checking the sign of the prevalence difference and by checking whether the two confidence intervals (for 1992 and 1997) overlap. Non-overlapping

Figure A8.2. Estimated temporal trend (solid line) with 90 % confidence intervals (dotted lines) of lymphocystis prevalence in common dab (*Limanda limanda*) for females in the 20–24 cm length class, by statistical rectangle. This trend is based on the empirical sampling dates. The season-adjusted estimated prevalence is shown as a dashed line.

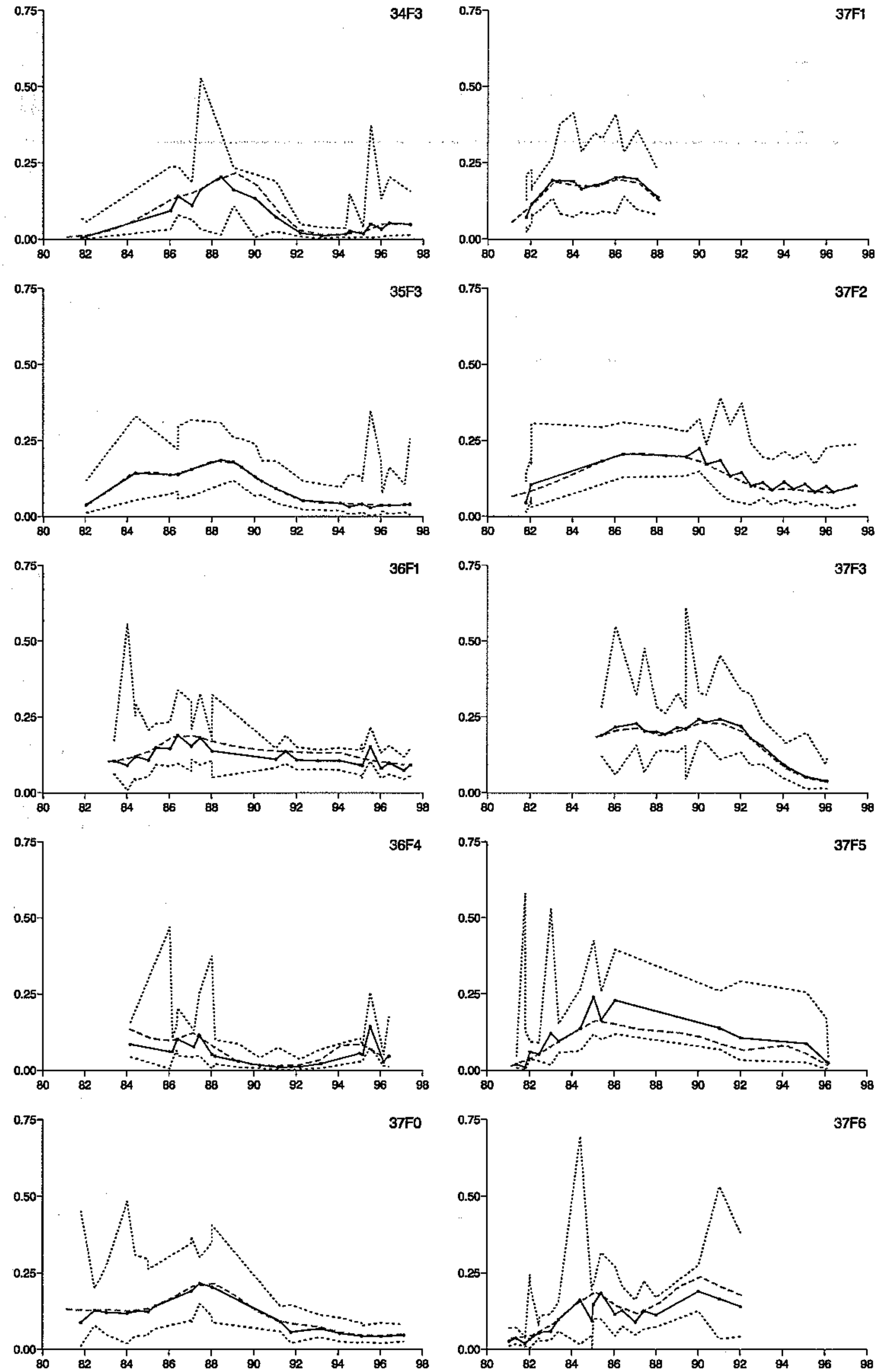


Figure A8.2. Continued.

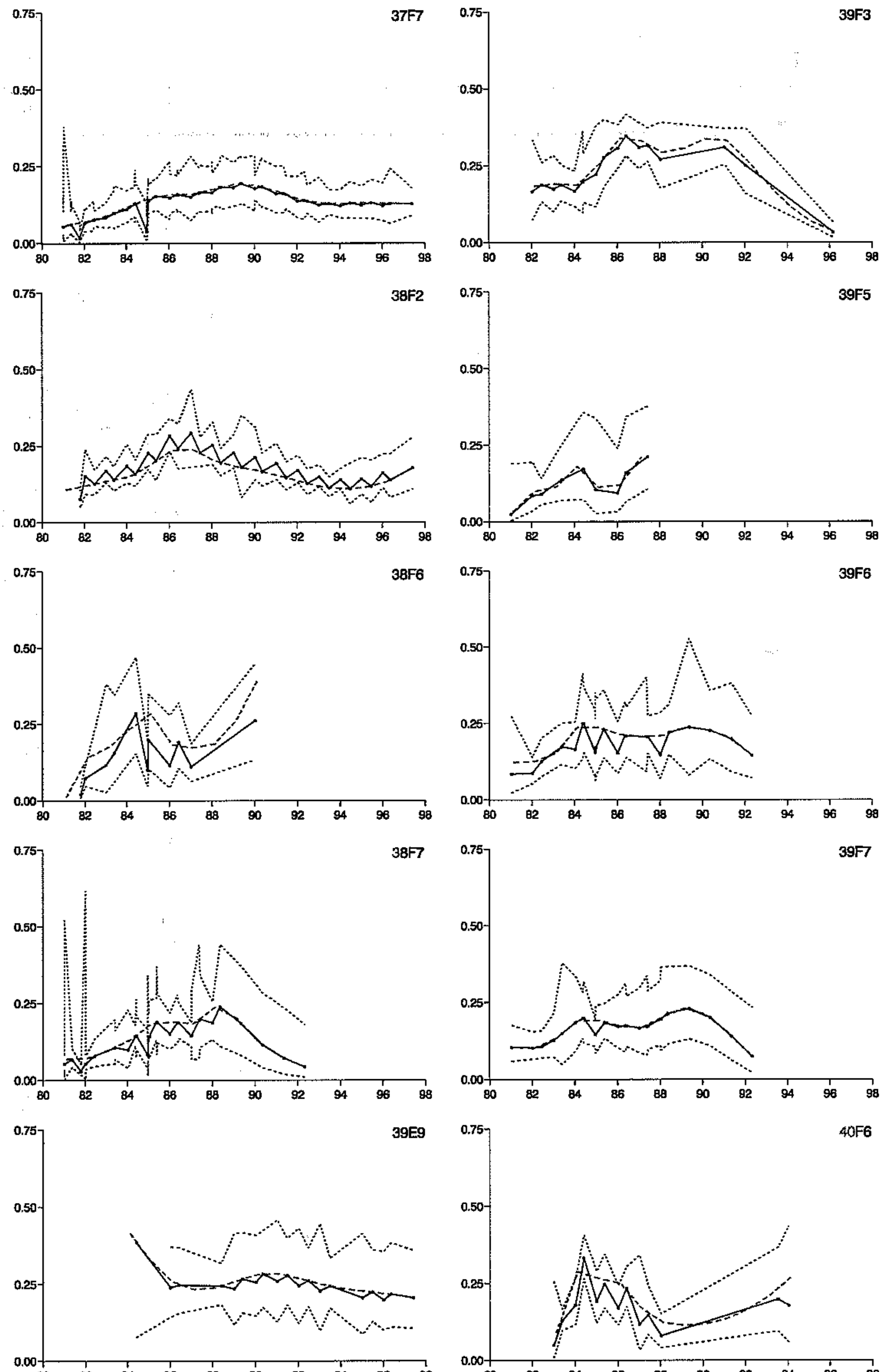


Figure A8.2. Continued.

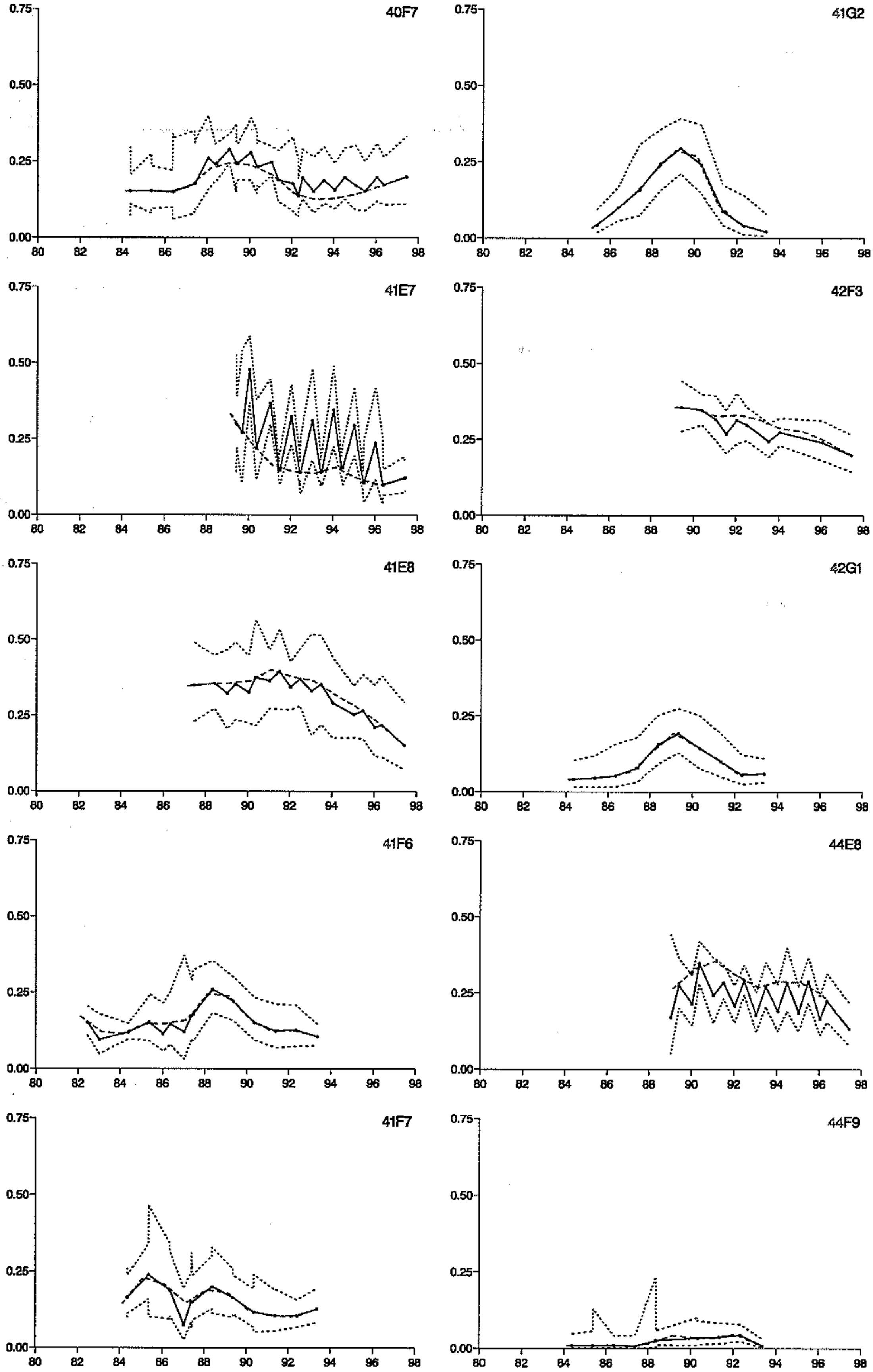


Figure A8.2. Continued.

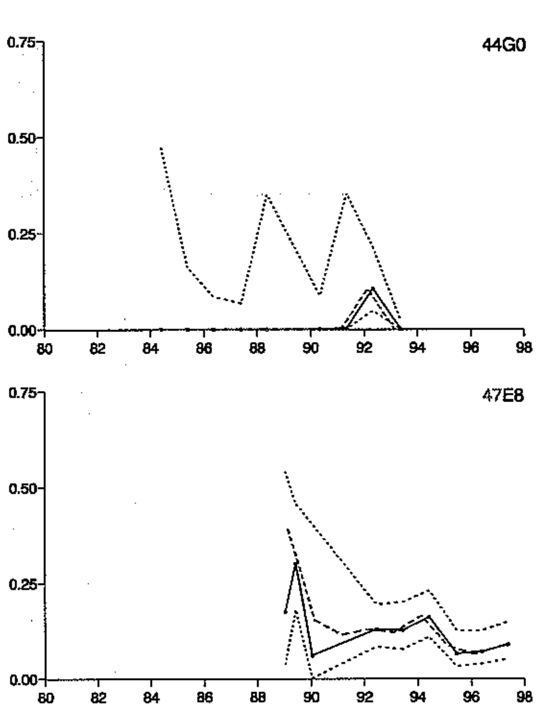


Figure A8.3. Estimated temporal trend (solid line) with 90 % confidence intervals (dotted lines) of epidermal hyperplasia/papilloma prevalence in common dab (*Limanda limanda*) for females in the 20–24 cm length class, by statistical rectangle. This trend is based on the empirical sampling dates. The season-adjusted estimated prevalence is shown as a dashed line.

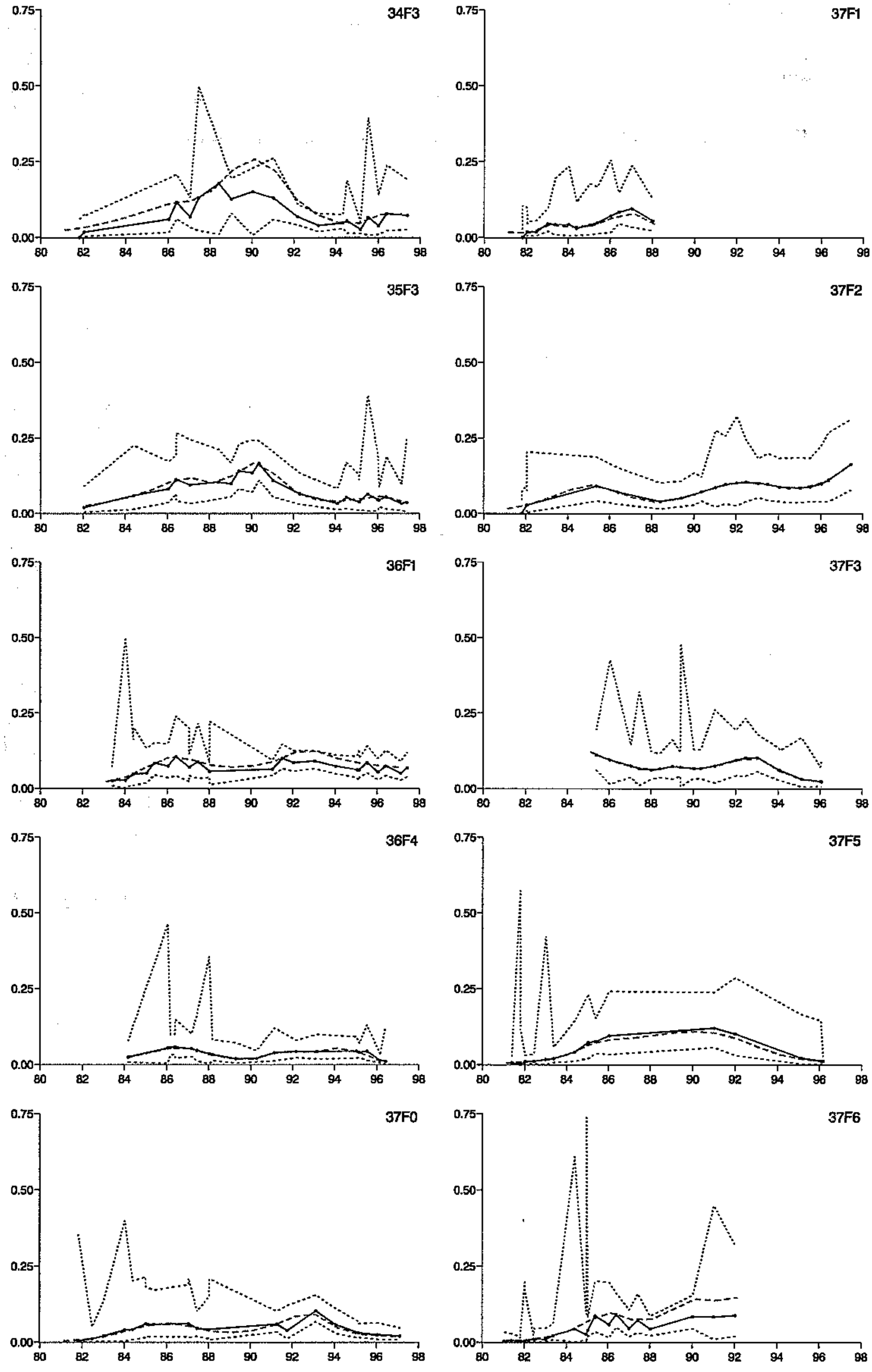


Figure A8.3. Continued.

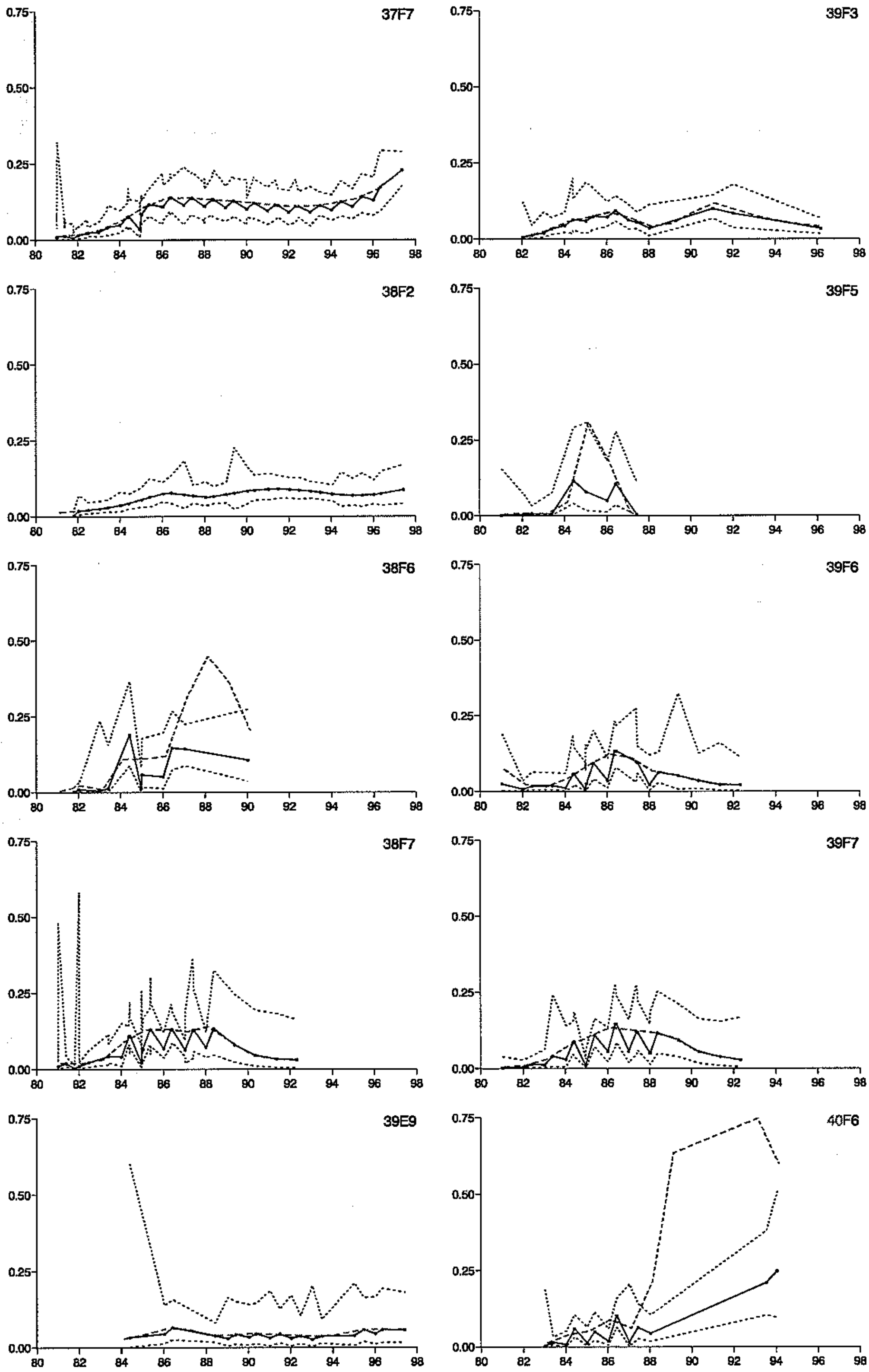


Figure A8.3. Continued.

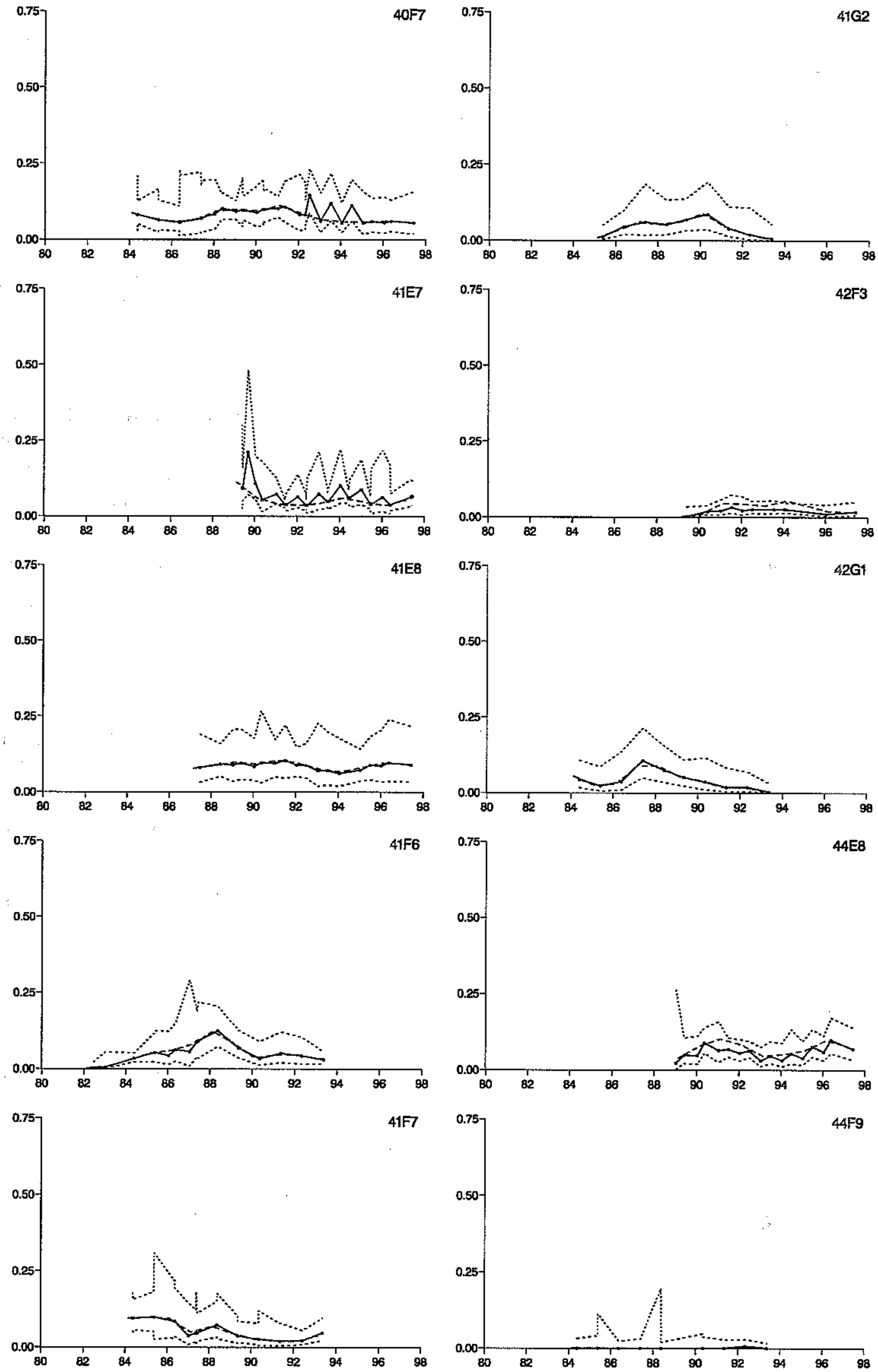


Figure A8.3. Continued.

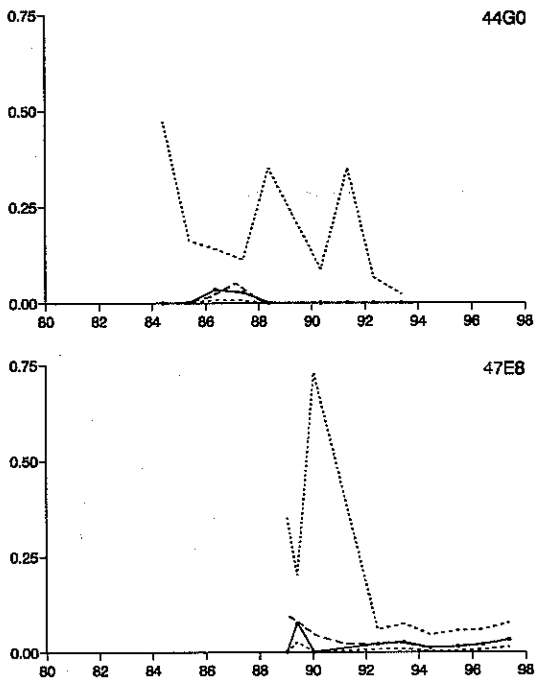


Figure A8.4. Estimated temporal trend (solid line) with 90 % confidence intervals (dotted lines) of acute/healing skin ulcer prevalence in common dab (*Limanda limanda*) for females in the 20–24 cm length class, by statistical rectangle. This trend is based on the empirical sampling dates. The season-adjusted estimated prevalence is shown as a dashed line.

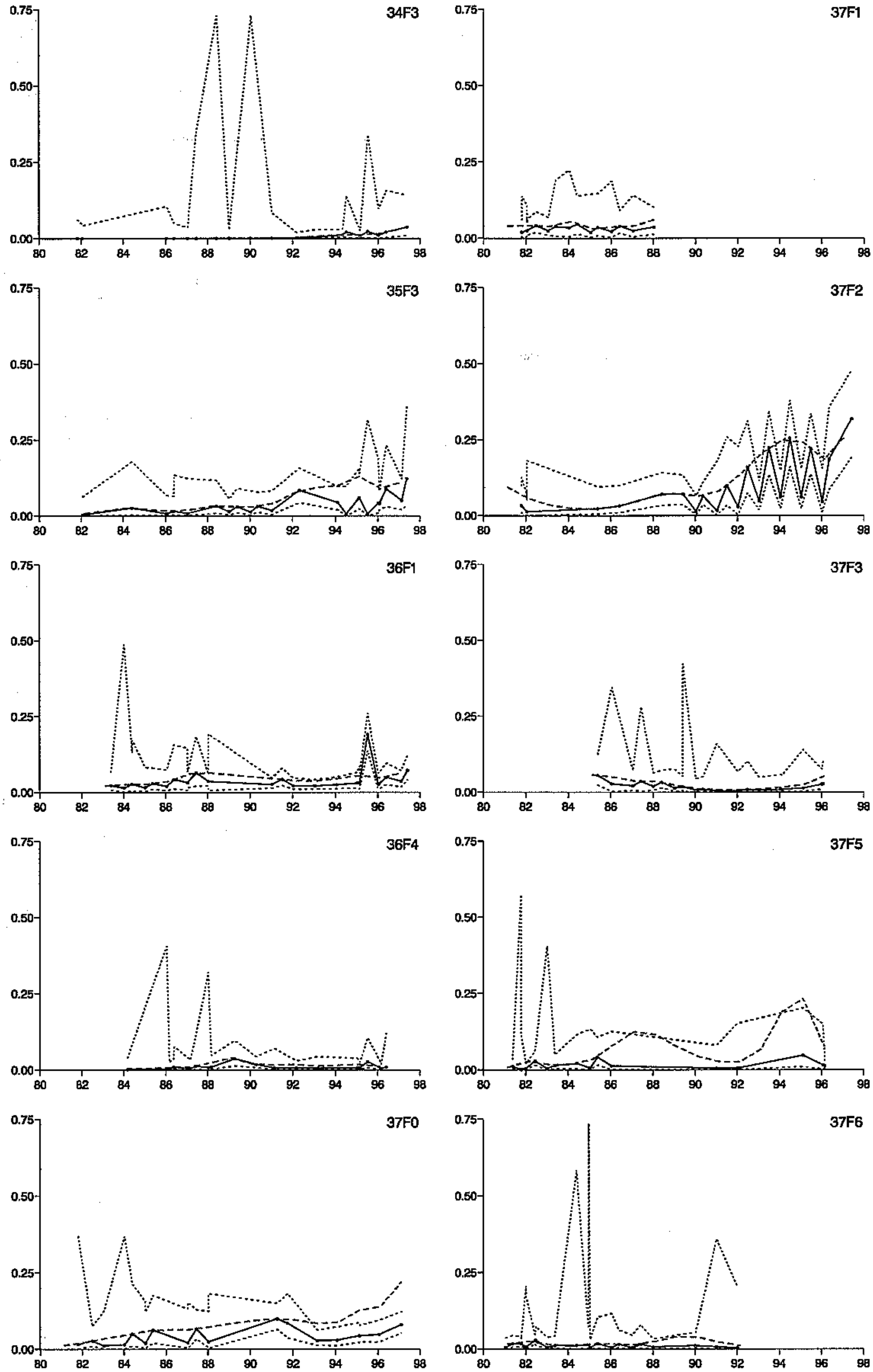


Figure A8.4. Continued.

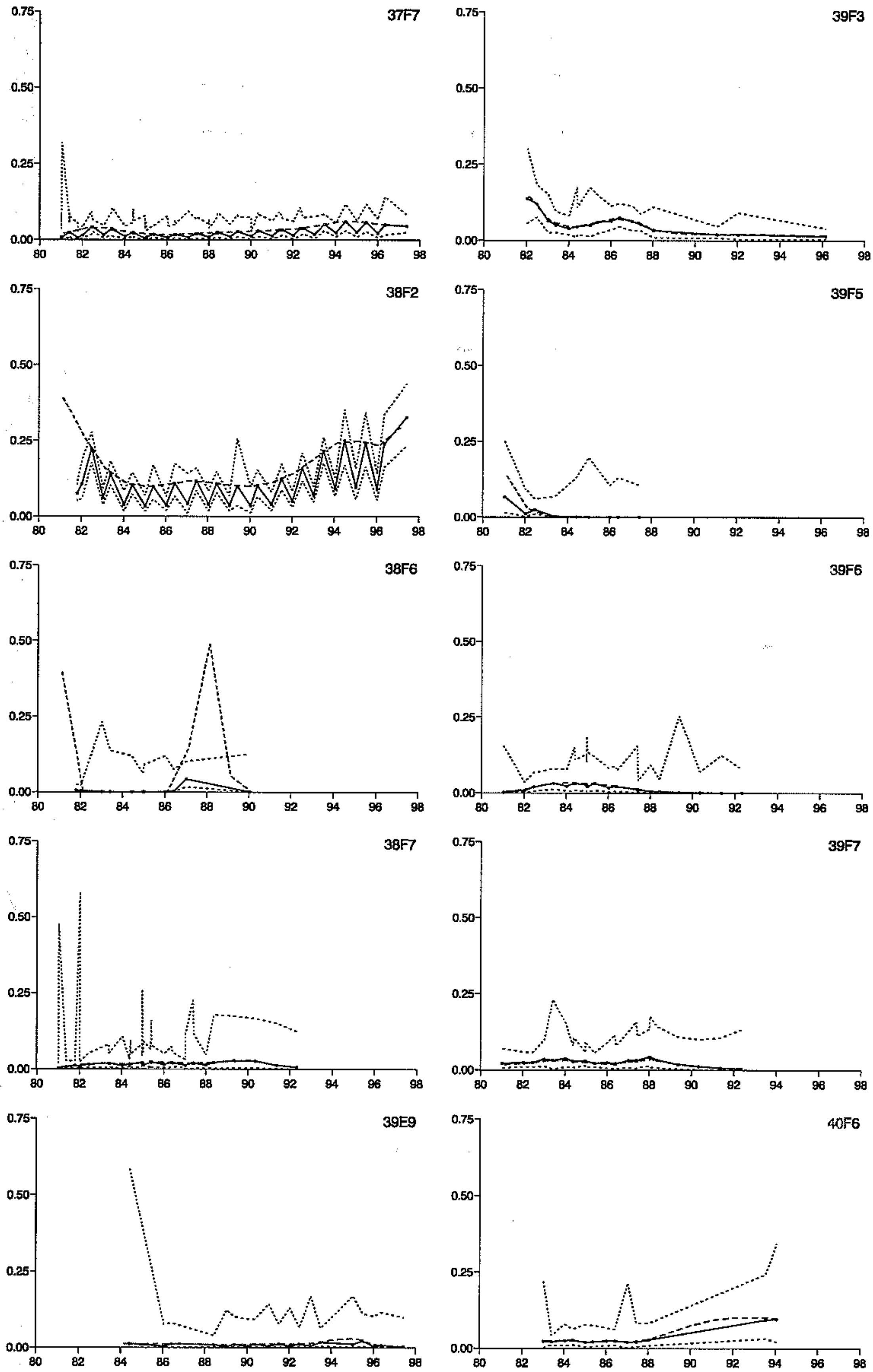


Figure A8.4. Continued.

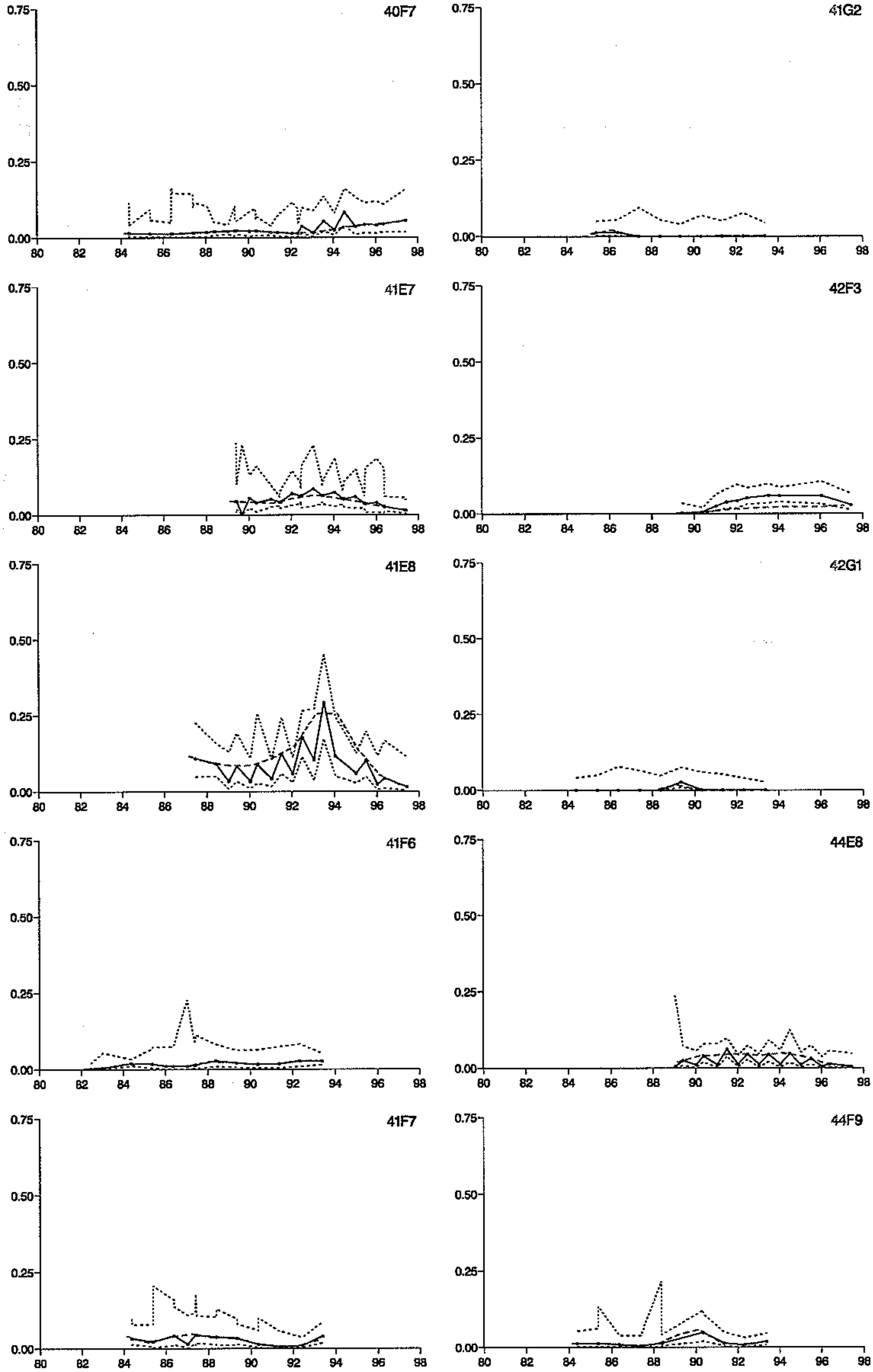


Figure A8.4. Continued.

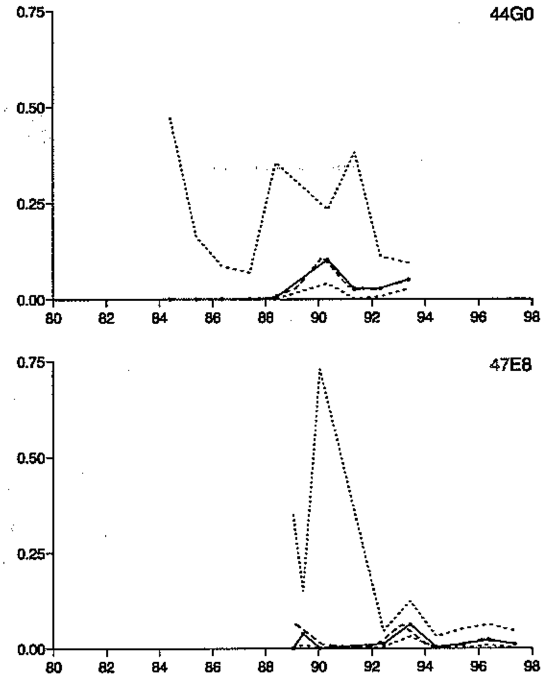


Figure A8.5. Estimated temporal trend (solid line) with 90 % confidence intervals (dotted lines) of lymphocystis prevalence in flounder (*Platichthys flesus*) for females in the 25–29 cm length class, by statistical rectangle. This trend is based on the empirical sampling dates. The season-adjusted estimated prevalence is shown as a dashed line.

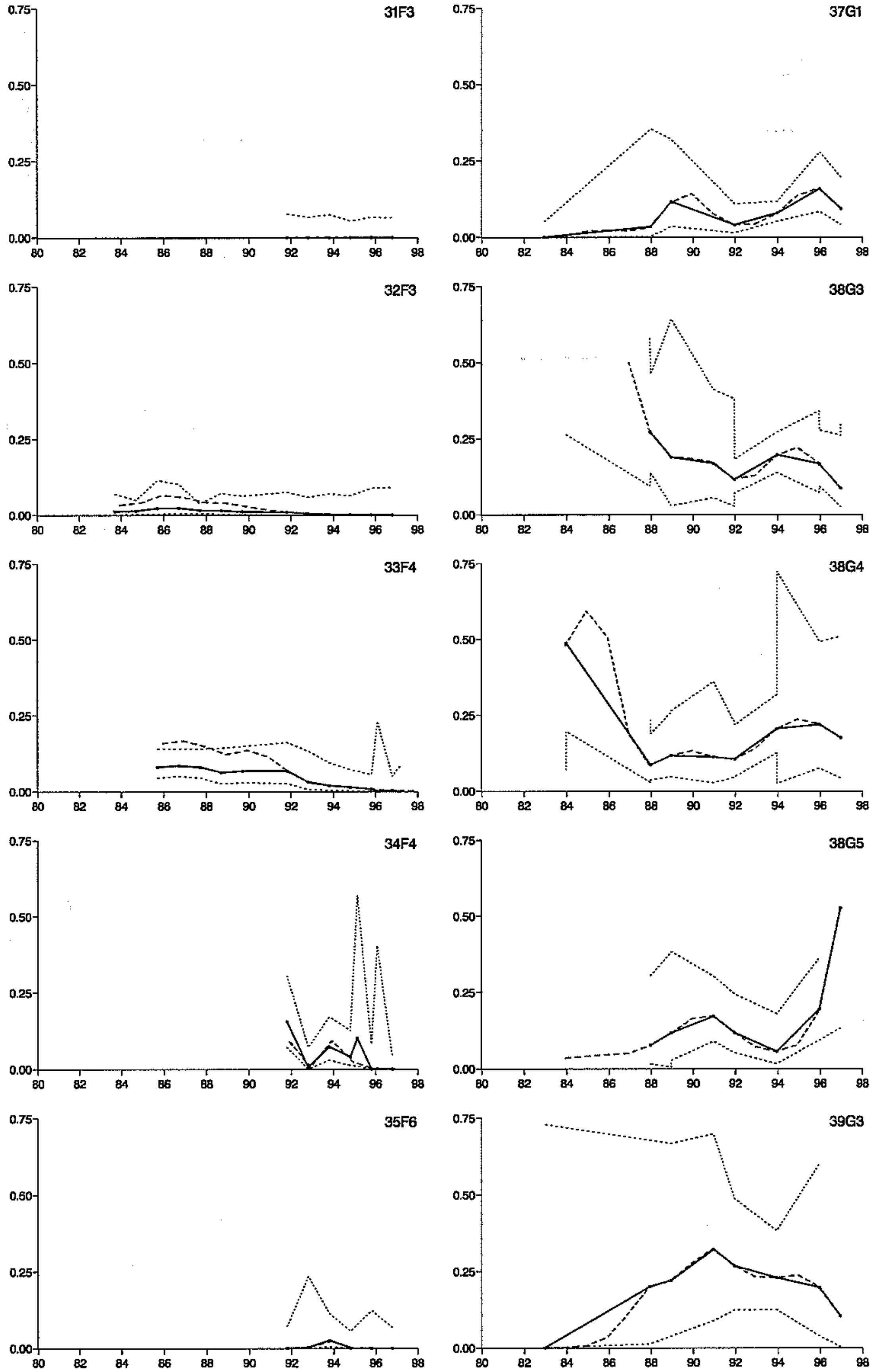


Figure A8.5. Continued.

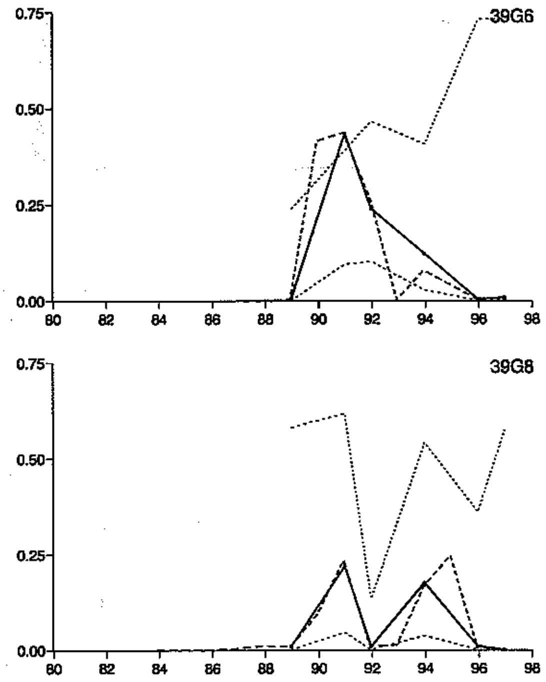


Figure A8.6. Estimated temporal trend (solid line) with 90 % confidence intervals (dotted lines) of skin ulcer prevalence in flounder (*Platichthys flesus*) for females in the 25–29 cm length class, by statistical rectangle. This trend is based on the empirical sampling dates. The season-adjusted estimated prevalence is shown as a dashed line.

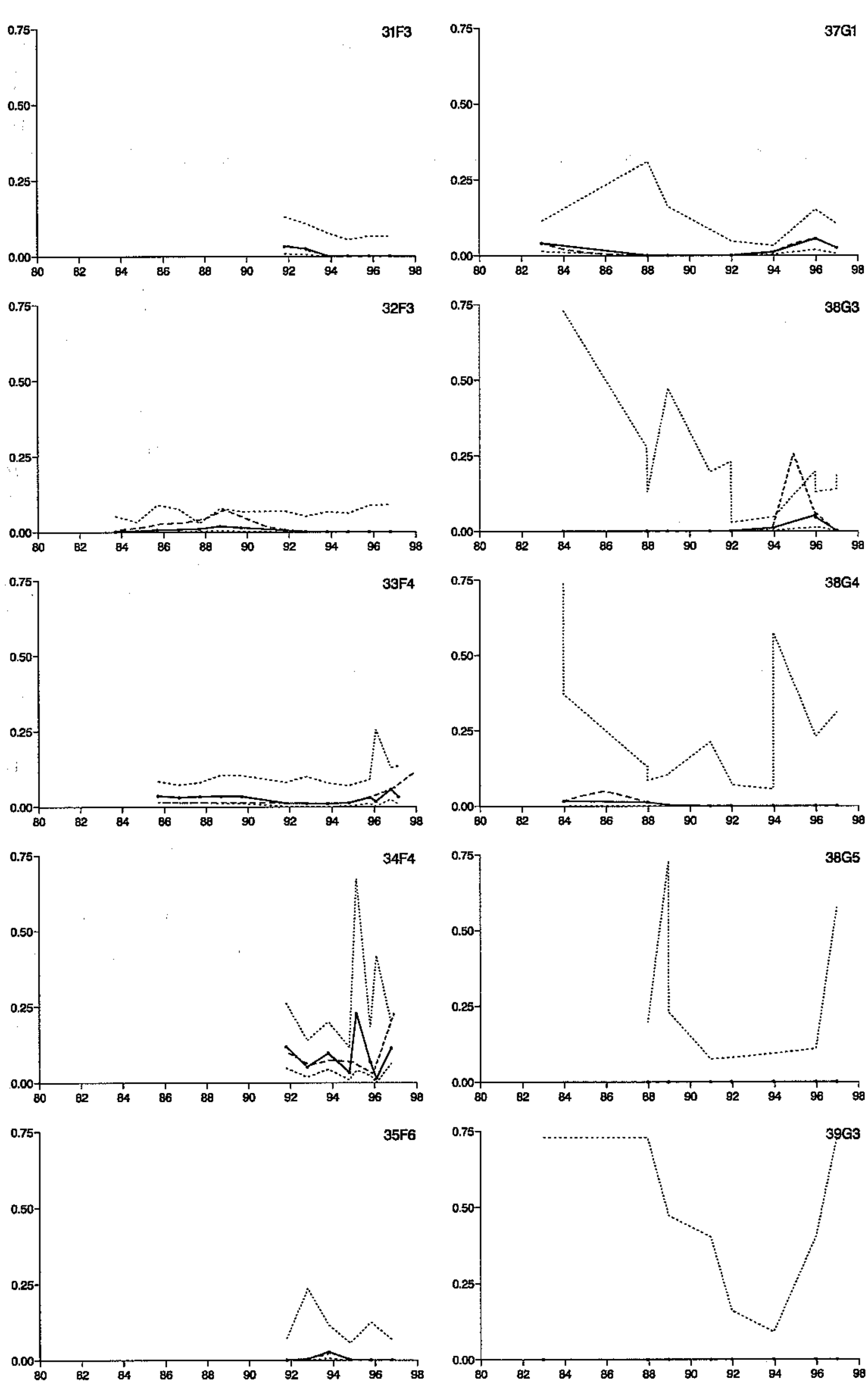
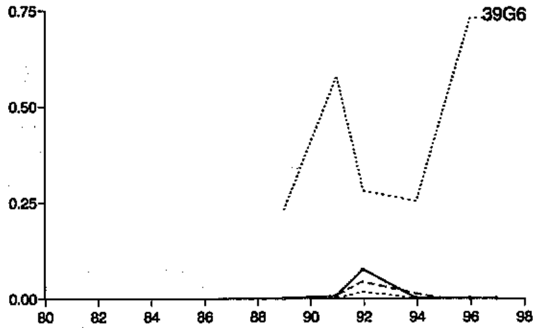


Figure A8.6. Continued.



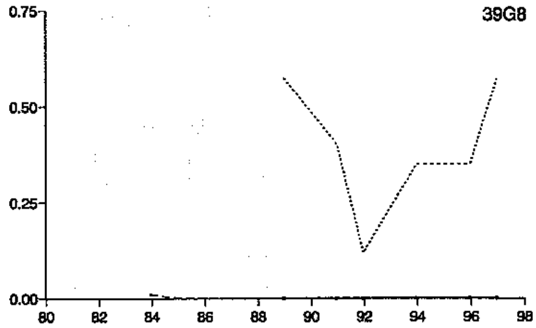


Figure A8.7. Recent developments in the prevalence of lymphocystis, epidermal hyperplasia/papilloma, and acute/healing skin ulcers in common dab (*Limanda limanda*). Only trends from 1992 onwards until the last observation in a rectangle are considered. The trend is checked for significance (p = 10 %) by the Mann-Kendall test for non-parametric trends. Significant increasing/decreasing trends are marked with arrows (up/down) and areas with no changes are marked with circles.

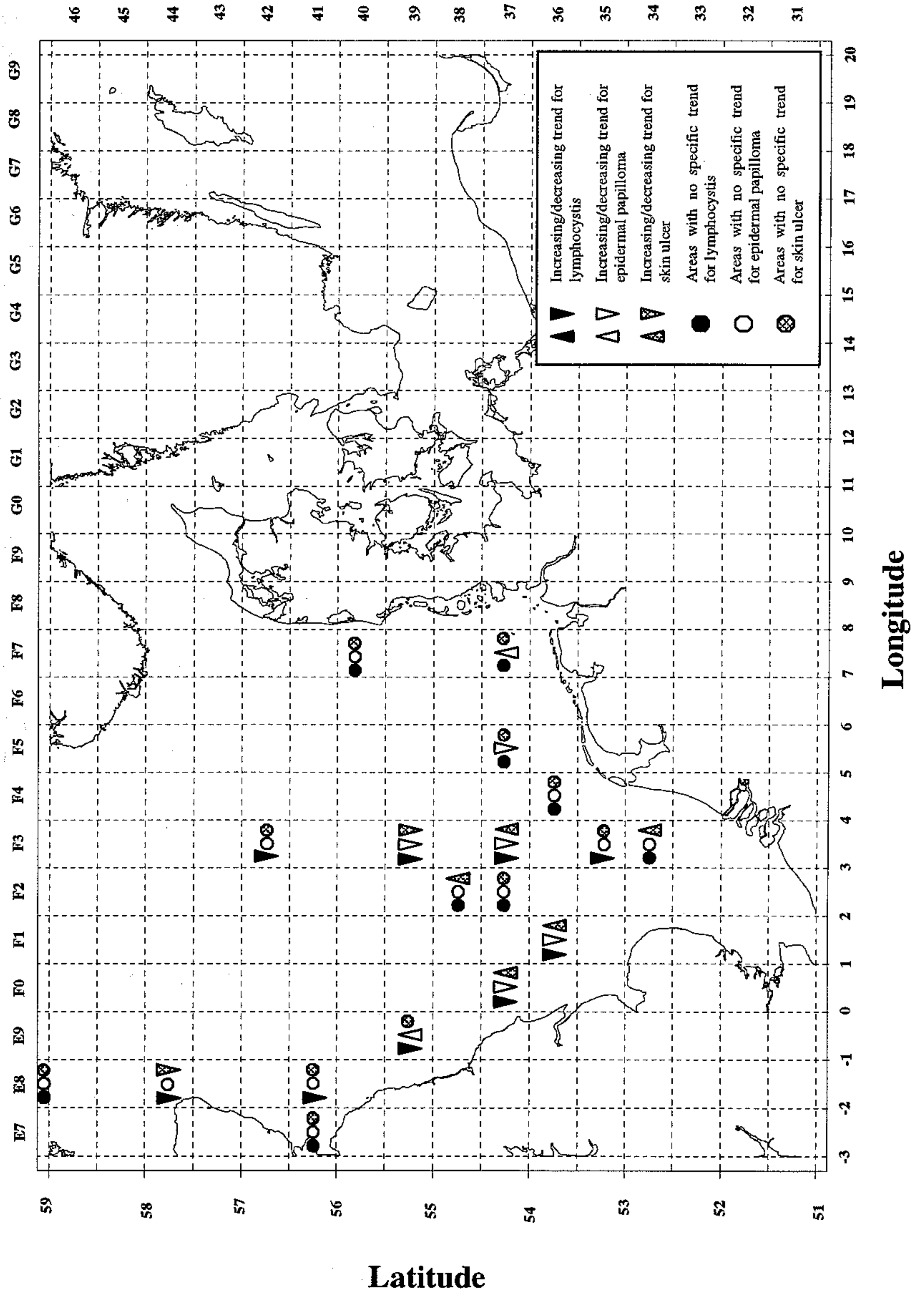
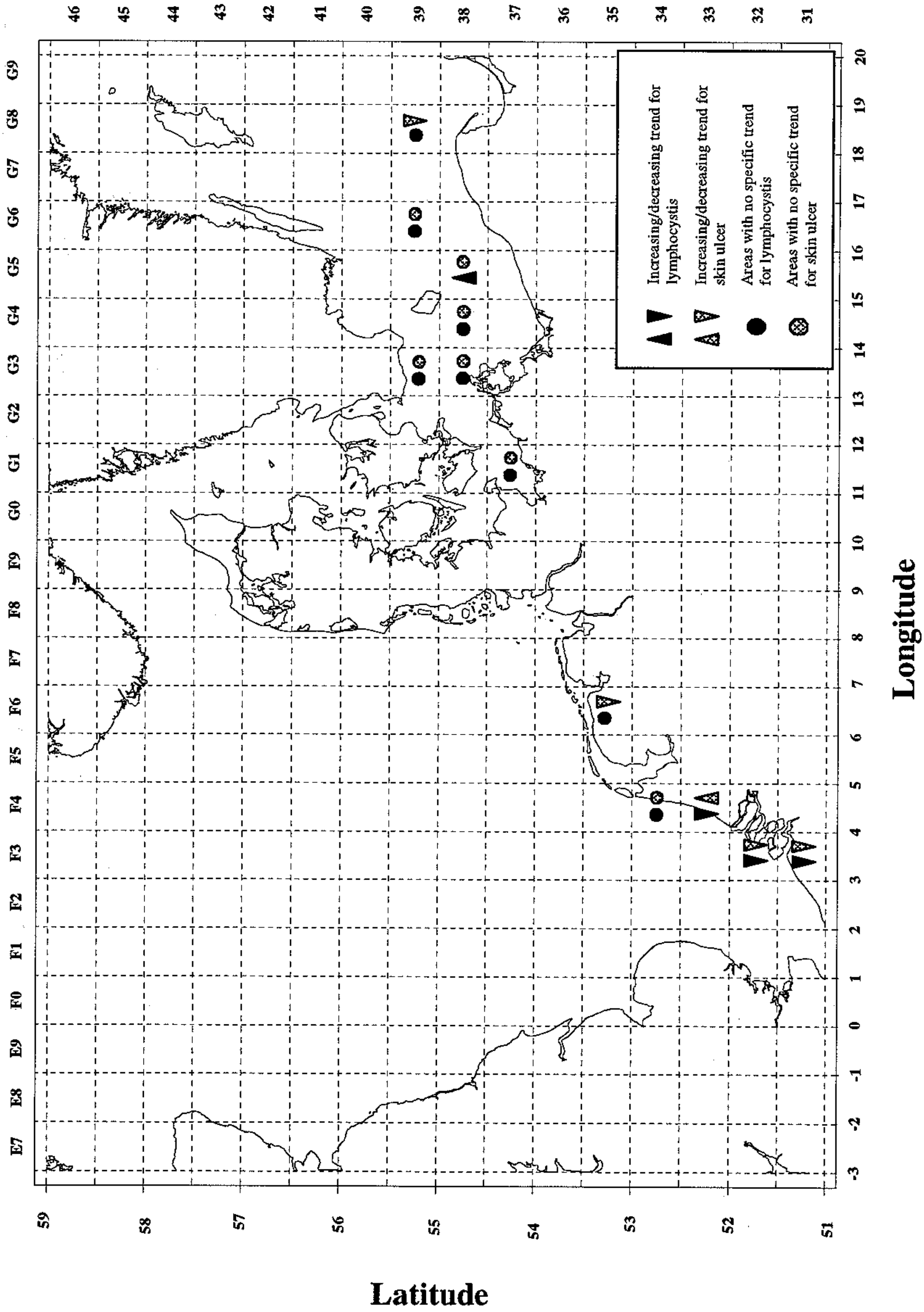


Figure A8.8. Recent developments in the prevalence of lymphocystis, epidermal hyperplasia/papilloma, and acute/healing skin ulcers in flounder (*Platichthys flesus*). Only trends from 1992 onwards until the last observation in a rectangle are considered. The trend is checked for significance (p = 10 %) by the Mann-Kendall test for non-parametric trends. Significant increasing/decreasing trends are marked with arrows (up/down) and areas with no changes are marked with circles.



confidence intervals indicate a clear increase or decrease in the trend.

However, even if confidence intervals overlap, a significant trend can exist. The reason is that a confidence band as a graphical tool can only give point-wise information, i.e., it does not say that an arbitrary shape of curve within the band would be compatible with the structure of the trend. Hence, an additional test is needed to derive a more concise statement about recent trends. The Mann-Kendall test for the existence of a non-parametric trend is an appropriate tool for this purpose (see ICES, 1997b, p. 121). Only prevalence data from 1992 onwards was used to consider recent trends. Locations for which the observation series had no or too few data in that time window could clearly provide no information in this respect. The Mann-Kendall test needs at least four observations to be able to detect a trend with a two-sided error level of 10 %. The two-sided test level of 10 % is certainly only an approximate value, as there is no adjustment for the fact that the test is performed on quantities which were already estimates based on the observational data. The predicted prevalences for 15 February of each year were used for the trend tests, thus providing a common time pattern for all rectangles and avoiding quarter effects. Figures A8.7 and A8.8 summarize the results of the tests for dab and flounder, respectively, with the same sex/length properties as in Figures A8.2–A8.6. Rectangles and diseases with significant trends over the period 1992–1996/1997 are marked with upward or downward arrows; rectangles and diseases showing no trends are marked with circles.

In the following paragraphs, the main results of the analyses carried out are summarized for dab and flounder, separately.

6.1 Dab (*Limanda limanda*)

- There is a sufficient spatial coverage of disease prevalence data in the southern and central North Sea, whilst data for areas in the northern North Sea and outside the North Sea are relatively scarce (exception: areas 41G2, 42G1, 44F9, and 44G0 in the Kattegat and Skagerrak). Disease data covering the whole period from 1981–1997 are only available from the southern part of the North Sea (south of 55 °N), whilst data series from the northern parts (north of 55 °N) and the Skagerrak/Kattegat area either started later or were terminated earlier.
- For all diseases considered, the absolute levels of the estimated disease prevalence differ considerably between ICES rectangles. Some areas are characterized by a consistently high prevalence of a particular disease, such as areas 37F2 and 38F2 (Dogger Bank) for acute/healing skin ulcers and area 37F7 (German Bight) for epidermal hyperplasia/ papilloma. Others are characterized by a consistently low prevalence, such as areas 44F9 and 44G0 (Skagerrak) for lymphocystis and epidermal hyperplasia/papilloma.
- In some areas, the prevalence over time is rather uniform with a relatively low degree of variation (e.g., 37F7 and 39E9 for lymphocystis and 41E8 for epidermal hyperplasia/papilloma), whilst prevalence shows strong variation in other areas (e.g., 40F6 for lymphocystis and 37F2 for acute/healing skin ulcers). Sometimes, even sudden and marked changes occur within a relatively short period of time (e.g., 41G2 for lymphocystis and 37F2 for acute/healing skin ulcers).
- The prevalence undergoes significant and consistent seasonal change in some diseases and areas (first quarter of the year vs. second quarter of the year, e.g., areas 37F7 and 38F2 for acute/healing skin ulcers, areas 38F2 and 41E7 for lymphocystis).
- The prevalence of lymphocystis was either stable or decreasing in all rectangles with sufficient data for the period 1992–1996/1997. In none of the areas was a significant increase detected.
- The prevalence of epidermal hyperplasia/papilloma was more variable. Although stable or decreasing trends dominated, there are two areas (37F7, 40F6) which showed a significant upward trend. Especially the marked increase in rectangle 37F7 situated in the German Bight requires further attention.
- Acute/healing skin ulcers are the only diseases which increased in prevalence in a considerable number of rectangles, particularly in the southern and central North Sea.

6.2 Flounder (*Platichthys flesus*)

- There are many fewer data for diseases of flounder compared to dab, resulting in a lack of spatial and temporal coverage. In the North Sea, data are restricted to Dutch coastal areas and in the Baltic Sea to the southwestern part.
- The prevalence of lymphocystis is considerably higher in the Baltic Sea areas than in the North Sea areas analysed. In most areas in Dutch coastal waters, there was a significant downward trend in the prevalence of lymphocystis since 1992. No significant trends were found in the Baltic Sea, except in area 38G5, which was characterized by a significant increase.
- For acute/healing skin ulcers, no marked differences seem to exist in the prevalence between the North Sea and Baltic Sea areas considered. Both the North Sea and Baltic Sea data do not reveal any consistent temporal trends.

7 DISCUSSION AND CONCLUSIONS

The present status of the fish disease section of the ICES Environmental Data Centre database can be considered sufficiently complete to allow a statistical analysis of spatial and temporal characteristics of disease prevalences. However, most data available refer to diseases of dab and only limited information exists so far on flounder diseases. This is mainly due to the fact that the submission of flounder data has only started recently. Up to now, only two laboratories have provided information, but there are a number of others (particularly in the Baltic Sea countries) which, in the future, will contribute to the data bank by submitting historic and current data. Once the flounder data bank has been completed, a more comprehensive data analysis similar to the one carried out with the dab data will be possible, providing better insight into spatial and temporal trends.

Furthermore, it has to be taken into account when assessing the results of the analyses that the disease data analysed only represent a part of the whole set available in the ICES Environmental Data Centre, since data from male fish and from female fish outside the size groups 20–24 cm (dab) and 25–29 cm (flounder) were not included in the analysis. More complete information on spatial and temporal trends will be available once these data have been considered in the analysis as well.

The results of the analysis indicate that there are marked spatial differences with respect to the absolute levels and the temporal changes in the prevalences of the diseases under study (see Figures A8.2 to A8.6). For dab, with only one exception (39F3), there are no rectangles where the diseases considered followed the same temporal trend (either an increase or decrease in prevalence) over the period 1992–1996/1997 (see Figure A8.7). The situation is different for flounder, for which in rectangles located in the southern North Sea a corresponding decrease in both lymphocystis and skin ulcer disease was detected (see Figure A8.8). For dab, there is no clear indication of consistent temporal trends in larger areas comprising a number of neighbouring rectangles. However, it seems that the southeastern part of the North Sea is characterized by a more stable situation with respect to possible changes in the disease prevalence than the other areas.

However, the dab data also reveal the occurrence of some common temporal features. For instance, the prevalences of lymphocystis and epidermal hyperplasia/papilloma in dab seem to follow a similar temporal pattern in some areas (e.g., areas 34F3, 35F3, and 36F1) (see Figures A8.2 and A8.3). Furthermore, there are some common peaks in disease prevalence in the majority of areas in the period 1988–1992 for both dab and flounder (see Figures A8.2 to A8.6), indicating the presence of underlying mechanisms driving the development of all diseases in the same way, possibly linked to general ecosystem change.

In general, the analysis for temporal trends for the period 1992–1996/1997 showed that, when combining the disease data, stable to decreasing trends in the prevalence dominate (see Figures A8.7 and A8.8). The only disease which increased in a considerable number of rectangles is acute/healing skin ulcers of dab. This disease, however, occurs at a low prevalence in most areas.

In addition to the establishment of these general trends, the results of the analysis help to identify areas of concern which differ from other areas with respect to both the absolute disease prevalence levels and the temporal changes in disease prevalence and, therefore, deserve particular attention (examples: exceptionally high prevalence levels of acute/healing skin ulcers in dab from areas 37F2 and 38F2 since 1990, the increasing prevalence of epidermal hyperplasia/papilloma in dab from area 37F7 in the German Bight, and the increase in the prevalence of lymphocystis in Baltic flounder from area 38G5).

It must be emphasized, however, that the results of the analyses do not provide information on possible natural and/or anthropogenic causes of the observed spatial and temporal trends. This will only be possible when environmental parameters known or suspected to be involved in the disease aetiology (e.g., oceanographic, contaminant, fish stock assessment and fishery intensity data) are included in a holistic data analysis. Since these kinds of data are available in the different ICES data banks, it will be a desirable future task to assess the suitability of these data and, if considered appropriate and promising, to undertake a holistic data analysis providing information on cause/effect relationships between fish diseases and environmental factors.

8 ACKNOWLEDGEMENT

This material was prepared by W. Wosniok (Institute for Statistics, University of Bremen, Bremen, Germany), T. Lang (Federal Research Centre for Fishery, Institute of Fishery Ecology, Cuxhaven, Germany), A.D. Vethaak (RIKZ, Ecotoxicology Section, Middelburg, The Netherlands), S. des Clers (Ecosystem Analysis and Management Group, Biological Sciences, University of Warwick, Coventry, United Kingdom), and S. Mellergaard (Danish Institute for Fisheries Research, Charlottenlund, Denmark).

9 REFERENCES

Bucke, D., Vethaak, A.D., Lang, T., and Mellergaard, S. 1996. Common diseases and parasites of fish in the North Atlantic: Training guide for identification. ICES Techniques in Marine Environmental Sciences, No. 19. 27 pp.

Hastie, T.J., and Tibshirani, R.J. 1990. Generalized additive models. Chapman and Hall, London.

ICES. 1989. Methodology of fish disease surveys. Report of an ICES Sea-going Workshop held on U/F 'Argos' 16–23 April 1988. ICES Cooperative Research Report, No. 166. 33 pp.

ICES. 1997a. Report of the ICES/OSPAR Workshop on the Identification of Statistical Methods for Trend Detection. Annex 6. ICES CM 1997/Env:11, p. 32.

ICES. 1997b. Report of the Study Group on Statistical Analysis of Fish Disease Data in Marine Stocks. ICES CM 1997/F:3.

MathSoft Inc. 1997. S-PLUS 4.0 User's Guide. MathSoft Inc., Seattle, Washington.

McCullagh, P., and Nelder, J.A. 1989. Generalized linear models. Second edition, Chapman and Hall, London.