

Development of tools for logbook and VMS data analysis

Studies for carrying out the common

fisheries policy (No MARE/2008/10 Lot 2)

D.J. Beare, N.T. Hintzen, Francois Bastardie, G.J. Piet,
Clara Ulrich, Fabrizio Manco, Hans Gerritsen, Sebastian
Demaneche, Stephanie Mahevas, Henrik Degel, Josefine
Egekvist, Martial Laurans, Neil Campbell, Nicolas
Deporte, Rasmus Nielsen, Stuart Reeves, Colm Lordan,
Ryszard Grzebielec.

IMARES Wageningen UR

(IMARES - Institute for Marine Resources & Ecosystem Studies)

European Commission
Directorate General Fisheries
Mr. Joao Eliseu de Bianchi-Pinto
Rue de la Loi 200
B-1049 Bruxelles~
BELGIUM

<https://docs.google.com/?tab=mo&authuser=0>

1*	Wageningen IMARES	Institute for Marine Resources and Ecosystem Studies	The Netherlands
2	CEFAS	Centre for Environment, Fisheries & Aquaculture Science	England (UK)
3	IFREMER	Institut Français pour la Recherche et l'Exploitation de la Mer	France
4	DTU-AQUA	Technical University of Denmark, National Institute of Aquatic Resources	Denmark
5	FRS	Fisheries Research Service	Scotland (UK)
6	SFI	Sea Fisheries Institute	Poland
7	MI	Marine Institute	Ireland

* Coordinator

IMARES is:

- an independent, objective and authoritative institute that provides knowledge necessary for an integrated sustainable protection, exploitation and spatial use of the sea and coastal zones;
- an institute that provides knowledge necessary for an integrated sustainable protection, exploitation and spatial use of the sea and coastal zones;
- a key, proactive player in national and international marine networks (including ICES and EFARO).

P.O. Box 68
1970 AB IJmuiden

Phone: +31 (0)317 48 09 00

Fax: +31 (0)317 48 73 26

E-Mail: imares@wur.nl

www.imares.wur.nl

P.O. Box 77
4400 AB Yerseke

Phone: +31 (0)317 48 09 00

Fax: +31 (0)317 48 73 59

E-Mail: imares@wur.nl

www.imares.wur.nl

P.O. Box 57
1780 AB Den Helder

Phone: +31 (0)317 48 09 00

Fax: +31 (0)223 63 06 87

E-Mail: imares@wur.nl

www.imares.wur.nl

P.O. Box 167
1790 AD Den Burg
Texel

Phone: +31 (0)317
48 09 00

Fax: +31 (0)317 48
73 62

E-Mail:
imares@wur.nl

www.imares.wur.nl

© 2010 IMARES Wageningen UR

IMARES, institute of Stichting DLO is registered in the Dutch trade record nr. 09098104, BTW nr. NL 806511618

The Management of IMARES is not responsible for resulting damage, as well as for damage resulting from the application of results or research obtained by IMARES, its clients or any claims related to the application of information found within its research. This report has been made on the request of the client and is wholly the client's property. This report may not be reproduced

and/or published partially or in its entirety without the express written consent of the client.

A_4_3_2-V11.2

Contents

Executive Summary	6
1. Introduction	7
1.1 Context	
1.2 Analysis of logbooks data	7
1.3 Analysis of VMS data.....	8
1.4 Project structure	9
2. Compilation of national data sets into a single database	11
2.1 Introduction	11
2.2 Material and methods.....	11
2.2.1 EFLALO2	11
2.2.2 TACSAT 2	13
2.2.3 Fish frame	14
2.3 Conclusions	18
3. Distinction of métiers from disaggregated logbook data	19
3.1 Introduction	19
3.2 Methods and application	20
4. Distinction of fishing from other activities	30
4.1 Introduction	30
4.2 Methods	31
4.3 Applications	37
4.4 Conclusions	38
5. Linking VMS to logbook data.....	39
5.1 Introduction	39
5.2 Material and methods.....	39
5.2.1 Classification of VMS data into trips.....	39
5.2.2 Quality flags	39
5.2.3 Coupling VMS and logbook data	40
5.2.4 Uploading data to FishFrame	45
5.2 Applications	46
5.3.1 Visualisation in the data tree	47
5.3.2 Output	48
5.4 Conclusions	50
6. Evaluation of the effect of recording rate	51
6.1 Introduction	51
6.2 Material and methods.....	51
6.3 Applications	51
6.3.1 Based on position registrations	51
6.3.2 Based on trawl tracks.....	54
6.3.3 Interpolation algorithm.....	56

6.3.4 Effect of recording rate.....	57
6.4 Conclusions	58
7. Describing the spatial distribution of fishing activity	59
7.1 Introduction	59
7.2 Material and methods.....	60
7.3 Applications	60
7.4 Conclusions	67
8. Synthesis: <i>Vmstools</i> software package.....	68
8.1 Introduction	68
8.1.1 Data issues and regulations.....	68
8.2 Material and methods.....	69
8.3 Applications	70
8.3.1 Regional database (FishFrame).....	75
8.3.2 FishFrame performance survey	79
8.4 Conclusion.....	79
8.5 General Conclusions.....	80
9. Project achievements	82
9.1 Quick Deliverables checklist	82
9.2 Dissemination	87
9.2.2 Manuscripts	87
9.2.3 Workshops	87
9.2.4 Virtual 88	
10. References.....	89
11. Quality Assurance	92
Justification.....	93
Appendices.	94

Executive Summary

This report describes the development of standard protocols for simultaneous analyses of EU logbook and VMS data. The process begins with the construction of standardised data formats for logbook (EFLALO) and VMS (TACSAT). The software for analysing the data takes the form of a fully documented package called **vmstools**¹, built using the freeware package, R (<http://cran.r-project.org/>). Once the data have been imported into R in the correct format, a series of R programs or 'functions', linked by 'scripts' enable all tasks necessary to be completed in a single software environment.

The software can 'clean' data and format input data, estimate distances between VMS positions, and métiers can be identified objectively from species assemblages in catch data using multivariate statistical techniques. We have included a range of complimentary methods for determining fishing activity from VMS position registrations. Positions at sea, for example, can be distinguished from vessels in harbour or erroneous positions on land. Position registrations of vessels actually fishing can be separated from those engaged in other activities (e.g. steaming) using their speed in conjunction with other information such as vessel size and gear being used. Logbook and VMS data can be merged such that high-resolution spatial maps of catches of various commercial species can be generated. Individual vessel tracks can be reconstructed for more realism through different interpolation techniques (both linear and non-linear, i.e. using Hermite spline functions). Further, all the fishing activity indicators required under the Data Collection Framework can be calculated using **vmstools**. The package can also be used to explore the impact of different spatial (grid size) and temporal aggregations (month, quarterly, annual) which need to be explicitly considered when assessing fishing impact on the seafloor. There are also scripts for displaying results using *Google Earth* which is a useful aid for dissemination. The combination of all these routines 'under one roof' then permits the construction of 'Regional' databases (i.e. FishFrame a regional database hosted by one of the current project partners) and scripts to produce output suitable for this are included with the **vmstools** package. As proof of concept, all analyses performed within each work package have been tested, using the **vmstools** package, against national datasets with contributions from the French, Danish, Irish, UK and Dutch institutes. As an example, FishFrame has been populated with Dutch and Danish combined VMS and logbook data for 2005-2009. The project demonstrates emphatically that logbook and VMS data from disparate countries with often different data collection regimens can be combined and compared using generic tools and that the output can be sent to regional databases permitting more holistic assessments of fishing activity.

The project has generated considerable interest in the marine/fisheries scientific community and the tools are already being used to advise on marine spatial management, e.g. construction of wind farms and the instigation of protected areas.

¹ The most recent version of **vmstools** can be downloaded from:
<http://code.google.com/p/vmstools/downloads/list>

1. Introduction

1.1 Context

The Common Fisheries Policy (CFP) requires the progressive implementation of an ecosystem-based approach to fisheries management (EBFM). This will include the integration of environmental protection requirements into the CFP, including measures to 'limit the environmental impact of the CFP'. To that order the collection of data that describe the impact of fishing by the European Member states is required to fulfil the obligations defined; amongst others under the current Data Collection Framework. Currently, two main sources of data that describe the impact of fishing activities on the ecosystem can be distinguished: logbook and VMS (Vessel Monitoring through Satellite).

As part of the revision of the CFP and implementation of EBFM, the need to move toward a fleet and area-based management (Council Regulation N° 199/2008²) as well as the development of indicators of fishing impact were identified. Both topics require additional and more detailed information on the fishing activities of the fleets such as can be obtained from logbook and VMS data.

1.2 Analysis of logbooks data

In order to operationalize fleet-based management more detailed information is needed on the fishing activities of the fleets as foreseen in the forthcoming Data Collection Framework (DCF)³. However, in order for these data to be useful for advice it is extremely important to assure that each member state uses consistent procedures for allocating trips to métiers (groupings of fishing operations of similar exploitation pattern). The collection and the analyses of logbook data will then have to be adapted and will have to consider the métier and fleet segments (groupings of fishing vessels) as recommended by the experts^{4,5,6}. This, in turn necessitates the development of common agreed methods and tools in order to be able to categorize the information provided by the logbooks into métier. Taking into account the information at the basic level of the fishing operation ensures that the outcome of this project will be compatible with any further development in the precise reporting of catches, such as electronic logbooks.

Logbook data provide information on the quantities of the main species caught and kept on board; the related date; location (ICES statistical rectangle); and type of gear used. The new DCF will follow a fleet-based approach, which implies that fishing operations and fishing trips showing similar exploitation patterns can be grouped by métier for the purposes of the new regulation, allowing more accurate estimates of fishing impact. The information currently available in the logbooks does not contain the information on (assemblage of) targeted species as required by the definition of the métier (see chapter 4). The targeted species, which is key information, must then be derived from all the other variables reported using a precise methodology. The objective of this study is to develop a tool allowing the classification of every trip into métiers under the new DCF fleet based approach on a regional level. This

²<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:060:0001:01:EN:HTML>

³ Proposal for a Council Regulation concerning the establishment of a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy (COM (2007)196 final)

⁴ Commission Staff Working Paper: Report of the Ad Hoc Meeting of the independents experts on the Fleet-Fishery based sampling, Nantes, France, 23-27 May 2005, 34p.

⁵ Commission Staff Working Paper: Report of the training Workshop on Fleet-based Approach, Nantes, France, 13-17 March 2006, 31p.

⁶ Commission Staff Working Paper: Report of the Ad Hoc Meeting of the independents experts on the Fleet-Fishery based sampling, Nantes, France, 12-16 June 2006, 98p.

tool will be the property of the European Commission and is to be used freely by Member states when implementing their obligations regarding the DCF. To this end the study shall:

1. Develop procedures based on statistical methods to allocate trips described in logbooks to métiers.
2. Test the above procedures to assure that the allocations of trips to métiers are consistent among Member states.
3. Develop follow-up methods to assure that the allocation procedures are consistent along time.
4. Describe a generic method to automatically run the above procedures, suitable to be implemented by each Member state.

The algorithms and methods will be designed in view of filling the métier/fleet matrices (Annex 1) as developed by EU experts and finally proposed by STECF/SGRN, in November 2006, for implementation in the next DCF. An important issue raised concerns the discrimination between métiers targeting one group of species and métiers targeting a mixture of species. As in a certain number of cases, the choice is given to report both the single and the mixed species, a special attention will be given to set up clear guidelines and allocation rules. The double dimension of the DCF matrix will be addressed, as methods for allocating vessels to fleet segments will be considered. Ultimately, the general philosophy put forward by STECF/SGRN (November 2006) will be used as a guideline to the project:

- all fishing practices are meant to be registered in the matrix *"the new DCF should address all removals from fish and shellfish stocks, regardless who or what is at their origin"*,
- a fleet/métier approach can only be defined on a given area and on a given time period, all fishing activities are reported, *"the collection of fishery-related data should comprise all fishing activities that cause such removals"*
- the elementary unit for defining the métier is the fishing operation, *"grouping fishing operations into strata with identical features"*
- the reporting of information in the matrix is done with common rules at a regional level *"the métiers attempt to harmonize the stratification of fishing operations at the regional level"*

1.3 Analysis of VMS data

When implementing an Ecosystem Approach to Fisheries Management (EAFM), indicators are required to describe the pressures affecting the ecosystem, the state of the ecosystem, and the response of managers (Jennings 2005). Such indicators can be used to support management decision making, track progress towards meeting management objectives and to aid communication with non-specialist audiences ([Garcia et al. 2000](#), [Rice 2000](#), [Rochet & Trenkel 2003](#)). Many indicators have been proposed (e.g. [Rice 2000](#), [Link 2002](#), [Link et al. 2002](#), [Rochet & Trenkel 2003](#)), but few (if any) of those that track changes in the "state" of the marine environment or of different ecosystem components (e.g. fish, benthos, habitat) can support management directly ([Rice 2000](#)). This is largely because the precise causes of any changes in "state" may be poorly understood, making it difficult to identify appropriate management action. To implement an EAFM successfully therefore, it is not only necessary to have a suite of indicators that accurately and comprehensively portray the "state" of various ecosystem components, but it is also critical that we have indicators that describe changes in the level of different manageable anthropogenic activities, and which indicate the impact of each activity on the various ecosystem components. Only by adequately covering both aspects will the mechanistic links between "cause" and "effect" be well enough understood so as to provide the advice required ([Daan 2005](#)).

In order to link the state of ecosystem components such as habitat, benthos or the demersal fish community to fishing, the following pressure indicators were suggested (EC 2008/MARE/020):

- Distribution of fishing activities: Indicator of the spatial extent of fishing activity. It would be reported in conjunction with the indicator for 'Aggregation of fishing activity'.
- Aggregation of fishing activities: Indicator of the extent to which fishing activity is aggregated. It would be reported in conjunction with the indicator for 'Distribution of fishing activity'.

- Areas not impacted by mobile bottom gears: Indicator of the area of seabed that has not been impacted by mobile bottom fishing gears in the last year. It responds to changes in the distribution of bottom fishing activity resulting from catch controls, effort controls or technical measures (including MPA established in support of conservation legislation) and to the development of any other human activities that displace fishing activity (e.g. wind farms).

In order to quantify these indicators there is a need for high-resolution spatial data of fishing activity such as provided by VMS. With these data the study shall:

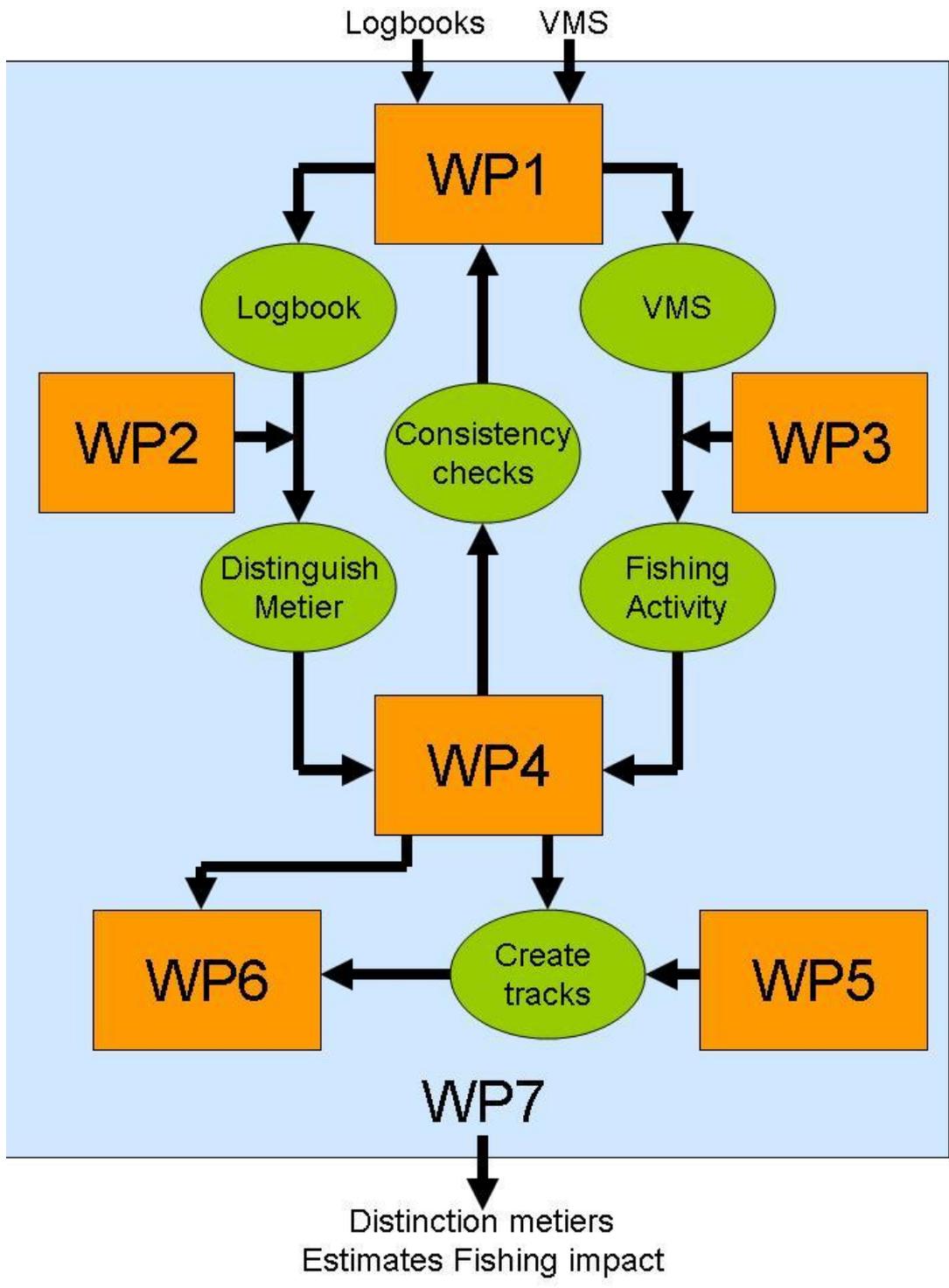
1. Develop and test methods and produce protocols on how to link VMS databases to Logbooks and calculate these indicators using GIS.
2. Model the dependence of recording rate on the precision of the suggested indicators

1.4 Project structure

The project consists of seven work-package structured such that in each of the work-packages tools *and/or protocols are developed that contribute to the process that starts with two sources of data: logbooks and VMS should result in a classification of trips into métiers based on the species composition* in the logbooks as well as the development and estimation of pressure indicators in support of an ecosystem approach to fisheries management. The aim is to apply all the tools that will be developed in this process on as many data sets across different European waters. The tools together with protocols describing their use will become available in the public domain. The project structure and brief description of work-packages below should explain how the different work-packages fit into this process.

- In WP1 the various data sets of logbooks and VMS will be compiled separately into common formats and stored into one logbook and one VMS database.
- In WP2 tools will be developed that can classify trips into métiers based on the species composition in the logbooks.
- In WP3 tools will be developed that can distinguish fishing from other activities using VMS data.
- In WP4 tools will be developed that combines matching logbook and VMS recordings with the aim of checking the consistency of the two data-sources (and feeding back into WP1) but also to allow each source to benefit from the additional information provided by the other source. For example this should allow VMS trips to be allocated to métiers or to allocate catches to the actual fishing positions at a higher spatial resolution than ICES rectangle. The results of the consistency checks flow back to WP1.
- In WP5 the effect of the VMS recording rate will be evaluated and to that end a tool that reconstructs the fishing track from the VMS positions will be developed.
- In WP6 a suite of pressure indicators in support of an ecosystem approach to fisheries management will be developed and estimated based on logbook, VMS data and the reconstructed tracks coming from WP5.
- Finally in WP7 the tools en knowledge generated in the previous work-packages are integrated into one package "**vmstools**" consisting of software routines that can be applied to any regional dataset in the common format.

These work-packages are described in each of the subsequent chapters of the report (i.e. work-package 1 becomes chapter 2,..... until work-package 7 becomes chapter 8)



2. Compilation of national data sets into a single database

2.1 Introduction

In this chapter is described how common data formats were created which the **vmstools** package exploits. Due to confidentiality issues, neither raw VMS nor logbook data could be stored in one multinational database during the project (see below). Instead, data had to be stored nationally in common exchange formats such as ASCII text files. Common data exchange formats for logbook and VMS data have already been developed during other EU funded projects (e.g. 'TECTAC' and 'CAFÉ') called 'EFLALO' (logbook data) and 'TACSAT' (VMS data) and the current project team agreed to use them as a starting point. The motivation for choosing these particular formats is that many Fisheries Institutes around Europe have 'expertise' in extracting such data from their respective national databases. Ultimately these 'formats' were modified slightly in order to include all the information needed, while data irrelevant to this project were excluded (see Tables 1.1 and 1.2). A further requirement for work package 1 was to create a common exchange database format into which aggregated VMS and logbook data could be inserted. The regional database, FishFrame (<http://www.fishframe.org>) hosted by DTU-AQUA was chosen for this purpose and the format required described below and in Appendix III.

2.2 Material and methods

2.2.1 EFLALO2

Logbook data are stored in the EFLALO2 format (Appendix I). Originally EFLALO was aggregated on a trip by trip basis. For combining logbook data with VMS data, the fishing date and other data from the logbooks is also useful, however, and was included. Sometimes within a trip a vessel may change gears or mesh size, for example, and target another species and this can be important. Therefore, the "log event" concept was introduced which is defined when there is a change of date, gear, area or haul during a trip. The format is described in Table 2.1.1 The data should be saved in a .csv format (comma separated file) and uploaded to **vmstools** using `readEflalo()` (see **vmstools** manual). The `formatEflalo()` function ensures that the data is formatted correctly. When information is missing, an NA field can be inserted between the commas. In some countries such as Denmark, haul information is present in the logbooks, whereas in other countries (e.g. the Netherlands) it is absent. Therefore, haul information was made optional and **vmstools** can cope with either. This slightly modified EFLALO format was named EFLALO2.

Table 2.1.1 Table showing column definitions for the EFLALO2 format

Type	Variable	Code	Format/Unit
Vessel	Vessel ID	VE_REF	20 character string
	Fleet	VE_FLT	DCF regulation
	Home country	VE_COU	ISO 3166 – 1 alpha-3 codes.
	Vessel length	VE_LEN	Oal (m)
	Vessel power	VE_KW	kW
	Tonnage	VE_TON	GT (<i>optional</i>)
Fishing trip	Fishing trip reference number	FT_REF	20 character string
	Departure country	FT_DCOU	ISO 3166 – 1 alpha-3 codes.
	Departure harbour	FT_DHAR	International harbour codes. UN LOCODE
	Departure date	FT_DDAT	DD/MM/YYYY
	Departure time	FT_DTIME	HH:MM
	Landing country	FT_LCOU	ISO 3166 – 1 alpha-3 codes.
	Landing harbour	FT_LHAR	International harbour codes. UN LOCODE
	Arrival date	FT_LDAT	DD/MM/YYYY
	Arrival time	FT_LTIME	HH:MM
Log event	Log event ID	LE_ID	25 character string FT_REF_number (1,2,3,etc.)
	Catch date	LE_CDAT	DD/MM/YYYY
	Log event start time	LE_STIME	HH:MM (<i>Optional</i>)
	Log event end time	LE_ETIME	HH:MM (<i>Optional</i>)
	Log event start position latitude	LE_SLAT	Decimal degrees (<i>Optional</i>)
	Log event start position longitude	LE_SLON	Decimal degrees (<i>Optional</i>)
	Log event end position latitude	LE_ELAT	Decimal degrees (<i>Optional</i>)
	Log event end position longitude	LE_ELOK	Decimal degrees (<i>Optional</i>)
	Gear	LE_GEAR	3 character string. DCF métier level 4
	Mesh size	LE_MSZ	mm stretched mesh
	ICES rectangle	LE_RECT	37F5, NA=unallocated
	ICES division	LE_DIV	10 character string (see codes in annex 1)
	Fishing activity (métier)	LE_MET	FishFrame level 6 code
	Fishing activity from cluster analysis	LE_MET_CLUST	
	Landing weight estimate of species SP1 (FAO species codes)	LE_KG_<SP1>	Kg
	Landing value of species SP1 (FAO species codes)	LE_EURO_<SP1>	EURO

	Landing weight estimate of species SPn (FAO species codes)	LE_KG_<SPn>	Kg
	Landing value of species SPn (FAO species codes)	LE_EURO_<SPn>	EURO

At the vessel level, information relating to the vessel is stored, e.g. Vessel ID, fleet, home country, vessel length, vessel power (kW) and tonnage. At the fishing trip level, the information relating to the fishing trip is stored, e.g. fishing trip reference number, departure country, departure harbour, departure date, departure time, landing country, landing harbour, arrival date and arrival time. At the log event

level information relating to the log event is stored, e.g. catch date, log event start time, log event end time, sequence number, gear, mesh size, ICES rectangle, fishing activity (métier), landing weight, and landing cash value of all species. The 'fishing activity field' is derived from the output of work-packages 2 and 3 (chapters 3 and 4 of this report).

In table 2.1.2 below a short example of an EFLALO dataset is shown.

Table 2.1.2. Example of EFLALO dataset.

VE_REF	VE_FLT	VE_COU	VE_LEN	VE_KW	VE_TON	FT_REF	FT_DCOU	FT_DHAR	FT_DDAT	FT_DTIME
238	TBB	Atlantis	24	221	NA	271253	nld	SCH	01/05/1800	8:00:00
731	TBB	Atlantis	23	221	NA	271268	nld	SL	01/05/1800	15:00:00
742	TBB	Atlantis	22.0728	221	NA	271270	nld	SCH	01/05/1800	4:00:00
742	TBB	Atlantis	22.0728	221	NA	271270	nld	SCH	01/05/1800	4:00:00
873	TBB	Atlantis	23	221	NA	271285	nld	SCH	01/05/1800	7:00:00
1505	TBB	Atlantis	25	221	NA	271507	nld	IJM	01/05/1800	21:00:00
FT_LCOU	FT_LHAR	FT_LDAT	FT_LTIME	LE_ID		LE_CDAT	LE_GEAR	LE_MSZ	LE_RECT	LE_DIV
nld	SL	01/05/1800	23:00:00	271253-TBB-32F3		01/05/1800	TBB	80	32F3	IVc
nld	SL	02/05/1800	7:00:00	271268-TBB-33F2		01/05/1800	TBB	80	33F2	IVc
nld	SL	02/05/1800	4:00:00	271270-TBB-32F3		01/05/1800	TBB	80	32F3	IVc
nld	SL	02/05/1800	4:00:00	271270-TBB-33F4		01/05/1800	TBB	80	33F4	IVc
nld	SL	01/05/1800	15:00:00	271285-TBB-33F4		01/05/1800	TBB	80	33F4	IVc
nld	WR	02/05/1800	10:00:00	271507-TBB-34F4		02/05/1800	TBB	80	34F4	IVc
LE_MET_level6		LE_KG_BLL	LE_KG_COD	LE_EURO_BLL	LE_EURO_COD					
TBB_DEF_70-99_0_0		0	0	0	0					
TBB_DEF_70-99_0_0		2.15429	7.938833	20.09722	24.05407					
TBB_DEF_70-99_0_0		0	0	0	0					
TBB_DEF_70-99_0_0		9.980239	39.889678	101.03526	125.86955					
TBB_DEF_70-99_0_0		0	0	0	0					
TBB_DEF_70-99_0_0		20.130751	0	202.2027	0					

2.2.2 TACSAT 2

The group agreed that VMS data would be stored in the TACSAT2 format (Table 2.2.1). It includes the following variables: vessel id, latitude, longitude, date, time, instantaneous speed delivered and instantaneous heading delivered. In order to know if a point is inside a harbour (and therefore not relevant) an extra variable can be added (SI_STATE) which indicates if the position is in harbour or at sea. Again, the data should be saved in a .csv format (comma separated file) and can be read into **vmstools** using `readTacsat()` (see Appendix I) where the `formatTacsat()` function ensures that the data is formatted correctly. When information is missing, an empty zero-length string should be added between the commas.

Table 2.2.1 TACSAT2 format (VMS data)

Type	Variable	Code	Unit
Vessel	Vessel ID	VE_REF	20 character string
Sighting operation	Latitude	SI_LATI	Decimal degrees
	Longitude	SI_LONG	Decimal degrees
	Date	SI_DATE	DD/MM/YYYY
	Time	SI_TIME	HH:MM
	Instant speed delivered	SI_SP	Knots
	Instant heading delivered	SI_HE	Degrees
	At Sea/In Harbour	SI_HARB	0: In harbour 1: At sea
	Fishing/Steaming	SI_STATE	0: Steaming 1: Fishing
	Fishing trip reference (FT_REF)	SI_FT	20 character string

Note: an 'example' dataset, loosely based on Dutch data, is available in the **vmstools** package for which permission for common use was gained from relevant fishing industry representatives (See Appendix VI). An example TACSAT dataset is given below in table 2.2.2

Table 2.2.2. Example TACSAT dataset.

VE_COU	VE_REF	SI_LATI	SI_LONG	SI_DATE	SI_TIME	SI_SP	SI_HE
Atlantis	10	51.44772	3.583731	04/05/1800	13:32:00	0	25
Atlantis	10	51.44067	3.583847	04/05/1800	15:28:00	0	25
Atlantis	10	51.44074	3.595529	04/05/1800	17:22:00	0	25
Atlantis	10	51.44315	3.586862	04/05/1800	19:18:00	0	25
Atlantis	10	51.44351	3.586948	04/05/1800	21:12:00	0	25
Atlantis	10	51.44811	3.593958	04/05/1800	23:08:00	0	25

2.2.3 Fish frame

The FishFrame format (see Tables 2.3 and 2.4 below) is described in more detail in Appendix III and there is no need to describe it further here. Once EFLALO and TACSAT data have been extracted FishFrame compatible data can be created with **vmstools** using the functions *mergeEflalo2pings()*, *bindAllMergedTables()*, and *pings2Fishframe()* (see pages 58 – 60 of Appendix I for a complete description of the process).

Table 2.3. Commercial fisheries effort statistics record (VE) based on VMS data

Order	Name	Type	Req.	Basic checks	Comments
1	Record type *	String	M		Fixed value VE
2	Vessel Flag Country *	String	M	Code list	ISO 3166-1 alpha-3 codes. The flag country of the vessel.
3	Year *	Integer	M	Code list	1900 to 3000
4	Quarter *	Integer	M	Code list	1 to 4.
5	Month *	Integer	M	Code list	1 to 12
6	Area*	String	M	Code list	Area level 3 (level 4 for Baltic, Mediterranean, Black Sea) in the data Collection regulation (EC, 2008a, 2008b).
7	C-square *	String	M	Code list	0.05x0.05 degree, C-square reference XXXX:XXX:XXX:X
8	Fishing activity category National *	String	O	Code list	Fishing activity category – <i>National</i> coding system. Bound to the Nantes matrix level 4 as children i.e. an alternative level 5+6.
9	Fishing activity category European level 6 *	String	M	Code list	Fishing activity category – <i>Level 6</i> in the Nantes matrix (SGRN 06-03)
10	Fishing hour	Decimal numeral	M	1 to 9999999999	Fishing hour calculated from VMS data.
11	kW*fishing hour	decimal numeral	M	1 to 9999999999	
12	Tot weight	Decimal numeral	M	1 to 9999999999	Total landings of all species caught. In kg
13	Tot value	Decimal numeral	M	1 to 9999999999	Total value of all species caught. In Euro

* = The field is a key field.

The references to code lists are the lists in the regular FishFrame exchange format description (ICES Cooperative Report 296, 2009)

Table 2.4. Species List record (VSL) in commercial fisheries based on VMS and logbook data

Order	Name	Type	Req.	Basic checks	Comments
1	Record type *	String	M		Fixed value VSL.
2	Vessel Flag Country *	String	M	Code list	ISO 3166-1 alpha-3 codes. The flag country of the vessel.
3	Year *	Integer	M	Code list	1900 to 3000
4	Quarter *	Integer	M	Code list	1 to 4.
5	Month *	Integer	M	Code list	1 to 12
6	Area*	String	M	Code list	Area level 3 (level 4 for Baltic, Mediterranean, Black Sea) in the data Collection regulation (EC, 2008a, 2008b).
7	C-square *	String	M	Code list	0.05x0.05 degree, C-square reference XXXX:XXX:XXX:X
8	Fishing activity category National *	String	O	Code list	Fishing activity category – <i>National</i> coding system. Bound to the Nantes matrix level 4 as children i.e. an alternative level 5+6.
9	Fishing activity category European level 6 *	String	M	Code list	Fishing activity category – <i>Level 6</i> in the Nantes matrix (SGRN 06-03) see appendix I.
10	Species	String	M	Closed list of species ¹⁾	Scientific name in Latin (<i>genus species</i>).
11	Weight	Integer	M	1 to 9999999999	Whole weight in gram. Decimals not allowed.
12	Value	Integer	O	1 to 9999999999	In Euro. Decimals not allowed.

* = The field is a key field.

Other codifications

The codifications for countries, areas, harbours, gears, and species are as follows (see also Appendices for more details). Country codes are ISO 3166 – 1 alpha-3 codes. Area codes are described in Table 2.5. codes use the International harbour codes based on the UN LOCODE format. These codes and harbour positions have been collected for the EU project ERS (Electronic Reporting System), and are available on the webpage: http://ec.europa.eu/fisheries/cfp/control_enforcement/ers_en.htm. Datasets for the locations of all EU harbours have been included in **vmstools**, type *data(harbours)* in R after installing the **vmstools** package. Throughout the project we have used FAO gear codes and species codes: <ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/AnnexM1fishinggear.pdf> <http://www.fao.org/fishery/collection/asfis/en>. Fishing activity codes used in FishFrame for Nantes matrix level 6 are used (see Appendix IV for details).

Table 2.5. Area codes used during the project

Code	Description
2a	Norwegian Sea (Division IIa)
2b	Spitzbergen and Bear Island (Division Iib)
3an	Skagerrak (Division IIIa)
3as	Kattegat (Division IIIa)
3b	Sound or the Transition Area (Divisions IIIb)
3c	Belt Sea or the Transition Area (Divisions IIIc)
3d	Baltic Sea (Division IIIId)
4a	Northern North Sea (Division IVa)
4b	Central North Sea (Division IVb)
4c	Southern North Sea (Division IVc)
5a	Iceland Grounds (Division Va)
5b	Faeroes Grounds (Division Vb)
6a	Northwest Coast of Scotland and North Ireland or as the West of Scotland (Division VIa)
6b	Rockall (Division VIb)
7a	Irish Sea (Division VIIa)
7b	West of Ireland (Division VIIb)
7c	Porcupine Bank (Division VIIc)
7d	Eastern English Channel (Division VIId)
7e	Western English Channel (Division VIIe)
7f	Bristol Channel (Division VIIf)
7g	Celtic Sea North (Division VIIg)
7h	Celtic Sea South (Division VIIh)
7j	Southwest of Ireland - East (Division VIIj)
7k	Southwest of Ireland - West (Division VIIk)
8a	Bay of Biscay - North (Division VIIa)
8b	Bay of Biscay - Central (Division VIIb)
8c	Bay of Biscay - South (Division VIIc)
8d	Bay of Biscay - Offshore (Division VIIId)
8e	West of Bay of Biscay (Division VIIe)
9a	Portuguese Waters - East (Division IXa)
9b	Portuguese Waters - West (Division IXb)
14b	Southeast Greenland (Division XIVb)

2.3 Conclusions

The project was successful in adopting and further developing two useful exchange formats for raw VMS and logbook data (see Tables 2.1 and 2.2) and we have written scripts that uses them for all manner of useful output. All partners were capable of creating such data but the confidentiality issues made 'sharing' of raw data problematic. Data were submitted to FishFrame during the project and this process is on-going.

3. Distinction of métiers from disaggregated logbook data

3.1 Introduction

The goal of WP 2 was to investigate a range of multivariate statistical methods and to develop a tool for selecting the most appropriate one for assigning one or several métier(s) (as defined in the Data Collection Framework regulation, DCF) to each logbook data. A métier is a group of fishing operations targeting a similar (assemblage of) species, using similar gear types, during the same period of the year and/or within the same area and which are characterized by a similar exploitation pattern. The DCF has thus defined métiers according to a hierarchical structure using six nested levels (EC 2008) from level 1- Activity (fishing/non fishing) to level 5- EU level gear type (e.g., Bottom Otter Trawl OTB, Bottom Pair Trawl PTB) combined with Target assemblages of species (e.g., Demersal fish, Crustaceans). Level 6 includes information about mesh size class and presence/absence of a selective device. In WP2, we thereby refer to a more precise target assemblage defined as the true species level in opposition to the DCF Level5 dealing with assemblage of species only (DCF Level7 suggested in EC 2008 as national strata). The tool developed is directly incorporated within the **vmstools** package and allows the definition of métiers at DCF Level7 and their subsequent aggregation into DCF Level5.

Logbook data are the main source of data for getting information on aspects of fishing activities. From logbook data, information of EU level gear type (DCF Level4 e.g., Bottom Otter trawl OTB, Bottom pair trawl PTB) and selectivity's of the gear are directly available. However, fishers do not declare which species they are actually targeting when fishing, and therefore the DCF Level5 and the more precise DCF Level7, expected to reflect the fishing intention, has to be inferred from the available data; and in particular the landings data.

In WP2 a tool is developed that classifies trips (more precisely, logbook events defined in WP1) into métiers based on the species composition from the logbook landings declaration. The detailed objectives in this WP were to:

- Evaluate data quality and remove erroneous data.
- Identify a set of appropriate multivariate statistical methods and appropriate evaluation criteria.
- Develop corresponding scripts to apply these methods to international data of the format described in WP1, and run the analyses within this work-package on the available national data sets.
- Evaluate the sensitivity of results to the multivariate method used for different national datasets.
- Propose a final robust method for métier identification that is consistent across national data sets.
- Translate the results obtained from this method into fixed allocation rules, linking each fishing trip to a single métier (at DCF level 5).

The end product of this work package is a complete set of R scripts (available in the **vmstools** package; see also the manual in Appendix I) which provide a transparent operational tool that enabling the classification into métiers from a range of pre-selected methods and specific settings; and all starting from the data formats defined in the WP1(EFLALO). These tools also enable sensitivity analyses of métier definitions to be done along a variety of previously used classification methods and the selection of the most robust method based on objective statistical criteria. The tools also allow the user to predict the métier of a new dataset and to link the results obtained at DCF Level7 to the required DCF Level5. These routines are also accompanied with several documenting files which detail the specific underlying assumptions when applying them (see the manual in Appendix I).

WP2 consist of seven deliverables, described below in seven methodological sections.

3.2 Methods and application

Section 1: Criteria for each statistical method.

Since the seminal works of Biseau & Gondeaux (1988) and Laurec et al. (1991), the classification of heterogeneous fishing activities has been quite thoroughly investigated by the scientific community. A preliminary review of statistical methods for the classification of trips into métiers based on the landings per species declared in the logbooks was done (see references below). Two different approaches were used:

- i) The first approach is quantitative and consists of conducting multivariate analyses on the species composition of catch data (referred as *landings profiles*) per trip, logbook event or fishing operation and then groups similar landings profiles into métiers.

Principal Component analysis (PCA) combined with a Hierarchical Agglomerative Clustering (HAC) are the most used methods (Biseau and Gondeaux 1988, Laurec et al. 1991, Pelletier and Ferraris 2000, Ulrich and Andersen 2004, Hutton et al. 2004, Holley and Marchal 2004). But within this approach, a large number of statistical methods, settings and software have been used: Multiple Correspondence Analysis (MCA), Detrended Correspondence Analysis (DCA) (Rogers and Pickitch 1992), K-means algorithm (He et al. 1997, Bastardie et al. 2010), Partitioning Around Medoids (PAM Algorithm) (IBERMIX project, 2007) and its variant Clustering LARge Applications (CLARA Algorithm) (Punzon et al. 2010). These alternative methods to identify métiers raise a number of questions. First, it should be underlined that even with statistical methods, métier classification is somehow still subjective, while the robustness of the results to the methodological choices is largely unknown.

- ii) The second approach is more *ad hoc* and requires a qualitative *a priori* knowledge on the fisheries. It is largely based on a trial and error process.

The objective of WP2 was to design operational and objective tools as far as possible i.e. based only on various statistical criteria when carrying out the entire procedure defining the métiers. Based on this, this second approach was not included in the tool. However, "expert" knowledge (e.g. expertise from the scientists and information obtained from fishermen) still play an important role by judging the quality and the meaning and the relevance of the métiers clusters identified. For example, proposal could be done for changing the method-specific settings to increase the quality of the clustering obtained when, for example, a cluster gathers two important species which are known not to be caught together, or when small but very specific fisheries has to be identified or in case of clusters which seem redundant.

Section 2: List of criteria to evaluate the reviewed statistical methods with respect to the project objectives.

Criteria being investigated were:

- **Input data format.** The selected method must be applicable to the available logbooks data format (or the data format must be easily converted to be used with, see WP1 for EFLALO data format).
- **Availability of algorithm.** It is desirable to use a well-known method for which the algorithm has been implemented and validated, and for which the software is available free of charge.
- **Relevance.** A review of method application in published analyses to assess adequacy for the project objectives
- **Stability of the results.** The results should be replicable.
- **Consistency over time.** In case the method is calibrated over a calibration period, this must be robust enough to be not modified when the calibration period is extended.

- **Sensitivity to calculation method.** Methods where the results are insensitive to the calculation method will be favoured.
- **Robustness of thresholds.** The methods selected must stay within the robustness thresholds.
- **Validation against expert knowledge.** The validity, the meaning and the relevance of obtained métiers clusters resulting from the analyses must be confronted with previous knowledge on the concerned fisheries.

All the methods listed in section 1 have been tested and, at the end, based on objective statistical criteria (see below) the tool developed includes the following: a PCA analysis combined with clustering analysis using either HAC, or K-Means or CLARA algorithm.

The tool also enables the output of these analyses to be further compared with the simple and empirical ordination methods allowing the allocation of each logbook event to a DCF Level5 métier based on the first assemblage species in cash value: methods generally agreed at that time by EU member states (see Council Regulation (EC) No 1543/2000, Report of the 4th Liaison Meeting).

Section 3: Description of selected multivariate statistical methods; required data and format; assumptions

The EFLALO format defined in WP1 is used to perform the analyses throughout. This format standardizes national logbook information using common heading names. Each row represents a logbook 'event' (logbook events are sub-sequences within a trip characterized by changes of catch date, area or gear). Analyses have been made at this level of disaggregation implying that a vessel may potentially be engaged in several métiers within a trip. Columns include several trip descriptors, as well as all information related to landings (kg) and cash values of landings by species (information of values is usually derived from a cross checking with the corresponding auction sales). Following, the ICES (2003) recommendation, analyses have been done on landings composition expressed as cash value as this may reflect more accurately the actual targeting choices of the fishermen.

As the DCF aims at unifying métiers definitions from the national scale to the regional basin level, the tool was developed to perform analysis at the regional scale to take into account the consistency across neighbouring fisheries. This raised new computational issues due to larger datasets.

The choice was made to perform analysis at the EU level gear type level (DCF Level4 e.g., Bottom Otter Trawl OTB, Bottom Pair Trawl PTB) as this level could be derived directly from logbooks data.

The work consisted in designing several sequential steps to have a complete tool firstly allowing identification of métiers from raw logbook data and then assigning a métier to a new dataset.

These sequential steps aim at:

i) Identifying the main species out of all species recorded and reducing the dataset to these key species only (**Step 1**). The species were first selected by the application of three different filtering methods provided that species which succeed in at least one were kept for the further analyses. This conservative choice has been made in order to insure that no relevant species were left aside. Consequently, the initial dataset (logbook event x all species) was then reduced to the main species only.

ii) Investigating the gain when running a PCA on the quality of the final classification (PCA is likely to discard potentially irrelevant information) (**Step 2**). A number, but not all of the studies cited above, used a PCA prior to the clustering, and it was also decided to investigate the relevance of this choice. There are two reasons for applying a PCA. First, it helps reducing the multi-dimensional landings matrix to a smaller number of informative components represented by the first axes of the PCA transformation. Second, it enlightens the interactions among species across logbook events.

iii) Running a selection of clustering methods and settings to group of similar logbook events (**Step 3**); including the following: a PCA analysis combined with clustering analysis using either HAC, or K-Means or CLARA algorithm.

iv) Defining the characteristic species-based assemblage of the cluster (DCF Level7) and converting them to a corresponding DCF Level5 category (**Step 4**);

v) Comparison with simpler ordination method (**step5**);

vi) Predicting the classification of any new logbook event into the categories of defined métiers (**Step 6**).

Furthermore, the developed tool allows the automatic generation of a number of plots for investigating the relevance of the choices and settings (see Figures section 5).

A scientific paper (see the draft manuscript in Appendix IX) is being written to describe in more detail the entire workflow (incorporated in **vmstools**) for assigning métiers to each logbook event and testing and discussing the robustness of this classification illustrated from a regional 'test' dataset (see section 5). Each of these steps is summarized on Figure 1.

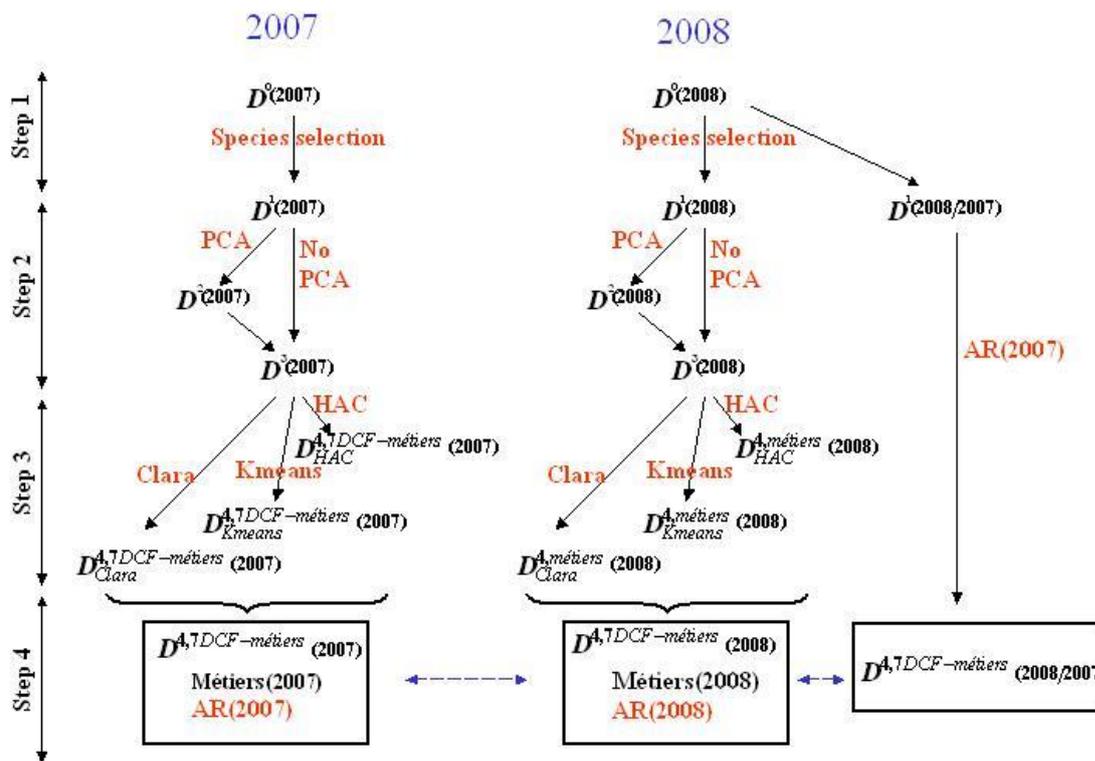


Figure 3.1: Flowchart of the analysis. Dotted arrows show comparisons performed between 2007 and 2008 outcomes.

Section 4: Evaluation of dataset quality in relation to the methods to be used, for each regional case study.

In October the project team held a workshop in order to do analyses at regional scales based on the results of WP1 (EFLALO format). The results obtained at this workshop have been an 'evaluation of dataset quality', and, at regional scale, the first outcomes obtained for the different methods selected.

Good quality of logbook data means among others (e.g. good reference for defining species caught and gear used, completeness of the data available, etc.), that the logbook event is the informed level of disaggregation (as defined in the WP1 for the EFLALO format) which is a preliminary for the procedure for assigning correctly one or several métier(s) to each fishing trip. On the tested data set, some of the logbook of particular countries had not yet this required level of disaggregation available (i.e. they only have the data informed at the fishing trip level) which could be problematic for vessels taking part in more than one métier within a same fishing trip.

Beside this, the size of the data sets (usually large when combining several countries) has proven to be the main limiting factor when applying these methods, given the demand in computation time. The K-Means algorithm, the CLARA algorithm and data sampling for the HAC method, however, have been able to overcome this potential problem to a certain extent.

Section 5: At regional scale, synthesis of the outcomes of the different methods applied.

The relevance of the whole procedure has been illustrated using the example of the international bottom otter trawl (OTB) fishery from detailed 2007-2008 logbook data gathered for the main countries (Denmark, England, France, Scotland and the Netherlands) fishing in the North Sea region, i.e. covering ICES areas IV (North Sea), IIIa (Skagerrak-Kattegat) and VIId (Eastern English Channel). The most salient points are summarised below.

For the North Sea example, the filtering methods (the combination of three sets) led to the keeping on board 69 main species, i.e. 36.3% of the initial number of species, representing 99.5% of the total value. On the basis of the results obtained here, as well as with similar trials performed on other data sets, both at national and regional levels, the most accurate classifications of logbook events into the DCF Level7 métiers were clearly obtained when combining a PCA, retaining at least 70% of the initial inertia, followed by a CLARA clustering. The combination of these two algorithms is then what we recommend. Figures 3.2 and 3.3 illustrate the results obtained when applying this recommended method on the North Sea regional dataset illustration. A comparative summary of the outcomes of the three methods (HAC, K-Means and CLARA) for the North Sea regional dataset illustration is given in the Table 3.1.

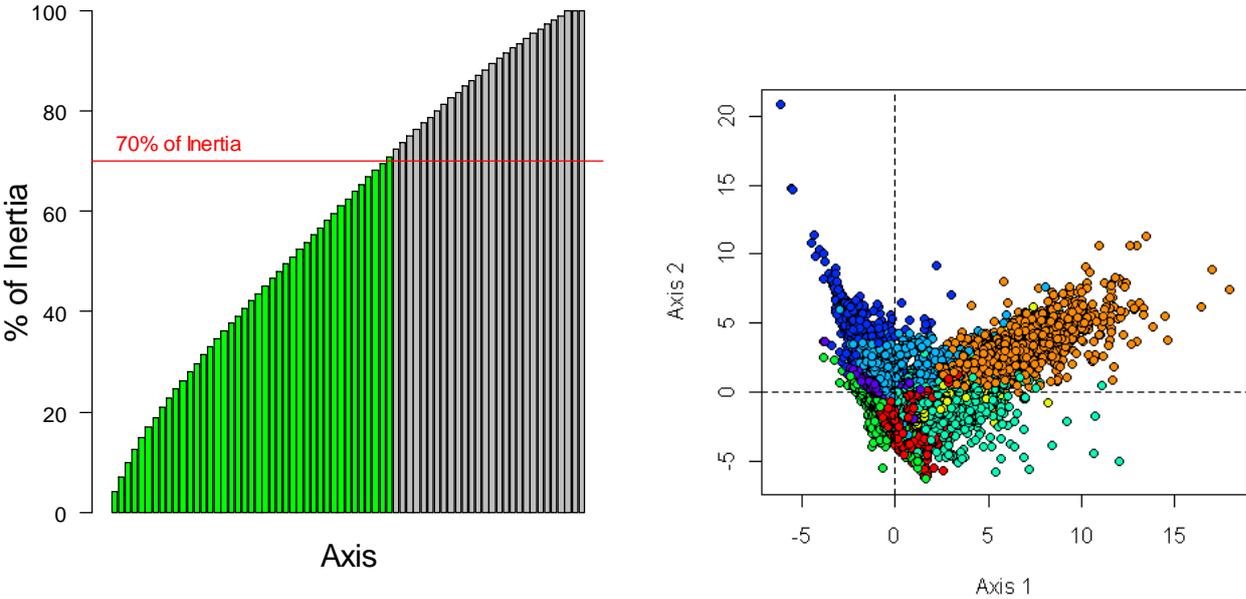
The results above were then used as a basis for evaluating the accuracy of the simple and empirical ordination methods mentioned before. Comparisons (see Table 3.2) between the two methods were pursued at the DCF level5 métiers. Characteristic species-based assemblages (defined by statistical criteria) of our original multivariate clusters were used to convert them into one DCF Level5 métier using the correspondence of species with their species type.

The mono-group métiers (demersal fish DEF, crustaceans CRU or small pelagic fish SPF) were generally well captured by the simple ordination method (over 95% of overlap), but obviously, these approaches could not capture clusters with mixed target type. This proves the usefulness of doing multivariate and clustering analyses when applied on such métiers.

The stability of the outputs over time is also shown comparing the year 2007 to the year 2008 (Table 3.3)

We performed a linear discriminant analysis to provide some allocation rules, using the outcomes of the clustering analyses in 2007. Monitoring the annual sampling program within the year is still a challenging issue. The allocation rules could be a relevant tool for checking the sampling intensity for each strata of the sampling scheme and deciding which métier should be sampled to operate efficiently. We mimicked this process predicting the allocation of each 2008 logbook event using the allocation rules of events to 2007 métiers classes. This step associated a 2007 métier class to each 2008 event. Then using the key table matching Métiers derived by the cluster analysis in 2007 and those derived in 2008 we provided a métier 2008 to each 2008 event from its predicted métier by the allocation rules. Except events classified in métier 9 (Nordic Shrimp) by the clustering algorithm applied in 2008, most of the

other 2008 events were rather well classified with respect to métier 2008 definition. Most of the events classified in class 9 in 2008 were classified by the allocation rules in the offshore métier (Class 6).



Figures 3.2. PCA outputs displaying (left) the cumulated % of inertia over axis and the 70% cut level, (right) the logbooks event in the space of the first two axes.

Table 3.1 – Comparison of the categories (clusters) obtained with three clustering methods, indicating for each category its size (number of logbook events), métier identified, and for each specie of the characteristic species-based assemblage defined, respectively, the mean catch percentage in value and the percentage of logbook events capturing it.

HAC	KMEANS	CLARA
CL4 - 9857 log events OTB flatfish - sole SOL (74%-98%)	CL7 - 20520 log events OTB flatfish – sole plaice SOL (37%-54%) PLE (19%-69%)	CL1 - 11642 log events OTB flatfish – sole SOL (66%-94%) CL5 - 6025 log events OTB flatfish – plaice lemon sole PLE (53%-98%) LEM (21%-61%)
	CL9 - 990 log events OTB flatfish - miscellaneous LEM (23%-81%) TUR (10%-58%) HAL (9%-80%) CAT (4%-53%)	
CL1 - 22746 log events OTB Nephrops NEP (85%-100%)	CL2 - 22493 log events OTB Nephrops NEP (85%-99%)	CL9 - 21387 log events OTB Nephrops NEP (85%-97%)
CL3 - 2233 log events OTB sandeel SAN (99%-100%)		CL10 - 2237 log events OTB sandeel SAN (99%-100%)
CL6 - 1914 log events OTB shrimp CSH (99%-100%)		CL4 - 1907 log events OTB shrimp CSH (100%-100%)
	CL10 - 555 log events OTB small pelagics SPR (62%-83%) HER (32%-91%)	CL11 - 556 log events OTB small pelagics SPR (61%-83%) HER (32%-91%)
CL2 - 20311 log events OTB inshore MUR (15%-65%) SQZ (14%-57%) WHG (12%-55%) IAX (9%-56%) BIB (2%-51%)	CL5 - 8751 log events OTB inshore – squid whiting mackerel SQZ (25%-73%) WHG (24%-80%) MAC (6%-57%)	CL6 - 7327 log events OTB inshore – squid whiting SQZ (30%-80%) WHG (28%-88%)
	CL6 - 5508 log events OTB inshore – surmullet cuttlefish MUR (40%-93%) IAX (27%-92%)	CL3 - 5675 log events OTB inshore – surmullet cuttlefish MUR (40%-94%) IAX (26%-92%)
	CL4 - 2886 log events OTB inshore – black seabream rays and skates smooth-hounds BRB (11%-69%) RAJ (11%-74%) SDV (6%-72%) SYC (5%-87%) GUR (4%-81%)	CL2 - 2761 log events OTB inshore – black seabream rays and skates smooth-hounds BRB (12%-72%) RAJ (11%-76%) SDV (7%-73%) GUR (5%-83%) SYC (4%-86%)
CL5 - 17321 log events	CL3 - 3717 log events OTB offshore – monkfish MON (39%-92%)	CL8 - 7169 log events OTB offshor – monkfish saithe witch flounder MON (29%-87%) POK (14%-78%) WIT (8%-58%)
	CL11 - 8464 log events OTB offshore – cod saithe haddock	CL7 - 7696 log events OTB offshore – cod saithe haddock

OTB offshore – cod plaice COD (19%-72%) PLE (19%-51%)	COD (30%-82%) POK (24%-68%) HAD (11%-63%)	COD (34%-85%) POK (20%-56%) HAD (11%-60%)
	CL8 - 365 log events OTB offshore – pollack anglerfishes POL (32%-75%) ANF (11%-59%)	
	CL1 - 133 log events OTB blue whiting norway pout WHB (47%-100%) NOP (21%-96%)	

23

Figure 3.3. Percentage of cash value per species (FAO code) per cluster (I to XI) applying the procedure PCA & CLARA Algorithm on the North Sea regional data set. Not displayed are the species labels with cash value not greater than 10% within at least one cluster.

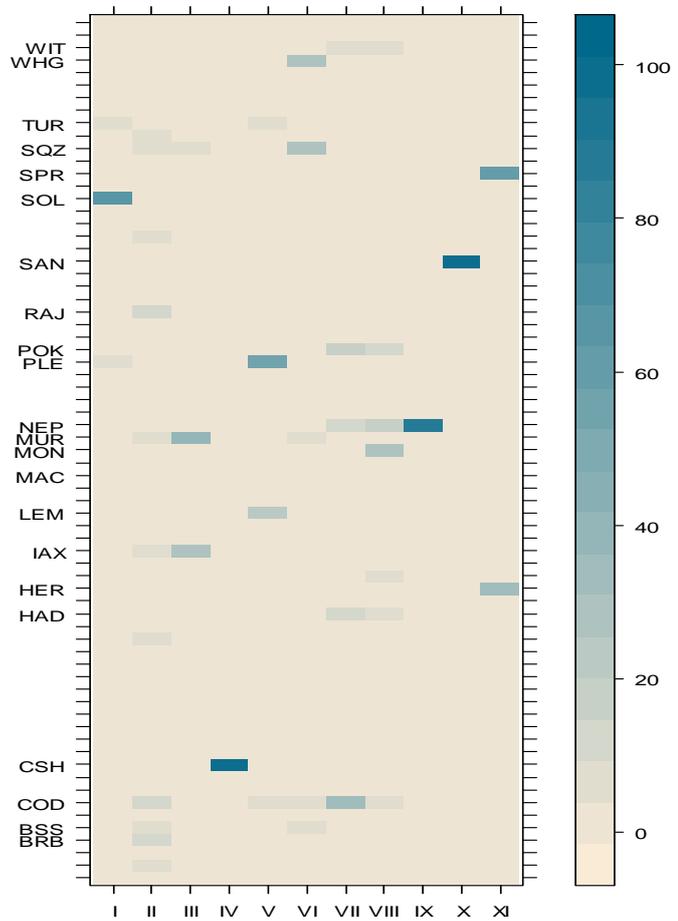


Table 3.2. Comparison of year 2007 and 2008 of the categories (clusters) obtained applying the CLARA clustering method on the landings profiles of the North Sea OTB logbook events

CLARA 2007	CLARA 2008
CL1 - 11642 log events OTB flatfish – sole SOL (66%-94%)	CL2 - 9787 log events OTB flatfish – sole SOL (65%-93%)
CL5 - 6025 log events OTB flatfish – plaice lemon sole PLE (53%-98%) LEM (21%-61%)	CL3 - 9276 log events OTB flatfish – plaice lemon sole turbot PLE (52%-96%) LEM (14%-51%) TUR (9%-52%)
CL9 - 21387 log events OTB Nephrops NEP (85%-97%)	CL7 - 39323 log events OTB Nephrops NEP (80%-94%)
CL10 - 2237 log events OTB sandeel SAN (99%-100%)	CL8 - 2878 log events OTB sandeel SAN (99%-100%)
CL4 - 1907 log events OTB shrimp CSH (100%-100%)	CL5 - 1513 log events OTB shrimp CSH (100%-100%)
CL11 - 556 log events OTB small pelagics SPR (61%-83%) HER (32%-91%)	
CL6 - 7327 log events OTB inshore – squid whiting SQZ (30%-80%) WHG (28%-88%)	CL1 - 9534 log events OTB inshore – squid whiting surmullet SQZ (23%-76%) WHG (21%-84%) MUR (13%-76%) GUU (2%-68%)
CL3 - 5675 log events OTB inshore – surmullet cuttlefish MUR (40%-94%) IAX (26%-92%)	
CL2 - 2761 log events OTB inshore – black seabream rays and skates smooth-hounds BRB (12%-72%) RAJ (11%-76%) SDV (7%-73%) GUR (5%-83%) SYC (4%-86%)	CL4 - 3785 log events OTB inshore – sea bass black seabream rays and skates smooth-hounds BSS (25%-78%) BRB (10%-72%) RAJ (9%-73%) POL (6%-52%) SDV (5%-64%) SYC (3%-75%)
CL8 - 7169 log events OTB offshore – monkfish saithe witch flounder MON (29%-87%) POK (14%-78%) WIT (8%-58%)	CL6 - 19117 log events OTB offshore MON (21%-65%) COD (21%-79%) POK (15%-67%) HAD (12%-70%)
CL7 - 7696 log events OTB offshore – cod saithe haddock COD (34%-85%) POK (20%-56%) HAD (11%-60%)	
	CL9 - 1545 log events OTB northern prawn PRA (89%-100%)

Table 3.3.-Correspondence of the DCF level7 métiers defined after the CLARA algorithm and aggregated at DCF level5 (in row) with métiers defined by the “first species” ordination method (in column) (upper) and “first group” ordination method (in column) (lower). Rows sums up to 100%.

	CEP	CRU	DEF	SPF
DEF (cl 1, 2, 5, 7, 8, 10)	1%	10%	87%	1%
MCF (cl 3, 6)	42%		56%	2%
CRU (cl 4, 9)		97%	3%	
SPF (cl 11)				100%

	CEP	CRU	DEF	SPF
DEF (cl 1, 2, 5, 7, 8, 10)		5%	94%	1%
MCF (cl 3, 6)	30%		69%	1%
CRU (cl 4, 9)		96%	4%	
SPF (cl 11)			1%	99%

Section 6: Table defining for each métier, its definition in term of robust percentage thresholds at regional scale.

Such tables have been produced by applying discriminant analysis on the output of the clustering analyses. Indeed, métiers are not only used to characterize a fishery from the historical point of view, but also to improve future sampling monitoring or fishing mortality estimates. Some quantitative rules are therefore useful to identify the métier of any new logbook event (for example during a new year). Discriminant analysis is an efficient technique for deriving such quantitative allocation rules. It allows classifying a set of observations into predefined classes, fitting a multi-choices model using one linear function for each class. This model is expected to predict the class of a new logbook events based on the species landings composition.

Allocation rules (named AR in Figure 3.1) derived from one year dataset have been tried and tested for the North Sea regional dataset illustration. First we performed a linear discriminant analysis to provide some LE allocation rules using the outcomes of the clustering analyses in 2007. We mimicked this process predicting the allocation of each 2008 log event using the allocation rules of LE to 2007 métier classes. This step associated a 2007 métier class to each 2008 LE. Then using the key table matching Métiers derived by the cluster analysis in 2007 and those derived in 2008 (Table 3.2) we provided a métier 2008 to each 2008 LE from its predicted métier by the allocation rules. With the exception of the LE classified as Métier 9 (Nordic Shrimp) by the clustering algorithm applied in 2008, most of the other 2008 LE were rather well classified with respect to the métier 2008 definition (see Figure 3.4). Most of the LE classified in class 9 in 2008 are classified by the allocation rules in the offshore métier (Class 6).

This comparison has demonstrated that the method is quite robust when allocating new logbook events to real time i.e. before that the year is finished and new analyses can be run.

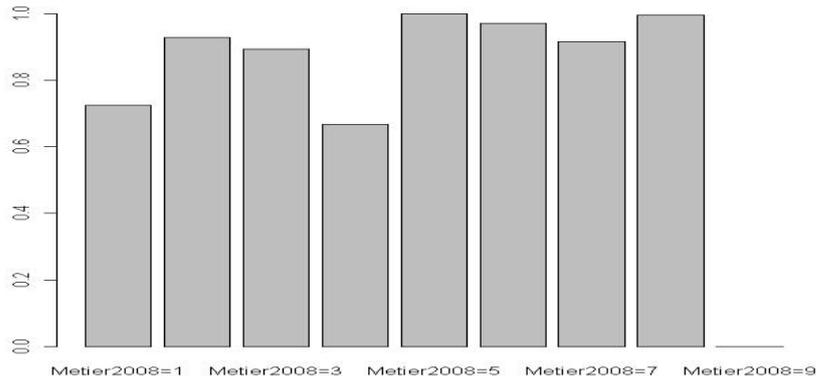


Figure 3.4. Percentage of well predicted logbooks events on the basis of the 2007 multivariate analysis compared with the cluster defined after the 2008 multivariate analysis (see Table 3.2 for correspondence between cluster 2007 and 2008).

Section 7: Protocols about how to use the methodological tools developed in order to distinguish métiers and allocate trips to them

The R functions and scripts (available in the **vmstools** package) are accompanied with several documenting files which detail their specific underlying assumptions (see also the manual in Appendix I and the draft of the intended scientific paper in Appendix IX)

Conclusion

This comprehensive study has attempted to resolve some issues which have been the source of potential concerns for the management of mixed demersal fisheries, since defining management units is the first step required before any further analysis can be conducted.

During the last two decades, some studies have been published on this issue which have used different methods on different fisheries. But none of them actually addressed the issue of the choice of the method itself in a quantitative manner; nor did they implement a prediction method to allocate new logbook events to a set of pre-defined métiers.

Our study links explicitly the outcome to the actual needs and levels defined by the Data Collection Framework. In this regards, we consider that the present analysis is a significant step forward towards:

- i) Addressing the long-lasting issue of métiers quantification in the most objective way;
- ii) Providing operational and generic tools directly applicable to any fisheries where logbooks data are routinely available (as far as the format defined in the WP1 is followed);
- iii) Contributing to the practical implementation of the Data Collection Framework in Europe.

4. Distinction of fishing from other activities

4.1 Introduction

The objective of WP3 was to develop tools capable of distinguishing fishing activity. At the very first kick-off meeting in October 2009 partners presented the work they had done on various treatments for distinguishing fishing from other activities using VMS data. As a first step each vessel 'trip' must be defined and this was discussed and included originally as part of WP3. The process of analysing VMS and logbook data together, however, is completely intertwined and hence difficult to separate sensibly and discuss on a work-package by work-package basis. 'Trip definition' is covered in the WP4 descriptions and (see also Bastardie et al. 2010; Appendix X) will, therefore, not be duplicated here. WP3 involves trip identification and definition of fishing activity but it also concerns the protocols that must be followed when analysing VMS data and combining them, i.e. what steps should be undertaken and in what order? The first step is data 'cleaning' (see below), and then identifying fishing activity which can be done in a variety of ways described below.

VMS data are relatively new to the world of fisheries management and gaining permission to use and publish them is and has been rather difficult. This means that surprisingly few scientific publications actually exist which describe VMS data analysis, and there are hence very few discussions available on how the different activities of fishing vessels can be distinguished.

In the very few VMS based studies that we were able to review, the speed (average or instantaneous) was always considered to be the most important variable. Vessel 'heading', or the bearing in which they are fishing or steaming, is also available (Table 4.1) and has been used too, but this information is generally considered to be less useful (Mills et al, 2007). In most cases, the vessel activity, once at sea (fishing versus steaming), is established by comparing the average speed to an "exogen" threshold value (Lee et al., 2010). This approach is based on the frequency distributions of the vessel VMS speeds per gear which usually appears as 'bimodal'. From programmes where observers were actually placed on fishing vessels recording activity as it happened (timing and location of fishing hauls, catches and discards) the first peak is associated with fishing activity and in general, speed thresholds are fixed for different types of gear (see also Fock, 2008, Rijnsdorp et al., 1998); each with an upper and lower boundary. The method developed by Bastardie et al. (2010) (described below and used in the **vmstools** package follows this approach (see *segmentTacsatSpeed()*, Appendix I)).

Bayesian approaches for identifying fishing activity from VMS positions have also been described in the scientific literature. Vermard, (2009), for example, used 'hidden Markov chains with a regular discrete time steps' to model transitions between successive behavioural states, e.g., fishing, steaming, stopping (at Port or at sea), for each trawler. The parameters of the transition process (speed and turning angles) are defined conditionally upon these 'behavioural states'. Bayesian methods are then used to integrate the available data (typically VMS position recorded at discrete times) and to draw inferences on any unknown parameters of the model. Results show that if VMS positions are recorded synchronously, at the instant at which the process switches from one behavioural state to another, precise and unbiased inferences on vessel activity and associated movements can be made. If, however, the observations are not gathered with a sufficiently high frequency, i.e. when the discrete observations are not synchronous with the switching instants, the performance of the estimation method can be drastically impaired. A further issue is that such Bayesian methods are highly computer-intensive and as such were of limited 'practical' use for inclusion in our package. In the current project they were, therefore, rejected in favour of the segmented regression approaches pioneered by Bastardie et al. (2010).

4.2 Methods

There can be a range of potential quality issues with VMS and logbook data and these should be addressed first. Table 4.1 shows the first few lines of a TACSAT (see WP1, Appendix I) dataset. Some VMS positions are impossible (e.g. greater than 90 degrees latitude, headings greater than 360 or less than 0 or speeds outside technical ranges possible, see Table 4.2), which can be eliminated easily (see Appendix V for a detailed script).

Table 4.1. TACSAT format of the VMS data.

VE_REF	SI_LATI	SI_LONG	SI_DATE	SI_TIME	SI_SP	SI_HE
8	51.445	3.591	18/01/2009	20:58:00	0.0	303
8	51.445	3.591	18/01/2009	22:54:00	0.0	303
8	51.428	3.532	19/01/2009	00:48:00	13.0	267
8	51.401	2.899	19/01/2009	02:44:00	11.4	265
8	51.380	2.293	19/01/2009	04:40:00	11.8	280

Table 4.2 : 'Impossible' VMS registrations, e.g. longitude = -202.567.

VE_REF	SI_LATI	SI_LONG	SI_DATE	SI_TIME	SI_SP	SI_HE
2463	33.250	-202.567	21/09/2009	19:23:00	NA	NA
33	51.648	3.183	27/01/2009	18:54:00	44.8	402
35	51.699	2.554	28/01/2009	13:52:00	10.4	393
631	54.003	1.832	26/01/2009	06:48:00	1.6	457
684	NA	NA	15/01/2009	00:56:00	NA	511
684	NA	NA	15/01/2009	23:59:00	NA	511

Duplicates are also often present in both logbook and VMS data (e.g. Tables 4.3) which can be removed using e.g. the *duplicated* function embedded within R.

Table 4.3. Removal of duplicate records.

VE_REF	SI_LATI	SI_LONG	SI_DATE	SI_TIME	SI_SP	SI_HE
8	51.428	3.532	19/01/2009	00:48:00	13.0	267
8	51.401	2.899	19/01/2009	02:44:00	11.4	265
8	51.401	2.899	19/01/2009	02:44:00	11.4	265
8	51.380	2.293	19/01/2009	04:40:00	11.8	280



VE_REF	SI_LATI	SI_LONG	SI_DATE	SI_TIME	SI_SP	SI_HE
8	51.428	3.532	19/01/2009	00:48:00	13.0	267
8	51.401	2.899	19/01/2009	02:44:00	11.4	265
8	51.380	2.293	19/01/2009	04:40:00	11.8	280

A useful feature of the *vmstools* package is the inbuilt datasets for harbour locations for all EU Member States which can be used to identify vessels clearly not fishing because they are in port. There are 4270 (e.g. Table 4.4 and Figure 4.1) harbours in the data set and the user can eliminate certain VMS positions according to any desired proximity to the harbour, e.g. 3kms.

Table 4.4. First two rows of dataset 'euharbours' in the vmstools package.

Harbour	Longitude	Latitude
Aberaeron	-4.25	52.23
Aberdaran	-4.72	52.80

The data can be accessed from the package by typing *data(euharbours)* at the command line in R and the function for determining which positions are close to harbour is *pointInHarbour()* (see Appendix I).

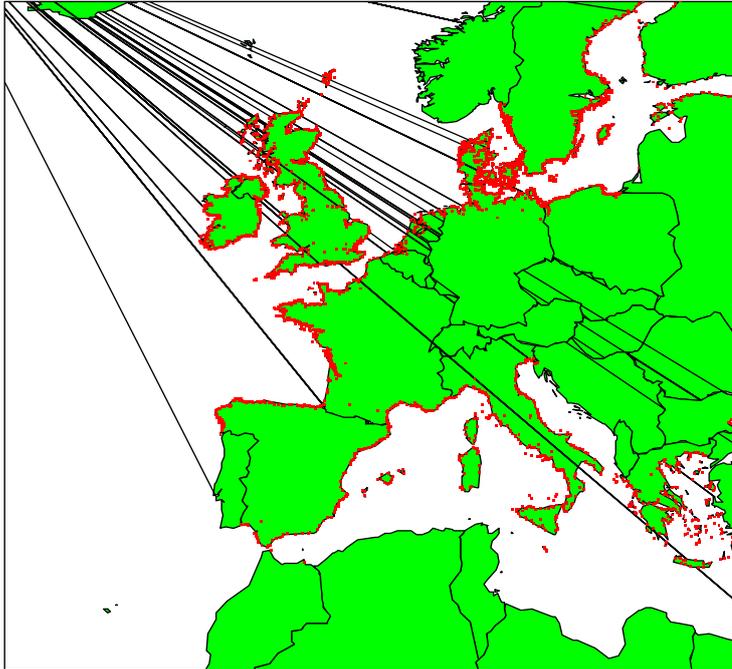


Figure 4.1. Harbour locations built into the *vmstools* package.

The next step is to remove VMS positions which are inland. Sometimes vessels really do steam inland up rivers for repairs, transit etc. where we can be certain they are not fishing. The *vmstools* package has inbuilt shapefile, "*europa*" for coast lines although the functions and data in the R packages such as *maps* and *mapdata* can also be used (see Appendix I and function *pointOnLand()*). To access these data, type *data(europa)*, convert *TACSAT* to "EVENT" format (see *PBSmapping*) and identify inland points using *pointOnLand()*.

Using the 'test' *TACSAT* data from Atlantis in the *vmstools* package then gives us the locations of vessels (Figure 4.2) from 'Atlantis' that are inland. Similarly we demonstrate the ease with generic output can be created using Irish, Dutch and French VMS data (Figures 4.3-4.5).



Figure 4.2. VMS positions in the 'Atlantis' dataset identified as being inland using the vmstools data sets europa and the function pointOnLand.

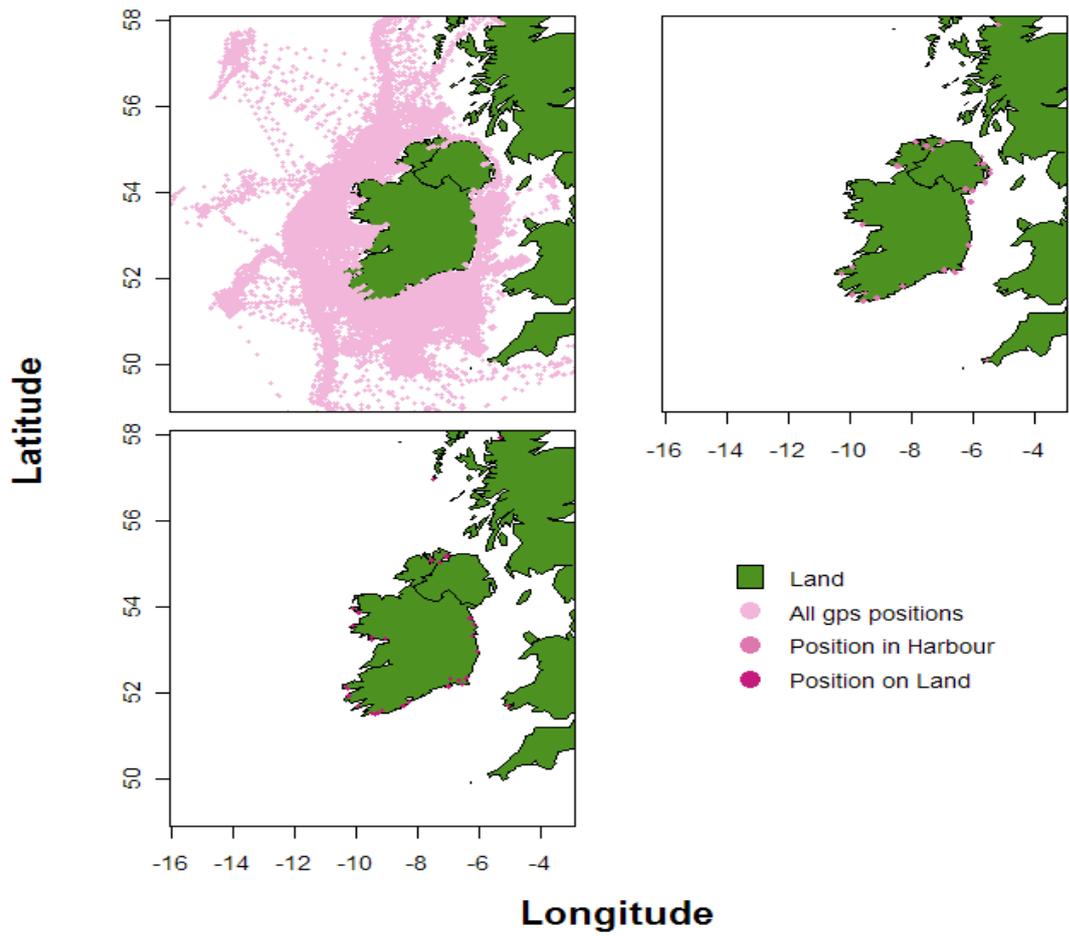


Figure 4.3. VMS data for Irish flag vessels showing positions at sea, on land and in harbour.

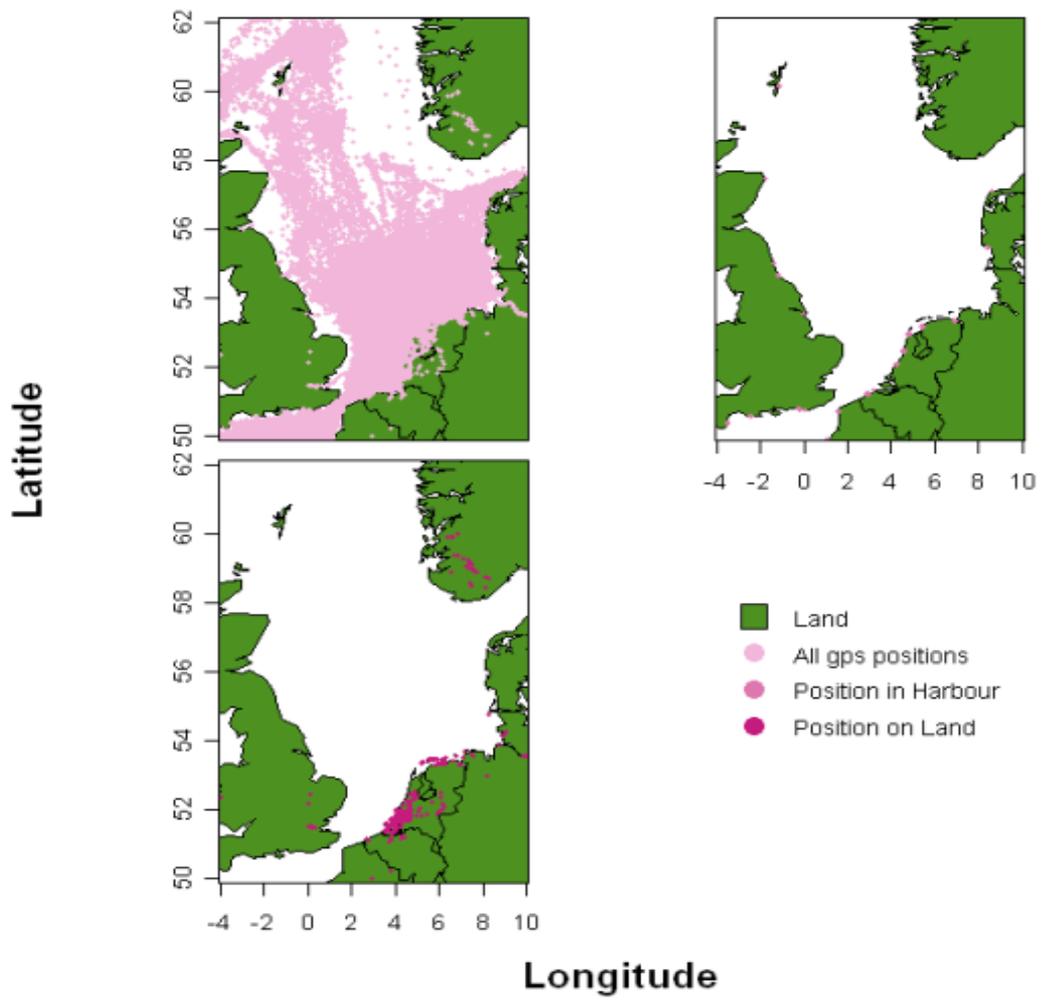


Figure 4. 4. VMS data for Dutch flag vessels showing positions at sea, on land and in harbour.

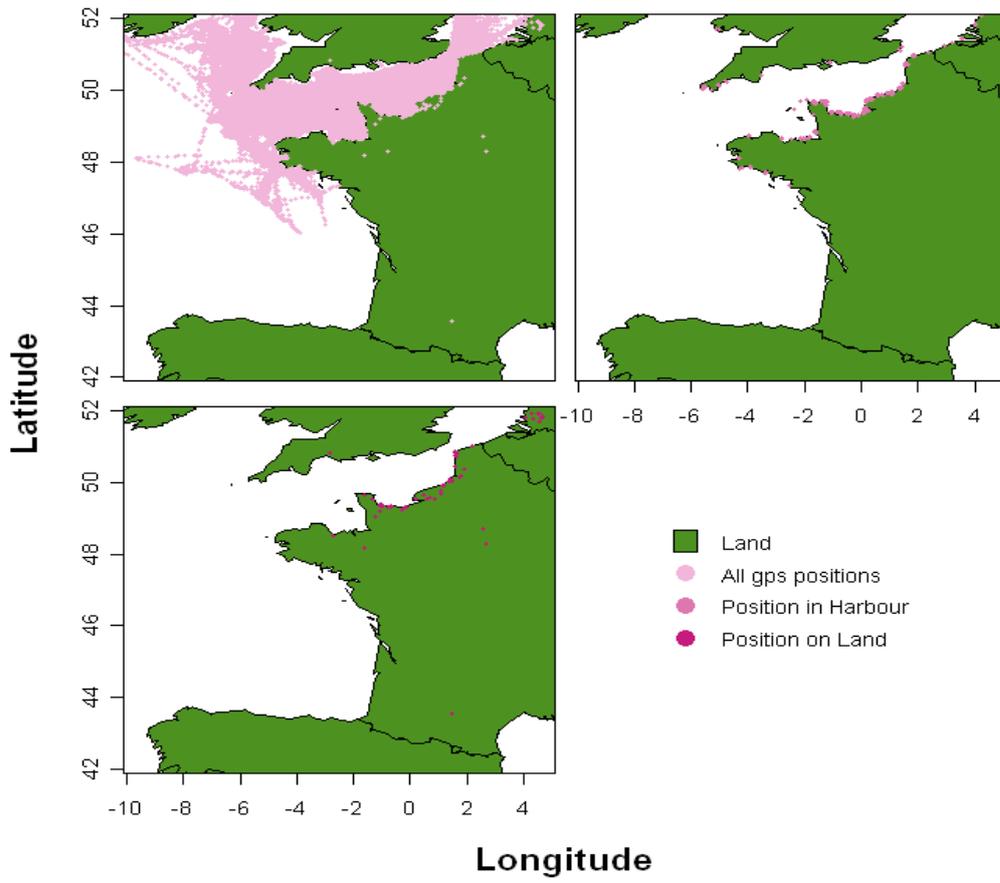


Figure 4.5. VMS data for Dutch flag vessels showing positions at sea, on land and in harbour.

4.3 Applications

After identification of points on land and those in harbour, fishing activity can be recognised best using speed thresholds (e.g. Fock 2008). For example large beam trawlers might be known to tow between 5 and 8 knots whereas they steam back to port at 12 knots. The first step in the procedure is to add the gear codes, mesh sizes, vessel main engine powers using the logbook data (EFLALO). Note: the EFLALO dataset may also require 'cleaning' since these can also contain duplicates etc. (see Appendix V for more details). Using the **vmstools** package they can be linked to the VMS data using functions such as *mergeEflalo2Tacsat* (see also WP4; Appendix I) which adds a column (FT_REF) for identifying the relevant trip, based on the rules described in WP4 and in Appendix I. These fishing trip 'flags' can then be used to find appropriate gears etc. from the EFLALO. Note: here, for example, you could also select a métier as might be identified using the methods described in WP 2 and Appendix I.

TACSAT data contain an instantaneous speed column (Table 4.1) and fishing activity can be identified using the **filterTacsat()** function (see Appendix I). A more sophisticated statistical method, based on frequency distributions of vessel velocities (see also Bastardie et al. 2010) is also available within the **vmstools** package and was developed during the project (*segmentTacsatSpeed()*). Data collected during fish trips with observers on board recording vessel movement at high spatial resolutions (e.g. a position recorded every minute) demonstrated that the vessel speed frequency histogram is usually bimodal for commercial trawlers. The first low speed peak corresponds to fishing activity, while the upper peak corresponds to activities such as steaming between fishing grounds and harbours. Low speeds can also correspond, for instance, to a regulated speed leaving a harbour.

The software (*segmentTacsatSpeed()*) determines speed boundaries by applying a segmented regression to the cumulative distribution of speeds (calculated speeds) as a practical way of automatically detecting break points between peaks in the speed frequency histogram. At this stage the computation is done using calculated speeds instead of instantaneous speeds to avoid possible non-representative records of speed (i.e. discrete records usually with large time-span). Peak detection is done for each vessel and for each of its gear types to account for individual skipper behaviour and gear-specific constraints on fishing activity. The present methodology is still, however, not well suited for many static gears such as gillnets and fly-seining because no peaks are found from observations of speed histograms. For these fishing activities, only the upper boundary of the automatic peak detection is retained, assuming that fishing starts at a given value in speed e.g. 0.5 nm. We can then use the *segmentTacsatSpeed()* function to estimate fishing activity for any vessel we chose, here vessel "1134"

The output which is sent by default to a text file to any directory specified says that the lower calculated speed boundary for this particular vessel is 2.7 knots, while the upper boundary is 6.65 knots.

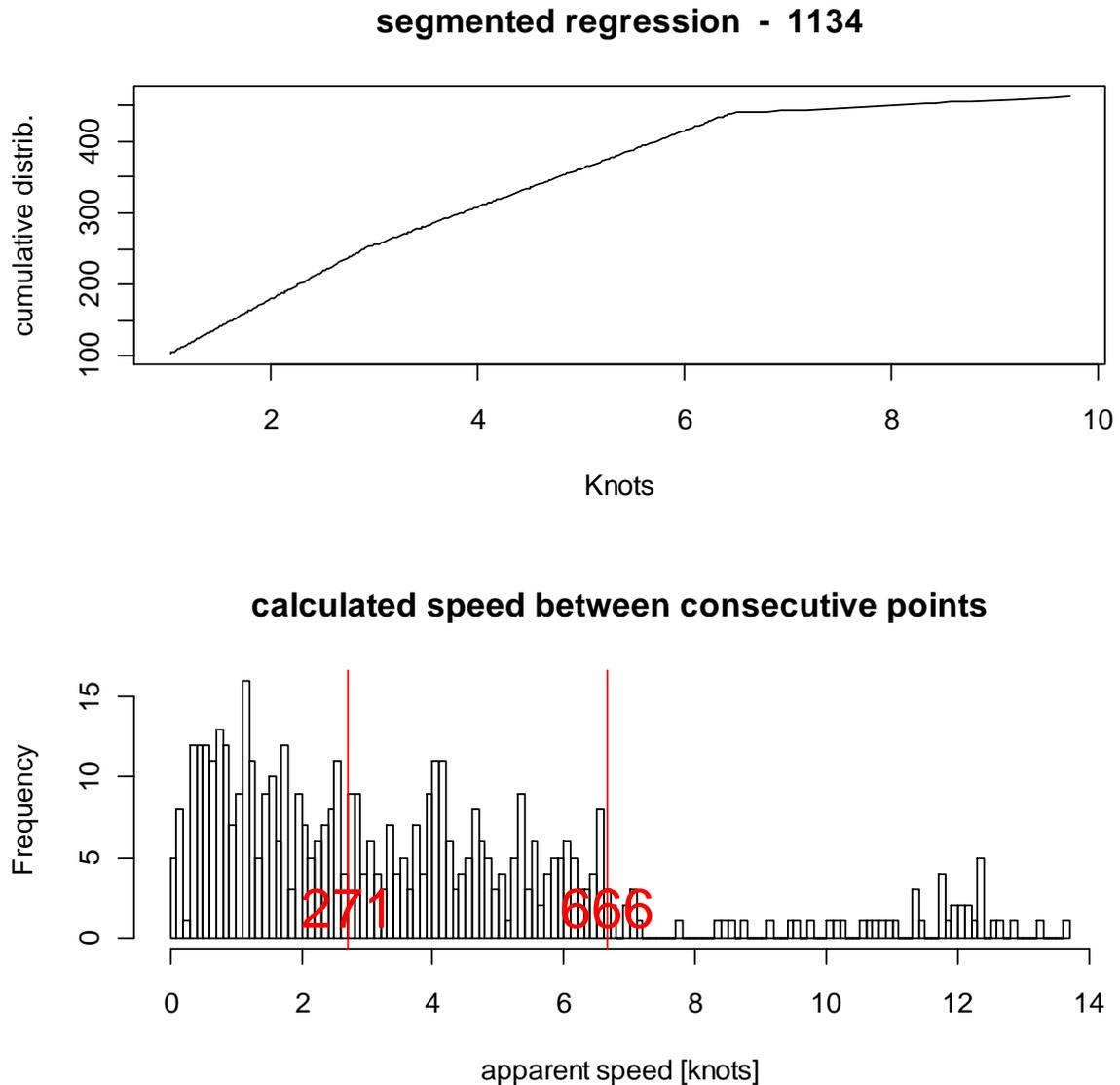


Figure 4. 5. Speed profiles from vessel 1134 from the test TACSAT dataset estimated with segmented regression (*segmentTacsatSpeed()*).

4.4 Conclusions

This work package is about identifying fishing activity but it also involves *protocols* for analysing the data which are very important. Some steps can be done in any order while others are crucially dependent on previous steps. Positions on land, for example, can be identified before or after those in harbours. Other steps, however, depend critically on the previous stage. Some of this is obvious but in other instances the analyst should think carefully about what the best order might be. Since VMS datasets are typically so large, care at this stage can save a great deal of computing time.

Clearly data must first be quality controlled and 'cleaned' before any analyses at all take place. The use of speed profiles for determining fishing activity, for example, depends critically on information on boat engine size and gear being available. Hence this step can only be reliably undertaken after TACSAT has been matched or linked with EFLALO (see WP 4). Protocols or scripts to see how these analyses can be done are included as examples with this report (see Appendix V).

5. Linking VMS to logbook data

5.1 Introduction

The steps describing the methods used by the *vmstools* package for coupling VMS registrations to logbook declarations, including the dispatching of landings between VMS fishing positions, are thoroughly described by Bastardie et al. (2010). Within the *vmstools* package the functions of relevance (*mergeEflalo2Tacsat*, *mergeEflalo2Pings*, *pings2Effort*, *pings2Fishframe*) are all described and documented in the manual (see Appendix I). Examples of their use with our 'test' EFLALO and TACSAT datasets are summarised in the script (Appendix V). The most salient points are summarised below.

5.2 Material and methods

5.2.1 Classification of VMS data into trips

The first step of the merging routine consists of first classifying VMS data into trips based on proximity of locations to ports (see also WP3). The linkage of each VMS trip to logbook trips is then done by assigning a 'common identifier' by searching for the nearest trip among all logbook trips, based on the temporal mid-point of trips. This circumvents problems where the start and end time of VMS and logbook trips do not match exactly, and ensures that each logbook trip will be merged with a given VMS trip: even if the reporting of the start and end date by fishermen is uncertain. We observed in the Danish data, for example, that these reporting dates may fluctuate with +/- 1 day around the VMS trip dates; possibly because of errors in reporting. We are not certain how much of a problem this may be for data from other countries. Additionally, the possible few remaining logbook trips are also linked with the nearest VMS trips. The latter occurs from unknown failures in the VMS devices while fishermen are still fishing. Depending on the situation, i.e. whether or not 'loss' of all reported landings is important, any remaining logbook landings that do not obviously match any VMS registrations can be distributed evenly among them.

After this first linkage, gear-specific methods for the distinction of the fishing positions from the non-fishing positions or steaming (see WP3), e.g. using method based on vessel speed profiles (see Bastardie et al. 2010), can be done with the knowledge that some gear types used during the trip are now assigned to each VMS trip. The output of this step is the identification of fishing positions among all the VMS positions.

5.2.2 Quality flags

The next step in the process is first an evaluation of the consistency between both types of data sources. Sequentially, a quality flag is issued for each recorded position depending on the various degree of matching, from finer to coarser, depending on whether both data have in common either:

- (i) the trip identifier, the same area (ICES rectangle), and the catch date;
- (ii) only the trip identifier and the area or;
- (iii) only the trip identifier.

The initial datasets (both VMS and logbooks) are then split into sub-blocks according to this flag. Each pair of sub-blocks is then joined separately and all merged sub-blocks are bound in one dataset afterwards (Table 5.1).

Table 5.1. Qualification of the 2010 Danish logbooks cod landings merged to the VMS data accounting for the 4 levels of matching in the procedure: (i) by trip only; (ii) by trip and rectangle; (iii) by trip rectangle and catch date; residuals; and per métier level 6 (see WP2). The category 'other' is the sum of all métiers having landed less than 100 tons within the year.

Metier level 6	Quality flag (%)				Weight (tons)
	1	2	3	4	
OTB_CRU_90-119_0_0	18	4	79	0	369
OTB_DEF_90-119_0_0	16	2	82	1	603
OTB_DEF_>=105_1_120	15	2	82	1	5 946
SDN_DEF_90-119_0_0	22	3	73	2	459
SDN_DEF_>=105_1_120	15	1	84	0	397
SDN_DEF_>=120_0_0	23	0	77	0	629
GNS_DEF_120-219_0_0	26	3	70	1	1 438
OTB_DEF_>=120_0_0	6	4	90	0	2 417
OTB_DEF_>=105_1_110	11	1	87	2	3 868
SSC_DEF_>=120_0_0	2	0	98	0	204
other	11	6	82	2	636

Since the qualification of the degree of matching is automatically done during the coupling procedure by attaching a quality flag to each logbook event, various subsequent tabulations can therefore be done to investigate poor matches, e.g. per métier/activity (Table 5.1) or per fishing area, per period, etc.

5.2.3 Coupling VMS and logbook data

In the third step, declared landings (in weight and cash value) at the ICES rectangle scale from logbooks can be further allocated to individual VMS registrations producing a 'new' data set at much more detailed spatial and temporal resolution. This was done by allocating a proportion of the landings to each detected fishing position (and only to fishing positions) depending on the level of matching between VMS and logbook data as described in the previous step above. In the case of full match, landings allocated to a given position are proportional to the number of detected fishing positions lying inside each declared logbook area for a given date, assuming that the total landings declared in this area at this date are divided equally among all of these positions. In the particular case of the use of several gear types within the same rectangle and also for the same catch date, an equal share of landings between gear types is also assumed for this day. In case of partial matching, the following procedure is used. First, when the match is correct by area but not by catch date, the landings are equally dispatched between all the fishing positions of this trip detected in this particular ICES rectangle. Second, when the match fails on both catch date and area, the landings concerned are equally allocated to all the detected fishing positions of the trip for which no previous match has been detected. Finally, residual landings can occur when the match fails both for the catch date and area, and no fishing position remained.

This method can be routinely applied to data from various EU countries using the **vmstools** package providing that the same data input format are provided (see WP1), i.e. TACSAT and EFLALO. So far, the method has been successfully applied to Danish, Dutch, Irish, French and British data sets (see Figures 5.1-5.4). The data output via this process can be further aggregated from discrete VMS positions by applying various gridding algorithms (e.g., by C-square cell, *CQuare()*, <http://www.marine.csiro.au/csquares/>). Furthermore functions within the **vmstools** package take the output from these procedures producing format compatible with the FishFrame data format (see Appendix III).

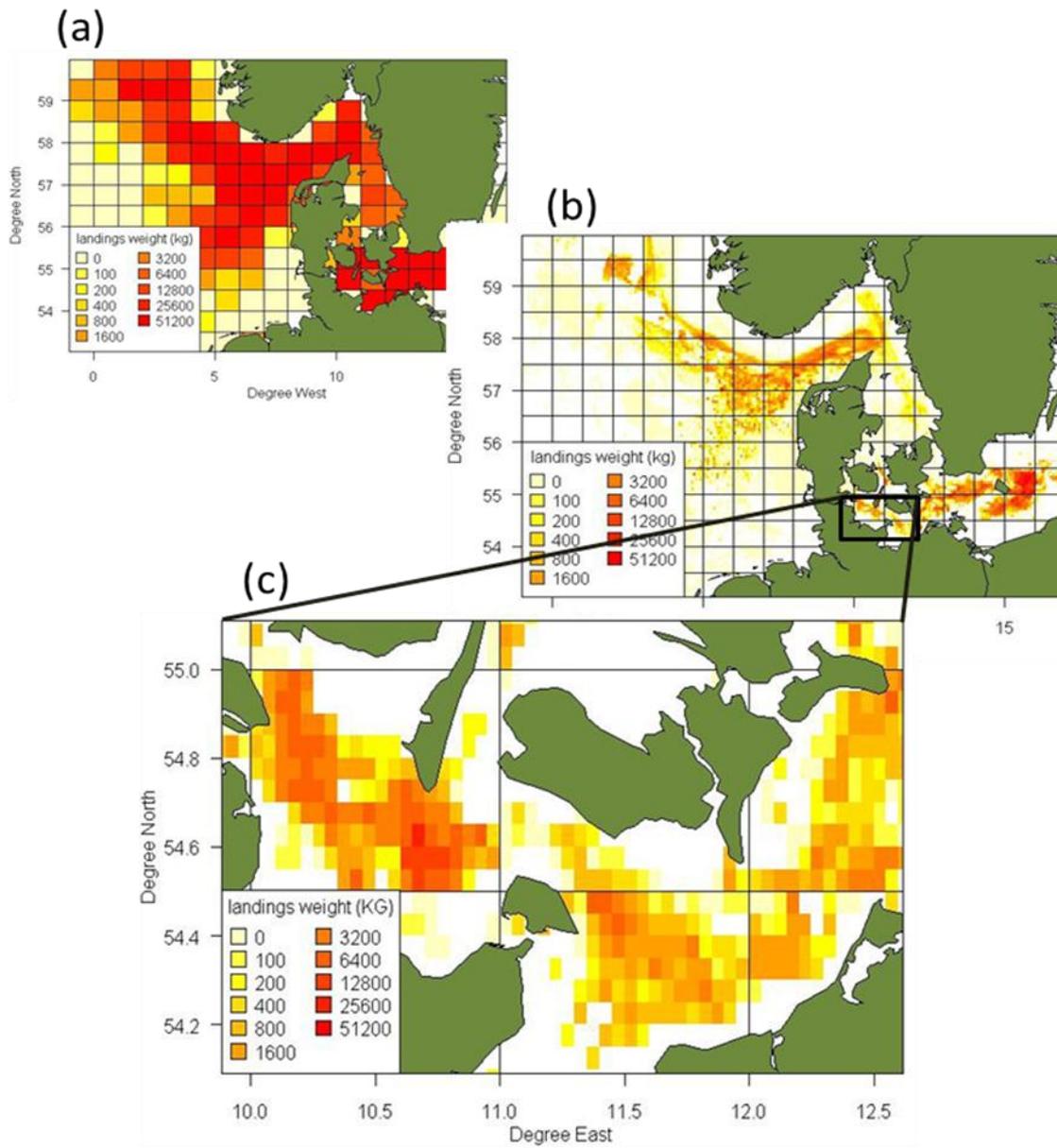


Figure 5.1: Illustration of the increased spatial resolution obtained for the 2010 Danish cod landings by vessels larger than 15 m when logbook data are distributed among VMS positions. (a) by ICES statistical rectangle; (b) on a 0.05 by 0.05 grid after coupling with VMS data; (c) zooming in on a particular area.

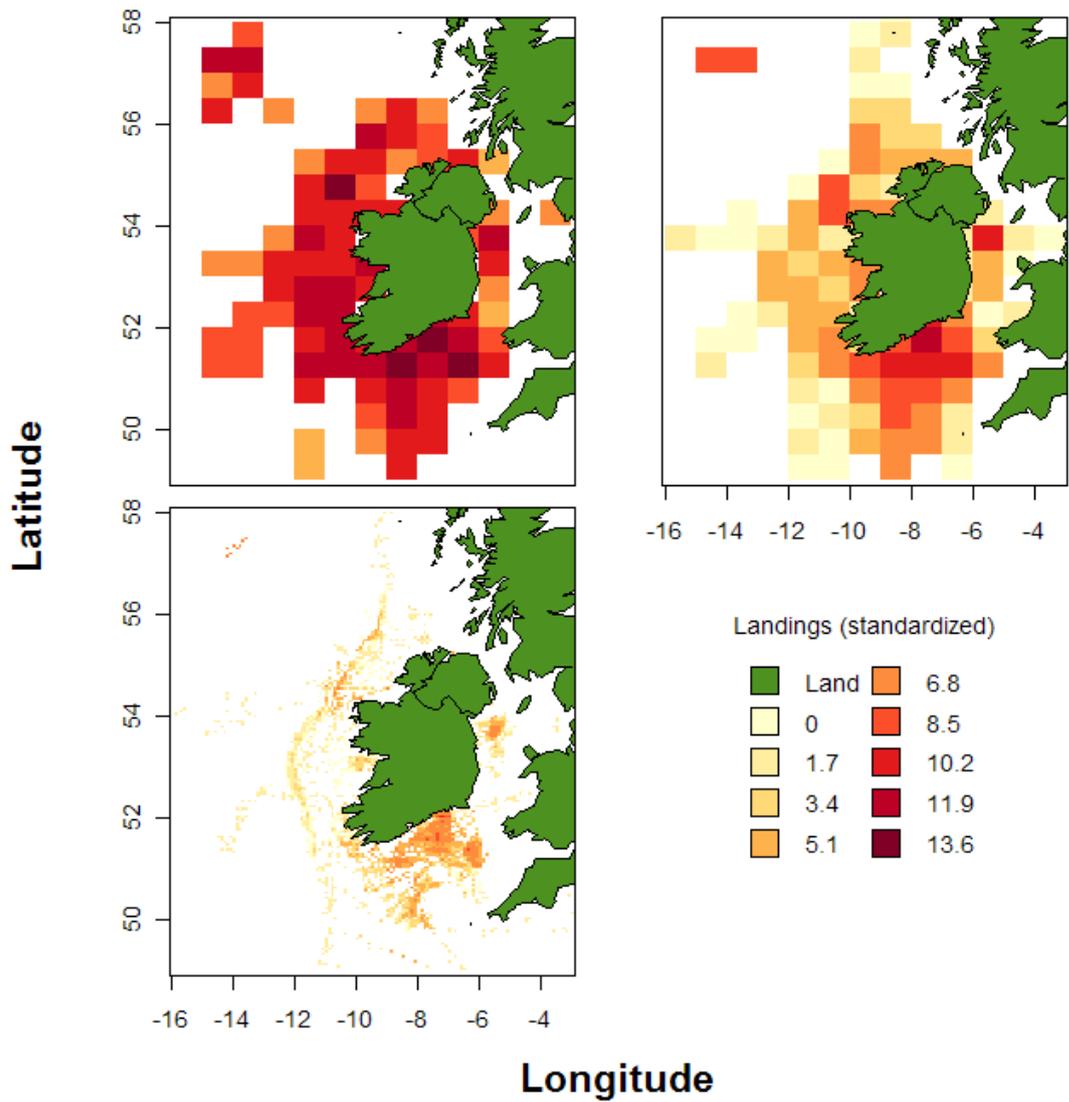


Figure 5.2: Illustration of the increased spatial resolution obtained for the Q1 2009 Irish cod landings by OTB vessels larger than 15 m when logbook data are distributed among VMS positions. (a) Landings from EFLALO by ICES statistical rectangle (the highest spatial detail possible); (b) Landings after linking TACSAT and EFLALO by ICES statistical. (c) Landings after linking TACSAT and EFLALO by 0.1 x 0.05 grid cell in respectively longitudinal and latitudinal degrees

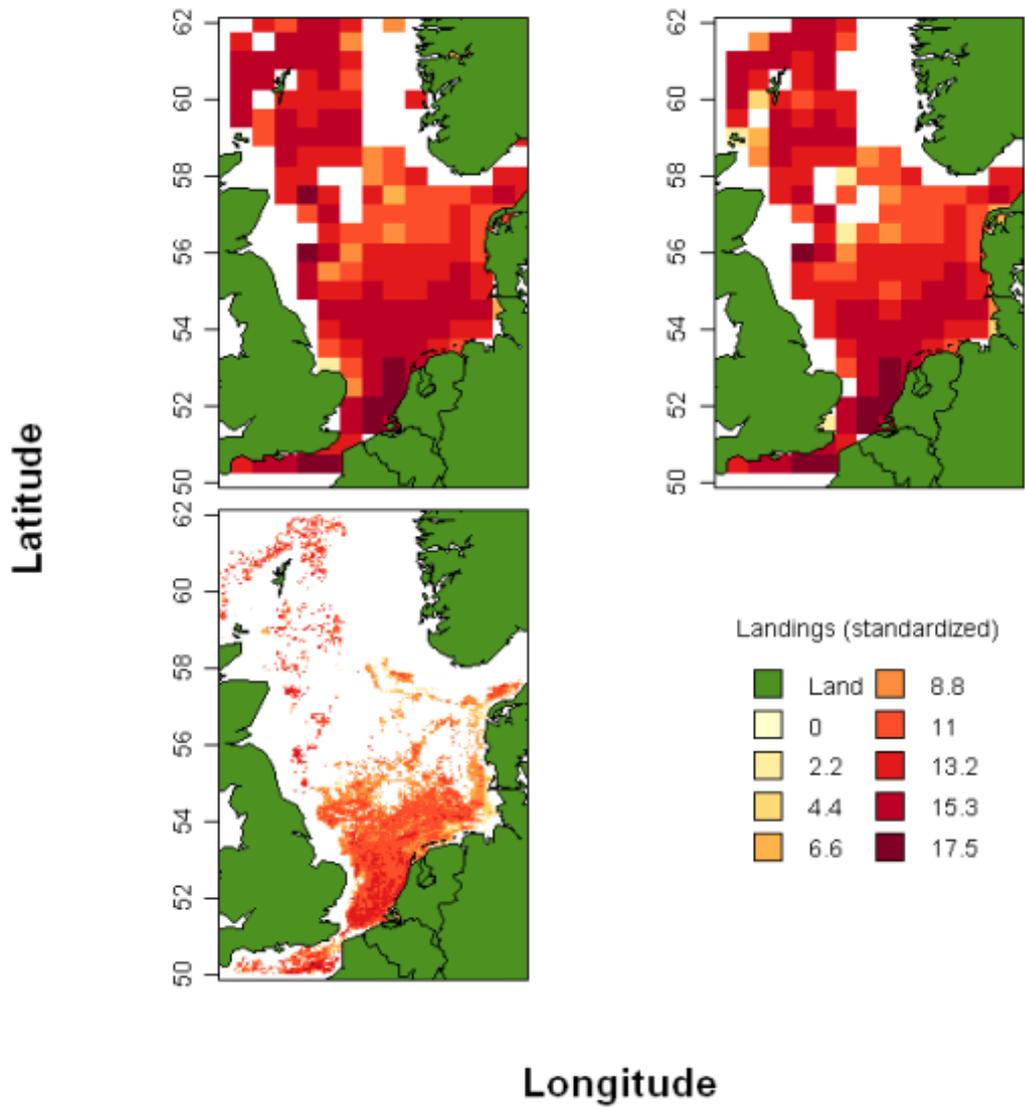


Figure 5.3. Illustration of the increased spatial resolution obtained for the Q1 2009 Dutch cod landings by OTB vessels larger than 15 m when logbook data are distributed among VMS positions. (a) Landings from EFLALO by ICES statistical rectangle (the highest spatial detail possible); (b) Landings after linking TACSAT and EFLALO by ICES statistical. (c) Landings after linking TACSAT and EFLALO by 0.1 x 0.05 grid cell in respectively longitudinal and latitudinal degrees.

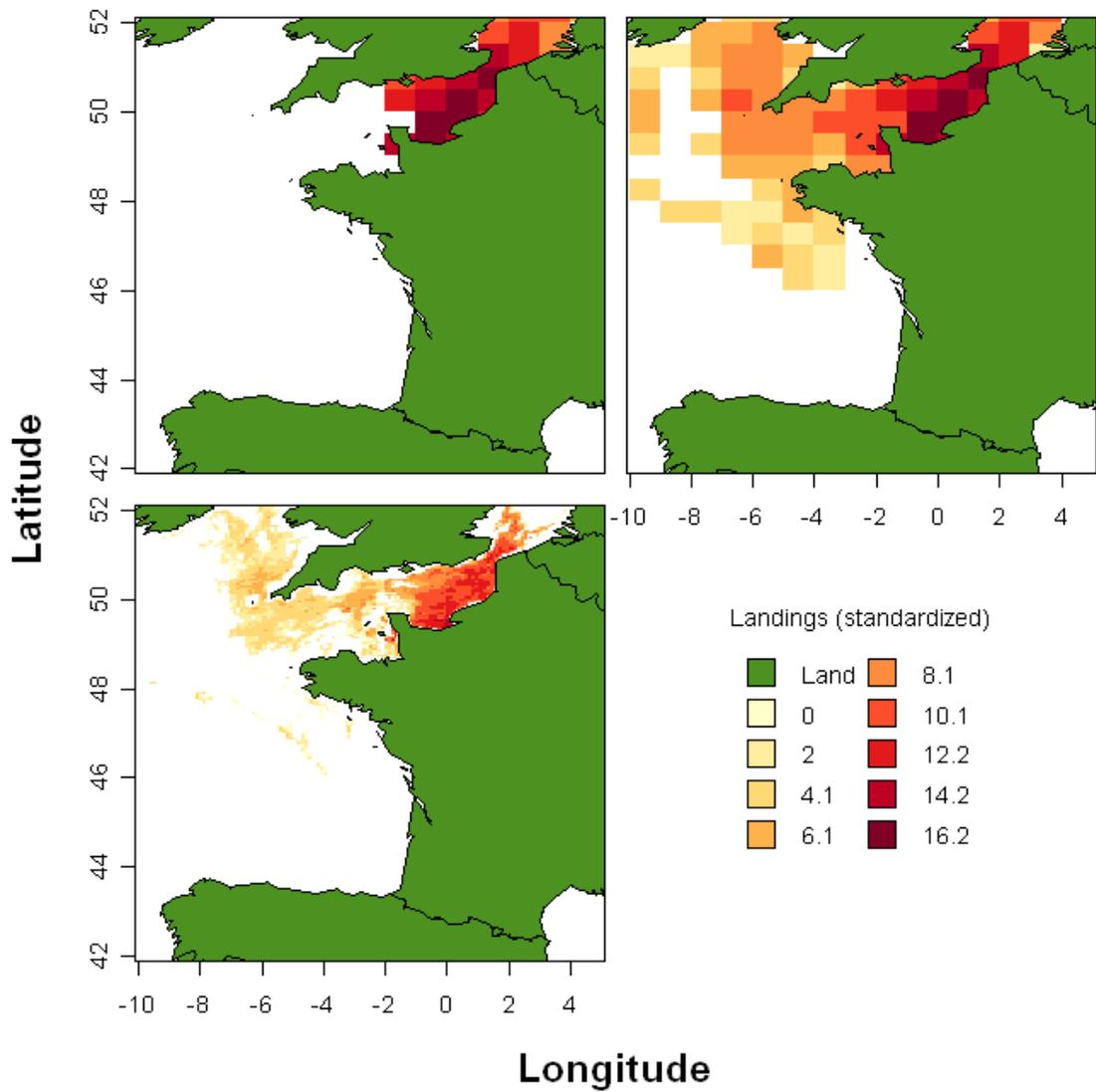


Figure 5.4. Illustration of the increased spatial resolution obtained for the Q1 2009 French cod landings by OTB vessels larger than 15 m when logbook data are distributed among VMS positions. (a) Landings from EFLALO by ICES statistical rectangle (the highest spatial detail possible); (b) Landings after linking TACSAT and EFLALO by ICES statistical. (c) Landings after linking TACSAT and EFLALO by 0.1 x 0.05 grid cell in respectively longitudinal and latitudinal degrees.

In Ireland, the Marine Institute has in the past applied other methods to analyse logbook and VMS data ([Gerritsen 2011](#)). The method of ([Gerritsen 2011](#)) assigns daily catches to VMS positions where vessels are deemed to be fishing. This was compared to the **vmstools** project approach described above and by Bastardie et. al. (2010) which assigns catches to the VMS positions on a trip basis (using the function *mergeEffalo2Pings*). An example dataset, consisting of cod catches in the first quarter of 2009 by OTB and TWR gears was used for the comparison. Simple speed criteria were applied for both methods. Figure 4.2 shows the distribution of the landings according to the two methods. The general pattern is very similar although small-scale differences do exist. The maps here have been included to demonstrate the utility of generic software such as **vmstools** which was applied easily to Irish and Dutch data to produce identical output based on standardised protocols (compare Figs. 5.2 and 5.3).

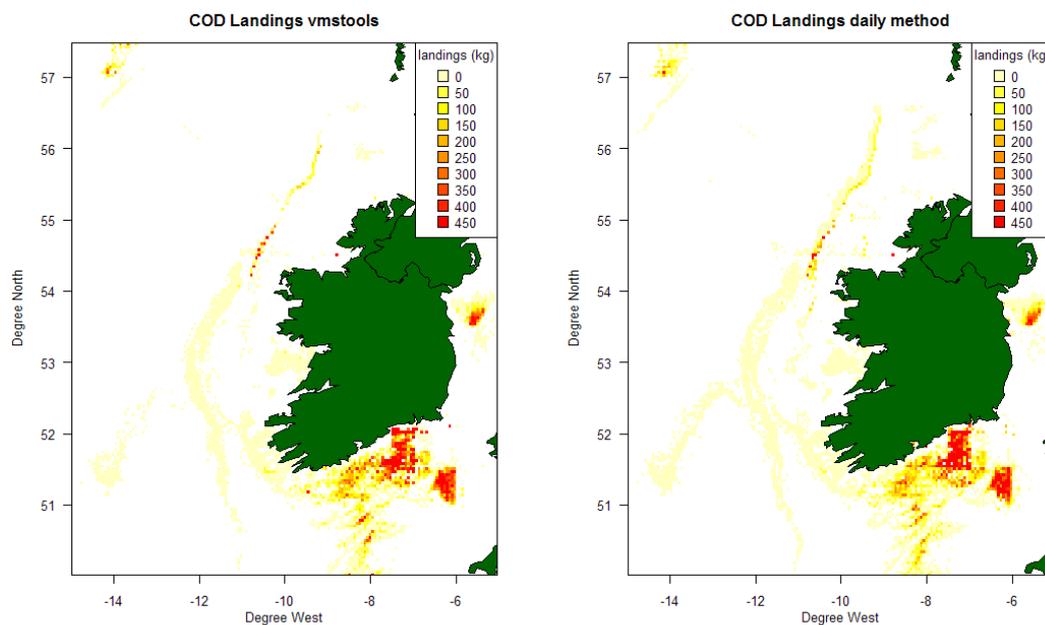


Figure 5.5. A comparison of the Irish cod landings in the first quarter of 2009. The **vmstools** method (left) was compared to the method by Gerritsen and Lordan (2011) that assigns daily catches to the VMS positions (right).

5.2.4 Uploading data to FishFrame

The final data step in the processing of raw disaggregated VMS and logbook data into merged and aggregated data suitable for presentation is carried out by FishFrame. FishFrame is the common Regional Database for three of the four regions defined by the EU. The Regional Database will, in the future, hold most fisheries data collected by most EU Member States and provides an easy way to access international information about the national and international fishery.

Raw VMS data are processed, aggregated and extracted to the FishFrame VMS data exchange format (See Chapter 2 for description of the exchange format) by use of some of the **vmstools** package developed in this project (Appendix I). In order to access FishFrame any user needs to be a registered user of FishFrame. After successful registration the user will be granted privileges according to role allocated to that particular user. The role defines to which data types, to which functionality and to which the level of dis-aggregation of international data the user have access to.

5.2 Applications

FishFrame is opened from fishframe.org and the front page below is shown (Figure 5.6).

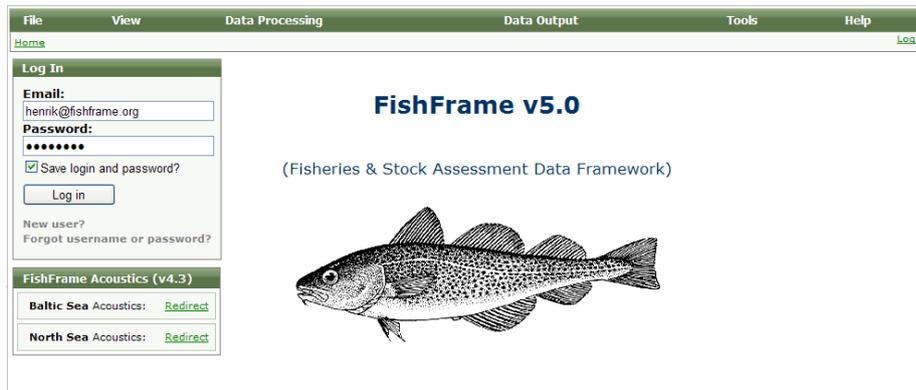


Figure 5.6. FishFrame front page

After login the "File"-tab is clicked in the green top menu bar and "Data upload" is selected. The data upload screen is then shown (below). Browse for the file you want to upload. Be aware of the file naming conventions in force (see WP1 and Appendices).

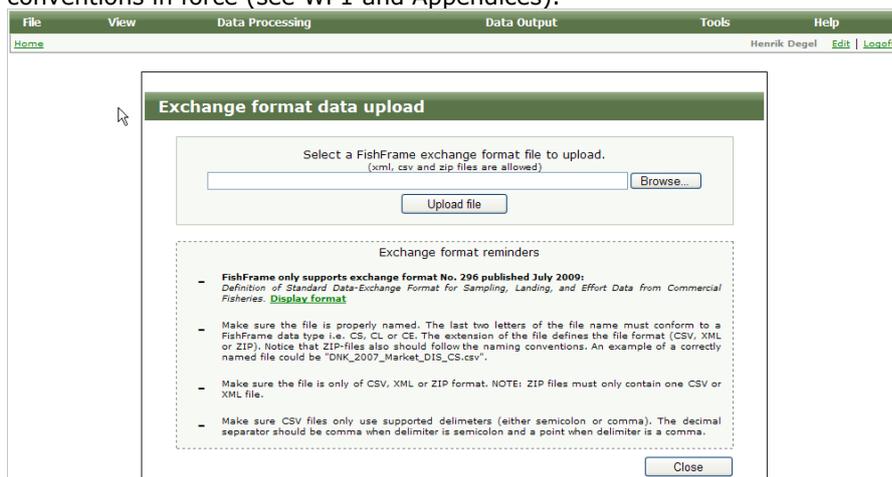


Figure 5.7. FishFrame upload page

After pressing "Upload file" (Figure 5.7) the data will be run through a number of checks (Figure 5.8 below). If the data fails to pass one of the checks, an error report will be shown giving precise indications of the type of the error, which variable and line number. The exchange file has then to be corrected and upload is re-executed.

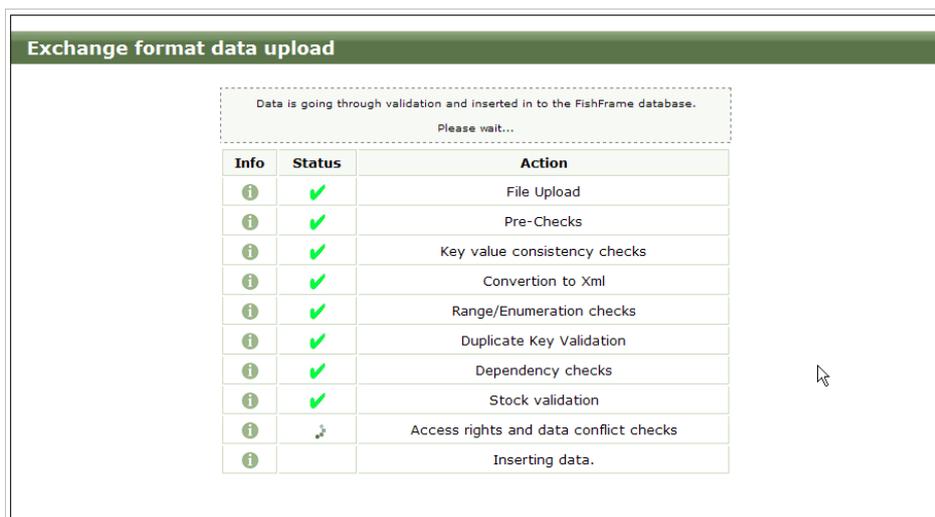


Figure 5.8. Checking the data

When the data has passed the checks it is included in the warehouse part of FishFrame and will, after automatic nocturnal processing, be available for output reports (maps and pivot tables). For immediate availability for the output module, (see the Output section below).

5.3.1 Visualisation in the data tree

By selecting "View" in the green main menu bar and hereafter "Data tree" the data can be seen in a tree structure. Select "VMS/Logbook data" in the bottom of the "Select tree type" field. Press the country or year for which you want to make a selection. An example is shown below (Figure 5.9) giving a subset of the Danish VMS data from 2010. The total data set consists of 2512 lines as indicated in the right bottom of the screen.

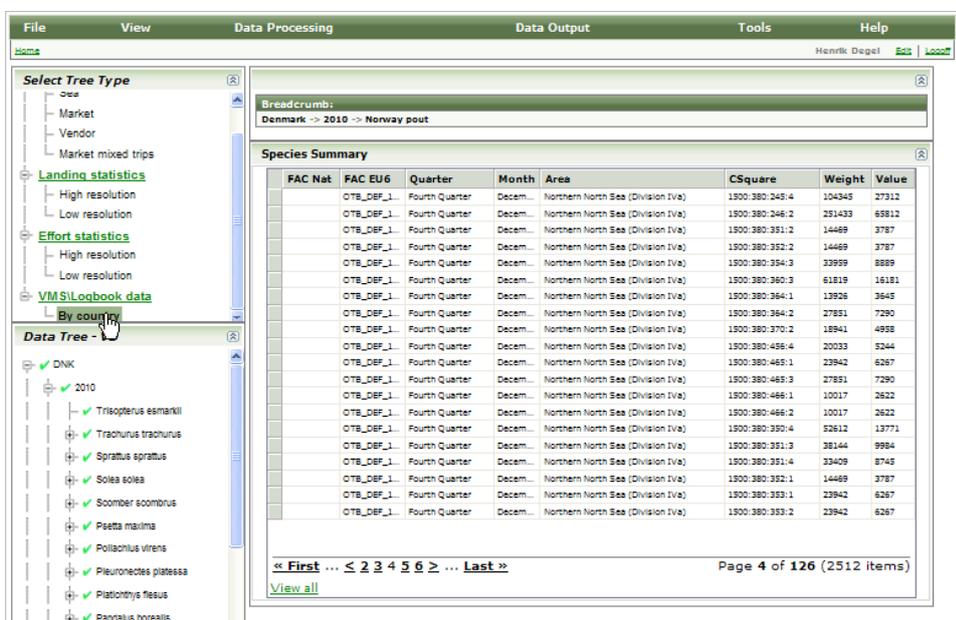


Figure 5.9. Visualising the data tree in FishFrame.

5.3.2 Output

Maps and pivot tables can be created (as can output for any other data type) by selecting "Data Output" and then "Reports and Analysis" and then "Select in tree" or "Select by keywords". The last one is preferable as long as the number of predefined reports do not exceed what can be overviewed.

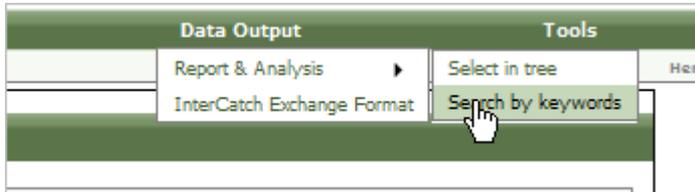


Figure 5.10. Getting maps and pivot tables in FishFrame

Then press "Show all reports" and select the report you want to create.

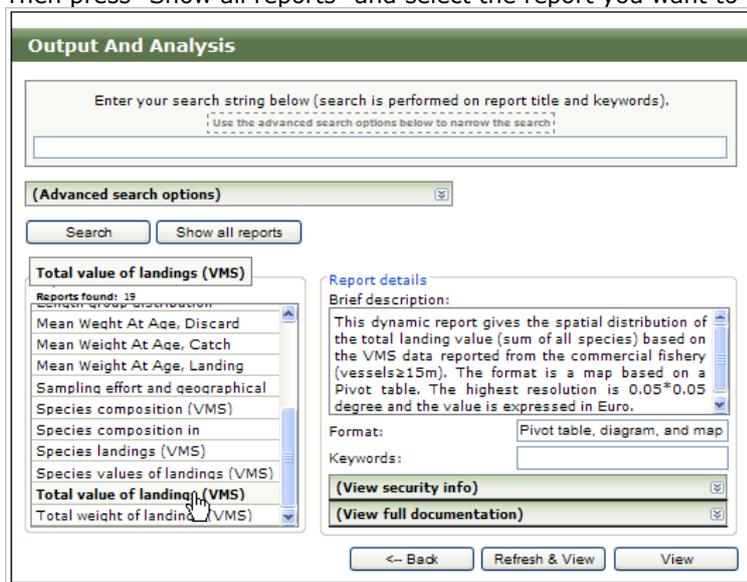


Figure 5.11. Selecting the desired report in FishFrame.

Select the report you want to create. (All reports using VMS data are indicated by "(VMS)") and see the description of the report (Figure 5.11). Press "View" and start to create the pivot table which is the basis for creation of the map. Initially, the "Dimension" is not defined in the pivot table (Figure 5.12). This might take a couple of minutes (see Response time discussion in WP7 for more information).

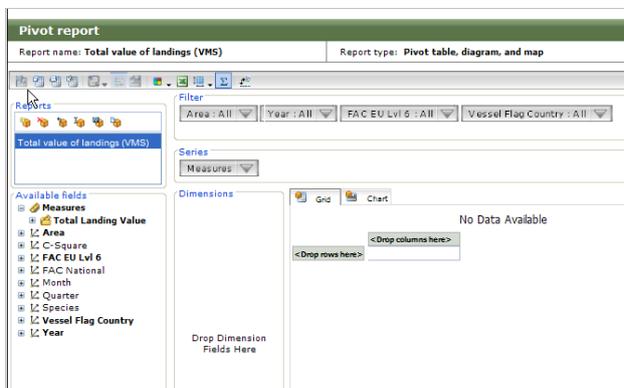


Figure 5.12. Selecting the 'dimension' in FishFrame

Frame the data you want to plot on the map by dragging variables to the "Filter" box and select a subset of the data.

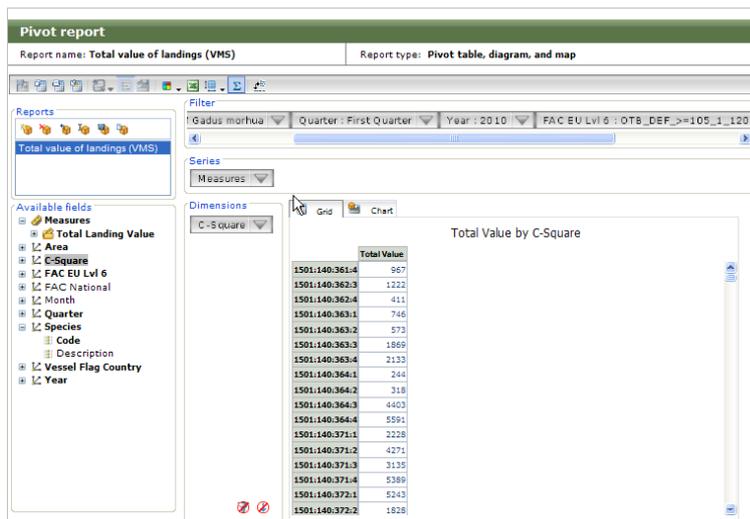


Figure 5.13. Obtaining the pivot table in FishFrame.

When the Pivot table (grid) is ready (the hour glass has disappeared) click "Refresh map with current data selection" below the table (Figure. 5.13) and the C-square values of the selected parameter will be plotted on the map.

The settings of the map can be modified using the settings field on the right side of the map. One of many possible final results is shown below (Figure 5.14).

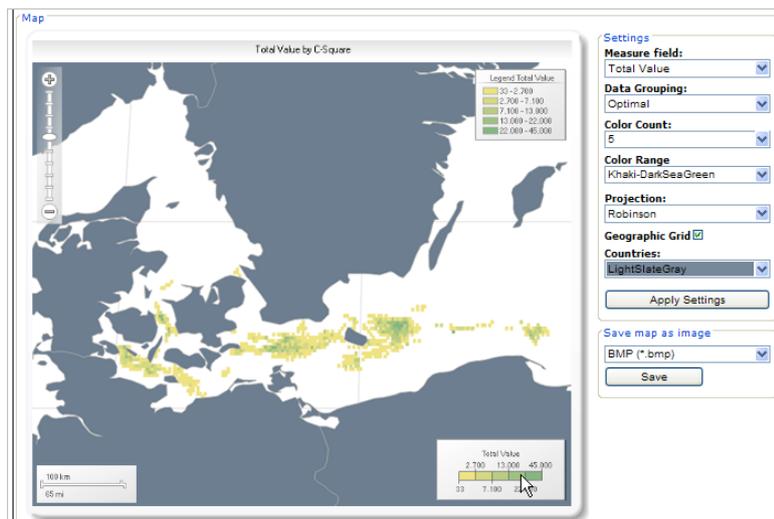


Figure 5.14. Total value of Danish landings in the Baltic from FishFrame.

The map can be exported by pressing "Save".

Alternatively, the 'pivot tables' and ordinary graphs can be created. When no map is required the pivot table can be designed using the list of parameters shown to the left of the table as the user prefers. An example is shown below. The table can be exported to EXCEL by pressing the EXCEL icon in the top.

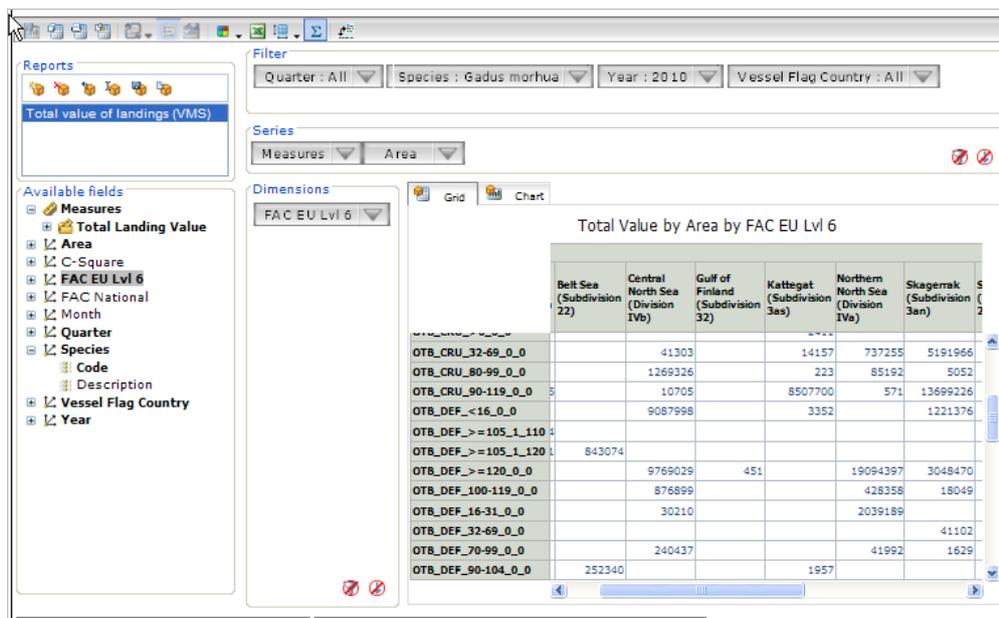


Figure 5.15. Total value of landings by area and métier (EU level 6).

These data can also be shown as a diagram by selecting the "Chart" banner. Data can also be filtered and sorted by clicking

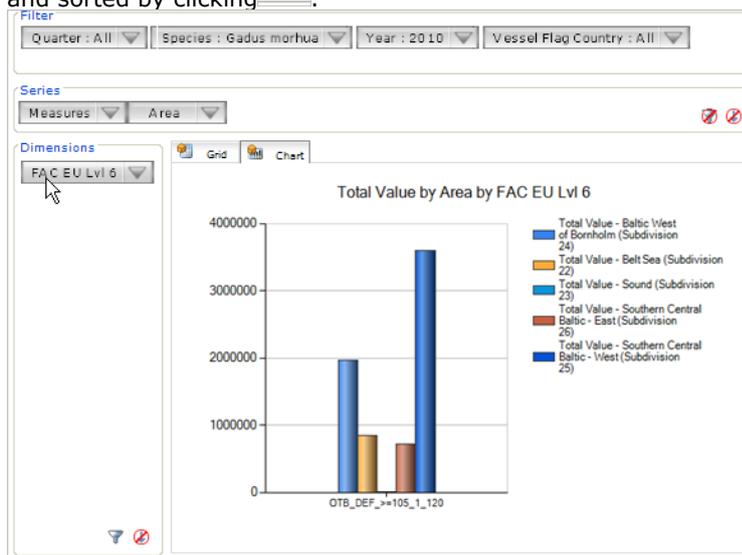


Figure 5.16. Total value of landings by area and métier (EU level 6).

5.4 Conclusions

Beyond the usual mapping of landings at the scale of the ICES statistical rectangle (Figure 5.1, top) from the logbook fishers declarations, the coupling of logbooks to the satellite-based VMS data enables the spatial origin of landings to be described at a much finer spatial and temporal resolutions (e.g. Figs 5.1-4). Furthermore, the output of WP2 assigning a particular activity/métier to each trip of a given vessel (more precisely, to each logbook event) can also form the basis for tabulating and mapping the origin of landings also for each specific fishing activity.

Work package 4 builds on output from WPs 1-3 creating methods for linking logbooks and landings and transferring those aggregated (i.e. by month, grid cell and métier) data to a Regional database. Note that these aggregated data can be exchanged more freely among nations. All methods are fully documented and described in detail in the *vmstools* manual (Appendix I).

6. Evaluation of the effect of recording rate

6.1 Introduction

The Vessel Monitoring by Satellite system (VMS) was originally introduced by the European Union in 2002 for fisheries 'control' purposes. It was primarily designed to check fishing activities and to verify whether vessels hold quotas and licences to fish in the areas in which they are active ([EC 2002](#)). Hence, the recording rate of the GPS transponder was adjusted to send its position, speed, heading and vessel information every one or two hours to a satellite, while in some occasions this rate could temporarily be increased.

Over the past years, many scientists have used VMS information to study fisheries, investigating fishers' behaviour and relating their activities to, for example, their impact on benthos and sea bed integrity. For many of these studies, however, higher recording rates should in theory result in better understanding of the vessels behaviour, enabling scientists to estimate the impacts of fisheries with more precision.

6.2 Material and methods

The effects of a higher recording rate are here examined following two different approaches:

(1) The first approach is based on the actual position registration data coming from specific high-frequency sets (i.e. with short intervals between the registrations). Only a few high-frequency datasets were available for this study: A Dutch dataset and an English dataset. Datasets varying in terms of frequency (and thus interval) are 'mimicked' by sequentially removing VMS records from these high-frequency sets. The Dutch position registration data are described in ([Rijnsdorp et al. 1998](#)) and were collected at a frequency of every 6 minutes, it follows that a 12, 18, 24 or 30 minute interval set can be mimicked by removing each 2nd, 2nd - 3rd, 2nd - 4th or 2nd - 6th record respectively. The English dataset was collected at a frequency of every 1 minute, but to enable a comparison between these datasets, this dataset has been subsampled at steps of 6 minutes as well. The area trawled can then be calculated for each of these datasets depending on the spatial resolution (i.e. grid size) given certain assumptions on gear width and effort allocation (fishing activity correctly assigned);

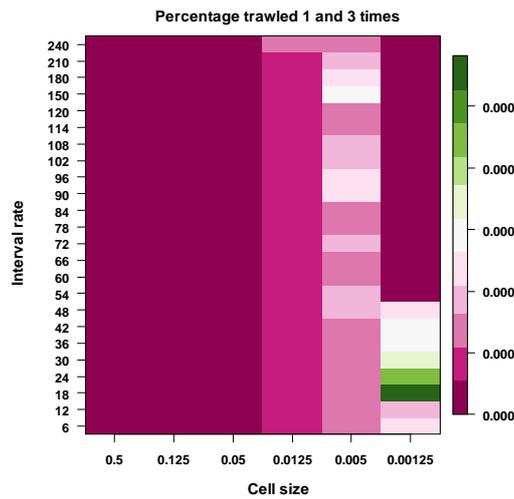
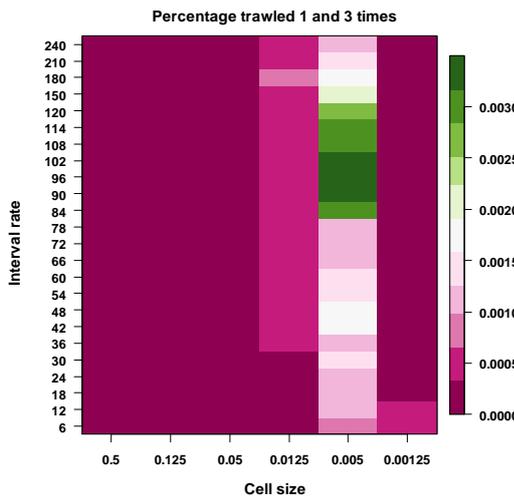
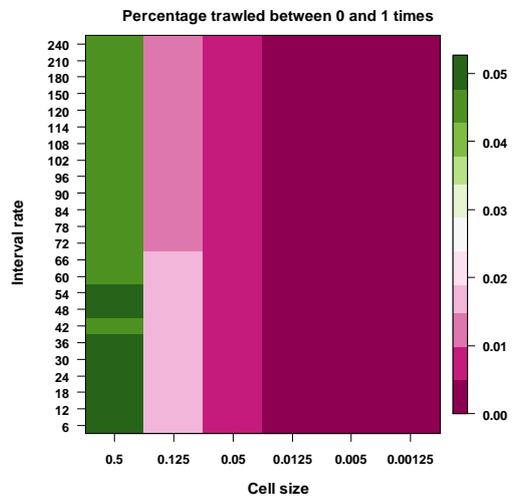
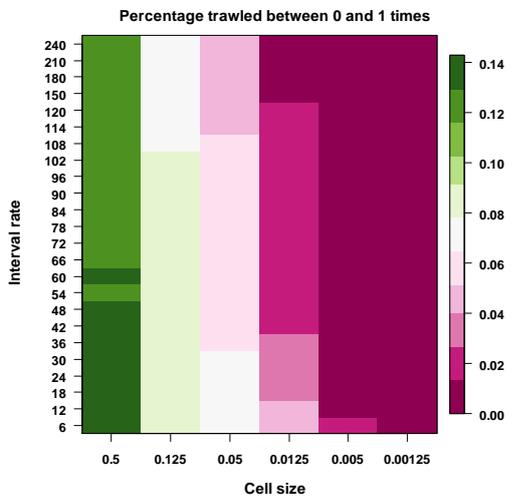
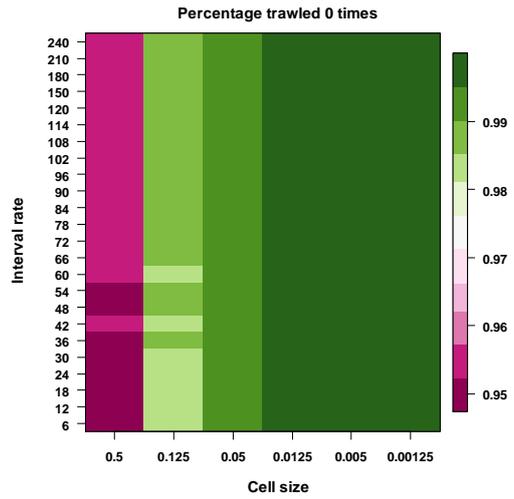
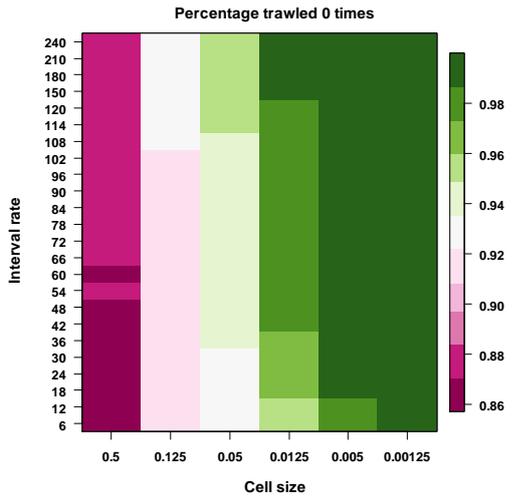
(2) The second approach is based on a reconstructed trawl track, created through an algorithm developed to interpolate fishing VMS recordings between 'known' VMS positions. Now the choice of the recording rate is determined by the performance of this algorithm at these specific recording rates.

All analyses performed within work package 5 and presented in this chapter of the report used the full functionality of the **vmstools** package as developed within this project.

6.3 Applications

6.3.1 Based on position registrations

The figures below show how the number of times a certain spatial grid cell is trawled is related to the recording rate. As the surface area of a particular cell size can be larger than the surface covered by a single trawling event, such a cell size can also be trawled e.g. 0.2 or 0.5 times.



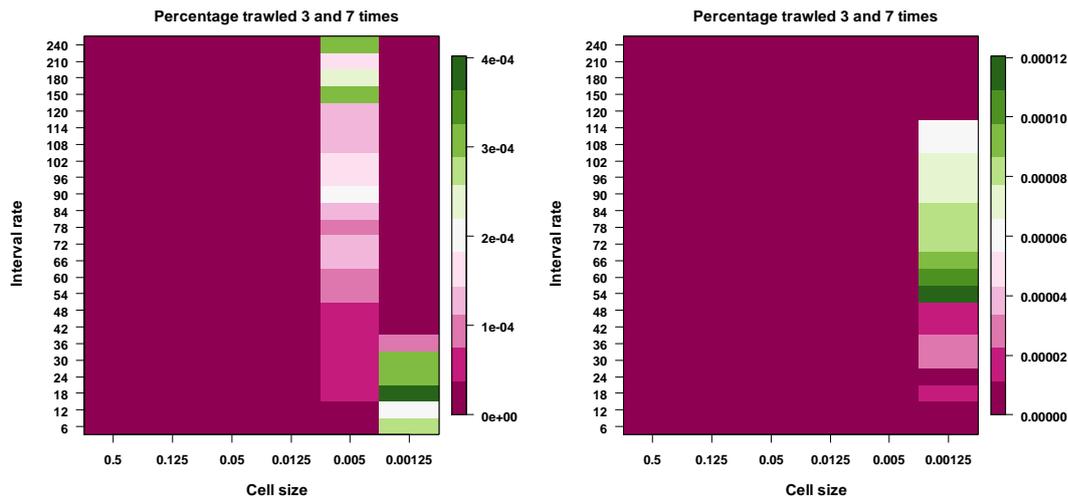


Figure 6.1. Percentage of area trawled depending on interval rate (y-axis) and spatial grid cell size (x-axis). Panels on the left show the results based on the Dutch high-frequency data while panels on the right show the results obtained from the English high-frequency data. Here, interval rate between VMS records follows a sequence between 6 and 120 minutes, in steps of 6 minutes, and thereafter steps of half an hour. The grid cell size is given as latitudinal degrees (where longitudinal degrees = 2* latitudinal degrees). I.e. a cell size of 0.5 results in spatial grid cell sizes of 1 degree longitude versus 0.5 degree latitude. The different colours indicate the percentage of area trawled a certain number of times. (1st row) percentage of total area trawled 0 times, (2nd row) percentage of total area trawled more than 0 times but less than 1 time, (3rd row) percentage of total area trawled at least 1 time but no more than 3 times and (4th row) percentage of area trawled at least 3 times but no more than 7 times

Figure 6.1 shows that for the different situations examined it is the spatial resolution that makes the largest difference, because the differences are greatest between the left and right cell sizes within a panel (Figure 1).

These findings are relevant when calculating indicators of fishing impact (i.e. the 'DCF indicators', see Chapter 7) which are described below.

- DCF Indicator 5: Distribution of fishing activities. The spatial extent of fishing activity based on the total area within which VMS records were obtained, each month.
- DCF Indicator 6: Aggregation of fishing activities. The extent to which fishing activity is aggregated based on the total area within which 90% of VMS records were obtained, each month.
- DCF Indicator 7: Areas not impacted by mobile bottom gears. The area of seabed that was not impacted by mobile bottom fishing gears in the last year.
-

In this example, DCF indicator 7 can be related directly to figure 1 (1st row), where the area not impacted by bottom gear is represented as the percentage of total area not trawled. Although there is considerable difference (~13%) among the spatial resolutions used (Table 6.1), there is hardly any difference by interval rate. Note that the large differences mainly occur in the upper graphs with low trawling intensities which are considered the most critical areas when it comes to estimating the effects of trawling.

Table 6.1. Relative difference in percentage trawled between the outer ranges of interval rate given for each spatial resolution scenario, as graphically presented in figure 1.

Spatial resolution	Relative difference in percentage trawled between outer ranges of interval rate			
	0 times trawled	1 times trawled	2-3 times trawled	4-7 times trawled
0.5	2.76%	17.00%	-	-
0.125	2.31%	23.67%	-	-
0.05	3.03%	41.02%	50.00%	-
0.00125	3.84%	83.90%	69.26%	64.29%
0.005	1.57%	100%	74.00%	91.08%
0.00125	0.10%	-	100%	100%

6.3.2 Based on trawl tracks

The effect of recording rate can also be assessed for reconstructed fishing tracks, based on an interpolation between succeeding fishing VMS recordings. Here we show how a cubic Hermite spline ([Hermite 1863](#), [de Boor 1978](#)) interpolation can be used to reconstruct fishing tracks providing that heading, speed and position are available at both the beginning and endpoint of the interpolation ([Hintzen et al. 2010](#)) (see Appendix XI). Details of this method are given below in chapter 6.4. The cubic Hermite spline (cHs) method is contrasted against a linear interpolation (SL) method. We try to answer the following questions:

- 1) To what extend does the actual spatial location compare to that based on the interpolation
- 2) To what extend does the real distance covered compare to that based on the interpolation

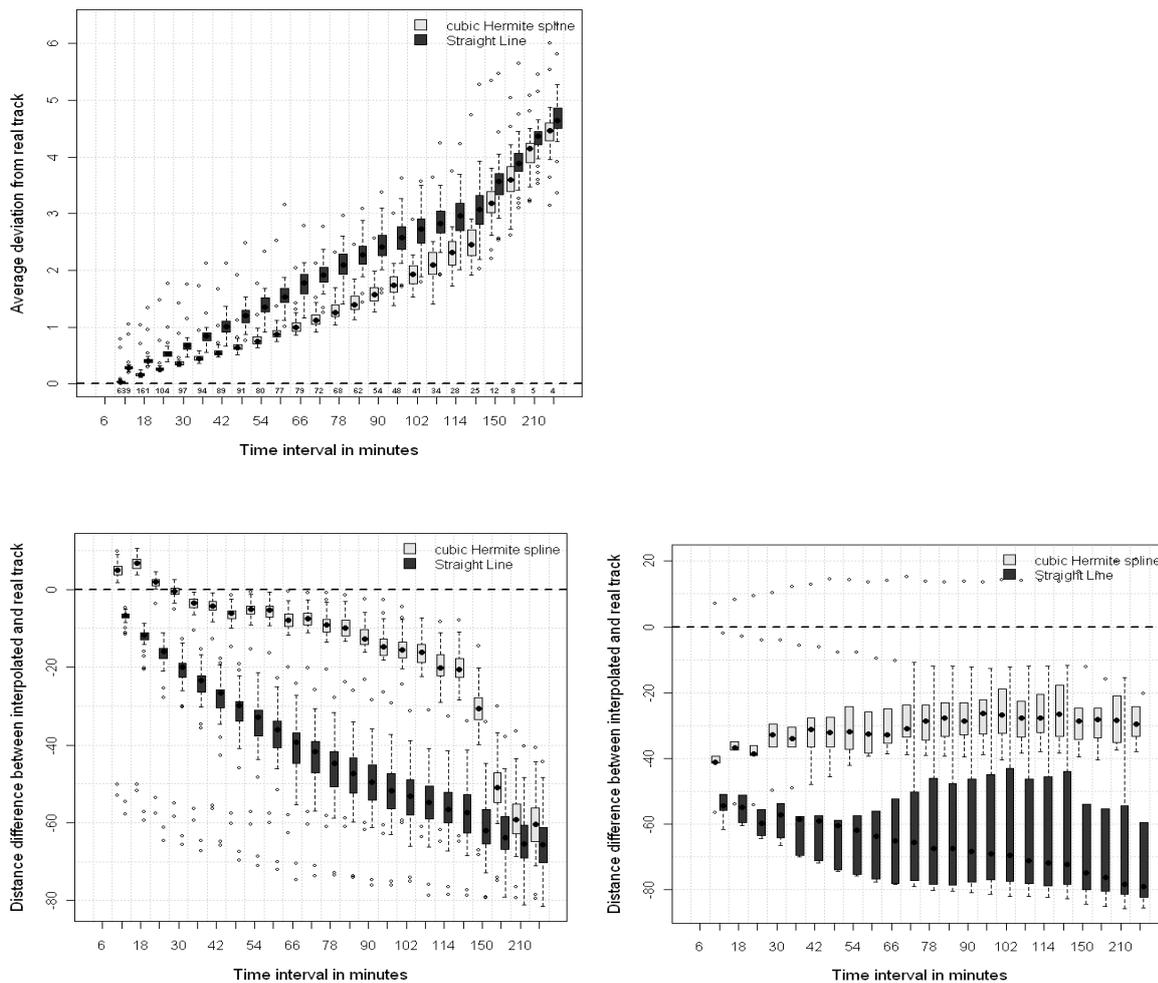


Figure 6.2. Comparison of performance of interpolation methods. The top panel indicates, for the Dutch dataset only, to what extent the interpolations deviate from the known tracks measured as the distance between the interpolated and known track at 10 intermediate points in kilometres (where the values on top of the x-axis represent the relative difference between the Straight Line and cubic Hermite spline interpolations). The bottom panels indicate the percentage of deviance in distance from the known tracks where the left panel shows the results obtained from the Dutch data while the left panels shows the results obtained using the English data.

The bottom panels of Figure 6.2 show that under all circumstances, the interpolated track is closer to the known track than the straight line track and therefore always results in a more precise identification of the spatial distribution of fisheries. Furthermore, the estimated distance of trawling tracks is much more precise when the cubic Hermite spline method is used as is given in the bottom panels. The results from the English dataset are less accurate compared to the results from the Dutch data because the absolute distance trawled is smaller in the English dataset and therefore results in higher relative differences. However, when recording rate changes to levels higher than 2 hours, a significant reduction in performance can be observed for the Dutch data in combination with the cubic Hermite spline method while the Straight Line method continues to decline almost linearly.

6.3.3 Interpolation algorithm

Existing algorithms to create tracks from VMS recordings are evaluated extensively and a promising newly developed algorithm is described in a paper by Hintzen et al. (2010) (see Appendix XI). This paper addresses and answers the Deliverable 5.2 (an algorithm that allows the best estimation of the actual vessel track from VMS position registration data). This algorithm was used for analyses on our perception of the spatial distribution of fisheries, as is presented in chapter 6.1 approach 2. It should also be noted that functions for estimation of these tracks, starting from TACSAT input format is available in the **vmstools** package (see Appendix I).

The figures below describe how the cubic Hermite spline method is developed, relying on two consecutive VMS pings while using their position, heading and speed to define an interpolation. A 24 period of trawling is given in the right hand figure where a straight-line interpolation and cubic Hermite spline interpolation are compared to the known track.

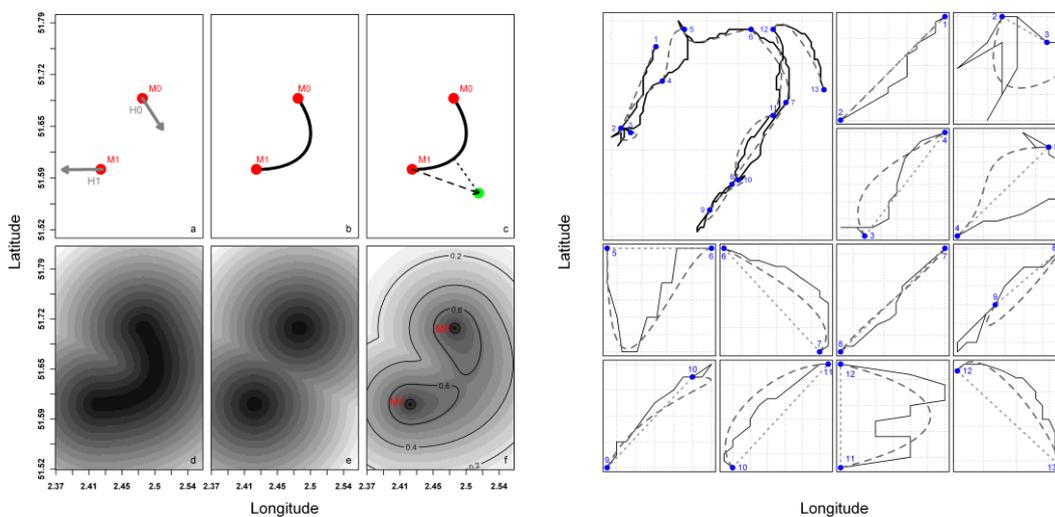


Figure 6.3 left. Schematic representation of the process starting with two succeeding VMS position registrations towards an estimated track surrounded by a confidence interval. (a) The start- and end-point of the vessel are represented by M0 and M1 respectively, the heading of the vessel at start- and end-point are represented by the small arrows H0 and H1. Based on the value of f_m these arrows become longer/shorter influencing the curvature of the interpolation (see panel b). (b) Interpolated track based on cubic Hermite spline (black solid line). (c) The parameter DSD for a random point on a grid (green dot) depends on the distance marked by the dashed arrow (black dashed arrow) from M1 to the green dot. The parameter D_{scale} for a random point on a grid (green dot) depends on the shortest distance from the green dot to the interpolated trajectory. This distance is represented by the dotted arrow (black dotted arrow). (d) Shortest distance from each point on a grid to the interpolated track. Lighter grey represents more distant grid cells. This distance is used to define the value of parameter D_{scale} for each point on a grid. (e) Shortest distance from each point on a grid to either M0 or M1. Lighter grey represents more distant grid cells. This distance is used to define the value of parameter DSD for each point on a grid. (f) Interpolation between two succeeding VMS data points surrounded by a confidence interval. At positions M0 and M1, values equal one.

Figure 3 right. Estimation of the real trawl track (top left, solid black line) consisting of 13 succeeding 2-hourly position registrations (blue points) obtained from APR dataset1. The chs interpolation is represented by the dashed dark grey line. (Smaller panels) Estimations of the real trawl track between two succeeding position registrations (blue points) (corresponding **to the tracks shown in the top-left panel**), the chs interpolation (dark grey dashed line) and straight line interpolation (light grey dotted line).

6.3.4 Effect of recording rate

Two different approaches were followed to assess the effect of recording rate and based on this, advice on an optimal recording rate. The first approach uses solely VMS records, while the second approach involves application of an interpolation method. In the first approach, no clear recording rate optimum can be found where the perception of spatial habitat use only differs slightly in absolute terms, given a certain spatial resolution. This is also shown by figure 4, where the relative difference between interval rates is graphically represented for those areas trawled zero times or less than once. In all occasions, the relative differences between recording rates are also small. Hence, we can conclude that given any interval rate between 6 and 240 minutes, our perception of the spatial distribution of fisheries will only change slightly given that indicators of trawling intensity have major importance.

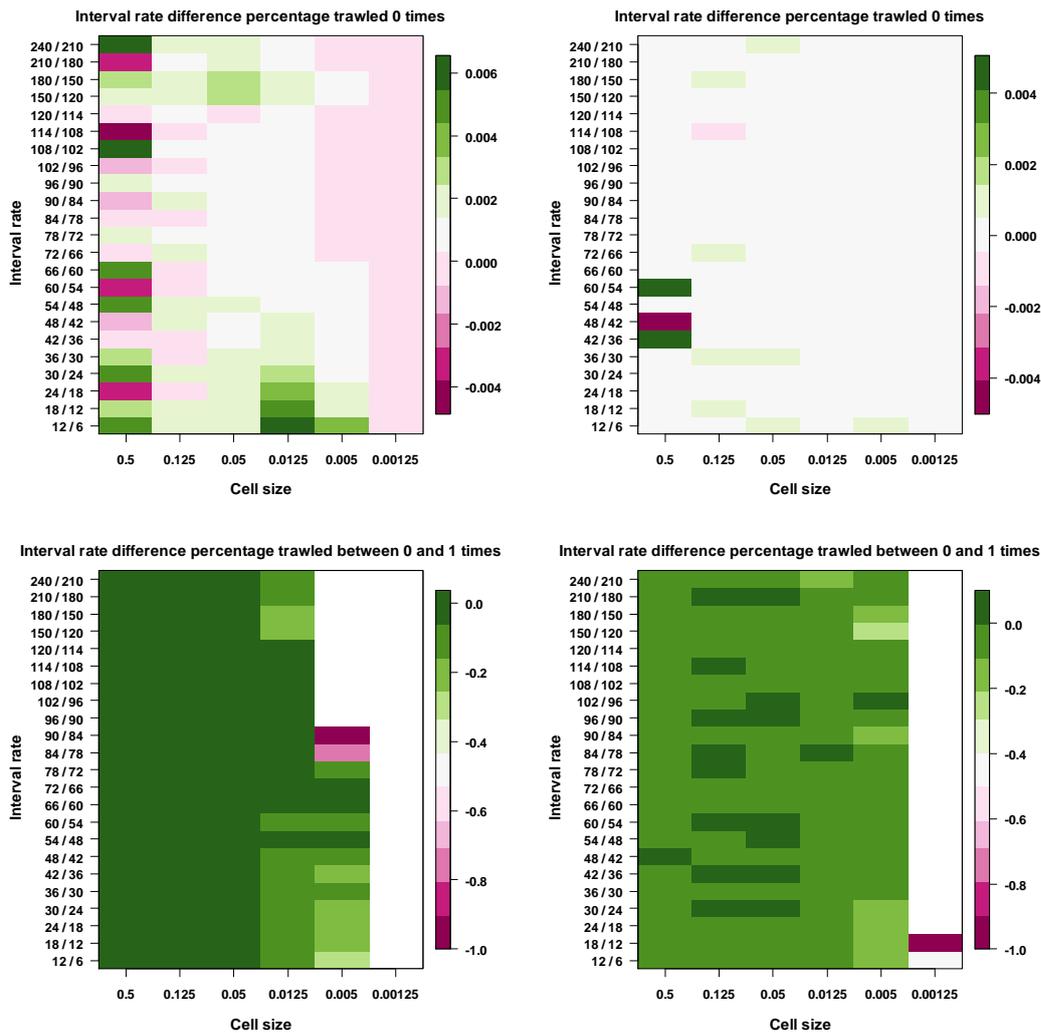


Figure 6.4. Proportional difference (per mill) in proportion trawled between interval rates (y-axis) and spatial grid cell size (x-axis). Left panels show the results as obtained from the Dutch data while the right panels show the results from the English data.

Figure 6.4 shows the relative difference between consecutive interval rates of the percentage of the total area trawled. This indicates where a relationship between interval rates might be strongest and could indicate as a preferred interval rate. In none of the figures above a clear transition is visible, indicating that there is no interval rate that markedly affects our perception of the spatial distribution of fishing.

In the second approach, the deviance from the known truth did not show an optimum but its performance decreased almost linearly. The relative difference in distance however showed for both the English as the Dutch data an optimum. The bottom row of Figures 6.2 has been combined into one figure to judge the combined performance where the variance of each of the datasets was weighted against the number of vessels present in each dataset (resp. 24 in the Dutch data and 6 in the English data).

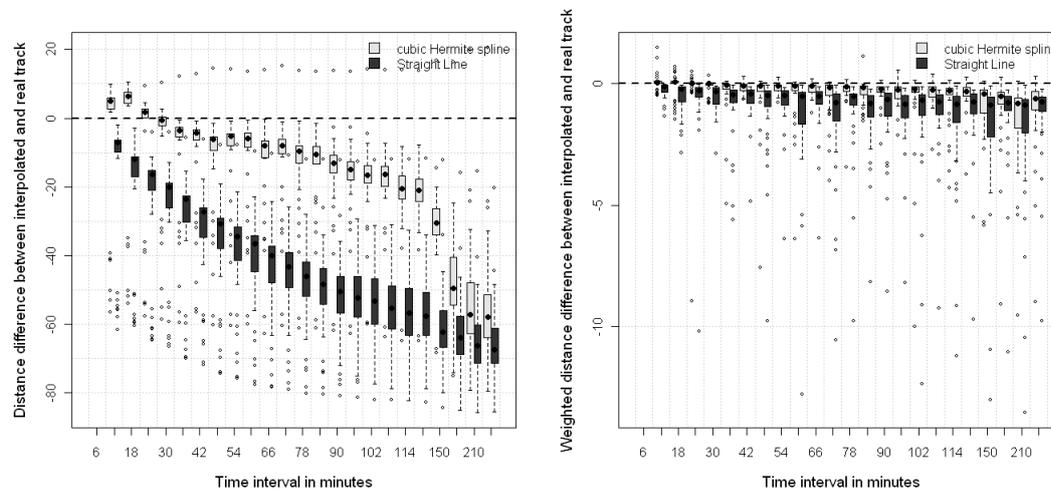


Figure 6.5. Indication of the percentage of deviance in distance from the known tracks for the combined Dutch – English dataset. Left panel indicates the absolute combined percentage of deviance while the right figure represents the combined weighted percentage of deviance.

Figure 6.5 left shows that at approximately 30 minutes, the difference in distance is estimated best, with slight overestimation of distance at lower interval rates and underestimation at higher interval rates. The right panel of Figure 5 shows the weighted relative difference where all observations have been divided by the interval rate. This is because increasing interval rate might also increase costs of storing and processing. In the right panel we can observe the best performance around an interval rate of 24 to 30 minutes.

6.4 Conclusions

The effect of recording rate was evaluated applying two approaches using datasets with high-frequency position registrations and reconstructed trawl tracks. The first approach showed that, not surprisingly, an increased recording rate results in more position registrations and thus more accurate estimates of fishing impact. For the specific dataset that was available to evaluate recording rate this only made a difference at specific grid sizes for specific trawling intensities. These differences, however, could amount up to several orders of magnitude.

When (reconstructed) trawl tracks based on different interpolation techniques were used a similar pattern emerged where increased recording rates result in more accurate estimates of fishing impact. As now the choice of recording rate was determined by the performance of the interpolation technique at the specific recording rates, two criteria were applied to evaluate the performance of the interpolation technique: (1) deviation from real track in terms of position and (2) difference from real track in terms of the estimated length of the track. The best interpolation technique (i.e. cubic Hermit spline) always performed better with increasing recording rate in case of the first criterion.

When aiming for the best recording rate to estimate the spatial extend of fishing, it can be concluded that the lower the recording rate the better. However, this might be very costly. Therefore, another analysis, indicating the track length deviation was performed and clearly showed an optimum in recording rate. Especially when weighted against a cost function, the performance was optimal at an interval rate between 24 and 30 minutes.

7. Describing the spatial distribution of fishing activity

7.1 Introduction

All of the main marine policy frameworks have adopted an ecosystem approach to the management of the anthropogenic activities and conservation of their resources. For ecosystem-based management (EBM), indicators are required to describe the pressures affecting the ecosystem, the state of the ecosystem, and the response of managers ([Jennings 2005](#)). To implement EBM successfully it is also critical that we understand the links between these indicators so that the policy objectives on the state of the ecosystem can be achieved by managing the human pressures through an appropriate response.

As the ecosystem approach is about the conservation of ecosystem integrity including all the different structural components and their functioning the various legislative frameworks operating on the marine system have all adopted different ways of distinguishing the various components and attributes of the ecosystem that need to be conserved in order for the ecosystem to be in a healthy state or achieve "Good Environmental Status"(GES) ([CEC 2007](#)). An important component that according to all of these frameworks needs to be conserved is the habitat ([CEC 2002](#), [2007](#)). Different habitats can be distinguished, e.g. water column, rock, biogenic reef or sediment (sand, mud, gravel) each with their specific communities of living organisms. For habitats there are few, if any, appropriate indicators of state for lack of sensitivity or responsiveness ([Hiddink et al. 2006a](#)). Therefore a different and complementary perspective is required where pressure indicators are used ([Hiddink et al. 2006a](#)) and the absence of human activities that disturb these habitats is considered the best indicator of GES. The advantage of this approach being that the absence of human activities by definition qualifies this habitat as having GES. For bottom habitats the main human activity known to affect them is fishing, specifically bottom trawling ([Duplisea et al. 2001](#), [Hiddink et al. 2006b](#)).

Information on the spatial and temporal distribution of fishing effort are required for pressure indicators describing the impact of fishing on the seafloor ([Piet et al. 2007](#)). The introduction of vessel monitoring by satellite systems (VMS) as a surveillance and enforcement tool has revolutionized the study of the spatial and temporal distribution of fishing effort, providing high-resolution real-time data for most of the larger fishing vessels ([Lee et al. 2010](#)). For that reason the European Union adopted VMS as part of their Data Collection Framework and in order to assess fishing impact on the wider ecosystem (i.e. other than the fish stocks) adopted a suite of indicators of which some described the spatial extent of fishing and the impact on the seafloor ([EC 2008](#)). The tools developed within work-package 6 and presented in this chapter of the report are intended to facilitate the calculation of three indicators to describe the spatial distribution of fishing effort. They are summarised as follows:

- DCF Indicator 5: Distribution of fishing activities. The spatial extent of fishing activity based on the total area of grids within which VMS records were obtained, each month;
- DCF Indicator 6: Aggregation of fishing activities. The extent to which fishing activity is aggregated based on the total area of grids within which 90% of VMS records were obtained, each month.
- DCF Indicator 7: Areas not impacted by mobile bottom gears. The area of seabed that has not been impacted by mobile bottom fishing gears in the last year. Could be reported annually and would state the total proportion of the area by depth strata in each marine region.

7.2 Material and methods

In order to calculate these indicators some functions were developed that are now part of the **vmstools** package. As a first step the TACSAT point data need to be aggregated into a gridded data frame using the functions *mapGrid()* and *vmsGridCFeate()* or *createGrid()* and therefore the resolution of the grid (cell size) must be defined.

DCF 5 calculates the total area of a grid of cells with fishing activity which is above a minimum threshold of number of pings or number of fishing hours (if the TACSAT data contains a field with time interval between two points called *SI_INTV*, then the threshold will be a minimal number of hours, otherwise it will be a minimal number of points). The area of each cell is calculated with the function *surface* either via a fast and rough method using a trapezoid approximation (option "Trapezoid"), either via a more precise but slow method using a Universal Transverse Mercator projection (option "UTM"). The first method is fine for big grids of small cell sizes, the second one is better for large cell sizes. This total fishing area is processed by month.

DCF 6 also calculates the total area of a grid with fishing activity but keeps only 90% of the points by discarding the outer 10% points (or any other specified percentage). It uses the function *tacsatMCP()* adapted from the *aspace* package. This function draws a minimum convex polygon around the central points to keep. Then these points are gridded and the total area of the cells is calculated with the *surface* function with the same optional methods as DCF 5. This total fishing area is processed by month.

DCF 7 calculates the total area of a specified polygon *not* impacted by mobile bottom gear. It therefore requires that the TACSAT data have been merged with the logbooks in order to have a gear code (or others) for each VMS point. The indicator needs a list of gear codes to include as mobile bottom gears (if empty, all the points will be included). The specified area to be processed is a polygon shapefile. This polygon (or group of polygons) is then clipped with the polygon marking the sea border of Europe to be sure that the indicator won't include land in its area calculation. If no shapefile is defined, the area of the bounding box containing all the VMS points will be considered. The result is the area of the polygon less the area of the grid where fishing activity occurs. The VMS pings are gridded with an optional threshold in either minimal of fishing hours or minimal number of points (see DCF 5). The area of each grid cell is calculated with the *surface* function (see DCF 5 or DCF 6).

7.3 Applications

To illustrate the use of the **vmstools** package and how the calculated indicators can be applied, the tools have been used to calculate the three DCF indicators using the following two example datasets: (1) activity by all VMS-equipped UK-registered fishing vessels during 2008; and (2) activity by Dutch beam trawlers >300hp in 2009. ([Anon 2007](#)) specifies that these indicators should be based on a grid size of 3km x 3km, with the proviso that this scale could be revised if necessary. The tools allow for flexibility in the grid-size selected, and for these examples a grid size of 0.05 degrees by 0.05 degrees has been selected. Similarly, the calculation of DCF 5 allows for a threshold number of VMS pings to be specified to define whether or not an individual grid square is classified as fished or unfished. The calculations here assume a threshold of zero pings. To aid interpretation, the tools include the option to map the data underlying the indicator values.

DCF indicator 5 is intended to be an indicator of the spatial extent of fishing activity in terms of the area of grid-squares where fishing activity with a particular gear (or all gears in the example used here) was recorded. The indicator would normally be used in conjunction with DCF 6 which is an indicator of the aggregation of fishing activity. It uses the same underlying data as DCF 5 then fits a polygon which contains the inner 90% of grid squares with fishing activity. The indicator is then the area of squares within this polygon where fishing activity has occurred during the specified month.

For the UK data, Figure 7.1 shows the maps of the fishing activity data used to calculate DCF indicator 5, with Figure 7.2 showing the same data after exclusion of the outer 10% for calculation of DCF indicator

6. Figure 7.3 shows the monthly values for DCF indicators 5 and 6, with DCF6 given as a percentage of DCF5 in Figure 7.4. In terms of total area affected, the distribution of fishing activity apparent from these data is relatively stable from month to month, although the area affected is lowest in December. DCF indicator 6 can be most usefully interpreted relative to DCF5 (Figure 7.4). In the example data set, the area corresponding to the inner 90% of fishing activity varies between 76% and 83% of the total area affected. The percentage is highest in January, suggesting that fishing is least aggregated at that time of year.

DCF indicator 7 is an estimate of the area not impacted by mobile bottom gears within a specified area and time period. Figure 7.5 shows the example area used in this case; a combination of two polygons covering areas off of the north-east coast of Scotland and the south-west of England. The value for this indicator for the test area for 2008 is 59842 km², which can also be given as a proportion when to total gridded area is taken into account.

Table 7.1, Monthly values for DCF indicators 5 and 6 for UK fishing vessels equipped with VMS equipment, 2008.

Month	DCF5 (km²)	DCF6 (km²)
Jan	262564	216701
Feb	270280	212817
Mar	261176	198838
Apr	276274	209969
May	279125	218802
Jun	261855	201029
Jul	281644	216982
Aug	287327	218267
Sep	284251	224024
Oct	262803	209111
Nov	266209	204542
Dec	223748	173616



Figure 7.1. Monthly distribution of fishing activity by UK-registered fishing vessels equipped with VMS, 2008. The total area of fishing activity in these maps is summarized as DCF indicator 5 in Table 7.1.



Figure 7.2. Monthly distribution of fishing activity by UK-registered fishing vessels equipped with VMS, 2008 showing only the inner 90% of fishing activity. The total area of fishing activity in these maps is summarised as DCF indicator 6 in Table 7.1. The plot for January also includes the fitted polygon used to define the inner 90%

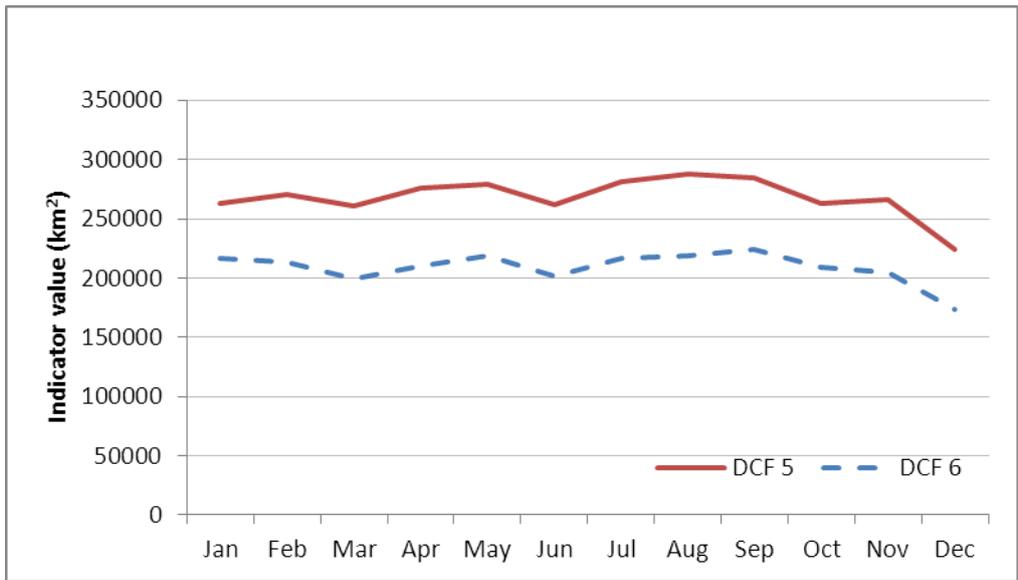


Figure 7.3, Monthly values for DCF indicators 5 and 6 for UK fishing vessels equipped with VMS equipment, 2008.

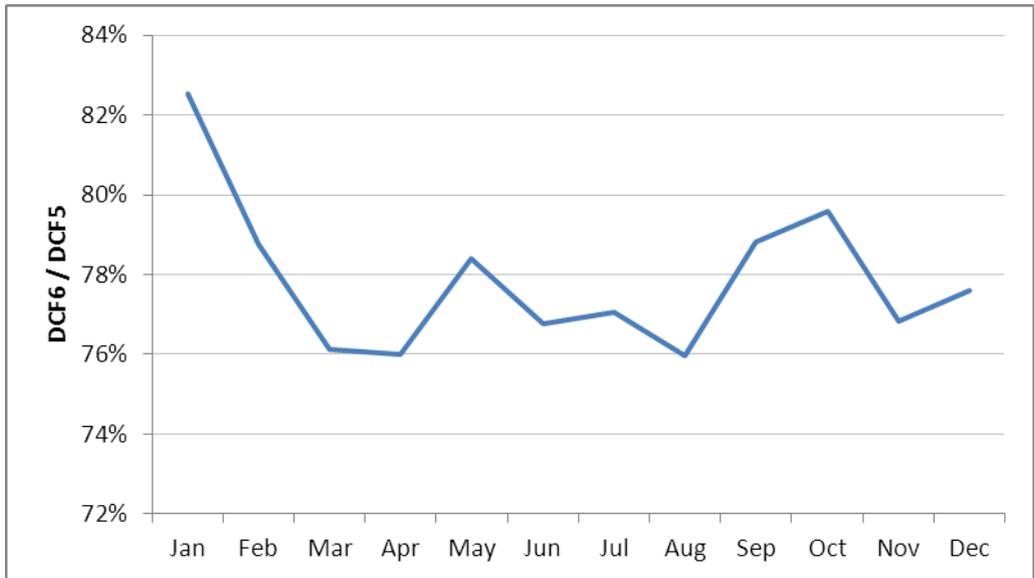


Figure 7.4. DCF indicator 6 as a percentage of DCF indicator 5. Data are for UK fishing vessels equipped with VMS equipment, 2008.

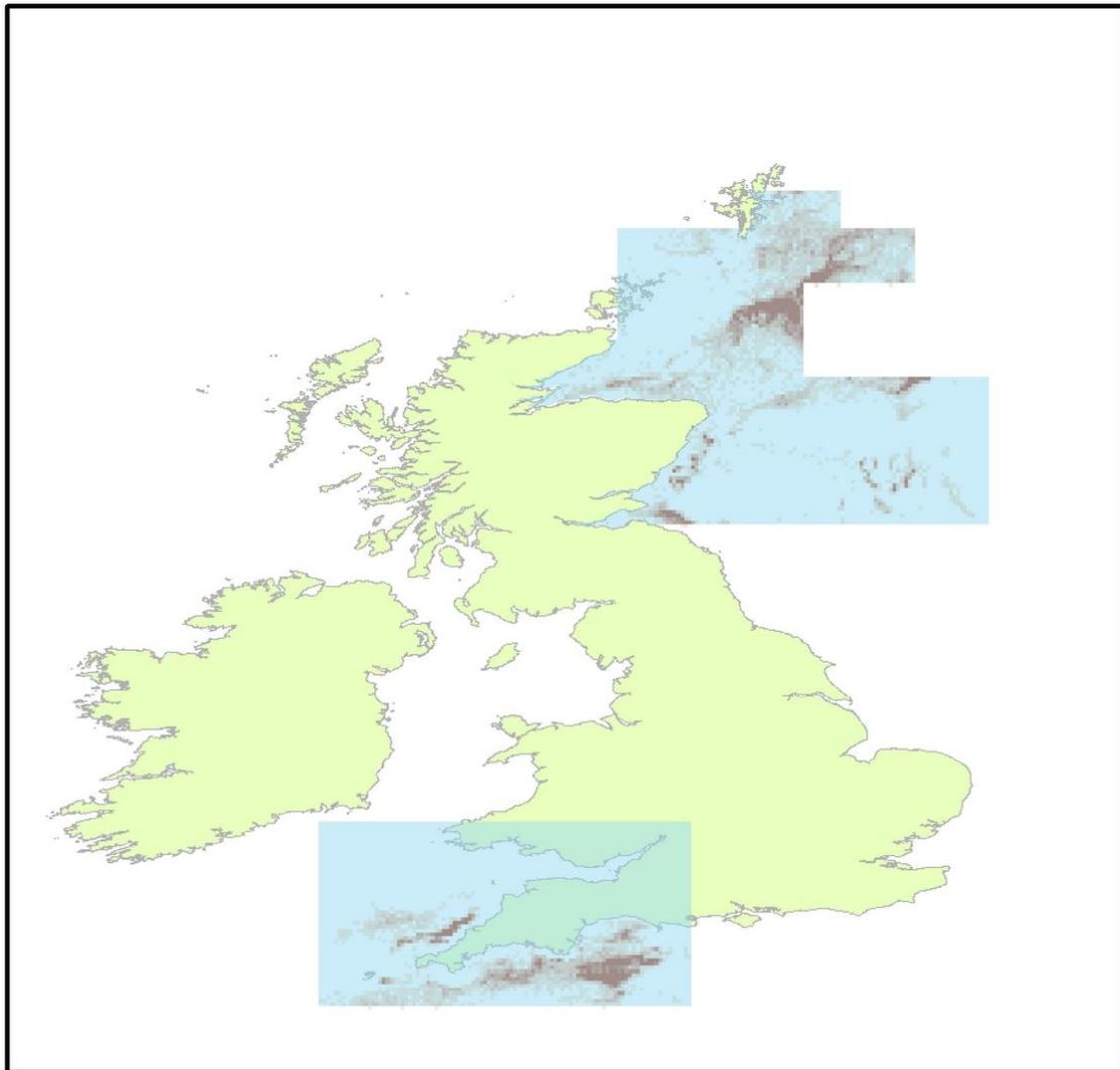


Figure 7.5, Fishing activity in the test area used for calculation of DCF indicator 7 for UK fishing vessels in 2008.

To further illustrate the functionality of **vmstools** DCF indicators 5 and 6 were also calculated for large (>300hp Dutch beam trawlers) and the results are displayed for February 2009 in Figures 7.6 and 7.7 and Table 7.2. The area covered by the Dutch fleet is much less than that covered by the British (~50000kms versus ~ 240,000kms) although the Dutch estimates appear to be more strongly seasonal with activity increasing steadily throughout the year (Tables 7.1 and 7.2).

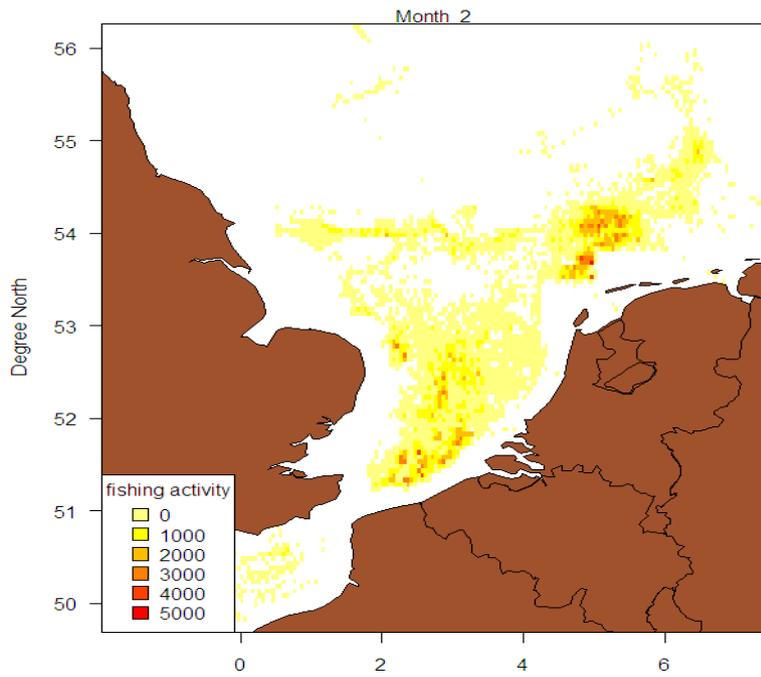


Figure 7.6. Distribution of fishing activity by Dutch-registered fishing vessels >300hp in February 2009 fishing beam trawls. The total area of fishing activity in these maps is summarized as DCF indicator 5 for every month in 2009 in Table 7.2.

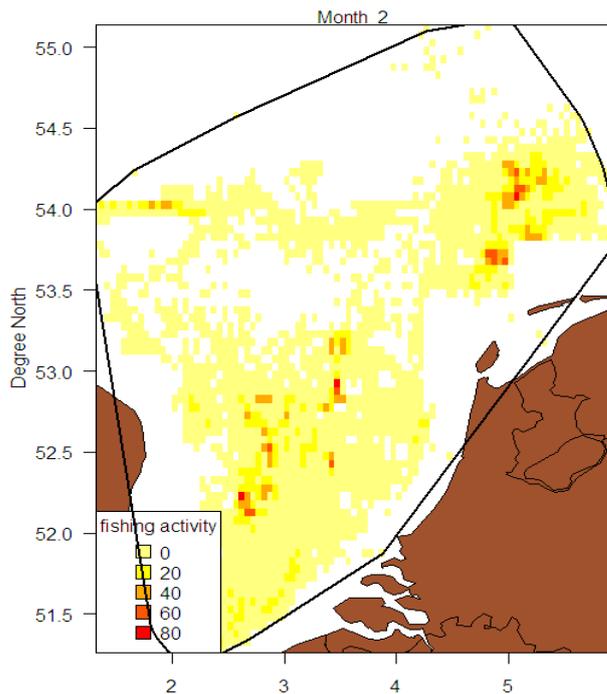


Figure 7.7. Distribution of fishing activity by Dutch-registered fishing vessels > 300HP in February 2009 showing only the inner 90% of fishing activity. The total area of fishing activity in these maps is summarized as DCF indicator 6 in Table 6.1. The plot also includes the fitted polygon used to define the inner 90%.

Table 7.2. DCF level 5 and 6 indicators for Dutch flag vessels > 300hp fishing beam trawls in 2009.

month	dcf5	dcf6
1	53717.47	42358.85
2	52570.39	40480.51
3	56592.65	45773.97
4	56110.5	47130.44
5	54279.34	47441.51
6	56641.46	51781.49
7	58234.82	52291.63
8	59526.54	50333.81
9	68028.12	56758.89
10	65095.59	54698.59
11	63474.39	53173.55
12	67042.65	57048.93

7.4 Conclusions

The tools developed in this project to describe the spatial distribution of fishing activity calculate the three indicators that were proposed as part of the EU Data Collection Framework. As required the tools are able to calculate these indicators on a monthly basis and also at different grid sizes even though initially only one specific grid size was proposed. The examples here are based on UK and Dutch data, but the tools can easily be applied to any set of National or regional data provided they are in the agreed format (see Chapter 2). It should be noted that the function *indicator* described above and also in the manual (Appendix I) outputs an ASCII matrix of ARCGIS format, and that another function in the **vmstools** package called *mosaicASCIIGrids()* can combine such matrices easily making 'integrations' of data between nations very easy.

8. Synthesis: *Vmstools* software package

8.1 Introduction

The Common Fisheries Policy (CFP) requires gradual implementation of an ecosystem-based approach to fisheries management (EBFM). This will include the integration of environmental protection requirements into the CFP, including measures to 'limit the environmental impact of the CFP'. Hence, the collection of data that describe the impact of fishing by the European Member states is clearly needed. Furthermore, as part of the revision of the CFP and implementation of EBFM, there is a requirement to move towards a fleet and area-based management (Council Regulation N° 199/2008⁷) system, and to develop indicators of fishing impact where identified. Detailed measures of fishing effort will also be needed to meet the requirements from the ICES Science Plan 2009-2013, especially for the subtopic 'Impacts of fishing on marine ecosystems' under the thematic area 'Understanding of interactions of human activities with ecosystems'.

Currently, there are two main data types for describing the impact of fishing activities on the ecosystem available: EU logbooks and VMS (Vessel Monitoring by Satellite). The logbooks summarise landings by vessel trip and have information on gear, vessel power, mesh size etc. The spatial resolution is coarse (ICES statistical rectangle) and there are instances of 'mis-reporting'. VMS data are simply vessel positions, recorded in most occasions by satellite every 2 hours, together with instantaneous speed, heading, and a vessel identifier, but there is no information, for example, on the gears used or fish landed. In order to address the questions described above we clearly require more detailed information on fishing activity than can be obtained from logbook and VMS data when each is considered in isolation.

8.1.1 Data issues and regulations

Throughout the project issues regarding the use and dissemination of VMS and logbook data have created problems, hampering our ability to undertake real 'regional' data analyses where combined data from multiple countries in a single database on a single computer would be necessary. It is also clear that each country interprets the rules differently. These regulations are summarised after SGVMS (ICES Study Group on VMS data, its storage, access and tools for analysis): "The EU Data Collection Framework (DCF), EC Regulation 199/2008, requires Member States to collect certain data under 'multi-annual national programmes', prescribes the process for collection, management and use of that data and provides for data collected in the framework of the Common Fisheries Policy (CFP), including VMS data, to be used for the purposes of such 'national programmes'. It requires Member States to provide anonymised data to 'end-users' to support scientific analysis as a basis for advice to fisheries management; in the interest of public debate and stakeholder participation in policy development, and for scientific publication (Article 18). 'End-users' are defined as bodies with a research or management interest in the scientific analysis of data in the fisheries sector. This regulation does not provide a guaranteed right of access to VMS data, which is generally considered personal information, obtained via surveillance. However, the right to withhold the data is limited, the most relevant reason being the risk of natural or legal persons being identified (Article 20). The DCF is concerned with improving the quality of information and scientific advice available for implementation of the CFP, therefore is entirely CFP-related. This obligation does not directly apply to data sharing for marine planning purposes, unless such marine planning is integral to the CFP as an environmental consideration or requirement.

Administrations of EU member states powers to share VMS data for non-CFP purposes is constrained by a combination of human rights law; data protection law; the law of confidence, and EU law - in particular the EU confidentiality obligation under Article 113 of EC Regulation 1224/2009 (the "Control Regulation"). When sharing VMS data without the sphere of the CFP, compliance with the EU confidentiality obligation cannot be guaranteed, however, it is arguable that sharing anonymised VMS

⁷ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:060:0001:01:EN:HTML>

data for marine planning purposes is not contrary to human rights law, data protection law or the EU confidentiality obligation if certain safeguards are put in place to protect the commercial value of VMS data and preserve confidentiality. Such safeguards could require a clearly defined and legitimate purpose to be defined before data would be shared, either aligning with the CFP's objectives or "conservation and protection of the wider marine environment" as described in the Control Regulation, a demonstration that disclosure is necessary to fulfil that purpose and proportionate (i.e. no alternative means could achieve the same aim); the data is anonymised and aggregated to prevent the identification of any natural/legal person, and access is restricted to individuals or bodies whose functions require them to have access and that adequate safeguards are in place to prevent further unauthorized disclosure. In the EC No 1077/2008 on electronic recording and reporting of fishing activities and on means of remote sensing, it is stated that "In each Member State, a single authority shall be responsible for transmitting, receiving, managing and processing all data covered by this Regulation." It is unclear how processing for scientific purposes relates to this. Regulation EC No 1077/2008 also states that: "A flag Member State shall ensure that a coastal Member State has on-line access in real time to electronic logbook and landing declaration data of vessels flying its flag when conducting fishing operations in the waters under the sovereignty or jurisdiction or entering a port of the coastal Member State." In the EC No 2244/2003 laying down detailed provisions regarding satellite-based Vessel Monitoring Systems, no reference is made to access to the data for third parties (except for specific requests of the Commission). The regulation does state that the all VMS-related data should be treated in a confidential manner." The purpose of this project, therefore, was to develop tools that integrate both logbook and VMS data in order to extract more information than each data source can provide alone. In particular much greater spatial and temporal detail in the impact of fishing activity is available when the two data sources are combined. In this report we show how we approached the problem by building a single software package capable of undertaking the tasks required to simultaneously merge, analyse, and visualize VMS and log book data.

8.2 Material and methods

The package that allows an integration of (inter-)national logbook and VMS data in order to provide (regional) descriptions of fishing impact is built using R (<http://cran.r-project.org/>) and is called **vmstools**. **vmstools** exploits two standardized data formats, EFLALO (EU logbook data) and TACSAT (the VMS positions) which build on work done and 'agreements' made during other EU funded projects such as CAFÉ, AFRAME, and TECTAC (see Chapter 2 and Appendices III and IV). Once the data have been imported to R, a series of R tools linked by 'scripts' enable a range of tasks to be completed in a unified software environment. *Métiers*, for example, can be identified objectively from species assemblages using multivariate statistical techniques; fishing activity can be distinguished logbook and VMS data can be merged (Bastardie et al. 2010) and individual vessel tracks can be interpolated both linearly and non-linearly using Hermite spline functions (Hintzen et al. 2010). The package can also be used to explore the impact of differing spatial (grid size) and temporal aggregations (month, quarterly, annual) which can be extremely important for gauging the real effects of fishing activity in any particular area (Piet and Quirijns 2009).

Several recent papers have demonstrated the utility of combined VMS and logbook data for summarizing the activities of European fisheries in unprecedented spatial and temporal detail (Witt and Godley 2007, Mills et al. 2007) using *ad hoc* mixes of software. The **vmstools** package can be used to standardize these analyses leading to compilation of regional, (combined VMS and logbook) databases. As an example we demonstrate how the **vmstools** package can be used to create data suitable for upload to FishFrame (<http://www.fishframe.org/>). FishFrame provides a database and online reporting tool that can facilitate the storage, retrieval and analysis of combined VMS and logbook data from multiple countries. Once data from different EU Member States are in FishFrame, data can be integrated to produce regional summaries using the tools available in the **vmstools** package. The **vmstools** package environment also ensures that the complicated processing procedures required for this type of analyses are identical, or at least compatible. So far data from multiple countries have, indeed, been uploaded to FISHFRAME where they can be accessed and visualized.

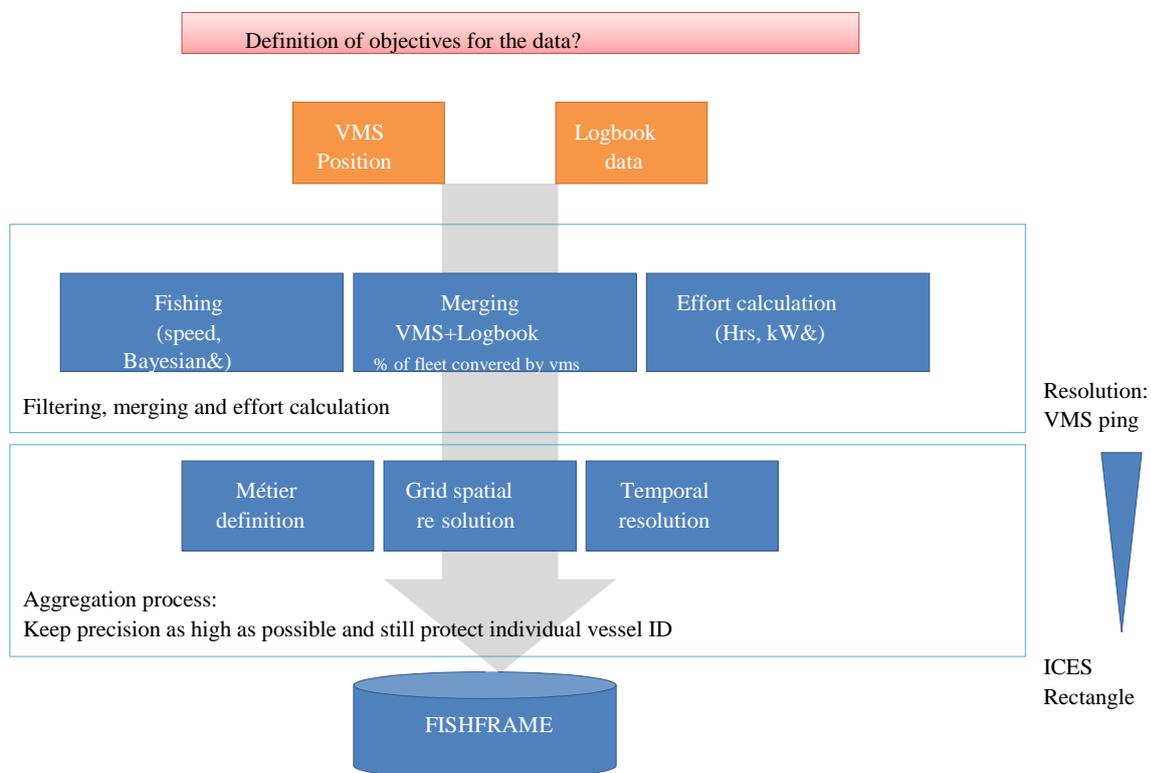


Figure 8.1. Schematic summarizing processes undertaken by the *vmstools* package before data are uploaded to FISHFRAME. The objectives include selections made on period, area and fleet characteristics. The data can be filtered using speed thresholds or more complex Bayesian analyses determining vessel activity, while effort from the logbooks can be calculated in hours fished or kilo watt hours fished.

8.3 Applications

During the project *vmstools* was extensively used in a wide range of projects and consultancies. In the Netherlands IMARES used *vmstools* to combine VMS and logbook data to produce estimates of fishing activity (various measures) for many clients. Maps of fishing effort components of the Dutch fleet were done for the Fish Product Board (e.g. Figure 8.2);

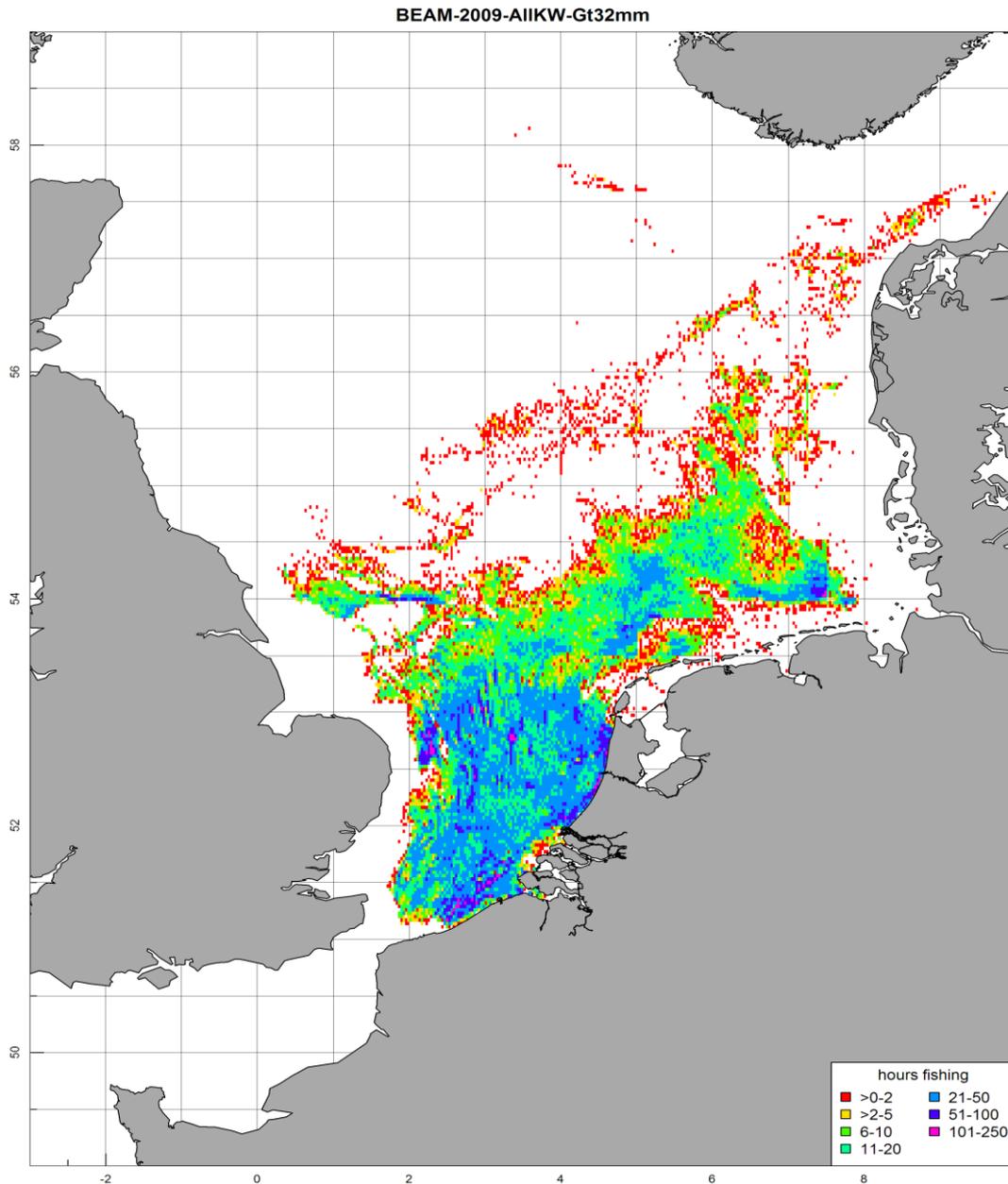


Figure 8.2. Fishing effort in the North Sea (hours fishing) by Dutch flag registered vessels working mesh sizes > 32mm estimated and plotted using *vmstools* for the Fish ProductBoard.

The impact of *INTERNATIONAL* fishing activity in various proposed protected areas within the Dutch EEZ (see Figure 8.3) was assessed using the *vmstools* package during the FIMPAS project (Fishery Impact and Conflicts with Conservation Objectives); the impact of fishing activity on benthos relating to the future development of Rotterdam Harbour was mapped with *vmstools* during the 'Passende Beoordeling Project' and 'Western Scheldt Estuary projects (see Figure 8.4). In a smaller project long term changes (2002-2010) in fishing activity in gullies around the Frisian Island was also assessed (Figure 8.5) while the ability of *vmstools* to output tow tracks in Google Earth (e.g. Figure 8.6) is proving extremely useful for attracting attention to the current project and colleagues are planning to use the methods for assessing the effects of fishery-free areas (see Appendix VII for a discussion).

In early January the *vmstools* package was instrumental in providing advice to the Dutch Ministry on fishing activity by TR2 (otter trawlers and seiners) in the Botney Cut area of the North Sea in relations to 'Seasonal Closures' intended to protect cod (Figure 8.7). In many situations advice on such matters is

required immediately and such rapid responses would not be possible without the **vmstools** package created during this project.

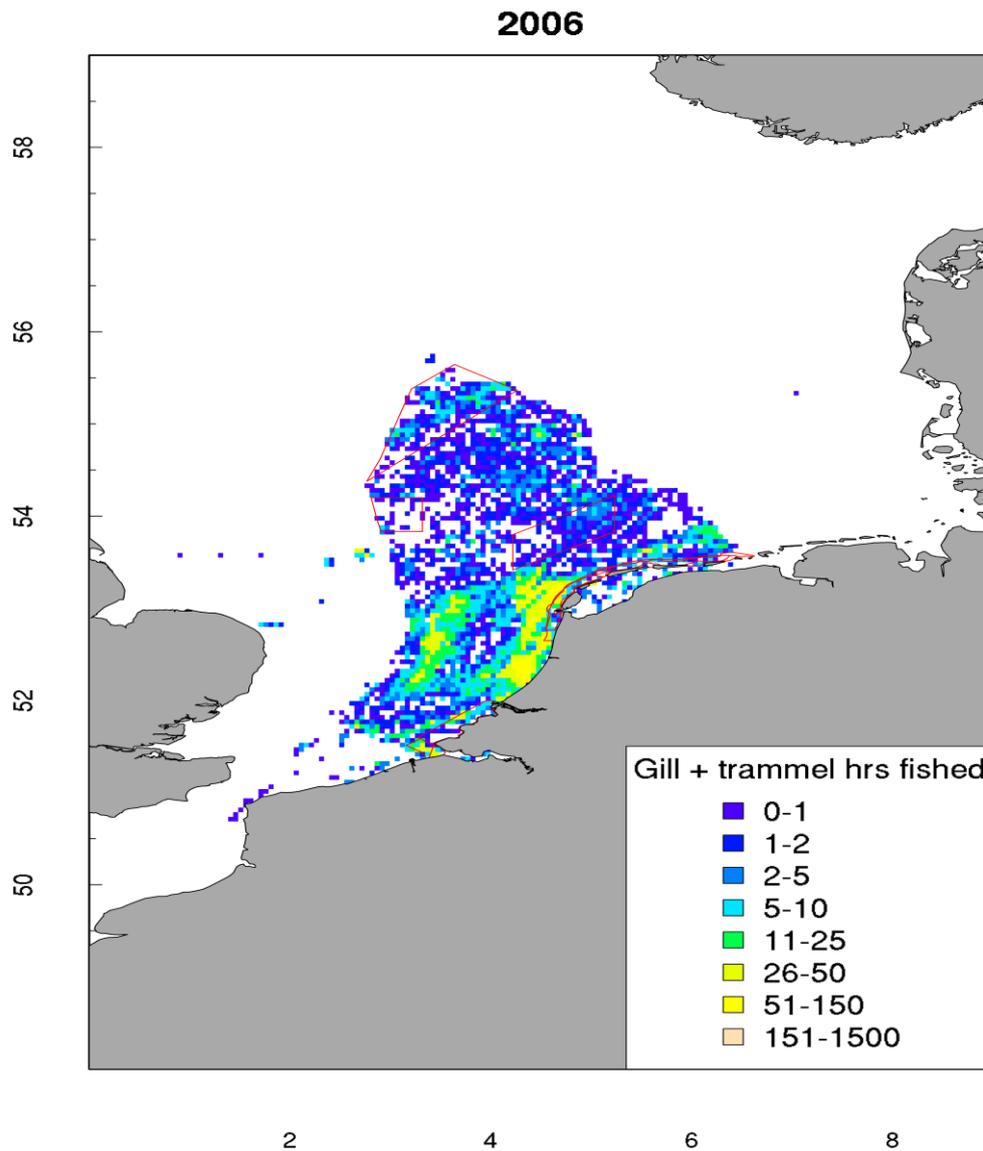


Figure 8.3. Total international fishing effort in the North Sea in the Dutch EEZ in 2006 working gill and trammel nets estimated and plotted using vmstools for the FIMPAS project. Red borders denote proposed protected areas. Note: the summaries here include information from vessels from all EU member states plus Norway that fish in the area. Since many of the ships land in the Netherlands, logbooks can be used to determine gears, mesh-sizes etc. For those that did not the EU fleet register was used.

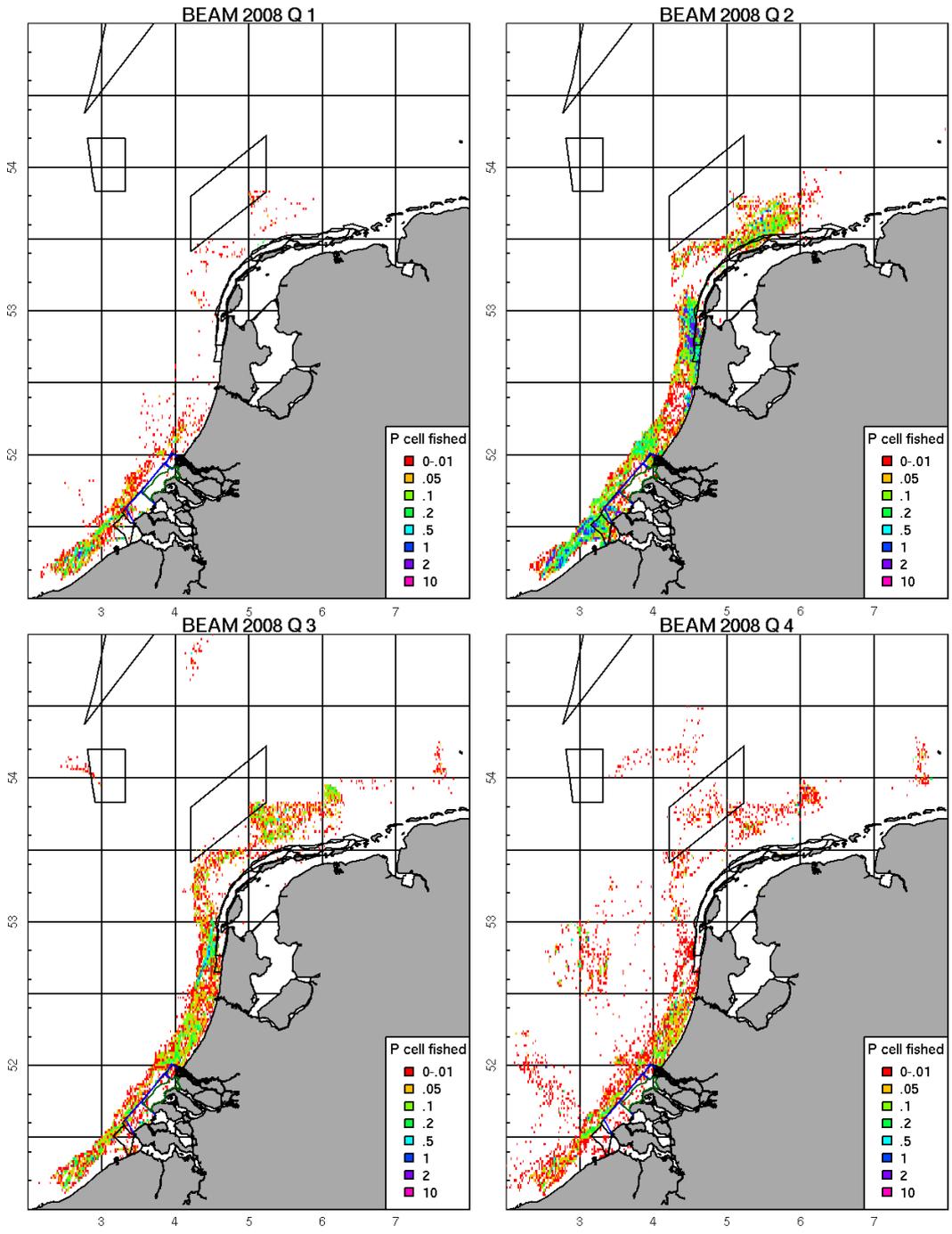


Figure 8.4. Proportion of grid cell fished by ≤ 300 HP Dutch registered beam trawlers by quarter in 2008 estimated using vmstools.

All Mobile Gears 2004

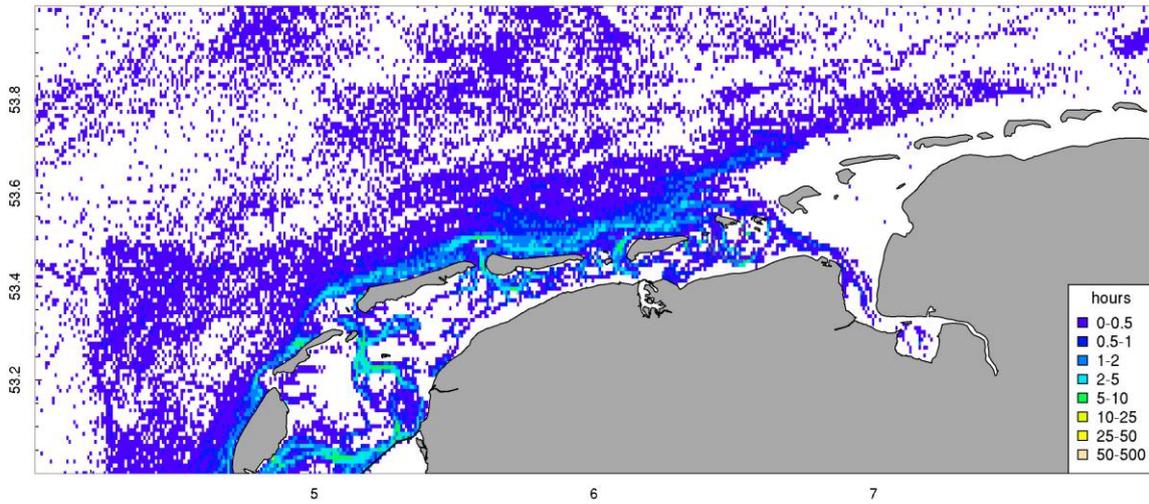


Figure 8.5. Fishing activity around the Frisian Islands by Dutch registered vessels working mobile gears in 2004 estimated and mapped with vmstools.

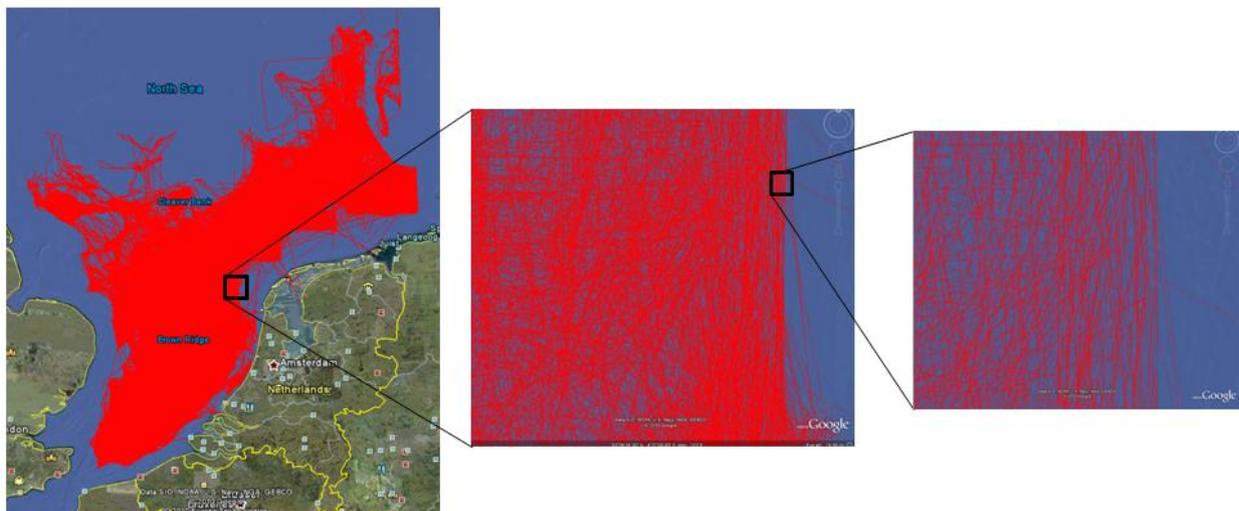


Figure 8.6. Trawl tracks by the Dutch beam trawl fleet (>300HP; year = 2009) displayed in Google Earth© and created using vmstools.

Danish colleagues have used **vmstools** to produce maps summarising the impact of fisheries and fishing effort allocation in the Femern Belt area where a submerged tunnel (initially a bridge) is being proposed to connect the German offshore island of Femern with the Danish Island of Lolland. Further it has already been used extensively for a project on plaice in ICES area IIIa which was collaboration between DTU Aqua and Danish Fishermen. The aim was to improve the stock assessment of plaice and to advice for future possible MSC certification.

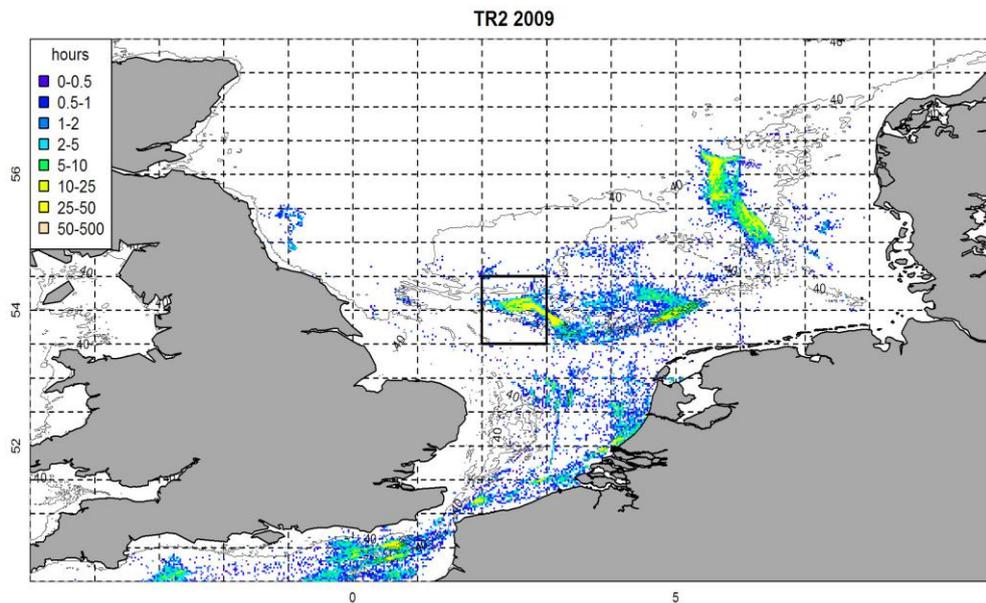


Figure 8.7. Fishing activity by Dutch registered TR2 (otter trawlers and seiners) in the North Sea and English Channel in 2009.

As well, scientists from Bangor University, UK are currently using the **vmstools** package to study the last disturbance event occurred at several stations sampled around the Isle of Man, Irish Sea in relation to benthic production.

The project has been, and will be disseminated in a range of the usual media, i.e. presentations, seminars, workshops and published manuscripts (described below). In addition we have created websites and the **vmstools** package manual (Appendix I). Recently the chairs of PGCCDBS have asked for a workshop (**Workshop on the utility of commercial CPUE and VMS data in assessments – WKCPUEFFORT**) on the use of VMS data, and other 'novel' e-data sources for spatially and technically refining commercial LPUE/CPUE series used as tuning fleets in stock assessment (see Appendix VIII). The workshop is scheduled for 5-7 April 2011 and will be held at ICES in Copenhagen. Clearly, there is very strong linkage between the objectives of the workshop and this project on the Development of Tools for Logbook and VMS analysis will form a central part of the guidelines that we aim to develop. One of the project partners will attend and present this report.

8.3.1 Regional database (FishFrame)

Progress towards a Regional database capable of storing output resulting from combining logbook and VMS data has been made during the current project. Early on the partners agreed to use the FishFrame system (<http://www.fishframe.org/>) developed in Denmark by DTU-Aqua. The input format required for upload to FishFrame is described in WP 1 (see also Appendices I, III and IV). So far data from a number of countries have been uploaded to FishFrame and instructions in its use are described in WP4. Examples of the sort of output that can be achieved are displayed here in Figures 8.8-11. which shows landings of plaice by Dutch TBB_DEF_100-119 (Figure 8.8) and landings of herring by Dutch (Figure 8.9), Danish (Figure 8.10), and Danish and Dutch (Figure 8.11) midwater otter trawl fleets (OTM_SPF_32_69_0_0) combined. One should note that, in FishFrame DCF level 6 fishing activity

métier definitions are used (Appendix IV) but that it would also be possible to select these using statistical criteria instead (see WP2). We have just chosen these fleets and species arbitrarily to demonstrate the functionality. It is becoming increasingly important that 'regional' estimates of fishing effort are made since it is not usually sufficient to have data from only one country. Since most countries usually only allow 'aggregated' data to be exchanged, centralised databases such as FishFrame will be instrumental in providing 'regional' data summaries of fishing activity for many countries as required during projects such as FIMPAS.

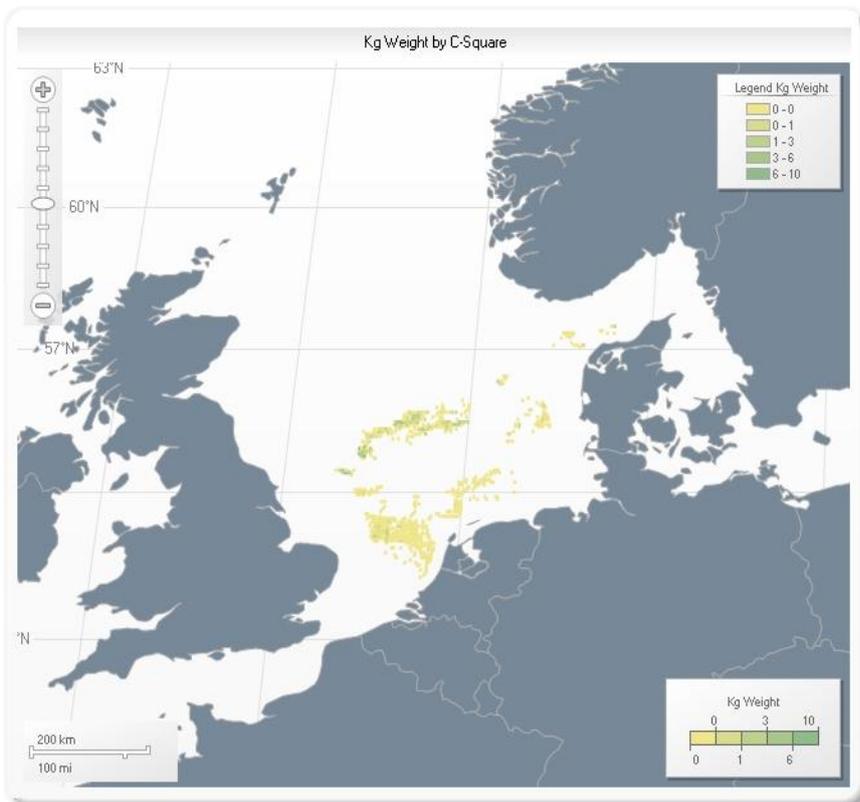


Figure 8.8. Plaiice landings in the North Sea by Dutch TBB_DEF_100-119 during 2009. Note the vmstools package was used to combine VMS and logbook data, identify fishing activity, distribute the logbook landings by VMS positions, and create output suitable for upload to FishFrame.

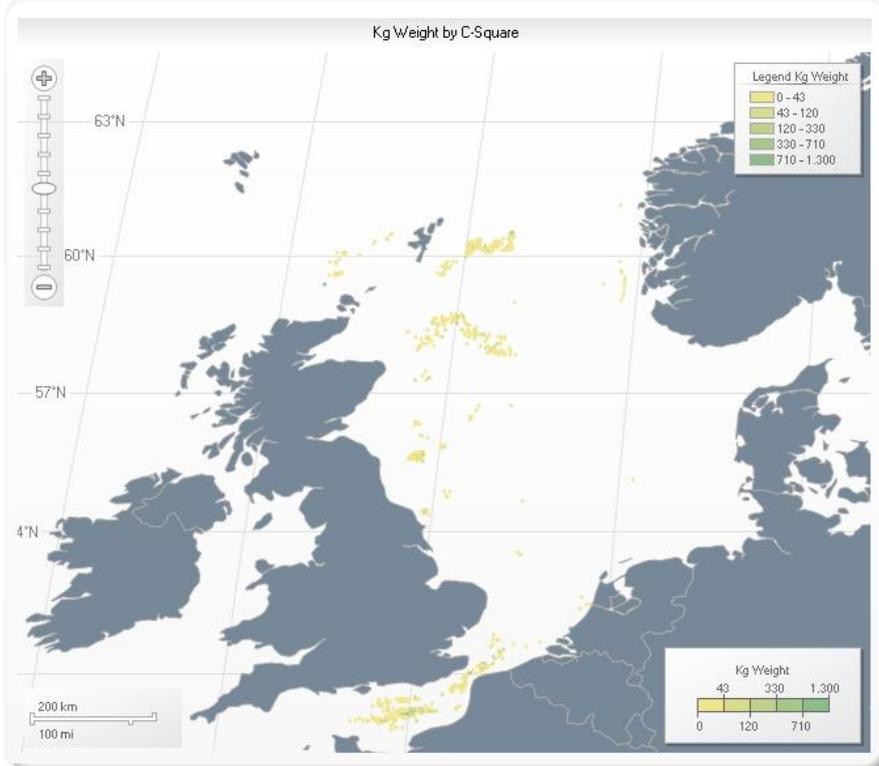


Figure 8.9. Herring landings in the North Sea by Dutch otter trawlers (OTM_SPF_32-69_0_0) during 2009 from FishFrame.

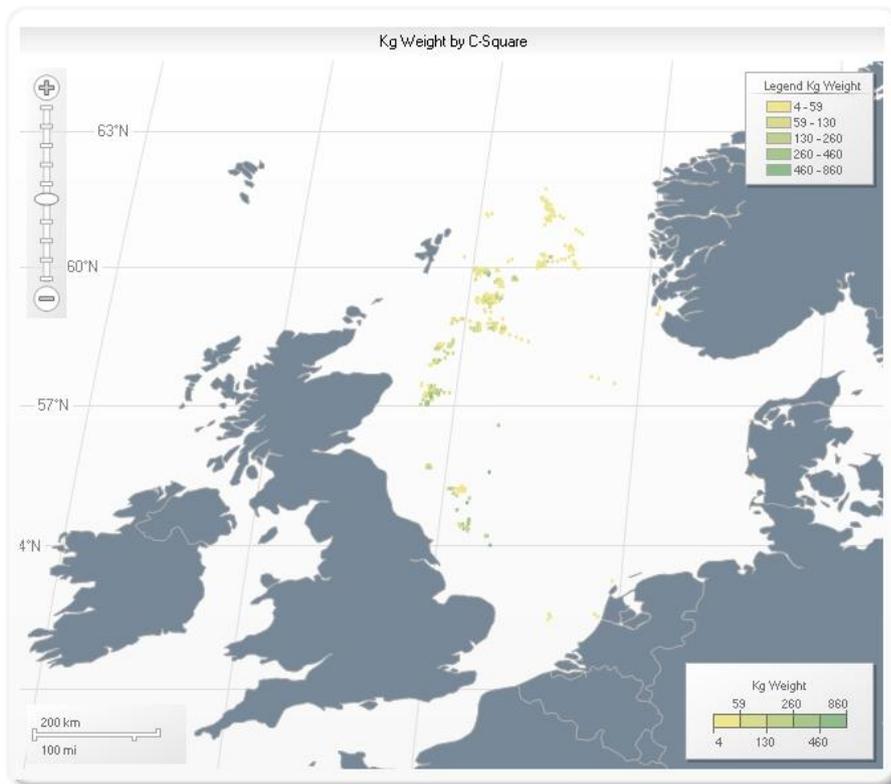


Figure 8.10. Herring landings in the North Sea by Danish otter trawlers (OTM_SPF_32-69_0_0) during 2009.

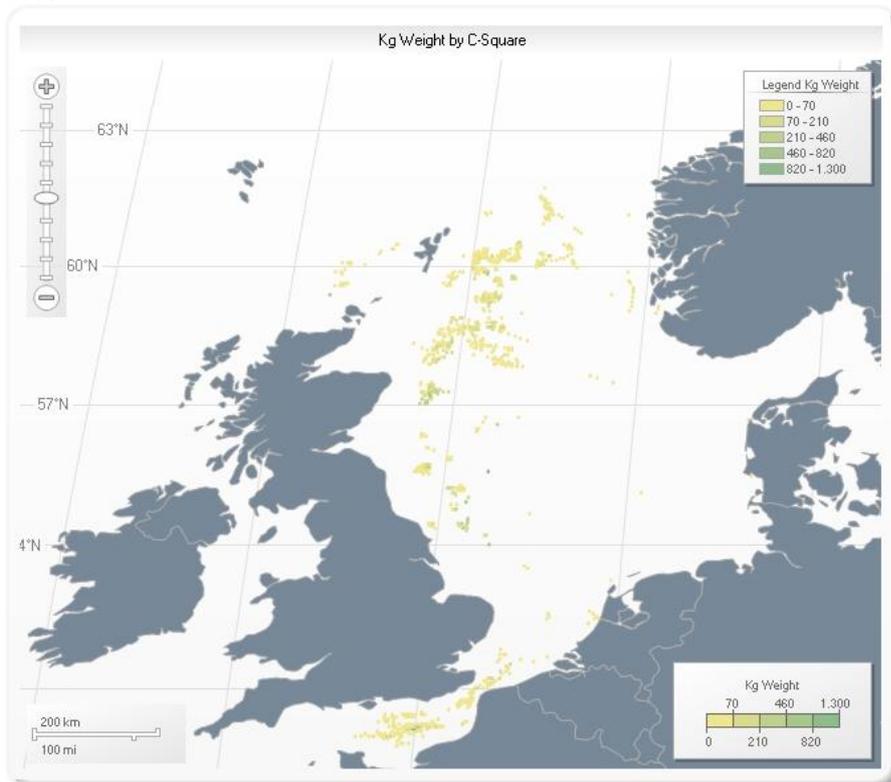


Figure 8.11. Herring landings in the North Sea by Danish and Dutch otter trawlers (OTM_SPF_32-69_0_0) during 2009.

8.3.2 FishFrame performance survey

VMS datasets are, of course, large and that can cause problems with database and website performance. In order to investigate the performance of FishFrame VMS and logbook data output 'reports', a number of experienced FishFrame users were contacted and kindly asked to create a well-defined output report (in this case a map) from their particular 'platforms'. They could use any internet browser they wanted. The test was sent to 41 persons. We received 11 answers including two persons who were not able to carry out the test due to local setting of the network. The response time of FishFrame is shown below (Figure. 8.12). Response times varied between 5 and 12 minutes. Work is under way to maximise the database performance and shorten these times.

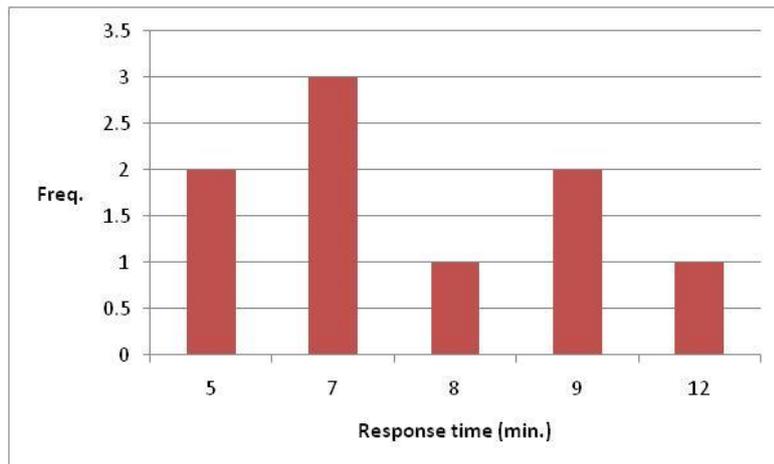


Figure 8.12. FishFrame performance according to the 11 potential users contacted.

Note: we inserted this small survey here in WP7 since it is a form of 'dissemination' and shows that many people use FishFrame and have been made aware of the **vmstools** package. Feedback has been extremely positive reflecting the general interest in VMS data and the information it can convey.

8.4 Conclusion

The synthesis of the project was based on the following objectives: (1) Integrate the tools developed in the previous work packages into a common framework (2) Apply this framework to datasets across Europe (Baltic Sea, North Sea, Bay of Biscay, East Atlantic) and show the possible output of this framework; (3) Identify gaps in knowledge that may hamper the estimation of fishing impact on the ecosystem now or in the future and (4) Make this framework including all tools and protocols available at a public website.

The project has been successful in meeting all four of these objectives and fulfilling each of the four deliverables listed in the original proposal (see Deliverables Checklist). The integration of the tools developed during the project is embodied in the **vmstools** package itself (see Appendix I). The **vmstools** package is already being widely used and discussed in the marine scientific community. It is being used to produce academic publications and also finds many more practical uses, satisfying the needs of various clients, from wind farm companies, to fishers, to those responsible for examining the impacts of 'no-take' zones.

The utility of **vmstools** cannot be questioned. Because of this we anticipate that its use among the marine scientific community will expand rapidly. For this reason the current team will continue to develop and maintain the software indefinitely. There are many improvements that we have identified for the future. Firstly we would like to incorporate more shapefiles, e.g. for depth or boundaries such as the 12

mile limit into the package which can be extremely useful for identifying fishing activity as we demonstrate for VMS positions on land and near harbours in WP3. Similarly depth contours would be a useful addition, together with shapefiles of fishing 'grounds' and protected areas.

Identifying the fishing activity of boats working static gear, for example, is still a problem and cannot be done particularly reliably. The segmented regression method described by Bastardie et al. (2010) performs best for towed gears. During the project an attempt was made to explore the utility of Bayesian methods (e.g. Vermard et al, 2010) but time was insufficient to allow thorough testing and incorporation and incorporation into the **vmstools** package.

It should also be noted that there can be dangers interpreting data in pan-European databases. For instance the distribution of logbook landings among VMS positions is a rather novel procedure. In certain cases certain species might be assigned to a certain position where it cannot be caught and erroneous impressions of fish distributions can easily be made. Ultimately given the utility of the spatial detail in VMS data this procedure, however, cannot be ignored and we believe that the potential benefits outweigh the costs.

8.5 General Conclusions

The spatio-temporal data available in VMS data provides invaluable information in the sphere of fisheries and marine spatial management. We are now able to examine fishing activity and the landings and values made by fishers in detail that would have been impossible only a few years ago. The usual 2 hour time period that elapses between the recording of each registration still causes some difficulties, especially in the area of separating true fishing activity in vessels working static gear such as crab pots or gillnets. Estimating fishing 'effort' by vessels working such gears, however, is a perennial problem in fisheries management: simply because a vessel is hauling a net does not mean that she is 'fishing' which clearly happened continuously between the time the net was shot and hauled.

The recent invitation of project partners to join the new ICES Study Group on the utility of commercial CPUE and VMS data in assessments and describe **vmstools** gives an indication of the direction in which the fisheries management community is headed; i.e. the detailed landings per unit effort data from combining VMS and logbook data will be used increasingly to inform stock assessments.

The VMS regulations were originally instigated for 'enforcement' purposes and the regulations concerning their use are complicated and open to various interpretations (see work-package 7 for more details). Traditionally such regulations caused problems with ordinary logbook data and the situation is the same in the case of VMS and it seems very unlikely that EU member states will ever allow 'raw' data for individual position registrations to be exchanged. The combination of logbook and VMS data and then subsequent aggregation over months or grid cells using common protocols (see scripts in Appendix V) as we describe here is, in our opinion, the only plausible way forward.

Herein we demonstrate how VMS and logbook analyses can be standardised using our R software package, **vmstools**. Although others have also show how this might be done ([Bastardie et al., 2010](#), [Gerritsen and Lordan, 2011](#), [Hintzen et al., 2010](#), [Lee et al., 2010](#), [Fock, 2008](#), [Pedersen et al., 2009](#), [ICES, 2010](#)), no single yet exists to do all the tasks necessary from analysing the data to visualising them. There are other examples of successful projects for formalising various analyses in fisheries science, e.g. FLR developed for management strategy evaluation, and it is not unthinkable that links might eventually be established between **FLR** (Kell et al., 2007) and **vmstools**.

It was not an original objective of the project to develop a fully documented software package. At the outset it was intended more as a kind of 'glue' linking a range of methodologies done in different software systems (e.g. Matlab and SAS). This means, however, that you might end up merging VMS and logbook in SQL, any statistics in SAS and subsequent visualisation in Matlab and mapping in ARCGIS. This would clearly be a difficult and contorted process. After much discussion at the beginning of the project, then, we decided to try and do all the work within one main package which of course makes it

far more powerful. R was chosen as the development environment for a variety of reasons: in Europe, many fisheries institutes are familiar with its operation; secondly it is freely available, and thirdly it has an ever-growing list of additional packages which can be utilized. We therefore commend **vmstools** wholeheartedly to you.

Doug Beare on behalf of the EU project No MARE/2008/10 Lot 2 Team.

9. Project achievements

9.1 Quick Deliverables checklist

We provide a 'quick deliverables checklist' below enabling the reader to see at a glance how we have dealt with each one.

Work package 1. Compilation of national data sets into a single database (DTU-AQUA)

D1.1 An agreed format for logbooks data to be used by each partner

EFLALO2 format (see Table 2.1.1).

D1.2 An agreed format for VMS data to be used by each partner

TACSAT2 format (see Table 2.2.1).

D1.3 A database structure that can hold these two data sources

FISHFRAME format (see Table 2.3).

Work package 2. Distinction of métiers from disaggregated logbook data

D2. A list of criteria to evaluate the reviewed statistical methods with respect to the project objectives.

A review of statistical methods to classify trips into métiers based on the species composition observed in the logbooks has been performed.

To evaluate these statistical methods with respect to the project objectives and to select the most appropriate ones, three criteria have been first proposed i.e. conformity with the Input data format, availability of the algorithm and relevance of the method. Further criteria have then been proposed to assess the performance of the methods selected, i.e. stability of the results, consistency over time, sensitivity to calculation method, robustness of thresholds and qualitative expert validation.

The statistical methods for the classification of fishing activities into métiers finally selected after the evaluation of some alternatives (e.g. the de-trended correspondence analysis, DCA, or the multiple correspondence analysis, MCA, and qualitative approaches) were finally the multivariate analysis method Principal Components Analysis (PCA), associated with clustering methods such as K means, PAM and CLARA algorithms, which are now all incorporated in the *vmstools* package.

D2.2 Table with the criteria to be filled for each statistical method.

The suitability of the methods selected in D2.1. together with their particular settings have then been examined in a systematic way by applying each of them to an extended logbook data set using the format defined in WP1 (i.e. EFLALO).

A comparative table of the performance of these methods has been produced to qualify each of them in terms of feasibility, computation time, stability of the results and over years, sensitivity to calculation method and relevance of the output classification given the expert knowledge.

D2.3 Description of selected multivariate statistical methods; required data and format; assumptions.

The *vmstools* package is a set of R functions and scripts providing a transparent operational tool starting from the data format defined in the WP1 that allows the classification into métiers to be done from the range of selected methods and by investigating the potential effect of particular specific settings. These tools enable the performance of sensitivity analyses of métier definition along a variety of previously used classification methods and selection of the most robust methods based on objective statistical criteria. These tools are also accompanied with several documenting files which detail their specific underlying assumptions.

D2.4 Evaluation of dataset quality in relation to the methods to be used, for each regional case study.

Good quality of logbooks data means that the logbook event is the informed level of disaggregation (as defined in the EFLALO format) and is a preliminary during the procedure to assign correctly one or several métier(s) to each fishing trip. On the tested data set, some of the logbooks of particular countries did not have this required level of disaggregation available (i.e. they only have data at the fishing trip level) which could be problematic for vessels taking part in more than one métier during a fishing trip.

Beside this, the size of the data sets (usually large when combining several countries) has proven to be the main limiting factor when applying these methods, but the CLARA algorithm has been able to overcome this potential problem.

D2.5 At regional scale, synthesis of the outcomes of the different methods applied.

The relevance of the whole procedure and the respective output of each of the investigated method have been illustrated using the example of the international bottom otter trawl (OTB) fishery from logbook data gathered from the main countries (Denmark, England, France, Scotland and the Netherlands) fishing in the North Sea region, which all together cover a large part of the main demersal fisheries in the area.

D2.6 Table defining for each métier, its definition in term of robust percentage thresholds at regional scale.

Such tables have been produced by applying a model of regression (discriminant analysis) on the output of the clustering analyses. It enables standard prediction of the métier of a new dataset (e.g., the one obtained the next year) and to link the results obtained to the DCF level 5 métier definitions.

D2.7 Protocols how to use the methodological tools developed in order to distinguish métiers and allocate trips to them.

A scientific paper is being written for describing the entire workflow dedicated to assigning a métier to each logbook event and testing and discussing the robustness of this classification applied to the regional North Sea demersal fisheries.

Work Package 3. Distinction of fishing from other activities (IFREMER)

D3.1 Review of and evaluation of existing methods that establish the main activities of a vessel trip and distinguish spatial distribution of fishing from other activities;

A review of the literature available is presented. Since VMS is such a new technology rather few papers exist on the identification of fishing activity from VMS position registrations.

D3.2 Application of the most promising methods to available regional datasets and evaluation of the results obtained.

The best methods have been incorporated into the *vmstools* package and have been used in the datasets uploaded to FishFrame

D3.3 Protocol for the application of the methods that should establish the main activities of a vessel trip and distinguish spatial distribution of fishing from other activities.

Protocols have been established starting with data 'cleaning' and identification of points on land and near harbours. Speeds are then used to determine fishing activity with simple speed thresholds for specific métiers and vessel sizes or more objectively using segmented regression.

Work package 4: Linking VMS to logbook data (DTU-AQUA)

D4.1 A list of quality flags that may be identified during evaluation

The evaluation of the consistency of the VMS data with the logbooks data has been handled by defining various degree of matching between both types of data sources, from finer to coarser, depending on whether both data have in common either (i) the trip identifier, the same area (ICES rectangle), and the catch date; (ii) only the trip identifier and the area; or (iii) only the trip identifier.

D4.2 A tool that will:

- Perform the evaluation of fishing activities and return these with attached quality flag
- Perform the merging of métier data and VMS data
- Output the merged data in FishFrame VMS/ Métier exchange format at the requested aggregation level

The *vmstools* package provides tools to merge the VMS tracks to the logbooks data from which the fishing effort and the origin of landings are mapped. The qualification of the degree of matching is automatically done during the coupling procedure by attaching a quality flag to each logbook event, and various subsequent tabulations can therefore be done to investigate poor matches, e.g. per métier/activity or per fishing area, per period, etc.

D4.3 Definition of the FishFrame VMS/ Métier exchange format and the required features in FishFrame to allow the upload and validation of this data

A FishFrame data format has been defined to handle the new available data (see Tables 1.3 and 1.4 and Appendix III). R functions within the *vmstools* package take the output from the merging procedures producing format compatible with this FishFrame format, also accounting for the aggregation of data at the requested level.

D4.4 Reporting section of FishFrame dedicated to the visualization and analysis of VMS/ Métier data

Users can request access to FishFrame from Henrik Degel (hd@aqua.dtu.dk) where there is a reporting tool (e.g. Figure A below) which can be used to make data selections, plot maps graphs etc. This feature is described with more examples in Work package 7.

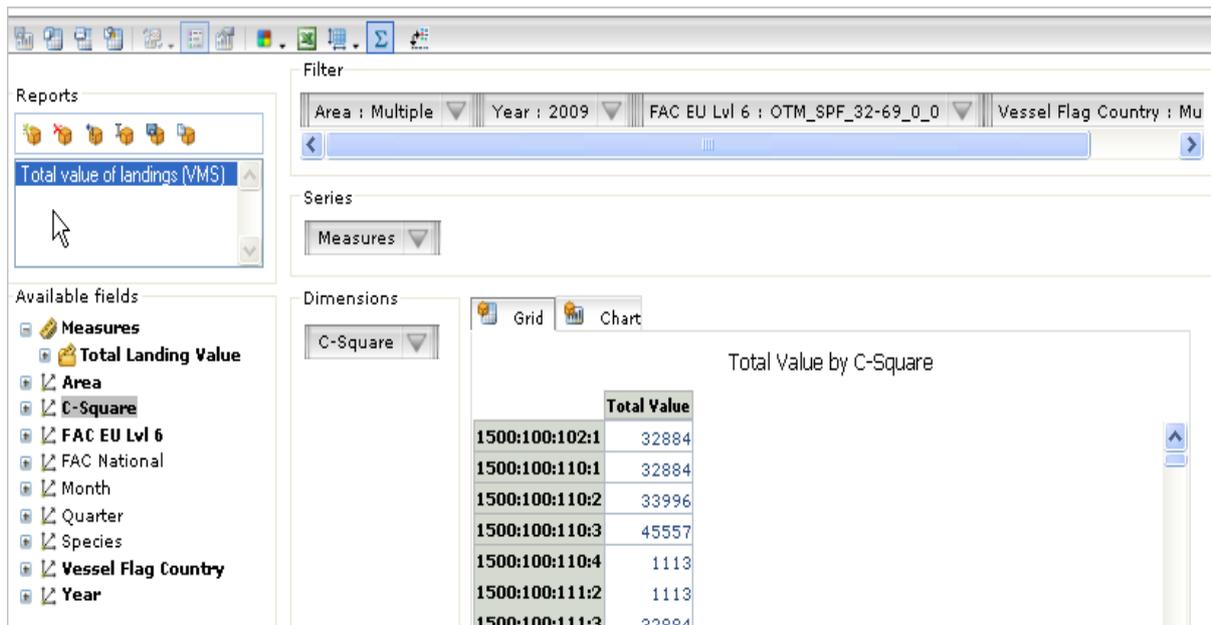


Figure A. Example of the FishFrame reporting tool interface.

Work package 5: Evaluation of the effect of recording rate

D5.1 Description of the effect of the recording rate on our perception of the spatial distribution of a fleet.

Two types of analyses have been performed to evaluate the perception of spatial distribution of a fleet. First, a dataset with much higher interval rate than regular VMS has been subsampled to mimic VMS sets with different interval rate. Based on these datasets, trawling intensity has been calculated, also varying the spatial scale. Second, based on an interpolation algorithm, fishing tracks have been reconstructed which allow us to estimate the spatial extent of fishing given a recording rate. The performance of the interpolation method has been tested against two criteria: the performance of estimating absolute distance trawled and performance of estimating spatial trawling location. These criteria have been evaluated against different recording rates.

D5.2. An algorithm that allows the best estimation of the actual vessel track from VMS position registration data.

An algorithm has been developed based on cubic Hermite spline interpolation to reconstruct trawling tracks of fishing vessels. The method performs markedly better on estimating total distance trawled and estimating spatial trawling location, given towed gears have been used. The algorithm has been published as a scientific paper.

D5.3. Advice on the preferred recording rate after an evaluation of the results of D5.1 against a set of criteria.

Based on the results obtained from the analyses described under D5.1. Also, two conclusions can be drawn. First, the perception of spatial distribution of a fleet will not change under an increase of recording rate. Second, if spatial distribution of trawling is to be estimated, the highest ping rate affordable can be advised while distance of trawling tracks can also well be estimated with a recording rate of +/- 30 minutes, taking into account the potential costs of a higher recording rate.

Work Package 6: Describing the spatial distribution of fishing activity

D6.1 Development of pressure indicators based on the spatial distribution of fishing activity

As the emphasis of this project was on the development of tools for operational use, this work package has been focused on pressure indicators which have been identified for implementation elsewhere. In particular, the EU Data collection framework defines a number of indicators of the effects of fisheries on the marine ecosystem, and these include three pressure indicators reflecting the spatial distribution of fishing activity. The three indicators summarise respectively the extent of the distribution of fishing activities, the aggregation of fishing activities, and the extent of areas not impacted by towed fishing gears.

D6.2 Tools and standardized protocols for producing these pressure indicators from VMS data

The *vmstools* package includes routines for calculating the three pressure indicators given the relevant VMS, logbook and other data. The routines allow some flexibility in how each of the indices are calculated, e.g. in terms of grid size or how the total area is calculated, and they also give the option of plotting the data on a map.

Work Package 7: Synthesis, output and dissemination

D7.1 Workshop with participants from various EU member states holding VMS and/or logbook data.

A workshop was held in October 2010 in Ijmuiden. Prior to that the *vmstools* package was described at the ICES SGVMS in Hamburg. More details are given in Work Package 7.

D7.2 Results from the application of the tools to the available data sets showing, e.g. different métiers per region and estimates of pressure indications on the spatial distribution of fishing activity.

Examples of the output from the tools (*vmstools* package) are displayed throughout this report (WPs 2-7).

D7.3 Dissemination of the tools and protocols.

The current project has generated a great deal of interest and the *vmstools* package is being used to service the requirements of a range of clients (see Work Package 7). Several peer-reviewed manuscripts, including one to describe *vmstools* itself, are in preparation. The FishFrame regional database is already being populated with combined logbook and VMS data from multiple countries. Many of the partners have given presentations on the *vmstools* package and its output and we have proposed to ICES that we give courses in the use of the *vmstools* package which is currently being considered.

D7.4. Public website containing data exchange formats, scripts and full documentation for the analyses.

Data exchange formats, scripts and full documentation for our analyses can be obtained from the following website: <http://code.google.com/p/vmstools/>.

9.2 Dissemination

Presentations

8th to 9th September 2010. Presentation (Niels Hintzen) on the use and potential of *vmstools* to the ICES Study Group on VMS data in Hamburg.

October 2010. Presentation (Francois Bastardie) of *vmstools* and output to the Chief Executive with the responsibility for designing and planning the link between Femern and Lolland.

15th November 2010. Presentation (Fabrizio Manco) on *vmstools* package to a combined meeting of the Cefas R-User group and GIS user group.

1st December 2010. Presentation (Paula Harrison) on *vmstools* package to the CoralFISH Science Meeting in Milan.

1st January 2011. Presentation (Clara Ulrich) on *vmstools* to ICES WGCHAIRS where the Strategic Initiative on Area-based Science and Management (SIASM) was discussed. [Note: In this forum fisheries scientists were under-represented compared to those from the environmental sphere] .

January 2011. Presentations (Francois Bastardie) to Danish Fishermen's Association of fishing vessel activity and landings using maps made with *vmstools*.

7th February 2011. Presentation (Niels Hintzen) to staff at IMARES on "How to analyse VMS (GPS) and logbook data using the *vmstools* package.

9.2.2 Manuscripts

Bastardie, F., Nielsen, J. R., Ulrich, C., Egekvist, J. and Degel, H. 2010. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. *Fisheries Research*, 106: 41-53.

Nicolas Deporte, Francois Bastardie, Sébastien Demanèche, Stephanie Mahévas, and Clara Ulrich-Rescan. (in prep). Comparative investigation of statistical methods for identifying fishing activities from logbook data- The international otter trawl fisheries in the North Sea.

Niels T. Hintzen^{1*}, Francois Bastardie², Doug Beare¹, Clara Ulrich-Rescan², Fabrizio Manco³ (in prep). *vmstools*: open-source software for the simultaneous processing, analysis and visualization of logbook and VMS data.

G.J. Piet, N. Hintzen and D. Beare. (in prep). Indicators for the spatial distribution of fishing effort.

Gerritsen, H.D. and Lordan, C., 2011. Integrating Vessel Monitoring Systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES J. Mar. Sci.*, 68(1): 245-252.

9.2.3 Workshops

8th -9th September 2010. Study Group on VMS data, its storage, access and tools for analysis (SGVMS) took place in Hamburg, 8-9 September 2010. Heino Fock and Vanessa Stelzenmüller chaired the meeting with 16 participants from 8 nations. Niels Hintzen and Doug Beare (IMARES) were also present. Hintzen gave a presentation on the *vmstools* package and distributed scripts which explain how to use it. See <http://www.ices.dk/reports/SSGSUE/2010/sgvms10.pdf>. Ultimately SGVMS recommended that:

- a centralised data base should be established in the ICES Data Centre in order to deliver standardised data products to Working Groups and clients.
- a working group should be established to regularly work on VMS methodologies and standards, and to provide data to the ICES Data Centre.
- a consultation process among WG chairs and ACOM and SCICOM should be undertaken by a competent authority to find out needs on VMS data in the ICES community to better determine the involvement of ICES in the medium and long term.

25th – 28th October 2010. Workshop with all project partners and other invitees. Each participant came with EFLALO and TACSAT format data from their respective countries (Denmark, The Netherlands,

France, Ireland, Scotland, England, and Poland). Participants were given instructions in how to use the **vmstools** libraries to undertake simultaneous analyses of VMS and logbook data. Finally each participant was shown how to create output for 'Regional' databases (FishFrame).

9th December 2010. Workshop to discuss **vmstools** package and its use with staff from LEI.

9.2.4 Virtual

FishFrame database. A Regional database capable of storing output resulting from combining logbook and VMS data has been adopted and further developed during the current project. FishFrame provides a database and online reporting tool that can facilitate the storage, retrieval and analysis of combined VMS and logbook data from multiple countries. The input format required for upload to FishFrame is described in WP 1 (see also Appendices I, III and IV). So far data from a number of countries have been uploaded to FishFrame and instructions in its use are described in WP4. Since most countries usually only allow 'aggregated' data to be exchanged, centralised databases such as FishFrame will be instrumental in providing 'regional' data summaries of fishing activity.

Vmstools package. This package allows the integration of (inter-)national logbook and VMS data in order to provide (regional) descriptions of fishing impact. **Vmstools** is built using the freeware package, R (<http://cran.r-project.org/>). Once the data have been imported into R in the correct format, a series of R programs or 'functions', linked by 'scripts' enable all tasks necessary to be completed in a single software environment. The tool has already been used by various [partners within the project to advice on marine spatial management, e.g. construction of wind farms and the instigation of protected areas.

10. References

- Anon (2007) Commission Staff Working Paper: Report of the Ad Hoc Meeting of independent experts on Indicators and associated data requirements to measure the impacts of fisheries on the marine ecosystem. Brussels, 25-27 June 2007, 32p
- Bastardie, F., Nielsen, J. R., Ulrich, C., Egekvist, J. and Degel, H. 2010. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. *Fisheries Research*, 106: 41-53.
- Biseau, A., and E. Gondeaux. 1988. "Apport des methodes d'ordination en typologie des flottilles." *ICES J. Mar. Sci.* 44:286-296.
- Council Regulation (EC) No 1543/2000, Report of the 4th Liaison Meeting between the Chairs of the RCMs, the chair of ICES PGCCDBS, the chair of PGMED, the ICES representative, the Chair of SGRN and the European Commission Brussels, 20-22 February 2008.
- CEC (2002) Council Regulation 2371/2002 of 20 December 2002 on the conservation and sustainable exploitation of fisheries under the Common Fisheries Policy. . OJ L 358/59 31.12.2002.
- CEC (2007) European Parliament legislative resolution of 11 December 2007 on the Council common position for adopting a directive of the European Parliament and of the Council establishing a Framework for Community Action in the field of Marine Environmental Policy. (2007) b. Marine Strategy Framework Directive 9388/2/2007 - C6-0261/2007 - 2005/0211(COD).
- Council Regulation (EC) No 949/2008 of 6 November 2008 adopting a multiannual Community programme pursuant to Council Regulation (EC) No 199/2008 establishing a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy
- Daan N (2005) An afterthought: ecosystem metrics and pressure indicators. *ICES Journal of Marine Science* 62:612-613
- de Boor C (1978) *A Practical Guide to Splines*, Vol. Springer-Verlag, Berlin
- Duplisea DE, Warr KJ, Dinmore TA, Jennings S (2001) A size-based model to predict the impacts of bottom trawling on benthic community structure. *Journal of Applied Ecology*
- Fock H. Fisheries in the context of marine spatial planning: defining principal areas for fisheries in the German EEZ. *Marine Policy* 2008;32:728-739
- EC (2002) Council Regulation (EC) No. 2371/2002 of 20 December 2002 on the conservation and sustainable exploitation of fisheries resources under the Common Fisheries Policy. *Official Journal of the European Communities* L358:59-80
- EC (2008) Commission Decision of 6 November 2008 adopting a multiannual Community programme pursuant to Council Regulation (EC) No 199/2008 establishing a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy (2008/949/EC). *Official Journal L 346* , 23/12/2008 P 0037 - 0088
- Everitt, B.S., Landau, S., Leese, M., 2001. *Cluster Analysis*. Arnold, London, UK.
- Garcia SM, Staples DJ, Chesson J (2000) The FAO guidelines for the development and use of indicators of sustainable development of marine capture fisheries and an Australian example of their application. *Ocean and Coastal Management* 43:537-556
- Gerritsen HD, and Lordan, C., (2011) Integrating Vessel Monitoring Systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES J Mar Sci*, 68(1): 245-252.
- Hartigan, J. A. (1975). *Clustering Algorithms*. New York: Wiley.
- Hartigan, J.A. and Wong, M.A. (1979). A K-means clustering algorithm. *Applied Statistics* 28, 100-108.
- He, X., K.A. Bigelow, and C.H. Boggs. 1997. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. *Fisheries Research* 31:147-158.
- Hermite C (1863) Remarque sur le développement de Comptes Rendus de l'Académie des Sciences. 57:613-618

- Hiddink JG, Jennings S, Kaiser MJ (2006a) Indicators of the ecological impact of bottom-trawl disturbance on seabed communities. *Ecosystems* 9:1190-1199
- Hiddink JG, Jennings S, Kaiser MJ, Queiros AM, Duplisea DE, Piet GJ (2006b) Cumulative impacts of seabed trawl disturbance on benthic biomass, production, and species richness in different habitats. *Canadian Journal Of Fisheries And Aquatic Sciences* 63:721-736
- Hintzen NT, Piet GJ, Brunel T (2010) Improved estimation of trawling tracks using cubic Hermite spline interpolation of position registration data. *Fisheries Research* 101:108-115
- Holley, J.-F., and Marchal P. 2004. Fishing strategy development under changing conditions: examples from the French offshore fleet fishing in the North Atlantic. *ICES J. Mar. Sci.* 61: 1410-1431.
- Hutton, Trevor, Simon Mardle, Sean Pascoe, and Robin A. Clark. 2004. "Modelling fishing location choice within mixed fisheries: English North Sea beam trawlers in 2000 and 2001." *ICES J. Mar. Sci.* 61:1443-1452.
- ICES 2003. Report of the Study Group on the Development of Fishery-based Forecasts Boulogne, France 18-21 February 2003, CM 2003/ACFM:08, Ref. D, 37 p. <http://www.ices.dk/reports/ACFM/2003/SGDFF/SGDFF03.pdf>
- Jennings S (2005) Indicators to support an ecosystem approach to fisheries. *Fish and Fisheries* 6:212-232
- Kaufman, L. and Rousseeuw, P.J. (1990). *Finding Groups in Data: An introduction to Cluster Analysis.* Wiley, New York.
- Laurec, A., Biseau, A. and Charuau, A. 1991. .Modelling technical interactions.. *ICES mar. Sci. Symp.*193.
- Lee J, South AB, Jennings S (2010) Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES Journal of Marine Science* 67:1260-1271
- Lewy, P., and Vinther, M. 1994. Identification of Danish North Sea trawl fisheries. – *ICES Journal of Marine Science* 51 , 263-272.
- Link JS (2002) What does ecosystem-based fisheries management mean? *Fisheries* 27:18-21
- Link JS, Brodziak JKT, Edwards SF, Overholtz WJ, Mountain D, Jossi JW, Smith TD, Fogarty MJ (2002) Marine ecosystem assessment in a fisheries management context. *Canadian Journal of Fisheries and Aquatic Science* 59:1429-1440
- Marchal, P. 2008. A comparative analysis of métiers and catch profiles for some French demersal and pelagic fleets. - *ICES Journal of Marine Science*, 65: 674-686
- Mercer, D.P., Linacre College (2003). Clustering large datasets.
- Mills C. M., Townsend S. E., Jennings S., Eastwood P. D., Houghton C. A. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. *ICES Journal of Marine Science* 2007;64:248-255.
- Pelletier D and Ferraris J, 2000. A multivariate approach for defining fishing tactics from commercial catch and effort data *Can. J. Fish. Aquat. Sci.* 57(1): 51-65
- Piet GJ, Quirijns FJ, Robinson L, Greenstreet SPR (2007) Potential pressure indicators for fishing, and their data requirements. *ICES J Mar Sci* 64:110-121
- Punzón, A., Hernández, C., Abad, E., Castro, J., Pérez, N., and Trujillo, V. 2010. Spanish otter trawl fisheries in the Cantabrian Sea. – *ICES Journal of Marine Science* 67: 000-000
- Rice JC (2000) Evaluating fishery impacts using metrics of community structure. *ICES Journal of Marine Science* 57:682-688
- Rijnsdorp AD, Buys AM, Storbeck F, Visser EG (1998) Micro-scale distribution of beam trawl effort in the southern North Sea between 1993 and 1996 in relation to the trawling frequency of the sea bed and the impact on benthic organisms. *ICES journal of marine science* 55:403
- Rochet MJ, Trenkel VM (2003) Which community indicators can measure the impact of fishing? A review and proposals. *Canadian Journal of Fisheries and Aquatic Sciences* 60:86-99
- Rogers, J. B, and E. K Pikitch. 1992. "Numerical definition of groundfish assemblages caught off the coasts of Oregon and Washington using commercial fishing strategies." *Canadian Journal of Fisheries and Aquatic Sciences* 49:2648-2656.
- Ulrich, Clara, and Bo Solgaard Andersen. 2004. "Dynamics of fisheries, and the flexibility of vessel activity in Denmark between 1989 and 2001." *ICES J. Mar. Sci.* 61:308-322.

- Ulrich, C., Gascuel, D., Dunn, M., Le Gallic, B. and Dintheer, C. (2001). Estimation of technical interactions due to the competition for resource in a mixed-species fishery, and the typology of fleets and métiers in the English Channel. *Aquatic Living Resources*, 14, 267-281.
- Vermard Y., Rivot E., Mahevas S., Marchal P., Gascuel D. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. *Ecological Modelling [Ecol. Model.]*. Vol. 221, no. 15, pp. 1757-1769. 24 Jul 2010.
- Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 58: 236-244.

11. Quality Assurance

IMARES utilises an ISO 9001:2008 certified quality management system (certificate number: 57846-2009-AQ-NLD-RvA). This certificate is valid until 15 December 2012. The organisation has been certified since 27 February 2001. The certification was issued by DNV Certification B.V. Furthermore, the chemical laboratory of the Environmental Division has NEN-AND-ISO/IEC 17025:2005 accreditation for test laboratories with number L097. This accreditation is valid until 27 March 2013 and was first issued on 27 March 1997. Accreditation was granted by the Council for Accreditation.

Justification

Rapport C019/11

Project Number: 430.1100.701

The scientific quality of this report has been peer reviewed by the a colleague scientist and the head of the department of IMARES.

Approved: Dr. T.H. van der Hammen
Researcher

Signature:

Date: 28 Februari 2011

Approved: Dr. ir. T.P. Bult
Head of Fisheries Department

Signature:

Date: 28 februari 2011

Appendices: