[Catégorie]

Report of acoustic processing routines & quality checking methods

Deliverable Lead:	CLS
Reference:	MESOPP-18-0003
Dissemination Level:	public
Issue:	1. 1
Date:	2018,Jun.01

Chronology Issues:

Issue:	Date:	Reason for change:	Author
1.0		First Version of the document (D3.1)	S. Fielding
1.1	1/06/2018	Inputs from UPMC and review by CLS	A. Conchon; B. Calmettes; P. Lehodey

Distribution:

Company	Means of distribution	Names
CLS	Notification	

List of tables and figures

List of figures:

List of Contents

Overview

The underlying concept of MESOPP is the creation of a collaborative network and associated e-infrastructure (marine ecosystem information system) between European and Australian research teams/institutes sharing similar interests in the Southern Ocean and Antarctica, its marine ecosystem functioning and the rapid changes occurring with the climate warming and the exploitation of marine resources.

In the past 30 years, facing global knowledge issues, lacking data, addressing huge modelling challenges, we observed the successful world organisation of meteorology. These past 15 years, Europe has kick started and demonstrated similar successful structuring of the operational oceanography fostered by the Copernicus initiative (http://marine.copernicus.eu/), today worldwide used and recognised, fully anticipated and integrated in GOOS (Global Ocean Observing System), IOSS (Integration ocean observation system), SOOS (Southern Ocean Observing System; see Rintoul et al 2011 and Meredith et al 2013), GODAE (the international global ocean data assimilation experiment), and IMBER (integrated marine biogeochemistry and ecosystem research).

A major R&D strategic challenge is to connect the marine ecosystem community across the fields of meteorology, climate, oceanography and biology. Lack of data, development of accurate high-end models, global coverage and need for exchange are issues that need to be overcome.

The objective of the MESOPP project is to meet this challenge and is threefold:

1. Make an inventory of science challenges, stakes and existing policies and develop tools to federate and structure the community;

2. Start to organise the related marine ecosystem community between EU and Australia through two implementation actions:
   - the specification and prototyping of an international e-infrastructure for marine ecosystems data
   - the development of best practices, R&D governance in relation with existing policy instruments

3. Propose a R&D roadmap to support a large international cooperation on marine ecosystems based on a e-infrastructure with additional countries such as USA, New Zealand, Canada (in the Frame of the Galway statement), Brazil and all active countries already involved in large organisations such as IMBER, CCAMLR or IMOS.

While MESOPP will focus on the enhancement of collaborations by eliminating various obstacles in establishing a common methodology and a connected network of databases of acoustic data for the estimation of micronekton biomass and validation of models, it will also contribute to a better predictive understanding of the SO based on furthering the knowledge base on key functional groups of micronekton and processes which determine ecosystem dynamics from physics to large oceanic predators. This first project and associated implementation (science network and specification of an infrastructure) should constitute the nucleus of a larger international program of acoustic monitoring and micronekton modelling to be integrated in the general framework of ocean observation following a roadmap that will be prepared during the project.

1. Introduction

The mesopelagic (200-1000 m depth), is one of the most understudied regions in the world oceans (St John et al. 2016). Micronekton (~1 to 20 cm in length, Kloser et al. 2009) are an ecologically important component of the mesopelagic community, having potentially large biomasses (Irigoien et al. 2014), high nutritional value (Lea et al. 2002), transferring carbon from the surface to depth (Anderson et al. 2018), and of commercial interest (Gjøsæter and Kawaguchi 1980; St John et al. 2016).

Notoriously hard to sample, due to poor sampling efficiency of nets, observations within the mesopelagic zone are frequently made using active acoustics (Simmonds and MacLennan 2005). Whereby, echosounders produce a pulse of sound and receive echoes backscattered from organisms, objects and discontinuities in the water. Measurement of the time delay of the received acoustic signal and quantification of the intensity of the returned sound reveals information about the source of the scattering and where it is in the water column (Benoit-Bird and Lawson 2016). Integrated into marine vessels, echosounders offer the ability to make measurements spanning high and wide spatial and temporal scales.

Acoustic methods are widely using in fisheries research for pelagic fish estimation and ecosystem-based management (Bertrand et al. 2003; Simmonds and MacLennan 2005). Dedicated acoustic survey programmes to count, map and predict fishing conditions commenced in the 1970s (Fernandes et al. 2002), and have expanded now to multi-national surveys covering sea and basin scales such as the International Blue Whiting Spawning Stock Survey (ICES 2018; WGIPS 2017) and the CCAMLR synoptic survey for Antarctic krill (Hewitt et al. 2004). In addition, as well as fisheries research vessels, many oceanographic research vessels and fishing vessels are equipped with hull mounted echosounders, operating at a variety of frequencies (e.g., Erreur ! Source du renvoi introuvable.). Acoustic data from these and other vessels have been collected for targeted reasons (ecosystem surveys, examples) or opportunistically (as part of transit routes, Kloser et al. 2009; Behagle et al. 2016; Escobar-Flores et al. 2018). As a result acoustic data exist in vast quantities, with extensive geographical and temporal coverage, and could be considered as “big” data (Colosi & Worcester, 2013) within environmental sciences.

Modern acoustic data are stored digitally and collected data are archived in data centres (e.g. NOAA National Centers for Environmental Information (https://www.ngdc.noaa.gov/mgg/wcd/), NERC data centres (http://www.datacentres.nerc.ac.uk) and Integrated Marine Observing System (www.imos.au). Raw acoustic data are typically stored in a proprietary format that requires specialized acoustic processing software (e.g. Echoview (www.echoview.com), LSSS (https://www.marec.no) or MOVIES 3D (Trenkel et al. 2009) or a knowledge of the file format and a scientific programming language. As a result, both IMOS and NOAA identified that enabling open-access to quality-checked, calibrated acoustic data would allow greater exploitation by non-acousticians (Kloser et al. 2009; Wall et al. 2016). Stored with a metadata convention to ensure proper documentation of how, when, why and where the data were collected, ensures consistency across datasets (ICES 2014).

In order to convert raw acoustic data to a quality-checked, calibrated acoustic data, a number of steps are required (**Figure 1**). The quantitative use of data from more than one sensor requires that the acoustic instrument is calibrated to allow comparison. This involves characterisation of measurement accuracy and precision, and best practise is a sphere calibration (Foote et al. 1987) that measures the overall performance of an echosounder using reflections from a solid sphere of known backscattering strength (σ_{bs} (m^2)) (Demer et al. 2015).

Table 1 - Vessels contributing acoustic data to the MESOPP project (www.mesopp.eu). Note not all frequencies currently available as calibrated post-processed Sv.

Vessel Name	Operator	Instrument/Transducer	Frequencies (kHz)	Ocean Sector
FV Janas	Sealord, NZ	ES60/ ES38B	38	Pacific
RV Kaharoa	National Institute of Water and Atmospheric Research, NZ	ES60/ ES38B	38	Pacific
FV Antarctic Discovery	Australian Longline Pty Ltd, AU	ES60/ ES18, ES38-7	18, 38	Pacific
RV Aurora Australis	Australian Antarctic Division, AU	EK60/ ES38B	38	Pacific
RV Marion Dufresne	CMA-CGM, France (until 17th of May 2017), LDA, France (then)	EK80/ES120-7C	18, 38, 70, 120, 200	Indian
RV Marion Dufresne	CMA-CGM, France	EK60, ES120-7	38, 120	Indian
RRS James Clark Ross	British Antarctic Survey, UK	EK60/ES38, ES120-7, ES200-7	38, 120, 200	Atlantic

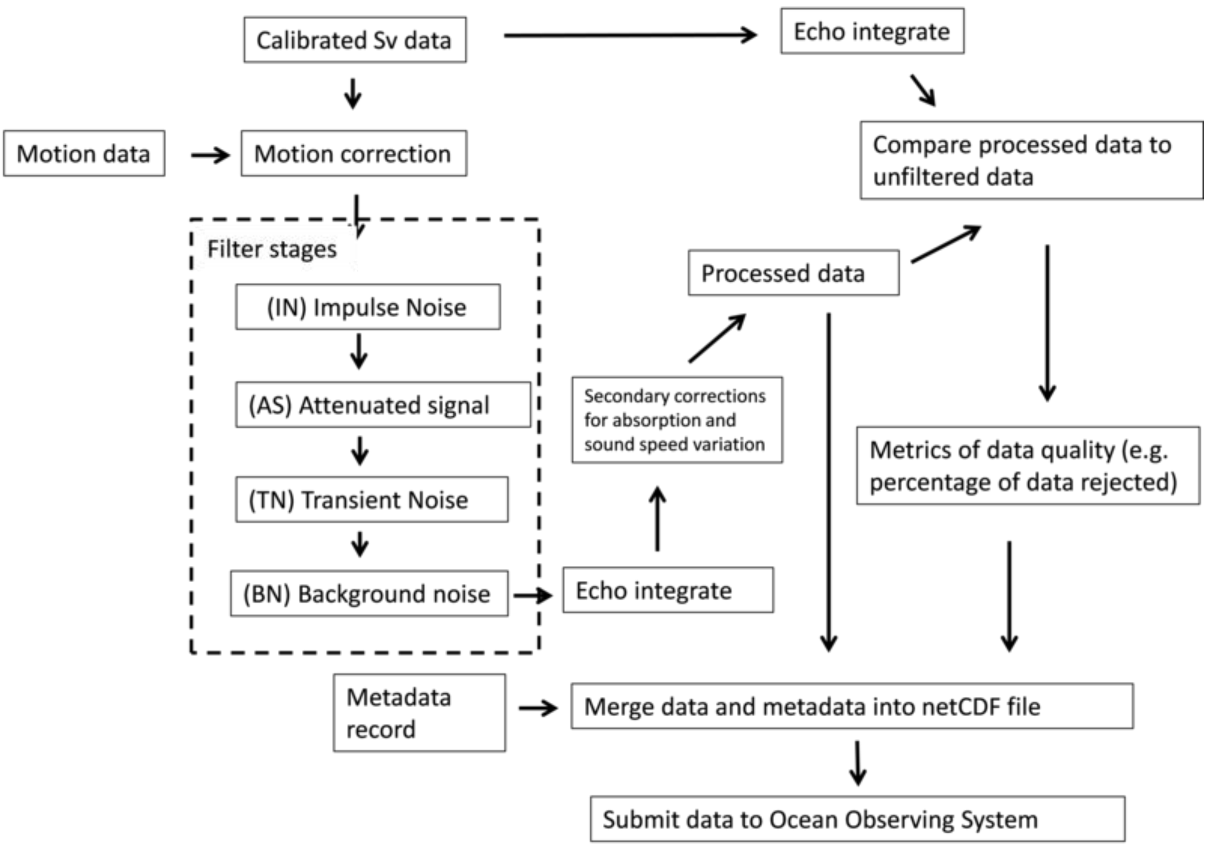

Figure 1 - Figure showing the data flow schema for production of mean volume backscattering strength (S_v, dB re 1 m^{-1}) data (Ryan et al., 2015).

Acoustic data from different vessels can vary significantly in quality, due to noise or signal attenuation. Noise is defined as unwanted contributions to the signal from mechanical, electrical or biological sources that positively bias echo integration results. Attenuation is the blocking of the acoustic signal by air bubbles or transducer motion that cause negative bias in echo integration results (Ryan et al. 2015, references there-in). Ideally mitigation measures are employed to reduce these sources of noise and attenuation (e.g. optimising hull design, electrical shielding and drop-keels, Simmonds and MacLennan 2005; Ona 1991), alternatively post-processing techniques are used to improve data quality. Ryan et al. (2015) identified four types of degradation to echosounder data:

(i) "spike" or impulsive noise (IN), where the duration of IN is less than one ping (Vaseghi 2009) – frequently caused by the transmit pulse from another unsynchronised echosounder

(ii) Transient noise (TN), is broad-spectrum high-energy sounds – typically generated in bad weather from waves colliding with the ship's hull

(iii) Background noise (BN), signal present at the receiver output in the absence of any transmission (Simmonds and MacLennan 2005) – caused by for example propeller noise (Mitson and Knudsen 2003)

(iv) Attenuated signal (AS), is attenuation of the acoustic signal – typically generated by air bubbles under the keel of a vessel or by vessel motion.

Several post-processing methods have been developed to identify and/or remove these different types of degradation. BN can be estimated by monitoring passive acoustic data during a survey (Nunnallee 1990) or calculated from the active data in post-processing (Watkins and Brierley 1996; Korneliussen 2000; De Robertis and Higginbottom 2007). Likewise, a number of filters exist to account for IN (Anderson et al. 2005; Fielding et al. 2014, Wang et al. 2016), AS (Dalen and Lovik et al. 1981; Cox et al. 2006; Honkalehto et al. 2011; Ryan et al. 2015) and TN (Ryan et al. 2015). Of note is that many of these filters still require input of user-defined parameters (Ryan et al. 2015) that could introduce variability between users as well as operating systems used to undertake the processing.

1.1. This report

This report describes the methods used to process acoustic backscatter data for the reference datasets available from the South Atlantic, South Indian and South Pacific Ocean sectors of the Southern Ocean. These data have been made available through the MESOPP Central Information System (CIS) at www.mesopp.eu/data/catalogue/ and users need to register on the MESOPP website (http://www.mesopp.eu/data/registration/) prior to downloading the data. British Antarctic Survey (BAS), Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) and French institutes CNRS/IRD/IFREMER provided the reference data, post-processed using a combination of in-house tools and commercially available software. Supplementary information on data quality is calculated during the post-processing and is discussed here. This report also contains a preliminary comparison of different institute's processing techniques and analysts interpretations to investigate the impact of post-processing techniques and agreement of data products. We highlight the need for ongoing data quality and assurance meetings at international fora to ensure best practice is being done (e.g. International Council for Exploration of the Seas Working Group on Fisheries Acoustics Science and Technology (ICES WGFAST) and Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) meetings).

2. Processing methods

The MESOPP reference datasets comprise of 67 transects (hereafter transits) of calibrated and post-processed 38 kHz volume backscattering strength (S_v, dB re 1 m^{-1}) collected from 6 vessels (both research and fishing). The reference datasets also contain calibrated and post-processed 18 kHz S_v from 3 transits, 70 kHz S_v from 2 transits, 120 kHz S_v from 4 transits and 200 kHz S_v from 2 transits (**Figure 2**Figure 2). The reference datasets come from three sectors of the Southern Ocean (South Atlantic, South Indian and South Pacific). In each case raw acoustic data (S_v (dB re 1m^{-1})) were collected using a Simrad EK60 or Simrad ES60 echosounder in Simrad proprietary format **.raw* files (http://www.simrad.net/ek60_ref_english/default.htm).

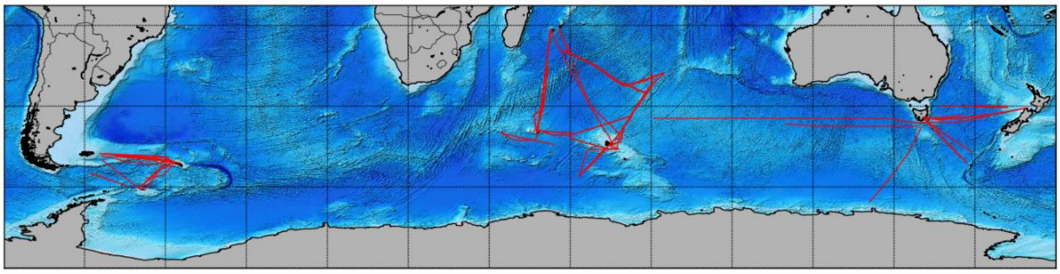

Figure 2 - MESOPP transits of calibrated and post-process acoustic backscatter

Terms and definitions used within the text below are given in Erreur ! Source du renvoi introuvable.**,** and are consistent with Ryan et al. (2015) and MacLennan et al. (2002). Data post-processing was undertaken using a combination of in-house tools developed by CSIRO, BAS, and/or IRD and used to assist with data management, identifying subsets of data and for post-processing. These utilised software packages including Java, Matlab, Movies 3D and Echoview.

Table 2 - Terms and definitions used within this report

Term	Description
S_v	Volume backscattering strength (dB re 1m^{-1})
i	Vertical sample number
j	Ping number
n	Number of pings
m	Specific vertical range (m)
$\overline{S_{v_{up}}}$	Mean volume backscattering strength obtained through echointegration of calibrated but un-post-processed S_v
$\overline{S_{v_p}}$	Mean volume backscattering strength obtained through echointegration of calibrated and post-processed S_v A cell of $\overline{S_{v_p}}$ data will span an interval based on either distance travelled, elapsed time or number of pings
$\widetilde{S_{v_p}}$	Median of S_v typically calculated within a vertical range denoted by an upper (R_1) and lower (R_2) depth limit
$\widetilde{S_{v_n}}$	Median of $\widetilde{S_{v_p}}$ for a block of n pings
δ	Threshold value
TVG	Time Varied Gain

The following post-processing steps were applied:

1) Identify on-transit data
2) Apply calibration offsets
3) Remove unwanted data (semi-autonomous), such as near-surface (including ringing, non-linearity and bubbles), data deeper than the seabed or 1000 m (whichever shallower), non-transit periods (periods where ship's speed was less than 4 knots) and aliased seabed (false bottom).
4) Apply semi-automated filters to identify noise and reject bad data
5) Grid data and apply correction for sound speed and absorption
6) Output data in NetCDF format for mean echointegrated volume backscattering strength (S_v, dB re m^{-1}) and quality control metrics.

2.1. On-transit data

On all vessels the echosounder was operated continuously, and therefore the raw data includes periods when the vessel undertakes long transits as well as periods when the vessel was stationary. Transits were selected because of data quality, long linear range, optimal vessel speed and water depth range, with criteria varying between ocean sectors according to **Erreur ! Source du renvoi introuvable. Erreur ! Source du renvoi introuvable.**.

For each identified transit, data management procedures followed those established by CSIRO, where the temporal and geographic extent of each *.raw* file was captured in a text format file (*inf* files) created during a data registration process and used to identify *.raw* files that belonged to an individual transit and add additional metadata such as vessel name, frequency, calibration information and echosounder serial number (Ryan 2011).

2.2. Apply calibration offsets

Each vessels echosounder was calibrated according to procedures documented in the ICES Cooperative Research Report 144 (Foote et al. 1987) and recent update CRR 326 (Demer et al. 2015). Measurements of target strength from a reference sphere were either made on-axis or throughout the transducer beam. Either Echoview software was used to calculate transducer on-axis gain and transducer directivity was taken from factory tests, or the ER60 software was used to estimate gain, transducer directivity adjusted to the local sound speed, and the difference in energy between the nominal and actual received pulse (Demer et al. 2015). The data from the ES60 and ES70 systems have a systematic triangle wave error sequence, which can cause biases of up to +12% (Ryan and Kloser, 2004). The triangle wave error was removed from calibration data and calibration offsets were computed using the corrected data. This error will average to zero over large datasets, and thus, it was not necessary to remove it from the identified transits. Calibration parameters from the nearest (in time) calibration were applied to the dataset during post-processing.

2.3. Removal of unwanted data

When the seabed was present and shallower than 1000 m an automated bottom detection algorithm was used to identify the seabed (e.g. best bottom candidate algorithm, https://support.echoview.com/WebHelp/Echoview.htm). Data below at least 10 m above the seabed and below 1000 m were excluded from further post-processing. In addition, data closer than 10 m to the echosounder were also excluded. In addition, regions of "false bottom" (seabed reverberations from preceding ping transmissions overlapping the current ping reception) were manually identified and removed.

2.4. Motion correction

Transducer motion reduces the received signal (Stanton, 1982) and increases inter-ping variability. When vessel-motion data were available at a suitable sampling rate (≥5 Hz), transducer motion effects were corrected using the Dunford (2005) filter, before application of semi-autonomous filters described below.

2.5. Semi-autonomous noise and bad data removal

Background noise was estimated and removed from the acoustic data. In addition, a combination of four semi-autonomous filters were used to identify bad data and remove it: a transient noise filter, an impulse noise filter, an attenuated signal filter and a signal-to-noise ratio filter. Ocean sector specific settings are detailed in table 3.

2.5.1. Background noise correction

Background noise was estimated and removed from the acoustic data by either collecting passive acoustic data manually (Nunnallee 1990) or following the automated method outlined by De Robertis and Higginbottom (2007). A key requirement of the method outlined by De Robertis and Higginbottom (2007) is that at some point in the measured cycle of a ping, the measurement is dominated by contributions from background noise. Acoustic data from both the water column and below bottom can be used to maximise the probability that the echosounder measurement was dominated by noise, so this procedure was undertaken before data below the seabed was removed, or in deep water. Transient noise, typically observed as high backscatter at range, also needs to be removed prior to background noise estimation as it causes overestimation of noise.

2.5.2. Impulse noise filter

The IN filter as described by Ryan et al. (2015) was used to remove impulse noise on vertically resampled data, where data were identified as bad data when the difference between a ping and its neighbours was greater than a threshold (δ) and if the value was greater than -80dB. Where:

$$S_{vij} - S_{vi(j-1)} > \delta \ and \ S_{vij} - S_{vi(j+1)} > \delta \ and \ S_{vij} > -80 \quad (1)$$

Rejected samples were set to "no data".

2.5.3. Attenuated signal filter

The AS filter as described by Ryan et al. (2015) was used to detect and remove pings that contained attenuated signal. Whereby each ping was compared with the median of n number of pings within a reference layer of 100 m (identified manually) and rejected if it was a threshold value of δ lower. Where:

$$\widetilde{S_{v_p}} - \widetilde{S_{v_n}} < \delta \quad (2)$$

Rejected samples were set to "no data".

2.5.4. Transient noise filter

The TN filter as described by Ryan et al. (2015) was used to detect and remove pings that contained transient noise. The TN filter identified samples which exceed the median values in a surrounding region of n pings at a vertical resolution of m or within a reference layer and rejected if they exceed a threshold value of δ. Where:

$$S_{v_{ij}} - \widetilde{S_{v_{nm}}} > \delta \qquad (3)$$

Rejected samples were set to “no data”.

2.5.5. Signal-to-noise filter

The signal-to-noise (SNR) ratio was calculated for each sample and used to objectively identify data that contained sufficient signal to warrant further analysis, following the method of De Robertis and Higgenbottom (2007). A minimum SNR threshold of 6 dB was set, and all samples that fell below this were set to “-999 dB”.

2.6. Grid data and speed of sound and sound absorption correction

After all filtering, data were gridded into 1000 m horizontal distance and 10 m vertical cells ($\overline{S_{v_p}}$). Data were either collected using sound speed and absorption values estimated from in situ environmental data, or were collected at nominal values corrected for in post-processing. In the latter case, sound speed (Mackenzie 1981) and absorption (Francois and Garrison 1982) were estimated for each cell based on temperature and salinity climatology (CSIRO Atlas of Regional Seas (CARS)), and used to correct S_v and range.

2.7. Data output and quality control metrics

Project, vessel and calibration metadata, mean volume backscattering strength obtained through echointegration of calibrated but un-post-processed S_v, mean volume backscattering strength obtained through echointegration of calibrated and post-processed S_v, with relevant CARS environmental data and metrics of data quality were stored in MESOPP NetCDF files (Appendix 1). Specific metrics of data quality include for each sector are given in Table 3 and could include: an estimate of background noise, the percentage of data retained and a range detection limit, depending on method.

2.7.1. Background noise

De Robertis and Higginbottom (2007) was used to estimate background noise, these values are provided.

2.7.2. Percentage of data retained

This was calculated as the percentage of original data retained per 1000 m horizontal and 10 m vertical grid cell before binning, derived from the ratio of the number of samples in the $\overline{S_{v_p}}$ cell to the number of samples in the $\overline{S_{v_{up}}}$ cell (**Figure 3**). Cells of $\overline{S_{v_p}}$ where the percentage of rejected data was greater than 50% may automatically be marked as no-data (-999) values. The total amount of data retained, after each filter has been applied, may also be provided as a single number for the whole transit.

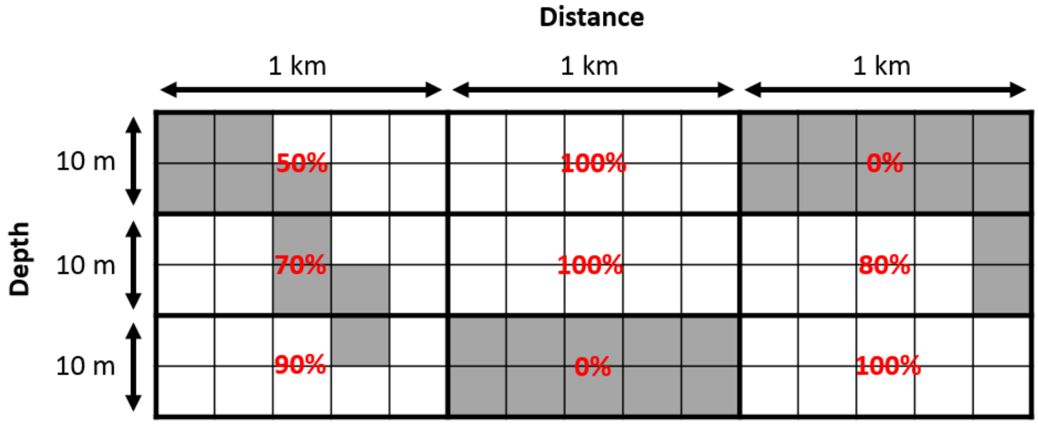

Figure 3 - The percentage of original data retained in a grid cell after a filter (e.g. IN filter). White cells represent retained good samples, grey represent bad data discarded using filters (and set to "no data")

2.7.3. Range Detection Limit (RDL)

In order to identify the range of usable data, an RDL was calculated. The RDL is an isoline drawn from the background noise (plus TVG) at a -80 dB threshold and indicates the depth below which noise dominates and targets of -80 dB or lower cannot be detected over background noise (**Figure 4**).

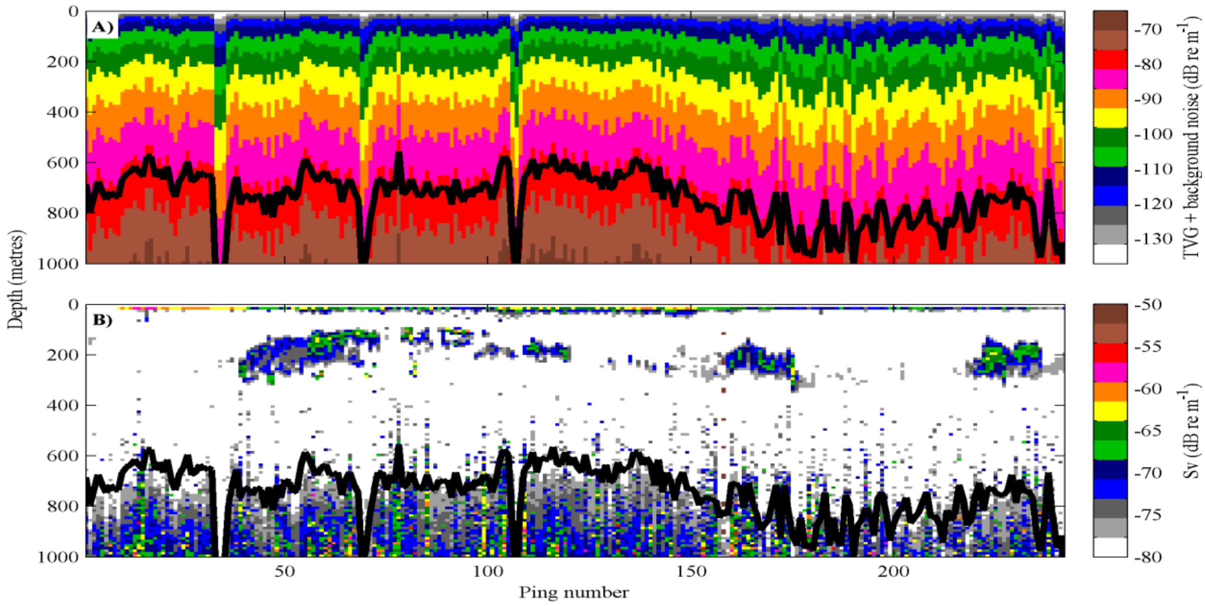

Figure 4 - A) Background noise + TVG calculated according to De Robertis and Higginbottom (2007), the black line represents the Range Detection Limit (RDL) of a -80 dB target. B) the calibrated but unprocessed gridded data, with RDL demonstrating usable data range overlaid as a black line.

3. Comparison of post-processing techniques

In order to use large datasets of acoustic data, collected from multiple platforms and post-processed using more than one technique, an understanding of how different processing methods and decisions influence data and data quality is required. To highlight this issue three transits from three different vessels were compared using the processing methods applied to the Atlantic (BAS-SONA) and Pacific (CSIRO-IMOS) sector acoustic data. The two processing methods generated estimates of water column mean S_v that were similar and follow similar patterns with time along the transits from each of the three vessels. From each transit total water column Nautical Area Scattering Coefficient (NASC, m^2 nmi^{-2}) were highly correlated (Erreur ! Source du renvoi introuvable.**-7**) and close to one-to-one for the RRS James Clark Ross transect. The BAS-SONA processing method retained less backscatter for the RV Jana transect (where in general water column NASC was higher) and more backscatter for the RV Tangaroa transect (where in general water column NASC was lower). Bland-Altman plots (**Figures 5-7**g-i) identify that there is no systematic difference between the measurements (i.e. the bias between measurements is consistent across the scale of measurements made). The mean bias between the two methods was close to zero, 4.38, 2.23 and 9.65 m^2nmi^{-2} NASC integrated from 0 to 800 m for the transits from the RRS James Clark Ross, FV Janas and RV Tangaroa respectively (**Table 4**). As a proportion of total NASC it represented 1 to 16 percent potential bias and contributes to the range of potential errors/uncertainty in calibration, motion correction, absorption corrections estimated to be 12 to 25%. Therefore, modelling approaches using total water column S_v or NASC as a proxy for mid-trophic level organisms may use data processed from either method.

Examining the difference between the two methods with depth indicates there is a depth-dependency to the differences between the two methods. The BAS-SONA method retains more acoustic backscatter in the surface 200, the two methods retain similar amounts between 200 and 400m, and the CSIRO-IMOS method retains more acoustic backscatter below 400 m (**Table 4**).

The processing methods differ in two places. First, the application of a transient noise filter: the BAS-SONA method removes the whole ping where transient noise is identified, whereas the CSIRO-IMOS method can remove transient noise at depth whilst retaining signal in the surface waters. The second difference is within the background noise calculation. The use of the 90th percentile for noise estimation in the BAS-SONA method, rather than the mean, within the De Robertis and Higginbottom (2007) noise calculations, would result in the BAS-SONA method having lower values of S_v at range, as observed here.

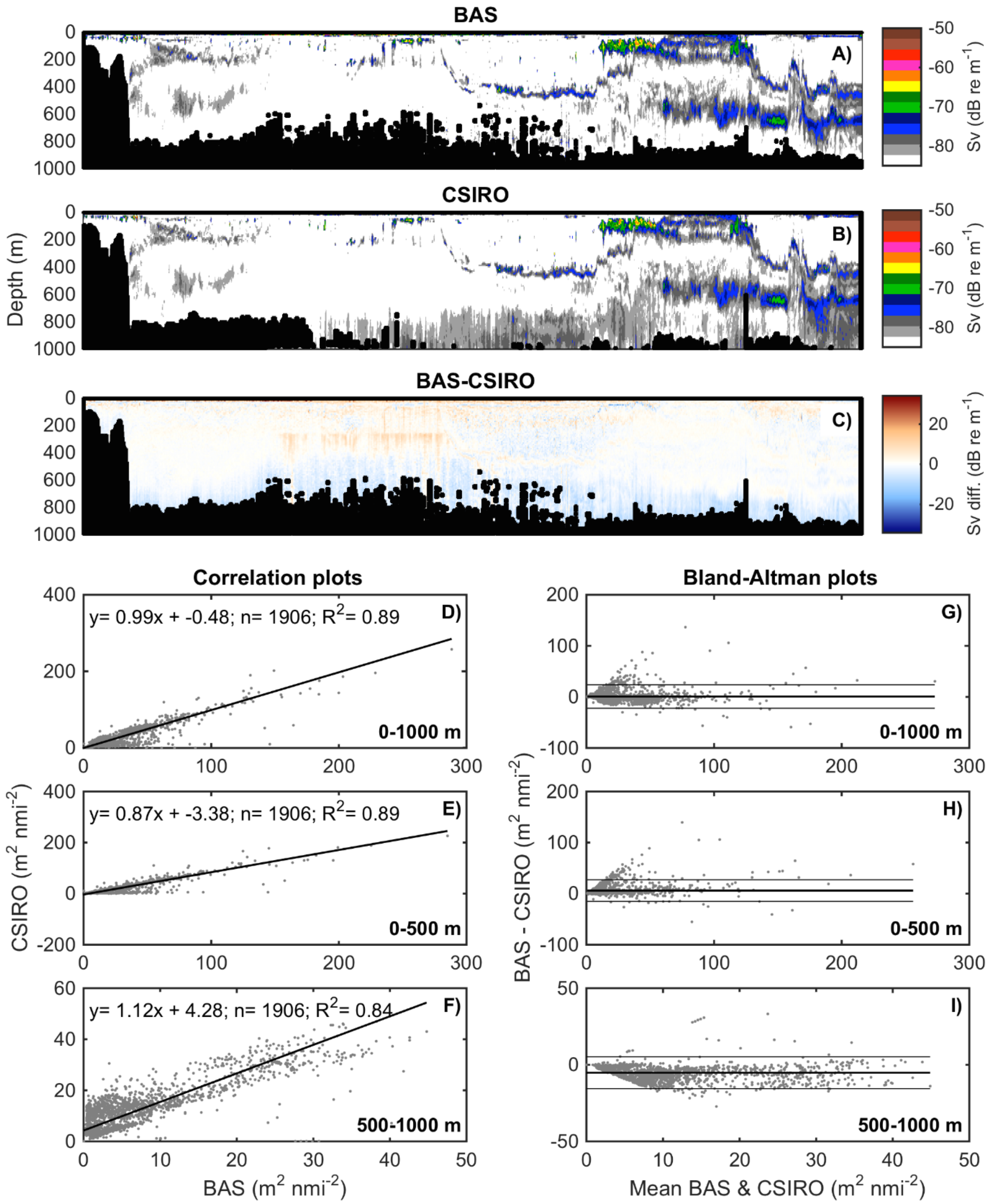

Figure 5 - South Atlantic Sector transit on the RRS James Clark Ross (JR260) a) calibrated post-processed S_v using BAS-SONA processing, b) calibrated post-processed S_v using CSIRO-IMOS processing, c) difference in dB between processing methods, Correlation plots of all (d), top 500m (e) and bottom 500m (f) water column NASC (Nautical Area Scattering Coefficient), and g-i) Bland-Altman plots of agreement between the two different methods for the same depth strata.

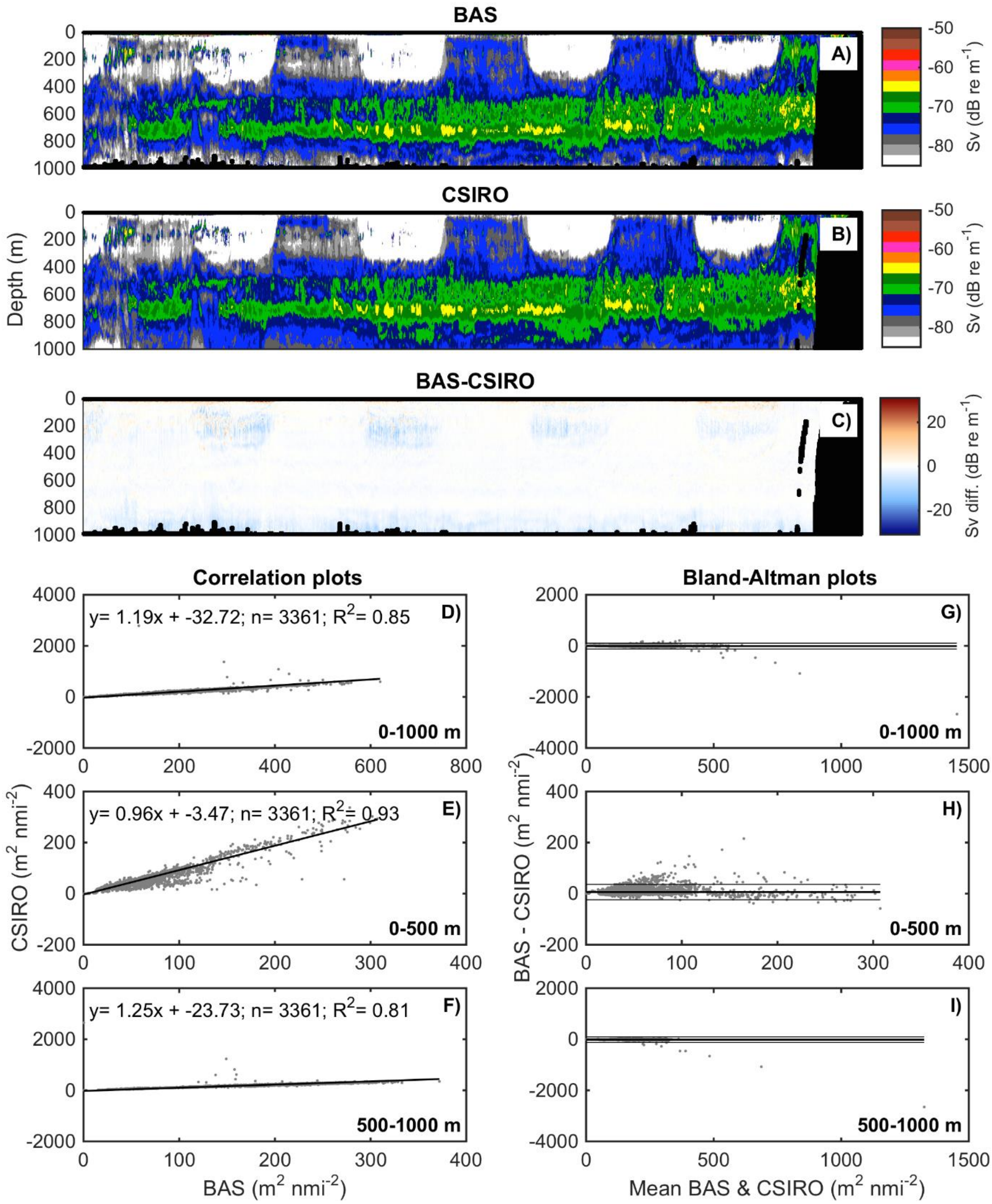

Figure 6 - South Pacific Sector transit on the FV Jana a) calibrated post-processed S_v using BAS-SONA processing, b) calibrated post-processed S_v using CSIRO-IMOS processing, c) difference in dB between processing methods, Correlation plots of all (d), top 500m (e) and bottom 500m (f) water column NASC (Nautical Area Scattering Coefficient), and e-i) Bland-Altman plots of agreement between the two different methods for the same depth strata.

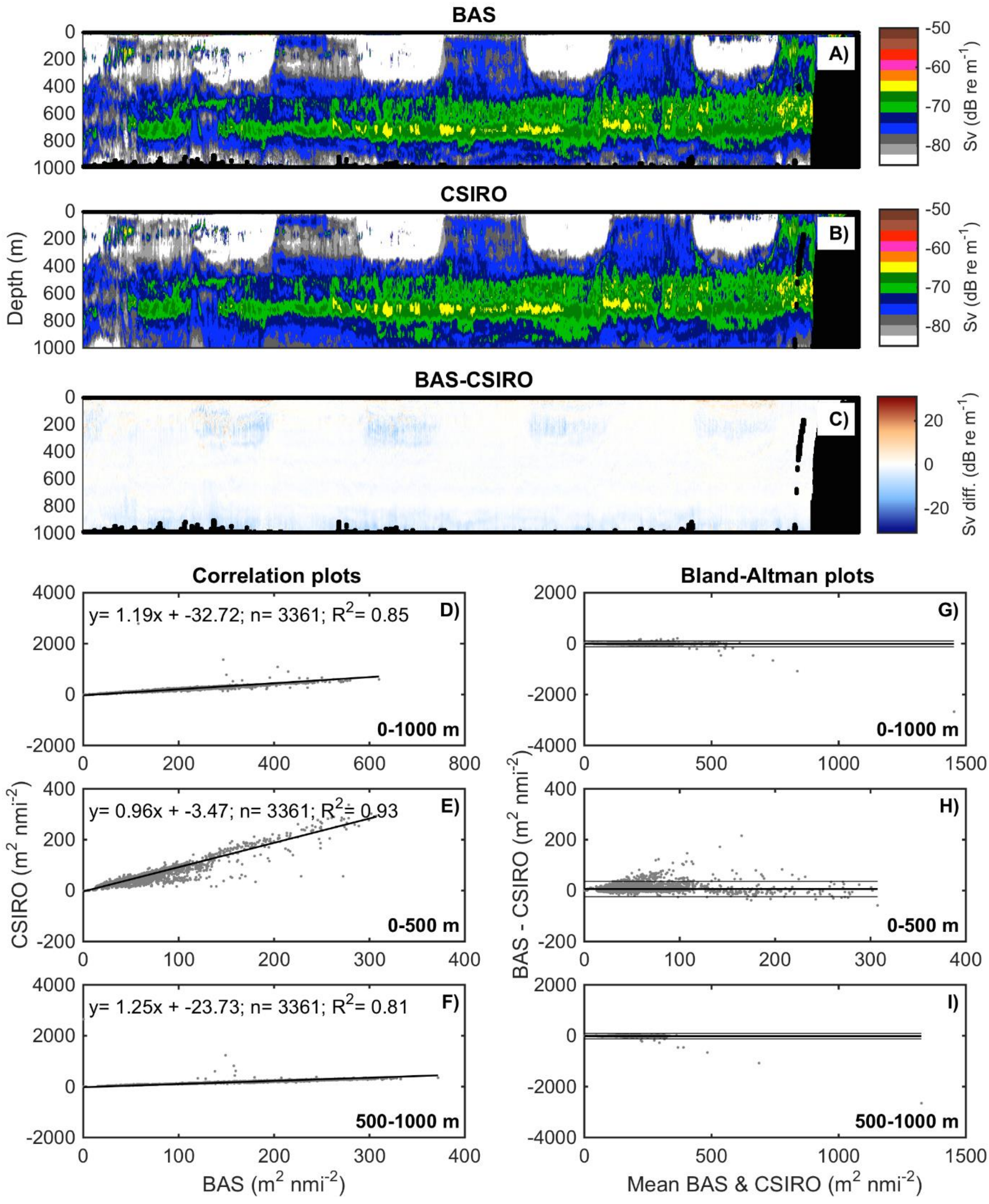

Figure 7 - South Pacific Sector transit on the RV Tangaroa a) calibrated post-processed S_v using BAS-SONA processing, b) calibrated post-processed S_v using CSIRO-IMOS processing, c) difference in dB between processing methods, Correlation plots of all (d), top 500m (e) and bottom 500m (f) water column NASC (Nautical Area Scattering Coefficient), and e-i) Bland-Altman plots of agreement between the two different methods for the same depth strata.

Table 3 - Sector specific settings for background noise detection and filters applied to acoustic backscatter data

Method	South Atlantic Sector	South Indian Sector	South Pacific Sector
Transit selection	1. The vessel travelled at speeds greater than 4 knots 2. The vessel travelled for a period longer than 6 hours 3. The vessel travelled in a consistent direction A matlab tool (SelectEK60track.m) was developed to visualise the vessel track (using gps data from the *.raw* files) during a cruise and interactively select suitable periods of data where the vessel travelled in a consistent direction for long periods.	1. Data quality 2. Vessel travelled a long linear latitudinal transect 3. Vessel travelled between 6 and 16 knots 4. Water depth was ≥1000m	1. Data quality 2. Vessel travelled a long linear latitudinal transect 3. Vessel travelled between 6 and 16 knots 4. Water depth was ≥1000m In-house tools have been developed to assist with data management and help identify and prioritise subsets of data for post-processing, where the temporal and geographic extent of each *.raw* file was captured in a text format file (*inf* files) using the tool *ES60_register.jar* and visualised as geo-referenced rectangle blocks using *Dataview.jar www.imos.org.au.*
Transient noise filter	A 300 ping wide median filter within a user defined layer ($\widetilde{S_{v_n}}$) in the lower 500m of data, data were identified as bad where δ was > 6 dB. Note not applied to data from the RRS James Clark Ross.	A 150 ping wide sliding median filter within a user defined layer ($\widetilde{S_{v_n}}$) where δ varied between 3 and 6 dB between transits	Filter not applied to data shallower than 250m. A 50 ping wide, 20m vertical range median ($\widetilde{S_{v_{nm}}}$) was calculated through the water column. Data were identified as bad where δ was > 12 dB.
Impulse noise filter	Data were vertically averaged to 1 m resolution and values removed if δ exceeded 10 dB. The IN filter was implemented step-by-step as described by Ryan et al. (2015) rather than using the Echoview single variable IN filter, as the Echoview IN filter did not remove impulse noise from the dataset.	Interference were replaced by the mean of the values either side.	Data were vertically averaged to 5 m resolution and values removed if δ exceeded 10 dB.
Attenuated signal filter	A 300 ping wide median filter within a user defined layer ($\widetilde{S_{v_n}}$) spanning a mesopelagic scattering layer, data were identified as bad where δ was < -6 dB.	A 150 ping wide sliding median filter within a user defined layer ($\widetilde{S_{v_n}}$) where δ varied between -3 and -6 dB between transits	A 30 to 300 ping wide median filter within a user defined layer ($\widetilde{S_{v_n}}$) spanning a mesopelagic scattering layer, data were identified as bad where δ was < -8 dB.

Method	South Atlantic Sector	South Indian Sector	South Pacific Sector
Background noise removal	Method following De Robertis and Higgenbottom (2007), modified to estimate noise by selecting the minimum 90th percentile of Power$_{cal}$ (equivalent to $S_{vmeasured}$-TVG (Time Varied Gain) from gridded data more similar to the method reported by Korneliussen (2000).	A passive recording of background noise was undertaken every 4 hours and subtracted from the active acoustic data.	Method following De Robertis and Higgenbottom (2007) estimated noise by selecting the minimum average Power$_{cal}$ (equivalent to $S_{vmeasured}$-TVG (Time Varied Gain) from gridded data.
Signal-to-noise-filter	A minimum SNR threshold of 6 dB was set, and all samples that fell below this were set to “-999 dB”.	No SNR threshold was applied.	A minimum SNR threshold of 6 dB was set, and all samples that fell below this were set to “-999 dB”.
Order of application	TN filter, Background noise, IN filter, AS filter, SNR filter	Background noise, IN filter, TN filter, AS filter	Figure 1

Table 4 - Bias in NASC (m^2 nmi^{-2}) between BAS-SONA and CSIRO-IMOS processing methods. NASC was integrated over 0-200, 200-400, 400-800, and 0-800 m range intervals, and results are shown for the three transects processed (RRS James Clark Ross, FV Janas, and RV Tangaroa).

	RRS James Clark Ross					FV Janas					RV Tangaroa				
Range (m)	Mean	Median	Std	Min	Max	Mean	Median	Std	Min	Max	Mean	Median	Std	Min	Max
0-200	5.33	1.95	10.29	-56.20	139.39	6.98	2.07	13.89	-50.62	215.26	4.12	1.09	13.18	-301.98	151.98
200-400	0.36	0.35	1.15	-27.68	11.98	0.06	-0.44	3.00	-11.23	82.35	1.35	0.23	12.12	-255.56	554.83
400-800	-1.31	-1.34	2.97	22.56	34.84	-4.81	-3.29	51.34	-2655.78	57.60	4.18	0.45	19.08	-164.02	1009.07
0-800	4.38	1.61	10.64	-56.44	136.63	2.23	-0.58	54.55	-2667.54	222.15	2.96	2.96	27.56	-464.41	1036.22

4. Impact of the differences in processing on data assimilation results in SEAPODYM-MTL

SEAPODYM-MTL is a spatial ecosystem and population dynamics model of mid-trophic levels (see Lehodey et all, 2010). It models micronekton organisms grouped into six functional groups according to their diel vertical migration (DVM) behaviour. The definition domain of the model is a three-layer ocean with a realistic description on the horizontal plan. The model represents the trophic web as an ensemble of energy fluxes from primary producer toward the upper trophic levels. The key parameters to estimates are those controlling these energy fluxes: one (*E*) is used to estimate the total transfer of energy from phytoplankton toward micronekton and six others (E'_n) represent the distribution of energy between the six functional groups.

These parameters cannot be measured directly *in-situ*. They must be estimated indirectly. To meet this purpose, a data assimilation framework has been developed (Lehodey et al, 2015). This framework used acoustic data to optimise the model parametrisation, especially the $E'_{n.}$ parameters.

In this study, we used three transit processed with the CSIRO-IMOS or BAS-SONA methods (see previous sections) to assess the impact of the difference in data processing to the data assimilation results. Two diagnostics are made: one is based on the operator of observations (the outputs of a processing methods used to compare acoustic data to modelled biomass, see Lehodey et al, 2010 for more details) and the other on results of data assimilation (namely optimal estimates of $E'_{n.}$ parameters).

We used the three transits processed with both methodology to create the operators of observations ρ_{epi}, $\rho_{u\text{-}meso}$ and $\rho_{l\text{-}meso}$ (ratio of NASC by layers and transit, calculated for each day or night periods).

As shown in **Figure 8** most of the observations come close to the first bisector and thus are similar whatever the methodology used. However, there are a few outliers in terms of NASC ratio (e.g., one portion of the JCR transect has 70% of the NASC in the first layer using the BAS-SONA methodology and only 10% using the IMOS one). A statistical comparison (Table 5: Difference in NASC ratio (mean plus or minus the standard deviation) for each layer and each period of the day. Difference are calculated by subtracting pairwise BAS-SONA ratio to CSIRO-IMOS ratio.**Table 5**) indicates that differences between the two methodologies are greater in the epipelagic and lower mesopelagic layers.

Table 5: Difference in NASC ratio (mean plus or minus the standard deviation) for each layer and each period of the day. Difference are calculated by subtracting pairwise BAS-SONA ratio to CSIRO-IMOS ratio.

Layer	Night	Day
Epipelagic	-0.027 ± 0.072	-0.078 ± 0.158
Upper-mesopelagic	0.009 ± 0.031	0.008 ± 0.096
Lower-mesopelagic	0.019 ± 0.074	0.070 ± 0.097

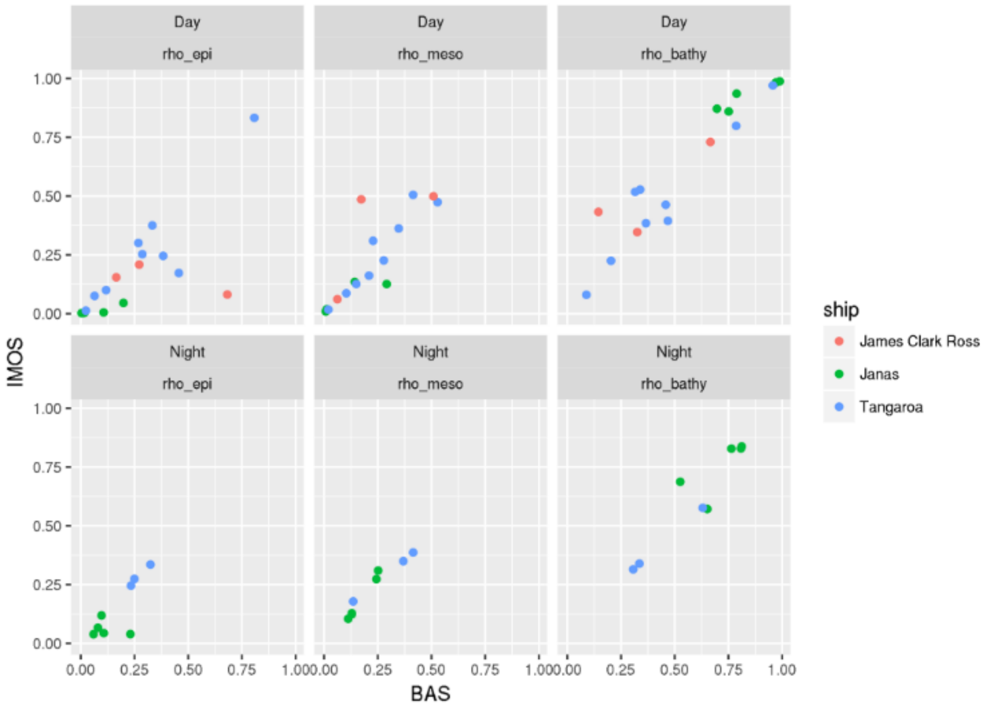

Figure 8: Comparison between SEAPODYM-MTL operators of observation using the transects processed with the CSIRO-IMOS and SONA-BAS methods.

Once aggregated at the resolution of the model, the number of observations (N=26) from these 3 transects appeared too small to run an optimisation experiment. Therefore, a test was conducted by applying the mean difference calculated above to all available transects available in the MESOPP database (www.mesopp.eu) processed by BAS-SONA or CSIRO-IMOS and acquired during the temporal extent of the model simulation (2006-2015).

The model configuration was:

- Physical fields from Glorys2V4 ocean circulation reanalysis from COPERNICUS-MEMS catalogue for currents and temperature at resolution ¼ ° x week http://marine.copernicus.eu/;
- Biogeochemical fields from VGPM model (Berhenfeld and Falkowski 1997) for primary production and euphotic depth; (https://www.science.oregonstate.edu/ocean.productivity/);
- Global domain and time series from January of 2006 to December of 2015;

Three runs were produced:

- The control run: unbiased ratios directly calculated from transects available online;
- The IMOS-based optimisation: using CSIRO-IMOS processed transects and modifying those from BAS-SONA by adding the mean difference between ratio shown in Table 5;
- The BAS-SONA-based optimisation: using BAS-SONA processed transects and modifying those from CSIRO-IMOS by adding the mean difference between ratio shown in Table 5.

The three optimisation experiments successfully converged and produced slightly different final log-likelihood, i.e., the measure of the distance between observations and model prediction. The log-likelihood achieved with the CSIRO-IMOS configuration is slightly lower (**Table 6**). Energy transfer parameters are estimated in the same proportions between epipelagic, upper- and lower mesopelagic groups with almost half of the energy allocated to the non-migrant lower-mesopelagic group, approximately 20% for the non-migrant upper-mesopelagic and epipelagic groups, just a few percent to the lower-mesopelagic migrant and highly migrant groups and no energy at all allocated to the migrant upper-mesopelagic group.

There are significative differences within two specific groups: the epipelagic non-migrant group and the highly migrant lower-mesopelagic group (**Figure 9**). These preliminary results indicate that the model is relatively sensitive to the methodology used for processing acoustic data. The bias applied to the observations in the two experiments led to a corresponding re-allocation of biomass in the equivalent allocation by layers, meaning that the model is behaving well. Two parameters seem more sensible to the difference in methodology: the energy allocation for the non-migrant epipelagic group and for the highly-migrant lower-mesopelagic group, which is a logical result as the major difference in the transect processed with the two methods are at depth and at the surface.

The operator of observation has been created with a very simple assumption (a constant bias between the outputs of the two processing methodologies) and also assumes that 38kHz NASC is a direct proxy for mid-trophic level biomass. New experiments with larger real datasets should be conducted. A more detailed sensitivity analysis could be done also to test the impact of the different stages in the chain of processing.

Table 6. Value of final log-likelihood for the three experiments

Experiment	***Likelihood***
Control	31.1
CSIRO-IMOS	27.1
BAS-SONA	29.6

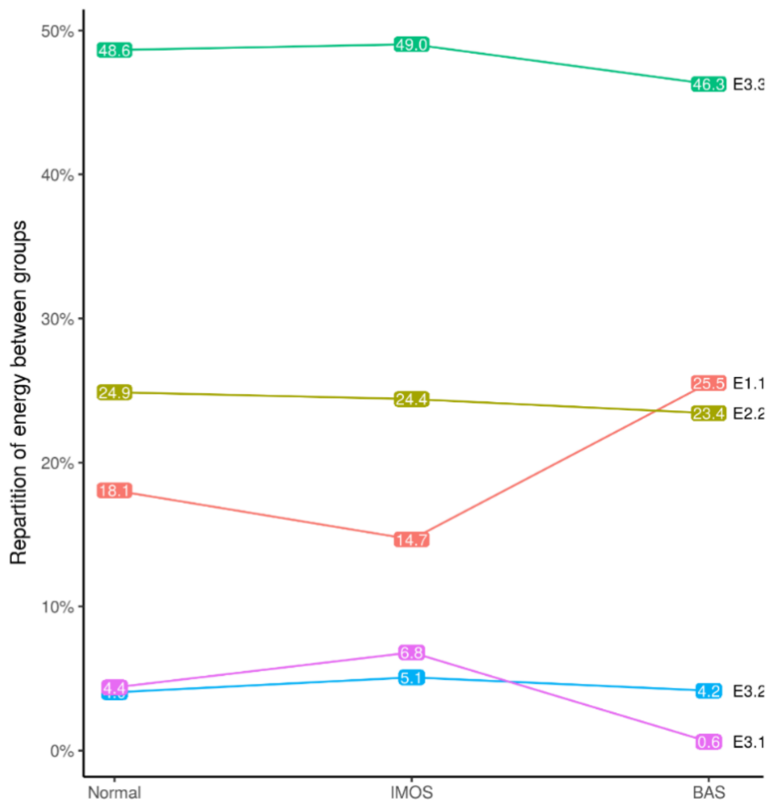

Figure 9. Value of 'optimal' parameters for each experiment. For clarity, the $E_{2.1}$ parameter is not represented as its 'optimal' value is null in each experiment.

5. Discussion

Active acoustic systems have the potential to deliver valuable information for aquatic ecosystem science and management. However scientific studies are limited by practical and financial constraints leading to small data coverage in time and/or space. Collection and storage of digital acoustic data from both research and fishing vessels during both dedicated surveys and opportunistically has rapidly expanded the temporal and spatial coverage as well as the amount of data collected (Escobar-Flores et al. 2018; Behagle et al. 2016). Larger datasets of basin to global scale acoustic data sets have enabled new insights into mesopelagic fish, such as greater biomass estimates of mesopelagic fish (Irigoien et al. 2014), intensification of open-ocean oxygen depleting by vertically migrating animals (Bianchi et al. 2013) and inferred increase in mesopelagic biomass and trophic efficiency by 2100 (Proud et al. 2017).

Ocean observation science is entering into the big data era (Liu et al. 2016), and active acoustic data will contribute to this. Big data has one or more of the following characteristics: volume; velocity (speed of acquisition); variety and veracity (Beyer and Laney 2012). Acoustic data are voluminous and complex such that traditional data processing software are becoming inadequate to deal with them and traditional relational databases struggle to capture, manage and process. They have the potential to have high velocity. For example, if every powered fishing vessel (225,000 in the world, FAO 2016) and every research vessel (836, https://www.researchvessels.org/qryshipinfo.asp) had an echosounder (Exemplar echosounder: Simrad EK60, 1 frequency, 1ms pulse, data saved to 500 m), they could generate ~2Gb of data per ping, and ~100Tb per day (2 second ping rate). In addition, oceanography, and therefore acoustics, is evolving from a ship-based science to a distributed, observatory-based approach with data collected from multiple platforms (Handegard et al. 2012). These include acoustic instruments on ships, autonomous underwater vehicles and gliders (Fernandes et al. 2002; Guihen et al. 2014), towed and lowered bodies (Kloser et al. 1996), moorings (Saunders et al. 2007), and seafloor electro-optic cables (Godo et al. 2014). Resultant data must have appropriate metadata, be stored, calibrated and processed for quality control, and importantly be discoverable and accessible.

The biological acoustic community has started to address these requirements: The ICES metadata standards for acoustic data has been expanded to allow for non-vessel platforms as data collectors (ICES 2014). The EU project MESOPP (www.mesopp.eu) and observing systems such as IMOS (www.imos.au) are now distributing calibrated, quality-controlled, post-processed data. The benefits of these repositories are global access to data, cross-institution collaboration and the ability to address cross-cutting scientific questions. However, robust algorithms and methods need to be developed so automated processing becomes less time intensive, more affordable and therefore more typical. This has led to a new focus on the use of and development of open-source software to read, process and visualise acoustic data allowing efficient large-scale processing (e.g. PyEcholab, Anderson et al. 2018b). An example of large-scale analyses using open-source software is the automated identification of sound-scattering layers undertaken by Proud et al. (2015) on 40 surveys of data obtained from several data centres. The remaining challenge will be the ability to locate and use all acoustic data, regardless of where it is stored and maintained, through one or a network of data portals, with post-processing undertaken on demand with either pre-defined or user-defined filters.

6. Summary and conclusion

This report summarises the post-processing steps used to generate calibrated, post-processed S_v freely-available from the MESOPP website (www.mesopp.eu). Datasets from 6 ships covering three sectors of the Southern Ocean are available. Data were collected from research vessels during research cruises and logistical transits and from fishing vessels during logistical transits. In each case standardised data collection settings (Ryan 2011) were required in order for the data to be considered comparable and available for inclusion in a larger observing network. Well-specified standard settings will enable the methodology to be adapted to the development of a large network of sampling using both vessels of opportunity and autonomous platforms.

In each case the echosounder was calibrated and then post-processed following similar steps, but with different software or settings. Many parts of the post-processing is now undertaken with semi-automated filters (as described above), only requiring a user to provide parameterisation of the filters depending on sampling platform and weather conditions. Remaining steps to enable full automated processing is identification of “false bottom” in the acoustic data and automated parameterisation of the filters mentioned above.

A preliminary comparison indicated that two of the three post-processing methods used generated similar values of water column NASC integrated from 0 to 800 m. There were some differences between biases with depth, although all still low. SEAPODYM LMTL modelling is optimised using the ratio between layers and therefore this discrepancy in bias with depth may cause greater uncertainty when using data post-processed with different methods. As a result it is clear that to ensure good data quality and assurance processes, data processing techniques should be compared on common data sets. Estimates of uncertainty to highlight precision and accuracy of the measurements and whether those variances matter in models need to be assessed so that potential errors are identified, and new and emerging methods are evaluated, in order to recommend best practice.

7. References:

Anderson C, Wall C, Cutter G, Towler R, Jech JM (2018b) PyEcholab: a toolkit for water column echosounder data analysis. Ocean Sciences Meeting, 11-16 February, Portland, Oregon.

Anderson CIH, Brierley AS, Armstrong F (2005) Spatio-temporal variability in the distribution of epi- and mesopelagic acoustic backscatter in the Irminger Sea, North Atlantic, with implications for predation on Calanus finmarchicus. Marine Biology, 146: 1177–1188.

Anderson TR, Martin AP, Lampitt RS, Trueman CN, Henson SA, Mayor DJ (2018a) Quantifying carbon fluxes from primary production to mesopelagic fish using a simple food web model. ICES Journal of Marine Science, doi: 10.1093/icesjms/fsx234.

Béhagle N, Cotté C, Ryan TE, Gauthier O et al. (2016) Acoustic micronektonic distribution is structured by macroscale oceanographic processes across 20–50°S latitudes in the South-Western Indian Ocean. Deep-Sea Research I, 110: 20–32.

Benoit-Bird KJ, Lawson GL (2016) Ecological insights from pelagic habitats acquired using active acoustic techniques. Annual Reviews of Marine Science, 8: 463-490.

Bertrand A, Josse E, Bach P, Dagorn L (2003) Acoustics for ecosystem research: lessons and perspectives from a scientific programme focussing on tuna-environment relationships. Aquatic Living Resources, 16: 197-203.

Beyer MA, Laney D (2012) The Importance of ‘Big Data’: a definition. Gartner Inc, Stamford

Bianchi D, Galbraith ED, Carozza DA, Mislan KAS, Stock CA (2013) Intensification of open-ocean oxygen depletion by vertically migrating animals. Nature Geoscience: NGEO1837. 4 pp.

Colosi J.A., Worcester P.F. (2013). Introduction to special issue on deep water ocean acoustics. The Journal of the Acoustical Society of America, 134 (4).

Cox MJ, MacKenzie ML, Watkins JL, Brierley AS (2006) The effect of missing acoustic observations (dropped pings) on mean area density estimates of Antarctic krill (Euphausia superba). ICES CM /I: 16.

Dalen J, Lovik A (1981) The influence of wind-induced bubbles on echo integration surveys. The Journal of the Acoustical Society of America, 69: 1653–1659.

De Robertis A, Higginbottom I (2007) A post-processing technique to estimate the signal-to-noise ratio and remove echosounder background noise. ICES Journal of Marine Science, 64: 1282-1291.

Demer DA, Berger L, Bernasconi M, Bethke E, Boswell K, Chu D, Domokos R, et al. (2015) Calibration of acoustic instruments. ICES Cooperative Research Report No. 326: 133 pp.

Dunford AJ (2005) Correcting echo-integration data for transducer motion. The Journal of the Acoustical Society of America, 118: 2121–2123.

Escobar-Flores PC, O'Driscoll RL, Montgomery JC (2018) Spatial and temporal distribution patterns of acoustic backscatter in the New Zealand sector of the Southern Ocean. Marine Ecology Progress Series, 592: 19-35.

FAO (2016) The state of world fisheries and aquaculture 2016. Contributing to food security and nutrition for all. Rome. 200 pp.

Fernandes PG, Gerlotto F, Holliday DV, Nakken O, Simmonds EJ (2002) Acoustic applications in fisheries science: the ICES contribution. ICES Marine Science Symposium, 215: 483-492.

Fielding S, Watkins JL, Trathan PN, Enderlein P, Waluda CM, et al. (2014) Inter-annual variability in Antarctic krill (*Euphausia superba*) density at South Georgia, Southern Ocean: 1997-2013. ICES Journal of Marine Science, 71: 2578-2588.

Foote KG, Knudsen HP, Vestnes G, MacLennan DN, Simmonds EJ (1987) Calibration of acoustic instruments for fish density estimation: a practical guide. ICES Cooperative Research Report No. 144: 69 pp.

Francois RS, Garrison GR (1982) Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption. The Journal of the Acoustical Society of America, 72: 1879-1890.

Gjøsæter J, Kawaguchi K (1980) A review of the world resources of mesopelagic fish. Food and Agriculture Org.

Godø OR, Johnsen S, Torkelsen T (2014) The LoVe Ocean Observatory is in Operation. Marine Technology Society Journal 48(2): 24-30.

Guihen D, Fielding S, Murphy EJ, Heywood KJ, Griffiths G (2014) An assessment of the use of ocean gliders to undertake acoustic measurements of zooplankton: the distribution and density of Antarctic krill (*Euphausia superba*) in the Weddell Sea. Limnology and Oceanography: Methods, 12: 373-389.

Handegard NO, du Buisson L, Brehmer P, Chalmers SJ, De Robertis A, Huse G, Kloser R et al. (2012) Towards an acoustic-based coupled observation and modelling system for monitoring and predicting ecosystem dynamics of the open ocean. Fish and Fisheries, 14:605-615.

Hewitt RP, Watkins JL, Naganobu M, Sushin V, Brierley AS, et al. (2004) Biomass of Antarctic krill in the Scotia Sea in January/February 2000 and its use in revising an estimate of precautionary yield. Deep-Sea Research II, 51: 1215-1236.

Honkalehto T, Ressler PH, Towler RH, Wilson CD (2011) Using acoustic data from fishing vessels to estimate walleye Pollock (*Theragra chalcogramma*) abundance in the eastern Bering Sea. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1231-1242.

ICES (2014) A metadata convention for processed acoustic data from active acoustic systens, SISP4 TG-AcMeta. ICES WGFAST Topic Group, TG-AcMeta: 35 pp.

ICES (2018) WGIPS 2017

Irigoien X, Klevjer TA, Røstad A, Martinez U, Boyra G, Acuña JL, et al. (2014). Large mesopelagic fishes biomass and trophic efficiency in the open ocean. Nature Communications 5: 3271. doi: 10.1038/ncomms4271

Kloser RJ (1996) Improved precision of acoustic surveys of benthopelagic fish by means of a deep-towed transducer. ICES Journal of Marine Science, 53: 407-413.

Kloser RJ, Ryan T, Young J, Lewis ME (2009) Acoustic observations of micronekton fish on the scale of an ocean basin: Potential and challenges. ICES Journal of Marine Science 66: 998-1006.

Korneliussen RJ (2000) Measurement and removal of echo integration noise. ICES Journal of Marine Science, 57: 1204-1217.

Lea MA, Nichols PD, Wilson G (2002) Fatty acid composition of lipid-rich myctophids and mackerel icefish (*Champsocephalus gunnari*) – Southern Ocean food-web implications. Polar Biology 25, 843-854.

Liu Y, Qiu M, Liu C, Guo Z (2016) Big data challenges in ocean observation: a survey. Personal Ubiquitous Computing, 21: 55-65.

Mackenzie KV (1981) 9-term equation for sound speed in the oceans. The Journal of the Acoustic Society of America, 70:807-812.

MacLennan DN, Fernandes PG, Dalen J (2002) A consistent approach to definitions and symbols in fisheries acoustics. ICES Journal of Marine Science, 59: 365-369

Mitson RB, Knudsen HP (2003) Causes and effects of underwater noise on fish abundance estimation. Aquatic Living Resources, 16: 255-264.

Nunnallee EP (1990) An alternative to thresholding during echo-integration data collection. Rapports et Procès-Verbaux des Réunions du Conseil International pour l'Exploration de la Mer, 189: 92-94.

Ona E (1991) Vessel heave, an index for air bubble attenuation correction. ICES Document CM 1991/B:37.

Proud R, Cox MC, Wotherspoon S, Brierley AS (2015) A method for identifying Sound Scattering Layers and extracting key characteristics. Methods in Ecology and Evolution, doi: 10.1111/2041-210X.12396.

Ryan T, Downie RA, Kloser RJ, Keith G (2015) Reducing bias due to noise and attenuation in open-ocean echo integration data. ICES Journal of Marine Science, 72(8): 2482-2493.

Ryan T. (2011). Overview of Data collection, management and processing procedures of underway acoustic data. IMOS BASOOP Sub-facility. http://imos.org.au/fileadmin/user_upload/shared/SOOP/plugin-IMOS_data_collection_and_processing_v1.01.pdf

Ryan TE, Kloser RJ (2004) Quantification and correction of a systematic error in Simrad ES60 Echosounders. In ICES FAST. Gdan´ sk. Copy available from CSIRO Marine and Atmospheric Research. GPO Box 1538, Hobart, Australia.

Saunders RA, Brierley AS, Watkins JL, Reid EJ, Murphy EJ, Enderlein P, Bone DG (2007) Intra-annual variability in the density of Antarctic krill (*Euphausia superba*) at South Georgia, 2002-2005: Within-year variation provides a new framework for interpreting previous 'annual' estimates of krill density. CCAMLR Science, 14: 27-41.

Simmonds EJ, MacLennan DN (2005) Fisheries Acoustics Theory and Practise. Blackwell Science, Oxford: 437 pp.

St John MA, Boja A, Chust G, Heat M, Grigorov I, Mariani P, Martin AP, Santos RS (2016) A dark hole in our understanding of marine ecosystems and their services: perspectives from the mesopelagic community. Frontiers in Marine Science 3:31.

Stanton TK (1982) Effects of transducer motion on echo-integration techniques. The Journal of the Acoustical Society of America, 72: 947.

Trenkel VM, Berger L, Bourguignon S, Doray M, Fablet R, Massé J, Mazauric V, Poncelet C, Quemener G, Scalabrin C, Villalobos H (2009). Overview of recent progress in fisheries acoustics made by Ifremer with examples from the Bay of Biscay. Aquatic Living Resources 22, 433-445

Vaseghi SV (2009) Advanced digital signal processing and noise reduction. Wiley: 544 pp.

Wall CC, Jech JM, McLean SJ (2016) Increasing the accessibility of acoustic data through global access and imagery. ICES Journal of Marine Science, 73(8): 2093-2103.

Wang X, Zhang J, Zhao X (2016) A post-processing method to remove interference noise from acoustic data collected from Antarctic krill fishing vessels. CCAMLR Science, 23: 17-30.

Watkins JL, Brierley AS (1996) A post-processing technique to remove background noise from echo integration data. ICES Journal of Marine Science 53: 339-344.

8. Appendix: NetCDF structure for acoustic data in MESOPP project

NetCDF structure for acoustic data in MESOPP project

Deliverable Lead:	CLS
Reference:	MESOPP-17-0001
Dissemination Level:	Public
Issue:	5. 0
Date:	2017,Nov.16

Horizon 2020. Grant agreement No 692173

Chronology Issues:

Issue:	Date:	Reason for change:	Author
0.0	21/07/2017		B. Calmettes
1.0	10/08/2017	Update with comments from: Sophie Fielding Gavin Macaulay Carrie C. Wall Tim Ryan	B. Calmettes A. Conchon
2.0	01/09/2017	Changes in sampling design approach	B. Calmettes
3.0	11/10/2017	Update with comments from Sophie Fielding and Alejandro Ariza	B. Calmettes
4.0	06/11/2017	One file per frequency	B. Calmettes
5.0	17/11	Comments from Haris Kunnath and Roger Proctor	B.Calmettes

Distribution:

Company	Means of distribution	Names
CLS All MESOPP partners	Notification and sharepoint	

List of tables and figures

List of tables:

List of Contents

Overview of MESOPP project

The underlying concept of MESOPP is the creation of a collaborative network and associated e-infrastructure (marine ecosystem information system) between European and Australian research teams/institutes sharing similar interests in the Southern Ocean and Antarctica, its marine ecosystem functioning and the rapid changes occurring with the climate warming and the exploitation of marine resources.

In the past 30 years, facing global knowledge issues, lacking data, addressing huge modelling challenges, we observed the successful world organisation of meteorology. These past 15 years, Europe has kick started and demonstrated similar successful structuring of the operational oceanography fostered by the Copernicus initiative (http://marine.copernicus.eu/). Today, it is worldwide used and recognised, fully anticipated and integrated in GOOS (Global Ocean Observing System), IOOS (Integrated ocean observation system), SOOS (Southern Ocean Observing System), GODAE (the International Global Ocean Data Assimilation Experiment), and IMBER (Integrated Marine Biogeochemistry and Ecosystem Research).

The next major R&D strategic challenge is to connect the marine ecosystem community across the fields of meteorology, climate, oceanography and biology. Lack of data, development of accurate high end models, global coverage and need for exchange are issues that need to be overcome.

The objective of the MESOPP project is to meet this challenge and is threefold:

1. Make an inventory of science challenges, stakes and existing policies and develop tools to federate and structure the community;
2. Start to organise the related marine ecosystem community between EU and Australia through two implementation actions
3. Propose a R&D roadmap to support a large international cooperation on marine ecosystems based on a e-infrastructure with additional countries such as USA, New Zealand, Canada (in the Frame of the Galway statement), Brazil and all active countries already involved in large organisations such as IMBER, CCAMLR or IMOS.

MESOPP will focus on the enhancement of collaborations by eliminating various obstacles in establishing a common methodology and a connected network of databases of acoustic data for the estimation of micronekton biomass and validation of models. It will also contribute to a better predictive understanding of the SO based on furthering the knowledge base on key functional groups of micronekton and processes which determine ecosystem dynamics from physics to large oceanic predators.

This first project and associated implementation (science network and specification of an infrastructure) should constitute the nucleus of a larger international program of acoustic monitoring and micronekton modelling to be integrated in the general framework of ocean observation following a roadmap that will be prepared during the project.

1. Introduction

The purpose of this document is to define the NetCDF4 classic model structure of the processed acoustic data in the context of the MESOPP project. This definition is crucial to allow a better understanding of the information included in each file and a better dissemination via the CIS (Central Information System).

The metadata for the MESOPP project must comply with the CF convention (http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.pdf) and with the ICES convention (SISP4 - A metadata convention for processed acoustic data from active acoustic systems)

It is important to recall the acquisition framework of the acoustic data: a NetCDF **dataset file** includes one acoustic channel from one mission (a mission could be constituted by multiple files, each containing data acquired at a specific frequency from a specific instrument). A mission consists in a set of acoustic data from a platform (the mission platforms are listed in the ICES convention). And the platform has a sampling design that can be either stationary (acoustic data acquired in a single location by ships, moorings, acoustic stations, …) or mobile (acoustic data acquired along a transect by ships, drifters, gliders, drones, ..)

The information in the "Transects metadata" or "Mooring metadata" described in the ICES recommendation is already included in other attributes and can be limiting for new sampling platforms.

This document describes the file name and the structure of the acoustic files for MESOPP standard. It details the required global attributes, dimensions and variables in the NetCDF files. The structure follows globally ICES convention but requires additional mandatory attributes, supresses other ones and simplified the platform attributes. They are highlighted in red in the following text

To be fully equivalent with MESOPP standard, the ICES convention should now include the sampling design.

The data in the "**data_created**" attribute can be included in the mandatory history attribute and the "data_created" attribute removed.

2. File names

Prior to define the internal structure in the acoustic files, it is important to determine the naming convention of the acoustic data files. Its aim is to allow an early identification of the relevant data included in the file.

For the MESOPP project, the acoustic files in NetCDF format must have the following name type:

[Owner]_[Mission]_[Type of sampling design]_[Platform name]_[Regional location name]_[Code instrument]_[Frequency]_[First date]_[End date].nc

Where:

Owner: the owner of the acoustic data (e.g., BAS / IMOS / UPMC)

Mission: Short name of the mission (e.g., BASOOP, MyctO3DMAP)

Type of sampling design: S (Stationary) or M (Mobile)

Platform name: name of the platform

Regional location name: a short name for the geographical domain

Code instrument: a short name for the acoustic instrument model (EK60, EK80…)

Frequency: The frequency in kHz

First date: The beginning **UTC** date in ISO8601 format

End date: The ending **UTC** date ISO8601 format

EXAMPLE:

Filename:

IMOS_SOOP-BA_M_SouthernSurveyor_M_TAS_EK80_38kHz_2015-12-05T04Z_2016-02-01T20Z.nc

Owner: IMOS

Mission: SOOP-BA

Type of sampling design: Mobile (M)

Platform name: Southern Surveyor

Regional location name: Tasmania (TAS)

Code instrument: EK80

Frequency: 38 kHz

First date/location: 5th December 2015 at 04 hours UTC

End date/location: 1st February 2016 at 20 hours UTC

3. Metadata

In the ICES Convention, there are four kinds of attributes depending on their obligation: Mandatory (M), Mandatory if applicable (MA), Recommended (R) and Optional (O).

This document describes the mandatory attributes in the context of the NetCDF structure for acoustic data in the MESOPP project files. It includes all the mandatory attributes in the ICES convention and some attributes in the other categories.

4. Global Attributes

The global attributes contain all the general information about:

- The metadata record
- The project
- Instruments
- Calibration
- Data acquisition
- Data processing
- Data attributes

4.1. Metadata record

This is the information concerning the ICES convention. The following table contains the mandatory attributes for the metadata record:

Table 1. Metadata record attributes

Attribute name	Comment
convention_name	ICES Convention
convention_version	The ICES Convention version (ex 1.10)
convention_reference	Record the reference for the ICES convention. (for example, ICES 2016. A metadata convention for processed acoustic data from active acoustic systems, SISP 4 TG-AcMeta Version 1.10, ICES WGFAST Topic Group, TG-AcMeta 47pp)

4.2. Dataset Attributes

The metadata of the project is included as global attribute in the NetCDF file (Category Dataset attributes in the ICES convention). The following table contains the mandatory attributes for MESOPP.

Table 2. Metadata for dataset information

Attribute name	Comment
title	Following ICES and CF Conventions. Short description of the dataset
history	Following the CF convention. Each line must begin with a timestamp indicating the when the data was generated
institution	Following ICES and CF Conventions. It indicates where the original data was produced
source	This attribute does not exist in the ICES convention and it is mandatory in the CF convention. The source corresponds to the method of production of the original data.
references	Following ICES and CF Conventions. Published or web-based references that describe the data or the methods used to produce the data
project	Following ICES convention. The scientific project that produced the data
abstract	Following ICES convention. A paragraph describing the dataset
keywords	Following ICES convention. A comma separated list of keywords and phrases. The use of keywords from the Global Change Master Directory (GCMD) is recommended
doi	Following ICES convention Digital Object Identifier
citation	Following ICES convention. The citation to be used in publications
license	Following ICES convention. Describes the restrictions to data access and distribution
author_email	Following ICES convention. Email address of the person responsible for the creation of the dataset
author	Following ICES convention. Name of the person responsible for the creation of the dataset
distribution_statement	Following ICES convention. Statement describing data distribution policy

4.3. Mission Attributes

The metadata of the mission is also included as global attribute in the NetCDF file (Category Dataset attributes in the ICES convention).

Table 3. Metadata for mission information

Attribute name	Comment
mission_name	Following ICES Convention. The name of the mission
mission_abstract	Following ICES Convention. Free text description of the mission
mission_start_date	Following ICES Convention. Start date of mission in ISO 8601 including local time zone
principal_investigator	Following ICES convention. Name of the principal investigator in charge of the mission
principal_investigator_email	Following ICES convention. Principal investigator email address
data_centre	Following ICES convention. Data centre in charge of the data management or party who distributes the resource
data_centre_email	Following ICES convention. Data centre contact email address
mission_platform	ICES convention. Platform type following Appendix B.1 in the ICES convention ("Ship, research", "Ship, fishing", "Ship, other", "Buoy, moored", "Buoy, drifting", "Glider", "Underwater vehicle, autonomous, motorized", "Underwater vehicle, towed", "Underwater vehicle, autonomous, glider")
sampling_design	Not in the ICES convention. Possible values: Stationary (S) or Mobile (M)
ancillary_instrumentation	Mandatory if sampling design is stationary. List suit of instruments and other equipment potentially relevant to the acoustic dataset

4.4. Instrument Attributes

Table 4.Metadata for instrument information

Attribute name	Comment
instrument_frequency	Following ICES Convention. Frequency (in kHz) of the transceiver/transducer combination
instrument_transducer_location	The values of this global attribute are specified in the ICES convention: "Hull, keel", "Hull, lowered keel", "Hull, blister", "Hull, gondola", "Towed, shallow", "Towed, deep", "Towed, deep trawlnet attached" or "Ship, pole"
instrument_transducer_manufacturer	Following ICES Convention. Transducer manufacturer
instrument_transducer_model	Following ICES Convention. Transducer model
instrument_transducer_beam_type	The values of this variable are specified in the ICES convention: "Single-beam", Single-beam, split-aperture", "Multibeam", "Multibeam, split-aperture"
instrument_transducer_orientation	Following ICES Convention. Direction perpendicular to the face of the transducer. For example: "Downward looking", Upward looking"
instrument_transceiver_manufacturer	Following ICES Convention. Transceiver manufacturer
instrument_transceiver_model	Following ICES Convention. Transceiver model

4.5. Calibration attributes

Table 5. Metadata for calibration information

Attribute name	Comment
calibration_date	Following ICES Convention. This variable must be in ISO 8601 extended format including local time zone (e.g. 2013-07-30T10:15:00Z -3)
calibration_acquisition_method	The values of this variable are specified in the ICES convention: “Standard sphere, in-situ”, “Standard sphere, tank”, “Standard sphere, other”, ”Reciprocity”, “Hydrophone”, “Seafloor reflection”, “Nominal”, “Intership”
calibration_processing_method	Following ICES Convention. Describe the method of processing used to generate calibration offsets
calibration_accuracy_estimate	As defined in the ICES convention, this attribute contains the description and units of the calibration
calibration_report	As defined in the ICES convention, this variable contains a reference to external documents.

4.6. Data acquisition attributes

Table 6. Metadata for data acquisition information

Attribute name	Comment
data_acquisition_stored_data_format	Following ICES Convention. Name of the format in which data are stored (for example, HAC)
data_acquisition_ping_duty_cycle	Following ICES Convention. Free text to describe ping duty cycle (for example 10 minutes pinging at 1 ping per second, followed by 50 minutes sleep mode)

4.7. Data processing attributes

Table 7. Metadata for data processing information

Attribute name	Comment
data_processing_software_name	Following ICES Convention. In the MESOPP context, it could be Echoview, MatEcho or EchoMyc
data_processing_software_version	Following ICES Convention. Data processing software version
data_processing_frequency	Following ICES Convention. Data processing frequency in kHz
data_processing_transceiver_power	Following ICES Convention. Data processing transceiver power in W
data_processing_transmit_pulse_length	Following ICES Convention. Data processing transmit pulse length in ms

Attribute name	Comment
data_processing_on_axis_gain	Following ICES Convention. Total system gain value when calibration sphere is on axis in dB
data_processing_transducer_psi	Following ICES Convention Transducer equivalent beam angle, expressed as $10\log_{10}(\psi)$, where ψ has units of steradians

4.8. Data attributes

Table 8. Metadata for data attributes

Attribute name	Comment
data_ping_axis_interval_origin	Following the ICES convention: Start, Middle, End
data_ping_axis_interval_value	Following the ICES convention. In the MESOPP context, this value is set to 1000
data_ping_axis_intervale_type	Following the ICES convention. In the MESOPP context, this value is set to "Distance (metres)"
data_range_axis_interval_origin	Following the ICES convention. Start, Middle, End
data_range_axis_interval_value	Following the ICES convention. In the MESOPP context, this value is set to 10
data_range_axis_interval_type.	Following the ICES convention. In the MESOPP context, this value is set to "Range (metres)"

5. Dimensions

For acoustic NetCDF files, three dimensions are mandatory: time, depth, and channel

6. Variables

The variables in the MESOPP file are classified in 6 groups:

- Variables associated with dimensions
- Variables associated with locations
- Variables associated with micronekton

The standard_name attribute for some variables is not still defined in the CF convention. When it does not exist, we respected the guidelines for the construction of standard names of the CF convention. They consist of lower-letters, digits and underscores, and begin with a letter. Upper case is not used.

The type of the variables can be double or float.

6.1. Variables associated with dimensions

double time(time)

Attribute name	Value
standard_name	time
long_name	time
units	days since 1950-01-01 00:00:00 **UTC**
calendar	gregorian
axis	T

float depth (depth): this variable contains the underwater depth (and not recording range)

Attribute name	Value
standard_name	depth
long_name	Mean depth of integration layer
units	metres or m
axis	Z
valid_min	-5
valid_max	12000

6.2. Locations variables

float latitude(time)

Attribute name	Value
standard_name	latitude
long_name	latitude
units	degree_north or degrees_north
valid_min	-90
valid_max	90

float longitude(time)

Attribute name	Value
standard_name	longitude
long_name	longitude
units	degree_east or degrees_eastt
valid_min	-180 (or 0)
valid_max	180 (or 360)

6.3. Acoustic Variables

float Sv(depth, time)

Attribute name	Value
standard_name	mean_volume_backscattering_strength
long_name	Mean volume backscattering strength
units	dB re m-1, dB re 1 m-1
_FillValue	9999
valid_min	-9999
valid_max	20

Appendix A - **List of acronyms**

TBC	To be confirmed
TBD	To be defined
AD	Applicable Document
RD	Reference Document