TNO Defence, Security and Safety

Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek/Netherlands Organisation for Applied Scientific Research

Oude Waalsdorperweg 63
P.O. Box 96864
2509 JG The Hague
The Netherlands

www.tno.nl

T +31 70 374 00 00
F +31 70 328 09 61
info-DenV@tno.nl

TNO report

TNO-DV 2008 A202

Improving an empirical formula for the absorption of sound in the sea

Date	May 2008
Author(s)	C.A.M van Moll, MSc Dr M.A. Ainslie J. Janmaat, BSc
Classification report	Ongerubriceerd
Classified by	LTZ1 drs. R.P.A. Dekeling
Classification date	20-05-2008
Title	Ongerubriceerd
Managementuittreksel	Ongerubriceerd
Abstract	Ongerubriceerd
Report text	Ongerubriceerd
Appendices	Ongerubriceerd
Copy no	10
No. of copies	19
Number of pages	63 (incl. appendices, excl. RDP & distribution list)
Number of appendices	3

The classification designation Ongerubriceerd is equivalent to Unclassified, Stg. Confidentieel is equivalent to Confidential and Stg. Geheim is equivalent to Secret.

Improving an empirical formula for the absorption of sound in the sea

Laag frequent geluid plant zich in de zee over grote afstanden voort. Een model hiervan dient de absorptie van dat geluid mee te nemen. De standaard absorptieformule van François en Garrison is erg complex. Het doel is deze formule te vervangen door een eenvoudigere, zonder verlies van nauwkeurigheid. Het onderzoek dient ook het inzicht in de absorptiemechanismen en de ervaring met inversiemethoden te vergroten.

William of Ockham (1285-1349):
"plurality should not be posited without necessity"

Probleemstelling

Voor de Koninklijke Marine (KM) komt de dreiging met name uit het water. Voor de waarneming van mijnen en onderzeeboten is kennis van onderwaterakoestiek essentieel. De KM rekent hiervoor op ondersteuning door TNO Defensie en Veiligheid te Den Haag. Het onderzoek van dit rapport richt zich op het bevorderen van de kennis van TNO van akoestische propagatie over grote afstanden, die relevant is voor laagfrequente actieve sonar systemen. Hierbij is de absorptie van geluid erg belangrijk. Kennis van deze absorptie heeft zich verdicht tot absorptieformules, waarvan de standaard formule tamelijk complex is. Met het oog op kennis-bevordering is gezocht naar een vereenvoudiging van deze formule zonder dat dit ten koste van de nauwkeurigheid gaat.

Beschrijving van de werkzaamheden

De kwaliteit van een empirische formule is afhankelijk van de gemeten data. Hiervoor is uitgegaan van de data die door François en Garrison bijeen zijn gebracht in twee artikelen. Bovendien zijn nu ook absorptiemetingen uit de Baltische Zee gebruikt, die veel informatie toevoegen. Van de vele wiskundige constructies die mogelijk zijn, is de niet-lineaire formule van Ainslie-McColm gegeneraliseerd, zodat deze afgestemd kan worden op de data. Omdat hierbij tegelijkertijd tien parameters worden aangepast, is een automatisch zoekalgoritme gebruikt. De formule die het dichtste bij de gemeten absorptiewaarden komt, is gevonden. Tegelijkertijd is de onzekerheid van deze formule onderzocht door na te gaan hoe de beste parameter-vector verandert als slechts een deel van de datapunten gebruikt zou worden.

Resultaten en conclusies

Het onderzoek laat zien dat 'de beste' formule niet bestaat, omdat deze afhankelijk is van de gebruikte data en van de manier waarop de afstand tussen formule en data wordt gemeten. Wel is een formule gevonden die aanmerkelijk eenvoudiger is dan die van François en Garrison (25% minder rekentijd) en daardoor bijdraagt aan overzicht en inzicht. Deze eenvoud is niet ten koste van de nauwkeurigheid gegaan. Deze is zelfs iets verbeterd, wat met een statistische test is aangetoond. Deze formule is ook van toepassing op omstandigheden als in de Baltische zee, waar het zoutgehalte erg laag is.

Toepasbaarheid

Een eenvoudige formule verkleint de kans op fouten en bevordert het overzicht binnen ingewikkelde akoestische propagatie-modellen. De verbeterde formule kan gebruikt worden bij updates van het operationele model ALMOST, binnen het

onderzoeksprogramma voor RUMBLE2 en bij toepassingen als bodemclassificatie. Verwacht kan worden dat publicatie tot gebruik ervan zal leiden in plaats van de standaardformule.

PROGRAMMA	PROJECT
Programmabegeleider LTZ1 drs. R. Dekeling, DMO	Projectbegeleider LTZ1 drs. R. Dekeling, DMO
Programmaleider dr. J.C. Sabel, TNO Defensie en Veiligheid	Projectleider M. Colin, TNO Defensie en Veiligheid
Programmatitel Sonar en Onderwater propagatie	Projecttitel Advanced acoustic modelling
Programmanummer V512	Projectnummer 032.11648
Programmaplanning Start 01-05-2005 Gereed 31-12-2008	Projectplanning Start 01-08-2007 Gereed 31-12-2008
Frequentie van overleg Met de programma/project-begeleider werd 2 maal gesproken over de invulling en de voortgang van het onderzoek.	Projectteam ir. C.A.M van Moll dr. M.A. Ainslie ing. J. Janmaat dr. R. Vossen

Contact en rapportinformatie

Oude Waalsdorperweg 63
Postbus 96864
2509 JG Den Haag

T +31 70 374 00 00
F +31 70 328 09 61

info-DenV@tno.nl

TNO-rapportnummer
TNO-DV 2008 A202

Opdrachtnummer
-

Datum
mei 2008

Auteur(s)
ir. C.A.M van Moll
dr. M.A. Ainslie
ing. J. Janmaat

Rubricering rapport
Ongerubriceerd

Contents

Appendices

1 Introduction

The absorption of compressional waves in water depends mainly on frequency. It also depends on temperature, salinity, acidity and pressure. Empirical formulae are derived to express this relation. The accuracy of such a formula is important, mainly for sonar performance modelling, but also for seabed classification and the estimation of fish abundance. In their papers [1] [2] François and Garrison present such an empirical formula, given in Appendix A. Ainslie and McColm have derived a simpler formula [3] by approximating that of François and Garrison. It is our objective to tune the formula of Ainslie and McColm to in situ data in a sophisticated way. We want to derive an empirical formula for the absorption results whose *simplicity* and *accuracy* exceeds that of the formula of François and Garrison. Since this is a matter of tuning a model to measured data, inverse theory will be used for this. The parameters of the formula (the model) are tuned such that the samples of a dataset are fitted best. A cost function is defined to quantify this fit, automated search is applied and issues of sensitivity and uncertainty are addressed.

One should be aware that it is the increased computer power that allows us to apply a different approach, which was not available to François and Garrison twenty five years ago. Their major effort of collecting the measurements still deserves appreciation.

Overview
After a description of the data processing in Chapter 2, three cost functions are defined in Chapter 3. The conclusion is in Chapter 8. Chapters 4, 5 and 6 present the results of several approaches to the search for the best formula and are described hereafter in more detail. Chapter 7 applies statistical tests.

Chapter 4
At the start of this investigation it was assumed that the formula of Ainslie-McColm is already accurate. If the optimal tuning of the formula can be expected to be in the vicinity of their setting, a local search suffices to find the optimum. Chapter 4 presents this approach.

The results depend on the data set that is used. Since the Baltic Sea has a very low salinity, Baltic data differ considerably from the rest. Therefore two cases are distinguished: one wherein all available data, including the Baltic, are used in the search and a second case that only uses all non-Baltic data. As a consequence the local search comes up with two different 'best' parameter vectors.

The absorption formula consists of three components, with the fresh water absorption as one of them. During local search, this part was tuned too. Since fresh water absorption can be measured in a laboratory tank, it is already accurately known. However, the derived values of the local search deviate too much from these known values.
This could have been expected; absorption measurements at sea are not suited for the tuning of the fresh water absorption part. Therefore, from Chapter 5 onward, it is decided not to search for parameters of the fresh water absorption part of the formula any more, reducing the number of parameters to search for from 13 to 10. In the chapters 5 to 7 we accept the fresh water absorption part of the François-Garrison formula and don't use the simpler approximate fresh water part from Ainsly-McColm. This results in a hybrid model that needs to be tuned.

Chapter 5
To test the assumption that the tuning of Ainslie-McColm is already nearly optimal, a global search has been applied. It is described in Chapter 5. The resulting best parameter vector differed considerably from the local search result. More important, however, is the sensitivity information that comes easily with the global search results. The cost function, that quantifies the distance between the calculated and the measured absorption values, showed to be very insensitive to changes of some parameters. Without the Baltic data, parameters coupled to salinity, for instance, are very insensitive and therefore poorly determined.

Chapter 6
The question remains to what extent the derived results depend on the dataset that is used. Instead of looking for extra measured data, we choose a pragmatic solution by considering what happens if we invert a random *subset* of the available data.
In Chapter 6 random subsets of the data are taken and on each subset a global search is applied. The variation of the best parameter vectors of these search runs provides information about the uncertainty of the tuning parameters. The variation of the minimum cost value per subset gives a feeling for the significance of variations of the cost. Facing this uncertainty and significance, a single tuning for the Ainslie-McColm formula, combined with the François-Garrison fresh water part, is chosen.

Chapter 7
The impressionist approach of uncertainty gets a more thorough basis in Chapter 7, where statistics is applied on the chosen formula. For the given dataset the simple formula is proven to be more accurate than that of François and Garrison. However, the resulting errors are not normally distributed, underscoring the need for new and better empirical absorption data.

2 Data processing

The derivation of an empirical formula for the absorption of acoustic power in water depends on the availability of a proper data set. As a start, we use data provided by François and Garrison [1] [2] and Schneider [4], trying to find better and more recent data later. The selection of these data is given here. The derived data are superficially compared with calculated absorption values from different formulae.

2.1 Selection of usable data from several papers

The data that are used come from the papers of François and Garrison [1] [2] and that of Schneider [4]. François and Garrison have collected and processed measurements given in other papers, Schneider has done measurements in the Baltic Sea. The details of these in situ measurements are given in the papers or in the papers to which they refer, but the basic principle is as follows. Given the distance between the source and the receiver, the propagation loss is calculated as if there were no attenuation. The difference with the measured propagation loss is attributed to attenuation. Uncertainties result from a multitude of sources. Variation in source levels, calibration errors of the receiver, variation of electrical current to the devices, disturbing noise, variations in the distance, small errors in the measurement of frequency, salinity, temperature, acidity and depth. For the Schneider data we have estimated the error. For the papers of François and Garrison we assume that the numbers provided by ±... present the standard deviation of the measured absorption value.

The first step to get data was to digitize measurement values given in the papers of François and Garrison [1] [2]. Although the pdf-files of these papers are essentially scanned images of the original hard copy versions, it was possible to copy-paste the values from the tables using text recognition software supplied with Acrobat Reader. The conversion was not perfect, but it was faster than copying the data manually.

Since we require that with the measurements also a measure of uncertainty is provided, we selected the following data. From [1] Table I (Bezdek), Table II (APL) and Table IV (Greene, but not Schulkin and Marsh). From [2] we use the tables I and II.

For [1] the following pH values have been added: 8 for the Atlantic and 7.7 for the Pacific. For the arctic region, north of the Bering Strait (Chukchi Sea and Beaufort Sea) a value of 8.0 is chosen. They are taken from [6]. In Table II of [1] we have used the 'Adjusted α' numbers if present, otherwise the 'Measured α' values are taken.
The measured absorption values for Dabob Bay have been modified too, by dividing the Uncorrected measurements of Table I and II of Murphy [5], using a more precise transformation from dB/kyd to dB/km (division by 9.144 instead of 9.1). The sample of 10 May 1956 is removed, because it has zero salinity, which is unlikely. The Bering Sea sample of 2 Apr 1973 is not removed, although it deviates very much from the Calculated value of François and Garrison.

Data from the Baltic Sea come from Schneider [4]. Although they don't have a specified error, it was possible to estimate it, because four measurements are given per combination of variables[1]. Per setting for the measured attenuation values (in dB/km) the average and standard deviation is calculated. Four summer data points are removed, viz. point 4, 5, 6 and 7 of Figure 4 from [4], since the high values are attributed to resonance from fish. These data are very valuable, because the Baltic Sea has a very low salinity and provides an exceptional situation in this respect. Therefore we distinguish the situation that the Baltic data are used together with the other available data ('inc Baltic'), from that where the Baltic data are excluded form the available data ('exc Baltic').

The data have been placed in the single tab-delimited text file "alldata_prepared2.txt". They are given in Appendix C. Listed are location, investigator and year, followed by depth [m], range [km], sound speed [m/s], temperature [°C], salinity [ppt], pH value, frequency [kHz], measured alpha [dB/km] and the accompanying error. In some cases, no value was given for a quantity; in those cases, a 'NaN' has been filled in. The data file alldata_prepared2.txt can be read easily using the Matlab script readtables.m.

2.2 Measured versus calculated absorption

As an example the low-frequency *measurements* (Figure 1) in the Mediterranean Sea by Skretting and Leroy (Table I in [2]) are plotted against frequency, with error bars. In the same Figure some *calculated* absorption values for the corresponding circumstances (see caption) are given. As can be expected, the calculated values of Skretting and Leroy match the measurements best.

[1] Maybe we have used a biased estimator of the variance, taking 1/N instead of 1/(N-1). Since we don't expect this error to be of main importance and correcting it takes very much time, we don't correct this.

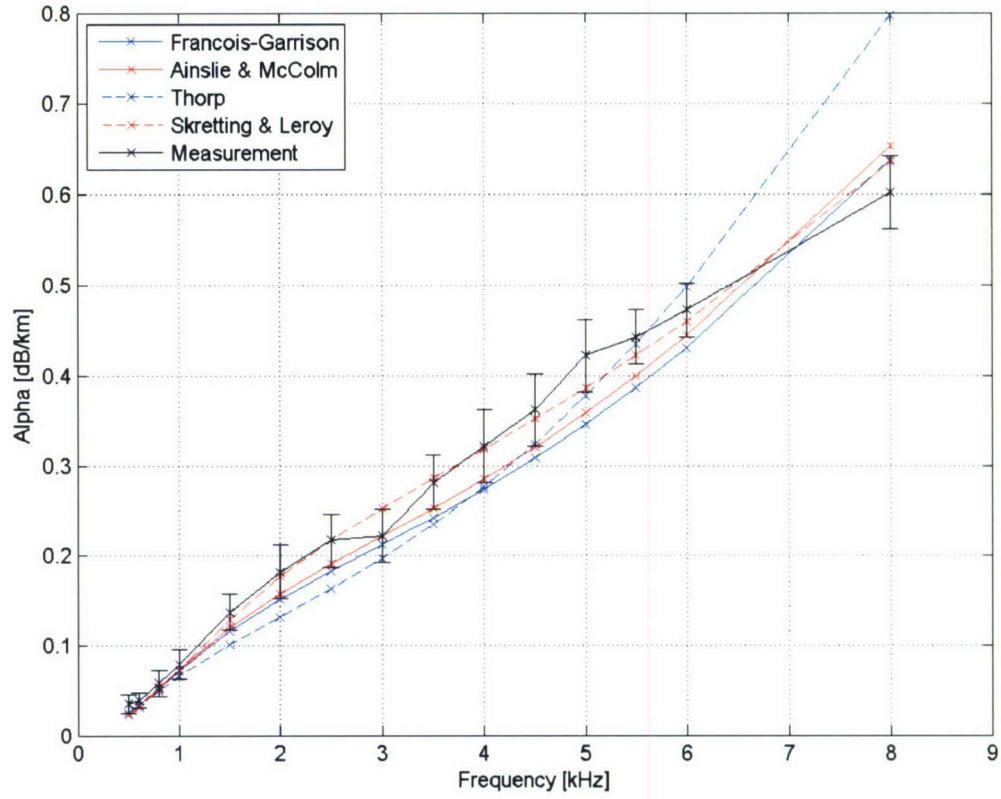


Figure 1 Absorption Alpha [dB/km] in the Mediterranean at low frequencies.
T = 13°C, S = 38.0 ppt, pH = 8.15, c = 1517 m/s, depth = 800 m2.

A sample of higher frequency measurements together with computed absorption values are shown in Figure 2. The data come from Table I in [1] and are the Pacific Ocean shallow measurements by Bezdek. Above 80 kHz the discrepancy between the calculated and measured absorption values is very big.

[2] There is an ambiguity in the formula of Skretting and Leroy. Commonly a factor $0.007f^2$ is used, but in their original paper this factor was $0.006f^2$. The latter is used here.

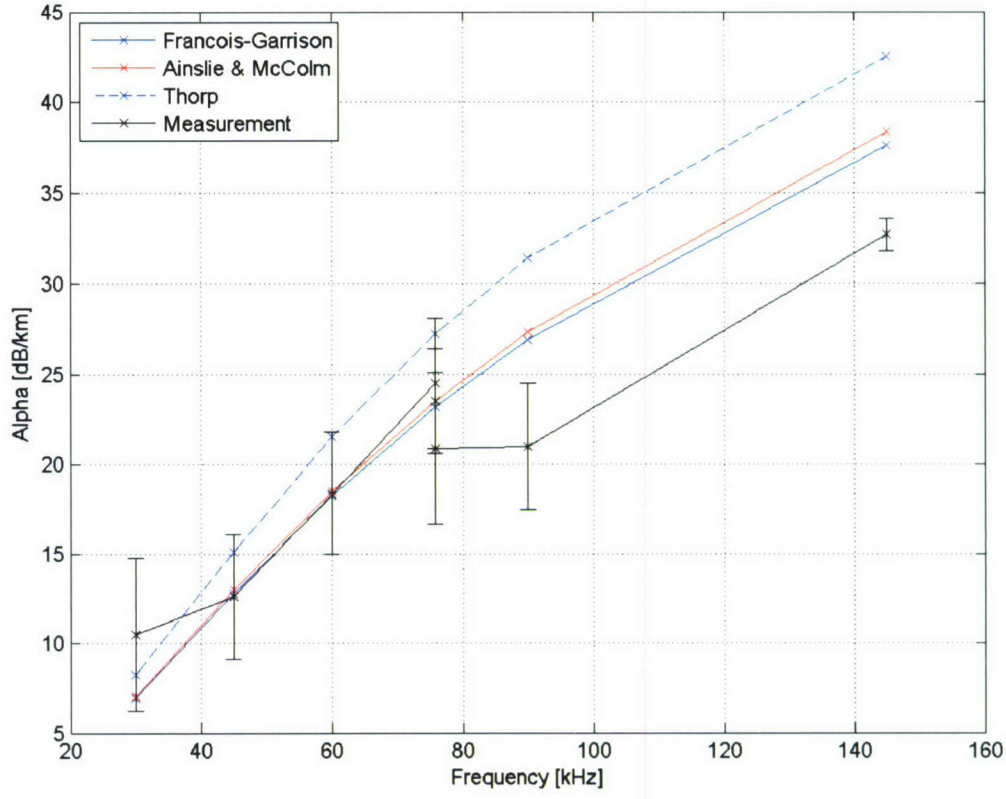


Figure 2 Absorption Alpha in the Pacific Ocean at high frequencies.
T = 7.0°C, S = 34.0 ppt, assumed[3] pH = 8.08, depth = 200 m.

For various areas with different pH values in the measurements taken from Table 1 in [2], root mean square (RMS) errors for different models with respect to the measurements have been calculated, and are shown in Figure 3. The data come from the North-East Pacific (pH = 7.69) measured by Chow and Turner, the Atlantic (pH = 8.03) measured by Thorp, the Mediterranean Sea (pH = 8.15) measured by Skretting and Leroy, the Red sea (pH = 8.18) by Browning and the Gulf of Aden (pH = 7.72) also by Browning.

[3] Approximation to an interpolated value of the Pacific acidity. N. Pacific data from Mellen et. al. 1987, p 44-48 referred to by Ainslie, Table 4; depth 0 m, pH = 8.23 and depth 500 m, pH = 7.70.

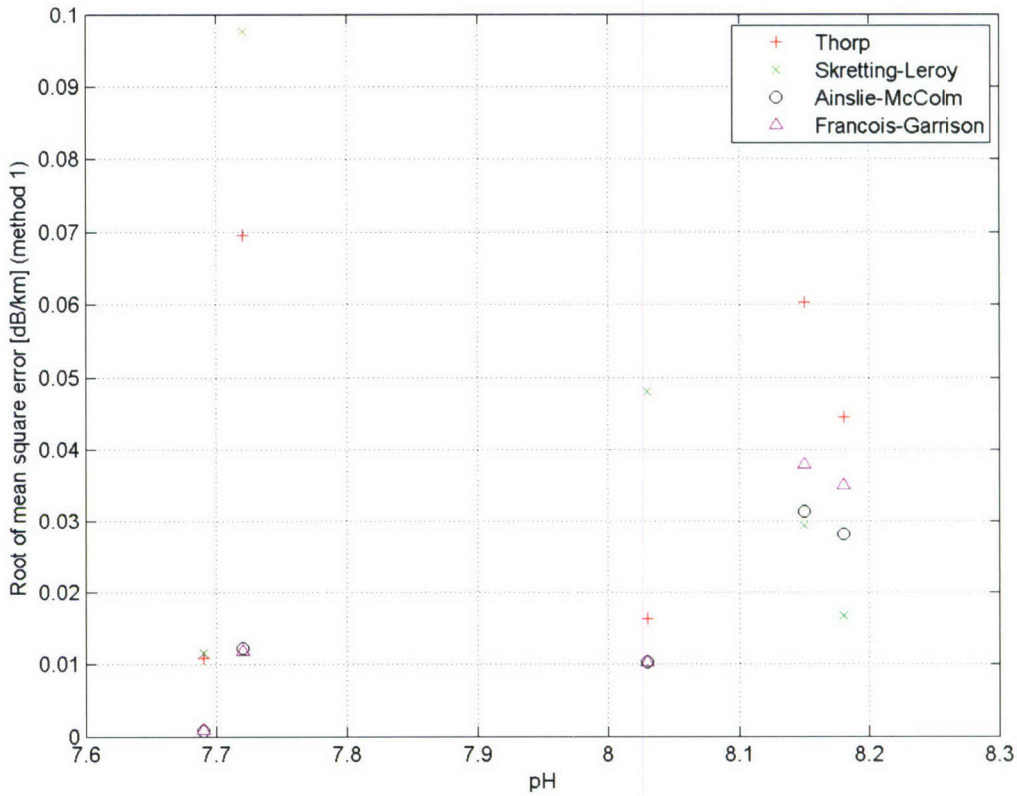


Figure 3 RMS errors ε_1 of the different models in various pH regions.

Each RMS value has been determined using the following distance measure ε_1, with α the calculated and β the measured absorption and N the number of samples.

$$\varepsilon_1 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\alpha_i - \beta_i)^2}$$

Figure 4 shows the result of using the following definition of the fractional mean square error ε_2 as a measure of the distance of the models from the measurements.

$$\varepsilon_2 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i - \beta_i}{\beta_i}\right)^2}$$

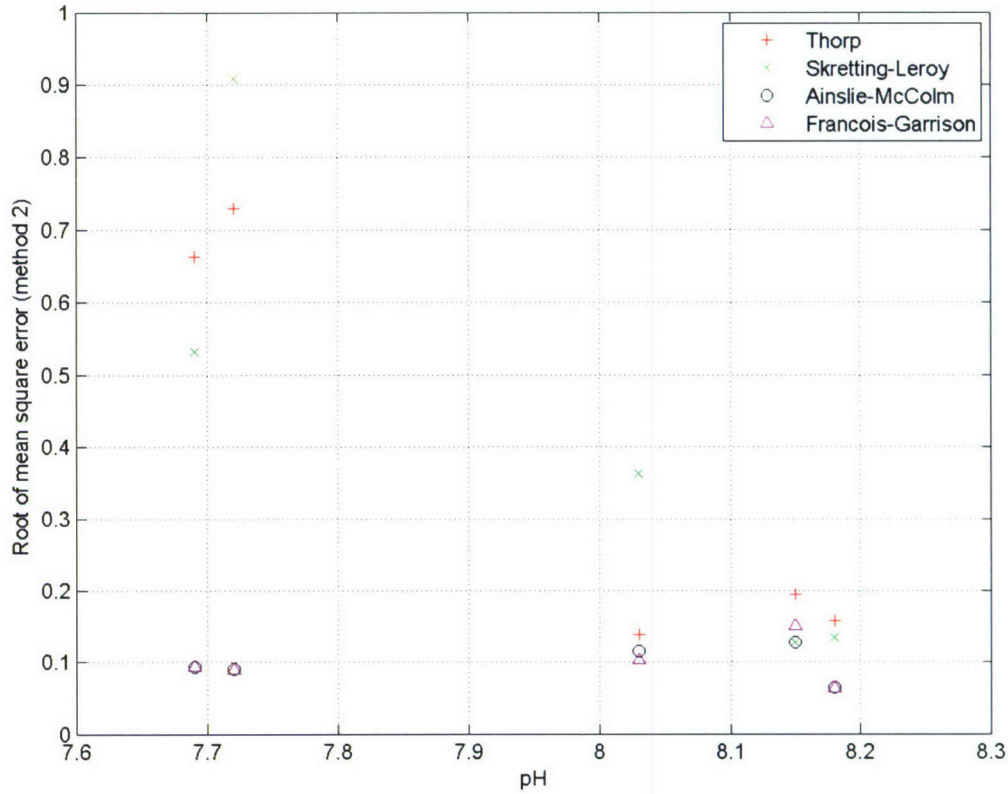


Figure 4 RMS errors ε_2 of the different models in various pH regions.

Comparison of these figures shows that the distance measure can exert considerable influence on the judgement of the quality of the fit of absorption formulae.
Other distance measures are presented in Section 3.2 about cost functions.

From these examples follows that it is impossible to analyse the performance of a formula for each combination of frequency, temperature, salinity, etc. The more systematic approach of inverse theory is needed.

3 Inverse theory

A specific model provides the absorption, given the values of the independent variables (frequency, salinity, etc). In inverse theory the measured absorption is given together with the settings of these variables, and the best tuning of the model is searched for. We are looking for the best coefficients of the Ainslie-McColm absorption formula and use the apparatus of inverse theory for this.

Presentation of the model is followed by a discussion of cost functions. The two main search methods are presented thereafter.

3.1 Model

The formula of Ainslie-McColm [3] - with independent variables frequency f [kHz], salinity S [$^o/_{oo}$] , temperature T [°C], pH and depth z [km] - is composed of three parts; the boric acid contribution α_1 , the magnesium sulphate contribution α_2 and the fresh water absorption α_3 .

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3 \quad \text{dB/km}$$

Boric acid contribution α_1 :

$$\alpha_1 = A_1 \frac{f_1 f^2}{f_1^2 + f^2} e^{\frac{pH-8}{P_1}} \qquad f_1 = F_1 \left(\frac{S}{35}\right)^{S_1} e^{\frac{T}{T_1}}.$$

Magnesium sulphate contribution α_2 :

$$\alpha_2 = A_2 \left(1 + \frac{T}{\theta_2}\right)\left(\frac{S}{35}\right)\left(\frac{f_2 f^2}{f_2^2 + f^2}\right) e^{-\frac{z}{Z_2}} \qquad f_2 = F_2 e^{\frac{T}{T_2}}.$$

Fresh water absorption α_3 :

$$\alpha_3 = A_3 f^2 e^{-\left(\frac{T}{T_3} + \frac{z}{Z_3}\right)}.$$

The 13 parameters F_1, S_1, T_1, F_2, T_2, A_1, P_1, A_2, θ_2, Z_2, A_3, T_3, and Z_3, are introduced here to tune this model. In their paper [3] this tuning was not an issue and the following values were inserted. We will call these values the 'Ainslie-McColm' parameter setting.

Table 1 Ainslie-McColm parameter setting.

F_1	S_1	T_1	F_2	T_2	A_1	P_1	A_2	θ_2	Z_2	A_3	T_3	Z_3
0.78	0.5	26	42	17	0.106	0.56	0.52	43	6	0.00049	27	17

It will be explained later that a second 'Hybrid model' is needed, that combines the first two parts of the fomula of Ainslie-McColm with the fresh water absorption part of François-Garrison. This hybrid model has the following fresh water absorption α_3, that will not be tuned.

$$\alpha_3 = A_3 P_3 f^2$$

$$A_3 = 4.937\ 10^{-4} - 2.59\ 10^{-5} T + 9.11\ 10^{-7} T^2 - 1.50\ 10^{-8} T^3 \quad \text{for} \quad T \leq 20^o C$$

$$A_3 = 3.964\ 10^{-4} - 1.146\ 10^{-5} T + 1.45\ 10^{-7} T^2 - 6.5\ 10^{-10} T^3 \quad \text{for} \quad T > 20^o C$$

$$P_3 = 1 - 3.83\ 10^{-2}\ z + 4.9\ 10^{-4}\ z^2$$

Although this formula is much more complicated than the Ainslie-McColm fresh water part, we assume that it is more accurate since it is derived from measurements in tanks. To distinguish both models, we call the formula with the Ainslie-McColm fresh water approximation the 'Full Ainslie-McColm model'.

3.2 Cost functions

The RMS errors ε_1 and ε_2 presented in the previous section quantify the distance between the measured and the calculated absorption and therefore can be chosen as cost functions. Careful consideration, however, leads us to use other definitions of the cost function. We define them by means of the following symbols.
α_i = the measured absorption of sample i (i = 1, ..., N),
σ_i = the standard deviation of the measured absorption provided with sample i and
$\alpha(\underline{v}_i,\underline{p})$ = the modelled absorption for the setting $\underline{v}_i$ of the independent variables of sample i, that depends on the model parameter vector $\underline{p}$.

The first cost function is the 'fractional mean error' (notation: L_{1F}-cost) and is defined as follows:

$$\text{cost}_{\text{L1F}}(\underline{p}) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\alpha(\underline{v}_i,\underline{p}) - \alpha_i}{\alpha_i}\right| .$$

The second cost function is called the 'normalised mean error' (notation: L_{1N}-cost) with definition

$$\text{cost}_{\text{L1N}}(\underline{p}) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\alpha(\underline{v}_i,\underline{p}) - \alpha_i}{\sigma_i}\right| .$$

The third cost function we consider is named the 'RMS normalised error' (notation: L_2-cost), which is defined as:

$$\mathrm{cost}_{L2}(\underline{p}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha(\underline{v}_i, \underline{p}) - \alpha_i}{\sigma_i}\right)^2} \quad .$$

The fractional mean error considers the error as a fraction of the measured value. The normalised mean error expresses the distance between the measured and calculated absorption values relative to the provided error of the sample. The RMS normalised error squares this normalised error to weight larger errors extra, whereas the normalised mean error gives equal weighting to each sample. If the measured data contains outliers, their influence is larger in the L_2- than in the L_{1N}-cost. The L_2-cost resembles the Mahalanobis distance. These cost functions will be applied on two different models; the Full Ainslie-McColm model and the Hybrid model.

3.3 Search methods

With 13 parameters to tune the full Ainslie-McColm model, the search space is huge. As a result the search process is automated by means of an algorithm. The main choice is between using a local search or a global search method. The local search method assumes that a parameter vector is given that is already in the vicinity of the best vector. It follows the gradient of the cost function so that the minimum is reached as fast as possible. A global search method starts without the assumed a priori information and samples the search space in such a way that it finds the global minimum after evaluating a limited but often huge number of parameter vectors.

As local search method we chose Downhill Simplex, which is available as the built-in Matlab script 'fminsearch.m'. Mainly because we are interested in the sensitivity of the parameters, we also decided to apply a global search. Differential Evolution is the algorithm used for this. The global search provides so much more information, that we changed our approach and used global search as the main instrument for our investigation from Section 5 onward. First we present the local search results.

4 Local search on the Full Ainslie-McColm model

The local search is started in the 'Ainslie-McColm' parameter vector. In this section we present the results of this search. A brief discussion follows thereafter. This discussion leads to two main improvements of the method. An overview of the Matlab m-files used is given in Appendix B.

4.1 Method

Downhill Simplex is a local search method that tries to find the vector with minimum cost as fast as possible. Starting in a point of the cost landscape, it explores the direction wherein the cost derivative is most negative. It follows the steepest downhill path. The disadvantage of such an algorithm is that if a local minimum is found, the method is not able to escape from it, possibly missing the global minimum. The state of the algorithm is specified by a single vector, in contrast with global search methods, whose state is often given by a population of vectors.

4.2 Numerical values

The best parameter vector depends on the data set used and on the cost function. For the data set we distinguish using a data set that contains all available data and the option of not using the Baltic data. The search can be done by means of the L_{1F}-, L_{1N}- or the L_2-cost function. This would lead to six 'best' vectors. Because this section mainly serves to show the flaws of this approach, we only apply the L_{1F}- and L_2-cost function. At the next stage of our investigation we exchange the L_{1F}- for the L_{1N}-cost function. The combination of two datasets with two cost functions provides us here with four 'best' vectors.

Table 2 Results of local search with 1000 function evaluations, starting at the 'Ainslie-McColm' setting. Values in brackets give percentage change relative to 'Ainslie-McColm'.

	inc Baltic			exc Baltic		
para-meter	Ainslie-McColm	best vector with L_{1F}-cost,	best vector with L_2-cost,	Ainslie-McColm	best vector with L_{1F}-cost,	best vector with L_2-cost,
F_1	0.78	0.86 (+10)	0.98 (+26)	0.78	0.88 (+13)	0.88 (+13)
S_1	0.5	0.46 (–7)	0.56 (+11)	0.5	0.6 (+20)	0.12 (–76)
T_1	26	30.5 (+17)	40.9 (+57)	26	30.1 (+16)	29.6 (+14)
F_2	42	48.8 (+16)	48.1 (+14)	42	51.0 (+21)	50.1 (+19)
T_2	17	20.6 (+21)	20.5 (+21)	17	21.1 (+24)	22.3 (+31)
A_1	0.106	0.108 (+1)	0.102 (–4)	0.106	0.109 (+3)	0.101 (–5)
P_1	0.56	0.57 (+1)	0.58 (+3)	0.56	0.58 (+4)	0.58 (+4)
A_2	0.52	0.51 (–2)	0.56 (+7)	0.52	0.51 (–2)	0.56 (+8)
θ_2	43	41.0 (–5)	77.9 (+81)	43	39.1 (–9)	86.8 (+102)
Z_2	6	4.8 (–20)	3.8 (–37)	6	4.5 (–25)	4.0 (–33)
A_3	0.00049	0.00048(–2)	0.00047(–4)	0.00049	0.00045(–8)	0.00046(–6)
T_3	27	22.1 (–18)	25.9 (–4)	27	23.5 (–13)	27.9 (+3)
Z_3	17	16.0 (–6)	5.0 (–70)	17	8.5 (–50)	6.6 (–61)
L_{1F}-cost	0.11436	0.10602	0.1155	0.11226	0.10244	0.11257
L_2-cost	2.2235	2.0995	1.8506	2.2892	2.1557	1.8651

To prevent misunderstanding, 'inc Baltic' means that the Baltic data are taken into account during the search and also for the calculation of the presented cost value. In the same way the Baltic data are left out during the search and are left out of the presented cost value for the 'exc Baltic' cases. The consequence of this is that the costs of 'inc Baltic' and 'exc Baltic' columns can not be compared, they will already differ only because they use different data sets.

4.3 Discussion

These results demonstrate that the best vector depends on the cost function used during the search (L_{1F}- or L_2-cost) and on the data set that is used (including or excluding the Baltic data). Comparison of the costs of the search results with that of the Ainslie-McColm parameter setting shows that a better choice of the parameters is possible. This means that an improvement of the accuracy of the Ainslie-McColm absorption formula can be derived.

The cost values that result are unexpectedly high. An L_{1F}-cost of 0.11 means an average fractional error of 11% of the measured value. This is much more than the claimed value of 5%. In the same way an L_2-cost of 1.9 means a root mean square error of the normalised deviation of the measurements from the model of 1.9. For normally distributed 'measurements errors' this would be considered a big Mahalanobis distance. This raises doubts about the appropriateness of the model or the quality of the dataset. Both issues will not be addressed here.

We compare the accuracy of the Ainslie-McColm formula with that of François-Garrison by means of their cost values.

Table 3 Comparison of costs.

	Ainslie-McColm inc Baltic	Ainslie-McColm exc Baltic	François-Garrison inc Baltic	François-Garrison exc Baltic
L_{1F}-cost	0.11436	0.11226	0.1157	0.11268
L_2-cost	2.2235	2.2892	2.1805	2.2305

These cost values suggest that the accuracy of the Original Ainslie-McColm formula is approximately the same as that of François-Garrison. However, costs are only auxiliary variables; the objective of the search are the derived parameter values and they deserve more attention.

The fresh water contribution in the François-Garrison formula is assumed accurately known, because it has been established in a laboratory tank. The original values for A_3, T_3 and Z_3 of Ainslie-McColm provides an approximation for this part of the absorption. However, the just derived values of A_3, T_3 and Z_3 deviate too much from these original values. It was therefore a mistake to search for these three parameters. We decide to stop improving the approximate fresh water part of the Ainslie-McColm formula and accept the more complicated fresh water absorption formula of François-Garrison as the most accurate.

This means that we stop using the Full Ainslie-McColm model and turn over to the Hybrid model. The parameters A_3, T_3 and Z_3 are removed from the search, reducing the search space to 10-dimensions instead of 13.

What is also missing in the results is how sensitive the costs are to small deviations of a parameter from its 'best' value. Marginal sensitivity is the effect of a small variation of a single parameter on the cost of the parameter vector. If such a change of a parameter has no effect on the cost, little value should be given to the precise value of this parameter. To get this sensitivity information it is required to calculate the cost of many vectors that differ slightly from the minimum vector.

Global search automatically provides the costs of many vectors, since it explores at random the region of a local minimum much more thoroughly than the efficient steep downhill local search. Global search even has the extra advantage that no information about the approximate position of the minimum in the parameter space is required. Because of these arguments, we decide to restrict ourselves from now on to global search.

A disadvantage of the L_{1F}-cost is that is does not take the error provided with the measurements into account. For this reason it is replaced by the L_{1N}-cost. This will be our preferred choice. Just like the L_2-cost it takes the given error into account, but in contrast to the L_2-cost it weights each sample equally, irrespective of the size of the error of the sample.

5 Global search on the Hybrid model

This section starts with a brief description of the inversion runs. Then the results are presented by means of sensitivity plots, followed by the derived numerical values. We conclude with a discussion of the results. An overview of the Matlab files used is given in Appendix B.

5.1 Method

Global search is applied on the Hybrid model, using the data file alldata_prepared2.txt (given in Appendix C). Because there are only 10 parameters left to search for, the settings of the search method are set to a population size of 16 and 150 generations. Here is an overview of the various runs.

Table 4 Files containing the results of the global search; each file contains 200 runs.

	L_{1N}-cost normalised mean error	L_2-cost RMS normalised error
Baltic data **in**cluded	MathieusRuns_300108	MathieusRuns_310108
Baltic data **ex**cluded	MathieusRuns_320108	MathieusRuns_330108

5.2 Sensitivity plots

An advantage of global search is that it can provide information about the sensitivity of the parameters. This sensitivity is visualised by means of 'red dot' pictures, that requires some explanation. In these pictures all evaluated vectors during the global search process are considered, together with their cost values. For each parameter each vector is represented by a single *red* dot in a picture that shows the parameter value against the cost. A second much smaller set of vectors consists of all parameter vectors from the last generation of each run. They usually will have low costs and are presented by *green* dots. *Black* dots show the 200 best vectors of the global search runs (after a local search of at most 1000 function evaluations on the single best vector of the last generation of each run). Collecting all vectors from the 200 files of MathieusRuns_300108 the red dot pictures look as follows. Vectors with an L_{1N}-cost higher than 1.6 are not presented.

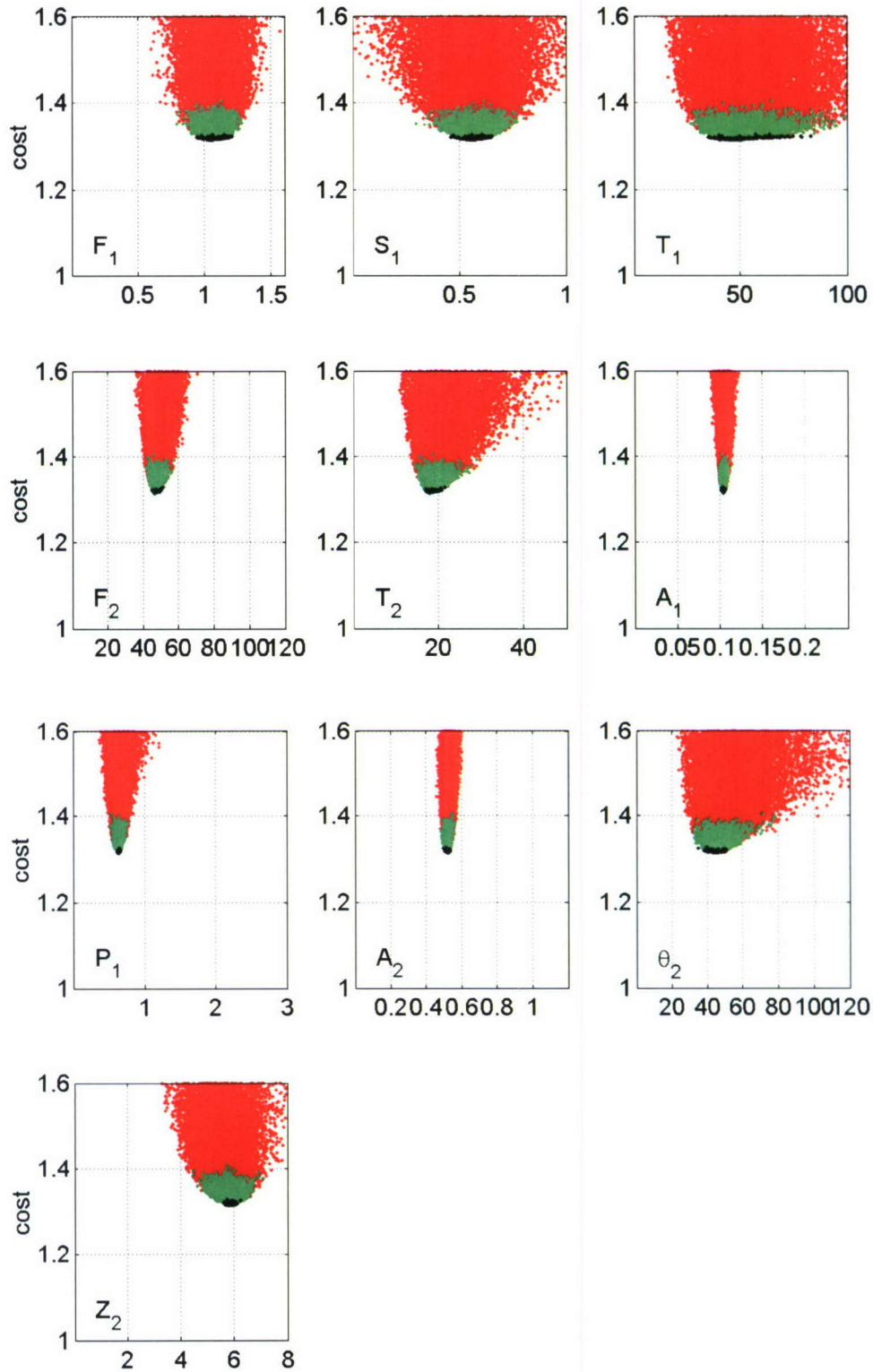


Figure 5 MathieusRuns_300108; Baltic data included, L_{1N}-cost.

Notice that the black dots are mostly very concentrated. The global search has found many parameters very precisely. There also is no sign of ambiguity of the solutions. Because the green dots represent vectors with very low costs too, they provide sensitivity information. It is seen that the cost is very insensitive to the parameter T_1.

The sensitivity plots for MathieusRuns_320108, that excludes the Baltic data, are as follows. Because the variation of the salinity in the measurements is very small without the Baltic data, the parameter S_1 can not be resolved.

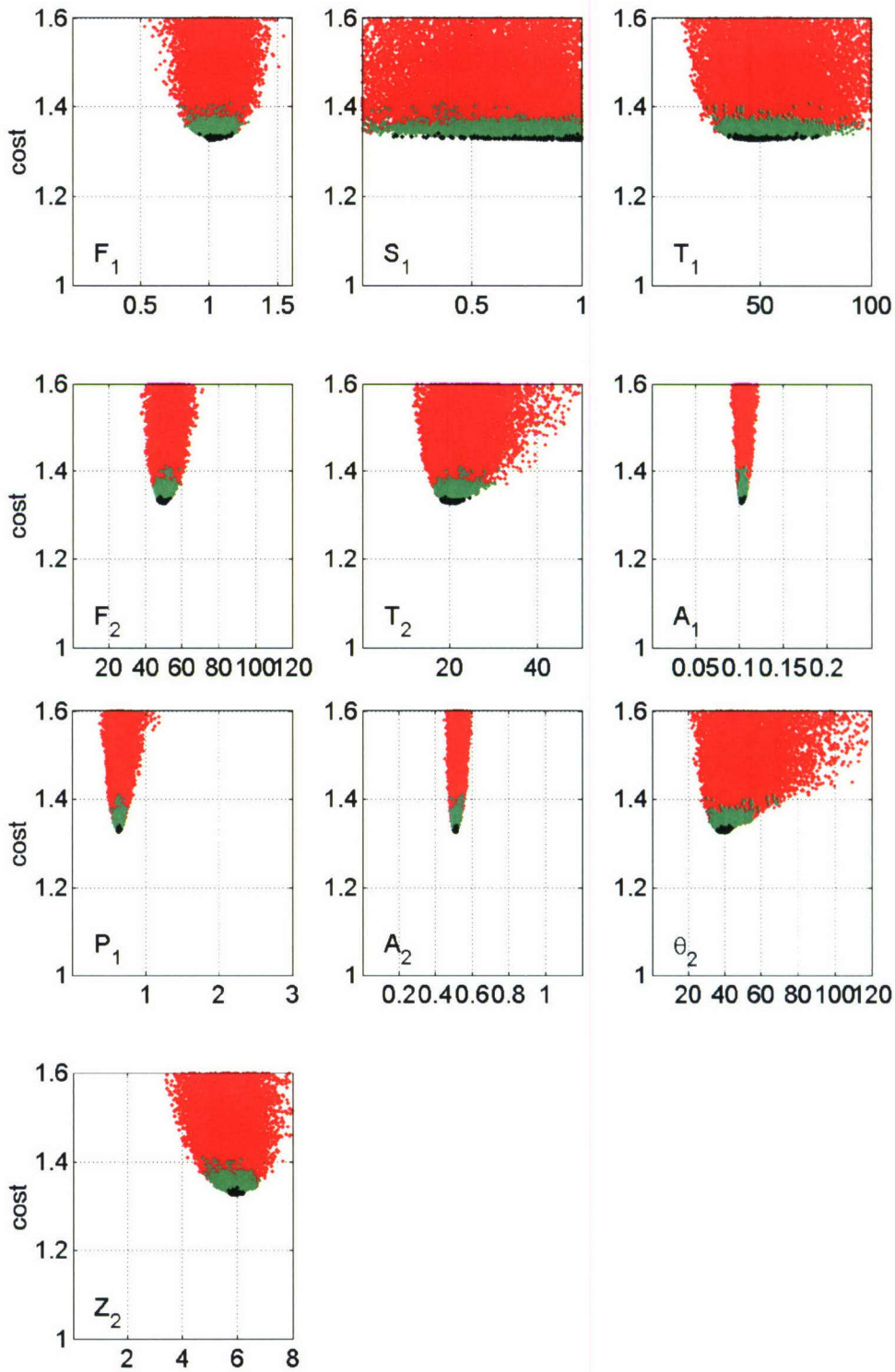


Figure 6 MathieusRuns_320108; Baltic data excluded, L_{1N}-cost.

The sensitivity of the parameters varies. The cost is very sensitive to the parameters A_1, P_1 and A_2 and very insensitive to S_1 and T_1. For the sensitive parameters a high precision can be expected.

Using the L_2-cost function during the inversion runs gives the following results.

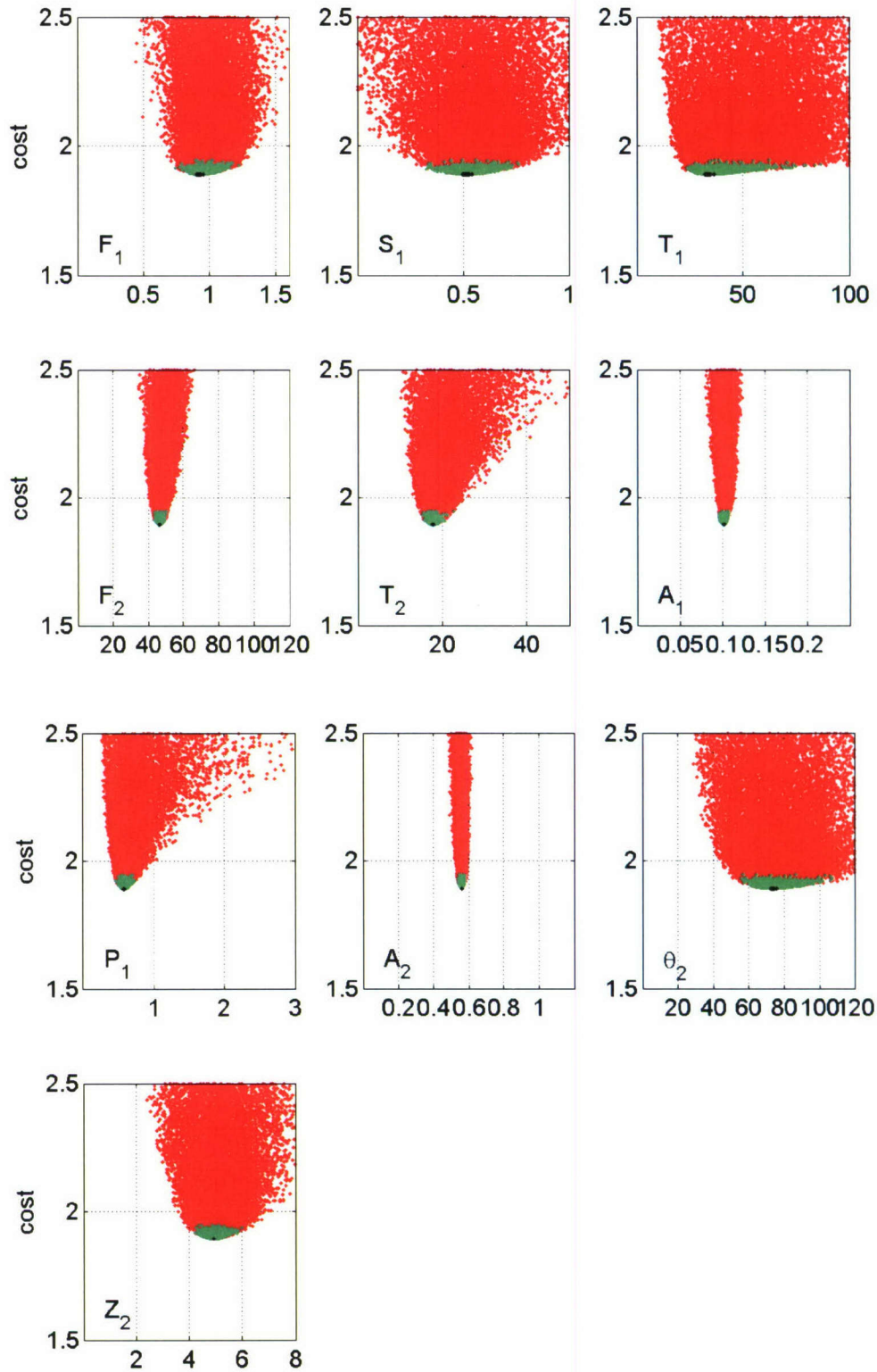


Figure 7 MathieusRuns_310108; Baltic data included, L_2-cost.

Notice that the scale of the vertical axis has changed. L_{1N} and L_2-cost values are incomparable. Besides, we have not discussed the significance of changes of the costs, an issue that will be addressed in Chapter 6. The scale of the axis is chosen such that approximately the same number of red dots as in Figure 5 are present.

In comparison with Figure 5, the sensitivity of the L_2-costs for the parameters S_1, T_1, P_1 and θ_2 have become smaller than was the case with L_{1N}-costs. Notice that the black dots are much more concentrated.

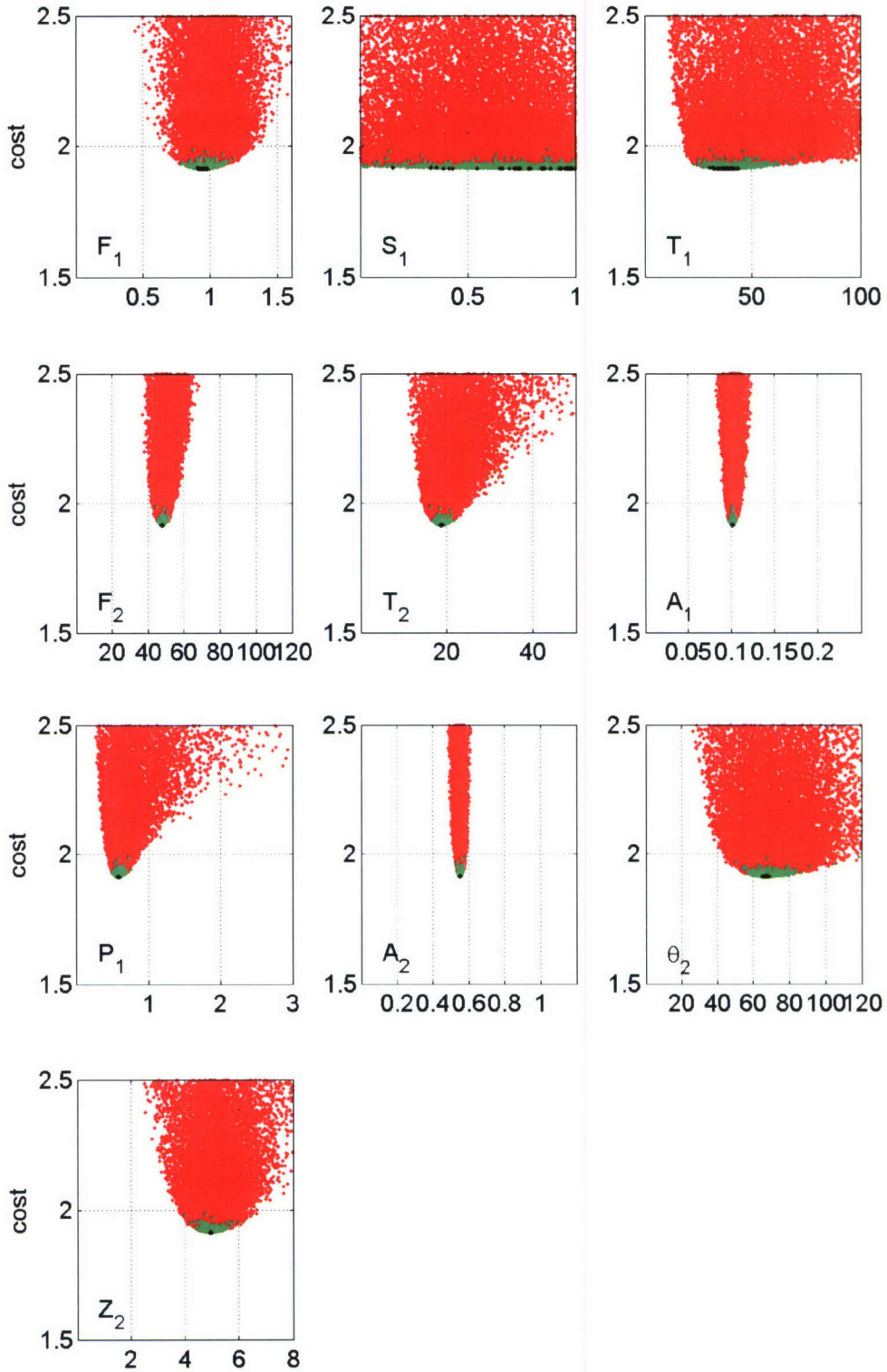


Figure 8 MathieusRuns_330108; Baltic data excluded, L_2-cost.

Comparison of Figure 8 with Figure 6 demonstrates that the sensitivity for S_1 and T_1 for inversions without Baltic data has deteriorated even further with L_2-costs.

5.3 Numerical values

For each parameter a search interval has to be chosen. We want to prevent that the optimal setting of a variable is found on the upper or lower bound (as still happens for S_1). After some trial and error the very high variable upper bounds (vub) given in the next table are used. The lower bound of the search interval of each parameter is taken the upper bound divided by 10000, which is effectively zero. The horizontal axes of the previous red dot pictures show the full search intervals. The numerical values of the best parameter vectors are presented in the table.

Table 5 Overview of parameter vectors and upper search boundaries.

par	variable upper bound	Ainslie-McColm	Best vector for data inc. Baltic; L_{1N}-cost	Best vector for data exc. Baltic; L_{1N}-cost	Best vector for data inc. Baltic; L_2-cost	Best vector for data exc. Baltic; L_2-cost
file nr	--	--	300108	320108	310108	330108
F_1	1.6	0.78	1.04	1.02	0.92	0.94
S_1	1.0	0.5	0.55	0.996	0.51	0.99997
T_1	100	26	46.4	50.6	33.3	39.1
F_2	120	42	46.6	49.3	46.4	47.9
T_2	50	17	17.9	20.1	17.7	18.7
A_1	0.25	0.106	0.104	0.103	0.101	0.101
P_1	3	0.56	0.62	0.62	0.57	0.58
A_2	1.2	0.52	0.52	0.51	0.56	0.55
θ_2	120	43	44.4	40.2	73.2	67.0
Z_2	8	6	5.8	5.8	4.9	4.9
L_{1N}-cost (inc. Baltic)	--	1.4851	1.3153	--	(1.3824)	--
L_{1N}-cost (exc. Baltic)	--	1.5304	--	1.3266	--	(1.3975)
L_2-cost (inc. Baltic)	--	2.2346	(1.9845)	--	1.8905	--
L_2-cost (exc. Baltic)	--	2.2988	--	(2.0153)	--	1.9122

The Ainslie-McColm parameter vector together with its cost is presented for comparison. The cost values given here belong to non-rounded parameter vectors. Since the vectors presented here are rounded, their costs will be slightly higher. This issue is addressed more thoroughly in Section 6.3.

5.4 Discussion

Now we have the best parameter values and know the sensitivity of each of them. Are we finished? No. This result relies fully on the supplied data set. Each time extra data come available, another parameter setting will become optimal. Our solution is not *robust*. What we want, is to derive a very good parameter setting that will not vary much when new data come available. Instead of waiting for these new data, we investigate the effect of leaving out a part of the available data. This gives an impression of the certainty or robustness of each of the parameters. Where *sensitivity* says something about the effect of a small change of a parameter for the cost (considering a single set of measured data), *uncertainty* of a parameter is caused by

variations of the measurements, considering multiple sets of data. These ideas must be distinguished.

In the next section the approach is adjusted a second time by varying the dataset to which the formula is fitted. Global search on the hybrid model will still be applied. From the previous Red Dot pictures (Figure 5 - 8) it is clear that the Baltic data are very important to derive sensitivity for the parameters S_1 and to a smaller extent T_1. Therefore we will consider results excluding Baltic data only briefly.

6 Robustness of global search

After a description of the way wherein the dataset is varied, the results of the global search on varying subsets are presented by means of sensitivity plots. Then a single parameter vector is selected as the best one, that represents the Improved Tuning formula. The improved accuracy is illustrated and the new formula is presented in its full glory.

6.1 Method

The *robust* global search is applied on the Hybrid model, but with *varying* data sets. For each inversion run we remove at random between 20 and 25% of the 166 samples of alldata_prepared2.txt, including the Baltic data. The resulting random subset contains at least 124, but usually 128 to 134 samples. The vectors of the inversion runs are collected to see the effects of the variation of the data set. Again the L_{1N}-cost and L_2-cost are used.

6.2 Sensitivity plots

The following red dot pictures show results of search runs on varying subsets.

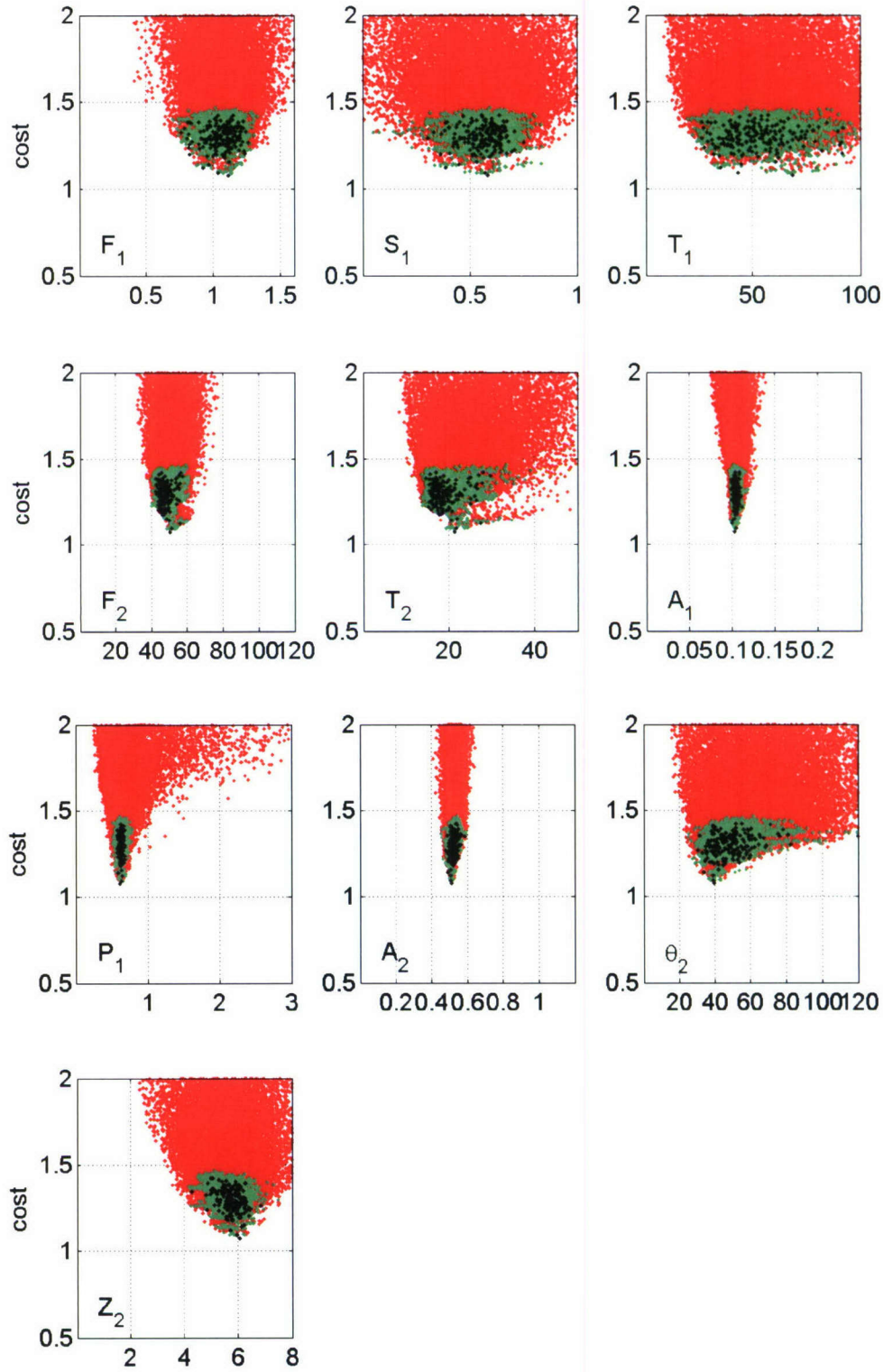


Figure 9 MathieusRuns_280108; Baltic data included, 200 runs, L_{1N}-cost.

Notice that lower cost values are derived now. This is easy to explain. Reduction of the set of measurements allows a better fit of the model. However, one must keep in mind that the cost values of two different runs are incomparable now. They give a measure of the distance of a single model to two different sets of data. The variation of the costs of

'good' vectors gives a feel for the *significance* of these variations. Looking at the costs of the best vectors only, it is clear that variations in the L_{1N}-costs of ca. 0.3 can result from slightly different subsets of the data. Smaller variations than 0.3 can not be considered to be significant. This means that all vectors below the minimum derived cost plus 0.3 could be considered to be as good. As a consequence it is better to concentrate on all vectors of the last generation (the green dots) than on the best vectors (the black dots). If we tune the hybrid model well, the L_{1N}-cost should be below 1.5, irrespective of the chosen subset.

A second observation is that there is no single model that is best for all datasets. For each subset of the data a best model is derived, but one such best model is only best for the particular dataset from which it comes. We must reduce our expectations and look for a model that is reasonably accurate for most datasets, in other words: which is *robust* to variation of the data set.

In the third place the results of these runs on random subsets of the data provide information about the *uncertainty* of the parameters. If we concentrate on the green dots of the vectors of the last generations, it can be seen there is very little uncertainty in for instance the parameter A_1, while T_1 is very uncertain.

Showing the red dots of all evaluated vectors is very useful to understand these pictures. However, the number of vectors is huge. Since we mainly want to know the spread of *very good and best* vectors, we have done 2000 runs, each one on a newly chosen random subset of the data. Showing all evaluated vectors of all these runs (red dots), would cause memory overload. Therefore hereafter only the vectors of the last generation (green dots) and the best vectors after local search (black dots) will be shown. After 2000 runs, the last generation and best vectors are stored in the file Summary_MathieuRuns_280108.mat, the files of the individual runs are deleted. The results are as follows.

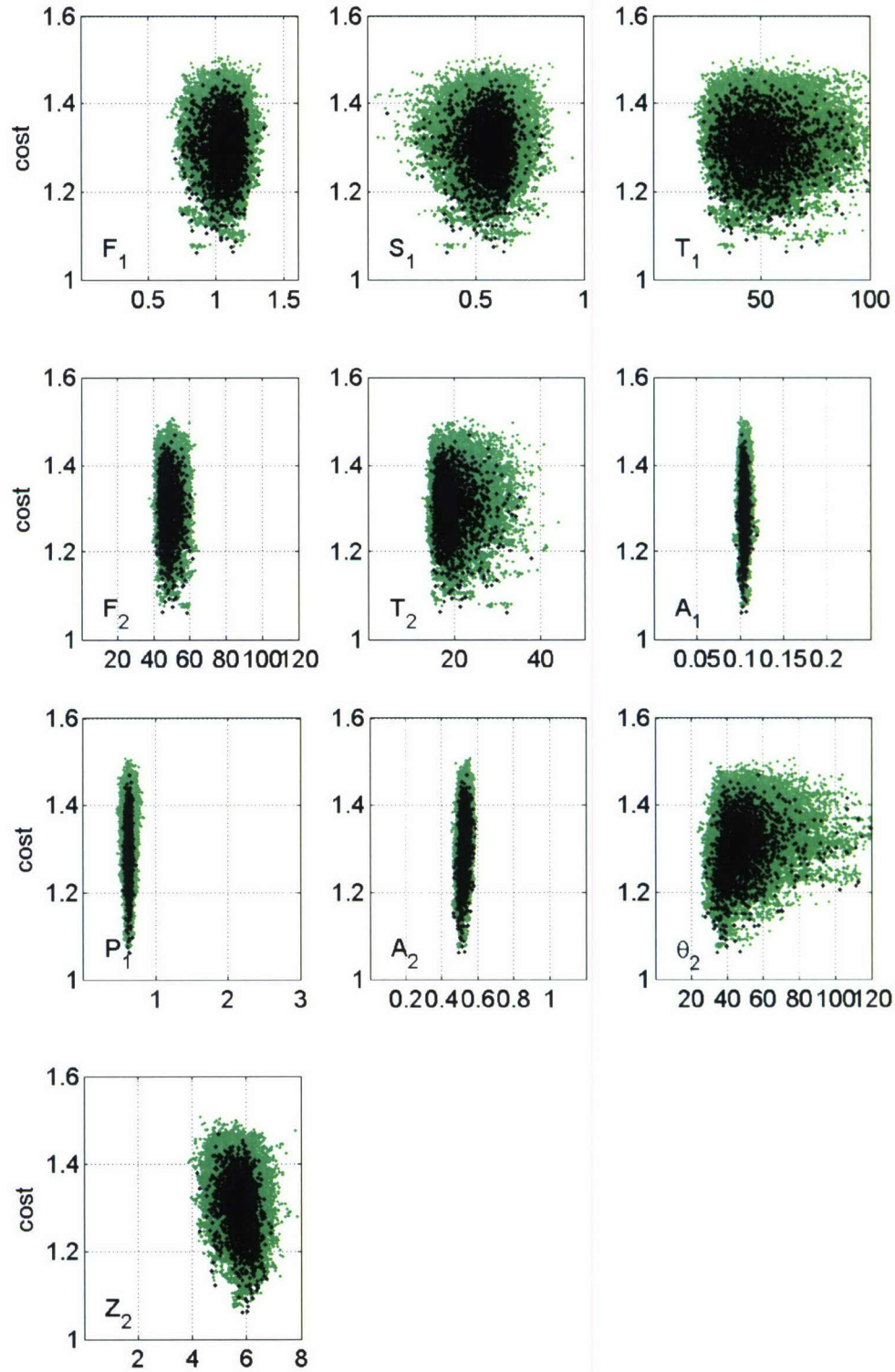


Figure 10 MathieusRuns_280108; 2000 runs; Baltic data included, L_{1N}-cost.

From these pictures follows that the parameters S_1, T_1 and θ_2 are very uncertain in comparison to their search regions and F_2, A_1, P_1 and A_2 are well determined. This can be compared with the sensitivity derived earlier (Figure 5). Often, but not always, uncertain parameters are also insensitive. The standard deviation σ_k of parameter k that results from the green dots is as follows.

Table 6 Uncertainty of the parameters from the last generations of 2000 runs of 280108 [4].

par nr		vub	minimum	maximum	mean	error σ_k	$\frac{\sigma_k}{vub}$
1	F_1	1.6	0.638	1.387	1.058	0.099	0.062
2	S_1	1.0	0.035	0.947	0.558	0.095	0.095
3	T_1	100	16.84	99.86	50.65	13.4	0.134
4	F_2	120	37.84	64.34	48.31	3.6	0.030
5	T_2	50	12.71	43.99	19.64	3.5	0.071
6	A_1	0.25	0.092	0.121	0.1049	0.0026	0.010
7	P_1	3	0.462	0.837	0.626	0.035	0.012
8	A_2	1.2	0.452	0.590	0.524	0.018	0.015
9	θ_2	120	23.96	119.99	50.8	12.3	0.103
10	Z_2	8	3.856	7.865	5.68	0.47	0.059
L_{1N}-cost		--	1.060	1.507	1.322	0.064	--

The last two columns of this table are the most valuable ones, the other columns are given for comparison only. The amount of uncertainty of a parameter is clearly visible in the last column.

The 2000 inversion runs with L_{1N}-cost and the hybrid model for the case that the Baltic data are *exc*luded, are stored as MathieusRuns_290108. Again the files are removed and bestVectors and lastGeneration are stored in Summary_MathieuRuns_290108.mat. The pictures can be compared with the previous ones to demonstrate the importance of the Baltic data. Without the Baltic data, there is very little variation in the salinity. As a result the parameter S_1 cannot be resolved; there is even a concentration on its upper bound. The parameter T_1 is poorly determined too, it has best vectors on the (very high) upper bound of its search interval. It is clear that the Baltic data are very valuable and should be included in the dataset.

[4] Because local search has been applied to derive bestVectors, they are not included in lastGeneration. Therefore the sets lastGeneration and bestVectors are unified to get this table. There is a problem in lastGeneration. 95 of 32000 vectors have NaN as cost. I don't understand this, but have removed these vectors from the set.

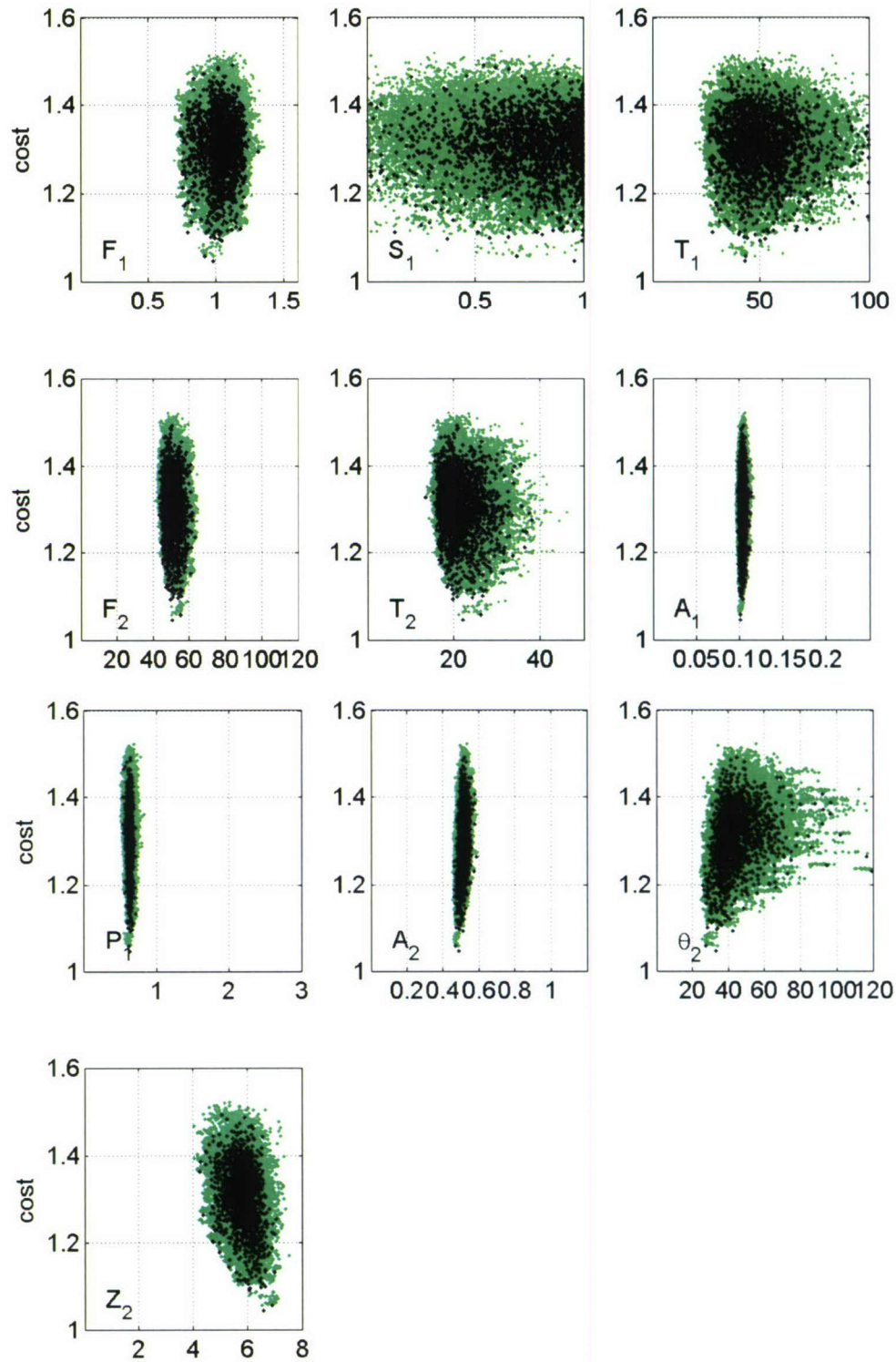


Figure 11 MathieusRuns_290108; 2000 runs, Baltic data excluded, L_{1N}-cost.

It is also interesting to investigate the effect of using the L_2-cost instead of the L_{1N}-cost. MathieusRuns_020208 has 2000 runs with the Hybrid model, random selection of subsets from the full dataset including the Baltic data. The L_2-cost is used. The spread of the bestVectors is smaller than was the case with the L_{1N}-cost (runs 280108). The plots for S_1, θ_2 and Z_2 give the strong impression that there are two ambiguous

solutions. This is most likely coupled to the subset of the data where the formula is fitted to.

This calls attention to the composition of the data set. With five independent variables (f, S, T, pH and z) 166 samples is an extremely small set. If there would be only 3 values per variable, $3^5 = 243$ samples are required for a *uniform spread* over the space of independent variables. Therefore 166 samples cannot have such a uniform spread. For a subset of 130 samples this non-uniform composition is even worse. This phenomenon is the most likely cause for the big variation in costs and parameter settings. Although the random subset approach exaggerates the uncertainty, it gives a feeling for the effect of a changing data set and of a non-uniform spread of the data over the space of all possible combinations of independent variables.

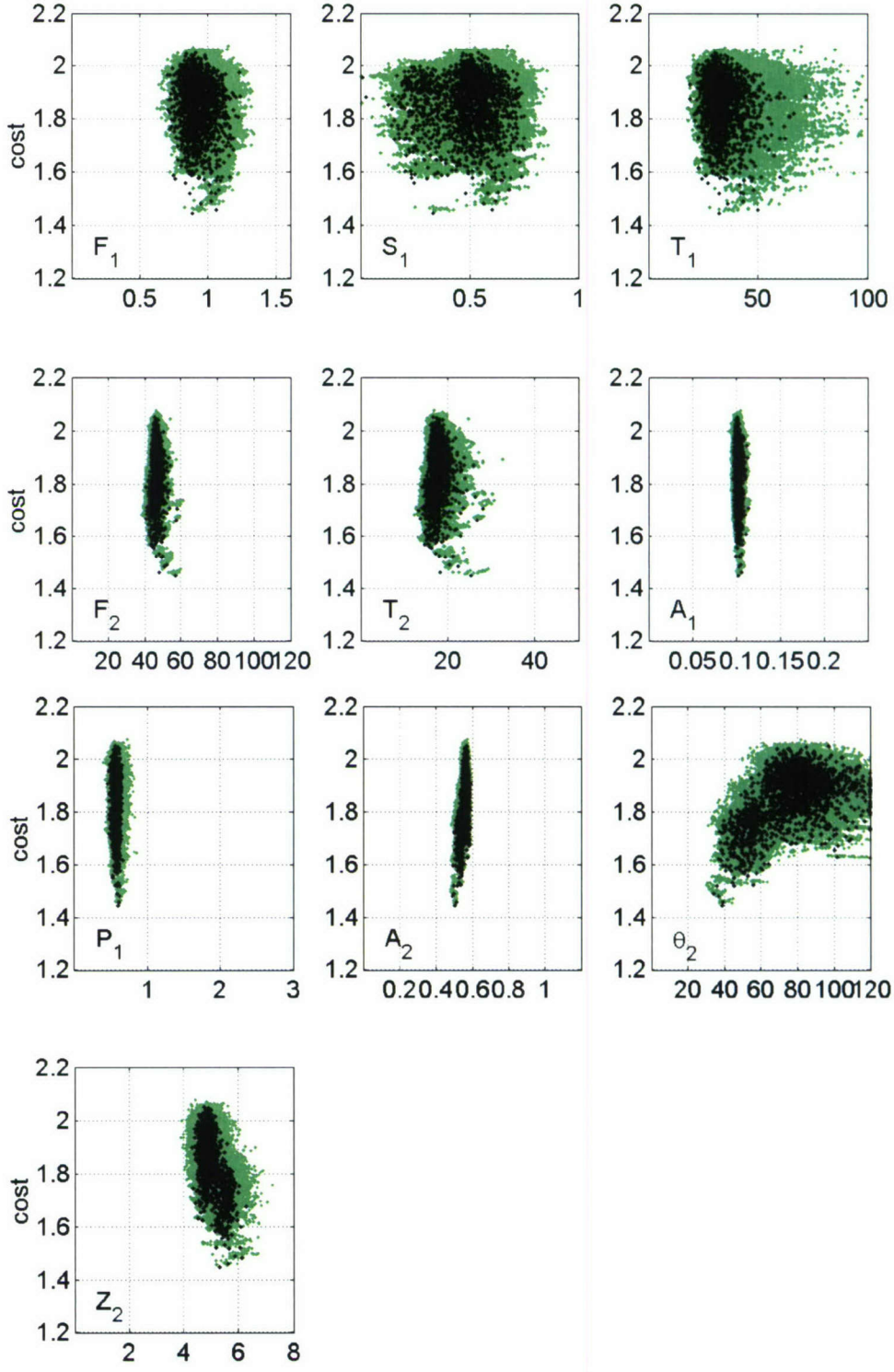


Figure 12 MathieusRuns_020208; 2000 runs, Baltic data included, L_2-cost.

6.3 Derivation of an Improved Tuning

Our search goes in the wrong direction. We are looking for one single best tuning of a formula, but at each step the number of candidates *in*creases instead of *de*creases. Which parameter vector shall we choose? As stated earlier, there is no parameter vector that has lowest cost for all possible subsets of the data. The best we can hope for is to select a parameter vector that has low costs for most of the datasets.

This suggests the need to define a new cost function. Because a parameter vector should perform well on all possible subsets of the data, a large number of random subsets can be selected, for instance 10 000. The cost of a single vector for each of these subsets can be calculated, whereafter the average of these subset costs gives the performance of the vector on all subsets. Raising the number of random subsets to 100.000 or more, leads to less statistical variation of this average cost value. However, with such huge numbers, each data sample will be included in this average cost calculation approximately as many times. As a consequence, the vector with the lowest average cost will be the same as the vector that minimized the cost on the full dataset. But this is a result we already derived in Chapter 5 and summarized in Table 5. The investigation of the robustness by means of random subsets does not lead us to the preferred setting, but has provided a feel for the *uncertainty* of the parameters and the *significance* of cost variations.

It has become clear that it is important to *include* the Baltic data in the dataset.
We still can choose to use the L_{1N}-cost or the L_2-cost. We prefer to use the L_{1N}-cost since it gives equal weight to each sample and is less sensitive to outliers in the data. Therefore we reconsider the results of the 200 runs of 300108 on the full dataset inc. Baltic (Figure 5) and combine it with the uncertainty information from the 2000 random subset runs of 280108 (Figure 10).

The set of best vectors of the 300108 runs has 200 members, with L_{1N}-costs varying from 1.3153 to 1.3277. In comparison to the size of 0.3 for significant differences in costs, the variation of the costs of the best vectors is negligible. All these vectors could be considered of equal quality. To choose a single vector, we explore the wealth of information that is available in this set. We want to choose a single vector which is *close to the 'middle'* of the set, hoping that this increases its robustness. Selecting this vector also means *discretisation* of the parameter values. The size of the discretisation step is implied by the search intervals of the parameters, because normally three digits for a parameter value suffices. The variable upper bounds (vub) and the chosen stepsizes are given in Table 7.

Since the set of best vectors has a considerable spread, the following heuristic approach is applied to select one vector. First 161 vectors that cost less than 1.32 are selected (with test180308B.m). The minimum and maximum value of each parameter in this set gives the range of that parameter, which is used together with the stepsize to make a histogram. The binned parameter values that appear most are selected for closer inspection (test220308.m). This results in the parameter intervals of Table 7.
With the given step sizes an exhaustive search (over 46 million vectors) has been applied (with test160308.m). The results are given in Table 8.

Table 7 Search intervals for exhaustive search with L_{1N}-cost.

param	vub	minimum parameter value	maximum parameter value	parameter interval	step size	number of steps
F_1	1.6	1.0011	1.2045	[1.03, 1.05]	0.01	3
S_1	1.0	0.5154	0.65375	[0.54, 0.57]	0.01	4
T_1	100	38.494	83.166	[46.0, 47.0]	0.1	11
F_2	120	45.29	48.952	[46.0, 47.0]	0.1	11
T_2	50	17.034	19.768	[17.5, 18.5]	0.1	11
A_1	0.25	0.10284	0.10656	[0.104,0.105]	0.001	2
P_1	3	0.61569	0.64968	[0.60, 0.65]	0.01	6
A_2	1.2	0.50491	0.5327	[0.50, 0.55]	0.01	6
θ_2	120	38.014	50.816	[44.0, 44.7]	0.1	8
Z_2	8	5.6077	6.0945	[5.6, 6.0]	0.1	5

The best non-rounded vector of 300108 has a cost of 1.3153, but rounding this vector to digits given by the step sizes increases its cost to 1.3168. The search on the discretised space has found a vector with a cost of 1.3157, just below this value but exceeding 1.3153. With a very small increase of the costs, the Improved Tuning vector is derived, that makes the formula look simpler and will be our final choice.

Table 8 Overview of parameter vectors and upper search boundary.

par	variable upper bound	Ainslie-McColm Hybrid	Best vector for data inc. Baltic; L_{1N}-cost	error σ_k	Improved Tuning (L_{1N}-cost)
file nr			300108	280108	
F_1	1.6	0.78	1.0392	0.099	1.04
S_1	1.0	0.5	0.54538	0.095	0.55
T_1	100	26	46.422	13.4	47
F_2	120	42	46.575	3.6	46.7
T_2	50	17	17.915	3.5	18
A_1	0.25	0.106	0.10387	0.0026	0.104
P_1	3	0.56	0.62167	0.035	0.63
A_2	1.2	0.52	0.52075	0.018	0.52
θ_2	120	43	44.405	12.3	44
Z_2	8	6	5.7855	0.47	5.8
L_{1N}-cost	--	1.4851	1.3153	--	1.3161
L_2-cost	--	2.2346	1.9845	--	1.9875
L_{1F}-cost	--	0.1134	0.1116	--	0.1115

Notice that for all parameters the Improved Tuning is very close to the non-rounded best vector. The Ainslie-McColm settings of F_1, T_1 and P_1 deviate considerably from these values, even more than one standard deviation. The improved accuracy of the simple Improved Tuning hybrid formula is demonstrated in the next section.

The following observation about the set op 200 best vectors of the 300108 runs deserves to be mentioned. Several parameters in this set are strongly coupled. The strongest correlations are the following.

Table 9 Strongest correlations of parameters in the set of best vectors (300108 runs).

	S_1	T_1	A_1	T_2	A_2	θ_2
F_1	0.94	0.95	0.87	--	--	--
F_2	--	--	--	1.00	-0.89	-0.88

The parameters S_1, T_1 and A_1 are strongly correlated to F_1; and T_2, A_2 and θ_2 are coupled to F_2. The linear least squares estimate of the relation between these coupled parameters for the best vectors of the 300108 runs, are derived with the m-file test170308.m. After removing 19 'eccentric' vectors from the set, the linear relations are as follows:

$S_1 = -0.12115 + 0.64651*F_1$
$T_1 = -137.49 + 176.19*F_1$
$A_1 = 0.088883 + 0.01457*F_1$
$T_2 = -16.455 + 0.73837*F_2$
$A_2 = 0.80182 - 0.0060232*F_2$
$\theta_2 = 174.39 - 2.7808*F_2$

This demonstrates that the number of parameters is too big and can be reduced. However, we don't think that this over parametrization hampers our investigation.

For later use we also present the results of a search that minimizes the L_2-cost on the full dataset inc. Baltic (220308 runs) instead of the L_{1N}-cost[5]. The 1140 best vectors are collected in the file Summary_MathieuRuns_220308.mat. The lowest L_2-cost is 1.8905, which is derived for a large number of vectors. The procedure of discretisation and exhaustive search is used to select one discretised vector that is close to the middle of the set of best vectors and has a low L_2-cost.

Table 10 Search intervals for exhaustive search with L_2-cost.

param	vub	parameter interval	step size	number of steps	Best L_2 discretised vector	Improved Tuning vector
runs					220308	300108
F_1	1.6	[0.89, 0.95]	0.01	7	0.91	1.04
S_1	1.0	[0.48, 0.54]	0.01	7	0.5	0.55
T_1	100	[31, 36]	1	6	33	47
F_2	120	[46.4, 46.9]	0.1	6	46.6	46.7
T_2	50	[16, 19]	1	4	18	18
A_1	0.25	[0.100, 0.102]	0.001	3	0.101	0.104
P_1	3	[0.55, 0.59]	0.01	5	0.57	0.63
A_2	1.2	[0.54, 0.58]	0.01	5	0.56	0.52
θ_2	120	[73, 77]	1	5	76	44
Z_2	8	[4.8, 5.0]	0.1	3	4.9	5.8
L_2	--	--	--	--	1.8913	(1.9875)
L_{1N}	--	--	--	--	(1.3865)	1.3161
L_{1F}	--	--	--	--	(0.1190)	(0.1115)

Notice that the best L_2 vector differs considerably in its parameter values from the Improved Tuning vector. We prefer to use the Improved Tuning vector (derived by minimizing the L_{1N}-cost) for the simple formula and will not use the L_2-vector until Chapter 7.

[5] These results agree with those of the 310108 runs; see Table 5.

It is worthwile to verify the position of these parameter vectors in the sensitivity plots of Figure 10 (for the Improved Tuning vector) and 12 (for the best L_2 discretized vector). The central position of both vectors demonstrates that the applied method has derived its objective and suggests robustness to variation of the data set.

6.4 Illustration of improved accuracy

Because there is no single parameter vector that has the lowest cost for all data subsets, it is very illustrative to compare the costs of the François-Garrison formula, the full Ainslie-McColm formula and the Improved Tuning hybrid formula for a large number of datasets. For this, 10.000 random subsets of the full dataset inc. Baltic are selected and for each of them the L_{1N}-costs of the three formulae is calculated and plotted. The formula that has on average the lowest cost is the most accurate one.

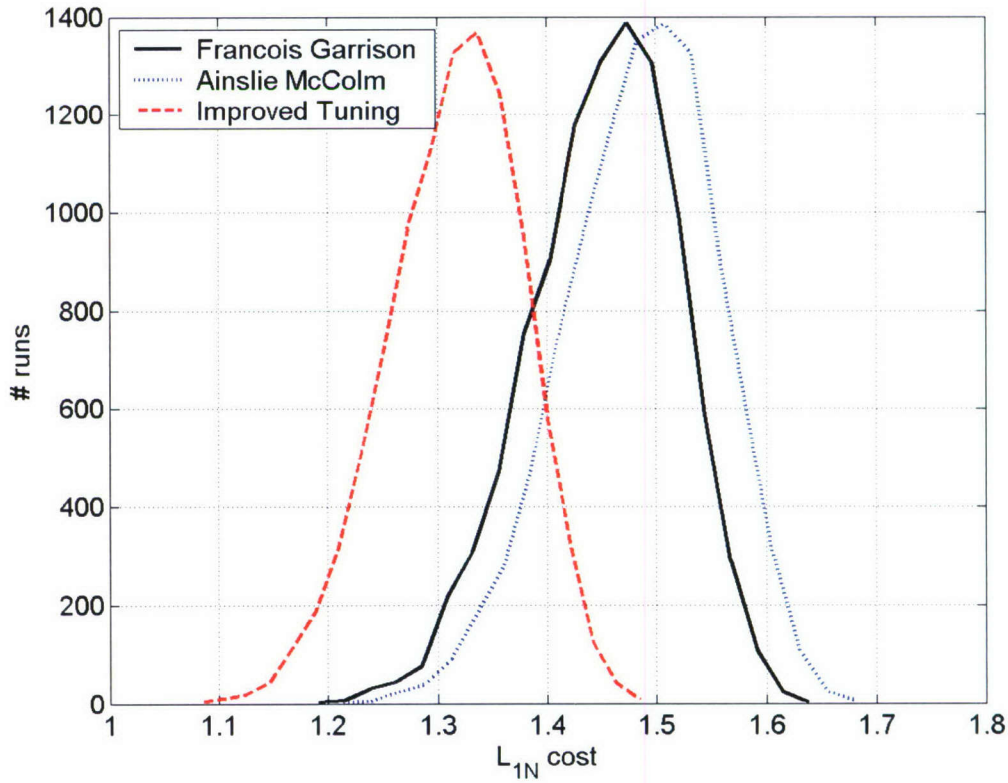


Figure 13 Comparison of the L_{1N}-costs of the François-Garrison formula, the Hybrid Ainslie-McColm formula and the Improved Tuning formula, for 10 000 random subsets of the data inc. Baltic.

The mean L_{1N}-costs for François-Garrison is 1.45, for the full Ainslie-McColm formula it is 1.47 and for the hybrid Improved Tuning formula 1.32. The first picture shows that the full Ainslie-McColm formula is an approximation to that of François-Garrison. The second figure shows that the hybrid Improved Tuning formula is more accurate than François-Garrison in most cases. In the third picture it is visible that for some subsets François-Garrison is more accurate than the Improved Tuning. The question remains, however, if these differences in L_{1N}-costs are significant, since they are far less than 0.3. But even if improved accuracy of the Improved Tuning formula can not be confirmed, it is sure that no accuracy has been sacrificed for the sake of simplicity.

Comparison of the L_2-costs for the three formulae is useful too. Instead of the previous pictures, the cost values of the formulae on the full dataset provides the information.

Table 11 Comparison of costs on the full dataset inc. Baltic.

	François-Garrison formula	Full Ainslie-McColm formula	Ainslie-McColm formula with François-Garrison fresh water part	Hybrid Improved Tuning formula	Best L_2 discretised vector
L_{1N}-cost	1.4507	1.4745	1.4851	1.3161	1.3865
L_2-cost	2.1805	2.2235	2.2346	1.9875	1.8913
L_{1F}-cost	0.1157	0.1144	0.1134	0.1115	0.1190

This comparison shows that the hybrid Improved Tuning, never has less accuracy than the François-Garrison formula. The formula is simpler without loss of accuracy.

6.5 New empirical formula and its characteristics

With the Improved Tuning parameter vector the new empricial formula is as follows.

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3 \text{ dB/km}$$

Boric acid contribution α_1 :

$$\alpha_1 = 0.104 \left(\frac{f_1 f^2}{f_1^2 + f^2} \right) e^{\frac{pH-8}{0.63}} \qquad f_1 = 1.04 \left(\frac{S}{35} \right)^{0.55} e^{\frac{T}{47}}$$

Magnesium sulphate contribution α_2 :

$$\alpha_2 = 0.52 \left(1 + \frac{T}{44} \right) \left(\frac{S}{35} \right) \left(\frac{f_2 f^2}{f_2^2 + f^2} \right) e^{-\frac{z}{5.8}} \qquad f_2 = 46.7 \; e^{\frac{T}{18}}$$

Fresh water absorption α_3 according to François-Garrison:

$$\alpha_3 = A_3 P_3 f^2$$

$$A_3 = 4.937 \; 10^{-4} - 2.59 \; 10^{-5} T + 9.11 \; 10^{-7} T^2 - 1.50 \; 10^{-8} T^3 \quad \text{for} \quad T \leq 20^o C$$

$$A_3 = 3.964 \; 10^{-4} - 1.146 \; 10^{-5} T + 1.45 \; 10^{-7} T^2 - 6.5 \; 10^{-10} T^3 \quad \text{for} \quad T > 20^o C$$

$$P_3 = 1 - 3.83 \; 10^{-2} z + 4.9 \; 10^{-4} z^2$$

The five independent variables are the frequency f [kHz], pH, salinity S [$^o/_{oo}$], temperature T [^{o}C] and depth z [km]. The formula should certainly not be used outside the range of the data, given in Table 12.

Table 12 Range of the data.

f	[0.16, 650]	kHz
pH	[7.69, 8.18]	
S	[8, 40.5]	ppt
T	[−1.75, 22]	°C
z	[0.013, 3.35]	km

A major deficiency of the data is their non-uniform distribution over the space of the independent variables, which can be seen from Figure 16.

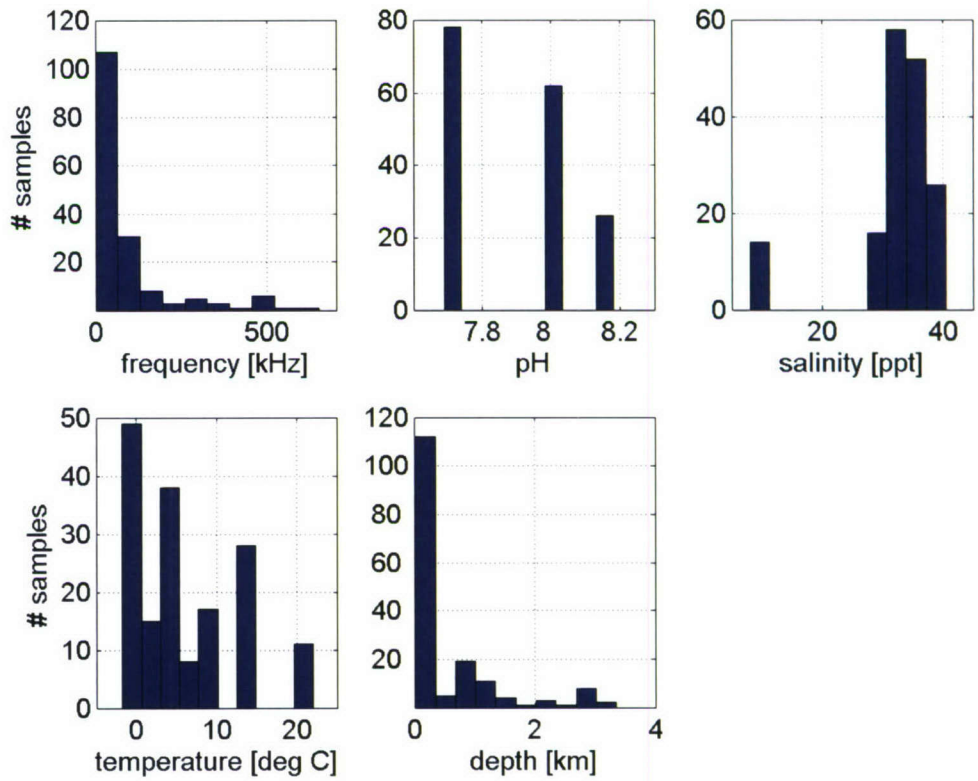


Figure 14 Histograms of the spread of the independent variables in the data.

6.6 Errors of the Improved Tuning formula

How big is the expected error? For this we need to take into account that the absorption values differ many orders of magnitude, depending on the setting of the independent variables $\underline{v}_i$. [6]

To illustrate this we recall the following definitions:

α_i = measured absorption sample i ,

$\alpha(\underline{v}_i)$ = calculated absorption for setting $\underline{v}_i$ and

σ_i = error provided for sample i of the measured data.

Table 13 provides some characteristics of the errors for the Improved Tuning formula.

[6] Made with ProofImprovedSetting3.m.

Table 13 Maximum and minimum of absorption values and errors; ImprovedTuning vector.

		minimum	maximum
measured absorption	α_i	0.0016	227
modelled absorption	$\alpha(\underline{v}_i)$	0.00164	247.64
provided error	σ_i	0.0009	17
fractional error	$\frac{\alpha_i - \alpha(\underline{v}_i)}{\alpha_i}$	-0.43	0.38
normalised error	$\frac{\alpha_i - \alpha(\underline{v}_i)}{\sigma_i}$	-7.24	11.74

A better overview is given by the two dimensional plot of the fractional error versus the normalised error of the data points in Figure 17 (left).

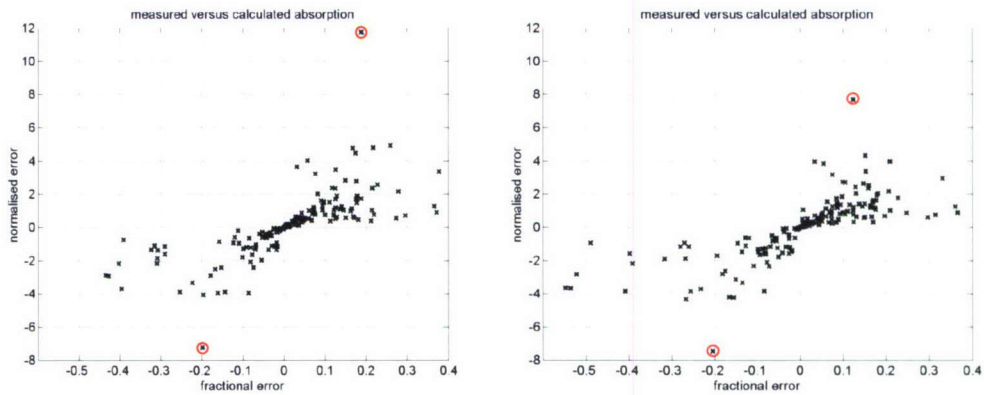


Figure 15 Fractional and normalised errors of all data points for Improved Tuning L_{1N} hybrid formula (left) and for the best L_2 vector formula (right).

Data points that are more than 5 times the provided error away from the calculated absorption can be considered to be outliers. There are 2 of them, sample nr 24 (normalised error -7.24) and nr 50 (+11.74). These samples are given in Table 14.

Table 14 Overview of the 2 outliers (sample numbers 24 and 50).

Investigator	location	year	α_i	error	*F*	*T*	*S*	*pH*	*Z*
Bezdek	Pacific	1972	32.7	0.9	145	7.0	34.0	7.7	0.200
APL	Bering Sea	1973	18.7	0.3	60	-1.75	32.9	7.7	0.045

With this kind of errors it is difficult to provide an error measure for the derived absorption formula.

6.7 Errors of the L_2 best vector formula

For the hybrid formula tuned by the best L_2 vector, the errors are as follows.

Table 15 Maximum and minimum of absorption values and errors; L_2 best vector.

		minimum	maximum
measured absorption	α_i	0.0016	227
modelled absorption	$\alpha(\underline{v}_i)$	0.00165	249.41
provided error	σ_i	0.0009	17
fractional error	$\frac{\alpha_i - \alpha(\underline{v}_i)}{\alpha_i}$	-0.55	0.37
normalised error	$\frac{\alpha_i - \alpha(\underline{v}_i)}{\sigma_i}$	-7.47	7.69

The picture of the errors is comparable to the previous one and given in Figure 16 (right). The same two samples are outliers here too, but with different normalised errors: -7.47 for sample nr 24 and +7.69 for sample 50. The quadratic nature of the L_2-cost function has resulted in smaller cost values for these outliers in comparison to the normalized errors of the Improved Tuning formula (that used the L_{1N}-cost), but this was only possible at the expense of larger fractional errors.

6.8 Alternative data set

Although the errors of the dataset are interesting, the cost values of the Improved Tuning formula given in Table 11 can not be used to provide an error estimate, since the formula has been tuned on these data. An estimate of the expected error should come from an alternative data set.

There are 48 measured absorption samples that have not been used during the search, because they have no measurement error provided with them. This dataset can be used to test the improved accuracy and to provide some estimate of the expected errors. With the L_{1F}-cost it is possible to express in a single number the distance of these samples to their calculated absorption value for each of the formulae. The following table demonstrates that - for this dataset - the accuracy of the hybrid Improved Tuning formula and the best L_2 formula is still at least as accurate as that of the other formulae[7].

Table 16 L_{1F}-cost for the dataset of 48 samples of FG provided without an error.

	François-Garrison formula	Full Ainslie-McColm formula	Ainslie-McColm formula with François-Garrison fresh water part	Hybrid Improved Tuning formula	Hybrid best L_2-cost formula
L_{1F}-cost	0.369	0.374	0.374	0.321	0.329

An estimate for the fractional error of 0.33 is huge and much bigger than the 5% claim of others.

[7] Is this new dataset reliable? It is worrying that the L_{1F}-costs are 3 times as high as for the original dataset, but this can be attributed to the tuning process.

6.9 Discussion

At last we have selected a tuning that results in a *simple* formula that is at least as *accurate* as the François-Garrison formula. Since its parameter values are in the centre of the distribution of the sensitivity plots of the random search runs, it is expected that this tuning is *robust* for changes in the data set. At this moment we prefer to use the L_{1N}-cost during the search process, leading to the Improved Tuning formula. However we also present the best L_2-cost hybrid model. It is worrying that both parameter vectors differ considerably in some of their values. The sensitivity to outliers of the L_2-cost function is a serious argument to prefer the Improved Tuning, especially if the quality of the in situ absorption data with their provided errors can be questioned. But the nice mathematical properties of the L_2-cost function allows for the application of tests, as will be explained in Chapter 7.

We are not able to provide an error for the simple formula. For a big part this is caused by the very small size of the data set. With 5 independent variables (f, T, S, pH and z) 166 samples covers only a negligible part of the variable space. These samples have therefore inevitably a non-uniform spread over this space, which will degrade the estimation performance of the formula. If some of these samples are dependent, this even becomes worse. Hopefully these effects of a limited number of samples is considerably compensated by the physical basis of the formulae.

However, all these issues also hold for the François-Garrison formula and in comparison to the latter, our formula can be considered to be *simpler*, *robust* and at least as *accurate*. In Chapter 7 we will prove its increased accuracy.

7 Testing for significance

Careful consideration of Chapter 6 reveals that the analysis of the significance of variations of the costs is *qualitative*. However, statistical testing can be applied.[8]
By means of the F-test we proof that the simple hybrid formula is significantly more accurate than the formulae of François-Garrison and Ainslie-McColm. Thereafter the distribution of the errors of the data with respect to the formulae is investigated.
The method presented in this report is suited for an extended improved dataset, if one becomes available in the future.

7.1 Statistical theory

From statistical theory the following is known. If N errors ε_i are independent and normally distributed with zero mean and standard deviation σ_i , the statistic X^2 has the central Chi-square distribution $\chi^2(N,0)$ with N degrees of freedom.

$$X^2 = \sum_{i=1}^{N} \left(\frac{\varepsilon_i}{\sigma_i} \right)^2$$

Since the errors $\varepsilon_i = \alpha(\underline{v}_i,\underline{p}) - \alpha_i$ that we consider are coupled to a fitting function $\alpha(\underline{v}_i,\underline{p})$ for which the m_a parameters of the vector $\underline{p}$ are estimated from the data, the statistic X_a^2 has a Chi-square distribution with $N–m_a$ degrees of freedom.

$$X_a^2 = \sum_{i=1}^{N} \left(\frac{\alpha(\underline{v}_i, \underline{p}) - \alpha_i}{\sigma_i} \right)^2$$

If the fitting function agrees with the parent function (the true function) of the data, the statistic X_a^2 represents the spread of the random data around this function. If there is a mismatch between both functions, X_a^2 combines the data quality with the misfit, and distinguishing the contribution of each is problematic. However, if two functions $\alpha(\underline{v}_i,\underline{p})$ and $\beta(\underline{v}_i,\underline{q})$ are fitted to the data, with m_a and m_b estimated parameters respectively, the statistics X_a^2 and X_b^2 differ in their fitting accuracy, but have the same data quality contribution. As a result, combining both statistics affords comparison of the fitting accuracy of the two functions. For this purpose the statistic F_{ab} is constructed, which has a central F-distribution with ($N–m_a$, $N–m_b$) degrees of freedom.

$$F_{ab} = \frac{X_a^2/(N-m_a)}{X_b^2/(N-m_b)} = \frac{1}{F_{ba}}$$

If the a-fitting function matches the parent distribution better than the b-function, it can be expected that on average X_a^2 is smaller than X_b^2 which results in small values of F_{ab}. Very large and very small values of F_{ab} show that it is very likely that one of the fitting functions represents a considerably better description of the data than the other function.

---

[8] The reviewer R. van Vossen has proposed this testing, leading to the addition of this Chapter. The test proves the improvement of the accuracy of the simpler formula.

With these statistical instruments three topics will be considered:

1 Use F_{ab} to test whether the Simple formula is significantly more accurate than the François-Garrison formula.
2 Use X_a^2 to test whether the Simple formula provides a sufficiently accurate description of the data.
3 Investigate the assumed independence and normality of the errors and the correctness of the assignment of uncertainties to the measurements.

7.2 Tests of accuracy; the F- and Chi-square tests

The L_2-cost is closely related to the statistic X_a^2. The definition of the L_2-cost

$$\mathrm{cost}_{L2}(\underline{p}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha(\underline{v}_i, \underline{p}) - \alpha_i}{\sigma_i}\right)^2} \quad \text{results in} \quad X_a^2 = N\left(L_2\mathrm{cost}\right)^2$$

The full dataset inc. Baltic has $N = 166$ samples. Ten parameters are derived from these measured absorption values. This leads to the following overview.

Table 17 Overview of statistics on the full dataset inc. Baltic.

	François-Garrison formula	Full Ainslie-McColm (AMC) formula	AMC formula with François-Garrison fresh water part	Improved Tuning formula	Best L_2 discretised vector
L_2-cost	2.1805	2.2235	2.2346	1.9875	1.8913
N–m	166	156	156	156	156
$X^2/(N-m)$	4.7546	5.2609	5.3135	4.2034	3.8063

Assuming that the Improved Tuning formula is more accurate than the other formulae, the statistics $F_{other,Improved}$ are calculated, each exceeding unity. How big is the probability that, given the respective degrees of freedom, F_{ab}-values or larger ones result from the data purely by chance, while both fitting functions are equally accurate. If this probability is smaller than 5%, it can be concluded that the assumption of equal accuracy is very unlikely and thus that the Improved Tuning formula is significantly more accurate than the formula to which it is compared. The values are as follows.

Table 18 Results of the F-test for L_{1N} Improved Tuning comparison.

Formula_a, Formula_b	F_{ab}	Degrees of freedom	probability
FG, Improved	1.1311	166, 156	0.22
AMC, Improved	1.2516	156, 156	0.08
AMChybrid, Improved	1.2641	156, 156	0.07

The values F_{ab} are small for more than 150 degrees of freedom. The probabilities that such a small F_{ab}-value or larger ones appear exceed 0.05, making it a common (not a rare) event. If, for instance, the François-Garrison and the Improved Tuning formulae are equally accurate, randomly selected data will produce an F_{ab}-value of 1.13 or more in 22% of the experiments. This means that the assumption of equal accuracy is quite likely to be true and therefore that the accuracy of the Improved Tuning formula is not proven to be significantly better than that of the other formula.

The X_a^2 statistic is smaller for the selected L_2-best discretized vector. If the Simple Formula is tuned with this vector, the comparison is as follows.

Table 19 Results of the F-test for L_2-best Simple Formula comparison.

Formula_a, Formula_b	F_{ab}	Degrees of freedom	probability
FG, Simple	1.2491	**166**, 156	0.080
AMC, Simple	1.3821	156, 156	0.022
AMC hybrid, Simple	1.3959	156, 156	0.019
Improved, Simple	1.1043	156, 156	0.268
FG, Simple	1.3292	**156**, 156	0.038

Since the probability of the François-Garrison comparison (with 166 degrees of freedom) exceeds 0.05, improved accuracy is not yet established with respect to François-Garrison. However, the low probability confirms the claim that the L_2-best Simple Formula is at least as accurate than François-Garrison. The value of 0.08 means that, when both formulae are equally accurate, the derived value (or larger ones) will occur only in 8% of the cases due to the random nature of the data.
The test demonstrates that the accuracy of the L_2-best Simple Formula is significantly better than the Ainslie-McColm (AMC) formulae. The comparison of the Improved Tuning and the Simple formula on the other hand shows that they are of comparable accuracy.

The previous test on François-Garrison is too conservative with respect to the assigned degrees of freedom. It is known that François-Garrison have used a part of the dataset to tune their more complex formula. This justifies a reduction of the number of degrees of freedom for François-Garrison too. If we assume that François-Garrison have also tuned 10 parameters from the data (resulting in 156 degrees of freedom), the probability of the F-value is 0.038, well below the 0.05 threshold. In this sense it is established that the L_2-best Simple Formula is not only simpler, but also more accurate than François-Garrison.

Is the L_{1N} Improved Tuning formula accurate enough to describe the data? Can it be considered to represent the parent function of the data? This can be investigated by means of the Chi-square test. A Chi-square probability distribution with 156 degrees of freedom has a probability of zero that the statistic $X_a^2/(N–m) = 4.2034$ or bigger. Such a big deviation of the data from the formula is therefore very unlikely. Can we conclude that the formula does not describe the data properly? The Chi-square test assumes normally distributed errors. Violation of this assumption, a bias in the data, or outliers, immediately result in big values of the Chi-square statistic. This makes it impossible to make a clear inference from the failure of this test. The conclusion that the formula is a poor description of the data is therefore not justified. In particular it is the reduced sensitivity of the F-test for these phenomena that makes this test so valuable, more than the Chi-square test.

7.3 Investigation of the errors

What causes the poor fits? The distribution of the errors provides clues. The histograms of the deviations of the samples from the François-Garrison formula and from the Improved Tuning simple formula are presented in Figure 19 together with the expected

number of observations per bin from the standard normal distribution (in total 166) for comparison.

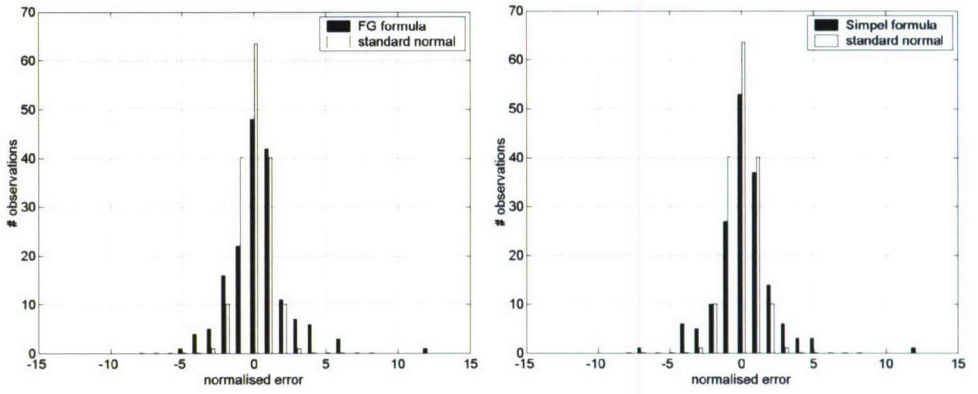


Figure 16 Comparing occurences of errors for the FG and the Simple formula.

The histograms of the normalised errors (ε_i/σ_i) deviate significantly from the normal distribution, especially their tails[9]. For the François-Garrison formula 5 samples have errors that are more than 5 times the provided uncertainty, while for the simple Improved Tuning formula this happens 2 times. The last 2 samples also differ more than 5 standard deviations from the calculated absorption values of François-Garrison. They are the samples number 24 and 50, which are already presented in Chapter 6. Five standard deviations can be considered statistically 'impossible'. These outliers suggest an explanation for the very big values of the X^2 statistic, because a quadratic distance measure (the L_2-cost function) is very sensitive to outliers.

We investigate if removing these two samples leads to considerably different results. Global search is applied on the full dataset inc. Baltic, but without the 2 samples nr 24 and 50, using the L_{1N}-cost (1000 runs of 230308) and the L_2-cost (1000 runs of 240308). From the sets of best vectors the procedure already applied in Chapter 6 results in discretised vectors, that are very good and near the center of the sets of best vectors. The results are as follows.

Table 20 Discretised parameter vectors with their costs.

param	vub	best discrete vector with L_{1N}-cost	best discrete vector with L_2-cost
runs		230308	240308
F_1	1.6	1.02	0.93
S_1	1.0	0.53	0.50
T_1	100	46.3	33.9
F_2	120	48.6	45.6
T_2	50	19.3	16.6
A_1	0.25	0.103	0.101
P_1	3	0.62	0.57
A_2	1.2	0.50	0.53
θ_2	120	37.5	50.1
Z_2	8	6.2	5.4
L_{1N}	--	1.2205	(1.2565)
L_2	--	(1.7189)	1.6559

[9] Testing is possible; see also probability plots mentioned by Rice [7].

With two very poor samples removed, the L_2-costs of all formulae decrease considerably. A new overview of statistics is given in Table 20.

Table 21 Overview of statistics on the full dataset inc. Baltic; exc 2 samples 24 and 50.

	François-Garrison formula	Full Ainslie-McColm formula	Ainslie-McColm formula with François-Garrison fresh water part	Selected vector L_{1N}-formula 230308
L_2-cost	1.8593	1.895	1.9314	1.7189
$N-m$	164	154	154	154
$X^2/(N-m)$	3.457	3.824	3.973	3.1465

Application of the F-test results in the following probabilities.

Table 22 Results of the F-test for L_{1N}-best Selected Formula comparison.

Formula_a, Formula_b	F_{ab}	Degrees of freedom	probability
FG, Selected L_{1N}	1.0987	164, 154	0.28
AMC, Selected L_{1N}	1.2153	154, 154	0.11
AMChybrid, Selected L_{1N}	1.2627	154, 154	0.075

The removal of two outliers has improved the fit of François-Garrison and Ainslie-McColm more than the L_{1N}-Selected simple formula. This decreases the difference in accuracy and thereby increases the probabilities of the F-statistics. The same happens with the best vector derived with the L_2-cost (240308 runs). The dubious approach of removing some unwelcome samples, does not help us to find a more accurate simple formula than the one we already have.

The investigation of the distribution of the errors demonstrates that some or all assumptions, needed to infer a Chi-square distribution of the errors, are violated. The samples are not independent, the normality assumption does not hold, or the estimated uncertainties of the data are too small. With more than 150 samples the law of big numbers compensates for a non-normal distribution of errors. This leaves the other two aspects of the data that deserve closer inspection, but not in this report.

We conclude that - if we accept that the François-Garrison formula should be assigned 156 degrees of freedom, instead of 166 - we have derived a significantly more accurate simple formula than François-Garrison (the L_2-best Simple formula). The major flaw is that the errors with respect to this formula are not normally distributed. However, the method presented here can be applied on a better dataset once it becomes available.

8 Applications and conclusion

Many sonar applications rely on a proper modelling of the propagation of sound in seawater. Since absorption of sound is one aspect of this propagation, an accurate absorption formula is needed. If it is also simple it improves understanding too. In this report an accurate and simple formula is derived. It is suitable for replacing the well known formula of François and Garrison in ongoing or future TNO projects, such as ALMOST updates (sonar performance prediction), RUMBLE2 and Mean Grainsize Mapping (geoacoustic parameter estimation).

The best absorption formula depends on the cost function used and the dataset that is considered. This is demonstrated by investigating the effect of either including or excluding the low salinity Baltic data on the best tuning of the model, and the effect of selecting a subset of the available data. The robust and simple formula which is derived, is shown to be more accurate than the standard formula of François and Garrison.
Only the boron and magnesium relaxations have been elaborated on; the fresh water part of the absorption is that of François and Garrison. While at the start the objective was to simplify the formula without loss of accuracy, it was found that a small, but statistically significant improvement of the accuracy could also be derived.
These characteristics of robustness, simplicity and accuracy provide strong arguments to use this formula instead of the standard formula of the last decades.

9 Literature

[1] R.E. François and G.R. Garrison,
'Sound absorption based on ocean measurements: Part I: Pure water and magnesium sulfate contributions',
Jasa 72 (3), September 1982, pp 896-907.

[2] R.E. François and G.R. Garrison,
'Sound absorption based on ocean measurements: Part II: Boric acid contribution and equation for total absorption', Jasa 72 (6), December 1982, pp 1879-1890.

[3] M.A. Ainslie and J.G. McColm,
'A simplified formula for viscous and chemical absorption in sea water', Jasa 103, March 1998, pp 1671-1672.

[4] H.G. Schneider, R.Thiele and P.C. Wille,
'Measurement of sound absorption in low salinity water of the Baltic sea', Jasa 77, April 1985, pp 1409-1412.

[5] S.R. Murphy, G.R. Garrison and D.S. Potter,
'Sound absorption at 50 to 500 kc from tranmission measurements in the sea', Jasa 30, September 1958, pp 871-875.

[6] R H Mellen, P M Scheifele and D G Browning,
'Global Model for Sound Absorption in Sea Water Part II: Geosecs PH Data Analysis', NUSC Technical Report 7925, May 1987.

[7] John A. Rice,
'Mathematical statistics and data analysis', Duxbury Press, ISBN 0-534-20934-3.

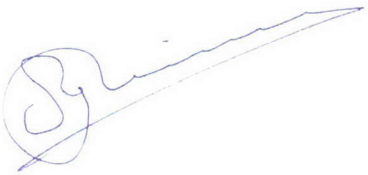
10 Acknowledgements

The authors are very grateful to the reviewer R. van Vossen, whose valuable comments resulted in a significant improvement of this report.

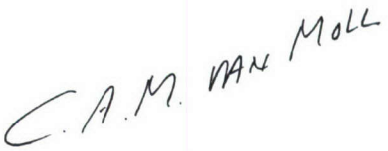
11 Signature

The Hague, May 2008

TNO Defence, Security and Safety



F.P.G. Driessen, MSc
Head of department



C.A.M van Moll, MSc
Author

A François-Garrison formula

The formula of François-Garrison is complicated and is as follows [2].

The independent variables are the frequency f [kHz], temperature T [°C], pH, salinity S [°/oo] and depth D [m]. The total absorption consists of three contributions: α_1 for boric acid, α_2 for magnesium sulphate and α_3 for fresh water.

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3 \qquad [\text{dB km}^{-1}]$$

Boric acid contribution α_1 :

$$\alpha_1 = \frac{A_1 P_1 f_1 f^2}{f_1^2 + f^2}$$

$$A_1 = \frac{8.86}{c} 10^{(0.78\,pH - 5)} \qquad [\text{dB km}^{-1}\ \text{kHz}^{-1}]$$

$$P_1 = 1$$

$$f_1 = 2.8 \left(\frac{S}{35}\right)^{0.5} 10^{\left(4 - \frac{1245}{\theta}\right)} \qquad [\text{kHz}]$$

$$c = 1412 + 3.21T + 1.19\,S + 0.0167\,D \quad \text{sound speed [m/s]}$$

$$\theta = 273 + T$$

$MgSO_4$ contribution α_2 :

$$\alpha_2 = \frac{A_2 P_2 f_2 f^2}{f_2^2 + f^2}$$

$$A_2 = 21.44 \frac{S}{c} (1 + 0.025\,T) \qquad [\text{dB km}^{-1}\ \text{kHz}^{-1}]$$

$$P_2 = 1 - 1.37\ 10^{-4}\ D + 6.2\ 10^{-9}\ D^2$$

$$f_2 = \frac{8.17\ 10^{\left(8 - \frac{1990}{\theta}\right)}}{1 + 0.0018\,(S - 35)} \qquad [\text{kHz}]$$

Fresh water absorption α_3 :

$$\alpha_3 = A_3 P_3 f^2$$

$$A_3 = 4.937\ 10^{-4} - 2.59\ 10^{-5}T + 9.11\ 10^{-7}T^2 - 1.50\ 10^{-8}T^3 \quad \text{for} \quad T \leq 20^{o}C$$

[dB km^{-1} kHz^{-2}]

$$A_3 = 3.964\ 10^{-4} - 1.146\ 10^{-5}T + 1.45\ 10^{-7}T^2 - 6.5\ 10^{-10}T^3 \quad \text{for} \quad T > 20^{o}C$$

$$P_3 = 1 - 3.83\ 10^{-5}\ D + 4.9\ 10^{-10}\ D^2$$

B Matlab m-files

During the investigation the following Matlab m-files have been used[10].

Table 23 M-files used during local search.

m-file	function	calls
Driver_AbsorptionEstimationParameters_local.m	stores measured data as global; applies local search on the original vector	setdataset.m fminsearch getcost_Mathieu_local
setdataset.m	calls measured data; choose including baltic data or not	readtable
readtable	sets the path to the digital data	--
fminsearch	applies Downhill Simplex (local search) on a parameter vector (it is a standard Matlab routine)	--
getcost_Mathieu_local.m	calculates the cost for a supplied parameter vector and the original Ainslie-McColm formula; choose between L_1 and L_2 norm	--

Table 24 M-files used for global search.

m-file	function	calls
Driver_AbsorptionEstimationParameters_global.m	set the tunings, calls the data and starts the global search	DE_parameters_Mathieu setdataset structure_DE
DE_parameters_Mathieu.m	sets a multitude of tunings: file names, energyfunction, vlb, vub and DE tuning parameters	getcost_Mathieu_global_Hybrid getcost_Mathieu_global_random_FullAtten
getcost_Mathieu_global_Hybrid.m	calculates the cost for a supplied parameter vector; choose between AMC and FG fresh water part; choose between L_{1N} and L_2 cost	check_gen FreshWaterAbsorptionFG
getcost_Mathieu_global_random_FullAtten.m	calculates the cost for a supplied parameter vector on a random subset of the data; choose between AMC and FG fresh water part; choose between L_{1N} and L_2 cost	check_gen FreshWaterAbsorptionFG
FreshWaterAbsorptionFG.m	calculates the fresh water absorption according to François-Garrison	--
RandomSubset.m	selects a random subset from the original data set	--

[10] These files are burned on a CD with the name '22 April 2008, Absorption formula estimation'.

Table 25 Global optimization m-files.

m-file	function	calls
structure_DE.m	starts the Differnetial Evolution global search; include 4 restarts? include local search at the end? Select random subsets? Save the global search results.	DE_parameters_Mathieu RandomSubset preprocess conditional_lhd process_DE fminsearch
preprocess.m	creates a starting population and applies global serach on it; selects a quarter of the last generation	conditional_lhd process_DE reduce_population
reduce_population.m	selects a quarter of the vectors from the present generation	--
conditional_lhd.m	creates a random new generation that satisfies the extra requirements, while there is not an existing generation	satisfy_requirements_Mathieu
process_DE.m	applies the DE search process on the populations	new_generation_DE
new_generation_DE.m	creates a new generation an calculates the costs of the vectors therein	create_descendants_DE
create_descendants_DE.m	creates a new generation from the present one, that satisfies extra requirements	satisfy_requirements_Mathieu
satisfy_requirements_Mathieu.m	no extra requirements are inforced; this file is only needed to prevent a lot of changes in the optimisation m-files	--
check_gen.m	checks if all parameters are between vlb and vub	--

Table 26 M-files used for the analysis of the results of globals search runs.

m-file	function	calls
AnalyseResults_Mathieu	collects the vectors evaluated during the search and presents red dot pictures for the 10-parameter global search	--
ProofImprovedSetting4.m	selects random subsets of the dataset and calculates several costs for a particular parameter vector, shows the results in a dot-picture, calculates the average costs	setdataset RandomSubset getcost_Mathieu_global_Hybrid getcost_Mathieu_global_random_FullAtten getcost_FrancoisGarrison_FullAtten
getcost_FrancoisGarrison_FullAtten.m	calculates the FG cost for a random subset of the dataset	abscoef
abscoef	calculates the absorption with the FG formula	--
CompareErrors.m	calculates the absorption for a single parameter vector and shows the fractional versus the normalised error for all data samples	setdataset FreshWaterAbsorptionFG
ProofImprovedSetting3.m	used to make the report; investigates the errors of different formulae	setdataset FreshWaterAbsorptionFG
test170308.m	used to make the report; establish best linear relation between several parameters	--
test180308.m and test180308B.m test220308.m	used to make the report; plot histograms for the parameter values in bestVectors	--
test160308.m	used to make the report; exhaustive search for discretised parameter values	DE_parameters_Mathieu setdataset getcost_Mathieu_global_Hybrid
test210308.m	used to make the report; select a single vector from the vectors derived from an exhaustive search, by means of 2 cost values	setdataset getcost_Mathieu_global_Hybrid

C Sample values

The contents of the file alldata_prepared2.txt is as follows

FG [1], Table I

Location	investigator	year	depth	range	cs[m/s]	temp	salinity	pH	frequency	attenuation	error
Pacific	Bezdek	1971	750	0.9	NaN	4.0	34.3	7.7	75.8	18.9	1.0
Pacific	Bezdek	1971	1350	0.9	NaN	3.4	34.4	7.7	75.8	17.0	1.0
Pacific	Bezdek	1971	1950	0.9	NaN	2.8	34.5	7.7	75.8	15.3	1.0
Pacific	Bezdek	1971	2550	0.9	NaN	2.1	34.6	7.7	75.8	15.1	1.0
Pacific	Bezdek	1971	3150	0.9	NaN	1.5	34.7	7.7	75.8	13.0	1.0
Pacific	Bezdek	1971	910	1.3	NaN	3.8	34.3	7.7	75.8	19.3	1.0
Pacific	Bezdek	1971	910	1.3	NaN	3.8	34.3	7.7	75.8	20.0	1.0
Pacific	Bezdek	1971	910	1.3	NaN	3.8	34.3	7.7	75.8	20.5	1.0
Pacific	Bezdek	1971	1520	1.3	NaN	3.2	34.4	7.7	75.8	17.8	1.0
Pacific	Bezdek	1971	1520	1.3	NaN	3.2	34.4	7.7	75.8	17.1	1.0
Pacific	Bezdek	1971	1520	1.3	NaN	3.2	34.4	7.7	75.8	17.6	1.0
Pacific	Bezdek	1971	2130	1.3	NaN	2.5	34.5	7.7	75.8	16.4	1.0
Pacific	Bezdek	1971	2130	1.3	NaN	2.5	34.5	7.7	75.8	16.8	1.0
Pacific	Bezdek	1971	2130	1.3	NaN	2.5	34.5	7.7	75.8	17.0	1.0
Pacific	Bezdek	1971	2740	1.3	NaN	1.9	34.6	7.7	75.8	14.4	1.0
Pacific	Bezdek	1971	3350	1.3	NaN	1.3	34.7	7.7	75.8	13.3	1.0
Pacific	Bezdek	1972	200	1.3	NaN	7.0	34.0	7.7	30	10.5	4.3
Pacific	Bezdek	1972	200	1.3	NaN	7.0	34.0	7.7	45	12.6	3.5
Pacific	Bezdek	1972	200	1.3	NaN	7.0	34.0	7.7	60	18.4	3.4
Pacific	Bezdek	1972	200	1.3	NaN	7.0	34.0	7.7	75.8	24.5	3.6
Pacific	Bezdek	1972	200	1.3	NaN	7.0	34.0	7.7	75.8	23.5	2.9
Pacific	Bezdek	1972	200	1.3	NaN	7.0	34.0	7.7	75.8	20.9	4.2
Pacific	Bezdek	1972	200	1.3	NaN	7.0	34.0	7.7	90	21.0	3.5
Pacific	Bezdek	1972	200	1.3	NaN	7.0	34.0	7.7	145	32.7	0.9
Pacific	Bezdek	1972	2800	1.3	NaN	2.5	34.6	7.7	30	3.6	0.8
Pacific	Bezdek	1972	2800	1.3	NaN	2.5	34.6	7.7	45	6.3	1.8
Pacific	Bezdek	1972	2800	1.3	NaN	2.5	34.6	7.7	60	9.9	1.8
Pacific	Bezdek	1972	2800	1.3	NaN	2.5	34.6	7.7	75.8	13.0	1.0
Pacific	Bezdek	1972	2800	1.3	NaN	2.5	34.6	7.7	75.8	10.8	2.6
Pacific	Bezdek	1972	2800	1.3	NaN	2.5	34.6	7.7	90	12.6	1.8
Pacific	Bezdek	1972	2800	1.3	NaN	2.5	34.6	7.7	145	25.3	2.6

From Table I and II Murphy, Garrison, Potter, Jasa 1958; corresponds to FG [1], table II, column measured α absorption and error adjusted 23-01-2008 by taking uncorrected and recalculate to m by division by 0.9144.

Dabob Bay	APL	1953	46	.820	NaN	9.9	30.4	7.7	60	16.95	0.11
Dabob Bay	APL	1954	91	.780	NaN	10.1	30.4	7.7	60	16.95	0.55
Dabob Bay	APL	1954	34	.650	NaN	8.0	29.1	7.7	60	15.42	1.31
Dabob Bay	APL	1954	128	.900	NaN	9.5	30.9	7.7	60	15.42	0.87
Dabob Bay	APL	1954	82	1.300	NaN	8.0	29.1	7.7	60	15.86	0.22
Dabob Bay	APL	1954	76	.900	NaN	9.0	29.5	7.7	60	15.09	0.33
Dabob Bay	APL	1954	76	.400	NaN	9.4	29.7	7.7	60	15.64	0.44
Dabob Bay	APL	1954	76	1.100	NaN	9.3	29.6	7.7	60	15.86	0.44
Dabob Bay	APL	1955	91	.580	NaN	9.0	29.8	7.7	142	37.95	1.20
Dabob Bay	APL	1955	91	.540	NaN	7.8	30.0	7.7	142	36.64	0.87
Dabob Bay	APL	1956	61	.560	NaN	8.0	29.8	7.7	142	38.93	0.55

Dabob Bay	APL	1956	149	.325	NaN	9.3	30.4	7.7	272	62.12	0.66
Dabob Bay	APL	1955	91	.270	NaN	9.0	30.0	7.7	467	102.36	3.17
Dabob Bay	APL	1955	91	.300	NaN	7.8	29.8	7.7	467	119.20	2.84
Dabob Bay	APL	1956	46	.280	NaN	8.3	29.4	7.7	467	107.83	1.42
Dabob Bay	APL	1956	67	.270	NaN	8.0	29.5	7.7	467	107.39	20.8
Dabob Bay	APL	1956	128	.310	NaN	9.2	30.4	7.7	467	113.74	0.98

From Table II FG [1]

whole line has been removed, because the zero salinity is suspicious

T3	APL	1972	35	.570	NaN	-1.62	31.9	7.7	50	12.0	1.1
Bering Sea	APL	1973	45	.9	NaN	-1.75	32.9	7.7	60	18.7	0.3

there is no justification for removing

Chukchi Sea	APL	1974	45	1.3	NaN	-1.6	32.3	8.0	10	1.43	0.55
Chukchi Sea	APL	1974	45	1.3	NaN	-1.6	32.3	8.0	20	3.62	0.55
Chukchi Sea	APL	1974	45	1.3	NaN	-1.6	32.3	8.0	30	7.78	0.55
Chukchi Sea	APL	1974	45	1.3	NaN	-1.6	32.3	8.0	40	10.40	0.44
Chukchi Sea	APL	1974	45	1.3	NaN	-1.6	32.3	8.0	60	13.90	0.55
Chukchi Sea	APL	1975	45	1.1	NaN	-1.6	32.0	8.0	7.1	0.82	0.44
Chukchi Sea	APL	1975	45	1.1	NaN	-1.6	32.0	8.0	20	4.86	0.22
Chukchi Sea	APL	1975	45	1.1	NaN	-1.6	32.0	8.0	30	8.47	0.33
Chukchi Sea	APL	1975	45	1.1	NaN	-1.6	32.0	8.0	60	14.93	0.55
Kane Basin	APL	1979	40	1.077	NaN	-1.7	33.8	8	10	0.88	0.45
Kane Basin	APL	1979	40	1.077	NaN	-1.7	33.8	8	20	3.11	0.78
Kane Basin	APL	1979	40	1.077	NaN	-1.7	33.8	8	30	5.25	0.97
Kane Basin	APL	1979	40	1.077	NaN	-1.7	33.8	8	60	10.89	1.64
Kane Basin	APL	1979	40	1.077	NaN	-1.7	33.8	8	75	13.00	1.90
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	100	21.7	1.4
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	150	30.2	2.5
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	200	38.6	2.6
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	275	55.6	2.6
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	300	71.5	7.1
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	350	93.0	3.8
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	420	106.0	7.0
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	500	135.0	5.0
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	550	131.0	14.0
Bering Sea	APL	1980	13	0.71	NaN	-1.7	32.1	7.7	650	227.0	17.0
Beaufort Sea	APL	1980	200	0.16	NaN	-1.5	33.6	8	59	24.1	2.7
Beaufort Sea	APL	1980	200	0.16	NaN	-1.5	33.6	8	84	26.6	1.4
Beaufort Sea	APL	1980	200	0.16	NaN	-1.5	33.6	8	107	28.4	2.9
Beaufort Sea	APL	1980	200	0.16	NaN	-1.5	33.6	8	161	36.6	2.7
Beaufort Sea	APL	1980	200	0.16	NaN	-1.5	33.6	8	251	60.6	2.2
Beaufort Sea	APL	1980	200	0.16	NaN	-1.5	33.6	8	297	66.4	1.1
Beaufort Sea	APL	1980	200	0.16	NaN	-1.5	33.6	8	347	89.9	7.3
Beaufort Sea	APL	1980	200	0.25	NaN	-1.2	33.6	8	59	21.0	2.7
Beaufort Sea	APL	1980	200	0.25	NaN	-1.2	33.6	8	84	24.2	2.9
Beaufort Sea	APL	1980	200	0.25	NaN	-1.2	33.6	8	107	30.6	2.7
Beaufort Sea	APL	1980	200	0.25	NaN	-1.2	33.6	8	161	41.7	3.8
Beaufort Sea	APL	1980	200	0.25	NaN	-1.2	33.6	8	251	60.4	3.2
Beaufort Sea	APL	1980	200	0.25	NaN	-1.2	33.6	8	297	77.8	4.1
Beaufort Sea	APL	1980	200	0.25	NaN	-1.2	33.6	8	347	100.9	6.4

FG [1], Table IV

Arctic	Greene	1965	122	1.2	NaN	-1.4	32.6	8	19	3.95	1.17
Arctic	Greene	1965	122	1.2	NaN	-1.4	32.6	8	30	5.45	0.98
Arctic	Greene	1965	122	1.2	NaN	-1.4	32.6	8	39	9.39	1.07
Arctic	Greene	1965	122	1.2	NaN	-1.4	32.6	8	49	11.29	1.28
Arctic	Greene	1966	122	.97	NaN	-1.3	32.3	8	31	8.04	0.71
Arctic	Greene	1966	122	.97	NaN	-1.3	32.3	8	41	8.84	0.59
Arctic	Greene	1966	122	.97	NaN	-1.3	32.3	8	52	11.19	0.69
Arctic	Greene	1966	122	.97	NaN	-1.3	32.3	8	72	15.06	0.95
Arctic	Greene	1966	122	.97	NaN	-1.3	32.3	8	84	19.23	0.96

FG [2], Table I; From here on the wrong column of the table was used. The right column is 'Adjusted to give ep. (7) f_1, Total α (dB/km)'. The proper values are inserted in alldata_prepared2.txt as given here.

NE Pacific	Chow and Turner	1973	505	1400	1467	4.6	34.05	7.69	0.160	0.0016	0.0009
NE Pacific	Chow and Turner	1973	505	1400	1467	4.6	34.05	7.69	0.250	0.0047	0.0012
NE Pacific	Chow and Turner	1973	505	1400	1467	4.6	34.05	7.69	0.400	0.0101	0.0016
NE Pacific	Chow and Turner	1973	505	1400	1467	4.6	34.05	7.69	0.630	0.0234	0.0027
NE Pacific	Chow and Turner	1973	505	1400	1467	4.6	34.05	7.69	0.800	0.0298	0.0033
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	0.354	0.0135	0.0010
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	0.446	0.0203	0.0010
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	0.562	0.0266	0.0010
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	0.707	0.0372	0.0010
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	0.891	0.0524	0.0010
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	1.120	0.0612	0.0025
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	1.410	0.0763	0.0037
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	1.780	0.1103	0.0120
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	2.240	0.1485	0.0170
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	2.820	0.1800	0.0200
Atlantic	Thorp	1962	1200	1800	1490	5.0	35.0	8.03	3.540	0.1905	0.0250

FG already corrected these values; correction according to Sketting and Leroy is already accounted for.

Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	0.500	0.0350	0.0100
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	0.600	0.0390	0.0080
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	0.800	0.0580	0.0150
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	1.0	0.0790	0.0160
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	1.500	0.1370	0.0200
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	2.0	0.1820	0.0300
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	2.500	0.2170	0.0300
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	3.0	0.2220	0.0300
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	3.500	0.2820	0.0300
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	4.0	0.3220	0.0400
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	4.500	0.3620	0.0400
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	5.0	0.4220	0.0400
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	5.500	0.4420	0.0300
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	6.0	0.4720	0.0300
Mediterranean Sea	Skretting and Leroy	1966	800	32	1517	13	38	8.15	8.0	0.6020	0.0400
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	0.570	0.0264	0.0022
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	0.720	0.0344	0.0044
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	0.890	0.0514	0.0044
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	1.150	0.0824	0.0044
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	1.400	0.1134	0.0077

Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	1.800	0.1434	0.0077
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	2.280	0.1814	0.0109
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	2.850	0.2244	0.0131
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	3.500	0.2804	0.0241
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	5.600	0.4494	0.0601
Red Sea	Browning	1971	200	280	1536	22	40.5	8.18	8.900	0.8324	0.2406
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	0.400	0.0072	0.0040
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	0.500	0.0122	0.0040
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	0.630	0.0182	0.0040
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	0.790	0.0272	0.0050
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	1.0	0.0412	0.0040
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	1.300	0.0542	0.0050
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	1.650	0.0702	0.0060
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	2.0	0.0922	0.0060
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	2.500	0.1252	0.0070
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	3.200	0.1422	0.0100
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	4.0	0.1532	0.0100
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	5.0	0.2672	0.0300
Gulf of Aden	Browning	1973	300	500	1510	14.31	35.8	7.72	6.0	0.3442	0.1000

FG [2], Table II

NE Pacific	Thorp	1965	753	3000	1479	4.25	34.1	7.67	0.120	0.0011	NaN
NE Pacific	Thorp	1965	753	3000	1479	4.25	34.1	7.67	0.150	0.0018	NaN
NE Pacific	Thorp	1965	753	3000	1479	4.25	34.1	7.67	0.200	0.0026	NaN
NE Pacific	Thorp	1965	753	3000	1479	4.25	34.1	7.67	0.250	0.0035	NaN
NE Pacific	Thorp	1965	753	3000	1479	4.25	34.1	7.67	0.300	0.0062	NaN
NE Pacific	Thorp	1965	753	3000	1479	4.25	34.1	7.67	0.400	0.0094	NaN
Pacific	Lovett (ISAR)	1969	700	245	1481	5.0	34.4	7.67	0.750	0.0221	NaN
Pacific	Lovett (ISAR)	1969	700	245	1481	5.0	34.4	7.67	1.500	0.0547	NaN
Pacific	Lovett (ISAR)	1969	700	245	1481	5.0	34.4	7.67	3.0	0.1470	NaN
Gulf of Alaska	Lovett	1971	75	270	1465	4.0	33.1	7.72	1.500	0.0744	NaN
Gulf of Alaska	Lovett	1971	75	270	1465	4.0	33.1	7.72	2.500	0.1290	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.050	0.0002	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.080	0.0004	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.100	0.0007	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.125	0.0012	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.160	0.0018	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.200	0.0028	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.250	0.0042	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.310	0.0068	NaN
NE Pacific	Morris	1975	505	2900	1476	4.60	34.1	7.69	0.400	0.0092	NaN
S Pacific (Line PA)	Kibblewhite and Denham	1971	1250	1150	1488	4.31	35.0	7.96	0.106	0.0007	NaN
S Pacific (Line PA)	Kibblewhite and Denham	1971	1250	1150	1488	4.31	35.0	7.96	0.212	0.0030	NaN
S Pacific (Line PA)	Kibblewhite and Denham	1971	1250	1150	1488	4.31	35.0	7.96	0.424	0.0083	NaN
S Pacific (Line PB1)	Kibblewhite and Denham	1971	1250	1700	1487	4.07	35.0	7.90	0.106	0.0002	NaN
S Pacific (Line PB1)	Kibblewhite and Denham	1971	1250	1700	1487	4.07	35.0	7.90	0.212	0.0032	NaN
S Pacific (Line PB1)	Kibblewhite and Denham	1971	1250	1700	1487	4.07	35.0	7.90	0.424	0.0084	NaN
S Pacific (KIWI WEST)	Kibblewhite and Denham	1974	1250	3000	1488	4.31	35.0	7.96	0.029	0.0015	NaN
S Pacific (KIWI WEST)	Kibblewhite and Denham	1974	1250	3000	1488	4.31	35.0	7.96	0.060	0.0012	NaN
S Pacific (KIWI WEST)	Kibblewhite and Denham	1974	1250	3000	1488	4.31	35.0	7.96	0.120	0.0021	NaN
S Pacific (KIWI WEST)	Kibblewhite and Denham	1974	1250	3000	1488	4.31	35.0	7.96	0.250	0.0034	NaN

S Pacific (KIWI WEST)	Kibblewhite and Denham	1974	1250	3000	1488	4.31	35.0	7.96	0.424	0.0079	NaN
S Pacific (TAS 1)	Kibblewhite and Denham	1963	1300	900	1489	4.35	35.0	7.87	0.450	0.0203	NaN
S Pacific (TAS 1)	Kibblewhite and Denham	1963	1300	900	1489	4.35	35.0	7.87	0.900	0.0246	NaN
S Pacific (TAS 2 Northwest)	Bannister et al	1975	1350	1800	1488	3.92	35.0	7.87	0.125	0.0016	NaN
S Pacific (TAS 2 Northwest)	Bannister et al	1975	1350	1800	1488	3.92	35.0	7.87	0.250	0.0063	NaN
S Pacific (TAS 2 Northwest)	Bannister et al	1975	1350	1800	1488	3.92	35.0	7.87	0.500	0.0167	NaN
S Pacific (TAS 2 West)	Bannister et al	1975	1100	2800	1484	3.94	35.0	7.87	0.125	0.0012	NaN
S Pacific (TAS 2 West)	Bannister et al	1975	1100	2800	1484	3.94	35.0	7.87	0.250	0.0055	NaN
S Pacific (TAS 2 West)	Bannister et al	1975	1100	1600	1484	3.94	35.0	7.87	0.500	0.0189	NaN
Baffin Bay	Mellen et al	1972	100	400	1442	-1.5	33.7	8.01	0.320	0.0170	NaN
Baffin Bay	Mellen et al	1972	100	400	1442	-1.5	33.7	8.01	0.410	0.0198	NaN
Baffin Bay	Mellen et al	1972	100	400	1442	-1.5	33.7	8.01	0.500	0.0246	NaN
Baffin Bay	Mellen et al	1972	100	400	1442	-1.5	33.7	8.01	0.640	0.0393	NaN
Baffin Bay	Mellen et al	1972	100	400	1442	-1.5	33.7	8.01	1.0	0.0683	NaN
Bismarck Sea	Mellen and Browning	1974	45	NaN	1546	30	36	8.20	0.560	0.0700	NaN
Bismarck Sea	Mellen and Browning	1974	45	NaN	1546	30	36	8.20	1.200	0.1000	NaN
Bismarck Sea	Mellen and Browning	1974	45	NaN	1546	30	36	8.20	2.300	0.1900	NaN
Bismarck Sea	Mellen and Browning	1974	45	NaN	1546	30	36	8.20	4.500	0.3600	NaN

Schneider, from Figure 4, removing 4 points summer 0.8-1.5 kHz; high values attributed to resonance from fish

Baltic	Schneider	1983	50	NaN	NaN	4	8	8	0.5012	0.0379	0.0158
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	0.6321	0.0455	0.0222
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	0.8036	0.0503	0.0143
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	1.0036	0.0422	0.0084
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	1.2555	0.0402	0.0027
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	1.6079	0.0535	0.0130
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	2.0098	0.0583	0.0124
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	2.5080	0.0762	0.0181
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	3.1578	0.0773	0.0110
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	4.0083	0.0962	0.0097
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	5.0304	0.1333	0.0212
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	6.3904	0.1908	0.0198
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	8.1380	0.2733	0.0298
Baltic	Schneider	1983	50	NaN	NaN	4	8	8	9.9838	0.3673	0.0129

REPORT DOCUMENTATION PAGE
(MOD-NL)

1. DEFENCE REPORT NO (MOD-NL)	2. RECIPIENT'S ACCESSION NO	3. PERFORMING ORGANIZATION REPORT NO
TD2008-0078	_	TNO-DV 2008 A202
4. PROJECT/TASK/WORK UNIT NO	**5. CONTRACT NO**	**6. REPORT DATE**
032.11648	-	May 2008
7. NUMBER OF PAGES	**8. NUMBER OF REFERENCES**	**9. TYPE OF REPORT AND DATES COVERED**
63 (incl 3 appendices, excl RDP & distribution list)	-	Final

10. TITLE AND SUBTITLE

Improving an empirical formula for the absorption of sound in the sea

11. AUTHOR(S)

C.A.M van Moll, MSc

Dr M.A. Ainslie

J. Janmaat, BSc

12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TNO Defence, Security and Safety, P.O. Box 96864, 2509 JG The Hague, The Netherlands

Oude Waalsdorperweg 63, The Hague, The Netherlands

13. SPONSORING AGENCY NAME(S) AND ADDRESS(ES)

Defence Research and Development Department of the Netherlands Ministry of Defence.

14. SUPPLEMENTARY NOTES

The classification designation Ongerubriceerd is equivalent to Unclassified, Stg. Confidentieel is equivalent to Confidential and Stg. Geheim is equivalent to Secret.

15. ABSTRACT (MAXIMUM 200 WORDS (1044 BYTE))

A new empirical formula for the absorption of sound in seawater is derived. The starting point of this investigation is the formula proposed by Ainslie and McColm for attenuation due to the boron and magnesium relaxations. Constants in the expression are treated as parameters to be determined. Since this is a nonlinear inverse problem, a global search is used to find the parameter setting that yields the best fit to in situ absorption measurements. The obtained formula is simpler than and at least as accurate as the widely used absorption formula of François and Garrison.

16. DESCRIPTORS IDENTIFIERS

Absorption, attenuation, inverse theory, global search, differential evolution, statistical testing, F-test

17a. SECURITY CLASSIFICATION (OF REPORT)	17b. SECURITY CLASSIFICATION (OF PAGE)	17c. SECURITY CLASSIFICATION (OF ABSTRACT)
Ongerubriceerd	Ongerubriceerd	Ongerubriceerd

18. DISTRIBUTION AVAILABILITY STATEMENT	17d. SECURITY CLASSIFICATION (OF TITLES)
Unlimited Distribution	Ongerubriceerd

Distribution list

1	DMO/SC-DR&D standaard inclusief digitale versie bijgeleverd op cd-rom
2/3	DMO/DR&D/Kennistransfer
4	Programmabegeleider Defensie LTZ1 drs. R. Dekeling, DMO
5	NATO Undersea Research Centre, dr. C.H. Harrison
6	Norwegian Defence Research Establishment, dr. T. Knudsen
7	Department of Underwater Research, FOI Kista, dr. S. Ivansson
8	SME UW Environment, DSTL Winfrith, mrs A. Midona
9/11	Bibliotheek KMA
12	Programmaleider TNO Defensie en Veiligheid, dr. J.C. Sabel
13/14	TNO Defensie en Veiligheid, vestiging Den Haag, Archief
15/19	TNO Defensie en Veiligheid, vestiging Den Haag, Business Unit Waarnemingssystemen, ir. C.A.M van Moll dr. M.A. Ainslie ing. J. Janmaat dr. R. Vossen M. Colin

The following agencies/people will receive the management summary and the distribution list of the report.

4 ex.	DMO/SC-DR&D
1 ex.	DMO/ressort Zeesystemen
1 ex.	DMO/ressort Landsystemen
1 ex.	DMO/ressort Luchtsystemen
2 ex.	BS/DS/DOBBP/SCOB
1 ex.	MIVD/AAR/BMT
1 ex.	Staf CZSK
1 ex.	Staf CLAS
1 ex.	Staf CLSK
1 ex.	Staf KMar
1 ex.	TNO Defensie en Veiligheid, Algemeen Directeur, ir. P.A.O.G. Korting
1 ex.	TNO Defensie en Veiligheid, Directie Directeur Operaties, ir. C. Eberwijn
1 ex.	TNO Defensie en Veiligheid, Directie Directeur Kennis, prof. dr. P. Werkhoven
1 ex.	TNO Defensie en Veiligheid, Directie Directeur Markt, G.D. Klein Baltink
1 ex.	TNO Defensie en Veiligheid, vestiging Den Haag, Manager Waarnemingssystemen (operaties), ir. B. Dunnebier PDeng
1 ex.	TNO Defensie en Veiligheid, vestiging Den Haag, Manager Informatie en Operaties (operaties), ir. P. Schulein
1 ex.	TNO Defensie en Veiligheid, vestiging Rijswijk, daarna reserve Manager Bescherming, Munitie en Wapens (operaties), ir. P.J.M. Elands
1 ex.	TNO Defensie en Veiligheid, vestiging Rijswijk, Manager BC Bescherming (operaties), ir. R.J.A. Kersten
1 ex.	TNO Defensie en Veiligheid, vestiging Soesterberg, Manager Human Factors (operaties), drs. H.J. Vink