SANDIA REPORT
SAND2023-00094
Printed March 2023

Sandia National Laboratories

New Tables for Infrasound Event Catalogs with a Focus on Machine Learning

Sarah Albert, Alexander Witsil[2], and Jeremy Webster[3]
[1]Sandia National Laboratories, Albuquerque, NM, USA
[2]Applied Research Associates, Greenville, SC, USA
[3]Los Alamos National Laboratory, Los Alamos, NM, USA

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Sandia National Laboratories
U.S. Department of Energy NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@osti.gov
Online ordering:	http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.gov
Online order:	https://classic.ntis.gov/help/order-methods/

ABSTRACT

Natural and anthropogenic sources such as volcanoes, earthquakes, auroral processes, chemical and nuclear explosions, rocket launches, and aircraft can generate infrasound, sound with frequencies less than 20 Hz. Both the availability of infrasound data and interest in machine learning (ML) applications have grown in recent years. Large, open-source datasets are essential to solving complex ML problems, but the field of infrasound is lacking in this arena. To increase the utility of ML for infrasound, here we present new tables for infrasound event catalogs. It is the aim that these tables be incorporated into both existing and future infrasound processing pipelines to generate large datasets ripe for use with ML/DL methods.

ACKNOWLEDGEMENTS

The authors would like to thank researchers who participated in candid discussions around the utility of ML for infrasound as well as best practices for generating infrasound event catalogs. Specifically, we thank: Sam Chodur, Elizabeth Berg, Jordan Bishop, Phil Blom, Fransiska Dannemann Dugick, David Fee, Mark Hale, Justin Hertzog, Clint Koch, Lisa Linville, Sarah McComas, David Norris, Kristine Pankow, Julie Schnurr, Katrina Smart, Richard Stead, Liam Toney, and Jesse Torres.

CONTENTS

LIST OF TABLES

This page left blank

ACRONYMS AND TERMS

Acronym/Term	Definition
CSS3.0	Center for Seismic Studies Version 3.0
DL	Deep learning
fk	Frequency-wavenumber
IDC	International Data Centre
IMS	International Monitoring System
KBCore	Knowledge Base Core

1. INTRODUCTION

1.1. Background

Sound less than 20 Hz, known as infrasound, is generated by both natural and anthropogenic sources including: volcanoes, earthquakes, auroral processes, chemical and nuclear explosions, rocket launches, and aircraft. These sources exert a force on the atmosphere, generating infrasonic waves that can travel up to thousands of kilometers due to their low frequency nature. Infrasound propagation largely depends on the wind and temperature structure of the atmosphere and waveforms may contain information about both the source and the wave's path.

The availability of infrasound data has grown exponentially in recent years, suggesting increased interest in both infrasound data collection and collaboration. Machine learning (ML) and deep learning (DL) have become increasingly popular across all science disciplines, and the field of infrasound is no exception. The advantages of using ML/DL models are that they may identify patterns in data that cannot be resolved using physics models alone. However, because ML/DL models do not always take physics into account, it is especially important that data are adequately curated. Data sets must be large enough and contain enough variation for models to capture patterns. There exist regional and global seismic monitoring centers that make generating large datasets possible, where analysts perform quality control on waveforms and automated processes form events. The field of infrasound lacks such dedicated centers that automatically process data, perform analyst quality control, and publish event bulletins. However, efforts should be made to establish them as this would increase the availability of data for ML/DL problems. To aid in this effort, this paper outlines best practices for generating infrasound event catalogs. It is the aim that these best practices be incorporated into both existing and future infrasound processing pipelines to generate large datasets ripe for use with ML/DL methods.

2. NEW TABLES FOR INFRASOUND EVENT CATALOGS

Many institutions processing infrasound data use some flavor of standardized database schemas originally designed for use with seismic data. These include the Center for Seismic Studies Version 3.0 (CSS3.0; Anderson et al., 1990) and the National Nuclear Security Administration Knowledge Base Core (KBCore; Carr, 2002) database schemas. However, there is no standard way to implement CSS3.0 or KBCore for use with infrasound data, which leads to differences in databases between data centers. This also makes it difficult for institutions that are in the process of standing up an infrasound processing pipeline. Therefore, the aim of this section is to establish similarity across infrasound studies by standardizing database tables. While the specific focus of implementing standardization is to benefit ML/DL applications, this information will be relevant for a variety of analytical methods across the field of infrasound.

CSS3.0 is a widely used and general data schema and will therefore be used as the baseline for this report. Instead of changing existing CSS3.0 tables, we recommend using existing CSS3.0 tables where possible and adding new tables for infrasound data processing (described below). This will streamline the standardization process and have the added benefit of allowing the schema to be used for seismoacoustic studies as well.

Following the convention of Anderson (1990) and Carr (2002), this paper lists table title, a short table description, columns, ORACLE storage type, and external format and character positions for flat files. For descriptions of existing CSS3.0 tables, the reader is referred to Anderson (1990).

2.1. Table Descriptions

2.1.1. fkdisc

Frequency-wavenumber (fk) processing is often applied to infrasound array data to determine signal direction of arrival and trace velocity. The fkdisc table contains information about fk processing.

This table is based on a table with the same name in the version 5.1.1 database schema used by the International Data Centre (IDC), which processes International Monitoring System (IMS) data for nuclear test verification (IDC User Guides Documentation).

Table 2.1. fkdisc

FIELD NUMBER	COLUMN	STORAGE TYPE	EXTERNAL FORMAT	CHARACTER POSITION	DESCRIPTION
1	*jdate*	number(8)	i8	1-8	julian date
2	*time*	float(53)	f17.5	10-26	epoch time of most recent processing
3	*wlen*	float(24)	f10.3	28-37	duration of measurement window
4	*sta*	varchar2(6)	a6	39-44	station identifier
5	*fktyp*	varchar2(4)	a4	46-49	f-k type
6	*arid*	number(8)	i8	51-58	arrival identifier
7	*maxkx*	float(24)	f7.4	60-66	maximum x-wavenumber

8	*maxsx*	float(24)	f7.4	68-74	maximum x-slowness
9	*nx*	number(4)	i4	76-79	number of x samples
10	*maxky*	float(24)	f7.4	81-87	maximum y-wavenumber
11	*maxsy*	float(24)	f7.4	89-95	maximum y-slowness
12	*ny*	number(4)	i4	97-100	number of y-samples
13	*fmin*	float(24)	f7.2	102-108	minimum frequency of filter band
14	*fmax*	float(24)	f7.2	110-116	maximum frequency of filter band
15	*commid*	number(8)	i8	118-125	comment identification
19	*lddate*	date	a17	127-143	load date

2.1.2. infra

This table contains information describing infrasound signal arrivals that are not captured in traditional CSS3.0 tables. For example, recording the maximum amplitude value and time can help analysts determine if the arrival is emergent, which has implications for infrasound signal characterization. In addition, event size is reported in CSS3.0 as magnitude, but here we provide columns for yield. A unique addition to this table includes the columns *addfile_dir* and *addfile_dfile*. These columns provide the path to a file that can be used to store information regarding machine learning architectures, testing/training datasets, data processing inputs, etc. This file should aid in open data sharing, increasing the utility of datasets for ML/DL applications.

Some entries in this table are based on the *infra_features* table in the version 5.1.1 database schema used by the IDC (IDC User Guides Documentation).

Table 2.2. infra

FIELD NUMBER	COLUMN	STORAGE TYPE	EXTERNAL FORMAT	CHARACTER POSITION	DESCRIPTION
1	*arid*	number(9)	i9	1-9	arrival identifier
2	*coh_time*	float(53)	f17.5	11-27	coherence start time
3	*coh_dur*	float(24)	f7.2	29-35	coherence duration
	coh_thresh	float(24)	f7.2	37-43	coherence threshold
4	*ford*	number(8)	i8	45-52	filter order
7	*mamp*	float(24)	f7.3	54-60	maximum amplitude
8	*mamp_time*	float(53)	f17.5	62-78	time of maximum amplitude
9	*yield*	float(24)	f7.2	80-86	yield
10	*y_method*	varchar2(16)	a16	88-103	yield estimation method

11	*atmos*	varchar2(16)	a16	105-120	atmospheric profile
12	*atmos_time*	float(53)	f17.5	122-138	atmospheric profile time
14	*auth*	varchar2(15)	a15	140-154	author
15	*commid*	number(9)	i9	156-164	comment identifier
16	*lddate*	date	a17	166-182	load date
17	*addfile_dir*	varchar2(64)	a64	184-247	path to file containing additional information
18	*addfile_dfile*	varchar2(64)	a64	249-312	filename

2.1.3. beam

In addition to fk processing, beamforming is also commonly used to determine signal direction of arrival and trace velocity. Therefore, the beam table includes information describing beams generated from accompanying data. To keep with standard convention, beams are described similarly to the way in which waveforms are described (i.e., backazimuth, slowness, etc.).

Table 2.3. beam

FIELD NUMBER	COLUMN	STORAGE TYPE	EXTERNAL FORMAT	CHARACTER POSITION	DESCRIPTION
1	*beamid*	number(8)	i8	1-8	beam identifier
2	*btype*	varchar2(1)	i1	10-10	beam domain type, t (time) or frequency (f)
3	*coh*	varchar2(1)	i1	12-12	coherence flag, c (coherent) or i (incoherent)
4	*baz*	float(24)	f7.2	14-20	beam backazimuth
5	*slow*	float(24)	f7.2	22-28	beam slowness
6	*ftype*	varchar2(16)	a16	30-45	filter type applied to beam
7	*fmin*	float(24)	f7.2	47-53	minimum filter frequency
8	*fmax*	float(24)	f7.2	55-61	maximum filter frequency
9	*auth*	varchar2(15)	a15	63-77	author
10	*commid*	number(9)	i9	79-87	comment identifier
11	*lddate*	date	a17	89-105	load date

2.1.4. *beamassoc*

This table associates a beam to the station, channel, and instrument that was used to generate it.

Table 2.4. beamassoc

FIELD NUMBER	COLUMN	STORAGE TYPE	EXTERNAL FORMAT	CHARACTER POSITION	DESCRIPTION
1	*beamid*	number(8)	i8	1-8	beam identifier
2	*sta*	varchar2(6)	a6	10-15	station identifier
3	*chan*	varchar2(8)	a8	17-24	channel code
4	*chanid*	number(8)	i8	26-33	instrument identifier

2.2. Column Descriptions

Here we describe columns not listed in previous work. For descriptions of columns not listed here, the reader is referred to Anderson (1990) and Carr (2002).

Name	***atmos***	
Table	infra	
Description	Model used to generate atmospheric profile used in data processing. May be related to yield calculation and/or detection/association/location methods. Examples: G2S, HRRR, GFS	
Format	varchar2(16)	a16
Unit	-	
NA Value	-	
Range	any string up to column size	

Name	***atmos_time***	
Table	infra	
Description	Epoch time corresponding to the model used to generate the atmospheric profile used in data processing, defined by *atmos*. Epoch time is given as seconds and fractions of a second since hour 0 January 1, 1970, and stored in a double-precision floating number.	
Format	float(53)	f17.5
Unit	s	
NA Value	-9999999999.999	
Range	*mamp_time* > -9999999999.999	

Name	***baz***	
Table	beam	
Description	Value identifying backazimuth of the beam in the station to source direction.	
Format	float(24)	f7.2
Unit	degree	
NA Value	-1	
Range	0 ≤ *baz* ≤ 360	

Name	***beamid***	
Table	beam	
Description	Beam identifier. Each beam is assigned a unique positive integer identifying it with a unique *sta*, *chan*, and *time*.	
Format	number(8)	i8
Unit	-	
NA Value	-	
Range	any string up to column size	

Name	***btype***	
Table	beam	
Description	Value identifying domain of the beam used in processing. Options: t (time) or f (frequency)	
Format	varchar2(1)	i1
Unit	-	
NA Value	-	
Range	{t, f}	

Name	***coh***	
Table	beam	
Description	Value identifying coherence of the beam used in processing. Options: c (coherent) or i (incoherent).	
Format	varchar2(1)	i1
Unit	-	
NA Value	-	
Range	{c, i}	

Name	***coh_dur***
Table	infra
Description	Duration that the spatial coherence exceeds the threshold defined by *coh_thresh*.
Format	float(24) External: f7.2
Unit	s
NA Value	-1
Range	*coh_dur* > 0.0

Name	***coh_thresh***
Table	infra
Description	Value defining the threshold for a coherent signal.
Format	float(24) External: f7.2
Unit	-
NA Value	-1
Range	0.0 < *coh_thresh* ≤ 1.0

Name	***coh_time***
Table	infra
Description	Epoch time corresponding to the time at which the spatial coherence exceeds the threshold defined by *coh_thresh*. Epoch time is given as seconds and fractions of a second since hour 0 January 1, 1970, and stored in a double-precision floating number.
Format	float(53) External: f17.5
Unit	s
NA Value	-9999999999.999
Range	*coh_time* > -9999999999.999

Name	***fktyp***
Table	fkdisc
Description	String identifying f-k spectrum type. Examples: monochromatic (mono), broadband (broa).
Format	varchar2(4) External: a4
Unit	-
NA Value	-

Range	lowercase string up to column size	

Name	***fmax***	
Table	fkdisc	
Description	Maximum frequency of filter band applied to data before computing the f-k spectrum.	
Format	float(24)	External: f7.2
Unit	Hz	
NA Value	-1	
Range	*fmax* > *fmin*	

Name	***fmin***	
Table	fkdisc	
Description	Minimum frequency of filter band applied to data before computing the f-k spectrum.	
Format	float(24)	External: f7.2
Unit	Hz	
NA Value	-1	
Range	0 < *fmin* < *fmax*	

Name	***ford***	
Table	infra	
Description	Value defining the filter order used. Filter type is defined by the *ftype* column in the *beam* table.	
Format	number(8)	External: i8
Unit	-	
NA Value	0	
Range	*ford* > 0	

Name	***ftype***	
Table	beam	
Description	Filter type applied to the beam. Additional descriptive values (*fmin, fmax*) can be found in the *infra* table. Examples: Butterworth, highpass, lowpass.	
Format	varchar2(16)	a16
Unit	-	

NA Value	-	
Range	any string up to column size	

Name	***addfile_dfile***	
Table	infra	
Description	Filename corresponding to file that contains additional processing information.	
Format	varchar2(64)	a64
Unit	-	
NA Value	-	
Range	any string up to column size	

Name	***addfile_dir***	
Table	infra	
Description	File path for file that contains additional processing information. This file can be used to store information regarding machine learning architectures, data processing inputs, etc.	
Format	varchar2(64)	a64
Unit	-	
NA Value	-	
Range	any string up to column size	

Name	***mamp***	
Table	infra	
Description	Measured maximum signal amplitude.	
Format	float(24)	External: f7.3
Unit	Pa	
NA Value	-999	
Range	*mamp* > -999	

Name	***mamp_time***	
Table	infra	
Description	Epoch time corresponding to the time at which the maximum amplitude of the signal occurs. Epoch time is given as seconds and fractions of a second since hour 0 January 1, 1970, and stored in a double-precision floating number.	
Format	float(53)	f17.5

Unit	s
NA Value	-9999999999.999
Range	*mamp_time* > -9999999999.999

Name	***maxkx***	
Table	fkdisc	
Description	X-axis maximum wavenumber in the f-k spectrum. Assumed to be symmetric, ranging from -*maxkx* to *maxkx*.	
Format	float(24)	External: f7.4
Unit	km^{-1}	
NA Value	-1	
Range	*maxkx* > 0.0	

Name	***maxky***	
Table	fkdisc	
Description	Y-axis maximum wavenumber in the f-k spectrum. Assumed to be symmetric, ranging from -*maxky* to *maxky*.	
Format	float(24)	External: f7.4
Unit	km^{-1}	
NA Value	-1	
Range	*maxky* > 0.0	

Name	***maxsx***	
Table	fkdisc	
Description	X-axis maximum slowness in the f-k spectrum. Assumed to be symmetric, ranging from -*maxsx* to *maxsx*.	
Format	float(24)	External: f7.4
Unit	s/km	
NA Value	-1	
Range	*maxsx* > 0.0	

Name	***maxsy***
Table	fkdisc
Description	Y-axis maximum slowness in the f-k spectrum. Assumed to be symmetric, ranging from -*maxsy* to *maxsy*.

Format	float(24)	External: f7.4
Unit	s/km	
NA Value	-1	
Range	*maxsy* > 0.0	

Name	***nx***	
Table	fkdisc	
Description	Number of sample points along the x-axis in the f-k spectrum. Sample points correspond to slowness or wavenumber, depending on *fktyp*.	
Format	number(4)	External: i4
Unit	-	
NA Value	-1	
Range	*nx* > 0	

Name	***ny***	
Table	fkdisc	
Description	Number of sample points along the y-axis in the f-k spectrum. Sample points correspond to slowness or wavenumber, depending on *fktyp*.	
Format	number(4)	External: i4
Unit	-	
NA Value	-1	
Range	*ny* > 0	

Name	***wlen***	
Table	fkdisc	
Description	Length of time window used in f-k processing.	
Format	float(24)	External: f10.3
Unit	sec	
NA Value	-1	
Range	*wlen* > 0.0	

Name	***y_method***
Table	infra
Description	Method used to calculate yield from infrasound signal. Examples: ANSI, BOOM

Format	varchar2(16)	a16
Unit	-	
NA Value	-	
Range	any string up to column size	

Name	***yield***	
Table	infra	
Description	Value corresponding to infrasound yield calculated using the method defined by *y_method*.	
Format	float(24)	f7.2
Unit	kg	
NA Value	-1	
Range	*yield* > 0.0	

REFERENCES

[1] Anderson, J., Farrell, W.E., Garcia, K., Given, J., and Swanger, H., 1990. Center for Seismic Studies Version 3 Database Schema Reference Manual, Technical Report SAIC-90/1235 C90-01, *Science Applications International Corporation*, San Diego, CA.

[2] Carr, D.B., 2002. National Nuclear Security Administration Knowledge Base Core Table Schema Document, Technical Report SAND2002-3055, *Sandia National Laboratories*, Albuquerque, NM.

[3] IDC User Guides Documentation, *https://swp.ctbto.org/web/swp/manuals.* Database Schema, IDC 5.1.1 (latest available revision).

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Daniel Gonzales	6752	dgonza2@sandia.gov
Jay Brotz	6831	jkbrotz@sandia.gov
Technical Library	1911	sanddocs@sandia.gov

Email—External

Name	Company Email Address	Company Name
Alex Witsil	awitsil@ara.com	Applied Research Associates
Jeremy Webster	jwebster@lanl.gov	Los Alamos National Laboratory
Kristine Pankow	Kris.Pankow@utah.edu	University of Utah

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.