SAND2023-05012C

SnT 2023
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE
HOFBURG PALACE - Vienna and Online
19 TO 23 JUNE

International Monitoring System 3-component seismic signal detection using the PhaseNet deep learning model

Stephen Heck, Jorge Garcia, Rigobert Tibi
Sandia National Laboratories

INTRODUCTION

Producing a complete and accurate set of signal detections is essential for automatically building and characterizing seismic events of interest for nuclear explosion monitoring. We explore training PhaseNet, a deep learning model using three-component sensor data from the IMS.

METHODS/DATA

We constructed five IMS training data sets by varying the ratio of noise windows to signal windows and compared the IMS-trained models to models trained from 8 non-IMS datasets available for download with SeisBench.

START

RESULTS

The IMS-trained models and SeisBench models were comparable on recall, but the IMS-trained models had 10x fewer false positives. Two of SeisBench models trained on regional signals performed well on teleseismic signals.

CONCLUSION

We found the primary advantage of training with IMS data over using the SeisBench models, was the suppression of the false detections on noise windows.

This Ground-based Nuclear Detonation Detection (GNDD) research was funded by the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development (NNSA DNN R&D).

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.

Please do not use this space, a QR code will be automatically overlayed

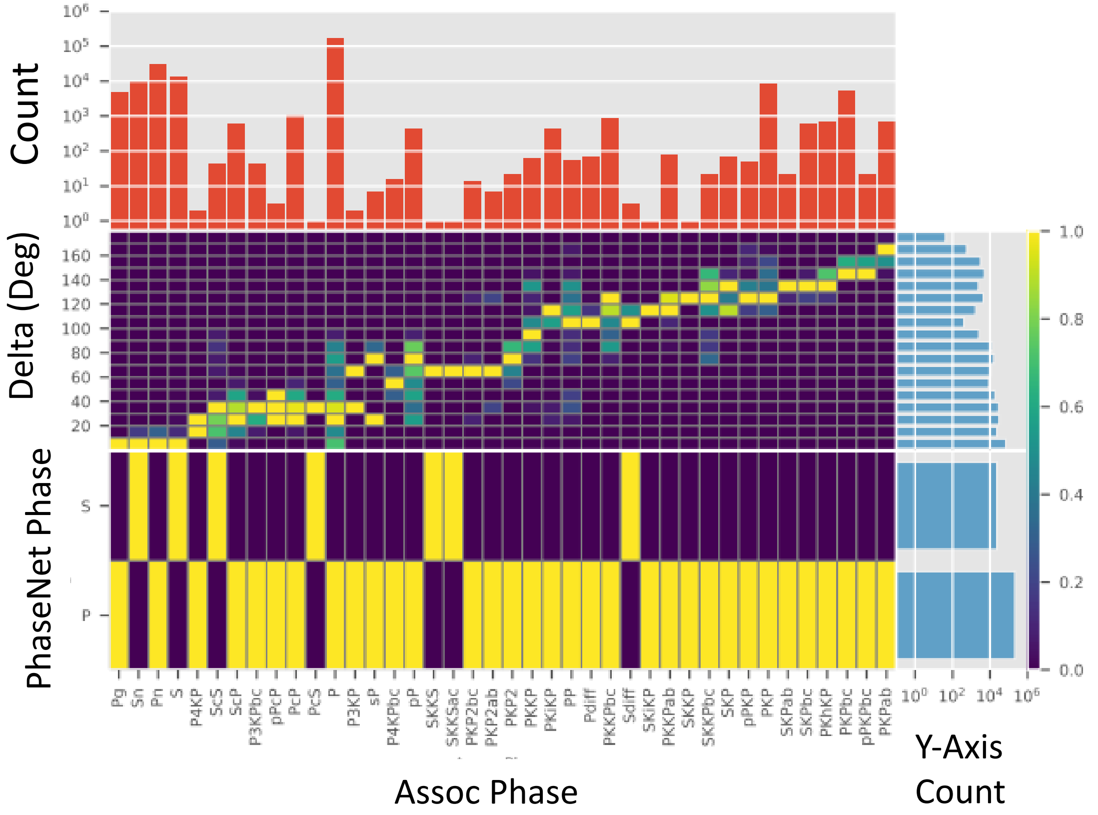
P3.5-365

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.

Introduction

We trained the deep learning model, PhaseNet, for seismic detection on 3-component stations from the International Monitoring System (IMS), and evaluated the results using the Unconstrained Global Event Bulletin (UGEB) [5]. Using 14 years of associated signals from the Late Event Bulletin (LEB), we auto-curated a training data set consisting of signal windows containing associated arrivals, and noise windows that contain no LEB associated signals. We constructed five training data sets by varying the ratio of noise windows to signal windows and found that increasing the number of noise windows increases the precision from .15 to .4 while reducing the recall from .6 to .5. Using the SeisBench Toolbox [2], we compared eight PhaseNet models trained on non-IMS data on the UGEB. The best SeisBench model achieved a .24 F1 score versus a .49 F1 score for our best IMS-models. However, we found that the primary benefit of training with LEB data is not in detecting more signals, but rather the suppression of noise detections.

IMS LEB Associated Phases Training Dataset



INTRODUCTION
OBJECTIVES
METHODS/DATA
RESULTS
CONCLUSION

Objectives

- Are PhaseNet models trained on non-IMS data transportable to IMS?

- Methods for building *effective* IMS training datasets:
    - How many samples are needed?
    - How does the ratio of noise windows to associated signal windows affect precision and recall?
    - What SNR threshold should be used for P and S phases?

- Characterize the relationship between signal SNR and PhaseNet's response.

- Identify technical gaps between seismic deep learning research and operational monitoring systems

Methods: Datasets and Experimental Setup

- IMS Dataset is auto-curated by thresholding waveforms by their estimated SNR using STA/LTA.
- Late Event Bulletin (LEB) was used to identify associated waveforms
- Multiple bandpass filters were applied to associated and non-associated waveform windows to identify *good* waveforms.
  - We removed associated windows *below* a minimum SNR of 4
  - We removed unassociated waveforms *above* a maximum SNR of 3
- Training Dataset Naming Convention:

LEB K *(ratio of unassociated waveforms to associated waveforms in %)*

IMS Training and Evaluation Datasets

Name	P	S	Noise	Total P	Total S	Total P+S	Total Noise
LEB 100	216611	23866	286942	673350	80584	753934	753934
LEB 80	216611	23866	275368	673350	80584	753934	603147
LEB 20	216611	23866	123926	673350	80584	753934	150787
LEB 0	216611	23866	0	673350	80584	753934	0
UGEB	3682	498	131747	14728	1992	16720	526988

Evaluation Dataset (see next slide)

"SeisBench" PhaseNet Baseline models [2]

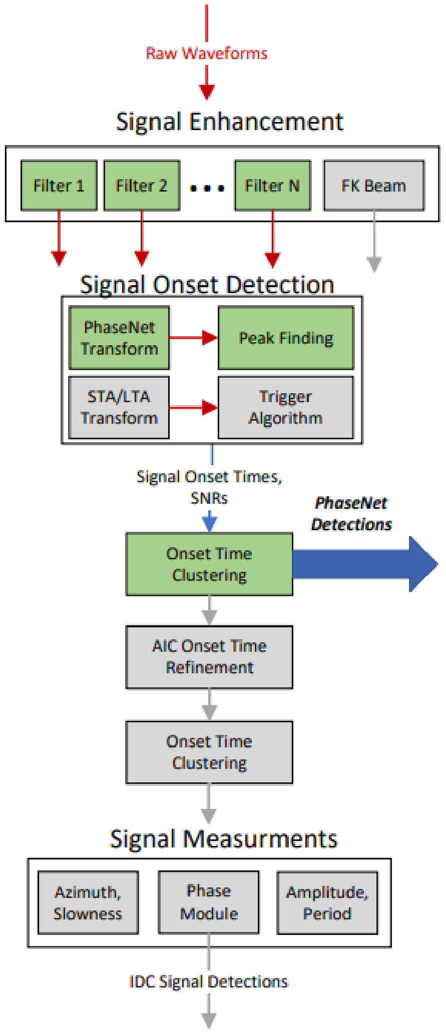
Name	Region	Source-Receiver Dist
STEAD	global	$\leq$ 350 km
ETHZ	Switzerland	$\leq$ 10 °
SCEDC	Southern California	$\leq$ 200 km
NEIC	global	$\geq$ 10 °
GEOFON	global	$\geq$ 10
Iquique	Chile	$\leq$ 10 °
LenDB	global	$\leq$ 120 km
INSTANCE	Italy	$\leq$ 600 km

We tested the "SeisBench" and LEB PhaseNet models on IMS associated signals from the UGEB using the our *Evaluation Framework*

Evaluation Framework

We used PhaseNet as a replacement for the STA/LTA transform in a processing pipeline similar to the IDC's signal detection software. Green boxes highlight our processing steps. Gray boxes indicate additional processing steps in the *traditional* signal detection pipeline. Parameters used in this study:

Parameter	Value
Bandpass Filter 1	[1Hz, 2.5Hz]
Bandpass Filter 2	[1.5Hz, 3Hz]
Bandpass Filter 3	[2Hz, 5Hz]
Bandpass Filter 4	[3Hz, 6Hz]
Peak Finding Width	4 sec
Min Peak Thresholds	.1, .2, .4, .6, .8
Onset Time Cluster Threshold	4 sec

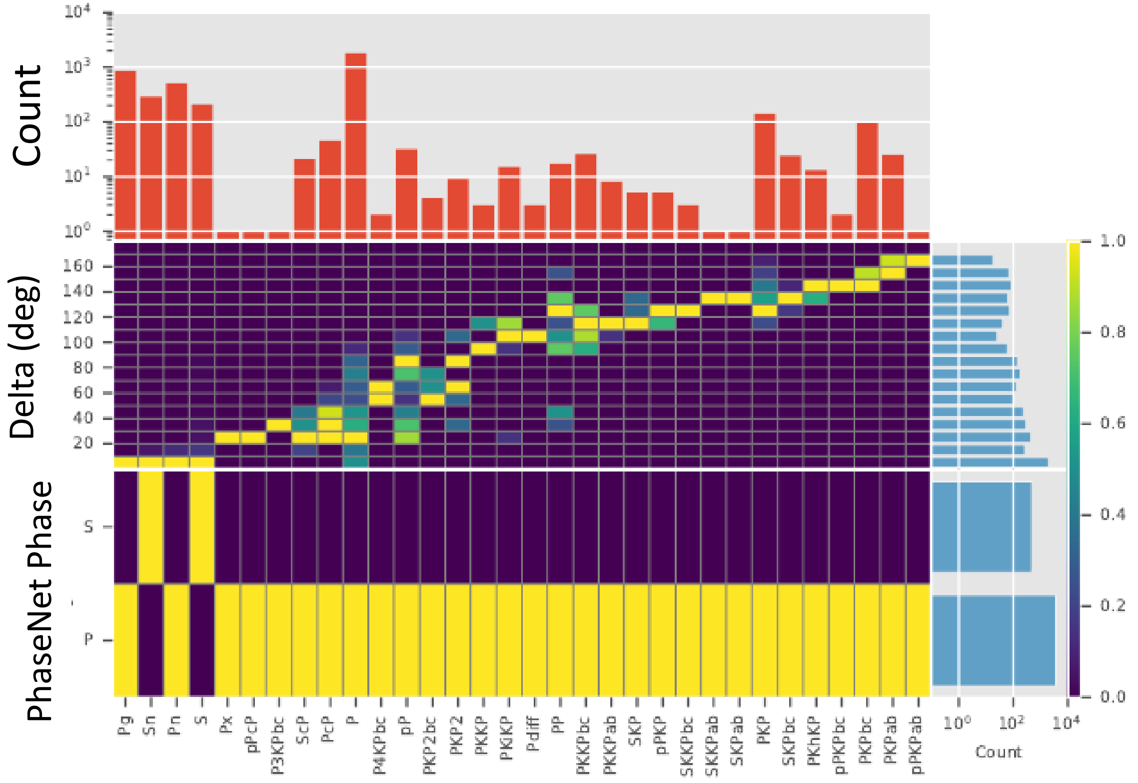


INTRODUCTION
OBJECTIVES
METHODS/DATA
RESULTS
CONCLUSION

Methods: Evaluation Dataset
Unconstrained Global Event Bulletin (UGEB) [3]

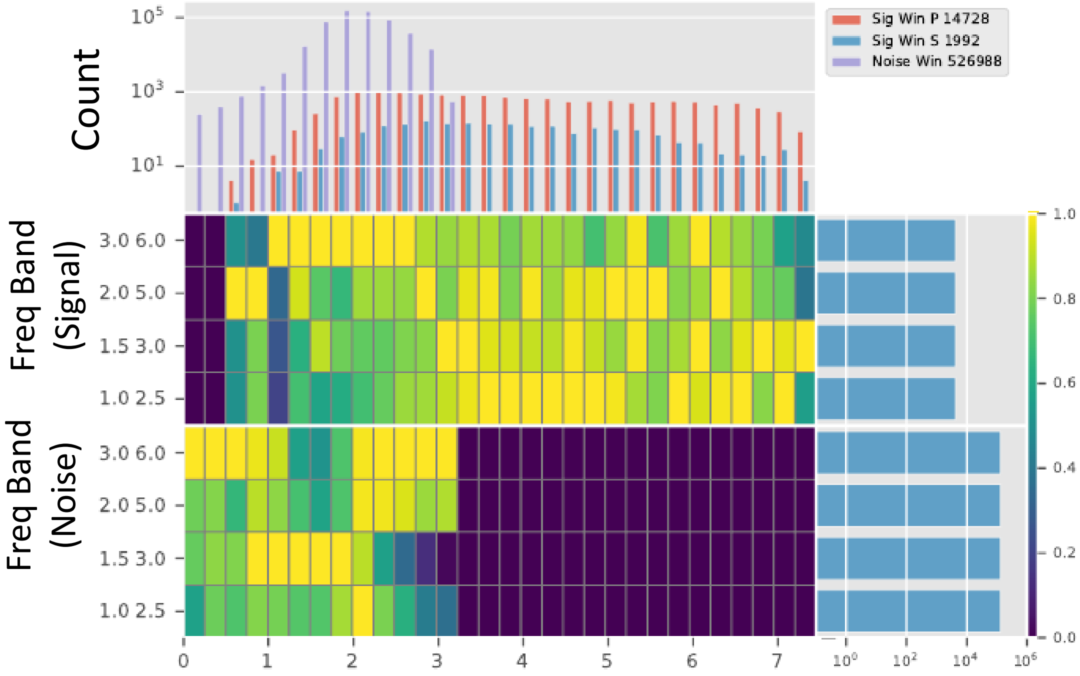
The UGEB spans May 15th 2010 – May 29th 2010 and consists of approximately 11,000 human-analyst-built events. The UGEB was constructed by first starting with LEB's 1650 events from the same time-period. Note that our LEB training datasets excluded waveforms from this two-week time period.

Phase Statistics



Associated Phase

Input Window Statistics



SNR (STA/LTA)

Freq Band Count

Each heat-map bin is normalized by the max element

INTRODUCTION
OBJECTIVES
METHODS/DATA
RESULTS
CONCLUSION

Sandia National Laboratories
U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration

Results: Precision and Recall

Precision and Recall versus Detection Threshold

Cell format
Precision, Recall, F1
False Count, True

IMS Models

model	PNR≥0.1	PNR≥0.2	PNR≥0.4	PNR≥0.6	PNR≥0.8
LEB 100	0.39 0.51 0.44 3289 2136	**0.55 0.43 0.49** 1468 1810	0.75 0.32 0.45 448 1346	0.86 0.20 0.32 130 825	0.96 0.05 0.10 9 209
LEB 100 SD	0.30 0.51 0.38 5054 2151	0.50 0.46 0.48 1910 1910	**0.73 0.36 0.49** 548 1517	0.89 0.25 0.39 133 1047	0.96 0.12 0.22 19 506
LEB 80	0.26 0.57 0.36 6836 2403	0.45 0.49 0.47 2494 2068	**0.73 0.37 0.49** 580 1555	0.87 0.25 0.39 155 1055	0.95 0.12 0.21 26 494
LEB 20	0.20 0.58 0.30 9601 2413	0.44 0.49 0.46 2613 2051	**0.73 0.37 0.49** 558 1526	0.86 0.25 0.39 172 1051	0.94 0.11 0.20 33 478
LEB 0	0.15 0.60 0.24 13813 2494	0.40 0.49 0.44 3063 2039	0.73 0.35 0.48 551 1482	0.86 0.26 0.40 172 1072	0.93 0.13 0.23 41 558

SeisBench Models

model	PNR≥0.1	PNR≥0.2	PNR≥0.4	PNR≥0.6	PNR≥0.8
ETHZ	0.01 0.62 0.02 284875 2586	0.01 0.56 0.03 170264 2325	0.03 0.45 0.05 69196 1883	0.06 0.33 0.11 20261 1398	0.31 0.17 0.22 1592 700
GEOFON	0.01 0.44 0.02 214340 1844	0.02 0.30 0.03 67997 1242	0.07 0.17 0.10 9893 718	0.28 0.09 0.14 969 374	0.50 0.01 0.02 47 47
INSTANCE	0.02 0.53 0.04 113316 2212	0.04 0.42 0.08 38908 1755	0.15 0.28 0.19 6962 1182	**0.41 0.17 0.24** 1053 730	0.77 0.07 0.12 80 275
Iquique	0.01 0.52 0.03 157733 2172	0.02 0.45 0.04 82038 1898	0.05 0.35 0.09 26966 1466	0.12 0.25 0.16 7476 1037	0.39 0.12 0.18 786 495
LenDB	0.02 0.28 0.04 60600 1167	0.04 0.18 0.07 18023 762	0.11 0.09 0.10 3061 374	0.19 0.03 0.05 566 129	0.34 0.01 0.02 73 38
NEIC	0.01 0.62 0.02 286193 2596	0.01 0.43 0.03 124767 1786	0.03 0.12 0.05 15733 513	0.06 0.00 0.01 231 16	N/A
SCEDC	0.01 0.58 0.01 372568 2416	0.01 0.46 0.02 223139 1917	0.01 0.27 0.03 80255 1120	0.03 0.12 0.04 19652 516	0.16 0.03 0.05 741 138
STEAD	0.01 0.62 0.02 261565 2585	0.02 0.53 0.03 136569 2231	0.04 0.39 0.07 40764 1618	0.09 0.20 0.13 8159 853	0.40 0.05 0.09 294 200

Notice the number of true positives remains high relative to other SeisBench models. See the next slide for a visualization.

True Positives by Phase

Cell format
True Count, Recall

model	P 1807	Pg 861	Pn 508	Sn 289	S 209	PKP 141	PKPbc 97	PcP 45
LEB 100	990 55%	454 53%	311 61%	104 36%	87 42%	61 43%	50 52%	18 40%
LEB 80	1074 59%	523 61%	327 64%	139 48%	116 56%	65 46%	57 59%	21 47%
LEB 20	1062 59%	527 61%	**328 65%**	155 54%	124 59%	62 44%	57 59%	21 47%
LEB 0	1111 61%	539 63%	**331 65%**	155 54%	123 59%	67 48%	56 58%	27 60%
ETHZ	**1150 64%**	**570 66%**	320 63%	152 53%	117 56%	**74 52%**	62 64%	**30 67%**
GEOFON	947 52%	333 39%	215 42%	56 19%	57 27%	64 45%	58 60%	23 51%
INSTANCE	1020 56%	485 56%	282 56%	112 39%	100 48%	53 38%	52 54%	22 49%
Iquique	915 51%	517 60%	293 58%	155 54%	114 55%	49 35%	43 44%	18 40%
LenDB	511 28%	269 31%	179 35%	51 18%	49 23%	32 23%	25 26%	10 22%
NEIC	**1161 64%**	561 65%	307 60%	177 61%	**132 63%**	69 49%	60 62%	26 58%
SCEDC	1048 58%	516 60%	292 57%	175 61%	130 62%	71 50%	53 55%	27 60%
STEAD	1127 62%	**570 66%**	307 60%	**186 64%**	130 62%	68 48%	**66 68%**	21 47%

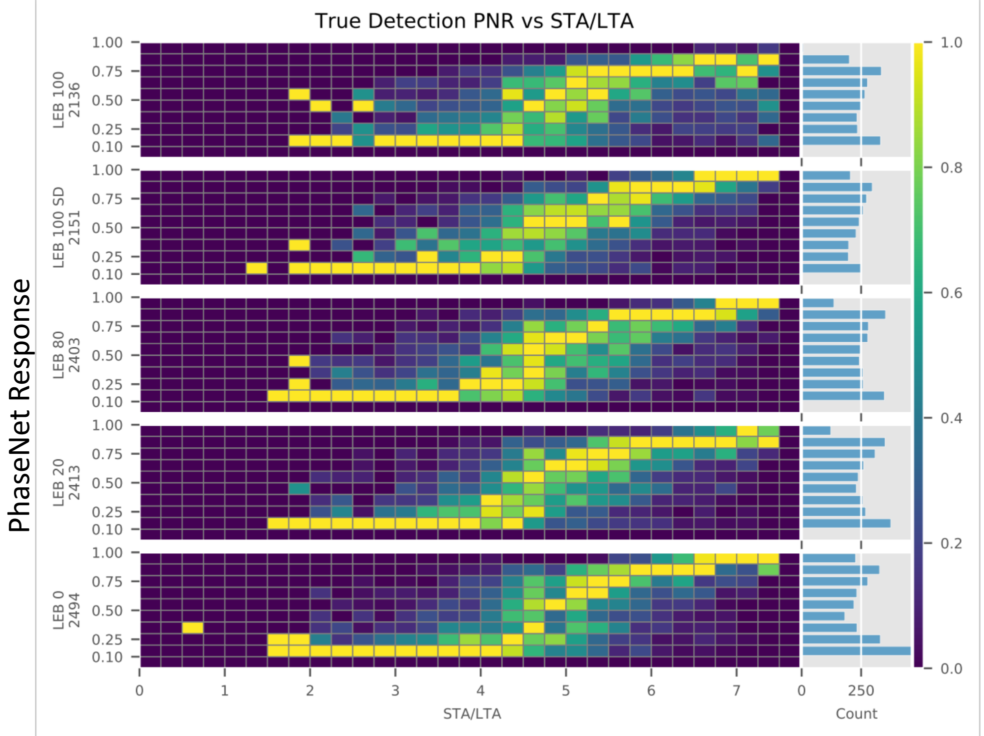
Unanticipated Result:
ETHZ and STEAD SeisBench models did well on teleseismic signals with respect to the LEB Models.

Results: PhaseNet Response versus Signal SNR for True Positives

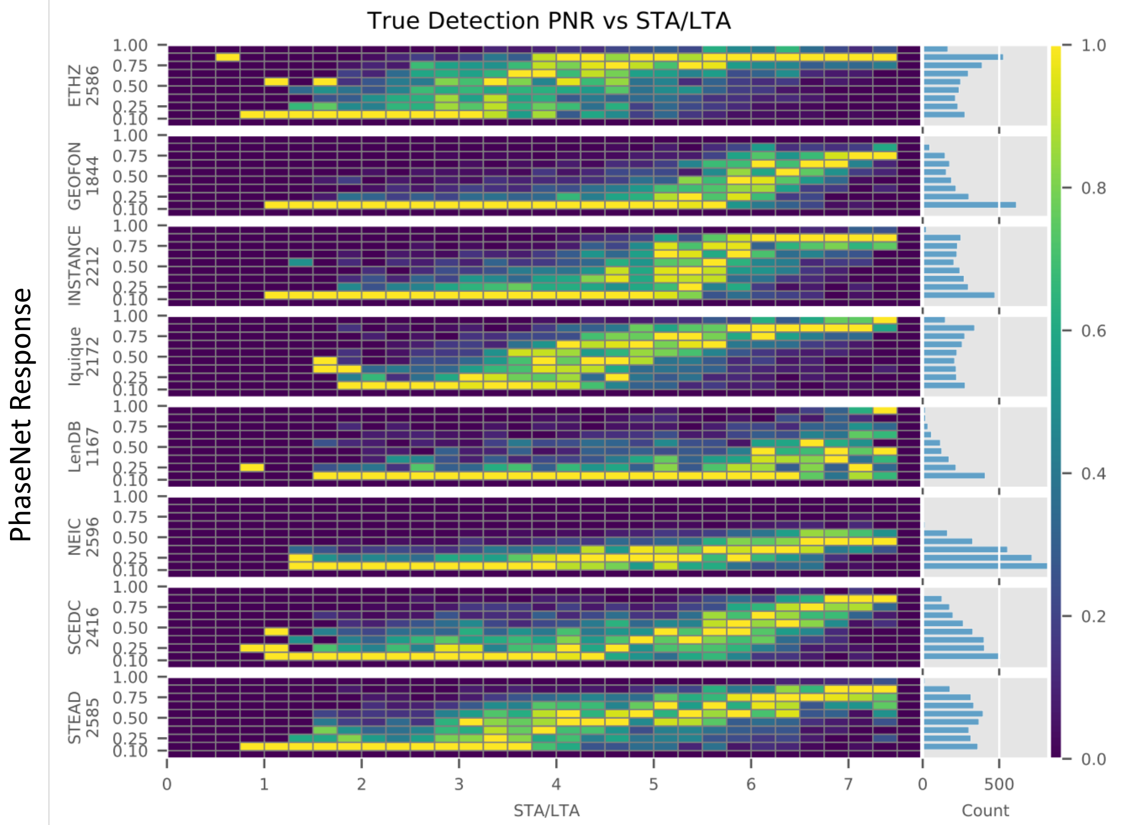
PhaseNet Response to signal SNR for True Positives Detections

Each heat-map bin is normalized by the max element

LEB PhaseNet Models



SeisBench Models



PhaseNet Response trends higher and the response spread decreases as the STA/LTA response increases. Stated in another way, as the signals become easier to detect, the PhaseNet response tends to have less spread, indicating the model is more consistent with respect to higher SNR signals. Notice the lower PhaseNet response for STA/LTA<=4. Our training dataset signal STA/LTA threshold was 4.

Seisbench models ETHZ, GEOFON, Iquique, SCEDC, and STEAD had response curves similar to that of the LEB models, however there are notable differences. We note that these SeisBench models have higher PNR values for lower STA/LTA response signals (compare the STA/LTA bins 2 to 3 for ETHZ with any of the LEB models). We qualitatively rank the ETHZ model as having the best PhaseNet response due to its higher PNR mass and less spread for STA/LTA bins >= 4, and its higher PNR mass for STA/LTA <= 4 relative to the other models.

INTRODUCTION
OBJECTIVES
METHODS/DATA
RESULTS
CONCLUSION

Results: False Positives

- We separated the false positive detections into two types: false positives created from signal windows, and false positive detections created from noise windows. The table to the right lists the false detections produced from noise windows, signal windows, and their detection rates.

- Given the previous slides on recall performance and the PhaseNet response to true detections, this table shows the LEB models' primary advantage over the SeisBench models: generating fewer false detections from noise windows.

False Positive Rates per input window

Name	Noise	Sig	Noise/Sig	UGEB Noise Win	Noise Rate	UGEB Sig Win	Sig Rate
LEB 100	1754	1535	1.14	131747	0.013	4180	0.367
LEB 80	4685	2151	2.18	131747	0.036	4180	0.515
LEB 20	7191	2410	2.98	131747	0.055	4180	0.577
LEB 0	11902	1911	6.23	131747	0.090	4180	0.457
ETHZ	278193	6682	41.63	131747	2.112	4180	1.599
GEOFON	209870	4470	46.95	131747	1.593	4180	1.069
INSTANCE	109511	3805	28.78	131747	0.831	4180	0.910
Iquique	151820	5913	25.68	131747	1.152	4180	1.415
LenDB	57655	2945	19.58	131747	0.438	4180	0.705
NEIC	279976	6217	45.03	131747	2.125	4180	1.487
SCEDC	364402	8166	44.62	131747	2.766	4180	1.954
STEAD	254494	7071	35.99	131747	1.932	4180	1.692

Column Descriptions
Noise: false detections within windows without signals
Sig: false detections within windows with signals
Noise/Sig: Ratio between noise window false detections and signal window false detections
UGEB Noise Win: number of unique noise windows in UGEB
Noise Rate: ratio between *Noise* to *UGEB* Noise Win columns The final two columns show analogous definitions for signal window false detections.

INTRODUCTION
OBJECTIVES
METHODS/DATA
RESULTS
CONCLUSION

Conclusion

Concluding Remarks

- We found the primary advantage of training with IMS data over using the SeisBench models, was the suppression of the false detections on noise windows.
- Unanticipated Result: Relative to the IMS-trained models, ETHZ and STEAD SeisBench models did well on teleseismic signals.
- Lowering the SNR threshold for the LEB dataset signals will likely lead to better performance on lower SNR UGEB evaluation. But, this will also likely lead to more false positives.

Thoughts towards building LEB training datasets for the operational monitoring mission:

Model-centric deep learning research has proven the viability of data-driven methods for seismic monitoring. ***With the assumption that*** *m*onitoring organizations will need to retrain models frequently, ***the challenge is*** in developing the data interrogation tools/“rules-of-thumb” to build *effective* training datasets. To address the dataset challenge, we will be focusing on applying *Data-centric AI* approaches [4, 5].

References

For additional information on the results presented in this poster:

Heck, et. al 2022 International Monitoring System 3-Component Seismic Signal Detection Using the PhaseNet Deep Learning Model, Sandia Technical Report, SAND2023-10785

References

[1] Zhu Weiqiang and Gregory Beroza. PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. Geophysical Journal International

[2] Münchmeyer, et. al Which picker fits my data? a quantitative evaluation of deep learning based seismic pickers. Journal of Geophysical Research: Solid Earth

[3] Linville, et. al Global- and Local-Scale High-Resolution Event Catalogs for Algorithm Testing. Seismological Research Letters, 90(5):1987–1993, 07 2019.

[4] Introduction to Data-Centric AI, https://dcai.csail.mit.edu/

[5] Curtis G. Northcutt, Anish Athalye and Jonas Mueller. 2021 Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks. arXiv:2103.14749

Please do not use this space, a QR code will be automatically overlayed

International Monitoring System 3-component seismic signal detection using the PhaseNet deep learning model

Stephen Heck, Jorge Garcia, Rigobert Tibi
Sandia National Laboratories

Sandia National Laboratories
SAND2023-03389O
Lightning Talks
P3.5-365

- Using 14 years of associated signals from the Late Event Bulletin (LEB), we auto-curated a training data set consisting of signal windows containing associated arrivals, and noise windows that contain no LEB associated signals.

- We construct five training data sets by varying the ratio of noise windows to signal windows. At the lowest detection thresholds, increasing the number of noise windows increases the precision from .15 to .4 while reducing the recall from .6 to .5.

- We found the primary advantage of training with IMS data over using the SeisBench models, was the suppression of the false detections on noise windows.

- Unanticipated Result: Relative to the IMS-trained models, ETHZ and STEAD SeisBench models did well on teleseismic signals.

False Positives versus True Positives