SnT 2025 8 SEPTEMBER ONLINE DAY 9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

Software Engineering Aspects of Modifying and Enhancing NET-VISA to Incorporate Long Term Priors

Alexander Shashkin[1], Urtnasan Khukhuudei[1], Mark Prior[1], Nimar Arora[2], Stuart Russell[3]

[1]CTBTO Preparatory Commission
[2]Bayesian Logic, Inc.
[3]University of California, Berkeley

INTRODUCTION AND MAIN RESULTS

The NET-VISA software package features a physics-based probabilistic model combined with a heuristic inference algorithm to identify the most likely set of seismic events corresponding to a series of detections by a global seismic network.
Results are presented that demonstrate the incorporation of long-term priors into the system. This approach generates priors from an extensive dataset spanning approximately ten years, capturing the average behavior of the network and stations across diverse environmental conditions and configurations. The extended training period yielded a comprehensive set of priors that were applied to all processing covering periods of a year or longer, ensuring greater consistency in event detection and analysis.
[

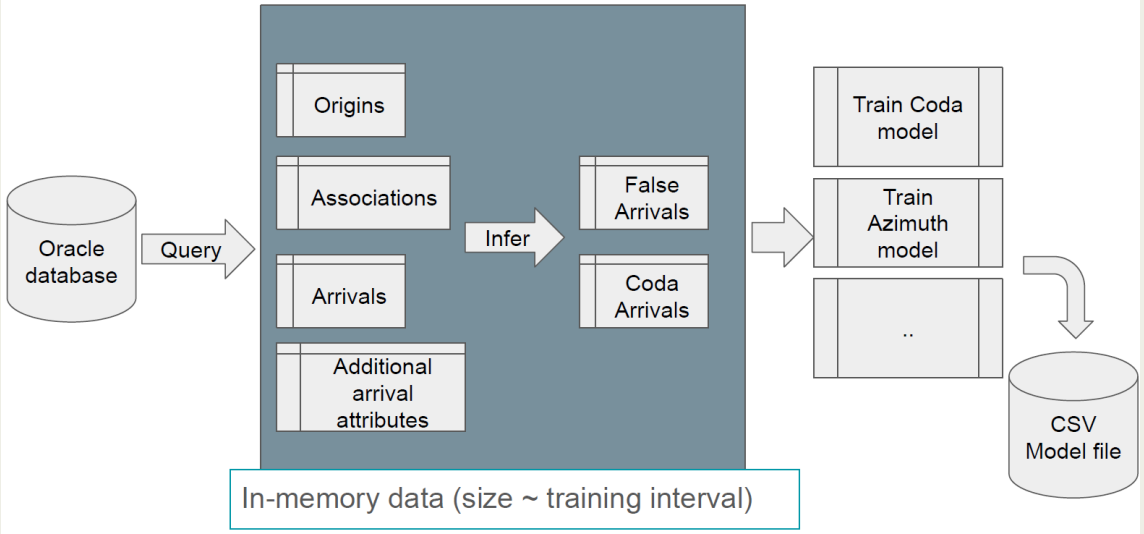
SnT 2025 CTBT: SCIENCE AND TECHNOLOGY CONFERENCE
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE

Software Engineering Aspects of Modifying and Enhancing NET-VISA to Incorporate Long Term Priors

Alexander Shashkin et al.

P3.5-553

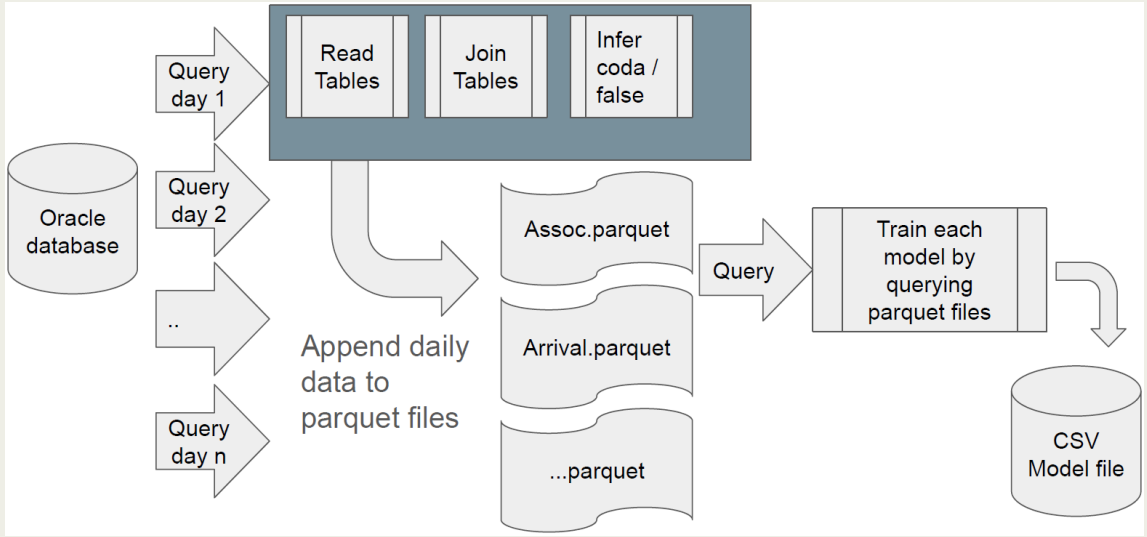
Current situation: NET-VISA Rolling Priors



NET-VISA is a package which is composed of a physics-based, probabilistic model and a heuristic inference algorithm designed to find the most probable set of seismic events which can explain a series of arrivals detected by a global IMS network. The basic components of NET-VISA are a Generative Model (GM) and an Inference Algorithm (IA). The GM contains the information necessary to evaluate the probability of events and arrivals. The IA searches for the set of events that has the maximum a posteriori (MAP) probability - that is the one with highest probability given the input data.

Experts were concerned that NET-VISA, which retrains every week, might encounter errors during training, which could break a smooth operational flow. It was suggested, that NET-VISA could benefit from training on stations with ten or more years of data.

Proposed Solution: NET-VISA Long Term Priors



In the long-term prior model, priors are derived from data accumulated over approximately **ten years**. This approach captures the average behavior of the network and stations across a wide range of environmental conditions and configurations. Because the training period is so extensive, the resulting priors are typically more stable and reliable than those from the rolling prior model. This makes them especially useful for processing data over long periods, helping ensure consistent and accurate results. Data is fetched from the Oracle database using SQL queries, converted into a Pandas DataFrame, serialized into a PyArrow Table, and saved to disk as a Parquet file.. This process is **modular, fast, and reproducible**—ideal for large-scale offline training of priors.

CTBTO Bayesian Logic Inc. UC Berkeley

SnT 2025 8 SEPTEMBER ONLINE DAY 9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

Software Engineering Aspects of Modifying and Enhancing NET-VISA to Incorporate Long Term Priors

Alexander Shashkin et al.

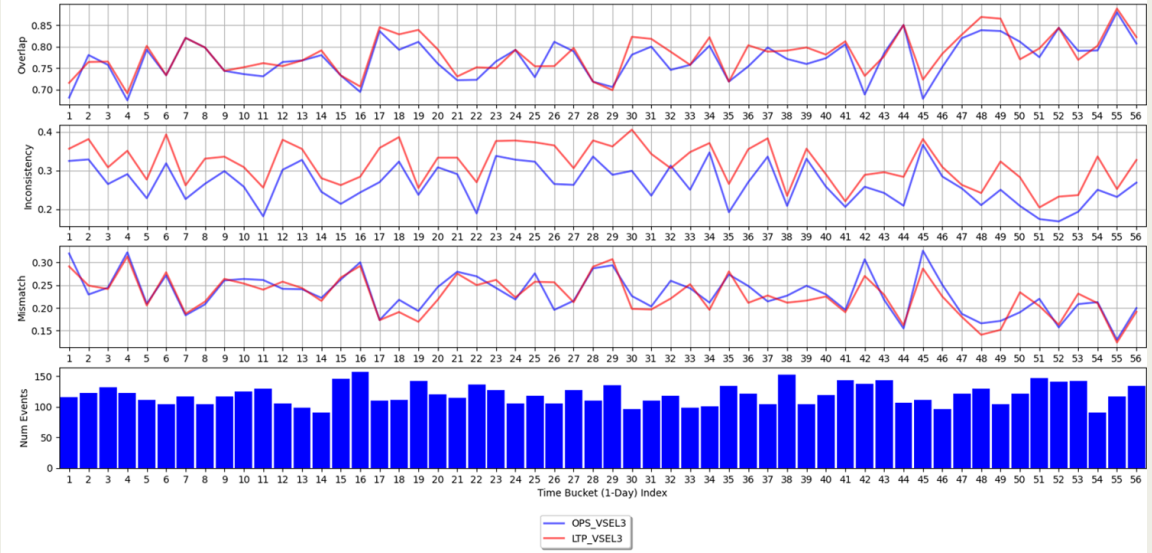
P3.5-553

Metrics

The key measures are:
Overlap, which evaluates the fraction of detected events match those in a reference bulletin.
Inconsistency, the fraction of events present in NET-VISA but absent in the reference.
New metric, **mismatch**, was introduced to provide a combined measure of false-positive and false-negative errors. Mismatch**,** explicitly weights missed events more heavily than false positives—by roughly a factor of 10, because in operational settings, missing real events can be far more costly than including a few uncertain ones.

$$\frac{10 \times (1 - overlap) + inconsistency}{11}$$

Results



The software engineering technologies applied made it possible for a training dataset of arbitrary length to be used, meaning that **10 years** of data could be utilized instead of **3 months** for in-memory training. As a test, seismic events from 13 May to 08 July were compared using the IDC Late Event Bulletin (LEB) as the reference. Long-Term Priors model was found to increase total Overlap from 0.770 to 0.780 and to decrease total Mismatch from 0.233 to 0.228

CTBTO Bayesian Logic Inc. UC Berkeley