

Project Report
TIP-159

QuakeCast: Earthquake Forecasting Using Machine Learning to Identify and Classify Preseismic Ionospheric Anomalies as Earthquake Signals: FY21 Homeland Protection and Air Traffic Control Technical Investment Program

Bhavani Ananthabhotla
Jeffrey Liu
Jessica D. Reid
Samuel Scheele
Dieter W. Schuldt
Matthew L. Weiss

20 October 2022

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS



DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001.

This report is the result of studies performed at Lincoln Laboratory, a federally funded research and development center operated by Massachusetts Institute of Technology. This material is based upon work supported by the United States Air Force under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force.

© 2021 Massachusetts Institute of Technology

Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

Massachusetts Institute of Technology
Lincoln Laboratory

QuakeCast: Earthquake Forecasting Using Machine
Learning to Identify and Classify Preseismic Ionospheric
Anomalies as Earthquake Signals: FY21 Homeland
Protection and Air Traffic Control Technical Investment
Program

Bhavani Ananthabhotla

Jeffrey Liu

Jessica D. Reid

Samuel Scheele

Dieter W. Schuldt

Group 21

Matthew L. Weiss

Supercomputing Center

Project Report TIP-159

20 October 2022

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported by the United States Air Force under Air Force Contract No.
FA8702-15-D-0001.

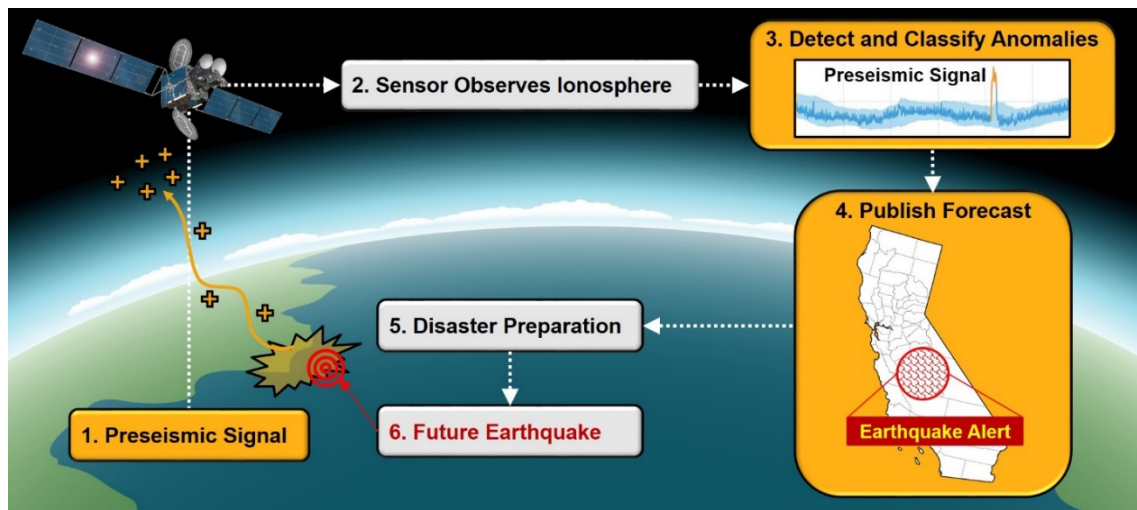
Lexington

Massachusetts

This page intentionally left blank.

EXECUTIVE SUMMARY

Earthquakes are deadly, expensive, and unpredictable. To date, the study of earthquakes has been predominantly through seismology, which has enabled researchers to describe the long-term progression of earthquakes, but is unable to provide early forecasting necessary for short-term disaster preparation. Global Navigation Satellite Systems (GNSS) satellites offer a new opportunity to study earthquakes and begin to unravel the complex interactions between preseismic processes and the global electric circuit that may be the key to earthquake early forecasting. The QuakeCast project is developing a novel machine learning method to identify ionosphere anomalies as preseismic earthquake signals and uses them to predict impending magnitude, geographical region, and timing.



In FY20, we presented a pair of prototype machine learning classifiers to explore the possibility of short-term predictions of impending earthquakes using ionosphere total electron content (TEC) data. We compiled a decade-spanning dataset of global ionosphere TEC data and earthquake events. We demonstrated that both a classical logistic regression model and a deep learning ConvLSTM Autoencoder Anomaly Detector trained on this dataset were able to predict future earthquakes in the time series significantly better than chance. These results strongly suggest that signals in the ionosphere TEC data contain information that may be used to predict earthquakes earlier than current warnings provided by existing systems.

The research completed in FY20 was exploratory in nature and the datasets and machine learning pipeline were ad hoc to support open exploration of potential concepts. In FY21, we worked to refine our machine learning infrastructure to enable us to design, test, and iterate with data and algorithms more quickly. This report will discuss the foundation built to support advanced algorithms, enhanced datasets to

include different sensing modalities and resolution, and the results of several new algorithms to test hypotheses based on data and detection.

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	i
LIST OF ILLUSTRATIONS	v
1. INTRODUCTION	1
1.1 Background and Motivation	1
1.2 QuakeCast Overview	1
2. DATA	5
2.1 Ionosphere Data	5
2.1.1 Global GNSS TEC Data	6
2.1.2 International Reference Ionosphere (IRI) Model	7
2.1.3 Disturbance Storm Time (Dst) Index	9
2.1.4 High-res GNSS TEC Data	9
2.1.5 Unified TEC Dataset	9
2.1.6 Feature Engineering	10
2.2 Earthquake Data	11
3. SUMMARY OF FY20 PROTOTYPE MACHINE LEARNING CLASSIFIERS	15
3.1 Background	15
3.2 Data Management and Training Sets	16
3.3 Logistic Regression Methods and Results	16
3.4 Convolutional Long-Short-Term-Memory (ConvLSTM) Autoencoder Anomaly Detector Methods and Results	19
3.4.1 Hypothesis Test and Results	20
3.4.2 Precision-Recall Curve and Results	22
3.4.3 Conclusion	23
4. PROBLEM FORMULATION	25
5. SUPERVISED PREDICTION USING DEEP NEURAL NETWORKS	27

TABLE OF CONTENTS (Continued)

	Page
5.1 Background	27
5.2 Methods and Results	27
6. SPHERICAL HARMONIC TRANSFORMATION (SHT) DIMENSIONALITY REDUCTION	33
6.1 Background	33
6.2 Method	33
7. VIDEO ACTION TRANSFORMER NETWORK	37
7.1 Action Detection Problem	37
7.2 Transformer Networks	37
8. PATH FORWARD	39
9. SUMMARY	41
REFERENCES	43

LIST OF ILLUSTRATIONS

Figure No.	Page
1 The QuakeCast project aims to provide early warning eight hours prior to an earthquake by making use of machine learning to identify and classify ionosphere anomalies.	2
2 Laboratory testing of granite blocks, such as those found in the lithosphere, demonstrated a current flow when stressed [Figure 2, Freund].	2
3 Current flow between the ionosphere, atmosphere, and lithosphere due to the electric current generated in the lithosphere by preseismic microcracking.	3
4 Data pipeline illustrating the data sources and preprocessing applied to construct the unified ionosphere dataset.	5
5 Percentage of missing UPC TEC data, monthly average.	7
6 The data sources used for the development of IRI with depictions of the instruments that accumulated these data sources: ionosondes (a), incoherent scatter radars (b), rockets (c), topside sounder and in situ satellites (d), Global Navigation Satellite Systems (e), and radio occultation (f). Data sources highlighted in green have been used for validation but not yet for model development [Figure 2, Bilitza].	8
7 Ionosphere TEC rising as sun rise terminator travels over the globe.	11
8 Spatial distribution of earthquakes in the ISC-GEM Global Instrumental Earthquake Catalogue between 2005 and 2015.	12
9 Distributions of magnitude of earthquakes between 2005 and 2015 in ISCGEM.	12
10 Our investigation of earthquake early warning is dual-prong; using a classical machine learning model.	
11 At each of the 96 timesteps within the 24-hour sample, the maximum of two pseudo-radial bins was computed in the 30-degree by 30-degree spatial window shown. The resulting 192 values were the feature set for the logistic regression.	17
12 Precision-Recall curve for a log regression classifier of preseismic data, using bootstrapping (sample size = 1000, number of bootstraps = 500, sampled data balanced and with replacement). The feature set was the max TEC value with radial bins at each timestep over a 24-hour period and within a 30-degree by 30-degree spatial window (192 features).	18
13 Confusion matrix for one representative classifier of preseismic data from the bootstrapping study using a threshold of 0.5, trained on 500 nominal and 500 imminent samples and applied to a test set of 500 nominal and 500 imminent samples. This plot shows the percent of all data that were (reading across, then down) true negatives, false positives, false negatives, and true positives when the model was applied to unseen test data. Accuracy = 0.578.	19
14 Deep learning approach using a ConvLSTM Autoencoder trained on nominal data to detect preseismic earthquake signals by looking for sequences with high reconstruction errors.	20

LIST OF TABLES (Continued)

Table No.		Page
15	Distributions of reconstruction errors suggest that the pre-earthquake data is different than nominal data for both a 15 minute lookahead (left) and 8 hour lookahead (right).	21
16	The precision-recall curves for both the 15 minute (left) and 8 hour lookahead (right) anomaly detectors outperform the null classifier; more work is needed to improve the performance for operational use.	23
17	Setup of time slices that comprise the datacube.	27
19	ROC curve from proposed deep neural network.	29
20	Neural network structure with two output nodes the attempt earthquake localization.	29
21.	Evidence of Overfitting in localization problem: validation loss increases while training loss decreases.	30
22	Estimated and true earthquake locations. Left column shows results early in the training process, right column shows results later in the training process. Top row shows performance on training set, and bottom shows performance on validation set.	30
23	(left to right) The original UPC data; the reconstructed function based on 16 coefficients; the difference between the original and the reconstruction.	33
24	Network used with the SHT data.	34
25	ROC curve for SHT network.	35

1. INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Earthquakes are deadly, expensive, and unpredictable. Earthquakes have accounted for six of the world's 10 deadliest natural disasters, and also result in U.S. annualized losses of \$6.1 billion per year. Today, the state of California utilizes an earthquake warning system called [ShakeAlert](#) that detects the earthquake seismically and can provide up to one minute of warning for nearby regions. With an earlier warning system, regions could stage recovery resources, evacuate, and put infrastructure like power plants into safe mode, saving lives and minimizing financial losses.

To date, the study of earthquakes has been predominantly through seismology. This method is well suited to describing the long-term progression of earthquakes, but is unable to provide early forecasting. Global Navigation Satellite System (GNSS) satellites offer a new opportunity to study earthquakes by monitoring atmospheric behavior in the days leading up to an earthquake. In the historical record, anomalous disturbances in total electron content (TEC) in the ionosphere have been observed in the days and hours preceding some earthquakes. Some theories suggest that microcracking in the lithosphere during preseismic processes generates stress-activated electrical currents due to lattice structure defects in the igneous rocks. The ionosphere is coupled to these electrical currents through the global electric circuit and results in a rapid change in TEC above the affected region. GNSS satellites can be used to monitor the ionosphere TEC globally by measuring the time delay between L1 and L2 frequencies.

1.2 QUAKECAST OVERVIEW

QuakeCast is a Homeland Protection, Air Traffic Control, and Humanitarian Aid and Disaster Response allocated-funded project that aims to provide earthquake forecasting at least eight hours prior to an earthquake (see Figure 1). This is accomplished by using machine learning to identify TEC anomalies and classify preseismic earthquake signals by magnitude, geographical region, and event window. The project has collected over 10 years of ionosphere TEC data, which contains over 10,000 earthquakes, and is utilizing machine learning approaches to analyze these data for patterns that indicated an impending earthquake. If successful, it may be possible to provide “shake forecasts” just as you would expect from a traditional weather forecast: “Today will be partly cloudy with a moderate chance of shaking.”

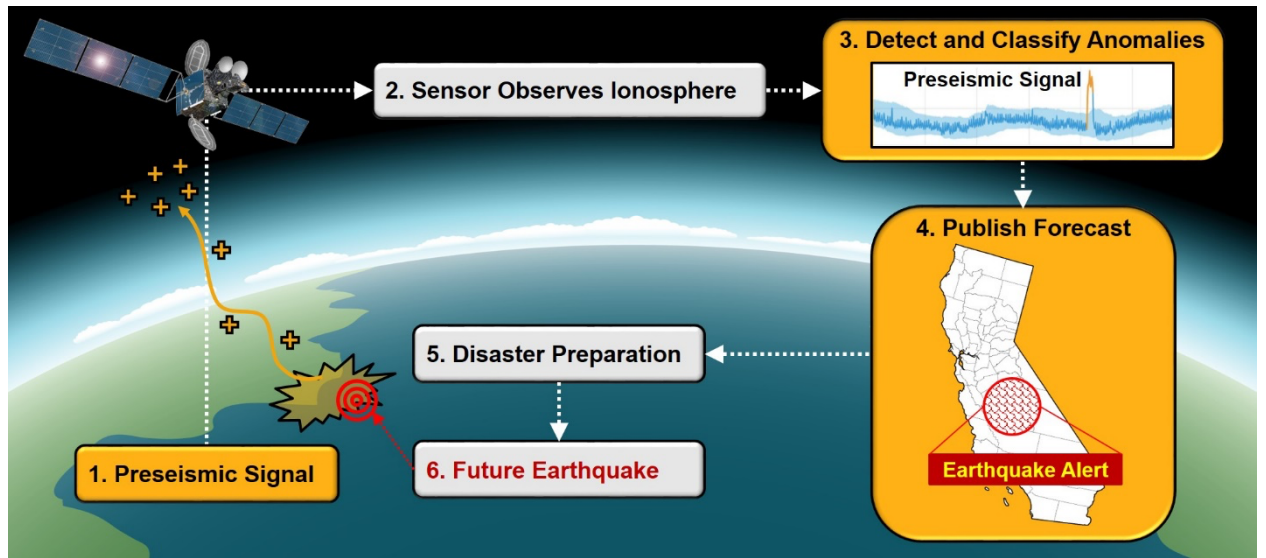


Figure 1. The QuakeCast project aims to provide early warning eight hours prior to an earthquake by making use of machine learning to identify and classify ionosphere anomalies.

1. *Preseismic Signal*: Anomalous disturbances in TEC in the ionosphere have been observed preceding strong earthquakes greater than magnitude 5 [Liu, Zhou, Zolotov]. Some theories suggest that microcracking in the lithosphere during preseismic processes generates stress-activated electrical currents due to lattice structure defects in the igneous rocks (see Figure 2).

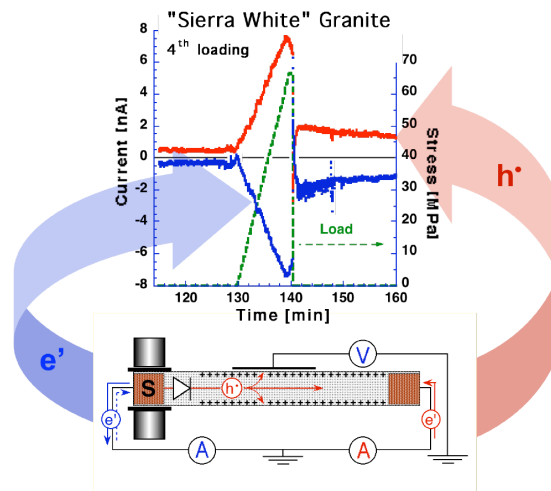


Figure 2. Laboratory testing of granite blocks, such as those found in the lithosphere, demonstrated a current flow when stressed [Figure 2, Freund].

As illustrated in Figure 3, the ionosphere is coupled to these electrical currents through the global electric circuit and results in a rapid change in TEC above the affected region [Freund, Kuo, Molchanov].

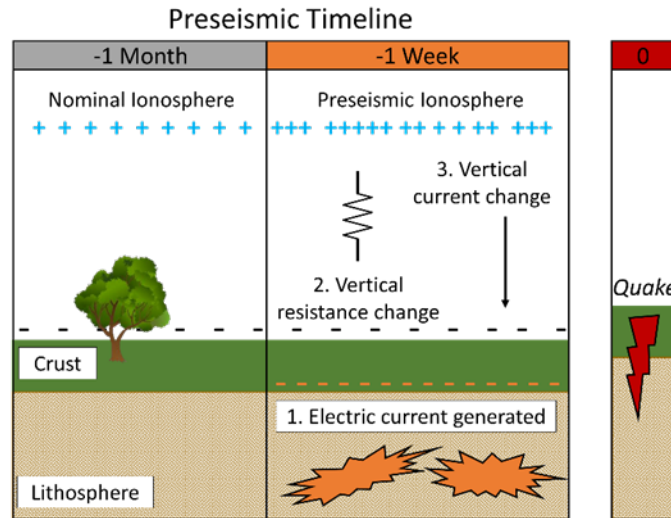


Figure 3. Current flow between the ionosphere, atmosphere, and lithosphere due to the electric current generated in the lithosphere by preseismic microcracking.

2. *Sensor Observes Ionosphere:* Global Navigation Satellite Systems (GNSS) monitors the ionosphere TEC globally by measuring the time delay between the L1 and L2 frequencies between satellites and ground receivers. Ionosphere propagation delay is persistently measured during normal operation because it affects positioning accuracy. TEC data collected by GNSS may be monitored for anomalous changes due to preseismic activity.
3. *Detect and Classify Anomalies:* QuakeCast is developing machine learning algorithms to identify these TEC anomalies and classify them as a preseismic earthquake signal.
4. *Publish Forecast:* By continuously monitoring the TEC above earthquake prone regions, QuakeCast aims to provide earthquake forecasts that will warn the area of an impending event and provide a projected risk region, event window, and event magnitude.
5. *Disaster Preparation:* Based on the earthquake forecast, the region can prepare by staging supplies, evacuating people, and putting some infrastructure into safe modes. This will save lives and reduce financial losses.
6. *Future Earthquake:* The earthquake occurs in the projected region and event window at the projected magnitude.

In FY20, we demonstrated that both a classical logistic regression model and a deep learning ConvLSTM Autoencoder Anomaly Detector trained on ionosphere TEC data were able to predict future earthquakes in the time series significantly better than chance. In the following sections we will discuss the FY21 research activities to improve upon the experimental results from FY20: the foundation built to support advanced algorithms work, enhanced datasets to include different sensing modalities and resolution, and the results of several new algorithms to test hypotheses based on data and detection.

2. DATA

Machine learning (ML) is a data-driven approach to model creation, requiring large amounts of training examples. The QuakeCast project leverages a number of data sources from two main categories: ionospheric measurements and earthquake data. In this section, we document the various data sources, the respective parsing and preprocessing steps, and the construction of the unified ionosphere dataset.

2.1 IONOSPHERE DATA

The ionosphere data characterizes the global dynamics within the ionosphere. Our preprocessing pipeline collects a number of different raw data sources, cleans them, and then combines them into a unified ionosphere dataset to use in our machine learning applications. Following this, we compute various derived features to provide alternate representations of the data for the ML inputs. Figure 4 outlines the various data sources and associated preprocessing steps applied to construct the unified ionosphere dataset. We will first discuss the various data sources and their relevant parsing and preprocessing steps; we will then discuss the dataset organization and derived features.

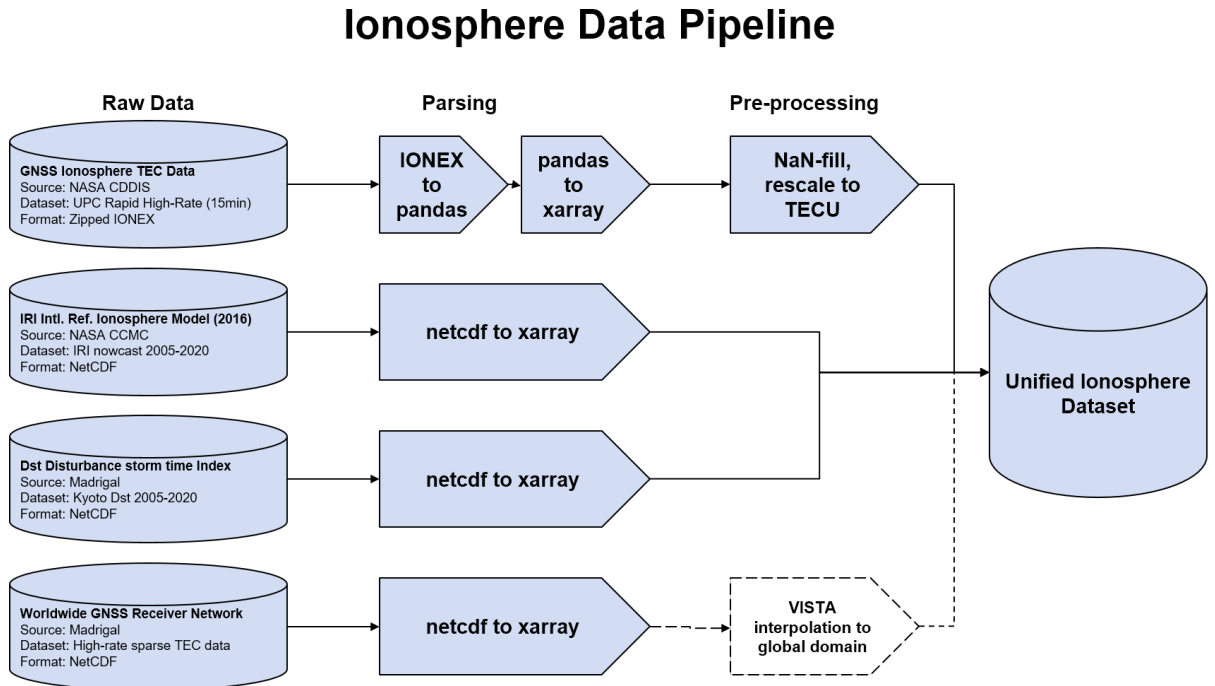


Figure 4. Data pipeline illustrating the data sources and preprocessing applied to construct the unified ionosphere dataset.

2.1.1 Global GNSS TEC Data

As discussed previously, TEC can be measured from GNSS positioning satellites by comparing timing differences between the L1 and L2 bands of the signal. Various organizations compile raw GNSS data and process them into global maps of TEC at regular intervals. We utilize the Rapid High Rate solution from Universidad Politecnica de Catalunya (UPC); this dataset served as the basis of the FY20 explorations.

The UPC high rate data solution has a temporal resolution of 15 minutes, and spatial resolution of 2.5° latitude and 5° longitude. This data is available from the NASA Crustal Dynamics Data Information System (CDDIS) platform from 2005 to present day. Data is presented in daily packages in the IONEX format, for which we developed a parser to convert to pandas (<https://pandas.pydata.org/>) dataframes (McKinney et al, 2010) in Python; TEC values are recorded as integers with units of 0.1 TECU, which we convert to floating point with units of TECU.

While the data are theoretically available from mid-2005 to present day, we found that in practice, there were many days that were missing data. Figure 5 shows the monthly average percentage of data that are missing over time. We can see that prior to 2011, a large fraction (40%–50%) of the data was missing each month. Between 2011 and 2016, the data are largely complete, with only a few months missing a small (<10%) amount of data. Starting in 2016, we see that a significant, but smaller, fraction (~20%) of the data are missing each month. We verified that these data packages were missing from the CDDIS source, and were not a problem caused by the data ingestion or parsing process. We have not yet reached out to UPC to determine the cause of this missing data; i.e., whether it is an issue with the raw data, their processing capability, or simply an issue with publishing/hosting.

The missing data creates a challenge for our machine learning approach, since it reduces the amount of data that we have to work with. Furthermore, since we are looking at time series, we require contiguous data, and data that are randomly missing throughout a period is more disruptive than contiguous sections of missing data, due to boundary effects. Our additional data sources seek to address some of these issues.

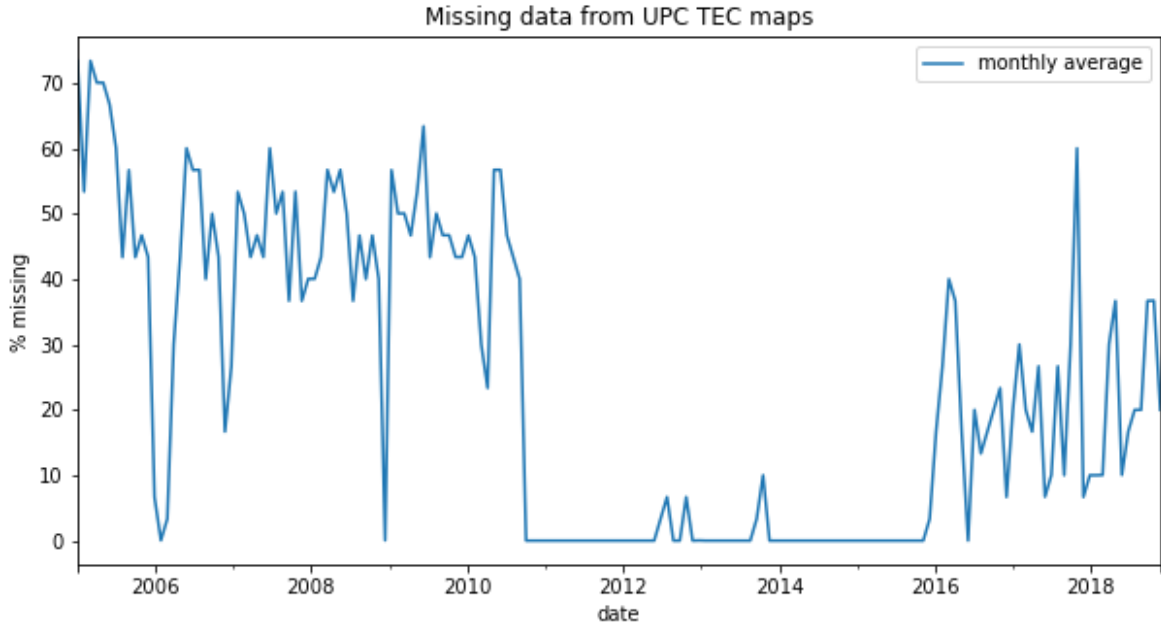


Figure 5. Percentage of missing UPC TEC data, monthly average.

2.1.2 International Reference Ionosphere (IRI) Model

Our second TEC data source is the International Reference Ionosphere (IRI) model, IRI Version IRI2016 (<http://irimodel.org/>). Since 1999 the IRI model has been the international standard empirical model of the ionosphere. The major data sources are the worldwide network of ionosondes, the powerful incoherent scatter radars (Jicamarca, Arecibo, Millstone Hill, Malvern, St. Santin), the ISIS and Alouette topside sounders, and *in situ* instruments flown on many satellites and rockets as shown in Figure 6. In addition, it incorporates data on solar weather, such as solar indices (F10.7 daily, 81-day, and 12-month running mean; sunspot number 12-month running mean), ionospheric index (ionosonde-based IG index 12-month running mean), and magnetic index (3-h ap, daily ap). Notably, the IRI model does not yet include GNSS data, so it can be seen as an independent measurement of the global TEC map. The IRI model is updated yearly during special IRI Workshops (e.g., during COSPAR general assembly). More information can be found in the [workshop reports](#). The resolution of the IRI data is 30 minutes temporally, and 5° lat and 5° lon. This is lower resolution than the GNSS TEC data, but unlike the UPC TEC data the IRI dataset is complete. We linearly interpolated the IRI data both spatially and temporally to align its dimensions to the UPC high rate TEC data. The IRI model data was attained through the “runs on request” feature from

the Community Coordinated Modeling Center (CCMC), <https://ccmc.gsfc.nasa.gov/>. The CCMC is a multi-agency partnership to enable, support and perform the research and development for next-generation space science and space weather models.

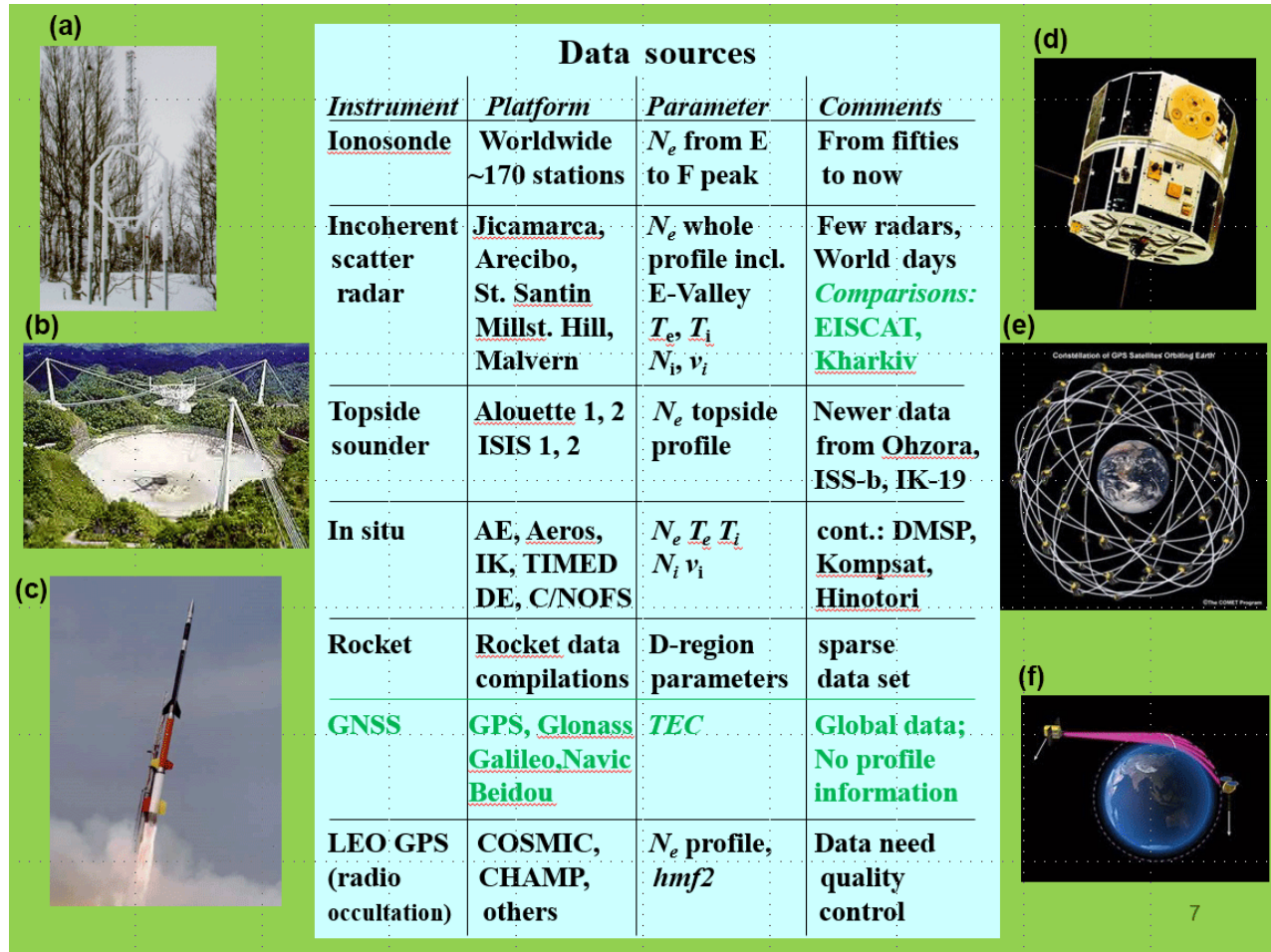


Figure 6. The data sources used for the development of IRI with depictions of the instruments that accumulated these data sources: ionosondes (a), incoherent scatter radars (b), rockets (c), topside sounder and in situ satellites (d), Global Navigation Satellite Systems (e), and radio occultation (f). Data sources highlighted in green have been used for validation but not yet for model development [Figure 2, Bilitza].

2.1.3 Disturbance Storm Time (Dst) Index

We additionally incorporated a measure of solar activity: the Kyoto Disturbance Storm-Time (Dst) index. This index has been identified as capturing dynamics between solar activity and the global TEC distribution [Blagoveshchensky]. TEC bursts in response to a geomagnetic storm can occur immediately or with a delay of hours or even days. Dst index data can help train our models to recognize bursts due to solar activity and not earthquake activity. The data is presented as a single scalar value for the entire globe, and is available at an hourly temporal resolution. We linearly interpolate the Dst to align with the 15-minute time resolution of the UPC high rate TEC data.

2.1.4 High-res GNSS TEC Data

We identified a higher resolution GNSS TEC data source, which provides data at five minute temporal resolution, with 1° latitude and 1° longitude spatial resolution. The higher resolution of the data is able to capture faster and smaller phenomena, which may help us better resolve the location and timing of potential earthquakes.

These data are closer to the raw data used in UPC high-rate products. However, they are spatially sparse—i.e., they only include TEC data within direct line of sight of ground receivers. The data availability depends on both the location of ground receivers, and the location of GNSS satellites at any given time; thus, many areas, such as the oceans, lack data. Global TEC maps can be generated from these sparse observations using various inference processes. The Solstice group from the University of Michigan has identified an approach to infer global TEC maps from such data called VISTA (Hu, 2020), and are planning to release global TEC maps at the high temporal and spatial resolution. These data are not yet available on the Madrigal Database at Millstone, <http://millstonehill.haystack.mit.edu>.

2.1.5 Unified TEC Dataset

Once the raw data was downloaded and pre-processed, we compiled the data into a unified, aligned dataset based on the NetCDF file format and xarray (Hoyer et al., 2017) Python library. The NetCDF file format is designed for spatiotemporal datasets, natively supporting high-dimensional datasets with spatial and temporal indexing (Unidata, 2021). We had previously used .csv and Apache Parquet file formats, but ran into performance issues when querying across many files or columns.

CSV files store data in a row-oriented manner, meaning all data within the same row are stored contiguously on disk. Our data were stored with timestamps on the row index, and lat/lon pairs on the column index. In this case, selecting a contiguous subset of columns (corresponding to lat/lon pairs) across files is inefficient, since each row needs to be loaded in memory in order to select the relevant columns.

To address this issue, we tried storing data in Parquet format and accessing it using Pyspark. Parquet data files are column-based, which allows for selecting contiguous subsets of columns. However, we ran into performance issues when the number of columns was large (~ 1000 s). Parquet tables are designed to support 2-dimensional data, and thus, for our spatiotemporal (3D) datasets, the spatial dimensions had to

be flattened. Since our data has 71×73 lat/lon grid point combinations, which resulted in over 5000 columns. Furthermore, our anticipated future support for the high resolution GNSS data will require us to support 181×361 grid points. Thus, the performance of Parquet was unable to scale to our anticipated needs.

Ultimately, we settled on NetCDF files for the unified data set, and the xarray Python library to access the data. This configuration allowed us to store data natively in spatial/temporal indexed form without needing to first combine the spatial dimensions into columns. NetCDF supports data annotation and metadata within the format, allowing the dataset to provide some of its own documentation. Furthermore, xarray provides some parallelization support for calculations via dask (Dask Development Team, 2016).

The native support for high-dimensional indexed data, such as spatiotemporal data, provided by NetCDF makes it easy and performant to index and slice into the data. NetCDF supports this by storing the index in headers separate from the main data, allowing queries to be done only on the indices without needing to load the rest of the files. We were able to align all of our available ionosphere datasets in space and time using the xarray resampling and interpolation methods. Once aligned, we were able to select from all of the data sources using the same query interface. We found that filter operations using the NetCDF and xarray implementation were over 20x faster than our csv implementation, and over 10x faster than our Parquet and PySpark implementation.

2.1.6 Feature Engineering

We explored various feature representations of the TEC data to refine our machine learning inputs. These alternate representations can provide additional information about the signal, enable dimensionality reduction to simplify the models and reducing their computational requirements, and also account for known dynamics of noise processes unrelated to the earthquake activity.

We computed gradient and derivative-based features to represent the change in TEC data over space and time. We also computed statistical measures to characterize the data distribution over time.

For dimensionality reduction, we explored various signal processing transforms, such as wavelets and curvelets, which did not show much potential in our initial analyses. We also considered the Spherical Harmonic Transform (SHT), which has been used in other TEC machine learning problems, and will be discussed in a later section. Spherical Harmonics serve as a set of basis functions on the surface of a sphere, and can be used to reconstruct signals on spherical surfaces, analogous to the use of circular functions to represent periodic signals in Fourier series.

To account for known dynamics, we looked to isolate the effects of known periodic trends in TEC signals. The ionosphere is not a static medium and TEC dynamics are caused by known conditions; time of day, time of year, sun activity, and even location. For example, as the sun rises the ionosphere is energized and the TEC rises; after the sun sets the ionosphere is de-energized and the TEC falls. Figure 7 is a visualization of the ionosphere TEC on a nominal day during which no > 5.0 magnitude earthquakes occurred. As can be seen, the TEC increases as the sunrise terminator travels over the globe. We

decomposed data into yearly, daily, and residual signals using signal filtering with periodic kernels. In addition, we computed residuals from time-lagged moving averages as alternate representations of data with periodic trends removed.

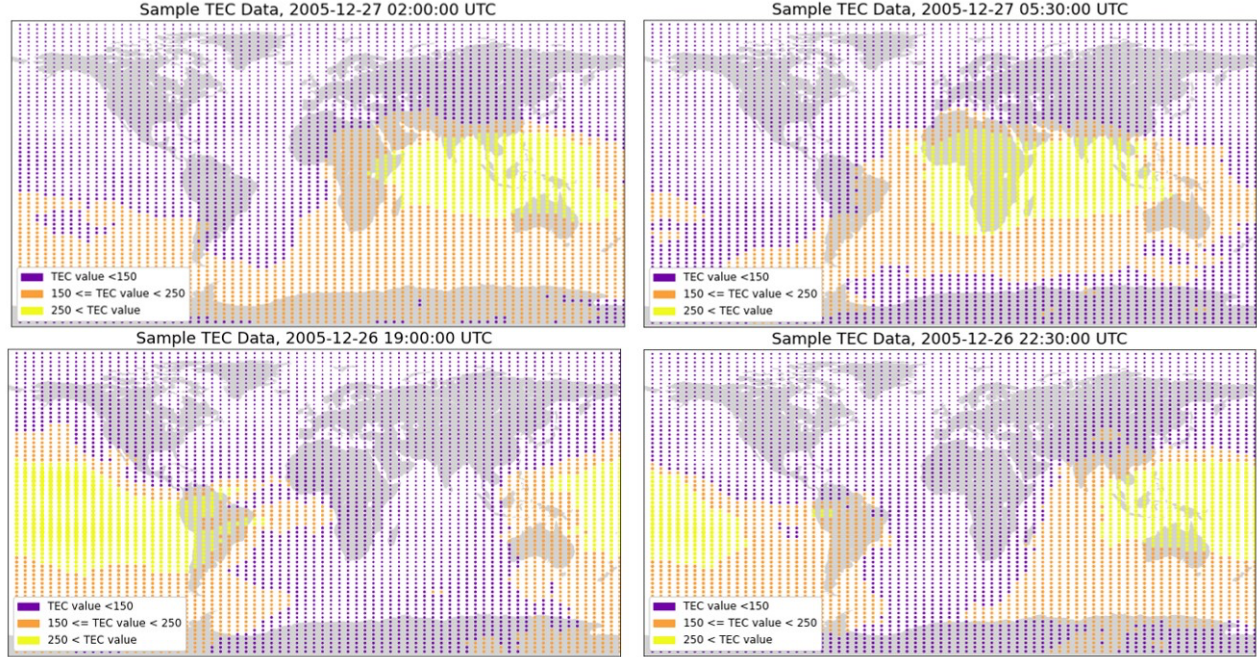


Figure 7. Ionosphere TEC rising as sun rise terminator travels over the globe.

2.2 EARTHQUAKE DATA

While the earliest recorded earthquake dates to 1831 BCE, modern methodology has provided a more detailed and systematized set of earthquake reports. Efforts by Storchak et al., (2013), Storchak et al., (2015), and Giacomo, Engdahl and Storchak (2018) culminated in the International Seismological Centre global catalogue of bulletin data (Figure 4 and Figure 5) for earthquakes of magnitudes greater than 5.5 and continental earthquakes of magnitudes greater than 5.0 specifically for hazard and risk studies. We use the 8th release, which contains data from 1904–2017. For this project, we focused on data from 2005–2017 to coincide with the availability of ionosphere TEC data. The ISC-GEM is highly calibrated and is meant to serve as a calibration reference for seismic measurements across the globe; for this reason, we use this dataset as our ground truth of seismic activity. Advances in machine learning present an opportunity to look for patterns in decades of past TEC and earthquake data to help unravel the complex interactions of the global electric system and look for preseismic earthquake signals.

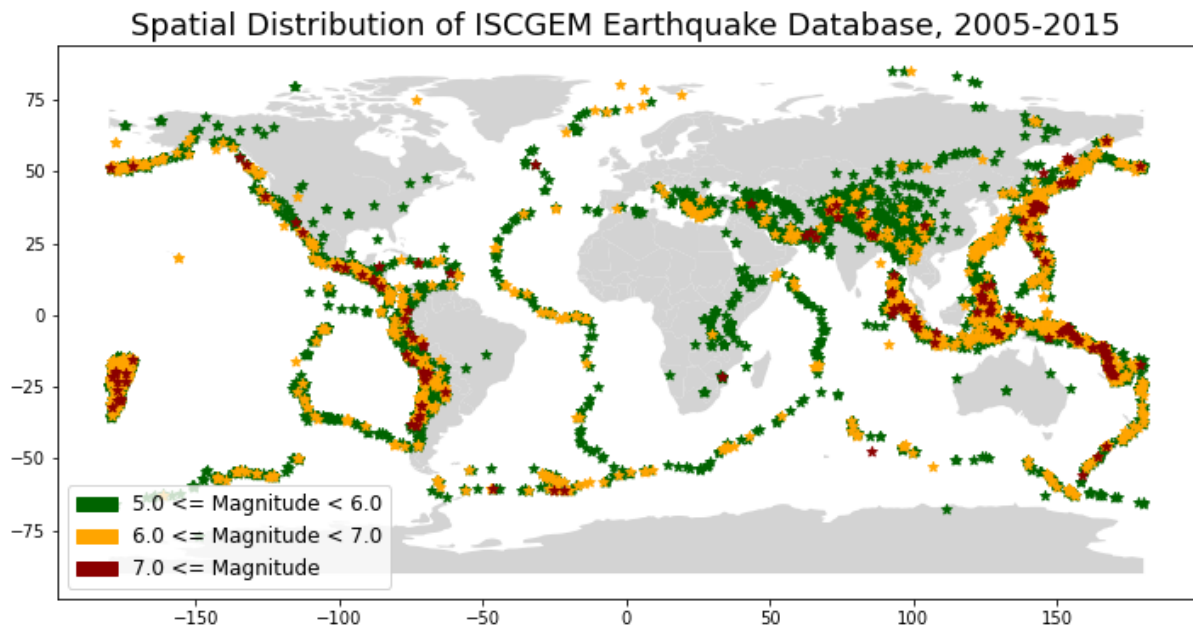


Figure 8. Spatial distribution of earthquakes in the ISC-GEM Global Instrumental Earthquake Catalogue between 2005 and 2015.

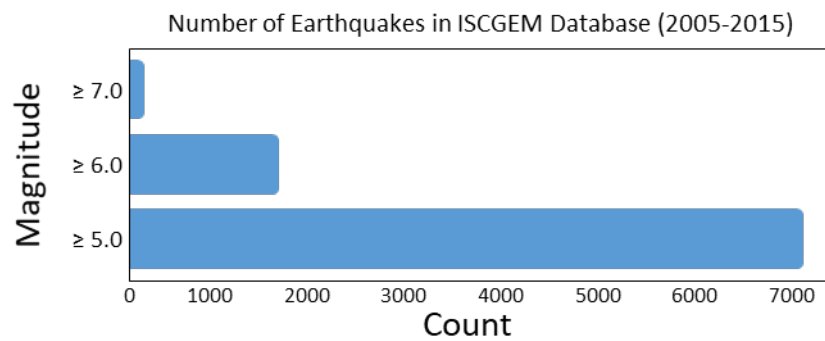


Figure 9. Distributions of magnitude of earthquakes between 2005 and 2015 in ISC-GEM.

Since GNSS measurements of TEC require line of sight between GNSS ground receivers and satellites, GNSS-derived TEC products only have direct measurements near land where the ground receivers are located. As such, TEC data over the ocean or land without many ground receivers must be inferred, and are potentially less accurate. We developed tools to allow the filtering of the earthquake data to any specified geographic regions, such as within a certain range from land, or by administrative boundaries using GIS products.

This page intentionally left blank.

3. SUMMARY OF FY20 PROTOTYPE MACHINE LEARNING CLASSIFIERS

3.1 BACKGROUND

Machine learning is the application of algorithmic and statistical models to derive inferences about data, and it allows us to approach the problem of earthquake early warning without extensive modeling of ionosphere dynamics. Instead, we frame it as the following classification problem: given a set of time-series TEC data, can we tell if it is preseismic, or within 24 hours of an earthquake event documented in the ISC-GEM catalog? We started the investigation using two methods: a classical method that is readily interpretable and a neural network approach that provides more opaque solutions. These approaches and others that were considered are summarized in Figure 10 below.

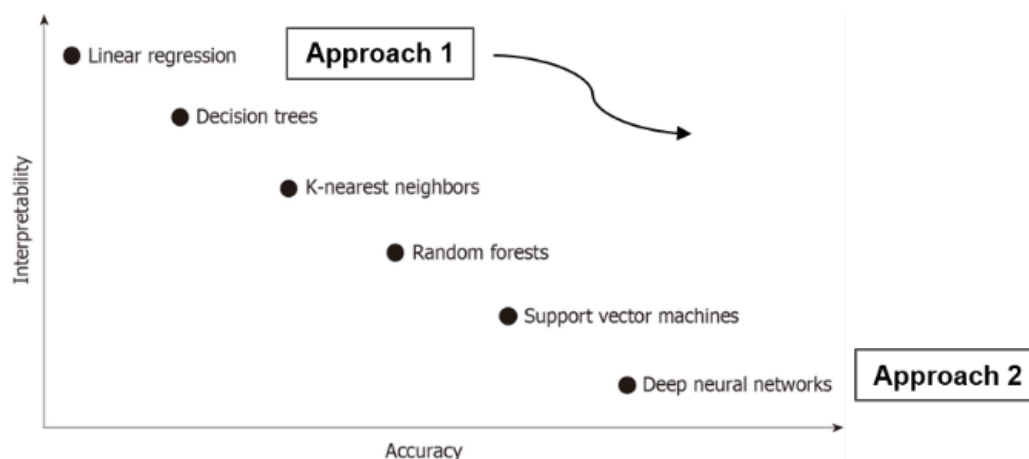


Figure 10. Our investigation of earthquake early warning is dual-prong; using a classical machine learning model that aims at interpretability, along with a deep-learning approach that aims at high accuracy.

For our classical machine learning approach we used logistic regression, which is a simple binary classification method that fits a logistic function to the feature set. The coefficients of the best fit curve allows us to understand the relative contribution of features to the model, making the results easily interpretable.

Our Deep Learning approach was a Convolutional Long-Short-Term-Memory (ConvLSTM) Autoencoder anomaly detector. ConvLSTM autoencoders are neural networks that learn to reproduce sequences of arrays, such as video data. They can learn to approximate system dynamics and predict the next frame of data based on the preceding frames. Data that look different from the data the ConvLSTM

Autoencoder was trained on will have higher reconstruction errors than data similar to the training data. This use of the ConvLSTM Autoencoder as an anomaly detector was proposed by Luo et al. (2017), and builds upon the development of the ConvLSTM by Shi et al. (2015) and the use of the autoencoder framework for anomaly detection in video by Xu et al. (2015). Unlike the classical approach there is a very limited set of ways that have been developed to visualize what features are most critical to the classification being made.

3.2 DATA MANAGEMENT AND TRAINING SETS

We considered 24-hour sequences of $30^\circ \times 30^\circ$ latitude/longitude windows of TEC data. The TEC data occasionally had missing observations and time steps; we considered only sequences that had a full 24-hours of observations.

Sequences belong in one of two categories: nominal, and pre-earthquake. Nominal data is defined as sequences which had no earthquakes within the spatial and temporal window, nor in the subsequent 24-hours. For the pre-earthquake sequences, we further defined sequences as X-lookahead pre-earthquake sequence, where X is a time interval. An X-lookahead pre-earthquake sequence means that an earthquake event occurs within the spatial window between X to X+15 minutes after the end of the sequence. In this paper, we discuss results for 15 minute- and 8 hour-lookahead pre-earthquake sequences.

The TEC values across the entire data were normalized to a mean of 0 and variance of 1, estimated from a random sample of 10k nominal sequences. We partitioned the data into a training set and a test set based on the timestamp of the data. Data from the years 2005-2014 was used as the training set, and those from 2015 as the test set. This is done to simulate a forward-looking prediction: had we made the predictions based only on the data from 2005–2014, how would it have performed in 2015? The following sections show the methods and results from the classical and deep learning approaches.

3.3 LOGISTIC REGRESSION METHODS AND RESULTS

The following feature reduction strategy was used in this initial approach: first, the TEC data was binned radially within each timestep of the 30° by 30° pre-event window, as can be seen in Figure 8. The maximum value of each bin at each timestep was used as input to the model, for a total of 192 features, rather than 8736 features.

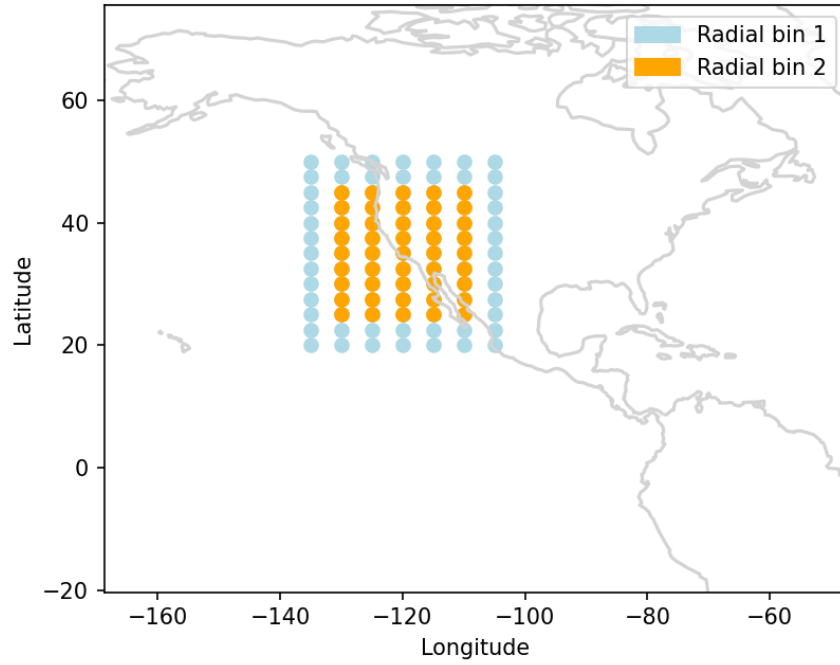


Figure 11. At each of the 96 timesteps within the 24-hour sample, the maximum of two pseudo-radial bins was computed in the 30-degree by 30-degree spatial window shown. The resulting 192 values were the feature set for the logistic regression.

Using a methodology called bootstrapping, 500 logistic regression models trained over class-balanced subsets of data from 2005–2014 were generated. They were each evaluated over the same set as well as an unseen class-balanced subset from 2015. The precision and recall of each model as a function of the threshold used to determine class labels is shown in Figure 9, as well as the average precision and recall of the models over 2005–2014 and 2015 data. For comparison, the results of a baseline classifier, which generates a random number to determine class label, are shown in Figure 10.

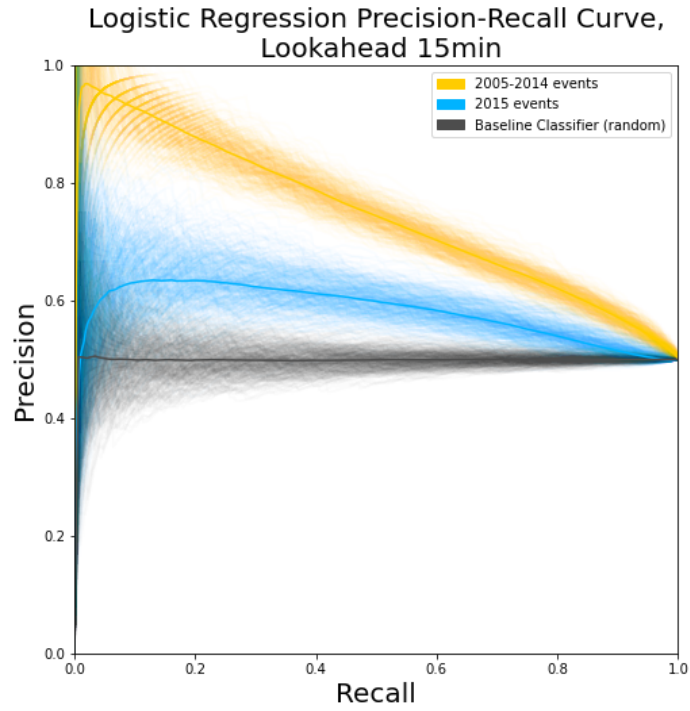


Figure 12. Precision-Recall curve for a log regression classifier of preseismic data, using bootstrapping (sample size = 1000, number of bootstraps = 500, sampled data balanced and with replacement). The feature set was the max TEC value with radial bins at each timestep over a 24-hour period and within a 30-degree by 30-degree spatial window (192 features).

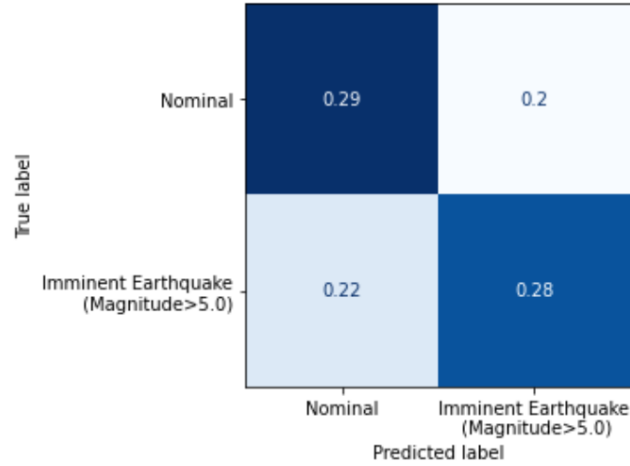


Figure 13. Confusion matrix for one representative classifier of preseismic data from the bootstrapping study using a threshold of 0.5, trained on 500 nominal and 500 imminent samples and applied to a test set of 500 nominal and 500 imminent samples. This plot shows the percent of all data that were (reading across, then down) true negatives, false positives, false negatives, and true positives when the model was applied to unseen test data. Accuracy = 0.578.

The results of the simple model provide compelling evidence for information content in TEC data for earthquake forecasting. Analysis of the results and why the feature reduction strategy was successful in retaining information relevant to preseismic classification would provide paths forward for improving classical machine learning models for preseismic classification in the future.

3.4 CONVOLUTIONAL LONG-SHORT-TERM-MEMORY (CONVLSTM) AUTOENCODER ANOMALY DETECTOR METHODS AND RESULTS

We trained a ConvLSTM Autoencoder on 100k samples of nominal data from our training set, which we call the Nominal ConvLSTM. By training only on nominal data, we can use the Nominal ConvLSTM as an anomaly detector. The theory behind this is that the Nominal ConvLSTM should be better at predicting frames of data which look similar its training data. Thus, if the pre-earthquake sequences look different from the nominal sequences, then pre-earthquake sequences will have a larger reconstruction error—measured as the Mean Absolute Error (MAE) between observed TEC sequence and that predicted by the Nominal ConvLSTM—than nominal sequences do. This approach is illustrated in Figure 11.

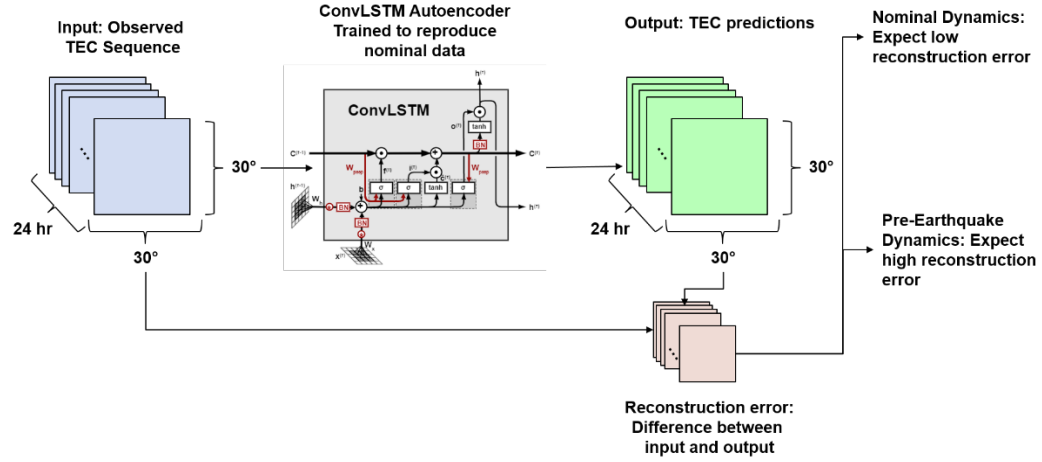


Figure 14. Deep learning approach using a ConvLSTM Autoencoder trained on nominal data to detect preseismic earthquake signals by looking for sequences with high reconstruction errors.

3.4.1 Hypothesis Test and Results

We test if the pre-earthquake sequences are different from the nominal sequences by posing it as a hypothesis test:

- H0: The distributions of reconstruction errors for nominal sequences and pre-earthquake sequences from the Nominal ConvLSTM are the same
- H1: The reconstruction errors from the Nominal ConvLSTM for pre-earthquake sequences are larger than those for nominal sequences

Under the null hypothesis, the Nominal ConvLSTM is equally accurate in reproducing both nominal and pre-earthquake TEC sequences. This would suggest that pre-earthquake sequences look sufficiently similar to nominal sequences that a ConvLSTM trained only on nominal data can predict the dynamics of pre-earthquake data. The alternative hypothesis states that the Nominal ConvLSTM has larger errors when trying to predict pre-earthquake dynamics. This would suggest that pre-earthquake sequences indeed look different from nominal sequences.

Note that since the Nominal ConvLSTM is trained exclusively on nominal data, we can use it to evaluate any 24-hour TEC sequence and look for high reconstruction errors. Thus, we are able to evaluate sequences with various lookahead values without any additional effort. In this section, we evaluate both 15 minute- and 8 hour- lookaheads pre-earthquake sequences, where a 15 minute lookahead would provide 15 minutes of warning and an 8 hour lookahead would provide 8 hours of warning.

We test the hypothesis using a one-sided Mann-Whitney U Test, which is a non-parametric test of whether two samples come from the same distribution. We chose a non-parametric test since we do not know the distribution of the reconstruction errors, and thus do not want to assume normality. We use the one-sided version because we are testing specifically if the reconstruction errors for pre-earthquake sequences are greater than those for nominal sequences.

We generated 1000 bootstrap samples, each with 500 nominal and 500 pre-earthquake sequences. For each sample, we computed the reconstruction error for every nominal and pre-earthquake sequence, and computed the Mann-Whitney U Test on the reconstruction errors for the two populations. Figure 15 shows an example of the distributions of reconstruction error for each of the 15 min- and 8 hr-lookahead pre-earthquake and nominal sequence populations for both the Training and Testing data splits. We can see that the distribution of reconstruction errors for the pre-earthquake sequences are skewed rightward compared to that of the nominal sequences.

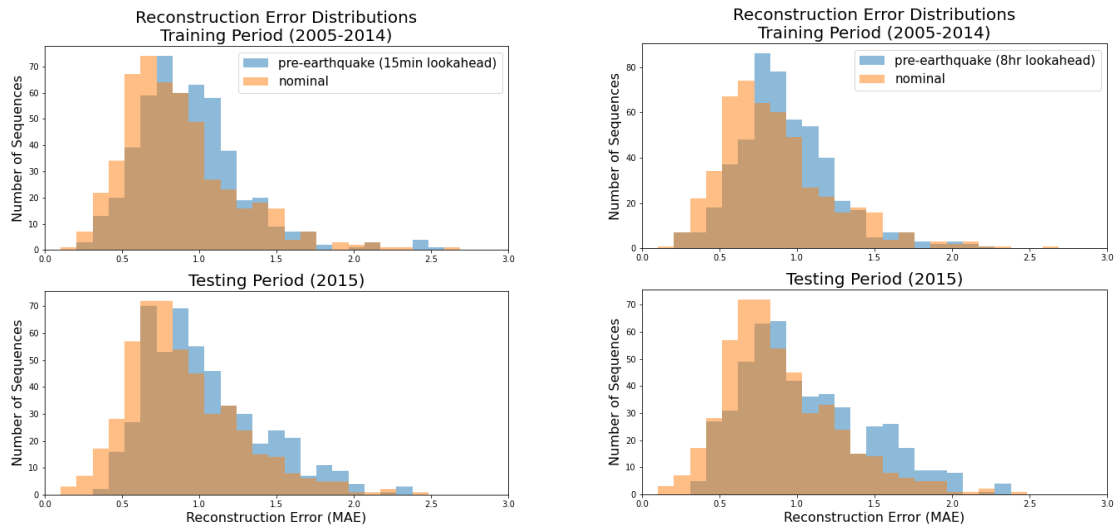


Figure 15. Distributions of reconstruction errors suggest that the pre-earthquake data is different than nominal data for both a 15 minute lookahead (left) and 8 hour lookahead (right).

We report the median bootstrap p-value in the Table 1. In both the Train and Test splits, we are able to reject the null hypothesis for all cases at the 99.99% confidence level. This means that the reconstruction errors for the 15 min- and 8 hr-lookahead pre-earthquake sequences are indeed larger than those for the nominal sequences, and that the pre-earthquake sequences look different from the nominal sequences.

TABLE 1

Both the train and test splits are able to reject the null hypothesis at the 99.99% confidence level, confirming the pre-earthquake data is different than the nominal data. These results hold for both 15-minute and 8-hour lookaheads.

Split	Lookahead	p-value	Significance Level
Train (2005-2014)	15 minute	1.8e-5	p<0.0001
	8 hour	4.5e-5	p<0.0001
Test (2015)	15 minute	8.9e-10	p<0.0001
	8 hour	8.7e-11	p<0.0001

3.4.2 Precision-Recall Curve and Results

In addition to the hypothesis test we present the precision recall (PR) curve for the anomaly detector below. The PR curve demonstrates the trade-off between precision (fraction of alarms that are correct) and recall (fraction of events that are detected). In Figure 16 the baseline null classifier (randomly guessing) is shown in grey as a comparison. Confidence intervals are generated by the same bootstrap resampling process described in the Hypothesis Testing section. We see that the anomaly detector outperforms the null classifier across the board. However, the precision is still relatively low, and additional work is necessary to bring the performance up to operationally-useful standards. Nevertheless, these results give us confidence that there are indeed some signals in ionosphere TEC data that can be used to predict some earthquakes before they happen.

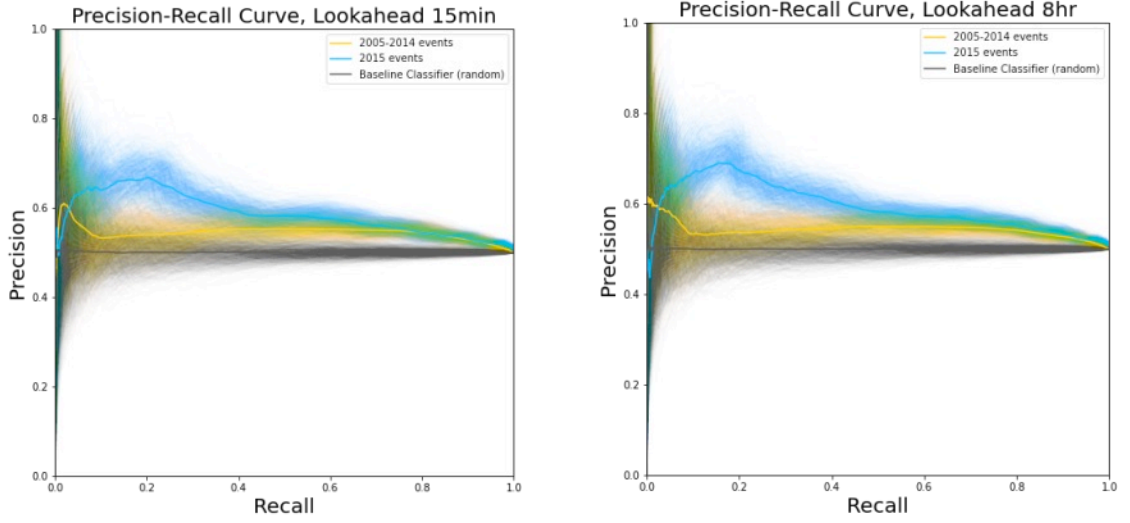


Figure 16. The precision-recall curves for both the 15 minute (left) and 8 hour lookahead (right) anomaly detectors outperform the null classifier, more work is needed to improve the performance for operational use.

3.4.3 Conclusion

We trained a ConvLSTM autoencoder on nominal TEC sequences to create a model of nominal TEC dynamics. This was used as an anomaly detector by comparing the reconstruction error of TEC sequences to the predicted dynamics from the Nominal ConvLSTM. We found that the reconstruction error of TEC sequences prior to an earthquake event were significantly higher than the reconstruction error for nominal sequences, suggesting that pre-earthquake sequences indeed look different than nominal sequences. We also show that these differences are apparent at least 8 hours prior to earthquake events.

This page intentionally left blank.

4. PROBLEM FORMULATION

While the FY20 approaches focused on the classification problem (*is this signal preseismic?*) on a local $30^\circ \times 30^\circ$ window of the globe, in FY21, we shift our focus to using the entire globe as our domain. Since the global ionosphere circuit is highly coupled, there may be non-local effects outside of the local window that may contribute toward the preseismic signal. By considering the entire globe as the domain, we are able to incorporate more data per prediction; however, the classification predictions now apply to the entire globe, as opposed to the local window. As such, we lose spatial specificity in the classification problem as a tradeoff. To address this, we need to perform a separate task of localization (*if this signal is preseismic, where will this earthquake occur?*).

We pivoted toward this formulation of the problem since we wanted more spatial precision than the original $30^\circ \times 30^\circ$ window provided already, so we would have had to incorporate a localization element anyway. The original $30^\circ \times 30^\circ$ window formulation had some issues – in particular, it treats all windows identically, which ignores differences in underlying geology and the effects of latitude. By considering the entire globe, we are able to represent absolute locations. The main tradeoff in switching to the global domain is the reduction in training data. We will further discuss this tradeoff in the *PATH FORWARD* section. We define the *classification* problem as follows: given a sequence of N consecutive global TEC maps spanning from times $[t - N \cdot dt, t]$, can we classify whether the signal is *preseismic with offset* T_{offset} , i.e., there will be an earthquake anywhere on the globe at $t + T_{offset}$, where T_{offset} is the “lookahead” amount of time after the end of the TEC map sequence. We define the *localization* problem as follows: given a signal classified as *preseismic with offset* T_{offset} , where on the globe will the earthquake occur? The output can be given as a coordinate (lat, lon), or a region $[lat_1, lat_2] \times [lon_1, lon_2]$.

In FY21, we focused on first refining performance in the classification problem, while also exploring possibilities for the localization problem.

This page intentionally left blank.

5. SUPERVISED PREDICTION USING DEEP NEURAL NETWORKS

5.1 BACKGROUND

Deep neural networks have established themselves as a ready solution to problems in machine vision and image recognition. Similar methods, though each with their own unique twists, include the VGG network (Simonyan & Zisserman, 2014), AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), and GoogLeNet (Szegedy, et al., 2015) among others. They work by discovering filters that detect textures at different hierarchical levels and using the response of these filters to predict which of several classes the input image belongs to. The word “deep” is in the name because of the large number of layers that the image passes through before reaching the output. Older neural networks typically only had a few hidden layers, while these modern architectures can be dozens of layers deep. These methods are considered ‘supervised’ because the weights associated with each connection are arrived at by minimizing the difference between the known class that a ‘supervisor’ has assigned to the image and the output prediction.

5.2 METHODS AND RESULTS

Typically, deep neural networks for image recognition operate on images formatted as $(u,v,3)$ matrices, where u and v are the height and width of the image and 3 accounts for the red, blue, and green channels. However, there is no reason that these networks cannot be applied to data consisting of many more layers. In our case we are using a datacube consisting of numerous time slices across the entire globe.

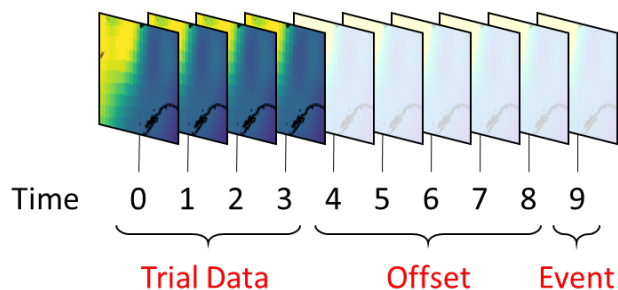


Figure 17. Setup of time slices that comprise the datacube.

Treating the sequence of frames as an image is a simplifying assumption. It does not explicitly represent the temporal relationships between the frames and treats them instead as independent “channels” of the image. Nevertheless, this simplifying assumption allows us to use well-established methods to approach the classification and localization problems. Figure 17. Setup of time slices that comprise the

datacube shows how the datacube is constructed. It consists of some number of “trial data” slices in time concatenated together. These are some amount of time (the “offset”) before the “event” which in our case is the measured earthquake. The length of the trial data and the offset are variables that can be adjusted to understand the amount of lead time afforded to decision makers.

Our proposed method for addressing the classification problem consists of nine convolutional layers separated into three groups of three banks, with each bank separated by a max pooling layer. This is referred to as the head of the neural network. The output, or “tail”, of the network is a single dense layer followed by a single node output layer. By using a sigmoid activation the output is constrained to the interval (0, 1) and can be interpreted as a probability. The max pooling layers take account of spatial variability and also reduce the overall number of weights in the system.

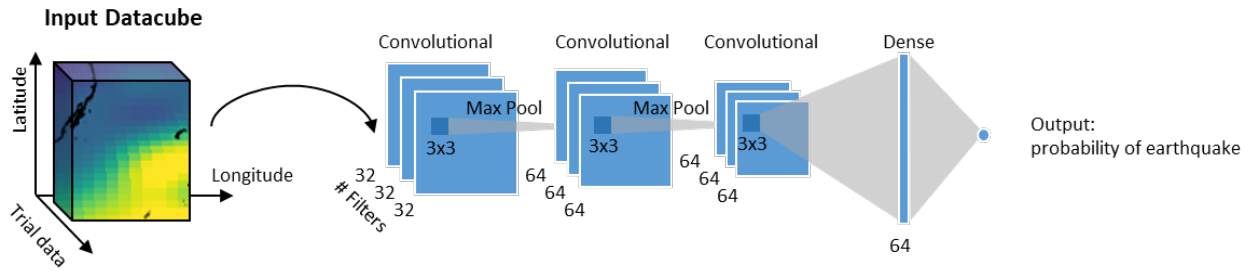


Figure 18. The proposed deep neural network for earthquake detection.

Training the network consists of minimizing the binary cross-entropy between the input labels and the output probabilities. Each data point that we used was 48 slices (12 hours) deep with a 48 slice offset. We set a threshold on positive examples at magnitude 6, which produces a training set consisting of 187 positive and 952 negative examples. Training was stopped after 200 epochs based on an observed plateauing of loss. In order to evaluate the performance of the network, another 86 positive examples and 475 negative examples were used. After training was complete, this test set was evaluated to create the ROC curve seen in

Figure 19.

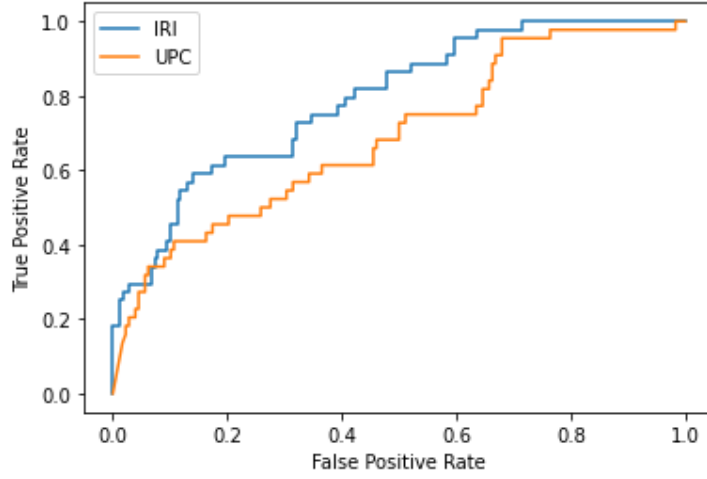


Figure 19. ROC curve from proposed deep neural network.

Figure 19 shows two different lines: IRI and UPC. As can be seen the IRI consistently outperformed the UPC, and did so across all cutoff values.

We also investigated the possibility of extending this network to addressing the localization problem. This attempt consisted of replacing the network tail with a two node layer and changing the loss function to the mean distance between predicted location and true location. Initially the distance used was the Haversine distance, but since we are treating the globe as a plane and the Haversine distance is only defined on a globe the distance metric was changed to Euclidean distance. This is permissible because large values of both metric are far from the correct location, and in the near limit the functions are identical.

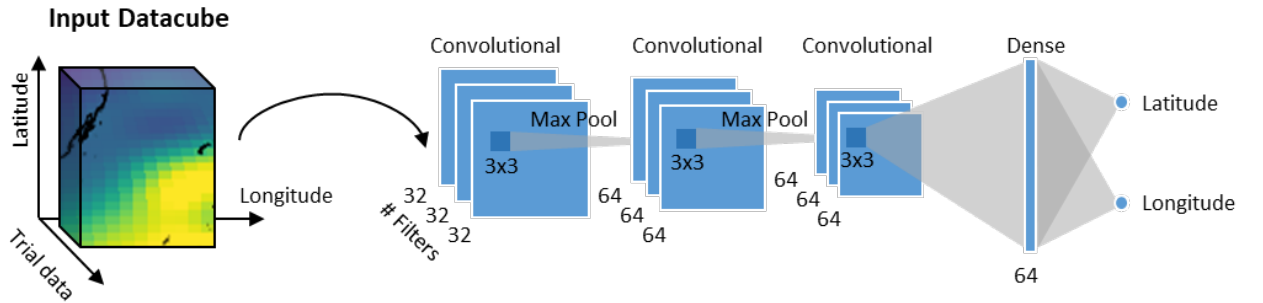


Figure 20. Neural network structure with two output nodes the attempt earthquake localization.

Unfortunately, the results of this experiment were not promising. After an initial training phase where some learning is taking place, there is a transition to overfitting. This is because of the relatively small number of training examples available. Visualizing evaluation of the test set shows no connection between the real earthquake locations and those that are predicted.

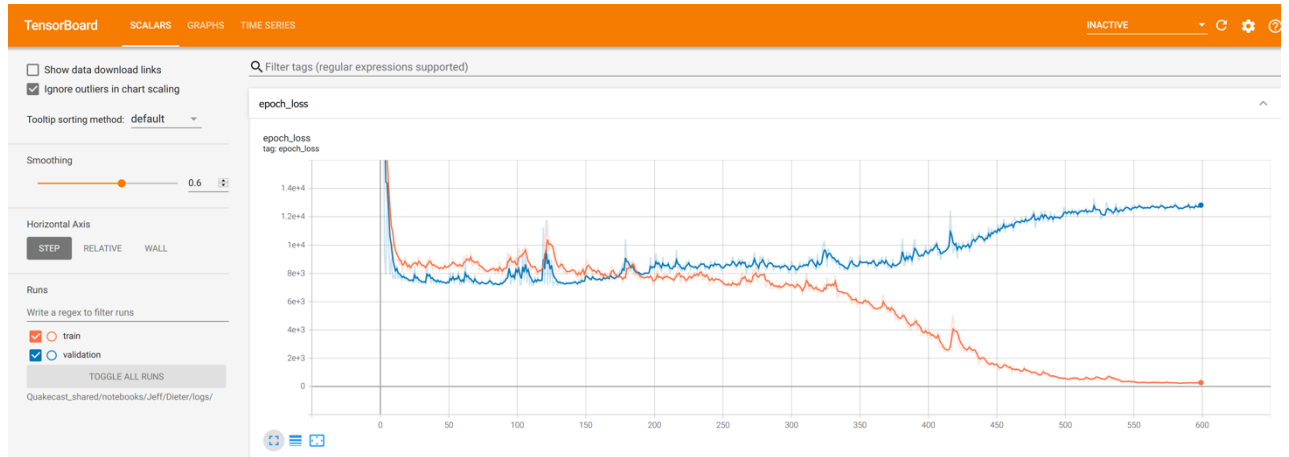


Figure 21. Evidence of Overfitting in localization problem: validation loss increases while training loss decreases.

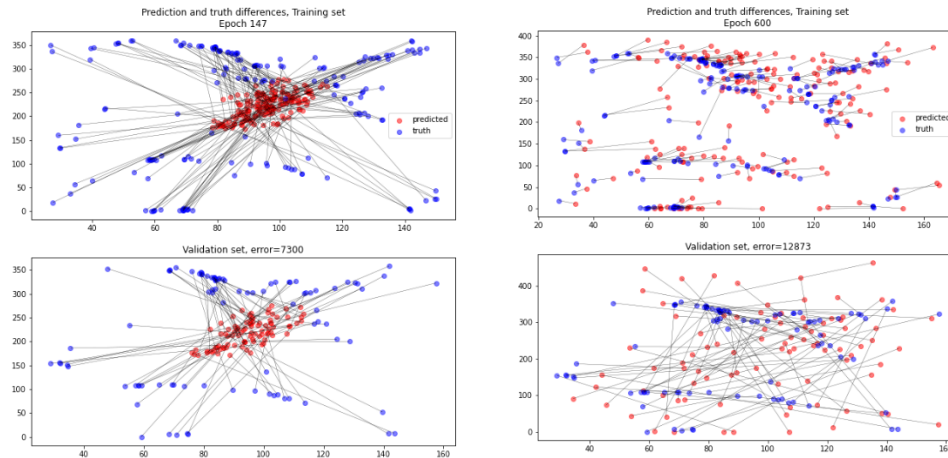


Figure 22. Estimated and true earthquake locations. Left column shows results early in the training process, right column shows results later in the training process. Top row shows performance on training set, and bottom shows performance on validation set.

As can be seen in Figure 22 the true locations and the estimated locations for the validation set have very little resemblance to each other. This is the result of the network being over-fitted to the training set. Unfortunately, the overfitting is driven by a lack of examples, which is not easily overcome. This is not to say that it is impossible to determine locations with the data that we have, but just that this method is unlikely to be fruitful.

This page intentionally left blank.

6. SPHERICAL HARMONIC TRANSFORMATION (SHT) DIMENSIONALITY REDUCTION

6.1 BACKGROUND

In the planar treatment of the globe that was used in Section 5 the space is $71 \times 72 = 5112$ dimensions. However, the presence of structures in the data means that good approximations can be made with a smaller dimensional representation. Using fewer dimensions has several advantages such as faster execution, smaller models, and reduced noise. There are many ways to reduce the dimension of a space, but for our purposes the most natural is to use the globe structure. To this end we chose a transformation called the spherical harmonic transform.

6.2 METHOD

Just like with the Fourier transform, the spherical harmonic transform treats a signal as periodic and fits it using a sum over some parameterized kernel. In the case of spherical harmonics the kernel is the product of a Legendre function and two trigonometric terms.

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{ml} Y_{ml}(\theta, \phi)$$

Here f_{ml} are the spherical harmonic coefficients (which are dependent on the function being decomposed) and Y_{ml} are the spherical harmonic kernels (which are independent). If the first summand is taken over some discrete set then $\tilde{f}$ will be an approximation of the function being modeled. A full treatment of the SHT process can be found in (Wieczorek & Meschede, 2018).

For any time slice we can transform the values into a vector of coefficients. These vectors can then be put through a neural network with a binary output. If we use 16 coefficients then the reconstruction has a mean squared error of 1.3 or about 7%.

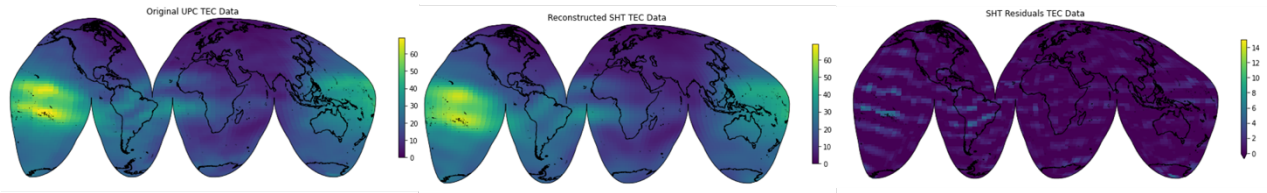


Figure 23. (left to right) The original UPC data; the reconstructed function based on 16 coefficients; the difference between the original and the reconstruction.

As mentioned previously, the SHT functions to remove noise. This is because the noise is typically high frequency and that content is contained in the higher order coefficients. By using only the first few coefficients this information is removed.

A slightly different network structure is required for using SHT data. The 2D convolutional filters from the deep network structure are replaced with 1D convolutional filters that operate in the direction of time. Similarly max pooling occurs in one dimension. This is depicted in Figure 24.

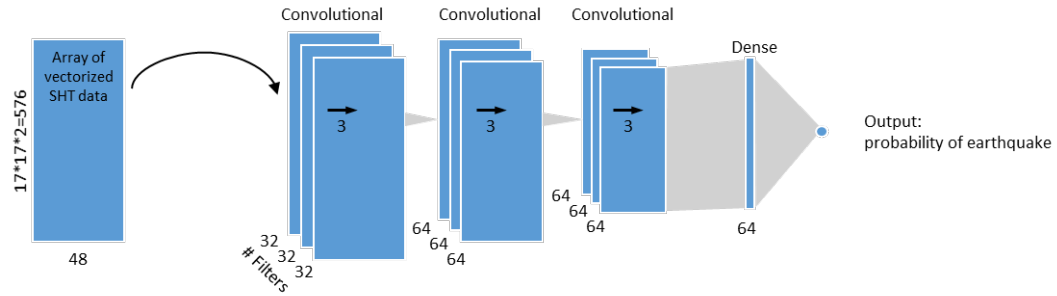


Figure 24. Network used with the SHT data.

Other than the slight changes to network structure, the training was performed according to the same protocol as in the deep CNN case. This structure reduces the dimension of the data from 245,376 dimensions to 27,648 dimensions. Even with the reduced dimensionality, the pooling only occurring in one direction means that the number of weights is about 6x what is seen in the deep CNN case.

The results of evaluation of the test set is depicted in Figure 25 as a ROC curve.

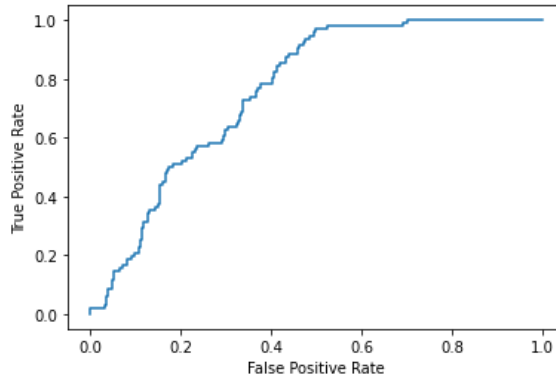


Figure 25. ROC curve for SHT network.

There does not appear to be any significant improvement in performance by the SHT network over the deep convolutional network described in Section 5. Execution time is not meaningfully reduced, but this is due in part to the relatively quick execution of both models. The SHT model has an added overhead of having to compute the SHT transform, but this can be done as a data preprocessing step prior to training and inference. While no immediate improvements have been found in using the SHT, this does not mean that it should be abandoned. For certain network architectures it may be possible that the SHT is more efficient. In addition, our planned use of the higher resolution VISTA TEC maps increases the number of grid points by an order of magnitude, which may make the SHT representation comparatively more advantageous. There are also different structures that are facilitated by this vectorization, such as the projection into some other space.

This page intentionally left blank.

7. VIDEO ACTION TRANSFORMER NETWORK

7.1 ACTION DETECTION PROBLEM

With the Convolutional Neural Network approach, we treated the entire sequence of frames as a single image. The convolution filters of the CNN operate on each frame independently, and thus only extract spatial features within each frame. There exist techniques developed for video data which additionally consider temporal relationships in addition to the spatial relationships already considered. This is important because the signature of the signal may not be apparent when looking at individual frames, but may appear when looking at the temporal relationship between features. We find that there is an existing problem in video data processing which provides a suitable analog to our preseismic signal detection problem: the action detection problem.

The action detection problem is the problem of identifying and localizing human actions within videos. Actions represent activities that a human may do, such as walking, speaking, carrying objects, etc. The action recognition problem differs from the image recognition problem in that some actions are indistinguishable when considering still frames, and only distinguish themselves when temporal relationships are considered. Take, for example, the actions of “running” and “skipping.” Any individual frame of each of the two actions will appear similar to one another; it is only in the temporal relationships between frames that the actions appear differently. As such, to distinguish between the two actions, the ML algorithm will have to have a way to encode the temporal relationships between frames. Recurrent Neural Networks, such as the LSTM, have been used to represent the sequence data; however, recently, Transformer architectures have become popular alternatives to RNNs in sequence modeling.

7.2 TRANSFORMER NETWORKS

Transformers are network architectures which include a mechanism called “attention,” which learns the contribution of the surrounding features to the classification of a given element. They have become popular in particular due to their success in Natural Language Processing, where the attention mechanism discovers which other words in a sentence modify a given word (Vaswani et al., 2017). The attention mechanism can be thought of as learning the “context” of the element being analyzed. This principle has been carried over to video analysis, where attention can learn which pixels and frames contribute toward the prediction of a given action. We consider an architecture called the Video Action Transformer Network (VATN) (Girdhar et al, 2019) which utilizes transformers to detect actions in videos. The VATN performs both the localization and classification task simultaneously, using a region proposal network (RPN) for localization, and the transformer for classification.

We were able to find a partial implementation of the VATN, which implemented the transformer classification portion of the network, but not the localization component. We performed initial studies of the classification performance, and found that it performed statistically significantly better than chance

(~5% margin), but addition work is required to fully characterize the performance and compare it to the other approaches. In addition, we plan to incorporate a region proposal network, which is available in other open source classifiers into the VATN architecture to enable localization capabilities.

8. PATH FORWARD

Work in FY21 was focused on building the ML infrastructure to accelerate data and algorithmic development. Moving forward, we will leverage and continue this work to integrate additional data sources, explore new algorithmic approaches, and refine predictive capabilities.

Incorporate of higher spatial and temporal resolution TEC data: We hypothesize that the ionosphere features related to preseismic processes are at the mesoscale or smaller and relatively short timescales compared to other known disturbance in the global electric circuit. Therefore, utilizing higher spatial and temporal resolution TEC data may enable the models to better resolve the instances of preseismic signals; including locality of the epicenter and magnitude. We plan to use the VISTA-processed high resolution global TEC maps as an additional data source.

Filter TEC data based on quality: Currently, limited sensors for measuring TEC data are available in very remote land regions and over ocean areas. Therefore, the quality of TEC measurements in these area is diminished. For the purposes of this program, our areas of interest for earthquake forecasting are in populated areas. By filtering out or down-weighting the importance of predicting earthquakes that occur in remote land regions and ocean areas where TEC sensor quality and data is poor, our models may show improved performance by training on only the high quality data.

Incorporate localization techniques: while our first attempt at localization failed, there are other methods of localization that have shown promise in other studies. One approach is a reduced dimensional space grid and try to localize to a grid cell. The proposed method would be similar to what is proposed in the YOLO architecture (Redmon, Divvala, Girshick, & Farhadi, 2016) where each grid cell is assigned a confidence of an earthquake occurring with that cell. Another approach uses the Region Proposal Network framework which is common in Regional-CNN architectures.

Windowed transforms: currently we are using transforms for dimensionality reduction that are applied in a holistic manner across globe (i.e., SHT). However, since the phenomenon is believed to be exhibited locally there may be discoveries to be made by analyzing the data in a sliding window that ranges over the spatial domain. Such techniques include the Gabor transform and wavelet analysis.

Speed up processes: Training the VATN and other large networks requires significant computational power. We are currently using a single node with two GPUs for network training. We are looking into performing training across multiple nodes and GPUs to accelerate the process. In particular, we are looking into the use of pytorch-lightning with the LL-developed Hydra-zen toolkit to simplify multi-node training and performing parameter sweeps.

Characterization of signal and physical models: The deductive, statistical nature of our approach means that while we are quite confident that a signal exists in the ionosphere TEC data, which can help predict earthquakes, we do not know about the physical characteristics of the signal. A characterization of

the physical characteristics and processes related to this signal would provide valuable insight for improving an earthquake prediction system. If a transformer-based approach proves capable of predicting the earthquakes, we can analyze the attention used in classification to identify which regions and times contribute the most toward the prediction. This can help inform possible physical theories and motivate future sensing efforts.

9. SUMMARY

In this report, we presented an overview of the QuakeCast project, including a summary of FY20 results, and a detailed explanation of FY21 efforts. Specifically, in FY21, we revamped our data workflow and created a unified ionosphere dataset which integrates multiple data sources into a single interface; refined the problem definition to include both the classification and localization problems; and explored several additional model architectures. The retooling of the dataset allowed us to more quickly iterate on algorithm development, compute derived features and transformations, and also facilitated onboarding additional staff members due to the simplified data access procedures. The formulation of the separate classification and localization tasks allowed us to study the problems independently and reduce complexity in our approach.

Nevertheless, we still face a number of challenges. We are consistently able to outperform random chance, but are still a ways away from generating foolproof predictions. This is due, in large part, to the limitations in both quantity and quality of data. We are hoping that the VISTA high-resolution dataset helps to address the completeness issue that we encountered with the UPC high-rate dataset, and the limited resolution issues with the IRI dataset. However, since the VISTA dataset is not yet available, we are unable to evaluate this. The IRI model can potentially help with some of the data limitation issues, in that we can use it to generate “fuzzed” observations corresponding to uncertainties in some parameters. However, this only allows us to smooth out existing examples in their local neighborhood, and cannot be used to generate new examples “from whole cloth.” Filtering or reweighting the importance of predictions to prioritize detecting earthquakes near high-quality TEC data can help address data quality issues in areas with limited TEC detection capability.

We also plan to further study the localization problem, and hope to introduce a magnitude estimation task, if possible. Our initial attempt at localization was prone to overfitting due to the limited amount of data available for our given approach. One cause of this is that using the entire data cube to estimate the location results in too many free parameters, enabling the network to just “memorize” training examples. Techniques which use bounding boxes and define a foreground and background class may be able to better focus the network into just learning the features relevant to the earthquake.

This page intentionally left blank.

REFERENCES

1. Blagoveshchensky, Donat, et al. "Impact of Magnetic Storms on the Global TEC Distribution." *Ann. Geophys.*, 36, 1057-1071, 2018, <https://doi.org/10.5194/angeo-36-1057-2018>
2. Bilitza, Dieter, "IRI the International Standard for the Ionosphere." *Adv. Radio Sci.*, 16, 1-11, 2018, <https://doi.org/10.5194/ars-16-1-2018>
3. Bondár, I., et al., "ISC-GEM: Global Instrumental Earthquake Catalogue (1900–2009), II. Location and Seismicity Patterns." *Physics of the Earth and Planetary Interiors*, vol. 239, Feb. 2015, pp. 2–13, doi:10.1016/j.pepi.2014.06.002.
4. Dask Development Team. *Dask: Library for Dynamic Task Scheduling*. 2016, <https://dask.org>.
5. Devlin, Jacob, et al. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *ArXiv:1810.04805 [Cs]*, May 2019. *arXiv.org*, <http://arxiv.org/abs/1810.04805>.
6. Di Giacomo, Domenico, James Harris, et al., "ISC-GEM: Global Instrumental Earthquake Catalogue (1900–2009), I. Data Collection from Early Instrumental Seismological Bulletins." *Physics of the Earth and Planetary Interiors*, vol. 239, Feb. 2015, pp. 14–24, doi:10.1016/j.pepi.2014.06.003.
7. Di Giacomo, Domenico, István Bondár, et al., "ISC-GEM: Global Instrumental Earthquake Catalogue (1900–2009), III. Re-Computed MS and Mb, Proxy MW, Final Magnitude Composition and Completeness Assessment." *Physics of the Earth and Planetary Interiors*, vol. 239, Feb. 2015, pp. 33–47, doi:10.1016/j.pepi.2014.06.005.
8. Di Giacomo, Domenico, E. Robert Engdahl, et al., "The ISC-GEM Earthquake Catalogue (1904–2014): Status after the Extension Project." *Earth System Science Data*, vol. 10, no. 4, Oct. 2018, pp. 1877–99, doi:<https://doi.org/10.5194/essd-10-1877-2018>.
9. Freund, et al., "Cracking the Code of Pre-Earthquake Low Frequency EM Emissions." 2005.
10. Girdhar, Rohit, et al. "Video Action Transformer Network." *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2019, pp. 244–53. *DOI.org (Crossref)*, <https://doi.org/10.1109/CVPR.2019.00033>.
11. Hoyer, Stephan, and Joe Hamman. "Xarray: N-D Labeled Arrays and Datasets in Python." *Journal of Open Research Software*, vol. 5, no. 1, 1, Ubiquity Press, Apr. 2017, p. 10. *openresearchsoftware.metajnl.com*, <https://doi.org/10.5334/jors.148>.
12. Kuo, Cheng-Ling, et al., "Electrical Coupling Between the Ionosphere and Surface Charges in the Earthquake Fault Zone." *Geophysical Monograph Series*, edited by Dimitar Ouzounov et al., John Wiley & Sons, Inc., 2018, pp. 99–124, doi:10.1002/9781119156949.ch7.
13. Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097-1105.

14. Liu, Cheng-Yan, et al., “Statistical Analyses on the Ionospheric Total Electron Content Related to $M \geq 6.0$ Earthquakes in China during 1998 - 2015.” *Terrestrial, Atmospheric and Oceanic Sciences*, vol. 29, no. 5, 2018, pp. 485–98, doi:10.3319/TAO.2018.03.11.01.
15. McKinney, Wes. “Data Structures for Statistical Computing in Python.” *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman, 2010, pp. 56–61, <https://doi.org/10.25080/Majora-92bf1922-00a>.
16. Molchanov, O., et al., “Lithosphere-Atmosphere-Ionosphere Coupling as Governing Mechanism for Preseismic Short-Term Events in Atmosphere and Ionosphere.” *Natural Hazards and Earth System Sciences*, vol. 4, no. 5/6, Nov. 2004, pp. 757–67, doi:10.5194/nhess-4-757-2004.
17. Noll, Carey E. “The Crustal Dynamics Data Information System: A Resource to Support Scientific Analysis Using Space Geodesy.” *Advances in Space Research*, vol. 45, no. 12, June 2010, pp. 1421–40, doi:10.1016/j.asr.2010.01.018.
18. Redmon, Joseph; Divvala, Santosh; Girshick, Ross; Farhadi, Ali (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779-788.
19. Simonyan, Karen; Zisserman, Andrew (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
20. Storchak, D. A., et al., “The ISC-GEM Global Instrumental Earthquake Catalogue (1900–2009): Introduction.” *Physics of the Earth and Planetary Interiors*, vol. 239, Feb. 2015, pp. 48–63, doi:10.1016/j.pepi.2014.06.009.
21. Storchak, Dmitry A., et al., “Public Release of the ISC–GEM Global Instrumental Earthquake Catalogue (1900–2009).” *Seismological Research Letters*, vol. 84, no. 5, Sept. 2013, pp. 810–15, doi:10.1785/0220130034.
22. Sun, Hu, et al., “Matrix Completion Methods for the Total Electron Content Video Reconstruction.” *arXiv:2012.01618v1*, 3 December 2020.
23. Szegedy, Christian; Liu, Wei; Jia, Wangqing; Sermanet, Pierre; Reed, Scott; Anguelov, Dragomir; Erhan, Dumitru; Vanhoucke, Vincent; Rabinovich, Andrew (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.
24. Unidata. “Unidata | NetCDF.” *NetCDF*, <https://doi.org/10.5065/D6H70CW6>. Accessed 3 Nov. 2021.
25. Vaswani, Ashish, et al. “Attention Is All You Need.” *Advances in Neural Information Processing Systems*, edited by I. Guyon et al., vol. 30, Curran Associates, Inc., 2017, <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>.
26. Wiczorek, Mark; Meschede, Mathias (2018). SHTools: Tools for Working with Spherical Harmonics. *Geochemistry, Geophysics, Geosystems*, 2574-2592
27. Zhou, Yiyang, et al., “Ionospheric Disturbances Associated with the 2015 M7.8 Nepal Earthquake.” *Geodesy and Geodynamics*, vol. 8, no. 4, 2017, pp. 221–28, doi:10.1016/j.geog.2017.04.004.

28. Zolotov, et al., "TEC Disturbances Prior to the Great Tohoku March 11, 2011 and October 23, 2011 Turkey Van Earthquakes and Their Physical Interpretation." Proceedings of the MSTU, Vol. 15, No. 3, 2012

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (10-20-2022)		2. REPORT TYPE Project Report, Technical Investment Program (TIP)		3. DATES COVERED (From - To)
4. TITLE AND SUBTITLE QuakeCast: Earthquake Forecasting Using Machine Learning to Identify and Classify Preseismic Ionospheric Anomalies as Earthquake Signals		5a. CONTRACT NUMBER		
		5b. GRANT NUMBER		
		5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Bhavani Ananthabhotla, Jeffrey Liu, Jessica Reid, Samuel Scheele, Dieter Schuldt, Matthew Weiss		5d. PROJECT NUMBER TI90-3201		
		5e. TASK NUMBER		
		5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02420-9108		8. PERFORMING ORGANIZATION REPORT TIP-159		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Laboratory Services The Air Force Life Cycle Management Center (AFLCMC/AZS), Bhupender Singh MIT Lincoln Laboratory 244 Wood Street, Lexington, MA, 02421		10. SPONSOR/MONITOR'S ACRONYM(S) AFLCMC/AZS; MIT LL		
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION / AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001.				
13. SUPPLEMENTARY NOTES				
14. ABSTRACT Earthquakes are deadly, expensive, and unpredictable. To date, the study of earthquakes has been predominantly through seismology, which has enabled researchers to describe the long-term progression of earthquakes, but is unable to provide early forecasting necessary for short-term disaster preparation. Global Navigation Satellite Systems (GNSS) satellites offer a new opportunity to study earthquakes and begin to unravel the complex interactions between preseismic processes and the global electric circuit that may be the key to earthquake early forecasting. The QuakeCast project is developing a novel machine learning method to identify ionosphere anomalies as preseismic earthquake signals and uses them to predict impending magnitude, geographical region, and timing.				
15. SUBJECT TERMS				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 57
a. REPORT	b. ABSTRACT	c. THIS PAGE		
				19a. NAME OF RESPONSIBLE PERSON
				19b. TELEPHONE NUMBER (include area code)

This page intentionally left blank.