SANDIA REPORT

SAND2020-3767
Unlimited Release
Printed March 2020

The use of seismic spatial gradients in a single layer neural network for seismic source discrimination: proof of concept

Christian Poppeliers

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

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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.

Sandia National Laboratories

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

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

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

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2020-3767
Unlimited Release
Printed March 2020

The use of seismic spatial gradients in a single layer neural network for seismic source discrimination: proof of concept

Christian Poppeliers

Abstract

This report describes a proof-of-concept method of seismic source discrimination using seismic gradiometry and a common machine learning technique. The tests described here are purely numerical, using synthetic seismic data and well understood mathematical techniques. The primary innovation described here is the application of a richer seismic data set derived from seismic gradiometry. Seismic gradiometry is a method to estimate the time variable spatial gradient of the wavefield to compute various wavefield attributes such as slowness, dynamic strain, and rotational motions. With the addition of these wavefield attributes, we are afforded up to twenty "components" of time series data measured at a single point on, or in, the Earth. This is in direct contrast to conventional three-component seismic data collected at several locations using a seismic network. Using the gradiometrically-derived wavefield components directly in a single-layer neural network, I show that it is possible to discriminate between three common seismic source types (earthquakes, explosions, and opening fractures) for various noise conditions and gradiometry configurations.

Acknowledgment

This work is funded by Sandia's Laboratory Directed Research and Development (LDRD), Earth Science Research Foundation program. L. Preston and B. Young (both of the Geophysics Department) provided an internal review, significantly improving the quality of this report.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Analysis of seismic data can reveal the nature of the seismic source. The motivation for seismic source studies can vary from academic studies of earthquake mechanisms (e.g. Dziewonski and Woodhouse, 1983; Ekstrom *et al*, 2012), seismic hazard assessment (Duputel *et al*, 2014), monitoring of fracture network development during hydraulic fracturing operations (Gu *et al.*, 2018), or nuclear nonproliferation (Dreger and Woods, 2002; Chouet *et al*, 2003). Regardless of the specific reason for conducting a seismic source study, the primary methods of data collection relies on a distributed network of seismic instrumentation that (ideally) surrounds the source. The ultimate goal of this type of analysis is to fit seismic amplitudes to a representative model of the seismic source, where the model predicts the radiation pattern of the seismic energy. The output of these types of analysis is typically a set of parameters that describe the seismic moment tensor. The seismic moment tensor is a simplified, linear-equivalent, representation of the seismic source where each term describes the amplitude and orientation of each moment term. The terms in the seismic moment tensor describe the orientation of the equivalent moment couples that produce the seismic waves. In the case of a pure earthquake source, the moment couples are related to the relative orientation of the fault plane, it's motions, and the subsequent relative amplitude and direction of the shear and compressional waves. It's important to emphasize that for an earthquake source, the radiation pattern for both compressional (P) and shear (S) waves is highly directional. Conversely, for an explosion source, there are no shear components and the (purely compressional) wave energy radiates outwards in all directions equally. These differences gives us a powerful tool to discriminate between earthquake and explosion seismic sources, but most analysis techniques to date rely on a distributed array of seismic instruments to give robust results.

The work I describe here is focused primarily on discrimination. The approach I take does not attempt to resolve the components of the seismic moment tensor, but rather treats the problem as a classification task. Specifically, I focus on developing a classification scheme to discriminate between various seismic source types, using a machine learning-based algorithm. The primary innovation presented here is that rather than use an array of distributed seismic instruments, I use a single, compact seismic array, termed a gradiometer (e.g. Langston, 2009), to extract tens of additional seismic wave attributes from the three-component seismic data. This type of analysis, termed seismic wave gradiometry, or gradiometry for short, uses a very dense array (e.g. a station spacing of only a few meters) of three-component seismic instruments to estimate the spatial gradients of the seismic wavefield. From the gradients, we can form additional "components" of data, such as dynamic strains, rotational motions, and estimations of the local change in the radiation pattern. I explore the use of a single layer neural network, where the data are the gradient-derived wave attributes. Preliminary results favorably suggest that using gradiometerically-derived wave attributes as input to a neural network can successfully classify seismic source types, even when the data contain a significant amount of correlated, colored noise.

Chapter 2

The seismic source

The basic premise of using seismic signals for source analysis is that the energy flux from a seismic event depends on the nature of the seismic source. A linear model describing seismic waves shows that seismograms in the far field are a convolution of a series of equivalent volumetric sources and the Earth's impulse response:

$$u_k(\mathbf{x}',t') = \int_{-\infty}^{\infty} \int_{V_0} g_{ki}(\mathbf{x}',t';\mathbf{x},t) S_i(\mathbf{x},t) dV dt \tag{2.1}$$

where u_k is the displacement seismogram at station k, g_{ki} is the Green's function describing the impulse response from source i located at $\mathbf{x}$ to receiver k at $\mathbf{x}'$, and S_i is the i^{th} finite volume general source of volume V_0 acting at point $\mathbf{x}$ (Aki and Richards, 2002; Stump and Johnson, 1977). By representing the volumetric source S_i as a series of time variable force couples m_i acting at point $\mathbf{x}$, equation 2.1 reduces to

$$u_k(\mathbf{x}',t') = \sum_{i=1}^{6} \int_{-\infty}^{\infty} g_{ki}(\mathbf{x}',t';\mathbf{x},t) m_i(\mathbf{x},t) dt. \tag{2.2}$$

In many cases, the term describing time time variable force couples m_i is simplified by assuming that the time dependence of the individual force couples are all similar. In this case,

$$m_i(\mathbf{x},t) = h(t)\widehat{m}_i(\mathbf{x}) \tag{2.3}$$

where $h(t)$ can be any arbitrary function and the term $\widehat{m}_i(\mathbf{x})$ is a rank-2 tensor:

$$\widehat{m}_i(\mathbf{x}) = \widehat{M}_{mn}(\mathbf{x}) = \begin{bmatrix} M_{xx} & M_{xy} & M_{xz} \\ M_{yx} & M_{yy} & M_{yz} \\ M_{zx} & M_{zy} & M_{zz} \end{bmatrix}, \tag{2.4}$$

which describes the relative scale and orientation of the source time function $h(t)$. The $m \times n$ tensor $\widehat{M}_{mn}(\mathbf{x})$ (of size 3×3) is often referred to as the seismic moment tensor. Note that the subscript i goes from 1 to 6, which I replace with the more explicit tensor coordinates, m and n. Thus the summation variable i in equation 2.2 includes only six terms, one for each independent and unique tensor component ($\widehat{M}_{mn}(\mathbf{x})$ is symmetric about the diagonal and therefore contains only six unique components: $M_{mn} = M_{nm}$).

A physical interpretation of the seismic moment tensor is that each term describes a force couple: the product of a physical quantity (in this case, a force vector) and a distance from an origin where the force is applied. The subscripts have the following physical interpretations: the on-diagonal term ($m = n$) indicates a force, applied at the origin, parallel to the axis indicated: e.g. M_{xx} denotes a force applied at the origin, in the x direction. Conversely, the off-diagonal terms are torques: for example, M_{xy} indicates force in the x direction applied at (and perpendicular to) the y axis but applied at unit distance from the origin (i.e. a torque). A visual representation of the force couples is shown in Figure 3.7 in Aki and Richards (1980).

A potential goal of seismic source studies is to estimate the terms in the moment tensor, which can then be used to quantify the relative contribution of each force couple. For example, an idealized explosion source consists of a rapid expansion of a volume of the Earth, where the expansion is equal in all directions. In this case, the moment tensor takes the form

$$A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.5}$$

where the term A relates to the total energy of the source. For this simple model, the source is termed isotropic, meaning that energy radiates outwards equally in all directions. Furthermore, there is no shear wave energy being generated, as all of the off-diagonal terms in the tensor are zero. Conversely, an earthquake results when two bodies of rock suddenly slip relative to each other along a fault. In this case, there is ample S wave energy generated, along with P wave energy and the radiation pattern of energy is directionally dependent. The moment tensor that describes this type of source, termed a double-couple source, is symmetric, with potentially non-zero terms on the off-diagonals (note: the off-diagonal terms indicate shear wave energy will be generated). For example, a vertically oriented, right lateral, strike slip earthquake fault with a strike of 0 degrees will have a normalized moment tensor of

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{2.6}$$

and a fault with a strike of 10 degrees, a dip of 85 degrees, and a rake of 5 degrees will have a normalized moment tensor of

$$\begin{bmatrix} -0.0864 & 0.4676 & -0.0353 \\ 0.4676 & 0.1624 & -0.0498 \\ -0.0353 & -0.0498 & -0.0760 \end{bmatrix}. \tag{2.7}$$

Note that for a pure double-couple source, the trace $tr = \sum_{i=1}^{3} M_{i,i}$ of the moment tensor will be zero. An important attribute of a double-couple source is that the energy flux of both P and S waves is highly directional which is in contrast to that of an isotropic source.

Based on this physics, there are two primary attributes that are commonly used to discriminate between source types. First, the relative energy content of P and S wavefields, and second, the azimuthal dependence of the wavefield's energy (i.e. the radiation pattern). For simple discriminants, the P/S energy is used, where this ratio will be much higher for explosion sources than for earthquake sources. More sophisticated discriminants operate by inverting the seismic data for the terms in the moment tensor, resulting in an estimate of six scalars that quantify the relative contribution of the tensor terms, which are often interpreted geometrically using "beach ball" graphics (Dziewonski *et al.*, 1981; Dziewonski & Woodhouse, 1983), so-called Hudson plots (Hudson *et al.*, 1989), or lunes (Tape and Tape, 2012).

There are many complications that can adversely impact the quality of the analysis described in the previous two paragraphs, such as noise in the data, scattering of the seismic wavefield, non-ideal or mixed source types, and overly sparse seismic networks. However, I will emphasize that both methods typically rely on a distributed array of instrumentation to perform well. This is even the case for methods that rely on the ratio of *S* and *P* energy, as radiation pattern effects must be considered. For example, for a given earthquake, the radiation pattern for *P* waves is "out of phase" with the radiation pattern of the *S* waves (e.g. the azimuth of the maximum energy flux for *P* waves corresponds to the minimum energy flux for *S* waves). In the following sections, I describe a different paradigm to address the seismic discrimination problem, which is based on attributes estimated from spatial gradients of the observed seismic wavefield.

Chapter 3

Seismic gradiometry

The discriminate that I develop here uses the spatial gradients of the seismic wavefield, as well as wave attributes computed from them. Therefore, in this chapter, I give a brief overview of the method of seismic gradiometry. I will begin by describing the underlying model, how to estimate the seismic gradients, and the quantities that can be derived from them.

The basis for seismic wave gradiometry is that a propagating wave may contain geometrical spreading and wave slowness that might change with distance. For example,

$$u(\mathbf{x},t) = \mathbf{G}(\mathbf{x})f(t - \mathbf{p}(\mathbf{x} - \mathbf{x}_0)) \tag{3.1}$$

where $\mathbf{x}$ is the position vector, $\mathbf{G}$ is the distance-dependent geometrical spreading of the wave, $\mathbf{p}(\mathbf{x})$ is the slowness (or reciprocal velocity), and $\mathbf{x}_0$ is the reference position. The spatial gradient of this wave is

$$\frac{\partial u}{\partial \mathbf{x}} = \mathbf{A}u + \mathbf{B}\frac{du}{dt} \tag{3.2}$$

where u is the displacement seismogram at point $\mathbf{x}$, the coefficient $\mathbf{A}$ relates to the geometrical spreading of the wave

$$\mathbf{A} = \frac{1}{\mathbf{G}(\mathbf{x})}\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}}, \tag{3.3}$$

and $\mathbf{B}$ relates to the wave slowness

$$\mathbf{B} = -\mathbf{p} - \frac{\partial \mathbf{p}}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_0). \tag{3.4}$$

The above description shows the relationships between the data and its spatial and temporal derivatives in any of three Cartesian dimensions. Note that these relationships are valid regardless of the component of seismic data used. Furthermore, there exist several methods of solving for the gradiometry coefficients $\mathbf{A}$ and $\mathbf{B}$ in one dimension ($\mathbf{x}$=x) (Langston, 2007a) two dimensions ($\mathbf{x} = [x,y]$) (Langston 2007b; Langston, 2007c) as well as three dimensions ($\mathbf{x} = [x,y,z]$) (Poppeliers *et al.*, 2013; Poppeliers and Punosevec, 2013).

Estimating Seismic Spatial Gradients and Gradiometric Wave Attributes

The fundamental quantities used in the analysis here are the temporal and spatial derivatives of the the various components of the displacement wave field, averaged over the spatial extent of the gradiometer. A gradiometer is essentially measuring the local slope of the displacement wavefield, and thus the aperture of the gradiometer must be scaled appropriately. Typically, a gradiometric array must have an aperture of less than 10% of the dominant wavelength of interest to avoid numerical artifacts and/or aliasing. However, if the gradiometer is too small, then the measurements become sensitive to uncorrelated noise in the data. Poppeliers and Evans (2015) showed that a gradiometer with an aperture in the range of 3% - 8% of the seismic wavelength of interest is best for optimizing the tradeoffs between noise sensitivity and numerical artifacts.

The individual elements of a gradiometry array typically measure the ground velocity as a function of position and time, and thus each seismic station is measuring the temporal derivative of the displacement wave field directly. To estimate the spatial gradient of the wave field at a single point in time t at surface location $[x_0, y_0]$ I use the method described by Liang and Langston (2009) where data from the N seismic stations of a gradiometer array are used to estimate the wave spatial gradients:

$$\begin{bmatrix} u_t(x_1,y_1) - u_t(x_0,y_0) \\ u_t(x_2,y_2) - u_t(x_0,y_0) \\ \vdots \\ u_t(x_N,y_N) - u_t(x_0,y_0) \end{bmatrix} = \begin{bmatrix} \delta x_1 & \delta y_1 \\ \delta x_2 & \delta y_2 \\ \vdots & \vdots \\ \delta x_N & \delta y_N \end{bmatrix} \begin{bmatrix} \frac{\partial u_t(x_0,y_0)}{\partial x} \\ \\ \frac{\partial u_t(x_0,y_0)}{\partial y} \end{bmatrix}, \tag{3.5}$$

where $u_t(x_i, y_i)$ is the displacement seismic wave measured at time t and location $[x_i, y_i]$ and the location $[x_0, y_0]$ indicates the location of the of a reference, or master, station from which the differences are obtained. The terms $\delta x_i and \delta y_i$ describe the geometry of the gradiometry (e.g. $\delta x_i = x_i - x_0$ for station i).

The term on the left hand side of equations 3.5 is a vector containing the waveform variations relative to a user-defined master station. On the right hand side, the $N \times 2$ matrix contains the coordinates of the individual seismometers in the gradiometer and the 2×1 vector contains the estimates of the horizontal spatial wave gradients at location $[x_0, y_0]$. This analysis can be extended to three dimensions as

$$\begin{bmatrix} u_t(x_1,y_1,z_1) - u_t(x_0,y_0,z_0) \\ u_t(x_2,y_2,z_3) - u_t(x_0,y_0,z_0) \\ \vdots \\ u_t(x_N,y_N,z_N) - u_t(x_0,y_0,z_0) \end{bmatrix} = \begin{bmatrix} \delta x_1 & \delta y_1 & \delta z_1 \\ \delta x_2 & \delta y_2 & \delta z_2 \\ \vdots & \vdots & \vdots \\ \delta x_N & \delta y_N & \delta z_N \end{bmatrix} \begin{bmatrix} \frac{\partial u_t(x_0,y_0)}{\partial x} \\ \\ \frac{\partial u_t(x_0,y_0)}{\partial y} \\ \\ \frac{\partial u_t(x_0,y_0)}{\partial z} \end{bmatrix}, \tag{3.6}$$

By solving equations 3.5 or 3.6 via standard matrix inversion techniques, we obtain an estimate of the wave spatial gradients for a given time point t. Although there are other methods to estimate the gradients (e.g. Langston, 2007), equations 3.5 and 3.6 offer two advantages: first, the array geometry can be completely arbitrary; however, it's desirable from an accuracy standpoint that the the array be roughly equal in the x and y (and potentially z) dimensions. Finally, the master station at $[x_0, y_0, z_0]$ is arbitrary. Therefore, it's possible to construct the matrix equations 3.5 N times, where for each case the master station is chosen to be one of each of the N possible stations. This results in N estimates of the spatial gradient for each time point, allowing for crude error estimates. A potential disadvantage of using this method to estimate the spatial gradients is that equations 3.5 or 3.6 estimate the gradients by fitting the gradient values to a plane, in an L-2 sense.

This could potentially result in single spurious stations having an unacceptably large influence on the calculated value of the gradient. Although not the goal of the research here, it's possible to solve equations 3.5 or 3.6 using a different method that would minimize the effects of outliers, such as robust least squares or methods that fit the data using the L-1 norm.

The spatial gradients of the wavefield allow us to estimate additional wave attributes such as strain and rotational motions. For example, the displacement **u** of a point **x** is related to a nearby point $(\mathbf{x}+\delta\mathbf{x})$ by

$$\mathbf{u}(\mathbf{x}+\delta\mathbf{x}) = \mathbf{u}(\mathbf{x}) + \varepsilon\delta\mathbf{x} + \vec{\omega}\times\delta\mathbf{x}, \tag{3.7}$$

where

$$\varepsilon = \begin{bmatrix} \frac{\partial u_x}{\partial x} & \frac{1}{2}\left(\frac{\partial u_x}{\partial y}+\frac{\partial u_y}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_x}{\partial z}+\frac{\partial u_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u_y}{\partial x}+\frac{\partial u_x}{\partial y}\right) & \frac{\partial u_y}{\partial y} & \frac{1}{2}\left(\frac{\partial u_y}{\partial z}+\frac{\partial u_z}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial u_z}{\partial x}+\frac{\partial u_x}{\partial z}\right) & \frac{1}{2}\left(\frac{\partial u_z}{\partial y}+\frac{\partial u_x}{\partial z}\right) & \frac{\partial u_z}{\partial z} \end{bmatrix} \tag{3.8}$$

is the strain tensor and

$$\vec{\omega} = \frac{1}{2}\nabla\times\mathbf{u}(\mathbf{x}) \tag{3.9}$$

is a pseudovector representing the angle of rigid rotation generated by, in this case, the passage of a seismic wave (Cochard *et al.*, 2006; Lee *et al.*, 2009). The rotational motions about the three orthogonal axis are given by

$$\begin{aligned} \omega_x &= \frac{1}{2}\left(\frac{\partial u_z}{\partial y}-\frac{\partial u_y}{\partial z}\right) \\ \omega_y &= \frac{1}{2}\left(\frac{\partial u_x}{\partial z}-\frac{\partial u_z}{\partial x}\right) \\ \omega_z &= \frac{1}{2}\left(\frac{\partial u_y}{\partial x}-\frac{\partial u_x}{\partial y}\right). \end{aligned} \tag{3.10}$$

The terms in equation 3.8 can be estimated by using the gradients estimated in equations 3.6. However, given a two-dimensional gradiometer deployed on the (approximate) horizontal plane of the Earth's surface, we can only estimate the horizontal components of strain parallel to the Earth's surface. In this case, equation 3.8 reduces to

$$\varepsilon = \begin{bmatrix} \frac{\partial u_x}{\partial x} & \frac{1}{2}\left(\frac{\partial u_x}{\partial y}+\frac{\partial u_y}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u_y}{\partial x}+\frac{\partial u_x}{\partial y}\right) & \frac{\partial u_y}{\partial y} \end{bmatrix} \tag{3.11}$$

and the rotational motions become

$$\begin{aligned} \omega_x &= \frac{\partial u_z}{\partial y} \\ \omega_y &= -\frac{\partial u_z}{\partial x} \\ \omega_z &= \frac{1}{2}\left(\frac{\partial u_y}{\partial x}-\frac{\partial u_x}{\partial y}\right) \end{aligned} \tag{3.12}$$

Note that the spatial gradients are functions of time, which we can use directly in equations 3.11 and 3.12 to compute time series of the horizontal strain and rotational motions. Likewise, there exist methods to compute the gradiometric parameters A and B for each time point (e.g. Langston 2007a,b; Langston and Liang, 2008; Poppeliers and Punosevac, 2013). However, experience shows that the parameters A and B are much more affected by data imperfections (e.g. coherent noise, superposed wavefields, etc.) so I employ a time-averaging method to estimate them. Specifically, equation 3.2 shows the relationship between any component of data and its derivatives, for a single time point. Assume that the data $\mathbf{u} = [u_1, u_2, ...u_M]$ is a vector of data samples (for arbitrary data component), where the window is of time length $t_{win} = M\Delta t$. Using, for example, the x-direction spatial derivatives of the data, we can form a system of linear equations over this time window as

$$\begin{bmatrix} \frac{\partial u_1(x_0,y_0)}{\partial x} \\ \frac{\partial u_2(x_0,y_0)}{\partial x} \\ \vdots \\ \frac{\partial u_M(x_0,y_0)}{\partial x} \end{bmatrix} = \begin{bmatrix} u_1(x_0,y_0) & \frac{du_1(x_0,y_0)}{dt} \\ u_2(x_0,y_0) & \frac{du_2(x_0,y_0)}{dt} \\ \vdots & \vdots \\ u_M(x_0,y_0) & \frac{du_M(x_0,y_0)}{dt} \end{bmatrix} \begin{bmatrix} \hat{A}_x \\ \hat{B}_x \end{bmatrix} \tag{3.13}$$

where $u_m(x_0,y_0)$ is the displacement wave field measured at the master station located at point $[x_0,y_0]$, $\frac{\partial u_m(x_0,y_0)}{\partial x}$ are its spatial gradients in the x direction, and $\frac{du_m(x_0,y_0)}{dt}$ are the time derivatives (e.g. the measured velocity seismogram at point $[x_0,y_0]$) for each time point $m = 1,2,...M$. An analogous system of equations can be constructed to solve for A_y and B_y (and A_z and B_z, in the case of a three-dimensional gradiometer) by using the y-direction spatial gradients in the left-hand side of the equations 3.13. The number of samples M can be chosen by the user, where so long as it's greater than the number of recording stations ($M > N$) the system is over-determined (and numerically stable). The result is that each gradiometer coefficient is a time-averaged estimate for the time window $t = M\Delta t$.

Table 3.1. The time series that can be extracted from the seismic wavefield collected on a 2D and 3D gradiometer. Note that for the spatial gradients, the derivative can be computed in the two horizontal directions and all three cardinal directions for the 2D and 3D gradiometers, respectively. Likewise, the quantities $A_j^{(i)}$ can be computed for the cardinal directions (indicated by the superscript) for each component j of data. Note that only unique wavefield attributes are listed here, so some entries may be omitted (e.g. $\varepsilon_{x,y} = \omega_y$).

2D gradiometer	3D gradiometer
u_x, u_y, u_z	u_x, u_y, u_z
$\frac{\partial u_i}{\partial x}, \frac{\partial u_i}{\partial y} \quad i = x, y, z$	$\frac{\partial u_i}{\partial x}, \frac{\partial u_i}{\partial x}, \frac{\partial u_i}{\partial z} \quad i = x, y, z$
$A_x^{(i)}, A_y^{(i)} \quad i = x, y, z$	$A_x^{(i)}, A_y^{(i)}, A_z^{(i)} \quad i = x, y, z$
$\omega_x, \omega_y, \omega_z$	$\omega_x, \omega_y, \omega_z$
ε_{xy}	$\varepsilon_{xy}, \varepsilon_{yz}$

Chapter 4

Gradiometric Attributes in a Machine Learning-based Discriminator

Machine Learning (ML) is an umbrella term that describes literally hundreds (if not more) ways to reduce data dimensionality. The primary goals of ML are to make data-based predictions of future data measurements and/or to classify data into sets. In all cases, however, ML revolves fundamentally around optimization: fitting data to some model (which may or may not be physics-based), according to some definition of goodness of fit. For example, the predictive capabilities of ML algorithms tend to focus on regression-type approaches, where the 'training data' are measured quantities. Geophysical data inversions are a specific case of ML algorithms, in which data are fit to a physics-based model of the Earth. The goal is to find a set of model parameters that will produce synthetic, or predicated, data that matches the observed data. Examples of this approach include seismic travel time tomography, source inversion, seismic imaging, and even the gradient estimation algorithms shown in equations 3.5-3.6.

In general, ML scenarios can be very similar to geophysical inversions, but the models are not physics-based. For example, a specific ML algorithm may attempt to find an arbitrary mathematical function that fits a specific data type (for example the daily stock market value). In this case, the ML algorithm will find parameters to a (for example) n^{th} order polynomial that minimizes the difference between the function and the data. The goal may be to use this polynomial strictly to make predictions rather than to understand the underlying mechanisms that control the system's behavior. There is no physics captured by the polynomial in this case, but the mathematical model **mimics** the system "good enough" to make predictions of future stock prices.

A different class of ML algorithms is concerned with classification. The goal of these algorithms is to separate data into groups that represent different scenarios. Note that classes may or may not be user defined. For the work presented here, the goal is to develop a method to classify seismic data into populations, where each population corresponds to a specific seismic source type. The classifier can be binary, (e.g. earthquake versus explosion source), trinary (e.g. earthquake vs. explosion vs. fracture source), or any number of classes that the user defines. The approach I explore here uses a neural network (NN) to build a classifier. I test both a binary and trinary classifier, where the data used to train the network are the wave attributes obtained from seismic gradiometry. This approach is different than most previous NN approaches to seismic data analysis, in that I do not use data that has been transformed into an image. Specifically, in almost every NN approach used to analyze or classify time series data, the data is first insulted prior to analysis by converting it into a spectrogram, which is a non-optimized time-frequency decomposition of the data. The spectrogram is then used as input in some sort of convolutional neural network. The tendency to convert time-series data into an image prior to application of a NN-type classifier is likely due to the fact that most NN research has been conducted with image processing applications in mind, and so it seems natural to convert time series into some sort of image. However, I argue that using spectrograms as input to a convolutional NN is non-ideal, or even advisable, as 1) spectrograms are mathematically objectionable (they divide the data into static time-frequency segments, so the time-frequency tradeoff is not optimally honored) and 2) it is a computationally expensive approach to the discrimination problem. Specifically, a one-dimensional seismogram, converted into a two-dimensional image takes vastly more computer memory to store. Subsequent processing, via a convolutional neural network, thus takes significant computer resources. I take a much more computationally efficient approach using a single-layer neural network as opposed to a convolutional neural network, where the data type are the gradiometric attributes, as opposed to spectrograms.

Neural Networks, Demystified

For classification tasks, the goal is to map a set of input data to a classification: we train the NN to accurately map the data $\mathbf{x}_j$ to their correct label $\mathbf{y}_j$. The input space has the dimension of the raw data $\mathbf{x}_j \in \mathbb{R}^n$. The output layer has the dimensions of the classification space. Figure 4.1 shows an example of a two layer NN, where we denote the various layers between input and output as $\mathbf{x}^{(k)}$ where k is the layer number. For a linear mapping between layers, the following relations hold

$$\mathbf{x}^{(1)} = \mathbf{A}_1\mathbf{x} \tag{4.1}$$

$$\mathbf{x}^{(2)} = \mathbf{A}_2\mathbf{x}^{(1)} \tag{4.2}$$

$$\mathbf{y} = \mathbf{A}_3\mathbf{x}^{(2)}, \tag{4.3}$$

which forms a compositional structure so that the mapping between input and output can be represented as

$$\mathbf{y} = \mathbf{A}_3\mathbf{A}_2\mathbf{A}_1\mathbf{x}. \tag{4.4}$$

This basic architecture can scale to M layers so that the general representation between input data and the output layer for a linear NN is given by

$$\mathbf{y} = \mathbf{A}_M\mathbf{A}_{M-1}\ldots\mathbf{A}_2\mathbf{A}_1\mathbf{x}. \tag{4.5}$$

Equations 4.1 - 4.5 show the general architecture for an M layer NN with linear mapping. Nonlinear mappings are also possible. In this case, equations 4.1 - 4.3 become

$$\mathbf{x}^{(1)} = f_1(\mathbf{A}_1, \mathbf{x}) \tag{4.6}$$

$$\mathbf{x}^{(2)} = f_2(\mathbf{A}_2, \mathbf{x}^{(1)}) \tag{4.7}$$

$$\mathbf{y} = f_3(\mathbf{A}_3, \mathbf{x}^{(2)}), \tag{4.8}$$

where the nonlinear functions f_i do not need to be the same between layers. Following the example in equation 4.5 we can derive an expression for an M layered, nonlinear NN as

$$\mathbf{y} = f_M(\mathbf{A}_M, \ldots, f_2(\mathbf{A}_2, f_1(\mathbf{A}_1, \mathbf{x}))\ldots). \tag{4.9}$$

The simplest discriminator will be a linear binary NN. In other words, the output of the NN will be only one of two possibilities where there is a linear mapping between the data and the binary output: for example earthquake source or explosion source. We can represent that as

$$\mathbf{y} = \{earthquake, explosion\} = \{+1, -1\} \tag{4.10}$$

which labels each data vector with the output $\mathbf{y} \in \{\pm 1\}$. In this case, the output layer is a single node (Figure 4.2). The mapping is linear between the input data $\mathbf{x}_j \in \mathbb{R}$ and the output layer. This gives a linear system of equations $\mathbf{AX} = \mathbf{Y}$:

$$\mathbf{AX} = \mathbf{Y} \rightarrow [a_1 a_2 \ldots a_n] \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_p \\ | & | & & | \end{bmatrix} = [+1 +1 \ldots -1 -1] \tag{4.11}$$

where each column of the matrix $\mathbf{X}$ is the data from the earthquake or explosion source and the columns of $\mathbf{Y}$ are its corresponding (and known) labels. For completeness, note that each column in matrix $\mathbf{X}$ is a time series of length N, of which there are P. Thus $\mathbf{X}$ is a $N \times P$ array, array $\mathbf{A}$ is of size $1 \times N$, and array $\mathbf{Y}$ is of size $1 \times P$

There is no need to restrict this classifier to bimodality. For instance, we can construct a trinary classifier by defining

$$\mathbf{y} = \{earthquake, fracture, explosion\} = \{+1, 0, -1\}. \tag{4.12}$$

In any case, since the output layer is a single node, both $\mathbf{A}$ and $\mathbf{Y}$ reduce to vectors. In this first case, the goal is to determine the matrix (vector) $\mathbf{A}$ with components a_j. Equations 4.11 can be solved using the pseudoinverse

$$\mathbf{A} = \mathbf{Y}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T, \tag{4.13}$$

however, some researchers have argued for methods that promote modality, such as the LASSO method for example (Brunton and Kutz, 2019, pages 198-202).

The method described in equations 4.1 - 4.11 is known as a supervised neural network, and thus needs to be “trained” using real data. A training data set is used to build the mapping matrix $\mathbf{A}$, which is then used to find the labels of the data being tested by using a single-layer version of equation 4.1:

$$\mathbf{y} = \mathbf{Ax}. \tag{4.14}$$

For the work here, we use the raw seismic data, its spatial gradients, and the attributes estimated from them as the data. For example, the vector $\mathbf{x}_1$ contains a vertical concatenation of all of the time series (e.g. $\partial u_x(t)/\partial x$, followed by $\partial u_y(t)/\partial x$, and so on) available for seismic event number 1. The second column $\mathbf{x}_2$ would contain a concatenated vector of all the data and gradiometric time series from the second seismic event, and so on. The only requirements are that each column $\mathbf{x}_i$ of the array $\mathbf{X}$ contain the same number of samples and that each column contains the data and gradiometric attributes in the same order. Note that for data collected on a two-dimensional gradiometer, there are fifteen independent time series that can be used for each training data set $\mathbf{x}_i$. There are twenty-six independent time series that can be extracted from a three-dimensional gradiometer (Table 3.1). Note that the ordering of the vectors $\mathbf{x}_i$ makes no difference to the results, so long as the ordering of the quantities in both the training data and test data are the same.

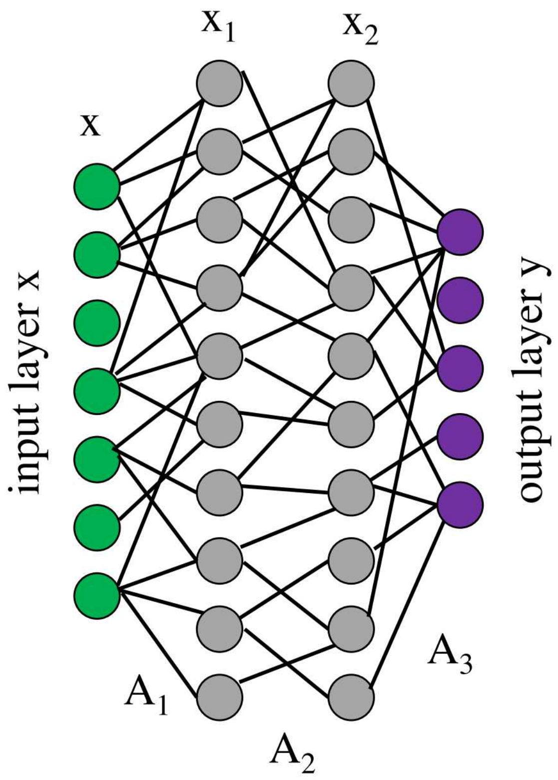

Figure 4.1. An example of a two layer neural network architecture which maps the input layer $\mathbf{x}$ to an output layer $\mathbf{y}$. The middle (hidden) layers are denoted $\mathbf{x}^{(i)}$ where j determines their sequential ordering. The matrices $\mathbf{A}_j$ contain the coefficients that map each variable from one layer to the next. The number of layers as well as the functional map between layers is chosen by the user.

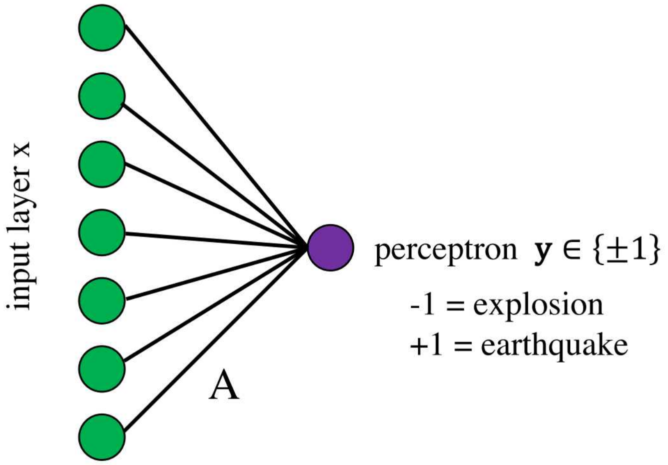

Figure 4.2. A single layer NN for a binary classifier between explosion and earthquake sources. The input layer **x** is a matrix containing the data, which consists of the seismic data and its gradiometric attributes. The output layer for this case is a perceptron with $\mathbf{y} \in \{\pm 1\}$. The linear mapping **A** is built using a training set of data (and its gradiometric attributes) where the labels are known, and solving the equations $\mathbf{A} = \mathbf{Y}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. After **A** is built, it can be applied to data to obtain a binary discriminant.

Chapter 5

Tests with synthetic data

The goal of this work is to conduct a proof-of-concept experiment to test whether seismic gradients and the attributes estimated from them, can be used in a simple neural network-based classification algorithm. I design the tests to discriminate between three common source types: earthquakes, explosions, and fractures. Although the original motivation for this work was to discriminate between earthquake and explosion sources, I realized that a potentially valuable application of this method could be for oil-field monitoring during hydraulic fracturing operations. Thus, I also test discriminating between earthquake and mode-1 opening fracture sources (i.e. a fracture that opens as a result of forces perpendicular to the fracture plane: there are no shear forces present). Finally, I conduct a test where I discriminate between all three source types simultaneously.

Geologic Model

The model consists of a layered sedimentary sequence overlying an igneous basement (figure 5.1) and is 2500m in length and width, and 1500m in depth. In addition to the deterministic geologic structure I superpose a stochastic impedance field to mimic geologic heterogeneity. The stochastic field is designed to induce seismic scattering, and I included it in order to make the simulations more realistic. Specifically, I model a stochastic wave speed model as

$$(x,y,z) = c_0(x,y,z)\,\delta c(x,y,z) \tag{5.1}$$

where $c_0(x,y,z)$ is the smooth (large scale) background wave speed, and $\delta c(x,y,z)$ is the stochastic background field. The stochastic field is a zero-mean perturbation of physical properties with predefined attributes that describe correlation length and variance. I use a von Karman autocorrelation function to describe the stochastic field:

$$P_{\delta v}(\mathbf{k}) = \frac{\Gamma(2v)\mathbf{a}}{[2^{(2v-2)}][(1+k^2\mathbf{a}^2)^{(v+1/2)}]} \tag{5.2}$$

where Γ is a gamma function, v is the Hurst exponent, $\mathbf{a} = [a_x, a_y, a_z]$ is the characteristic length of the wave speed perturbation in the x, y, and z directions, and $\mathbf{k} = (k_x^2a_x^2 + k_y^2a_y^2 + k_z^2a_z^2)^{1/2}$ is the wavenumber (Goff *et al.*, 1994). The characteristic length controls the roll-over of the stochastic field's power spectrum: for wavenumbers less than the characteristic length, the field is spectrally white. For wavenumbers greater than the characteristic length, the stochastic field has a power spectrum with a slope controlled by the Hurst exponent. Essentially, the characteristic length controls the “size” of the heterogeneities and the Hurst exponent controls the roughness. For our work here, I set the characteristic length to $[a_x, a_y, a_z] = [600, 600, 200]$m and the Hurst exponent to $\nu = -2.7$. To generate the stochastic field, I compute the von Karman power spectrum according to equation 5.2, then add a randomized phase in the form of Gaussian-distributed random numbers. Applying the inverse Fourier transform results in unscaled stochastic field, that I then normalize to have a variance that corresponds to the desired degree of wave speed perturbation (in this case, $\pm 5.0\%$ of the local P wave speed). The stochastic field is then added to the deterministic wave speed model. This is done similarly to the S wave speed and density.

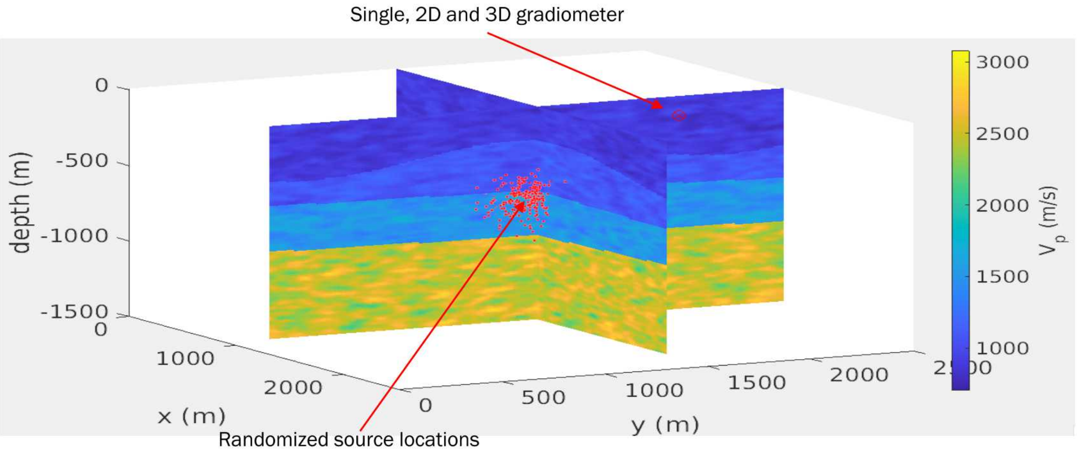

Figure 5.1. The geologic model used to create the synthetic data. The red dots indicate the location of the individual sources used to create the synthetic data. There are 330 total source locations for each source type, resulting in 990 total wave simulations.

Synthetic Seismograms

To generate the synthetic seismograms, I use a 3D finite difference scheme that’s formulated to solve for the three-component velocity field for all points in the Earth model. The system is a set of nine coupled, first order, linear, non-homogeneous partial differential equations:

$$
\begin{aligned}
\frac{\partial v_i}{\partial t} - \frac{1}{\rho}\frac{\partial \sigma_{ij}}{\partial x_j} &= \frac{f_i}{\rho} && \text{(3 equations)} \\
\frac{\partial \sigma_{ij}}{\partial t} - \lambda \frac{\partial v_k}{x_k}\delta_{ij} - \mu\left[\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right] &= \frac{\partial M_{ij}}{\partial t} && \text{(6 equations)}
\end{aligned}
\tag{5.3}
$$

where $v_i(\mathbf{x},t)$ is the velocity vector, $\sigma_{ij}(\mathbf{x},t)$ is the stress tensor, $\rho(\mathbf{x})$ is density, $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$ are the elastic moduli, $f_i(\mathbf{x},t)$ is a force vector, and $M_{ij}(\mathbf{x},t)$ is the moment tensor. To solve this system, I use a staggered-grid finite difference scheme and solve for the velocity and stress at each time step and grid node. For this work, I used a grid node spacing of 10 m and a discrete time step of 0.00074 s. Based on the wave speed model and these finite difference parameters, I can simulate wave fields that are numerically stable and free from significant numerical dispersion up to a frequency of approximately 10 Hz.

I constructed 330 sets of three-component seismograms for each seismic source type, for a total of 990 synthetic data sets. In each case, the seismic source location was chosen from a random distribution of subsurface locations, where the mean location was $[x,y,z] = [1250,1250,600]$m. The "cloud" of potential source locations had a standard deviation of 200m in the x, y, and z directions, respectively. Note that for each of the 330 randomized source locations, I generate three sets of seismograms; one set for each source type. To generate the explosion-source seismograms, I used a moment tensor of the form

$$M_{ij}^{expl} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5.4}$$

The earthquake source was defined by a double couple source, where the orientation of the fault plane is randomized. The mean tensor was defined as

$$M_{ij}^{EQ} = \begin{bmatrix} 0 & -0.27 & 0.67 \\ -0.27 & 0.48 & -0.50 \\ 0.67 & -0.50 & -0.48 \end{bmatrix}, \tag{5.5}$$

which defines a strike slip fault that's been rotated about the x ([1,0,0]) axis by +10 degrees and about the y ([0,1,0]) axis by +20 degrees. For each set of synthetic earthquake seismograms, of which there were 330, I chose a source location from the (randomized) source cloud similarly to the explosion sources. I then randomize the orientation of the earthquake fault by randomly rotating the tensor about each unit axis. The randomized rotation angles are chosen from a set of zero-mean, Gaussian-distributed angles where the standard deviation is 10 degrees.

Finally, seismograms from the mode-1 opening fracture sources were generated similarly to those of the earthquake sources, except that the mean fracture-type tensor was defined as

$$M_{ij}^{frac} = \begin{bmatrix} 1.48 & 0.46 & 0.19 \\ 0.46 & 1.45 & 0.18 \\ 0.19 & 0.18 & 1.07 \end{bmatrix}, \tag{5.6}$$

and the randomized rotation angles had a standard deviation of 20 degrees.

For all three synthetic data types, I computed Green's functions by solving equations 5.3 where the initial condition was a unit-amplitude delta function at t=0. To form the seismograms for the tests here, I then convolved each set of Green's functions with a Gaussian wavelet, where the center frequency was 5.0 Hz. I note that a Gaussian wavelet source time function is not strictly correct in terms of simulating various seismic sources, but I chose this to eliminate that possibility of having the discrimination algorithm discriminating on source wavelet type. In future tests, I will use more realistic source time functions, but I don't anticipate that this will significantly effect the results that I report here. An example of data for a single station is shown in Figures 5.1 and 5.2.

Finally, I add various degrees of noise to the data to determine the efficacy of my algorithm on non-perfect data. However, rather than simply adding uncorrelated "white" (or colored) noise to each seismogram, I used an approach that honors the nature of seismic noise. Specifically, I create a second wavefield through the model (using the finite difference approach described above) where the source is designed to mimic a surface mounted pump located on the Earth's surface. I modeled this type of source using a vertically directed force applied to the surface of the model. As before, I use the finite difference simulation to generate the Green's functions for this source, and then convolved the Green's functions with a band-limited (0.1 - 10 Hz) sequence of random numbers, which yielded a band-limited sinusoid-like signal. The wavefield generated from this type of source approximates a typical anthropogenic source such as an irrigation pump or a large air cooling unit. I then scaled the resulting wavefield such that its root mean square (RMS) amplitude was a defined percentage of the data's RMS amplitude and added it directly to the synthetic seismograms (Figure 5.2).

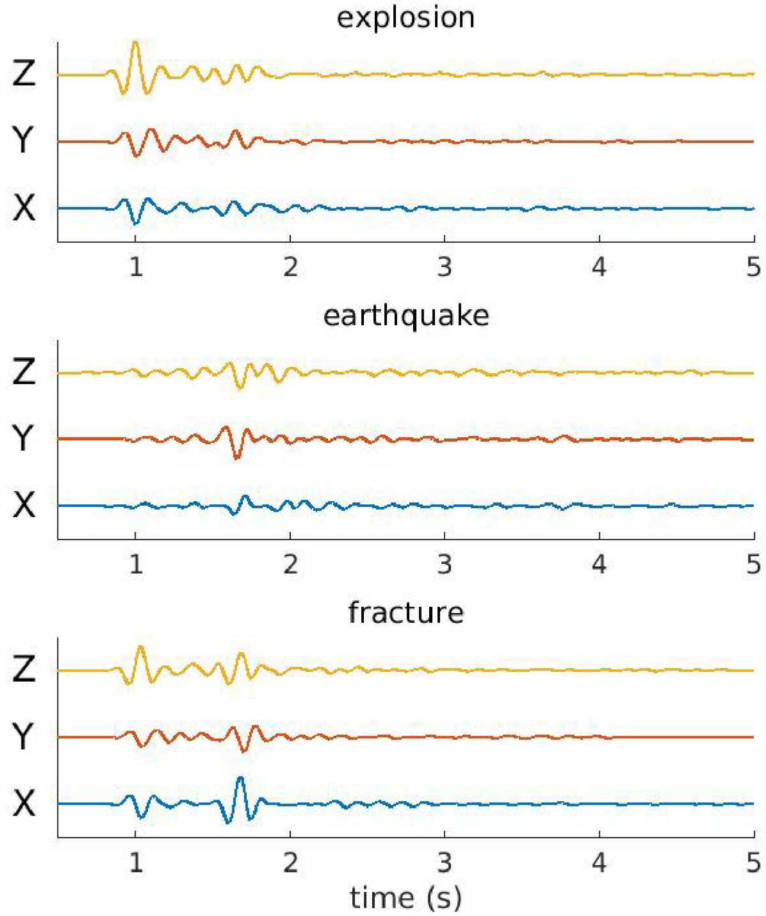

Figure 5.2. The synthetic data for a single station for the three different source types, where the data contain no noise.

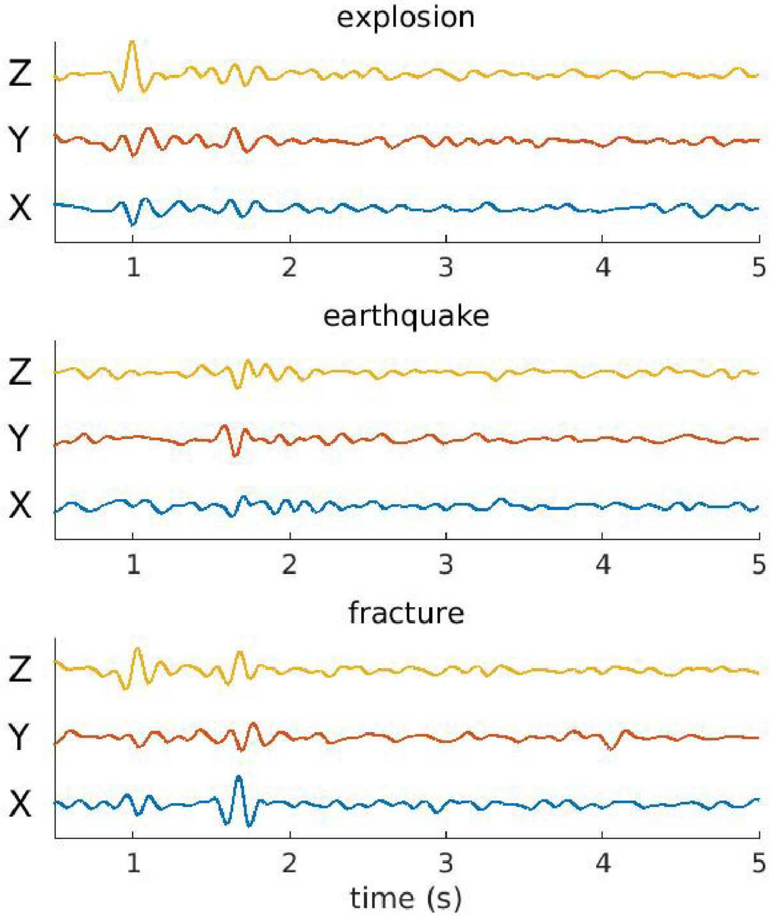

Figure 5.3. The synthetic data for a single station for the three different source types, where the data contains noise whose RMS amplitude is approximately half that of the data RMS amplitude.

Chapter 6

Results

Here I review the results of the classifiers based on a single-layer neural network. First, I test two binary classifiers. For the first type, I classify between earthquake sources and explosion sources. I also test a binary classifier designed to classify between earthquake and fracture sources. The fracture sources simulate a mode-1 opening fracture which may occur, for example, during a hydraulic fracturing operation. I also test a trinary classifier, where I test the algorithm's ability to discriminate between an earthquake, explosion, and fracture sources.

For all tests, I use a bootstrap error-estimation approach in order to estimate the reliability of the method. For each bootstrap test, select different data for the training and analysis groups. Recall that I generated 330 sets of synthetic seismograms for each source type. For a single classification task, I randomly select 90% of the data for each source type as a training set, and attempt to classify the remaining data. Then I repeat this test, but for a different selection of training/analysis data. I repeat this procedure 500 times, where each bootstrap test would ideally result in the same results. However, given that the data for training and analysis are (mostly) different for each test, there is a degree of variability of the results. However, because I know the actual seismic source in all cases, I can use the results for each individual bootstrap test to summarize the success rate for a given classifier (as well as the misclassification rate).

I perform the bootstrap tests assuming a two dimensional gradiometer as well as a three dimensional gradiometer. Furthermore, I add varying degrees of noise with the goal of determining the efficacy of the method with non-ideal data. The results are presented as graphs, where I report both the success rate and misclassification rate.

Binary discriminator

In this section, there are only two possible classes of data: an explosion source and a second source type. The second source type can be either an earthquake source or an explosion source.

Earthquakes versus explosions

These two source types represent the end-member of possible sources, and result in data with the greatest differences in both P/S energy ratio and radiation pattern. For both the 2D gradiometer (figure 6.1) and 3D gradiometer (figure 6.2) the classifier has more success in identifying the explosion source. The 3D gradiometer has a slightly higher success rate for high noise conditions.

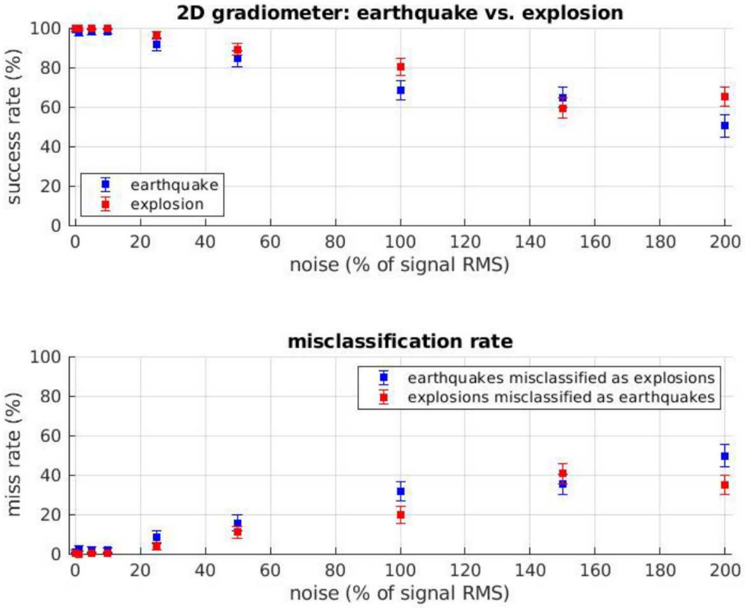

Figure 6.1. Top: the success rate for binary discrimination using a 2D gradiometer: earthquake versus explosion. The vertical axis is the success rate and the horizontal axis is the percent noise. Bottom: the misclassification rate. Note that for this test, the algorithm generally has a better success rate for identifying an explosion source.

Earthquakes versus fracture

For this test, the classifier has a much better success rate for identifying earthquakes than for the previous (earthquake-vs-explosion) test. Additionally, the classifier more often correctly classifies the earthquake source than it does the fracture source. However, as with the explosion-vs-earthquake case, the classifier generally works better for data collected by the 3D gradiometer (figures 6.3 - 6.4).

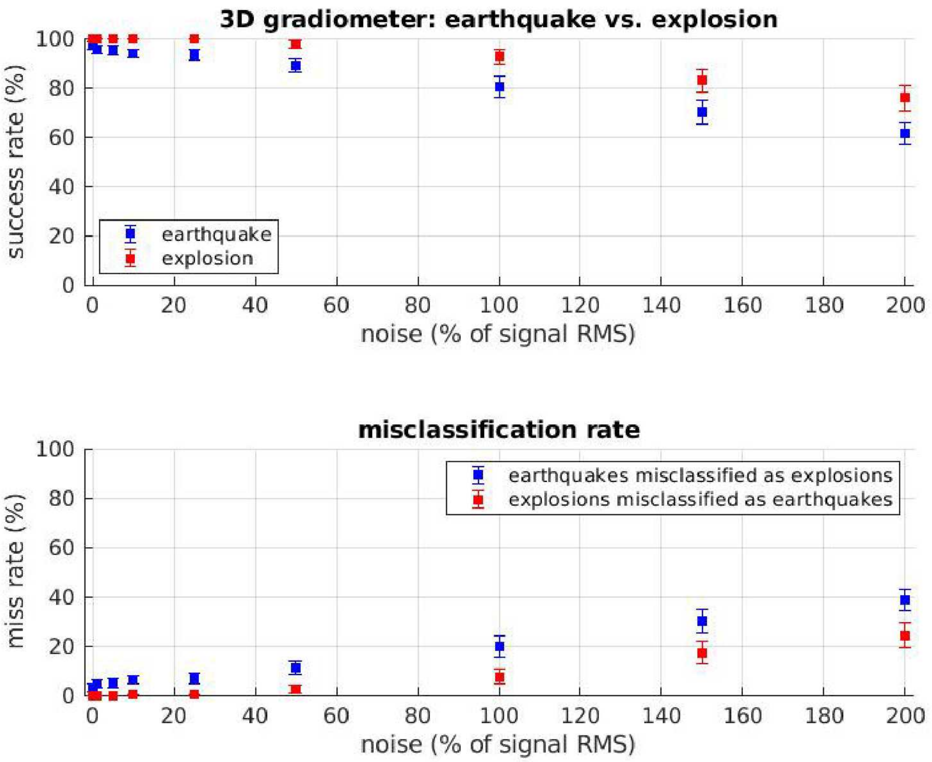

Figure 6.2. Same as figure 6.1, but using data from a 3D gradiometer. Note that in this case, the algorithm is still more successful at identifying explosion sources, but the difference between successful earthquake and explosion classifications is less than that of the 2D gradiometric test. Furthermore, the success rate for both source types is generally higher when using data collected by the 3D gradiometer.

Trinary discriminator

In this test, I perform the same bootstrap method but use three classes: earthquake source, explosion source, and fracture source. This is a more difficult classification task, and generally the success rate is lower than for the binary classifiers.

In general, the classification is more accurate when using data from the 3D gradiometer (figures 6.5 - 6.6). However, for both gradiometer types, the source type that is most accurately classified is the fracture source, with the earthquake source being the source type that is correctly identified the least accurately. This is the case for almost all noise conditions.

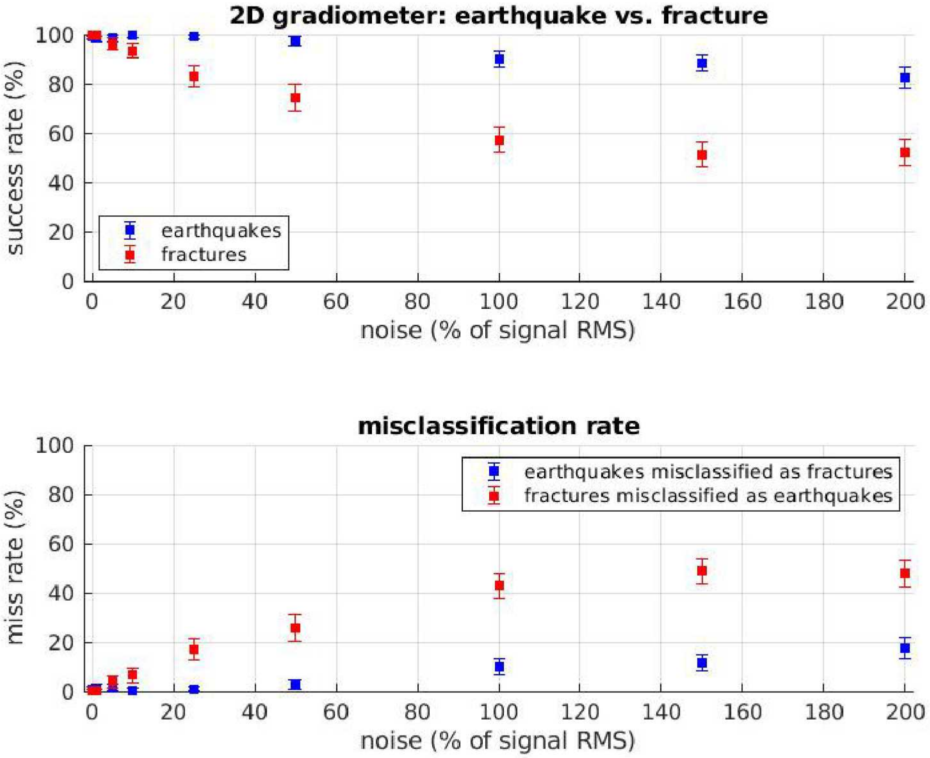

Figure 6.3. Top: the success rate for binary discrimination using a 2D gradiometer: fracture versus explosion. Note that for this test, the algorithm has a better success rate for identifying an earthquake source. Bottom, the misclassification rate.

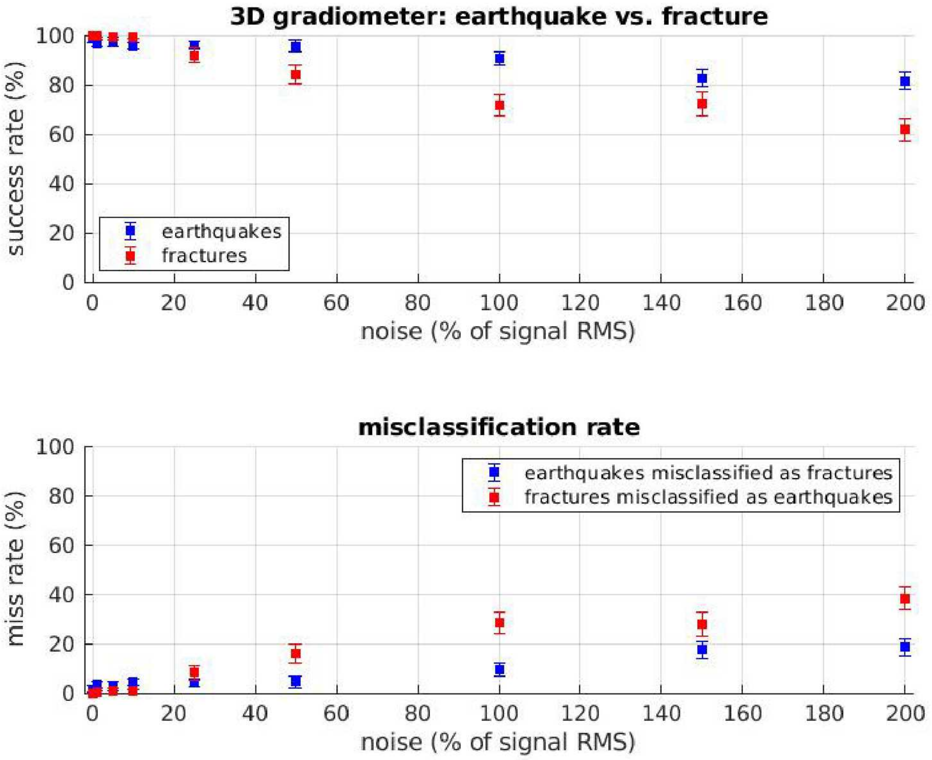

Figure 6.4. Same as figure 6.3, but using data collected on a 3D gradiometer.

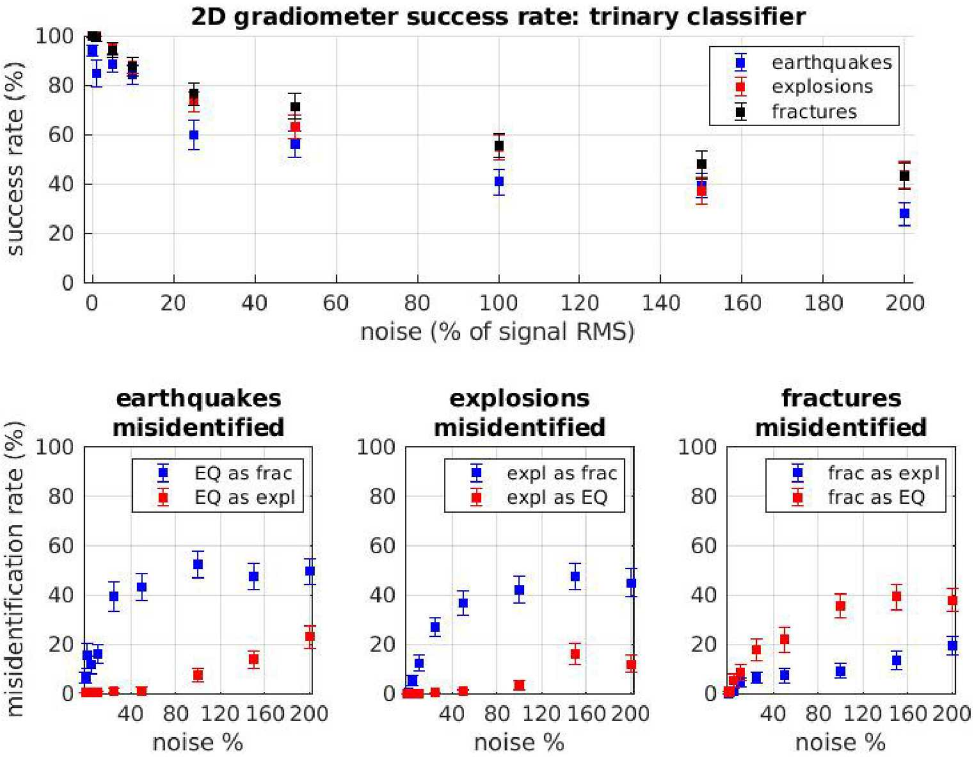

Figure 6.5. Top: the success rate for a trinary discriminator using 2D gradiometric data. Note that, in general, the fracture source is discriminated more accurately than either the earthquake or explosion source. Bottom: the misclassification rate for the three source types. For most noise conditions, the least accurate classification is the case where earthquakes are misclassified as fractures.

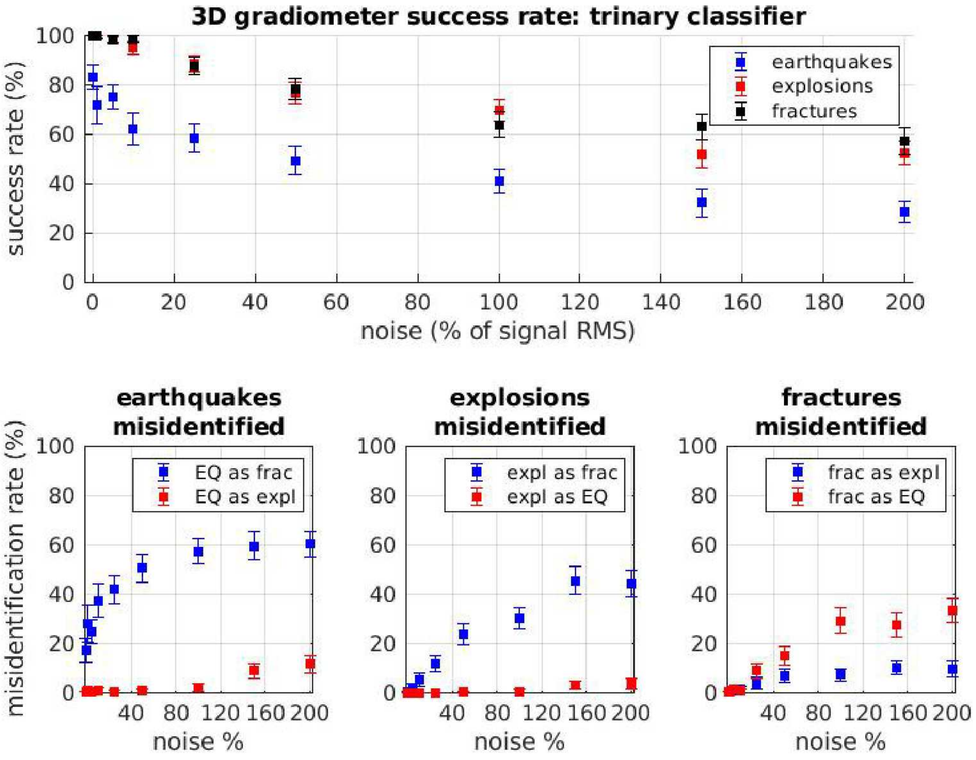

Figure 6.6. Same as figure 6.5, but for a 3D gradiometer. Note that in general using data from a 3D gradiometer results in better performance of the classifier.

Chapter 7

Summary

This report summarizes a new method for seismic source discrimination. The method uses a single-layer computational neural network to classify seismic data into pre-defined classes. However, the method I present is unique in that I use seismic spatial gradients, and the wave attributes computed from them, as input. The algorithm I present and test is different from previous machine learning-based seismic discriminants in two important ways. First, I do not need a regional-scaled seismic network to surround the source. Secondly, I use the data directly in the algorithm, circumventing the need for computationally expensive time series-to-spectrogram conversions.

The primary goal of this work was to provide proof-of-concept using simulated data: specifically, I wished to determine whether we can use a compact gradiometer for the seismic discrimination task rather than a distributed seismic network. I viewed this as the most important aspect of this project, as a gradiometer is much easier and cheaper form of instrumentation than a distributed network. For example, a typical seismic network will have a few dozen instruments, with a nominal aperture of several hundreds of wavelengths. This translates into a seismic network where the instruments are tens of kilometers apart. Conversely, a gradiometer designed to analyze a similar wavefield will be composed of roughly the same number of seismic instruments, but have a nominal aperture of a few tens to hundreds of meters. Permitting, deploying, and maintaining a gradiometric array is thus a much easier (and cheaper!) proposition. Furthermore, given the small size of a gradiometer, it becomes tractible to construct a three-dimensional gradiometric array by placing seismometers in boreholes. This provides an especially attractive possibility, as bore-hole deployments can have more optimal noise conditions when compared to surface-mounted instruments.

For the tests here, I showed that using gradiometric wave attributes in a neural network can discriminate between three possible source types. However, I note that the method required training data. In other words, for the algorithm to work, there had to be a volume of data where the source type was known. This is a potential weakness of this method, and any supervised machine learning technique. However, it may be possible to use a known catalog of seismicity in a region to train the algorithm, if the region contains an active mining operation where explosives are used as part of the operation. Regardless, further research as part of this project will explore so-called unsupervised machine learning techniques to classify source type.

Chapter 8

References

Aki, K., Richards, P. G., 1980, Quantitative Seismology: Theory and Methods, Volumes I and II, W. H. Freeman and Co., San Francisco, CA.

Aki, K., Richards, P. G., 2002, Quantitative Seismology, Second Edition, University Science Books, Sausalito, CA.

Brunton, S.L, Kutz, J.N., 2019, Data-driven science and engineering: machine learning, dynamical systems, and control, Cambridge University Press, Cambridge, United Kingdom.

Chouet, B., Dawson P., Ohminato, T., Macello, M., Saccorotti, G., Guidicepietro, F., De Luca, G., Minala, G., Scarpa R., 2003. Source mechanisms of explosions at Stromboli Volcano, Italy, determined from moment-tensor inversions of very-long period data. *J. Geophys. Res*, **108**(B1), 2019.

Duputel, Z., Agram, P.S., Simons, M., Minson, S.E., Beck, J.L., 2014. Accounting for prediction uncertainty when inferring subsurface fault slip, *Geophys. J. Int.*, **197**(1) 464-482.

Dziewonski, A.M., Chou, T.-A., Woodhouse, T.A., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. Geophys. Res.*, **86**, 2825-2852.

Dziewonski, Woodhouse, T.A., 1983. An experiment in systematic study of global seismicity: Centroid-moment tensor solutions for 201 moderate and large earthquakes of 1981, *J. Geophys. Res.*, **88**, 3247-3271.

Dreger, D., Woods, B., 2002. Regional distance seismic moment tensors of nuclear explosions, *Tectonophysics*, **356**(1-3), 139-156.

Ekstrom, G., Nettles, M., Driewonski, A.M., 2012. The global CMT project 2004-2010: Centroid-moment tensors for 13017 earthquakes, *J. Phys. Earth Plant. Inter.*, **200-201**, 1-9.

Goff, J.A., Holliger, K., Levander, A., 1994. Model fields: a new method for characterization of random seismic velocity heterogeneity, *Geophys. Res. Lett.* **21**, 493-496.

Gu. C., Marzouk, Y.M., Toksoz, M.N., 2018, Waveform-based Bayesian full moment tensor inversion and uncertainty determination for the induced seismicity in an oil/gas field, *Geophys. J.*

Int., **212**(3), 1963-1985.

Hudson, J.A., Pearce, R.G., Rogers, R.M., 1989, Source type plot for inversion of the moment tensor, *J. Geophys. Res.*, **94**(B1), 765-774.

Langston, 2009. Wave Gradiometry in Two Dimensions, *Bull. Seis. Soc. Am.* **97**(2), 401-416.

Poppeliers, C.J., Evans, E.V, 2015. The Effects of Measurement Uncertainties in Seismic-Wave Gradiometry, *Bull. Seis. Soc. Am***105**(6), 3143-3155.

Poppeliers, C.J., Punosevac, P., 2013. Three Dimensional Seismic-Wave Gradiometry for Polarized Seismic Waves, *Bull. Seis. Soc. Am.* **103**(4), 2161-2172.

Poppeliers, C.J., Punosevac, P., Bell, T., 2013. Three Dimensional Seismic-Wave Gradiometry for Scalar Waves, *Bull. Seis. Soc. Am.* **103**(4), 2151-2160.

Stump, B.W., Johnson, L.R., 1977. The determination of source properties by the linear inversion of seismograms, *Bull. Seis. Soc. Am.* **67**, 1489-1502.

Tape, W., Tape, C., 2012. A geometric comparison of source-type plots, *Geophys. J. Int.*, **190**, 499-510.

DISTRIBUTION:

1	MS 0735	Kyle Jones, 8861
1	MS 0735	Christian Poppeliers, 8861
1	MS 0359	D. Chavez, LDRD Office, 1911
1	MS 0899	Technical Library, 9536 (electronic copy)

Sandia National Laboratories