EMTF XML: New data interchange format and conversion tools for electromagnetic transfer functions

Anna Kelbert[1]

ABSTRACT

Initial processing of magnetotelluric (MT) data collected at a site results in a small data file that defines the MT transfer functions (MT TFs) or variants at a discrete set of frequencies. Although not data intensive, the TF represents almost all of the information about earth's conductivity found in the raw data and provides a key input to geophysical inversion and to geomagnetic hazard-mitigation procedures. However, to be useful for scientific interpretation, MT TFs must be accompanied by carefully recorded metadata. The value of these data is amplified if they are collected into a common database in a discoverable fashion; to achieve this, the metadata must also be searchable.

We have developed a new format for MT and related electromagnetic transfer function extensible markup language (EMTF XML), which is a novel, self-describing, searchable, and extensible way to store the data. We also evaluate our open-source, highly capable conversion tools that implement format conversion between historical MT TF formats (electrical data interchange [EDI], EDI SPECTRA, Z-files, and potentially others) and EMTF XML and we examine the associated issues of data rotation. We have paired our work with the open-source format conversion tools, EMTF FCU v4.0, with a user guide and a set of use case examples. We hope that this effort will assist the MT community in creating a more open data-sharing culture and practices.

INTRODUCTION

Magnetotellurics (MT) is a geophysical technique, complementary to seismology, which uses electromagnetic (EM) wave propagation to infer the electrical conductivity of the earth's crust and mantle. For this, MT uses very large-scale natural sources from the solar wind and lightning. Modern-day MT uses state-of-the-art instrumentation, data processing, and analysis tools to provide valuable information about the deep earth structure (e.g., Unsworth et al., 1997; Evans et al., 1999; Wei et al., 2001; Booker et al., 2004; Unsworth et al., 2005; Soyer and Unsworth, 2006; Wannamaker et al., 2009; Kelbert et al., 2012; Meqbel et al., 2014; Murphy and Egbert, 2017). However, there is a long-standing need to modernize deeply historical MT data formats to a common standard that is fully documented, platform-independent, extensible, and accessible to the broader community of geoscientists.

The traditional MT plane-wave source assumption allows for the ionospheric source to be decomposed into two polarizations: south to north and west to east. Transfer functions (TFs) cancel out the magnitude of the source; thus, they can be used under the plane-wave assumption without explicit estimation of the amplitudes of the natural sources. For this reason, the final product of MT data processing, MT impedances and tippers, takes the form of TFs: frequency-domain tensors that, by definition, relate output EM data channels to input EM data channels. Traditionally, the MT method uses two horizontal magnetic field components as the input. Multiple electric and magnetic field components measured locally and at a distance can serve as outputs of a TF. For example, the impedance relates two local horizontal electric field channels to two local horizontal magnetic field channels. Other TFs with different output channels are also commonly defined. Here, we focus on recording a general electro-

magnetic transfer function (EMTF) in a robust, well-documented, matrix-based formulation.

As the science of MT matured in the 1980s and 1990s, EMTFs have mostly been created and stored in the historical Society of Exploration Geophysicists (SEG) Data Interchange Standard 1987, also known as the Electrical Data Interchange (EDI) (Wight, 1988). This ASCII format was quite comprehensive at the time of its development, but it no longer supports many of the varieties and components of MT data and metadata that are encountered in modern geophysics. More importantly, the format and metadata contents of the EDI files are not particularly supportive of data sharing, which indeed has not been a common practice thus far within MT. In fact, prior to this effort, MT transfer functions (MT TFs) from historical observational campaigns have not been widely or easily available. Thus, a researcher planning an experiment was typically "in the dark," with no access to previous data from the same or nearby areas, except for information gathered and data obtained through direct personal communication.

More complete and free access to digital data available from peers and historical surveys is of great value in planning an observational campaign and then to leverage the spatial coverage of the new data at the inversion and interpretation stages. Open and direct access to historical MT data is now particularly valuable in view of the recently matured and widely available 3D MT inversion capabilities (Siripunvaraporn et al., 2005; Egbert and Kelbert, 2012; Kelbert et al., 2014) that allow for creative combinations of multiple data sets as part of an integrated interpretation using large-scale, high-resolution inversions for the ground structure. In some cases (e.g., data from the oil and gas or geothermal industries, as well as from some international collaborations), it is not immediately possible to make final data products freely available. In such situations, it would still be useful to store a searchable metadata header, so that a researcher planning a new experiment could easily find out which data exist and coordinate access with the data owner. Here, we discuss the technical and societal stumbling blocks for data sharing in MT and we discuss our solution, which includes a novel, flexible, and searchable data format; format conversion tools; an approach to credit attribution; and an openly accessible EMTF database. With a combination of these techniques, we hope to have provided a pathway for a new culture of data sharing in the MT community.

EFFORTS TOWARD A NEW EMTF STANDARD

Traditional MT data formats are varied and rather fluid. For the most part, no freely available readers are available: Instead, a great number of "homebrewed" reading and writing tools proliferate (a notable exception is a powerful, general-purpose open-source Python package MTpy by Krieger and Peacock [2014], which provides a partial solution). Part of the challenge lies in the great heterogeneity of the historical file formats. Specifically, the EDI files (Wight, 1988) allow enough file variability that, until now, no single tool existed that claimed to be able to read them all. The other part of the problem is metadata or lack thereof. Metadata are critical for long-term data archiving and sharing. Historical EDI files are notorious for not having enough metadata; for example, one common practice omitted the latitude and longitude of the measurement location, opting instead for the distance from a common reference point provided in either meters or feet (the projection used to compute this distance was not included). Very rarely, critical information such as the sign convention used for data processing, remote reference-site information, and

provenance are found in these historical files. This inherent heterogeneity of the data was one of the roadblocks that kept the community away from healthy data-sharing practices.

Another source of great ambiguity in historical data is orientation. All TFs relate pairs of channels. The channels are components of electric and magnetic fields, which are both vectors, and have an orientation. For most historical EMTFs, an ambiguity is present with respect to how the TFs are rotated. They can either match the channel orientations, as written in the file, or they can be rotated to an orthogonal set of coordinates, at an angle to geographic north (which could be zero, but it does not have to be). In general, the site layout does not need to be orthogonal, although it often is. However, the TFs can only be interpreted if they are in an orthogonal coordinate system, and the angle to geographic north is known. Among the historical EDI files, it is rare to find a record that is unambiguous with respect to the data orientation (see the examples that accompany this paper).

Beyond the metadata, estimation of data errors has been the most challenging component of MT data processing. Full covariance matrices may be defined to relate different TF components together, statistically (e.g., Eisel and Egbert, 2001). This information is of more than theoretical interest because it is necessary to oversee the rigorously estimated error bars through a data rotation. EMTFs are matrices, and allowing for a general statistical covariance would require storage of a matrix to accompany each of the data components. Traditional EDI MT files are not well-suited for matrix storage, and this information is invariably omitted. On the other hand, a data rotation is often necessary for reinterpretation of these historical data. This illustrates how a mundane issue of data format has long-term scientific implication. A variant of the EDI files, EDI SPECTRA files, stores EMTFs in a format that is not ready for interpretation; a processing code is necessarily applied before they can be used. However, in contrast to the more common EDI MT files, EDI SPECTRA contain sufficient information for the full EMTFs and the covariance matrices to be extracted, in any right-handed coordinate system. Nevertheless, this capability has not existed, to our knowledge, and it is first presented in this paper.

The other commonly used EMTF file formats, the J-format (Jones et al., 1983), EMTF Z-files (Eisel and Egbert, 2001), and BIRRP format (Chave et al., 1987), which are outputs of the corresponding processing codes, did not all suffer from the same lack of information on the error estimates. They have, however, all suffered from a near-absolute lack of metadata. This is only understandable: The files were intended for immediate analysis within a single research group and in the framework of a single project, never for sharing or reanalysis decades later. They served their purpose perfectly well. Only now, have the lack of metadata and variability of file formats turned into stumbling blocks for conceptually new science that massive data reanalysis would allow. In fact, it would not be an understatement to claim that all historical data lack one or another piece of critical information necessary to interpret data in a way that might be different from what the original project intended.

From the time series to the TFs, there is a real need to modernize this variety of MT data formats to common standards that are fully documented, platform-independent, and extensible. This need is widely recognized in the global EM community, which has paid substantial attention in recent years to MT and other EM standards modernization. In 2011, SEG formed an EM Data Standards

subcommittee to review self-describing data formats that would facilitate the exchange of marine controlled-source EM (CSEM) and MT field data. In October 2015, the EM field data component of the SEG-D 3.0 standard for the memory-intensive time series data were finalized. Meanwhile, the Australian SEG established a parallel working group for various kinds of EM data; to our knowledge, no report was published based on that effort. However, other than in our group (IRIS EMTF, 2018; Kelbert et al., 2011, 2018), no searchable metadata frameworks have been established to make noncommercial EM data sets discoverable online, and the modernization of EMTF data formats has not been addressed.

At the time of this publication, an international collaboration (Kelbert et al., 2018) has recently been established with the aim of developing new standards for EM-data sharing that will be self-descriptive, machine readable, and discoverable. In Europe, the European Plate Observing System (EPOS) is a research infrastructure through which the scientific communities will make data and services available (EPOS, 2018). Data provided through EPOS will conform to common standards and be available from a single portal to support cross-disciplinary research. This includes MT data and conductivity models. Luleå University of Technology is responsible for the implementation of MT Data And MOdels (MTDAMO) service within EPOS. The Geophysical Instrument Pool Potsdam, Germany (GIPP) experiment and data archive is the platform for long-term archiving of geophysical experiment data. In the United States, EarthScope MT (2006–2018) is a program of the National Science Foundation (NSF), led by Oregon State University (OSU). Collaboration between OSU, the U.S. Geological Survey, and the Incorporated Research Institutions for Seismology (IRIS) generated the first open database of searchable MT data, IRIS EMTF. NSF's EarthCube is an ongoing initiative focused on enabling data sharing, interoperability, and reproducibility across the geosciences. In Australia, AuScope (2007–2018) provides access to MT instruments to allow researchers to image the subsurface of the earth. Geoscience Australia and Australian State and Territory Geological Surveys and research organizations collect MT data under government programs. All make MT data accessible as traditional file downloads from their sites, and some use the National Computational Infrastructure for Virtual Laboratories and high-performance computing-based processing of MT data. Each of these projects aims to serve their data to the public in a self-describing, well-documented, efficient, and searchable manner. To that end, data formats and archiving strategies are being developed. International collaboration and convergence on metadata content and archiving strategies will allow interoperability between their respective formats and databases, to ensure a smooth data interaction experience for the user. Strategies involve common or consistent metadata models, mirroring of each other's servers, development of format-conversion tools, and other pathways to ensure enhanced backup and accessibility of international data.

We hope that our work can help inform these efforts in the realm of EMTF data sharing. The format that we propose in this publication may or may not become a formal international standard, but it will bring attention to the issues and subtleties of MT data and metadata, which do not depend on the details of the file format specification. Here, we provide a complete solution to these issues based on the extensible markup language (XML) and describe a format of our own, EMTF XML, capable of storing the information from any EMTF data file without loss of content. XML-based formats are extensible, platform-independent, and self-describing by design; they are also well-suited for archival systems and searchable access mechanisms. As opposed to the time series, the EMTFs are not storage- or memory-intensive; at the same time, they contain intricate metadata critical for long-term storage and further interpretation (i.e., inversion). The XML framework is an excellent fit to fully and concisely represent these TFs to enable sharing. The matrix-based approach to data storage allows, but does not require, the inclusion of the complete statistical covariance matrices in EMTF XML data files. We have accompanied this development with open-source conversion tools, written in Fortran 90. These conversion tools are very capable and go to great lengths to parse the complete variability of the historical file formats and convert them to the new EMTF XML data format. The tools also allow for the lossy conversion from EMTF XML to these historical formats, to comply with the formerly established workflows of the MT community. Storage of the EM time series is not a focus of the current publication, but we note that the metadata solutions developed for the TFs can also be applied, with slight modifications, to time series sharing and archiving.

CRITICAL METADATA FOR EM TFs

Metadata are critical for long-term data archiving and sharing. For XML metadata, we have defined a set of minimal fields that make the data self-descriptive. These include formal copyright and conditions of use, geographic name and location, data quality identifiers, data provenance, processing assumptions, and time period and period range. These fields are searchable, and many can be color-coded on a map, making it easy to view and pull the necessary data content. The structure is flexible and is set up for optional but useful additional metadata, including instrument settings and other (more MT-specific) details, provided through the optional XML lists (Example 12) as part of EMTF XML file creation using the supplied conversion software, the electromagnetic transfer function file conversion utilities (EMTF FCU).

Five elements are expected to contain metadata in EMTF XML: Provenance, Copyright, Site, FieldNotes, and ProcessingInfo. They are listed here in the order they are expected in EMTF XML. However, for discussion, we take a more freeform approach and cover the material in the order of relevance. XML is a flexible and extensible language, so as long as this information is present, it can typically be parsed; however, for optimal compatibility with archiving databases, we suggest that the element order be maintained.

In our experience, example XML is typically more descriptive for the community than an XML schema; for this reason, our XML schema is no longer supported. Instead, several complete, working examples are provided in Appendix C, as well as in the digital attachment to this paper, found in the footnotes on the first page. Here, we briefly describe each of these core metadata elements and we outline the minimal metadata requirements for each.

Site identification and location

Site defines the critical metadata that uniquely identify the TF within the project, as well as globally, by providing an unambiguous geographic location. Details are found in Appendix (Listing 3). Here, we specify the key **Project**, which is intended to provide a concise, ideally abbreviated, form for the large-scale data collection umbrella that is responsible for the data. This could be a major

government initiative, an institution, or a similar indicator. A Project could contain multiple Surveys.

The **Survey** is also provided, and is a free-form text. A Survey defines a collection of data united by common authors, years of data collection, and exploration goals. We also specify **YearCollected** and **Country** as critical metadata.

The site itself is identified within the Survey by a concise and unique identifier (**Id**) and a **Name**. The latter is a human-readable geographic location that could be used to roughly place the site; this typically is the full name of the nearest village or geographic feature (e.g., Little Moose Pond, ME, USA).

Finally, the complete geographic location is required for a valid EMTF XML. For completeness, we always specify the datum, and latitude, longitude, and elevation. Declination (with an epoch) is optional, but convenient, metadata. Note that capitalization matters in the XML.

```
<Location datum="WGS84">
  <Latitude>45.532102</Latitude>
  <Longitude>-69.864103</Longitude>
  <Elevation units="meters">361.725</Elevation>
</Location>
```

Timing and data quality

The timing of the data collection is not usually considered that important for EMTFs. However, it is a useful detail when searching for original time series collected during a specific magnetic event. We require that start and end times are specified; if unknown, we specify a reference date or the year of data collection.

Finally, we defined a simple scale for assessing data quality. Data quality **Rating** goes from 0 to 5, as described in Table 1 (a default value for the survey may be specified in the configuration file; see Listings 1 and 2). The data quality **Flag** is typically zero and is set to 1 if the data require manual attention due to cultural-noise issues or other nonphysical behavior. The **Comments** fields provide room for explanations. The **Rating** and the **Flag** values may be plotted on a map during a database search.

Data orientation

MT and other EM TF data are recorded in the field in a wide variety of orientations. Whereas the magnetic field channels are always at a right angle to each other, with the first magnetic field

Table 1. Data quality rating and assessment convention for MT TFs.

Data quality rating	Quality assessment
5	Great TF for the stated period range
4	Good TF for the stated period range
3	Could be noticeably improved by additional runs
2	Serious issues with one or both modes
1	Poor TF that can barely be used for inversion
0	No quality control available

Note: Quality is assessed by either an expert on an algorithm based on the smoothness of TF curves, which are usually scaled by the square root of the period (in seconds) and inspected against log period, as well as based on the amplitudes of the error bar estimates.

component HX most often pointing to geomagnetic north, the electric field channels are oriented in various configurations at an "X," "L," or "T" shape such that the first electric field component EX points as close to geomagnetic or geographic north as is practical, given the field conditions. In certain challenging field circumstances that occur more often than one might think, the channels are oriented to some other, arbitrary, directions that possibly deviate from orthogonal.

For correct interpretation of the data, it is critical that these channel orientations are correctly recorded. However, in an analysis that wishes to use the measurements jointly with other data, it is also critical that the user rotates the TFs to a common orthogonal coordinate system. For a modern 3D interpretation, it is most common to interpret the data in a coordinate system that is oriented to geographic north. Many old data files are ambiguous with respect to data orientation, and they often miss critical pieces of information (Example 5), or they contain several conflicting messages to that effect (Example 2).

Moreover, EDI data files (unless they include SPECTRA, as we describe in detail in the following section) do not contain sufficient information for statistically correct rotation of error bars. Our original intention was therefore to archive the data as they were provided, without a data rotation. In this strategy, we presumed that the data were oriented to the site layout, as defined by the input and output channel orientations, for example,

```
<InputChannels ref="site" units="m">
  <Magnetic name="Hx" orientation="0.0" x="0.0"
  y="0.0" z="0.0"/>
  <Magnetic name="Hy" orientation="90.0" x="0.0"
  y="0.0" z="0.0"/>
</InputChannels>
<OutputChannels ref="site" units="m">
  <Magnetic name="Hz" orientation="0.0" x="0.0"
  y="0.0" z="0.0"/>
  <Electric name="Ex" orientation="0.0" x="-50.0"
  y="0.0" z="0.0" x2="50.0" y2="0.0" z2="0.0"/>
  <Electric    name="Ey"    orientation="90.0"
  x="0.0" y="-40.0" z="0.0" x2="0.0" y2="40.0"
  z2="0.0"/>
</OutputChannels>
```

We have since come to realize that this strategy can result in severe data misinterpretation by scientists who are not well versed with certain subtleties of MT data because an arbitrary channel orientation would need to be corrected for each individual site to ensure consistent orientation of the TFs in a joint interpretation effort. Moreover, because the channel orientations in the file had to match the orientation of the TFs, this strategy did not support preservation of the original site layout information in a rotated EMTF XML file, making any rotation irreversible. Finally, a scientist could not quickly discern whether or not the data were oriented to orthogonal geographic or to some other layout, without checking the orientations of all channels. Even though the declination is specified in the files, if known, this strategy did not provide enough information in the rotated files to know whether or not the data were collected in geomagnetic coordinates in the first place.

To overcome these shortcomings, we adopted an adjusted XML schema to include unambiguous rotation information. We should note that the updated XML format has enough information to

unambiguously store data in any orientation and it also allows for reversible rotations (Example 10). For this, we had to redefine the meaning of the channel orientations, as stored in the file.

First of all, we added a critical new element to the XML `<Site>` header, right after `<Location>`. Two options are allowed:

```
<Orientation angle_to_geographic_north="0.0">
orthogonal</Orientation>
```

or

```
<Orientation>sitelayout</Orientation>
```

In the database, the files will be oriented to orthogonal geographic, as indicated by the first variant of the new `<Orientation>` element. The `orthogonal` and `sitelayout` are keywords, and no other keywords are supported. More generally, the angle to geographic north can, of course, be arbitrary. Alternatively, if the orientation is defined by the site layout, it no longer necessarily has to be (or presumed to be) orthogonal. Even as the data are archived in orthogonal geographic, we will always strive to archive the channel information in their original site layout. If the data are oriented to the site layout, then and only then will the channel orientations also define the data orientation. EMTF FCU v4.0 conversion codes have been generalized to revert any rotation back to the site layout, if needed, as well as to rotate to any other arbitrary orthogonal coordinate system.

To better match the new meaning of the channels, we have now encompassed them with a new element, `<SiteLayout>`. The channels are no longer rotated with the data because they merely indicate the original site layout, whenever this information is known. No other changes to the channel information were required to accommodate these improvements. The example channel block now looks like

```
<SiteLayout>
  <InputChannels ref="site" units="m">
    <Magneticname="Hx"orientation="0.0"
    y="0.0" z="0.0"/>
    <Magnetic   name="Hy"   orientation="90.0"
    x="0.0" y="0.0" z="0.0"/>
  </InputChannels>
  <OutputChannels ref="site" units="m">
    <Magnetic   name="Hz"   orientation="0.0"
    x="0.0" y="0.0" z="0.0"/>
    <Electric    name="Ex"    orientation="0.0"
    x="-50.0"   y="0.0"   z="0.0"   x2="50.0"
    y2="0.0" z2="0.0"/>
    <Electric    name="Ey"    orientation="90.0"
    x="0.0" y="-40.0" z="0.0" x2="0.0" y2="40.0"
    z2="0.0"/>
  </OutputChannels>
</SiteLayout>
```

Additional information (such as "geomagnetic") may be specified as a `SiteLayout` attribute, but we opted against providing this information; it can be discerned from the channel's orientations, which provide a significantly more general way to record this information. Duplication of information in verbal and numeric form leads to conflicts and confusion, rather than to clarity.

Finally, we have also added an optional descriptive element called `<RotationInfo>`. In our data-archiving practice, it has helped us to record any subtleties and ambiguities with respect to discerning the data rotation that we had to overcome before the data could be archived. Our human interpretation of human omissions from 20 to 30 years ago is not flawless. Therefore, these considerations are an important piece of information for anyone who has concerns about any particular data site and needs to find the root of the problem. Our new strategy is to always rotate the TFs to orthogonal geographic for archival at the IRIS EMTF (Kelbert et al., 2011) database.

Critical processing information

One of the persistent ambiguities in EM data processing is the sign convention, either $+i\omega t$ or $-i\omega t$, that is assumed when the Fourier transform is applied to the time series. Either of these is correct, and the choice of the sign is purely a matter of preference. However, it is critical that this choice is recorded along with the TF. None of the historical EMTF file formats recorded the assumed sign convention, causing much confusion for all but the creator of the files. Indeed, among some of the most common EM robust processing codes, EMTF (Egbert and Booker, 1986) and BIRRP (Chave et al., 1987) assume $+i\omega t$, while the Larsen et al. (1996) uses the $-i\omega t$ assumption.

We allow our `ProcessingInfo` element to be as detailed as desired, optionally including the full information about the remote reference site and the processing software. However, `SignConvention`, `ProcessedBy`, and `ProcessingSoftware` are considered critical information and are therefore included in Listing 3. An example of minimal processing details is as follows:

```
<ProcessingInfo>
  <SignConvention>exp(+ i\omega t)</SignCon-
  vention>
  <RemoteRef type="Robust Remote Reference"/>
  <ProcessedBy>Gary Egbert and Prasanta Patro
  </ProcessedBy>
  <ProcessingSoftware>
    <Name>EMTF</Name>
    <LastMod>1998-03-24</LastMod>
    <Author>Gary Egbert</Author>
  </ProcessingSoftware>
</ProcessingInfo>
```

Critical provenance information

Provenance documents the file history. This includes the creation time and application and the contact details of a **Creator** (the author who wrote the original TF) and of a **Submitter** (the person responsible for quality control and/or uploading the EMTF XML file to a database for long-term storage). In our case, the IRIS XML database is used. All of this is optional, but highly recommended, and is considered critical metadata for storage and data discovery. Create time is particularly useful for versioning, allowing an updated variant of the same data to be submitted at a later time.

Credit attribution for data collection

One of the significant advances of this work is the creation of the first **Digital Object Identifiers (DOIs)** for EMTFs. DOIs, attributed directly to data sets, allow us to give proper credit to the very significant effort of data collection, allowing users to cite the data sets directly in any new publications in lieu of (or in addition to) data interpretation papers. This provides a great incentive of data sharing

that was never before in place in the MT community. DOI attribution has additional benefits, which include the ability to track the use of any particular data set and notify the users of any updates as necessary.

To make this possible, we reached out to the authors of every data set that we archived or were planning to archive to obtain the critical information for the data citation, namely, the authors, years of data collection, title, acknowledgements, and selected publications. We found that this was a learning process for everyone involved and that the concept of data citation did not come naturally to the MT community, but was, in retrospect, appreciated. We also found that the authorship of many historical data sets is notably different from that in the final publications, if such publications even exist.

The greatest technical challenge was achieving the right granularity of the DOIs. Within the IRIS EMTF archive, a unique DOI is given and automatically included within any EMTF XML data file upon submission to EMTF database. However, for data citations to be practical in publications, one would need a data DOI that points to a collection of data, such as a survey, united by the common authors, dates, geographical area, and purpose. Jointly with IRIS, we have devised a strategy to attribute a unique DOI to a geophysical survey. This goes under a "Survey DOI" and is available along with the complete data citation, whenever any site is selected and opened in the database. Users are strongly encouraged to cite Survey DOIs in their publications, just like they would cite a paper.

In EMTF XML, **Copyright** is precisely what it says. It contains the **Title**, **Authors**, and **Year** for a data citation (a **DOI** should only be included if it already exists). The database may be searched for an author's name based on the **Authors** citation field. **Copyright** also contains information on the allowed usage of the data. **ReleaseStatus** element can currently take three values: "Unrestricted Release," "Academic Use Only," and "Restrictions Apply." Depending on this choice, the **ConditionsOfUse** is a concise license and disclaimer. Depending on the value of the **ReleaseStatus**, a version of this text is automatically included through EMTF FCU, and the default options are provided with the code. Other, more formal, licensing options may also be added. The **Copyright** element is required metadata and may not be omitted.

Optional field notes

EMTF XML and conversion codes also support the inclusion of detailed field notes into the body of the file. The corresponding
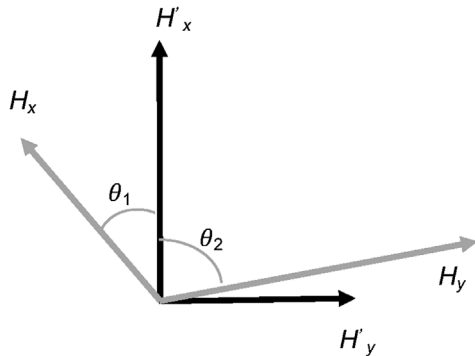

Figure 1. Schematic illustrating rotation of two arbitrarily oriented magnetic field channels H_x and H_y to an orthogonal coordinate system; this is oriented to geographic north ($\theta = 0$). The new magnetic field components are denoted by H_x' and H_y', respectively. Note that in this illustration, $\theta_1 < 0$.

element, **FieldNotes**, is completely optional. In our practice, this information is included from the Site, Run, and Channel lists as discussed in Appendix A (also, Example 12). For each site run, the field notes contain the information about the **Instrument** and the **Magnetometer** (**Manufacturer**, **Name**, **ID**, and **Settings**, the latter being as detailed as desired) and the multiple **Dipole** elements, which also include the **Length**, **Azimuth**, and detailed **Electrode** information.

ROTATION OF MT TRANSFER FUNCTIONS

This section provides the mathematical background for the novel spectra to EMTF conversion tool that also computes the full error covariance, allowing for future change of TF orientation without loss of information. This conversion tool is integrated into EMTF FCU v4.0. The code, as currently implemented, supports all variants of rotation described here and follows the same logic as outlined. It does not currently allow for channel tilt. Usage examples are provided (Example 7 for remote reference; Example 8 for a single station).

General formulation

Consider rotation of the two horizontal magnetic field components on the horizontal plane (no tilt). The original site layout is arbitrary, with H_x and H_y pointing at angles θ_1 and θ_2 (Figure 1), an arrangement that we wish to transform to an orthogonal coordinate system rotated by angle θ against geographic north. The magnetic field in the transformed, orthogonal coordinate system is denoted by H_x' and H_y', respectively.

Let us first write down the transformation to the new orthogonal system, from the original layout:

$$H_x' = \cos(\theta_1 - \theta)H_x + \cos(\theta_2 - \theta)H_y, \quad (1)$$

$$H_y' = \sin(\theta_1 - \theta)H_x + \sin(\theta_2 - \theta)H_y. \quad (2)$$

Similarly, for the output channels:

$$H_z' = H_z, \quad (3)$$

$$E_x' = \cos(\theta_4 - \theta)E_x + \cos(\theta_5 - \theta)E_y, \quad (4)$$

$$E_y' = \sin(\theta_4 - \theta)E_x + \sin(\theta_5 - \theta)E_y.. \quad (5)$$

Define

$$\mathbf{Q}_{lm} = \begin{pmatrix} \cos(\theta_l - \theta) & \cos(\theta_m - \theta) \\ \sin(\theta_l - \theta) & \sin(\theta_m - \theta) \end{pmatrix} \quad (6)$$

for pairs of channels l and m, such that $\mathbf{H}' = \mathbf{Q}_{12}\mathbf{H}$ for the input magnetic field channels $l = 1$ and $m = 2$.

Consider $\mathbf{U} = (\mathbf{Q}_{lm}^T)^{-1}$ and note that $\mathbf{Q}_{lm}^{-1} = \mathbf{U}^T$ (because the transpose and inverse commute). In general,

$$U = (\mathbf{Q}_{lm}^T)^{-1} = \frac{1}{\sin(\theta_m - \theta_l)} \begin{pmatrix} \sin(\theta_m - \theta) & -\sin(\theta_l - \theta) \\ -\cos(\theta_m - \theta) & \cos(\theta_l - \theta) \end{pmatrix}. \quad (7)$$

The inverse is well-defined unless the axes are parallel, which should never happen in MT. If the original site layout is orthogonal, $\theta_m = \theta_l + 90$. In this case, $\sin(\theta_m - \theta_l) = 1$, $\sin(\theta_m - \theta) = \cos(\theta_l - \theta)$, $\cos(\theta_m - \theta) = -\sin(\theta_l - \theta)$, and $\mathbf{Q}^{-1} = \mathbf{Q}^T$. Then, both reduce to the traditional rotation matrix:

$$\mathbf{U} = \mathbf{Q} = \begin{pmatrix} \cos(\theta_l - \theta) & -\sin(\theta_l - \theta) \\ \sin(\theta_l - \theta) & \cos(\theta_l - \theta) \end{pmatrix}. \quad (8)$$

Alternatively, accounting for the more general case that allows for channels to tilt at an angle ψ relative to the horizontal plane, $\mathbf{Q}_{lm}$ could be generalized to

$$\mathbf{Q}_{lm} = \begin{pmatrix} \cos(\theta_l - \theta)\sin(\psi_l) & \cos(\theta_m - \theta)\sin(\psi_m) \\ \sin(\theta_l - \theta)\sin(\psi_l) & \sin(\theta_m - \theta)\sin(\psi_m). \end{pmatrix} \quad (9)$$

Now, consider an impedance $\mathbf{Z}$ such that $\mathbf{E} = \mathbf{ZH}$ for arbitrary orientations of the magnetic and electric fields. Rotated, $\mathbf{H}' = \mathbf{Q}_{12}\mathbf{H}$ and $\mathbf{E}' = \mathbf{Q}_{45}\mathbf{E}$. Note, however, that this framework is easily adjusted to an arbitrary number of output channels by extending the definition of $\mathbf{E}$ and using the general formulation $\mathbf{E}' = \mathbf{Q}_E\mathbf{E}$. For example, we can also define the transformation matrix $\mathbf{Q}_E$ for the three output channels as

$$\mathbf{Q}_E = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{Q}_{45} \end{pmatrix}. \quad (10)$$

By itself, the tipper $\mathbf{T}$ tilted at an angle ψ relative to the vertical axis may be rotated by setting the scalar $\mathbf{Q}_E = \cos(\psi)$. For symmetry of expression, let us also define $\mathbf{Q}_H = \mathbf{Q}_{12}$. Then,

$$\mathbf{Q}_E^{-1}\mathbf{E}' = \mathbf{Z}(\mathbf{Q}_H^{-1}\mathbf{H}'), \quad (11)$$

$$\mathbf{E}' = \mathbf{Q}_E\mathbf{Z}\mathbf{Q}_H^{-1}\mathbf{H}', \quad (12)$$

$$\mathbf{Z}' = \mathbf{Q}_E\mathbf{Z}\mathbf{Q}_H^{-1}, \quad (13)$$

where $\mathbf{Z}'$ is the impedance in the new orthogonal coordinate system with orientation θ relative to geographic north. Then, the rotation of the impedance and tipper to an orthogonal coordinate system may be described in general terms with the formulation

$$\mathbf{Z}' = \mathbf{VZU}^T, \quad (14)$$

by setting $\mathbf{V} = \mathbf{Q}_E$ and $\mathbf{U} = (\mathbf{Q}_H^T)^{-1}$, as above. The same expression will hold for any EMTF. Similarly, the tipper $\mathbf{T}$ tilted at an angle ψ relative to the vertical axis may be rotated by setting $\mathbf{V} = \mathbf{Q}_E = \cos(\psi)$, $\mathbf{T}' = \mathbf{VTU}^T$.

To rotate from the orthogonal coordinates $\mathbf{E}' = \mathbf{Z}'\mathbf{H}'$ to the original layout, we would perform the opposite transformation, which in its general form looks like

$$\mathbf{Q}_E\mathbf{E} = \mathbf{Z}'(\mathbf{Q}_H\mathbf{H}), \quad (15)$$

$$\mathbf{E} = \mathbf{Q}_E^{-1}\mathbf{Z}'\mathbf{Q}_H\mathbf{H}, \quad (16)$$

$$\mathbf{Z} = \mathbf{Q}_E^{-1}\mathbf{Z}'\mathbf{Q}_H. \quad (17)$$

Thus, the reverse rotation of the traditional MT impedance from the orthogonal coordinate system to the original site layout would require setting $\mathbf{V} = \mathbf{Q}_E^{-1}$ and $\mathbf{U} = \mathbf{Q}_H^T$ in equation 14, but otherwise is amenable to the same treatment as the forward rotation.

Rotation of the error bars associated with the MT TFs is a less trivial matter. If the full error covariance is available, it is usually supplied in the form of two matrices: the inverse signal power $\mathbf{S}$ and the residual covariance matrix $\mathbf{N}$ (Eisel and Egbert, 2001). The full covariance is the Kronecker delta product of the two:

$$\mathbf{Cov}[\mathbf{Z}] = \mathbf{N} \otimes \mathbf{S}, \quad (18)$$

so that

$$\mathbf{Cov}[Z_{ij}Z_{i'j'}] = N_{ii'}S_{jj'}, j, j' = 1, 2 \text{ and } i, i' = 1, \ldots, N, \quad (19)$$

where N is the number of output channels. To compute the variance,

$$\sigma_{ij}^2 = \mathbf{Var}[Z_{ij}] = \mathbf{Cov}[Z_{ij}Z_{ij}] = N_{ii}S_{jj}, j = 1, 2$$
$$\text{and} \quad i = 1, \ldots, N. \quad (20)$$

Upon rotation,

$$\mathbf{Cov}[\mathbf{Z}\prime] = \mathbf{N}' \otimes \mathbf{S}' = [\mathbf{V} \otimes \mathbf{U}][\mathbf{N} \otimes \mathbf{S}][\mathbf{V} \otimes \mathbf{U}]^T, \quad (21)$$

$$\sigma_{ij}'^2 = \mathbf{Var}[Z_{ij}'] = \mathbf{Cov}[Z_{ij}'Z_{ij}'] = N_{ii}'S_{jj}', j = 1, 2$$
$$\text{and} \quad i = 1, \ldots, N. \quad (22)$$

The new inverse signal power $\mathbf{S}'$ and residual covariance matrix $\mathbf{N}'$ may thus be computed using the expressions $\mathbf{N}' = \mathbf{VNV}^T$ and $\mathbf{S}' = \mathbf{USU}^T$, as in Eisel and Egbert (2001).

If all that is available are the variances, a variance matrix

$$\mathbf{D} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix} \quad (23)$$

is formed, and the rotation is performed in much the same way as that for the impedance tensor: $\mathbf{D}' = \mathbf{VDU}^T$.

Computation of the full error covariances from cross-power spectra data

Consider a traditional linear regression model $\mathbf{Y} = \mathbf{Xb} + \epsilon$, where $\mathbf{X}$ and $\mathbf{Y}$ are two sets of measurements, $\mathbf{b}$ is the vector of unknowns, and ϵ are the error estimates. For this formulation, the estimate is $\hat{\mathbf{b}} = (\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*\mathbf{Y}$, where the inverse signal power matrix

$\mathbf{S} = (\mathbf{X}^*\mathbf{X})^{-1}$. The estimation errors $\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}}$, where $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}(\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*\mathbf{Y}$.

In this simple formulation, the so-called hat matrix $\mathbf{P}$ is defined by $\hat{\mathbf{Y}} = \mathbf{P}\mathbf{Y}$. Then, $\mathbf{P} = \mathbf{X}(\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*$. It can be easily seen that $\mathbf{P}$ is a projection matrix, i.e., $\mathbf{P} = \mathbf{P}^*$ and $\mathbf{P}$ is idempotent ($\mathbf{P}^2 = \mathbf{P}$).

Then, the residual covariance estimate is

$$\mathbf{N} = \hat{\Sigma} = \hat{\sigma}_\nu^2[(\mathbf{I} - \mathbf{P})\mathbf{Y}]^*[(\mathbf{I} - \mathbf{P})\mathbf{Y}] = \hat{\sigma}_\nu^2 \mathbf{Y}^*(\mathbf{I} - \mathbf{P})\mathbf{Y}, \quad (24)$$

because $\mathbf{I} - \mathbf{P}$ is a projection. Here, $\hat{\sigma}_\nu^2$ is a scalar error variance estimate factor.

To apply this analysis to MT, let us model the output channels $\mathbf{E}$ as the response variable ($\mathbf{Y}$), $\mathbf{H}$ as the input variable ($\mathbf{X}$), and the impedance $\mathbf{Z}$ as the unknown parameter matrix ($\mathbf{b}$), resulting in

$$\mathbf{E} = \mathbf{H}\mathbf{Z}^T + \epsilon, \quad (25)$$

where $\mathbf{E}$ and $\mathbf{H}$ are defined, for simplicity of notation, as `horizontal`, rather than vertical, vectors. If we were to transpose both sides of equation 25, we would arrive at the more traditional definition of the impedance. To further generalize the analysis, we note that the same expression may be used for a variety of output channels, so that the definition of $\mathbf{E}$ may include the vertical magnetic field component, as well as, or instead of, the electric field output channels, and the same analysis would apply.

Single station

The standard analysis above can be directly applied to single-station MT processing. Indeed, the inverse signal power matrix may be computed as

$$\mathbf{S} = (\mathbf{H}^*\mathbf{H})^{-1}, \quad (26)$$

and the residual covariance estimate is

$$\mathbf{N} = \hat{\sigma}_\nu^2 \mathbf{E}^*(\mathbf{I} - \mathbf{P})\mathbf{E}, \quad (27)$$

$$= \hat{\sigma}_\nu^2[\mathbf{E}^*\mathbf{E} + (\mathbf{E}^*\mathbf{H})(\mathbf{H}^*\mathbf{H})^{-1}(\mathbf{H}^*\mathbf{E})], \quad (28)$$

an expression that is easy to obtain from the cross-power spectra. The scalar variance $\hat{\sigma}_\nu^2$ is the inverse of `AVGT` value from the EDI SPECTRA file, the number of independent averages in time, at each frequency.

Then, the full error covariance matrix and the variances of TF components may be obtained via equations 19 and 20, respectively. This allows us to proceed with an arbitrary coordinate rotation directly from the impedance or vertical magnetic field TF estimates, without a loss of information.

Remote reference

Following Eisel and Egbert (2001), for the remote reference case, the inverse signal power $\mathbf{S}$ is written as

$$\mathbf{S} = (\mathbf{R}^*\mathbf{H})^{-1}(\mathbf{R}^*\mathbf{R})(\mathbf{H}^*\mathbf{R})^{-1}, \quad (29)$$

where $\mathbf{R}$ is the horizontal magnetic field at the remote site and $\mathbf{H}$ is the local horizontal magnetic field.

For the remote reference case, $\mathbf{P} = \mathbf{H}(\mathbf{R}^*\mathbf{H})^{-1}\mathbf{R}^*$ is no longer a projection matrix (nor is it Hermitian) and the full $\mathbf{N} = \hat{\sigma}_\nu^2[\mathbf{Y}^*(\mathbf{I} - \mathbf{P})^*(\mathbf{I} - \mathbf{P})\mathbf{Y}]$ needs to be evaluated:

$$\mathbf{N} = \hat{\sigma}_\nu^2[\mathbf{E}^*(\mathbf{I} - \mathbf{H}(\mathbf{R}^*\mathbf{H})^{-1}\mathbf{R}^*)^*(\mathbf{I} - \mathbf{H}(\mathbf{R}^*\mathbf{H})^{-1}\mathbf{R}^*)\mathbf{E}], \quad (30)$$

$$= \hat{\sigma}_\nu^2[\mathbf{E}^*\mathbf{E} - \hat{\mathbf{Z}}^*\mathbf{H}^*\mathbf{E} - \mathbf{E}^*\mathbf{H}\hat{\mathbf{Z}} + \hat{\mathbf{Z}}^*(\mathbf{H}^*\mathbf{H})\hat{\mathbf{Z}}]. \quad (31)$$

Cross-power spectra, as stored in a remote referenced EDI SPECTRA file, may be unpacked to comprise the six matrices $\mathbf{R}^*\mathbf{H}$, $\mathbf{R}^*\mathbf{E}$, $\mathbf{H}^*\mathbf{E}$, $\mathbf{R}^*\mathbf{R}$, $\mathbf{E}^*\mathbf{E}$, and $\mathbf{H}^*\mathbf{H}$. If only the electric field output channels are present, these matrices are 2×2. If the vertical magnetic field also exists, the vector $\mathbf{E}$ has three complex components, to also include the vertical magnetic field output channel. Then, $\mathbf{R}^*\mathbf{E}$ and $\mathbf{H}^*\mathbf{E}$ are expanded to form 2×3 matrices, and $\mathbf{E}^*\mathbf{E}$ is of size 3×3.

This expression, scaled by the number of independent averages in time (`AVGT`) at every period, provides us with the full covariance matrix for the EMTFs stored in the cross-spectra format. When this formulation is compared to the approximate Stodt (1983) variance expressions for remote reference that have been used in the past, the impedances, tippers, and impedance variances we obtain are exactly the same, within numerical precision. The tipper variances are very close at short periods but are significantly different at the longest periods. Remote reference formulation can, of course, also be used for single-station analysis, by substituting $\mathbf{H}$ for $\mathbf{R}$ in the above expressions. This substantially simplifies the analysis, so a general code might want to account for both cases, even at the expense of computation costs, which in any case are trivial on a modern personal computer.

The final point of note is that when we use the cross-power spectra to compute the full $\mathbf{S}$ and $\mathbf{N}$, these estimates in general also include the off-diagonal terms of the residual covariance that relate the vertical-field TF estimate to that of the impedance. In EMTF XML formulation, we chose to ignore these values, allowing for a cleaner archiving, so that there are distinct error covariance matrices for each data type.

HOW DATA FORMATS DETERMINE SCIENCE LIMITATIONS: THE ISSUE OF ERROR COVARIANCES

In view of the analysis developed above, here, we present a case for a matrix-based approach to data storage.

Matrix-based approach to data storage

We find that one of the critical flaws of the EDI format (Wight, 1988) — which made it so flexible in the first place — was storage of each of the data components as a separate line. While convenient for record-keeping, this storage format put a limit of the types of statistical estimates that could be conveniently recorded, resulting in a practice that omits statistical covariances from everyday use (variance was more commonly used). The actual MT TFs are matrices, and allowing for a general statistical covariance would require storing a matrix to accompany each of the data components. In addition, matrices are often defined to relate different TF components together statistically (e.g., Eisel and Egbert, 2001). Thus, we reasoned that a matrix-based approach to EMTF data storage would be more appropriate.

Conceptually, each of the MT data types (and their error estimates) could always be represented as either a real or a complex matrix. For primary data types, each data element in the matrix would correspond to a specific input and output field channel. Typically, input channels are always magnetic; output channels could be either magnetic or electric. The number and the names of the channels are kept flexible in the XML format and the conversion codes. These channels are typically described in the EDI files; however, due to the human factor, the EDI format was not always used consistently, in some cases describing only the channels present in the file, and in other cases also describing the remote reference channels that were, essentially, metadata for the purposes of TF parsing. Supporting an automated conversion between the two data formats requires that (1) a matrix-based storage would be allocated for each of the data types present in the EDI files and (2) each line of the EDI file would be read unambiguously into a component of the appropriate matrix.

In an EMTF XML document, the data are grouped by period. For each period, the data components are grouped by data type. The syntax for the element names echoes that of the EDI files: It is an uppercase abbreviation, as described in Tables 2 and 3. Within each data type, there may be multiple components, with an important restriction that all components within a type have the same units. For example, derived data types `apparent_resistivity` (RHO) and `impedance_phase` (PHS) are stored as two different data types, whereas all components of `phase_tensor` are stored together. Missing data are supported directly in the XML by omitting the relevant TF components. The matrix size is determined by the numbers of respective input and output — whether electric or magnetic — channels (these numbers default to 1, e.g., for `impedance_determinant`).

Any statistical error estimates are stored in a similar manner, with the element name formed of the data type abbreviation and the

Table 2. List of primary data types currently supported by the XML format and the conversion tools.

EDI/XML name	Real/complex	Output	Input	Units	Description	Tag
Z	Complex	E	H	[mV/km]/[nT]	MT impedance	Impedance
Z	Complex	E	H	[mV/km]/[nT]	MT impedance — off-diagonal components only	Off-diagonal_impedance
T	Complex	H	H	[]	Vertical field transfer functions (Tipper)	Tipper
Q	Complex	E	H	[mV/km]/[nT]	MT interstation impedance	Interstation_impedance
P	Complex	H	H	[]	Interstation magnetic field transfer functions	Interstation_transfer_functions

Note: Additional primary data types may be easily added without any modification to the conversion codes: A simple XML file that described the new data type must be added to the DATATYPES directory of the format conversion tools. The [] stands for no units.

Table 3. List of derived data types currently supported by the XML format and the conversion tools.

EDI/XML name	Real/complex	Output	Input	Units	Description	Tag	Derived from
PT	Real	E	H	[]	Phase Tensor	phase_tensor	Impedance
RHO	Real	E	H	Ohm	Apparent resistivity	apparent_resistivity	Impedance
PHS	Real	E	H	Degrees	Impedance phase	impedance_phase	Impedance
ZEFF	Complex	E	H	[mV/km]/[nT]	Effective impedance	effective_impedance	Impedance
ZDET	Complex	E	H	[mV/km]/[nT]	Impedance determinant	impedance_determinant	Impedance
ZSTRIKE	Real	E	H	[]	Impedance strike	impedance_strike	Impedance
ZSKEW	Real	E	H	[]	Impedance skew	impedance_skew	Impedance
ZELLIP	Real	E	H	[]	Impedance ellipticity	impedance_ellipticity	Impedance
TIPMAG	Real	H	H	[]	Tipper magnitude	tipper_magnitude	Tipper
TIPPHS	Real	H	H	Degrees	Tipper phase	tipper_phase	Tipper
INDMAG	Real	H	H	[]	Induction arrows: magnitude	induction_arrow_magnitude	Tipper
INDANG	Real	H	H	Degrees	Induction arrows: angle (Parkinson convention)	induction_arrow_angle	Tipper
TSTRIKE	Real	H	H	[]	Tipper strike	tipper_strike	Tipper
TSKEW	Real	H	H	[]	Tipper skew	tipper_skew	Tipper
TELLIP	Real	H	H	[]	Tipper ellipticity	tipper_ellipticity	Tipper

Note: Additional derived data types may be easily added without any modification to the conversion codes: A simple XML file that described the new data type must be added to the DATATYPES directory of the format conversion tools. This preliminary list is compiled based on Wight (1988), Hobbs (1992), and Becken et al. (2008).The [] stands for no units.

abbreviation of the statistical error estimate (e.g., Z.VAR, Z.INVSIGCOV, Z.RESIDCOV, and the others). A flexible suite of statistical estimates has also been defined. However, each statistical error estimate requires a specific matrix storage to be defined within the conversion codes; therefore, a simple code modification is required to add to these existing capabilities.

Lossless and lossy data rotation and format conversions

In addition to EDI files, another common way of storing MT data is known as EMTF Z-files (see Egbert and Booker [1986] and Gary Egbert's EMTF robust processing software developed subsequently). Unlike the EDIs, Z-files contain complete information about the error covariances in the form of matrices **S** and **N** (as discussed in the previous section). This allows for unambiguous rotation of Z-files to any coordinate system. However, channel and data orientation information is tightly linked. The data are always oriented to site layout, as defined by the channels. The channel orientations are, in turn, defined relative to the declination value (also usually specified in the file header), not relative to geographic north, as would be true for the EDI data format. If an EMTF Z-file has already been rotated to orthogonal geographic, for example, then the declination in the header is necessarily set to zero, and the orientations are set to 0° and 90° for the x- and y-channel components, respectively. Therefore, rotation back to "site layout" is not meaningful in this setting; we only have information about the current channel orientations, and we have no knowledge of the original site layout after an initial rotation has been performed. Therefore, rotation of Z-files is lossy in the sense that enough metadata are lost in rotation that reverting to the original coordinate system is no longer possible.

In contrast, the EMTF XML files have been designed such that, in theory, no information is lost if the data are rotated and any rotations are reversible. Conversion from the other formats to EMTF XML has been designed in such a way as to preserve all (or the vast majority of) of the data and metadata, i.e., lossless.

	To: EDI MT	EMTF Z-files	EMTF XML
	From:		
EDI MT	Lossy: upon rotation, error bars are no longer correct	Lossy: metadata are not saved	Lossless
EDI SPECTRA	Lossy: covariance matrices are not saved	Lossy: metadata are not saved	Lossless
EMTF Z-files	Lossy: covariance matrices are not saved	Lossy: upon rotation, declination is no longer known	Lossless
EMTF XML	Lossy: covariance matrices are not saved	Lossy: metadata are not saved	Lossless

EMTF FCU: ELECTROMAGNETIC TRANSFER FUNCTION FILE CONVERSION UTILITIES

Overall structure and capabilities

To make the use of the newly developed EMTF XML format practical, we have also released a set of highly capable, open-source format conversion tools, which we have called the Electromagnetic Transfer Function File Conversion Utilities (EMTF FCU), currently at v4.0. The codes are written in Fortran 90, can be compiled on any computational platform, and were originally developed to create the well-documented EMTF XML files now available at IRIS EMTF (2018); Kelbert et al., (2011, 2018).

As discussed earlier, the original EDI or Z-files, as well as the other formerly available EMTF data formats, do not contain sufficient metadata to fully populate these files with information; for example, details of field operating procedures, data processing, and copyright in most cases cannot be found in the original data files. To fully populate this information, we have developed a set of auxiliary XML file formats, including XML lists and a configuration file, which can be used in conjunction with existing data files to create a complete EMTF XML. Creation of the EMTF XML documents is discussed in the Appendices; working examples are also provided.

EMTF FCU is freely available through IRIS SeisCode repository (IRIS EMTF FCU, 2018). An up-to-date snapshot of EMTF FCU v4.0 is provided here. We use the FoX Toolbox for XML input and output. The FoX Toolbox is an open-source Fortran 90 library that implements all objects and methods mandated in Document Object Model (DOM) core levels 1 and 2. The DOM (2000) is a platform- and language-neutral application programming interface that underlies much of the interoperability on the Web, allowing programs and scripts to dynamically access and update the content, structure, and style of documents.

The following standalone programs are available. In addition to a data format conversion, all of these codes are intended to provide general rotation capabilities, as much as allowed by the data-rotation considerations discussed above.

- **edi2xml and z2xml**: Conversion utilities edi2xml and z2xml depend on the FoX library and require additional metadata files for full functionality. These are used to create EMTF XML, which are fully self-contained, searchable, and suitable for long-term storage.
- **xml2z and xml2edi**: Utilities xml2z and xml2edi also depend on the FoX library but do not require any additional files. They have usage restrictions in that they will only work correctly with those XML documents that contain the right data types and statistical error estimates for this type of output storage.
- **xml2xml, edi2edi, z2z, and z2edi**: These are stand-alone codes that require no additional inputs. They can ingest EMTF XML, EDI, EDI SPECTRA, and Z-files, and they can rotate the data to any orthogonal coordinate system specified by the angle to geographic north. Note that conversion from a traditional EDI file to a Z-file (with full error covariances) is not meaningful, and it is therefore not supported.

Support for primary and derived data types

In addition to supporting most standard data types adopted from the EDI format, we have made an effort to devise a scheme that would potentially be easily extensible to more general data types and statistical error estimates. To that effect, EMTF FCU code repository contains a subdirectory DATATYPES, which fully defines the data types that are supported by the code. For complete functionality, conversion codes need to have access to this directory, as well as its complement, COPYRIGHT, which externally defines a set of copyright options and may be just as easily extended.

To convert a survey from EDI to EMTF XML, a user would visually inspect the EDI files to see which data types were present. The user would then define a list of comma-separated "tags" in an XML configuration file, e.g., "impedance, tipper" (see Tables 2 and 3 for lists of allowed tags). As can be seen from the tables, 5 primary and 15 derived data types are currently allowed. However, a new data type (and the corresponding tag) may be added at any time by setting up a new data type XML definition. Importantly, such an extension would not require a code modification and would be completed entirely with the aid of external configuration scripts. Internally, the code structure is flexible enough to accept any EMTF for rotation and format conversion.

As can be seen from the XML definition examples and tables below, each data type tag has a corresponding EDI name that is used for the one-to-one conversion between the formats, as well as to store the actual data matrices in the XML. The format-conversion code then goes through the tag list to add the definitions of each of the data types to EMTF XML, points to the corresponding online documentation, and sets up the matrices for storage and XML writing.

EMTF FCU conversion codes and EMTF XML format restrict themselves to EMTFs, in the most general sense. To this end, every data type is explicitly organized by input and output channels. The common name of the data type (e.g., Zxy, for an impedance component) is also provided for user convenience, but it is not used in any of the conversion tools.

In the XML file, the data are grouped by period. For each period, the data components are grouped by data type. The syntax for the element names echoes that of the EDI files; it is an uppercase abbreviation, as described in Tables 2 and 3. Within each data type, there may be multiple components; however, each data type has a specific number of input and output field components, and all components within a type have the same units. For example, derived data types `apparent_resistivity` (RHO) and `impedance_phase` (PHS) are stored as two different data types, whereas all components of `phase_tensor` (PT) are stored together. Any statistical error estimates are stored in a similar manner, with the element name formed of the data type abbreviation and the abbreviation of the statistical error estimate (e.g., Z.VAR, Z.INVSIGCOV, Z.RESIDCOV, etc.).

Defining a new primary or derived data type is as simple as creating a new simple XML file in the DATATYPES subdirectory of EMTF FCU code, e.g.,

```
<DataType name="Z" type="complex" output="E"
input="H" units="[mV/km]/[nT]">
  <Description>MT impedance</Description>
  <Intention>primary data type</Intention>
  <Tag>impedance</Tag>
</DataType>
<DataType name="RHO" type="real" input="H"
output="E" units="[Ohm m]">
  <Description>Apparent resistivity computed
  from MT impedance</Description>
  <Intention>derived data type</Intention>
  <Tag>apparent_resistivity</Tag>
  <DerivedFrom>impedance</DerivedFrom>
  <SeeAlso>impedance_phase</SeeAlso>
</DataType>
```

Support for statistical estimates

The list of supported statistical error estimates is based on Wight (1988) and Eisel and Egbert (2001):

```
variance
covariance
inverse_signal_covariance
residual_covariance
coherence
multiple_coherence
signal_amplitude
signal_noise
```

These are defined, similarly to the data types, in the DATATYPES directory, e.g.,

```
<Estimate name="INVSIGCOV" type="complex">
  <Description>Inverse Coherent Signal Power
  Matrix (S)</Description>
  <Intention>signal power estimate</Intention>
  <Tag>inverse_signal_covariance</Tag>
</Estimate>
```

However, each statistical error estimate requires a specific matrix storage to be defined within the code. Hence, additional error estimates may be introduced with code modification.

Generality and limits of use

EMTF FCU v4.0 is a powerful set of EMTF format-conversion tools. At present, file conversions between EDI, Z-files, and EMTF XML are implemented with great generality, and an unambiguous coordinate conversion capability is included. EMTF FCU v4.0 may be used to create EMTF XML documents for archiving or self-describing data storage of MT and other EMTFs, using edi2xml or z2xml (Appendix A). These conversions are reversible. EMTF FCU v4.0 can also be used for coordinate conversion without a change of format, using edi2edi, z2z, or xml2xml.

To the authors' knowledge, the ability to compute full error covariances from EDI SPECTRA files has never been developed before. This allows for arbitrary rotations of the TFs without loss of information, even after conversion to EMTF XML. We have also now implemented frequency-by-frequency rotations that allow rotation, and conversion to EMTF XML, of data that went through a principal axis rotation algorithm — a 1970s practice of rotating impedances frequency-by-frequency to minimize diagonal-component amplitudes and make them more suitable for 1D interpretation. This can be undone for EDI files that include variable ZROT values, even though the resultant error estimates may suffer. Finally, we have implemented a general rotation algorithm that allows for rotation of TFs from an arbitrary (not necessarily orthogonal) set of channel orientations to an arbitrary orthogonal coordinate system, or back.

Typical usage:

```
./z2xml  VAQ60bc_R56cohC.zrr  VAQ60bc_R56cohC_
0.xml silent 0
```

should convert to EMTF XML and rotate to orthogonal geographic, while suppressing most output.

EMTF XML provides ample room for streamlined extensions. Additional MT TF data types or related, non-MT EMTFs may be readily recorded in the same framework, the only difference being in the number and types of input and output channels.

Wherever the source is known or estimated directly, the TF framework no longer holds. This concerns related EM applications such as the CSEM method. The framework for the XML format and conversion codes could be adapted to these different types of EM data. The XML metadata content would need to be modified to include detailed transmitter and receiver information, and the stored data components would need to follow a different paradigm. These would be composed of processed data from EM channels and include an explicit source and receiver reference. The data might be organized by the transmitters or receivers, whichever is more appropriate for a particular controlled-source method.

Any additional primary or derived data type is inherently supported **without modifications** to the XML format or conversion codes. This provides ample room for extensions. For example, related, non-MT EMTFs such as C- and D-responses for global induction studies may be readily recorded in the same framework, the only difference being that there would only be a single input channel.

Traditional derived data types are all supported by EMTF FCU and the EMTF XML format. However, correct rotation of derived data types requires recomputation of the primary data types (e.g., impedance, tipper), their rotation, followed by a recomputation of the derived data (e.g., apparent resistivities and phases, tipper magnitude, skew, phase, or ellipticity). General implementation of such a capability is a challenging task that is not currently warranted by data-archiving needs. Therefore, for now, we omit any additional derived data type products on conversion to XML files, unless configured otherwise by the user. If you have a need in any such products, the original EDI files are always included in the archive bundle.

Similarly, any general statistical estimate is supported by the XML format, as long as that estimate relates one or more of primary data type components to each other. However, the inhomogeneity with which, e.g., the coherence and predicted coherence are sporadically recorded in the EDI files is such that correctly reading and interpreting this information in general is challenging. This work is currently not warranted by the few occurrences of these estimates in the historical EDI files, and the coherences are therefore omitted from the XML files. As with the derived data types, for now, they can always be accessed by downloading the original EDI file.

There are also some inherent limitations of our format and approach that could be overcome by developing a similar (but different) framework. Specifically, **not supported** are joint statistical estimates across periods (transmitters), locations (receivers), or data types. The only cross-data type statistical estimates known to the authors that are currently available in MT are the components of the residual covariance matrix (commonly found in Z-files) that correspond to the residual covariance between the impedance elements and the vertical field TFs (tipper). These are not used for rotation, variance estimation, or in the MT inversion and do not need to be stored in the XML database (G. Egbert, 2015, oral communication) within the framework of this project. Whereas information loss, however intangible, is always undesirable, we recognize that this is a necessary step required for streamlined and user-friendly storage and sharing of modern EMTFs. If conceptually new statistical methods arise in the context of EM geophysics, these could be accommodated within a different framework. As a corollary, if an EMTF

XML conversion is converted back to a Z-file, some entries of the residual covariance matrix will be zero. This is known and not a problem.

Finally, during the archiving work, we have encountered several occasions of EDI SPECTRA files that have been edited, perhaps manually, to include fewer frequencies than the number of frequencies (NFREQ) value might suggest. Because we want the output XML file to contain a correct frequency count, we have not attempted to overcome this problem programmatically (although we may implement a workaround in the future). In these rare circumstances, conversion to XML will halt and the original EDI needs to be edited to fix the problem (by adjusting NFREQ value to the actual number of frequencies in the file).

Note also that due to the high variability of the historical EDI files, generalizations of the code to better parse any subset of historical data may be viewed as an ongoing process.

ARCHIVING HISTORICAL MAGNETOTELLURIC DATA

Using EMTF FCU and the new XML-based data format, we for the first time put together a searchable public resource for MT TFs, IRIS EMTF data product at the IRIS DMC (IRIS EMTF, 2018; see also Trabant et al., 2012; Kelbert et al., 2018a). This international database currently includes 5186 sites from 124 surveys collected in the scope of 16 projects. We have made it easy for the community to archive their MT data, and we continue to solicit and receive data contributions. This database is currently widely used by the international MT and space weather communities for teaching and research. Thus, our EMTF XML development has for the first time allowed global and extensive MT data sharing, allowing MT researchers easy access to modern and historical data worldwide.

We also worked with IRIS to develop a DOI strategy for MT data. Data collection and processing require a lot of effort and do not necessarily result in a publication. The capability to assign DOIs to data sets stored at IRIS, directly, thus making them citable for career-related purposes, provides an invaluable incentive to MT community members to continue sharing their data. Additionally, citing the data sets helps track the end-users of each of the data collection efforts, both to assess the usefulness of a project and to propagate corrections and data-related information to the end users. Initially, IRIS DMC could only effectively assign a DOI to (1) all of MT TFs together or (2) each specific data submission. Unfortunately, assigning a separate DOI to each of the data files would result in an unmanageable database; most end users would use hundreds of sites in each journal publication and would not be able to cite each data file independently (rendering the DOIs useless because they would not be used in practice). On the other hand, a citation to the IRIS EMTF database as a whole would not give the necessary credit to the principal investigators for their efforts in data collection. We have agreed on a strategy that would assign a separate DOI to each Project/Survey combination found in the database. This strategy is now implemented and supported by the EMTF XML data format.

CONCLUSION

We conceived and developed a novel, flexible and extensible, EMTF XML data format for MT and, potentially, other EMTFs. We also developed a set of open-source EMTF XML ↔ EDI converters for MT TFs, which we have called electromagnetic transfer

function format conversion utilities (EMTF FCU). For these tools, we used Fortran 90 programming language, making it straightforward for our community to embed the reading and writing routines in their codes. These utilities, which have now matured over a decade of active use, have a great deal of flexibility and correct a large number of inevitable historical metadata omissions, while also allowing the user to supply extensive optional metadata. Additionally, EMTF FCU support arbitrary coordinate rotations, as well as conversions between other common MT file formats, and they have the potential to be easily further extensible.

The novel MT format conversion utilities (EMTF FCU) and the corresponding EMTF XML data format have the following basic intentions:

1) Provide a more modern and general alternative to the EDI standard that would accommodate a wider range of EMTFs, and
2) allow for easy archiving and sharing of historical and modern EMTFs in a searchable, widely available online database.

We hope that our tools, presented here, facilitate these objectives. Nonetheless, there is much room for improvement, primarily to provide continued support for historical and emerging MT data types. EMTF FCU v4.0 is open source and easy to edit; please contribute your efforts back to the community. The value of MT among geosciences hinges on its accessibility. It is our hope that the new, self-descriptive, and searchable EMTF file format and, correspondingly, an openly available EMTF database may help facilitate better documentation and sharing practices within the MT community.

ACKNOWLEDGMENTS

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be accessed via the following URL: https://seiscode.iris.washington.edu/projects/emtf-fcu.

APPENDIX A

CONVERSION OF HISTORICAL FILE FORMATS TO EMTF XML

EMTF FCU v4.0 may be used to create EMTF XML documents for archiving or self-describing data storage of MT TFs, using z2xml or edi2xml.

To supply the metadata necessary for a complete EMTF XML, an XML configuration file (config.xml) needs to be put together and placed in the same directory as your input data. This is where any user-defined information about the experiment is stored. A documented example configuration file for EDI to XML conversion is provided in Listing 1. Similarly, an example configuration file for Z-file to XML conversion is provided in Listing 2.

Listing 1: "A documented sample config.xml needed to convert EDI to EMTF XML"

```xml
<Configuration>
  <!– set this to zero or omit if the time series
  are not archived at IRIS DMC –>
  <TimeSeriesArchived>0</TimeSeriesArchived>
  <!– Project / Survey combination used to allo-
  cate survey DOI; Project should contain no
  spaces –>
  <Project>Lithoprobe</Project>
  <Survey>AB</Survey>
  <!– Project.SiteID.YearCollected used to
  create a unique data ID in SPUD –>
  <YearCollected>1995</YearCollected>
  <Country>Canada</Country>
  <!– required, comma-separated list of data
  types. See list of possible types –>
  <Tags>impedance,tipper</Tags>
  <!– required, except for DOI which should be
  omitted if it does not yet exist –>
  <Citation>
    <Title>Lithoprobe, Canada Magnetotelluric
    Survey</Title>
    <Authors>Jim Craven and Alan Jones</
    Authors>
    <Year>1995</Year>
    <DOI></DOI>
  </Citation>
  <!– include some relevant text –>
  <THANKS>acknowledgements.txt</THANKS>
  <PAPERS>references.txt</PAPERS>
  <README>readme.txt</README>
  <!– supported options: Unrestricted Release/
  Academic Use Only/Conditions Apply –>
  <ReleaseStatus>Unrestricted Release</Relea-
  seStatus>
  <!– allows to supplement missing provenance and
  processing information from EDI files –>
  <AcquiredBy></AcquiredBy>
  <Creator>
    <Name>Jim Craven</Name>
    <Email>        Jim.Craven@nrcan-rncan.gc.
    ca</Email>
    <Org>Geological Survey of Canada</Org>
```

```
    <OrgUrl></OrgUrl>
  </Creator>
  <Submitter>
   <Name>Responsible Person</Name>
   <Email>person@responsible.institution.
   edu</Email>
   <Org>Responsible Institution</Org>
   <OrgUrl>http://responsibleinst.edu
   </OrgUrl>
  </Submitter>
  <ProcessedBy>Jim Craven</ProcessedBy>
  <ProcessingSoftware>
   <Name></Name>
   <LastMod></LastMod>
   <Author>WesternGeco</Author>
  </ProcessingSoftware>
  <!- tells the conversion codes how to parse the
EDI and whether to write the INFO block to XM ->
  <DateFormat>MM/DD/YY</DateFormat>
  <ParseEDIInfo>1</ParseEDIInfo>
  <WriteEDIInfo>0</WriteEDIInfo>
  <!- set some optional defaults ->
  <DefaultSiteName>Geographic Location, Can-
  ada</DefaultSiteName>
  <DefaultDataQuality>5</DefaultDataQuality>
  <DataQualityComment>Default  data  quality
  assigned   to   the   survey   at   archiving
  </DataQualityComment>
  <!- this option is for those who only wish to
  upload the metadata ->
  <MetadataOnly>0</MetadataOnly>
  <!- if these extensions are present, the corre-
  sponding files are also submitted to SPUD and
  displayed ->
  <Image>jpg</Image>
  <Original>edi</Original>
</Configuration>
```

Listing 2: "A documented sample config.xml needed to convert Z-file to EMTF XML"

```
<Configuration>
  <!- optional; used to indicate that time series
  are also archived at IRIS DMC; the network
  should match ->
  <TimeSeriesArchived>1</TimeSeriesArchived>
   <Network>EM</Network>
  <!-Project/Survey combination used to allo-
  cate survey DOI; Project should contain no
  spaces ->
  <Project>YSRP</Project >
  <Survey>Yellowstone-Snake    River    Plain
  </Survey>
  <!- Project.SiteID.YearCollected used to cre-
  ate a unique data ID in SPUD ->
  <YearCollected>2004</YearCollected>
  <Country>USA</Country>
  <!- required, comma-separated list of data
  types. See list of possible types ->
  <Tags>impedance,tipper</Tags>
```

```
  <!- required, except for DOI which should be
  omitted if it does not yet exist ->
   <Citation>
    <Title>Deep  Magnetotelluric  Sounding
    along the Yellowstone-Snake River hotspot
    track</Title>
    <Authors>Catherine deGroot-Hedlin, Ste-
    ven Constable, Karen Weitemeyer</Authors>
    <Year>2003-2004</Year>
    <DOI></DOI>
   </Citation>
  <!- include some relevant text ->
  <THANKS>acknowledgements.txt</THANKS>
  <PAPERS>references.txt</PAPERS >
  <README>readme.txt</README>
  <!- supported options: Unrestricted Release/
  Academic Use Only/Conditions Apply ->
  <ReleaseStatus>Unrestricted        Release
  </ReleaseStatus>
  <!- all provenance and processing information
  is optional but useful ->
  <AcquiredBy>UCSD/Catherine  deGroot-Hedlin
  </AcquiredBy>
  <Creator>
   <Name>Gary Egbert and Anna Kelbert</Name>
   <Email>egbert@coas.oregonstate.
   edu</Email>
   <Org>Oregon State University</Org>
   <OrgUrl>http://oregonstate.edu</ OrgUrl>
  </Creator>
  <Submitter>
    <Name>Responsible Person</Name>
    <Email>person@responsible.institution.
    edu</Email>
    <Org>Responsible Institution</Org>
    <OrgUrl>http://responsibleinst.edu
    </OrgUrl>
  </Submitter>
  <ProcessedBy>Gary  Egbert  and  Anna  Kelbert
  </ProcessedBy>
<ProcessingSoftware>
    <Name>EMTF</Name>
    <LastMod>2008-06-28</LastMod>
    <Author>Gary Egbert</Author>
  </ProcessingSoftware>
  <!- if these extensions are present, the
  corresponding files are also submitted and
  displayed ->
  <Image>png</Image>
  <Original>zrr</Original>
  <!- optional lists that allow to provide much
  additional metadata ->
  <RunList>Runs.xml</RunList>
  <SiteList>Sites.xml</SiteList>
  <ChannelList>Channels.xml</ChannelList>
  <!- this option is for those who only wish to up-
  load the metadata ->
  <MetadataOnly>0</MetadataOnly>
  </Configuration>
```

Note that you could also use the configuration file to specify one or all XML lists with additional metadata: Sites.xml, Runs.xml, and Channels.xml. The lists, if specified, should contain experiment metadata information about the runs and the sites (see also module `read_lists.f90` of the open-source codes).

The program looks for the configuration file and the optional XML lists in the same directory where the input file is located. If the lists are not found, the programs run without the additional information.

APPENDIX B

EXAMPLE EMTF XML FILE (METADATA ONLY)

Listing 3: "Sample metadata-only XML definition"

```xml
<?xml version="1.0" encoding="UTF-8"?>
<EM_TF>
  <Description>Magnetotelluric Transfer Func-
tions</Description>
  <ProductId>GrandProject.ID001.2006
</ProductId>
  <SubType>MT_TF</SubType>
  <Tags>impedance,tipper</Tags>
  <Provenance>
    <CreateTime>2011-09-28T17:35:59
    </CreateTime>
    <CreatingApplication>Harry's Ancient Con-
verter Program 1.0</CreatingApplication>
    <Creator>
      <Name>Harry   Author   and   Ronald   Post-
doc</Name>
      <Email> author@great. college.edu</Email>
      <Org>Great College</Org>
      <OrgUrl>http://greatcollege.edu</OrgUrl>
    </Creator>
    <Submitter>
      <Name>Responsible Person</Name>
      <Email>person@responsible.institution.
edu</Email>
      <Org>Responsible Institution</Org>
      <OrgUrl>http://responsibleinst.edu
      </OrgUrl>
    </Submitter>
  </Provenance>
  <Copyright>
    <Citation>
      <Title>Very  Interesting  Subduction  Zone
Magnetotelluric Transfer Functions</Title>
      <Authors>Harry Author, Ronald Postdoc and
Hermione Colleague</Authors>
      <Year>2006</Year>
      <DOI>10.1111/sample-doi</DOI>
    </Citation>
    <ReleaseStatus>Academic       Use       Only
</ReleaseStatus>
    <ConditionsOfUse>
```

All data and metadata in this survey are available free of charge for academic use only. If some or all of the data content is missing from this file, academic users may contact the submitter for data access. Commercial users should contact the author(s) of this data set for permission and conditions of use.

Although the author(s) strive to provide data and metadata of the best possible quality, neither the author(s) of this data set nor IRIS make any claims, promises, or guarantees about the accuracy, completeness, or adequacy of this information and expressly disclaim liability for errors and omissions in the contents of this file.

Guidelines about the quality or limitations of the data and metadata, as obtained from the author(s), are included for informational purposes only.

```xml
    </ConditionsOfUse>
  </Copyright>
  <Site>
    <Project>GrandProject</Project>
    <Survey>Very  Interesting  Subduction  Zone
</Survey>
    <YearCollected>2006</YearCollected>
    <Id>ID001</Id>
    <Name>Hoppin Springs, OR, USA</Name>
    <Location datum="WGS84">
      <Latitude>42.085064</Latitude>
      <Longitude>−117.552100</Longitude>
      <Elevation       units="meters">1978.750
      </Elevation>
      <Declination       epoch="1995.0">15.300
      </Declination>
    </Location>
    <Orientation     angle_to_geographic_nor-
th="0.000">orthogonal</Orientation>
    <AcquiredBy>Geophysics  Contractors,  Inc.
</AcquiredBy>
    <Start>2006-10-13T22:50:32</Start>
    <End>2006-10-31T17:47:39</End>
    <DataQualityNotes>
      <Rating>5</Rating>
      <GoodFromPeriod>8.000</GoodFromPeriod>
      <GoodToPeriod>20000.000</GoodToPeriod>
      <Comments author="Harry Author">great TF
from 10 to 10,000 s (or longer)</Comments>
    </DataQualityNotes>
    <DataQualityWarnings>
      <Flag>1</Flag>
      <Comments author="Harry Author">cultural
noise around 60 s</Comments>
    </DataQualityWarnings>
  </Site>
  <ProcessingInfo>
    <SignConvention>exp(+ i\omega t)</SignCon-
vention>
    <ProcessedBy>Ronald Postdoc</ProcessedBy>
    <ProcessingSoftware>
      <Name>EMTF</Name>
      <LastMod>2011-04-22</LastMod>
      <Author>Gary Egbert</Author>
    </ProcessingSoftware>
  </ProcessingInfo>
</EM_TF>
```

APPENDIX C

PRACTICAL DATA ROTATION AND FORMAT CONVERSIONS WITH EMTF FCU

Supplemental data provided in the EXAMPLES/ directory contain a set of practical usage examples for EMTF FCU v4.0. Here, we provide a brief reference to the specific problems addressed by these cases.

Each example contains a run script (run.sh) and an output file (output.*). All of the supporting files are also provided. All of these examples use real, unmodified data files. To run them, please ensure that the EMTF FCU conversion programs are compiled and in your system path.

Example 1: edi2edi (spectra to impedance)

Convert EDI SPECTRA to EDI impedance, with an arbitrary output orientation; this example uses geographic north (0.0) for output. If the output orientation is not specified, it is left unchanged.

This basic functionality requires no XML libraries; config.xml is not read.

Note that as is typical for many old EDI files, the original file does not contain the true geographic latitude and longitude. These are supplied separately in coords.txt. Instead, the original EDI file contains a reference coordinate and a distance, in meters, from the reference coordinate supplied for each of the channel sensors. The geographic site location can be reverse-engineered from this information under certain assumptions on the geographic projection used by EDI. However, edi2edi does not (by design) modify critical metadata.

Instead, to achieve the desired effect of geographic latitude and longitude computation, use the combination edi2xml → xml2edi in conjunction with config.xml that sets the key <ComputeSite-Coords> to 1.

Example 2: edi2xml (ambiguous orientation)

Convert EDI SPECTRA to XML, while rotating to geographic north. Here, I have selected one of the many examples where the site orientation information in the EDI file is found in numerous places and is conflicting.

For this reason, EMTF FCU conversion codes take extreme care when interpreting site orientation in EDI files. As is seen in this example, several options to that effect can be specified in config.xml.

Example 3: edi2xml (archive metadata only)

Convert EDI to an XML file, while withholding the data content. The output contains metadata, only. This option is useful for simplified sharing of metadata with colleagues even when the TFs are not ready for sharing.

This option can be exercised by setting the key <MetadataOnly> in config.xml to 1.

Example 4: edi2xml (no geographic location)

Convert EDI SPECTRA to XML, with an arbitrary output orientation; this example uses geographic north (0.0) for output. If the output orientation is not specified, it is left unchanged.

Similar to Example 1, this is yet another specimen of EDI that contains no site location in geographic coordinates. To estimate geographic coordinates on the fly, set the key <ComputeSite-Coords> in config.xml to 1.

This example also illustrates the behavior of conversion codes on occasions when certain missing values are encountered in the EDI.

Example 5: edi2xml (no orientation in the file)

Convert impedance EDI to XML. In this example, the orientation of the two magnetic field channels is specified as geographic. No orientation is supplied for electric field channels, and no ZROT rotations block (or other orientation indicators) is supplied with the impedances. In this example, HZ is also missing.

We deal with this situation by assuming that the E-fields are oriented the same as the H-fields and that no additional rotation is applied to the TFs. Note that in this case, one of the remote H channels is interpreted as an output channel in the XML. This is due to the ambiguity with which the channels are encoded in the EDI files. It can be fixed by changing the minimum number of output channels in the code (currently 3). It is a minor inaccuracy that I am leaving alone for now for the sake of code generality.

Example 6: edi2xml (principal axis rotation)

The old principal axis rotation technique was used to rotate the TFs frequency by frequency to make them more suitable for 1D interpretation. This method is not based on physics; principal axis rotated TFs cannot be used with modern techniques, unless this rotation is undone.

This idea of rotating frequency by frequency to minimize diagonal component amplitudes is an old (bad) idea that was often used when people had minimal modeling capabilities and limited understanding of MT. Too often, they then threw out the diagonal components, so we would not be able to recover the actual impedance.

EMTF FCU v4.0 codes are capable of undoing principal axis rotation, assuming that the diagonal components are still present in the file, so that the output TFs in the XML are oriented to the chosen direction (in this case, geographic north).

This example also contains multiple derived data types. Note that rotation of derived data types is presently not supported by the conversion codes, so that these values will need to be recomputed after rotation. It is therefore recommended that you omit reading these derived data types from files that require a coordinate rotation. The code writes out a warning to that effect. Here, we write them out, unrotated. This is fine for illustration only.

Example 7: edi2xml (spectra remote reference)

Simple EDI SPECTRA remote reference to XML conversion example; no rotation is needed. Note that the full error covariances are recorded in the XML, so that arbitrary rotations will be possible in the future.

If no rotation is specified, the <Orientation> tag in the output XML is recorded as oriented to "sitelayout." However, if the rotation is specified in the EDI file, <Orientation> defaults to "orthogonal." That is the case here.

It's best to always ensure that the <Orientation> is set to "orthogonal" in the XML, unless the TFs are truly oriented to a nonorthogonal site layout.

Example 8: edi2xml (spectra single station)

Simple EDI SPECTRA single station to XML conversion example, with rotation to geographic north. Note that the full error covariances are recorded in the XML, so that arbitrary rotations will be possible in the future.

Example 9: xml2edi (default file conversion)

Simple XML to EDI conversion; no rotation involved. This is precisely the conversion that runs behind the scenes in the EMTF XML database (IRIS EMTF, 2018) when an EDI download is requested.

Example 10: xml2z (revert to site layout)

Reverse conversion from XML to Z-file, with full recovery of the original site layout orientation. In this case, the original site orientation was complex; not only did the electric field point in different

directions than the magnetic fields, but also the Ex also exhibited reversed polarity, resulting in a −90 degree angle between the two, instead of the usual +90. All of these complications were appropriately corrected for by rotation to geographic north.

Note that some processing metadata (decimation level, frequency bands, number of data points, and sampling frequency) is lost in this conversion, relative to the original Z-file, which may be found in Example 11. In addition, the residual covariance between data types is lost, in this case, between impedances and vertical field TFs (aka tippers). Other than that, the conversion is quite accurate.

Example 11: z2edi (rotate to geographic)

EMTF FCU v4.0 can also be used for conversion between Z-file and EDI formats, no XMLs involved, with or without TF rotation. This example rotates to geographic coordinates. XML library is not required to run z2edi (also, z2z and edi2edi).

Example 12: z2xml (full metadata archiving)

A complete archiving example, with full metadata, was applied for the USArray archiving project. There, additional field metadata are included using the XML "lists": Sites.xml, Runs.xml, and Channels. xml. The EMTF Z-file is converted to XML with rotation to geographic north, and additional metadata are extracted from the lists.

REFERENCES

Becken, M., O. Ritter, and H. Burkhardt, 2008, Mode separation of magnetotelluric responses in three-dimensional environments: Geophysical Journal International, **172**, 67–86, doi: 10.1111/j.1365-246X.2007.03612.x.

Booker, J. R., A. Favetto, and M. C. Pomposiello, 2004, Low electrical resistivity associated with plunging of the Nazca flat slab beneath Argentina: Nature, **429**, 399–403, doi: 10.1038/nature02565.

Chave, A. D., D. J. Thomson, and M. E. Ander, 1987, On the robust estimation of power spectra, coherences, and transfer functions: Journal of Geophysical Research, **92**, 633, doi: 10.1029/JB092iB01p00633.

DOM, 2000, Document object model (DOM) level 2 core specification, https://www.w3.org/TR/DOM-Level-2-Core/, accessed 17 October 2019.

Egbert, G. D., and J. R. Booker, 1986, Robust estimation of geomagnetic transfer functions: Geophysical Journal of the Royal Astronomical Society, **87**, 173–194, doi: 10.1111/j.1365-246X.1986.tb04552.x.

Egbert, G. D., and A. Kelbert, 2012, Computational recipes for electromagnetic inverse problems: Geophysical Journal International, **189**, 251–267, doi: 10.1111/j.1365-246X.2011.05347.x.

Eisel, M., and G. D. Egbert, 2001, On the stability of magnetotelluric transfer function estimates and the reliability of their variances: Geophysical Journal International, **144**, 65–82, doi: 10.1046/j.1365-246x.2001.00292.x.

EPOS, 2018, EPOS data policy July 2018 web version: Technical report, European Plate Observing System, https://www.epos-ip.org/documents/.

Evans, R. L., P. Tarits, A. D. Chave, A. White, G. Heinson, J. H. Filloux, H. Toh, N. Seama, H. Utada, J. R. Booker, and M. J. Unsworth, 1999, Asymmetric electrical structure in the mantle beneath the East Pacific Rise at 17 degrees S: Science, **286**, 752–756, doi: 10.1126/science.286.5440.752.

Hobbs, B. A., 1992, Terminology and symbols for use in studies of electromagnetic induction in the earth: Surveys in Geophysics, **13**, 489–515, doi: 10.1007/BF01903487.

IRIS EMTF, 2018, EM transfer function product query, http://ds.iris.edu/spud/emtf, accessed 17 October 2019.

IRIS EMTF FCU, 2018, EMTF FCU, https://seiscode.iris.washington.edu/projects/emtf-fcu, accessed 17 October 2019.

Jones, A. G., B. Olafsdottir, and J. Tiikkainen, 1983, Geomagnetic induction studies in Scandinavia - III. Magnetotelluric observations: Journal of Geophysics, **54**, 35–50.

Kelbert, A., G. D. Egbert, and C. DeGroot-Hedlin, 2012, Crust and upper mantle electrical conductivity beneath the Yellowstone Hotspot Track: Geology, **40**, 447–450, doi: 10.1130/G32655.1.

Kelbert, A., G. D. Egbert, and A. Schultz, 2011, IRIS DMC Data Services products: EMTF, the magnetotelluric transfer functions, doi: 10.17611/DP/EMTF.1.

Kelbert, A., S. Erofeeva, C. Trabant, R. Karstens, and M. Van Fossen, 2018a, Taking magnetotelluric data out of the drawer: EOS, Transactions, American Geophysical Union, **99**, doi: 10.1029/2018EO112859.

Kelbert, A., N. Meqbel, G. D. Egbert, and K. Tandon, 2014, ModEM: A modular system for inversion of electromagnetic geophysical data: Computers & Geosciences, **66**, 40–53, doi: 10.1016/j.cageo.2014.01.010.

Kelbert, A., M. Y. Smirnov, O. Ritter, J. Duan, G. Heinson, and N. Rees, 2018, International collaboration for magnetotelluric data and model sharing and interoperability: Presented at the Fall Meeting, American Geophysical Union.

Krieger, L., and J. R. Peacock, 2014, MTpy: A Python toolbox for magnetotellurics: Computers & Geosciences, **72**, 167–175, doi: 10.1016/j.cageo.2014.07.013.

Larsen, J. C., R. L. Mackie, A. Manzella, A. Fiordelisi, and S. Rieven, 1996, Robust smooth magnetotelluric transfer functions: Geophysical Journal International, **124**, 801–819, doi: 10.1111/j.1365-246X.1996.tb05639.x.

Meqbel, N. M., G. D. Egbert, P. E. Wannamaker, A. Kelbert, and A. Schultz, 2014, Deep electrical resistivity structure of the northwestern U.S. derived from 3-D inversion of USArray magnetotelluric data: Earth and Planetary Science Letters, **402**, 290–304, doi: 10.1016/j.epsl.2013.12.026.

Murphy, B. S., and G. D. Egbert, 2017, Electrical conductivity structure of southeastern North America: Implications for lithospheric architecture and Appalachian topographic rejuvenation: Earth and Planetary Science Letters, **462**, 66–75, doi: 10.1016/j.epsl.2017.01.009.

Siripunvaraporn, W., G. D. Egbert, Y. Lenbury, and M. Uyeshima, 2005, Three-dimensional magnetotelluric inversion: Data-space method: Physics of the Earth and Planetary Interiors, **150**, 3–14, doi: 10.1016/j.pepi.2004.08.023.

Soyer, W., and M. J. Unsworth, 2006, Deep electrical structure of the northern Cascadia (British Columbia, Canada) subduction zone: Implications for the distribution of fluids: Geology, **34**, 53–56, doi: 10.1130/G21951.1.

Stodt, J., 1983, Bias removal for conventional magnetotelluric data: Technical report, U.S. Department of Energy, OSTI, https://www.osti.gov/servlets/purl/7145868-XTQItq/, accessed 27 September 2018.

Trabant, C., A. R. Hutko, M. Bahavar, R. Karstens, T. Ahern, and R. Aster, 2012, Data products at the IRIS DMC: Stepping stones for research and other applications: Seismological Research Letters, **83**, 846–854, doi: 10.1785/0220120032.

Unsworth, M. J., A. G. Jones, W. Wei, G. Marquis, S. G. Gokarn, J. E. Spratt, P. Bedrosian, J. R. Booker, C. Leshou, G. Clarke, L. Shenghui, L. Chanhong, D. Ming, J. Sheng, K. Solon, T. Handong, J. Ledo, and B. Roberts, 2005, Crustal rheology of the Himalaya and southern Tibet inferred from magnetotelluric data: Nature, **438**, 78–81, doi: 10.1038/nature04154.

Unsworth, M. J., P. E. Malin, G. D. Egbert, and J. R. Booker, 1997, Internal structure of the San Andreas Fault at Parkfield, California: Geology, **25**, 359, doi: 10.1130/0091-7613(1997)025<0359:ISOTSA>2.3.CO;2.

Wannamaker, P. E., T. G. Caldwell, G. R. Jiracek, V. Maris, G. J. Hill, Y. Ogawa, H. M. Bibby, S. L. Bennie, and W. Heise, 2009, Fluid and deformation regime of an advancing subduction system at Marlborough, New Zealand: Nature, **460**, 733–736, doi: 10.1038/nature08204.

Wei, W., M. J. Unsworth, A. G. Jones, J. R. Booker, H. Tan, D. Nelson, L. Chen, S. Li, K. Solon, P. A. Bedrosian, S. Jin, M. Deng, J. Ledo, D. Kay, and B. Roberts, 2001, Detection of widespread fluids in the Tibetan crust by magnetotelluric studies: Science, **292**, 716–719, doi: 10.1126/science.1010580.

Wight, D. E., 1988, SEG standard For MT and EMAP data: 58th Annual International Meeting, SEG, Expanded Abstracts, 249–251, doi: 10.1190/1.1892244.