**Recommendations for seismological data processing and management at AFAD**

**Main recommendations**

* Improve software system to enable user friendly output and more analysis tools
* Make sure all software is documented and independent of individuals
* Make the strong motion real time data available for weak motion network.
* Collect all data in one data base
* Recover all old data and put into system
* Invest more on capacity building (education, training) including international degrees

**Introduction**

This document is based on visits to AFAD by Lars Ottemöller in November 2012 and Jens Havskov in January. Both are part to a joint project with Erzurum (Ataturk) University and AFAD of using the AFAD database to develop magnitude scales for Turkey to be used by AFAD.

AFAD has a team of seismic analysts that process earthquakes in Turkey on a 24/7 basis. The seismic network comprises more than 200 real-time broadband seismic stations and more than 400 strong motion stations. It is no small undertaking to keep such a large network operating on a continuous basis and AFAD seems to be doing a good job in operating the network and doing basic processing.

The quality of the data in terms of completeness and noise levels was not looked at by us.

**Weak motion network**

The weak motion data are acquired from the field stations using Güralp’s Scream software which copies the continuous data into a windows based flat file structure of 15 minute data files. From there the data is processed manually. In addition the data is automatically processed using the SeisComp3 software.

Processing:

AFAD presently uses their in house developed data processing system called EA for the weak motion (seismometer) data. The software works with the continuous data that are divided into 15 minute long Guralp gcf –files (a MiniSeed reader has been developed for the next version of EA). For one 15 minute time window or longer, the analyst can create event entries in the database system and do the data processing. The continuous data is inspected manually for detections using a filter 1-10 Hz. All channels are checked. Plotting can be done sorting stations in different ways, e.g. EW, to better identify nearby stations with manual detection. When an event has been found, it is located and, if location is considered ok, also stored in the SQL data base together with relevant parameters. At the same time, the event is published on the web page, emails are sent out to specific agencies and an SMS is sent to interested parties. In addition, when detection on SeisComp3 is announced, the corresponding event is located immediately.

There is no waveform file created corresponding to the event, so any processing must be based on the continuous data. EA has online access to all continuous data since 2006. From 1997 to 2006, the waveform data is stored in SUDS and SEISAN format as event files (some continuous data is available) but it cannot be used with the current processing system.

The EA software is very user friendly and well adapted to working with continuous data in the current structure. It is simple to recall events (if the time is known) within a continuous segment and it is very fast to plot, once the data has been read in. Thus for detecting and locating events from a continuous data stream, it is difficult to make it much better.

The processing is presently limited to picking of P and S phases, and reading of amplitudes for magnitude calculated from the P waves (which is non-standard, a new version of EA will also be able to read standard amplitudes for Ml, mb, Ms and MS). From these data, hypocenter and magnitude are computed using Hypoinverse. At first all stations are read. However if large residuals are encountered with particular stations, they are not used for location.

This processing appears efficient, but has limitations in what can be done. Within the EA software it is not possible to extract event listings, which makes it difficult to work with event data retrospectively. In other words, putting data into the database with some limited processing works well, but extracting events or waveform data is not easily possible. Searching for earthquakes in the database can be done through a web interface but only gives parameters for events and no waveform data.

The current magnitude calculated immediately is the non standard Ml. Correct Mw from moment tensor inversion is calculated independently of EA after a few days since the data has to be taken to a different system and reprocessed. More standard magnitudes will probably be calculated by EA in the future.

So while EA is very good for initial processing, it cannot be used for further event analysis and it has no output of integrated data sets for other programs.

**The data base**

The EA SQL data base under Windows contains hypocenters and magnitudes (so far a nonstandard Ml) for local events and corresponding observations. Fault plane solutions made outside EA are also put into the data base, as well as the fault plane solutions from moment tensor solutions. All other standard parameters are missing: Teleseismic locations and magnitudes, result from spectral analysis, elements of moment tensor and felt (macroseismic) information. However the data base contains all instrumental information and is very useful for maintaining the network parameters and it seems to be well maintained.

On average 100 events are located per day and the majority have less than 20 stations used for the location. The number of fault plane solutions and moment tensor solutions per month is at least 10 and 10 respectively. The analysis workload is therefore reasonable on normal days. For teleseismic earthquakes, only P is read on some stations and no location or magnitude is made. Since the events are not marked as teleseismic events, they are difficult to find again in the data base.

The continuous data is stored in original raw file formatted gcf files. These files cannot directly be used with other programs since the channel information is not correct in all cases. Working with EA presents no problem since the correct channel information is stored in the SQL data base.

The in house general user has no access to the data outside the closed processing computer network. Waveform data can here be copied out in 15 min files but the readings must be taken from the web page in a comma separated format. An output in SEISAN format is also being made.

**Personnel**

There are 14 people running the network. Thirteen have BSc or MSc degrees and one person is a technician.

Four people are on permanent night and weekend shifts, one at a time, and do not participate in other activities. During this time they locate all events in their period.

One person (Tugbay) is leader of the whole group.

One person (Kenan) is programming the EA system and does general IT work.

Eight persons share the processing in the day time on week days, each working in 2 hour slots which leaves time for other tasks:

One person: Reports in case of larger earthquakes (in both Turkish and English), making a historical catalog and working on GZF deep borehole instrumentation project.

One person: Fault plane solutions with polarities and Focmec and moment tensor solutions with Isola. Making a historical catalog and working on GZF deep borehole instrumentation project.

One person: Moment tensor solution using (Swift), maintaining catalog and responsible for field operation (main job).

One person: Moment tensor inversion (Swift), also responsible for checking all routine locations. Also works with catalog maintenance.

One person: Main responsible field operation and processing.

Technician: Field work and logistics, minor electronic repairs.

3 newly recruited persons: Only doing locations and initial training.

Field work is shared within the group and takes about one complete position.

The work seems to be well organized and there seems to be good working relations among the personnel. It is good that the whole group is getting experience with both routine locations and field work, however it also to time away from more skilled work. There is some lack of seismological knowledge and speaking English is a problem for some.

**Strong motion network**

The strong motion network has about 300 Guralp stations with internet connection and about 100 older stations without internet connection. Event files (triggered files, trigger level 1 gal) are collected from the field using Scream. Of the internet connected stations, 50 sends real time data to a local Scream system. The remaining stations trigger locally and data has to be collected from the sites. In addition, 75 new GeoSig stations are now being installed. These stations (will be able to) transmit data in real time with SeedLink.

The strong motion data from all events are stored in a public data base, all in the same ASCII format. There are a total of 7800 events available since 1976 and currently about 25 events are recorded per month.

Events from the strong motion network from 2011 and 2012 were checked in SEISAN. It seems that many events are only recorded with one station which seems strange considering the density of the network.

The strong motion stations are generally not used for location although the nearest ones significantly could improve the epicenter location and depth.

**Based on the observations made during the two visits, we make the following comments and recommendations:**

* The main content of the AFAD data base should in the future be hypocenter locations of both local and teleseismic earthquakes, local magnitude Ml, broad band surface magnitude MS and moment magnitude Mw, spectral parameters, moment tensor solutions for larger events, fault plane solutions and information about felt events.
* It is difficult to extract complete sets of event, waveform and instrumentation data for the general user. This is a serious problem as it means that the data are not easily available for research work. An ongoing project to build a national data set may overcome this limitation. The data should be stored in a way that is easily searchable for all parameters with output of an integrated data set in international format directly to the user.
* The EA software has limitations both in terms of processing functionality and access of data. The software is written in Visual Basic (platform dependent system) which is not an obvious choice for developing software of this type for a network of more than 200 seismic stations. However, on the positive side, EA has many nice features, is very fast for initial processing, makes use of a relational database and it is an impressive one man product. The EA software is very useful software for maintaining network parameters. Considering that it is a one man product, it should be documented and a manual be written. We recommend to continue to use EA for initial processing provided standard magnitudes are implemented.
* We recommend installing a second interactive processing system to work on the database as the EA is lacking many advanced features. Options are software packages such as SEISAN, and Geotool. During out visits we have installed SEISAN and put all available data into the SEISAN data base for a limited time period. This was needed for the magnitude project, partly to do the analysis and partly to check the new magnitude facilities in EA. SEISAN can contain all the parameters mentioned above, calculate all parameters and store them. While SEISAN is not as efficient as EA for the initial analysis, it might bridge the gap between the EA data base and provide an easy way to access the data for research as well as providing more integrated analysis tools. It would also be possible to use it as a secondary processing system in case of a problem with EA. Whatever storage and processing system AFAD chooses in the future, it should at least have the capability of the SEISAN system, so SEISAN is a good start.
* Event detection is done manually. It is a reasonably good detector to have a person looking at all the data. However it is also unnecessary and a waste of time considering modern trigger systems. By providing the analyst with an automatic detection, manual processing can be done much faster since the analyst only have to look at the channels with detections at the correct time. In addition, an immediate notification is given when there is an event. SeisComP3 is doing this job and in addition provide automatic location and magnitude, however SeisComp3 is not so well suited for smaller events. It is recommend to install an automatic trigger system like EarthWorm which could send trigger times directly to EA or another processing system. Apparently a new commercial SeisComp has better detection capability and could be an alternative.
* Process teleseismic earthquakes since they are important for international data exchange as well as important for local crustal studies. In addition, calculating magnitudes of teleseismic earthquakes provide a good check on the network calibration.
* It is recommended to recover all old data like waveform data, readings and hypocenters to complete the data base. A test with both old SUDS and SEISAN files showed that they were readable by SEISAN. This also includes the 8000 events from the strong motion data base. As a test, all the events (>2000) from 2011-12 were put in.
* The waveform data should be stored in standard MiniSeed format (used by most institutions) and the data should contain correct station and channel information. We recommend to use the standard SeisComp data structure of day files in MiniSeed. This is already available in the current SeisComp system. Initial tests made during the visit indicate that reading speed would be the same or faster than with the current system.
* The complete EA system in terms of software and descriptive data is maintained by one person. This is not ideal for an institution such as AFAD. It is strongly recommended to involve additional qualified people to develop software and maintain the database of metadata. Not all software has to be developed from scratch, there is much public domain software available. If AFAD is to develop their own tools it is recommended to do this based on the SeisComp data structure for continuous data and to be written in C++ or other high level languages. In any case, writing all needed modules from scratch is several man-years of work. It is also recommended to use a Linux server for future storage and processing since Linux is much better suited for multiple users than Windows. We do not know of any institutions of the size of AFAD that use Windows.
* Comparing the location of the strong motions stations with the weak motion stations it is seen there is little overlap. The real time data from the strong motion network should therefore be added to the flow into the weak motion network. This would increase the station density significantly and would provide significantly improved locations, better depths and magnitudes for all important events. This is more important than adding more weak motion stations. Considering that the strong motion network already have 80 stations into a local Scream server and it is a question of less than one hour work to put the channels into the weak motion network, it seems incredible that it has not already been done!
* It is recommend to register the strong motion station names at ISC.
* It seems that the strong motion network records too few events. Maybe the trigger level should be lowered. It is recommended to check the continuous data from the strong motion stations to investigate if there is more data. This would be checked automatically if the strong motion signals were integrated with the seismic network.
* Considering the few recordings for some events in the strong motion network, it is recommended to use the data from the broad band stations. The velocity record can easily be converted to accelerations. This would significantly increase the strong motion coverage. The conversion was tested on a station having both accelerometer and seismometer and the acceleration record made from the seismic record was practically identical to the acceleration record from the accelerometer.
* The new GeoSig strong motion network should, using SeedLink, also be connected directly to the seismic network.
* The data quality in terms of completeness and noise levels should be investigated and regularly checked.
* We recommend that AFAD regularly (perhaps every 2 years) invites a small group of international experts in the field of seismic instrumentation, networking and processing to evaluate the operation to help AFAD to improve their operation.
* The resources spent on instrumentation and operation is disproportionate compared to the resources used for advanced analysis and personnel training. Considering the extremely large and valuable data base, AFAD should have had several internationally trained seismologists at MSc and PhD level. Considering that AFAD has several talented young people, we suggest allocating funds for sending the best staff abroad to get seismology degrees. If that is not possible, send people out to do specific research projects resulting in international publications. It is also possible to take courses abroad (with exam, just going to a training course might not be very useful). Lack of English skills is a serious limitation for some people so it might be worth investing in intensive English training.

These recommendations are prepared considering the importance of the earthquake monitoring service provided by AFAD. We see limitations in the practical implementation of this service related to data storage and processing. However, AFAD is well operated, has sufficient funds, a good network and good data, a group of talented young people. However, they do spend a large part of their time in routine work leaving too little time for research and hiring young people dedicated field work is advisable. The potential for improvement is very good.