[*Ruano et al.*, 2014] talks about building MLP and SVM method to classify seismic signals. They use sliding window to extract features from one station, and train a MLP and SVM method to classify different signals. For the SVM, they use an active learning method to retrain on the mis-classified results, and achieved better results. Besides, they also train a model with shorter time window which can be applied to EEW. Overall, it is a paper overlays the correct procedure to do signal classification. But a few things not correctly or clearly stated, for example, for the imbalanced dataset, how they deal with it.

[*Shirzaei et al.*, 2016] reports the surface uplift due to the injection in eastern Texas. Using time evolution InSAR images, they get a uplift of 3 mm/year over >8km area from the injection wells. They also build a poroelastic model to explain the uplift. They invested two wells, one in the west, and one in the east, and only the east shows the uplift. The west wells do not show the uplift, but accompanied by a sequence of earthquakes. They attribute this to the low compressibility of the rocks at the west wells. Therefore, the seismicity and the deformation behavior depends both on the injection activity and the local hydrogeological properties. Some interesting things from this paper: (1) Seismic activity increased even while the injection rates declined, owing to diffusion of pore pressure from earlier periods with higher injection rates. (2) Induced seismicity potential is suppressed where tight confining formations prevent pore pressure from propagating into crystalline basement rocks. (3) Over time, the increased pore pressure due to injection can spread to distances of many kilometers. (4) A localized increase in pore pressure shifts the circle (Mohr circle) to the left and changes its radius because of poroelastic strain, whereas a spatially uniform pore pressure increase only shifts the circle to the left until it touches the failure envelope.

[*Levander et al.*, 2011] proposed the hypothesis for the uplift of the Colorado plateau – delamination style convective lithospheric downwelling. The main method they used is from body wave tomography and receiver function. Based on the results from these two methods, they propose the process like this: The Colorado plateau lithosphere has been dydrated, and the small increase in density from the freezing metls, and the viscosicty reduction from hydration and advected heat, destabilizes the lithosphere and initiates a localized downwelling. The re-fertilized Colorado plateau mantle has been removed, delaminating the lowermost crust with it. The asthenosphere is invading the region from the beneath the drip and around the peripheries of the drip. They inferred that the lowermost crust involved in the dwonwelling has been modified by intrusion of basaltic melts that froze to produce high density eclogites. A series of these events have been removing the lithosphere from the Colorado plateau peripheries since the Farallon slab was removed 20-30 Myr ago, and causing the uplift.

[*Taira et al.*, 2009] reports the changes of fault-strength on the San Andreas fault at Parkfield by remote triggering. They argue that the seismicity of the repeating earthquakes at Parkfield have revealed a means of monitoring fault strength. For the Landers and Sumatra earthquake, the dynamic strain causes the changes in two manifestations: temporal variations in the properties of seismic scatters – probably reflecting the stress-induced migration of fluids – and systematic temporal variations in the characteristics of repeating-earthquake sequences that are most consistent with changes in fault strength. They also found the 2004 M6 Parkfield earthquake causing the two remote triggering different maybe due to: it damaged the fault zone by creating new fractures, it relieved most of the stress stored in the fault zone, and the absence of accompanying slip for the 2004 Sumatra earthquake can be explained by the low driving stress.

[*Bayrakci et al.*, 2016] reports the fault-controlled hydration of the upper mantle during continental rifting. They shoed the serpentinization at the rifted continental margin offshore from western Spain was probably initiated when the whole crust cooled to become brittle and deformation was focused along large normal faults. They use seismic tomography to image the 3D distribution of serpentinization in the mantle and find that the local volume of serpentinite beneath thinned, brittle crust is related to the amount of displacement along each fault. This implies that sea water reaches the mantle only when the faults are active. They also estimate the fluid flux along the faults and find it is comparable to that inferred for mid-ocean ridge hydrothermal systems. They conclude that brittle processes in the crust may ultimately control the global flux of sea water into the Earth. Some useful backgrounds can be found [here](https://en.wikipedia.org/wiki/Non-volcanic_passive_margins).

[*Lenardic*, 2017] gives a very nice review about a paper published in nature geoscience by Van Avendonk et al. The hypothesis proposed by the paper is that the changes of the earth’s internal energy cooling rate leaves a trace on the sea floor thickness. Oceanic crust forms dominantly by decompression melting of mantle rocks below mid-ocean ridges, so raised mantle temperatures can result in thicker crust forming at the ridge. Therefore, by examining the sea floor thickness, they can estimate the variations in mantle temperature through time. They used seismic data gathered over the past 40 years to determine the thickness of oceanic crust across the globe and show that, on average, the oceanic crust has thinned. They also notice that the cooling rate below the Atlantic and Indian mid-ocean ridges is about three times higher than that beneath the Pacific. Immediately prior to this time, the Atlantic and Indian oceanic basins were sites above which the supercontinent Pangaea resided. Therefore, they argue that the supercontinent has an insulating effect which the high temperature beneath it may cause the instability and initiates the break up of the supercontinent. Thus we see a transit signal in these places in the last 100 M years. The implications of such fluctuations go beyond internal Earth dynamics, it also link to the greenhouse gas released, since it has been argued that the greenhouse world that our planet experienced in the Cretaceous may be connected to a volcanic-tectonic forcing event associated with Pangaea’s break-up.

[*Mishra and Gordon*, 2016] talked about the debates between the rigid-plate and the shrinking-plate hypotheses, and try to provide more evidences to support the shrinking-plate hypothesis. They use model to predict the azimuth of the transform-fault distributed between 15 plate pairs, and test whether a significantly better fit to the data is obtained after correction for the predicted bias. The three key points they got from this paper are: (1) The shrinking-plate hypothesis predicts subtle differences in azimuths of right-lateral versus left-lateral transform faults; (2) Transform-fault azimuths observed globally indicate a statistically significant difference between right-lateral and left-lateral faults; (3) Transform faults do not precisely parallel plate motion, thus validating inferred quantifiable plate nonrigidity.

[*Zheng et al.*, 2014] reports a modified method to estimate absolute plate velocities from seismic anisotropy by correcting the correlated errors. There are two common methods to estimate the absolute velocity: trends of hot spot tracks, and seismic anisotropy from shear wave splitting. But the current seismic anisotropy method assumes the errors in the azimuths inferred from shear wave splitting are uncorrelated, which is they show find that residuals to azimuths from any one plate are strongly correlated with the residuals from the same plate. Therefore this paper proposed a new method for analyzing the azimuths inferred from shear wave splitting data. The key points of this paper are: Absolute plate velocities differ insignificantly from zero for eight plates and seismic anisotropy indicates plate motion direction if speed exceeds ~5 mm/a.

[*Niu et al.*, 2003] reports the evidence of crustal structural changes at Parkfield that detected from seismic data. They find a systematic temporal variation in the seismograms of repeat microearthquakes that occurred on the Parkfield segment of the San Andreas fault over the decade 1987 – 97. Their analysis reveals a change of the order of 10 m in the location of scatters which plausibly lie within the fault zone at a depth of ~3 km. The motion of the scatters is coincident, in space and time, with the onset of a well documented aseismic transient (deformation event). They speculate that this structural change is the result of a stress-induced redistribution of fluids in fluid-filled fractures caused by the transient event.

[*LeCun et al.*, 2015] gives a very nice review about Deep learning in nature. They first talked about the benefits of deep learning, the key advantage is that deep learning can learn good features automatically using a general-purpose learning procedure, thus avoid of feature engineering. Some key points from their reviews are:

1. Representation learning is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification.
2. Stochastic gradient descent called stochastic because each small set of examples gives a noisy estimate of the average gradient over all examples.
3. At present, the most popular non-linear function is the rectified linear unit, which typically learns much faster in networks with many layers compare with tanh and sigmoid.
4. The hidden layers can be seen as distorting the input in a non-linear way so that categories become linearly separable by the last layer.
5. In practice, poor local minima are rarely a problem with large networks.
6. The unsupervised ‘pre-training’ can reduce the overfitting problem when use small datasets, and it is needed when you have a small dataset
7. The first major application of this pre-training approach was in speech recognition, and it was made possible by the advent of fast graphics processing units (GPU).
8. By 2012, versions of deep net were already being deployed in Android phones.
9. The convolutional neural network (ConvNets) was much easier to train and generalized much better than networks with full connectivity between adjacent layers.
10. ConvNets are designed to process data that come in the form of multiple arrays. There are 4 key ideas that take advantage of the properties of natural signals: local connections, shared weights, pooling and the use of many layers.
11. New regularization method is to use the dropout.
12. Recurrent neural networks used for sequential data by unfolding the time into a very deep nets.
13. RNN is not good to store information for very long time, therefore, the proposal of long short-term memory networks is to solve this problem.

[*Long et al.*, 2014] gives a review of the USArray and the introduction of the papers in the special issue from this array. It summaries the exciting results from USArray mainly from the study of the structure. Therefore it provides a quick way to know what the papers in this issue about. But I am thinking the other type of study for USArray, maybe I should spend more time on the Machine Learning and see if I can identify some interesting projects.

[*Goodfellow et al.*, 2014] proposed the generative adversarial nets, a very interesting idea. It has two models that compete with each other, one generative model and one discriminative model, that being trained together. The generative model (or generator) is trying to produce fake data to fool the discriminative model, while the discriminative model is try to distinguish whether the input data is from the real world data or the generator. Competition in this game drives both teams to improve their methods until the generator generates data points so real that the discriminative model can not distinguish it. The argument is even we don’t know what exactly the generative model works inside, if it can generate realistic sample, we can think it has a very nice internal mechanism to model the world. You can see [*Focus*, 2017] for an example to generate realistic samples in the astronomy study. Personally, I think this is really a cool method that I can use to generate realistic data without build a parametric model!

[*Yoffie and Kim*, 2009] show the history of the E ink company. It is very interesting how a research project turn into a business. Something need me to think, for example, I should think what is my role in the value chain, since there are many ways to place your company at different position in value chain, but the higher end, the more money you can make. Also, at the beginning of the company, we should not work on multiple things, focus on one important thing, and overcome the challenges, then it is much easier to success. They scattered everywhere and later cut all the different projects. The main problem of them is they didn’t see clear what’s the business model.

[*Sahlman*, 2009] shows the history of the success of the SpinBrush. It is really cool to see how John Osher became success step by step. The focus on the development of the core business, know better of the market, use a small group of overqualified people etc. are all the factors to bring the business success. Also, it listed sixteen mistakes entrepreneurs do not have to make, which is really useful. For doing business in a mature market, the best thing is to have something new, and then collaborate with some giants. Patent the new developments is also very important factor we need consider to protect us from the giants.

[*Massari et al.*, 2017] reports the CSN network at the 16th world conference on earthquake. It describes the instrumentations of CSN on the JPL campus. This paper has a very nice discussion of the application of this dense data to the engineering community. Using the finite element model to compare with the observation and to explore some features of the building is another section in this paper. Also, it talks using the radon transform to find the reflection points of the waves due to impedance. The final part of the paper talks about the real-time platform that can be visually explored by the users.

[*Bletery et al.*, 2016] reports that the mega-earthquakes preferentially rupture flat (low-curvature) interfaces. They studied the curvature from the slab1.0, and overlay the rupture areas of the past large earthquakes. They found most of the rupture areas are within the places where the curvature is small, and small average dip angles. When they plot the magnitude and the average dip angle and curvature, they found a negative correlation, even though the variances are large. The build a simplified analytic model and demonstrates that heterogeneity in shear strength increases with curvature. Shear strength on flat megathrusts is more homogeneous, and hence more likely to be exceeded simultaneously over large areas, than on highly curved faults. This is a very interesting hypothesis, since it reduces the different factors into one factor – curvature. The high curvature will work as barriers for the rupture. But the other thing is how accurate is the slab1.0 model to study the curvature.

[*Aguiar and Myers*, 2016] talks about characterizing microseismicity at the Newberry volcano geothermal site using PageRank method they developed at Stanford. In this report, it mainly talks about the indirect links this method can identify, but fails to talk about the physical meaning of the indirect links. Besides, this report seems focus on the method itself instead of the results that characterize the microseismicity, maybe there will be another paper talk more about this.

[*NAEEM ZAFAR*, 2011] talks about the whole history of the company Veridicom, and some interesting things I learned are:

1. the fingerprinting sensor system is a very interesting system. It includes scan the image of the ridges and valleys of the finger using thousands of built-in capacitive sensors, the matching algorithm to match the pattern, and data protection software then erased the actual fingerprint image, but did store a set of characteristics unique to the fingerprint (that even if stolen could not be used) for future identification.
2. The different views of the founder and the technical lead is very interesting, the technical lead saw the founder as a way to get money for their projects, not necessarily because they wanted to build a company. This maybe not good for the company?
3. VC invest in technology, people, and markets order.
4. Timing of starting a specific company with the technology is really important, if you don’t have outside support, even your technology is really advanced, you can not success. Think about you are an smartphone app developer before there is an smartphone
5. “The number one lesson that I learned when you are starting a company with just the technology, is that you need to hire a marketing guy before you hire a sales guy”
6. The early fail of the company is due to they did not figure out “who exactly are the customers”, therefore, the transform from OEM manufacturer to a solution company that aiming for get BOA etc bank is a very wise move, and leading to some big customers first to advance the business.
7. It is a little crazy to see that the board fire the founder CEO, but I guess this is how the company evolve.
8. Identifying the existing competitors are important.

[*Kohler et al.*, 2013] reports the idea of using low-cost sensors to do structural health monitoring. In this paper, they showed that they can model the building using Timoshenko beam approach. In this case, knowing the first two eigenfrequencies of the building allows them to estimate the mode shapes of a bending, shearing, and rocking building. Then they use the first two modes and the traveling wave component, they can estimate the displacement around the fundamental frequencies. This let them to compare with the observations from the low-cost sensors and showing nice matching. They also have timing issue on different sensors, and they applied method to correct it. This approach uses the observation that the building’s response is dominated by the 1st modal response and the residual response which is the travelling wave due to the building’s transient response to earthquake forces exciting it at the base. Approximations are more accurate when the single record is obtained from near the top of the building.

[*Kohler et al.*, 2015] reports the ShakeNet project with the goal to record ambient vibrations for several days from new designed system. They designed the whole system to be a class A type of sensor, and have the capability to run on battery with several days. It can be used to quickly instrument large structures immediately after an earthquake to capture aftershocks and ambient motions. One thing during their test is that the 16-bit resolution sensors are inadequate for monitoring ambient vibrations in large structures. It was originally designed for high-frequency sensing in the kHz range, so its response at the sub-1 Hz modal frequencies of large structures was poor. Therefore, they are now using 24-bit acceleration data acquisition system that uses commercial low-noise MEMS. The timing in the system is sync through Flooding Time Synchronization Protocol (FTSP), but is infeasible because they were not able to estimate clock drift between master tier and sensor motes with millisecond accuracy. To overcome it, they use GPS to record the value of each ShakeBox’s internal clock at the beginning and end of data collection. This assumes linear drift between the two recording times, which may not always true. They also show some test results in the lab, and two prototype projects. It seems overall the system is good for quick structural monitoring, and I may use their system to do quick investigation of some of the monitoring in the future.

[*Cheng et al.*, 2015] reports the method to rapidly estimate the total displacement response of a building based on limited observational data, in some cases from a single seismometer. In general, the earliest part of the response is simulated by assuming a vertically propagating shear wave. Later motions are simulated using mode shapes derived from a beam model, the parameters of which are determined from the ratios of the modal frequencies and the building’s exterior dimensions. Then they verify the method by (1) comparing predicted and actual records from a 54-story building, and (2) comparing finite-element simulations of the 17-story UCLA Factor building. The method can be straightforwardly applied to multiple instrumented buildings, resulting in a tool to visualize linear elastic motions of those buildings.

[*Farrar and Worden*, 2007] gives a quick introduction of structural health monitoring. It talks the definition of damage, and monitoring. Then it discusses a brief history of the whole field. Afterwards, it talks the statistical pattern recognition paradigm, including (1) operational evaluation, (2) data acquisition, normalization and cleansing, (3) feature selection and information condensation, and (4) statistical model development for feature discrimination. It also talks the challenges for the SHM. This paper is just a quick introduction to the field, and after reading it, you will have a quick sense how people approach this problem.

[*Yin et al.*, 2016] talks about using the Palert system to do structural health monitoring. They talk about the hardware of the sensor, and use the method proposed by Nakata et al. to estimate fundamental frequency of a steel building in a laboratory to monitor the health of the building.

[*Duputel and Rivera*, 2017] reports main features of the M7.8 Kailoura earthquake (New Zealand) using long-period seismological observations. They calculated the apparent Rayleigh-wave moment-rate functions and found a clear northeastward directivity with an unusually weak rupture initiation during 60 s followed by a major 20 s burst of moment rate. They then did a Bayesian point-source inversion with 4 point-source events, and showed that rupture initiated as a small strike-slip rupture and propagated to the northeast, triggering large slip on both strike-slip and thrust faults. Thus the Kaikoura earthquake is a rare instance in which slip on intraplate faults trigger extensive interpolate thrust faulting.

[*Perol et al.*, 2017] explored using convolutional neural network to detect induced earthquakes. It seems it can achieve high detection rate while very efficient (since it is basically some matrix operation after training). What they did is to feed the waveforms of 3 components into the network, and train the convolutional network to recognize earthquakes. They first find the earthquakes from the catalog, and then in order to create a balanced dataset, they use Gaussian noise to perturb the waveform, which I am not sure if this is correct, I need test this approach. They also divide the interest region into 6 different regions, and use class 0 – 6 (0 means no earthquakes) to classify the earthquakes and decide which region they come from, they claim this as location (which I think is just group into different clusters, the detecting waveforms is more similar to the ones in the cluster). I think this method is promising, and I can try to build on top of their method, and make it a better one.

[*Background*, 2007] talks about how the virgin company entered into a crowd, mature market. The key points are:

1. Collaborate with a big company – Sprint
2. Identifying a Niche – the 15 – 29 segment, this is unmet need
3. Text Messaging
4. Online Real-Time Billing
5. Rescue Ring
6. Wake-up Call
7. Ring Tones
8. Fun clips
9. The hit list
10. Music Messenger
11. Movies
12. Special Advertising aiming for youth
13. Special pricing decision

This document has no results, which I really want to see how they did they success, but I guess if they fail, they probably not showing here as a case.

[*Noda et al.*, 2016] reports a new way to estimate earthquake magnitude that from the arrival time of the peak high-frequency amplitude. It uses the time different between the onset of the S-wave (which is estimated from the travel time from a velocity model) to the time of the peak high-frequency amplitude in an accelerogram. This parameter is designed to reflect the rupture duration of the rupture, and it should have weak link with epicentral distance, and can be ignored for distances < 200 km. It also uses high-frequency (>2 Hz) data. The results of the test using the Tohoku earthquake looks good.

[*Noda and Ellsworth*, 2016] proposes another relationship to estimate the magnitude from the beginning of P-wave from displacement records. What they found from the absolute average displacement is that the very beginning is similar for all the earthquakes, and then there’s a departure happens at different times for small and large earthquakes. By measuring this departure time, they found a relationship with magnitude. Here they argue that even though the similarity between small and large earthquakes can not be distinguishable from the very beginning, the departure time do provide a way to distinguish them. A couple of things need pay attention (a) the onset of the P wave in this study is all hand picked, what’s the uncertainty when apply the method to automatic pickings, (b) the way of measuring the departure is not so clear to me, and may not working in real time as well, (c) the filtering do seems have an effect on the measuring of the departure, and what is the uncertainty of that?

[*Niksarlioglu and Kulahci*, 2013] reports using ANN to predict magnitudes of the earthquakes along the East Anatolian Fault Zone. To estimate the magnitude, during the earthquake occurrence, the soil radon gas amount, latitude, longitude, stream pressure, wet bulb temperature, dry bulb temperature, temperature of the soil at 10, 20 and 50 cm depths used as the input to the ANN. They use Lavenberg-Marguart method to train ANN. Overall the paper is poorly written, but the method used is interesting to me.

[*Davidson*, 2017] gives the introduction of the special issue of The Leading Edge for Machine Learning. I really like some of the points in this introduction:

1. The complecx interactions between dynamic reservoir properties and the many-faceted well completions process are governed by complex physics that may be only partially understood.
2. Furthermore, our best attempts to perform controlled experiments can often produce highly varied results, suggesting the stochastic nature of production from unconventionals and that we may benefit from adding new tools to our toolkit of deterministic, physics-based analysis.
3. Many techniques used in these other industries can be directly utilized to understand oil field data and to optimize cost of supply; however, our industry has been relatively slow in adopting these new approaches.
4. These methods allow us to make sense of lots of data with many variables while avoiding the biases that humans can bring to any analysis where data set sizes may be reduced to just examining a few points or a few variables.

[*Rouet-Leduc et al.*, 2017] reports using machine learning algorithm to predict the occurrence of the earthquake in the lab environment. What they did is to collecting sensor data for the shear experiment, and solve a regression problem to predict the time until next earthquake. They use random forest to do the estimation, and used about 100 different features. In the paper, I didn’t see the details about the implementation, the test results seems amazing (but if I use such a flexible model to approximate a function, I can get that high precision as well). Anyway, this is based on lab experiment, I don’t think it will generalize well in the reality when applied to the waveforms of real recordings since they ignored too many factors.

[*De Wit et al.*, 2013] reports solve a general Bayesian non-linear inverse problem using artificial neural network approach, the science part of this paper is to investigate the constraint on spherically symmetric P-wave velocity structure provided by body-wave traveltimes from the EHB bulletin. They use a Mixture Density Network to obtain 1-D marginal posterior probability density functions. They first generate synthetic data using different earth models and adding noise to simulate real case. At the same time, they also used observational data from global events. Then they use the phase traveltime at different places as the input data to train a Mixture Density Network. The output is a 15 mixture Gaussian model for each velocity parameter (Note, they train a network for each velocity parameter). They found P-wave velocities in the inner core, outer core and lower mantle are resolved well. The data contain little or no information on P-wave velocity in the D” layer, the upper mantle and the homogeneous crustal layers. This is a very cool schema that I can use as an alternative to MCMC.

[*Chen and Bürgmann*, 2017] is a quick review of the creeping faults. It tries to answer the following 6 questions:

1. What are creeping faults?
2. Where do creeping faults occur and what damage do they cause?
3. Why do faults creep?
4. Are creeping faults a seismic hazard?
5. Can fault creep stop or limit earthquakes?
6. What further research is needed to better understand fault creep?

Specific types of rock, fluid pressure, temperature, the chemical environment, fault geometry, and sudden stress changes from nearby earthquakes all affect fault to creep. Also, creeping faults very often are found to not slip at their expected full long-term rate, when and how do creeping faults catch up with this slip deficit? A question is, does the fault creep reduce the amount of slip during a large earthquake and therefore cause less ground shaking? This is an interesting question to work on, [*Harris*, 2017] partially addressed it from SAF. The other question is why the creeping faults usually occur on strike-slip fault but only on a few thrust and normal faults?

[*DeVries et al.*, 2017] talks about using ANN (Artificial neural network) to speed up the large-scale viscoelastic calculations. What they did is quite simple, they train an artificial neural network to learn the mapping between the deformation and the model. Using a feed-forward model, they can train the ANN to output the displacement on the surface. They compare the results, and it is much faster than the viscoelastic code. They use Keras to do the job.

[*Kaiser et al.*, 2017] is the preliminary report of the 2016 Kaikoura New Zealand earthquake. It covers the initial results from earthquake source, ground motions and structural response, landslide and tsunami impacts, and aftershock forcasts and future earthquake scenarios.

[*Beans*, 2017] gives an overview of turning data into sound (music) as an alternative way to find insights, which is really interesting. It has already had many use cases, such as turning the climate temperature data, the tidal data, gene data, walking data and so on into sound (music), which is really cool. But as for the insights, I didn’t see a lot of insights from this way, it is just another representation, the things you can hear out, are similar to the things you can view out as well.

[*Hollinsworth et al.*, 2017] reports the investigation of rupture of the M7.8 2016 Kaikoura (NZ) earthquake by using optical satellite imagery and seismology. They first talked about using the Optical imagery and back-projection and finite fault model. They first try to use only one fault from the Global CMT to fit the data in finite fault model, but was unable to fit the long period observation and the uplift. Therefore, they need adding another fault to fit the data. Using only this two faults, they already can get a first order of fitting very well. They argue that the simultaneous coseismic slip occurred on the Kekerengu Fault (strike-slip) and a deeper shallow-dipping fault (oblique-slip).

[*Kanamori and Kikuchi*, 1993] reports a Tsunami earthquake – the 1992 Nicaragua earthquake. It has very rich long period wave with a moderate shaking. The duration of the source time function is about 100 s, with a very clear reverse mechanism. Two mechanism of the tsunami earthquakes were proposed, the first occurs in trenches with large amounts of sediment and an accretionary prism. Although the rupture of the individual thrust earthquake may not reach the ocean bottom, occasional slumping there may be the cause of large tsunami earthquakes. The second occurs in subduction zones without large amounts of sediment. In these subduction zones, the sediments are completely subducted and the plate interface is filled with soft sediments. The slip can extend to the surface, breaking through a relatively weak plate interface filled with sediments. Also, the difference between Mw and Ms are also very effective to recognize tsunami earthquakes.

[*Fujii et al.*, 2011] reports the tsunami waveform inversion for the March 11, 2011 Tohoku earthquake using Ocean-bottom pressure, GPS wave gauges, and coastal gauges. They reveal that the source of the largest tsunami was located near the axis of the Japan trench. They show the large tsunami was produced by both a very large displacement near the trench axis and a deeper interpolate slip in the southern Sanriku-oki, Miyagi-oki, and Fukushima-oki regions. The former explains the largest and impulsive tsunami waveforms, while the latter reproduces the initial part of the tsunami waveforms, as well as the large inundation on the Sendai plain. While they captured the first order of the waveform, there is a discrepancy near the central Sanriku-oki region, may require additional tsunami sources.

[*Newman and Okal*, 1998]

* Tsunami earthquake, characterized by significant deficiency of moment release at high frequencies.
* Difference between a tsunamigenic earthquake, which is merely an earthquake having generated an observable tsunami, and a ‘tsunami earthquake’, defined as an event whose tsunami is significantly larger than would be expected from its seismic waves.
* This paper starts with the two types of mechanisms generating tsunami from [*Kanamori and Kikuchi*, 1993], and due to the slower rupture of the tsunami earthquake, it is maybe easier to identify by purely seismic methods, this is the goal of this paper.
* They define a dimensionless parameter, log10(E/M0), similar to mb/Ms to discriminant tsunami earthquakes.

[*Shelly*, 2017] reports a nice dataset of a 15 year catalog of more than 1 million low-frequency earthquakes along the deep San Andreas Fault. In this paper, he talks about the method and procedures to create this catalog, as well as the uncertainties and difficulties associated with it. He also talks about the research effort on this dataset, particularly: Influence of two nearby larger earthquakes, recurrence pattern, bimodal 3 and 6 day recurring family, fast and slow migration of LFEs, and implications for Physics at the tremor source. Also, you can sort of using the correlation of the detections to monitor the network health status. Overall, this paper lay out a very nice dataset and I should work on it to draw useful insight.

[*Krizhevsky et al.*, 2012] reports using CNN to classify images. This is the first few papers that showed the amazing results of CNN. Here are some key points from the paper:

1. Deep convolutional neural network with ReLUs train several times faster than their equivalents with tanh/sigmoid units.   
2. ReLUs have the desirable property that they do not require input normalization to prevent them from saturating.   
3. Pooling layers in CNNs summarize the outputs of neighboring groups of neurons in the same kernel map.   
4. They maximizes the multinomial logistic regression objective, which is equivalent to maximizing the average across training cases of the log-probability of the correct label under the prediction distribution.   
5. Two ways they used for combating overfitting, (a) Data augmentation, (b) Dropout  
6. They found that their network's performance degrades if a single convolutional layer is removed.

[*Simard et al.*, 2003] reports some best practices using convolutional neural network in the tasks of image classification. The main finding they found are:

1. get a training set as large as possible, they expand the training set by adding a new form of distorted data
2. They found the convolutional neural networks are better suited for visual document tasks than fully connected networks.

[*Hinton et al.*, 2012] talks about the effectiveness of using dropout in the convolutional neural networks. They found using dropout are:

1. This prevents complex co-adaptations in which a feature detector is only helpful in the context of several other specific feature detectors  
2. Another way to view the dropout procedure is as a very efficient way of performing model averaging with neural networks

Also, in this paper, it shows the results that improved due to using dropout on various datasets. The most interesting part of the paper is the last paragraph, that link the findings to evolution.

[*Hensman and Masko*, 2015] did interest test of the influence of the imbalanced datasets on Convolutional neural network on CIFAR-10. They tried different distributions of different classes, and found that, the imbalanced datasets have a large effect on the performance. Then they also oversampled the minor classes by simply randomly duplicate the data samples. This simple oversample technique actually make the results comparable to the original balanced datasets, which I feel really amazed. Therefore, I can think of using the oversample even with a lot of duplicates to boost the performance. They also show that, after the oversample, the initial imbalanced distributions seems have no effect, which illustrate that the balanced distribution is more important than the unique data points.

Aguiar, A. C., and S. C. Myers (2016), *Characterizing Microseismicity at the Newberry Volcano Geothermal Site using PageRank*.

Background, C. (2007), Virgin Mobile USA : Pricing for the Very First Time, *Harv. Bus. Rev.*

Bayrakci, G. et al. (2016), Fault-controlled hydration of the upper mantle during continental rifting, *Nat. Geosci.*, (March), 1–6, doi:10.1038/ngeo2671.

Beans, C. (2017), Science and Culture: Musicians join scientists to explore data through sound, *Proc. Natl. Acad. Sci.*, *114*(18), 4563–4565, doi:10.1073/pnas.1705325114.

Bletery, Q., A. M. Thomas, A. W. Rempel, L. Karlstrom, A. Sladen, and L. De Barros (2016), Mega-earthquakes rupture flat megathrusts, *Science (80-. ).*, *354*(6315), 1027–1031, doi:10.1126/science.aag0482.

Chen, K. H., and R. Bürgmann (2017), Creeping faults: Good news, bad news?, *Rev. Geophys.*, 1–5, doi:10.1002/2017RG000565.

Cheng, M. H., M. D. Kohler, and T. H. Heaton (2015), Prediction of wave propagation in buildings using data from a single seismometer, *Bull. Seismol. Soc. Am.*, *105*(1), 107–119, doi:10.1785/0120140037.

Davidson, M. (2017), Introduction to this special section: Data analytics and machine learning, *Lead. Edge*, *36*(3), 206–206, doi:10.1190/tle36030206.1.

DeVries, P. M. R., T. Ben Thompson, and B. J. Meade (2017), Enabling large-scale viscoelastic calculations via neural network acceleration, *Geophys. Res. Lett.*, *44*(6), 2662–2669, doi:10.1002/2017GL072716.

Duputel, Z., and L. Rivera (2017), Long-period analysis of the 2016 Kaikoura earthquake, *Phys. Earth Planet. Inter.*, (February), doi:10.1016/j.pepi.2017.02.004.

Farrar, C. R., and K. Worden (2007), An introduction to structural health monitoring, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, *365*(1851), 303–15, doi:10.1098/rsta.2006.1928.

Focus, N. I. N. (2017), Astronomers explore uses for AI-generated images Hydrogen yet to prove it ’ s metal, *Nature*, 6–7.

Fujii, Y., K. Satake, S. Sakai, M. Shinohara, and T. Kanazawa (2011), Tsunami source of the 2011 off the Pacific coast of Tohoku Earthquake, *Earth, Planets Sp.*, *63*(7), 815–820, doi:10.5047/eps.2011.06.010.

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014), Generative Adversarial Networks, , 1–9.

Harris, R. A. (2017), Large earthquakes and creeping faults, *Rev. Geophys.*, 169–198, doi:10.1002/2016RG000539.

Hensman, P., and D. Masko (2015), The Impact of Imbalanced Training Data for Convolutional Neural Networks, KTH Royal Institute of Technology.

Hinton, G. E., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov (2012), Improving neural networks by preventing co-adaptation of feature detectors, , 1–18, doi:arXiv:1207.0580.

Hollinsworth, J., L. Ye, and J.-P. Avouac (2017), Dynamically triggered slip on a splay fault in the Mw 7.8, 2016 Kaikoura (New Zealand) earthquake, *Geophys. Res. Lett.*, (Figure 1), 1–9, doi:10.1002/2016GL072228.

Kaiser, A. et al. (2017), The Kaikoura (New Zealand) earthquake: preliminary seismological report, *Seismol. Res. Lett.*, (June), doi:10.1785/0220170018.

Kanamori, H., and M. Kikuchi (1993), The 1992 Nicaragua earthquake: a slow tsunami earthquake associated with subducted sediments, *Nature*, *361*(6414), 714–716, doi:10.1038/361714a0.

Kohler, B. M. D., S. Hao, N. Mishra, R. Govindan, R. Nigbor, S. Jewell, and U. S. G. Survey (2015), *ShakeNet : A Portable Wireless Sensor Network for Instrumenting Large Civil Structures*.

Kohler, M. D., T. H. Heaton, and M.-H. Cheng (2013), The community seismic network and quake-catcher network: enabling structural health monitoring through instrumentation by community participants, *Proc. SPIE - Int. Soc. Opt. Eng.*, *8692*, 86923X, doi:10.1117/12.2010306.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012), ImageNet Classification with Deep Convolutional Neural Networks, *Adv. Neural Inf. Process. Syst.*, 1–9, doi:http://dx.doi.org/10.1016/j.protcy.2014.09.007.

LeCun, Y., Y. Bengio, and G. Hinton (2015), Deep learning, *Nature*, *521*(7553), 436–444, doi:10.1038/nature14539.

Lenardic, A. (2017), PLATE TECTONICS A supercontinental boost, *Nat. Publ. Gr.*, *10*(1), 4–5, doi:10.1038/ngeo2862.

Levander, A., B. Schmandt, M. S. Miller, K. Liu, K. E. Karlstrom, R. S. Crow, C. T. A. Lee, and E. D. Humphreys (2011), Continuing Colorado plateau uplift by delamination-style convective lithospheric downwelling, *Nature*, *472*(7344), 461–465, doi:10.1038/nature10001.

Long, M. D., A. Levander, and P. M. Shearer (2014), An introduction to the special issue of Earth and Planetary Science Letters on USArray science, *Earth Planet. Sci. Lett.*, *1*(September 2014), 1–5, doi:10.1016/j.epsl.2014.06.016.

Massari, A., M. Kohler, R. Clayton, R. Guy, T. Heaton, J. Bunn, K. M. Chandy, and D. Demetri (2017), Dense Building Instrumentation Application for City-Wide Structural Health Monitoring, in *16th World Conference on Earthquake*.

Mishra, J. K., and R. G. Gordon (2016), The rigid-plate and shrinking-plate hypotheses: Implications for the azimuths of transform faults, *Tectonics*, *35*(8), 1827–1842, doi:10.1002/2015TC003968.

NAEEM ZAFAR, V. C. (2011), If you build it, they will come, *Berkeley - Haas Case Ser.*

Newman, A. V, and E. A. Okal (1998), Teleseismic estimates of radiated seismic energy: The E/M0 discriminant for tsunami earthquakes, *J. Geophys. Res.*, *103*(B11), 26885–26898, doi:10.1029/98JB02236.

Niksarlioglu, S., and F. Kulahci (2013), An Artificial Neural Network Model for Earthquake Prediction and Relations between Environmental Parameters and Earthquakes, *Int. J. Environ. Chem. Ecol. Geol. Geophys. Eng.*, *7*(2), 65–68.

Niu, F., P. G. Silver, R. M. Nadeau, and T. V. McEvilly (2003), Migration of seismic scatterers associated with the 1993 Parkfield aseismic transient event, *Nature*, *426*(6966), 544–548, doi:10.1038/nature02151.

Noda, S., and W. L. Ellsworth (2016), Scaling relation between earthquake magnitude and the departure time from P wave similar growth, *Geophys. Res. Lett.*, *43*(17), 9053–9060, doi:10.1002/2016GL070069.

Noda, S., S. Yamamoto, and W. L. Ellsworth (2016), Rapid Estimation of Earthquake Magnitude from the Arrival Time of the Peak High‐Frequency Amplitude, *Bull. Seismol. Soc. Am.*, *106*(1), 232–241, doi:10.1785/0120150108.

Perol, T., M. Gharbi, and M. Denolle (2017), Convolutional Neural Network for Earthquake Detection and Location,

Rouet-Leduc, B., C. Hulbert, N. Lubbers, K. Barros, C. Humphreys, and P. A. Johnson (2017), Machine Learning Predicts Laboratory Earthquakes, , 1–17.

Ruano, A. E., G. Madureira, O. Barros, H. R. Khosravani, M. G. Ruano, and P. M. Ferreira (2014), Seismic detection using support vector machines, *Neurocomputing*, *135*(January), 273–283, doi:10.1016/j.neucom.2013.12.020.

Sahlman, W. A. (2009), Dr. John’s Products, Ltd., *Harv. Bus. Rev.*, (January), 1–22.

Shelly, D. R. (2017), A 15 year catalog of more than 1 million low-frequency earthquakes: Tracking tremor and slip along the deep San Andreas Fault, *J. Geophys. Res. Solid Earth*, (Figure 1), 1–15, doi:10.1002/2017JB014047.

Shirzaei, M., W. L. Ellsworth, K. F. Tiampo, P. J. Gonzalez, and M. Manga (2016), Surface uplift and time-dependent seismic hazard due to fluid injection in eastern Texas, *Science (80-. ).*, *353*(6306), 1416–1419, doi:10.1126/science.aag0262.

Simard, P. Y., D. Steinkraus, and J. C. Platt (2003), Best practices for convolutional neural networks applied to visual document analysis, *Doc. Anal. Recognition, 2003. Proceedings. Seventh Int. Conf.*, 958–963, doi:10.1109/ICDAR.2003.1227801.

Taira, T., P. G. Silver, F. Niu, and R. M. Nadeau (2009), Remote triggering of fault-strength changes on the San Andreas fault at Parkfield, *Nature*, *461*(7264), 636–639, doi:10.1038/nature08395.

De Wit, R. W. L., A. P. Valentine, and J. Trampert (2013), Bayesian inference of Earth’s radial seismic structure from body-wave traveltimes using neural networks, *Geophys. J. Int.*, *195*(1), 408–422, doi:10.1093/gji/ggt220.

Yin, R.-C., Y.-M. Wu, and T.-Y. Hsu (2016), Application of the low-cost MEMS-type seismometer for structural health monitoring: A pre-study, in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pp. 1–5, IEEE.

Yoffie, D. B., and R. Kim (2009), E Ink in 2008, *Harv. Bus. Rev.*, (February), 1–5.

Zheng, L., R. G. Gordon, and C. Kreemer (2014), Absolute plate velocities from seismic anisotropy: Importance of correlated errors, *J. Geophys. Res. Solid Earth*, *119*(9), 7336–7352, doi:10.1002/2013JB010902.