Dear Dr. Jean-Philippe Avouac,

We thank both reviewers for their insightful comments on our manuscript EPSL-D-20-00481 entitled “Heterogeneous Dynamic Triggering of Earthquakes in the North Island of New Zealand Following the 2016 Mw7.8 Kaikoura Earthquake”. We are hereby submitting a new version that accounts for the comments. Below are the detailed responses to all comments. We mark the editor’s and reviewers’ comments as **bold**, our reply as regular font, and inserted/revised test in the paper as *italic*.

We thank you and both reviewers again and hope that the new version will be considered for further revision to Earth and Planetary Science Letters.

Best regards,

Dongdong Yao (dongdony@umich.edu), Zhigang Peng, Yoshihiro Kaneno, Bill Fry, and Xiaofeng Meng

**Reviewers' comments**:

Reviewer #1

**1. I am glad to see that the non-zero median was simply a plotting error, rather than something that should cause more concern. It is curious that this is a feature of the code: it seems to be a facility for covering up other internal problems with the code if it is required to give auto-correlations of 1.0.**

We thank the reviewer for pointing this out. The reason for not returning a 1.0 for self-detection is not due to other “internal problems” as the reviewer suspected. Instead it’s caused by the rounding issue when different traces starting with origin time with sub-sample differences [Meng et al., 2013]. The CC value at every single trace is 1.0. However, due to the rounding error, the different traces might have a one data point shift and result in a value of less than 1 after stacking and averaging.

**2. It is also good to see time-dependent Mc taken into account, although it isn’t clear to me that time-dependent Mc was taken into account for each spatial-bin. It would be good to confirm that and note it in the text. I would imagine that the Mc would be most elevated near the main- shock region.**

We didn’t apply the time-dependent Mc to each spatial-bin. Instead, we apply the method to the whole catalog. Since most initially mislocated aftershocks were removed by GNS analysts’ manual inspection and relocation (Figure S9), only a small to moderate elevation of the Mc values was observed following the mainshock.

3. **I continue to have concerns about the matched-filter detection process used. I plotted the first 100 detections in both the full and refined catalogues, using the bandpass-filters for both template sets. I have provided pdfs of those plots attached to this. While there are clearly some good detections in both sets, the number of false detections seems extremely high. This is per- haps unsurprising given the low correlation values (despite the high MAD thresholds). The cause of this false detection rate is unlikely to be related to the threshold used. The au- thors are using a very MAD high threshold, yet this is resulting in low-correlations: the MAD it- self must therefore be low. In my experience, template design plays a strong control on the quality of detections from matched-filters. I wonder how the authors decided on their chosen template parameters? Did any testing take place to check for false detection rates? In my ex- perience, removing low-frequencies often results in a significantly increased false detection rate. I would encourage the authors to check their detections for false detections, and if they find that the false detection rate is as high as it appears from the first 100 detections, that they reconsider their choices of template parameters. Furthermore, when I have run templates with bad processing parameters I have noted an increase in false detections and decrease in real detections. Because the false detection rate looks to be very high (I estimate 54 out of the first 100 detections in the full detection set are false and 46 out of the first 100 refined detections are false), and the subsequent work is based on assuming that the detections represent real events, the rest of the work is likely poorly supported.**

We thank the reviewer for exploring our MFT catalogs by plotting out and visually inspecting the waveforms. We generally agree with the reviewer that the peak amplitude ratio might be noisy especially for detections with very low CC values. However, we would argue that event without visible phase picks don’t necessarily lead to a false detection. For example, Li et al. (SEG [2015], <http://geophysics.eas.gatech.edu/zpeng/zpeng_paper/Li_etal_SEG_2015.pdf>) showed that when stacking with plenty of stations, it is possible to detect a seismic event that is completely buried in the background noise.

After evaluating our analysis procedures, we suspect that following parameters might contribute to seemly “high false rate” and problematic magnitudes. 1) We only utilized P waves, which would loosely match the P wave moveout especially when there are fewer stations. Similar to event location, this would match events originating from different locations. 2) we used a long template window (15s for previous refine detection), which would include a significant portion of noise. Hence, it would suppress the signal as well. 3) we applied a 10-30 Hz filter, rather than the typical frequency range of 2-8 or 2-16 Hz used in other recent studies. While the high-frequency range helps to suppress signals in the aftershock region, high-frequency local noise might affect the detection result.

As suggested by the reviewer, we conduct another analysis using the following parameters, as well as the updated GNS catalog and phase picks. We started with careful quality control of the updated catalog events (Last accessed: Sep 2020), and a total number of 1,466 templates between 2016/11/01 and 2016/11/30 were chosen to scan through continuous waveform. A 5s-long window starting 1s before P and S wave on vertical and horizontal components were utilized, respectively (see updated main text for detailed method). As done by the reviewer, we generated the event waveform for first 100 detections by carefully selecting corresponding traces based on their best-matched template (i.e., the template with higher CC value).

As mentioned before, Li et al. [2015] reported that the multiple-station matched filter technique could significantly enhance the detection with a very low SNR situation (up to a magnitude 2 difference) by adding more stations. We also designed a simple test here. We selected a ML2.1 earthquake in the GeoNet catalog (2016p834777: 2016/11/04 20:49:20), and gradually reduced its amplitude by multiplying 10^(-0.1), 10^(-0.2), …, until 10^(-3). These events with reduced amplitudes were added on top of continuous waveforms of the day 2016/11/04 before the catalog event’s origin time with a time step of 300s (Figure R1). We then used the ML2.1 template to scan the continuous waveforms on the day 2016/11/04, and saved positive detection above the 9MAD threshold. As expected, the MFT can still help to detect the buried signal even below the noise level (Figure R2).

We plotted out the histogram of the relative magnitude (magnitude difference) of our detected catalog (Figure R3). More than 97% of the detected events has magnitude difference above -2, which are within the magnitude range where the smaller signal could be uncovered by the MFT.

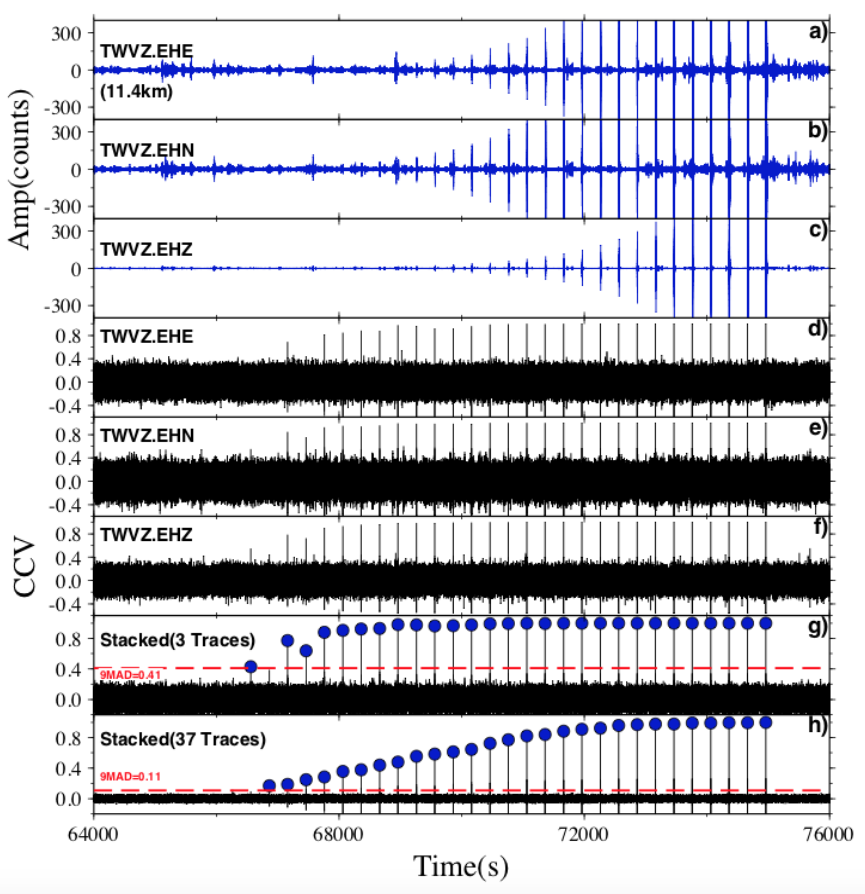


Figure R1. SNR synthetic test. a)-c): three component continuous waveform of the nearest station NZ.TWVZ (11.4km). A ML2.1 catalog event was added evenly every 300s with decreasing amplitudes. d)-f) CC function of above three components. g) stacked CC function for above 3 traces. Detections using 9MAD as the cut-off threshold (horizontal dashed line) is shown with blue filled circles. h) stacked CC function for all 37 traces for this template with SNR above 5 (see main text for choosing trace in template matching).

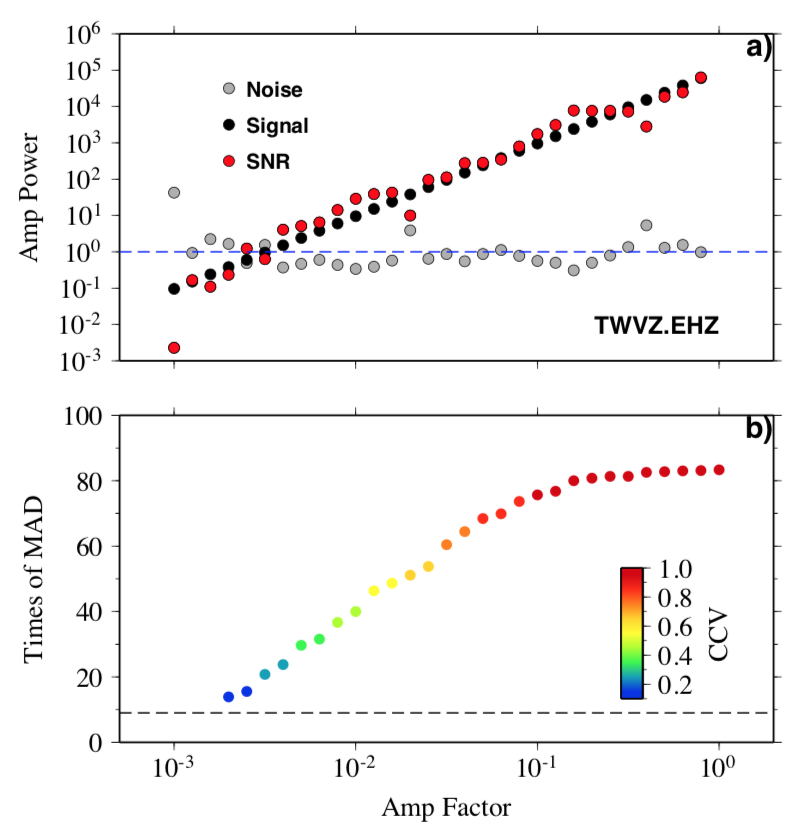


Figure R2. a) Signal-noise-ratio for “buried signal” of trace TWVZ.EHZ, where noise window is chosen as a 5s-long window right before P wave. b) Time of MAD (all traces stacked) as a function of amplitude factor applied to buried signal, color-coded with the CC value.

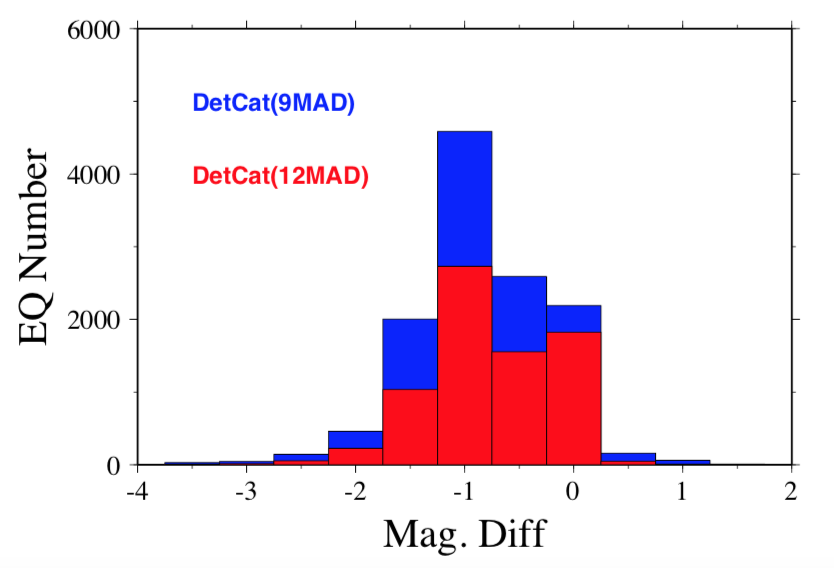


Figure R3. Magnitude difference (relative magnitude to corresponding template) histogram for events in detected catalog (9MAD as blue and 12MAD as red). 97% and 98% of detections with relative magnitude larger than 2 for each catalog.

**4. Completeness: figure S4, right hand panel. This Mc of 2.3 remains a very poor fit to the observations. I still suspect that this is due to incorrect magnitudes (as outlined above). Nevertheless, the method used to estimate the magnitude of completeness does not appear to be working as the authors expect and I suggest they try a different approach to finding the completeness.**

We agree with the reviewer that the magnitudes based on peak amplitude ratios might be contaminated by noisy traces, especially only P waves were used. In the updated version, we removed the result by using all catalog events, where many aftershocks from the Kaikoura were wrongly mis-catalogued as local events in the North Island. In the new analysis, we included both P and S waves, and the resulting cumulative frequency-magnitude distribution can be found in Figure R4 (also as the new Figure S3).

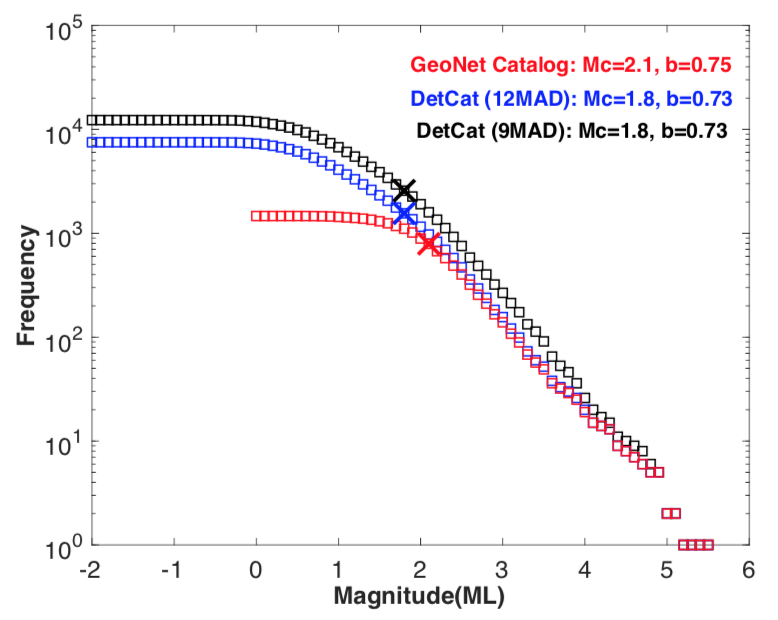


Figure R4. Cumulative frequency-magnitude distribution for the GeoNet Catalog (1,466 earthquakes, red squares), detection catalog of 12MAD (7,513 events, blue), and detection catalog of 9MAD (12,291 events, black). Magnitude of completeness (Mc, marked with cross symbol) and b value for each catalog are labeled.

We also test other magnitude calibration methods. As noted in Data and Method section, a peak amplitude ratio (PAR) method was applied to obtain the relative magnitude between every detection and its best-matched template, and we further obtain its local magnitude using the catalog magnitude of the best-matched template (Mag(PAR\_all)). Firstly, instead of using all available traces, we applied a signal-noise-ratio threshold to each trace and only used traces with SNR above 1.5. The resulting magnitude Mag(PAR\_SNR15) is shown in Figure R5a & R5b for both detection catalog of 9MAD and 12MAD. Moreover, we adopted a similar principal component fit (PCF) method as used in several recent studies [Shelly et al., 2016; Meng et al., 2018; Yao et al. 2020]. Specifically, we run another cross-correlation between the best-matched template and the detected event around its origin time (1minute before and after). Only trace with CC at the matched time above a threshold was used for the PCF analysis, and the threshold is defined as the median + 7\*MAD of the 2-minute long CC function. We required a minimum number of 3 traces to find the median amplitude ratio. The obtained Mag(PCF) is also shown in Figure R5c & R5d. The magnitudes of different methods show generally consistency, while the PCF analysis would result in smaller amplitude ratios when compared to peak amplitude ratios. In summary, the choices of different magnitude calibration methods should not change our current observations and major conclusions.

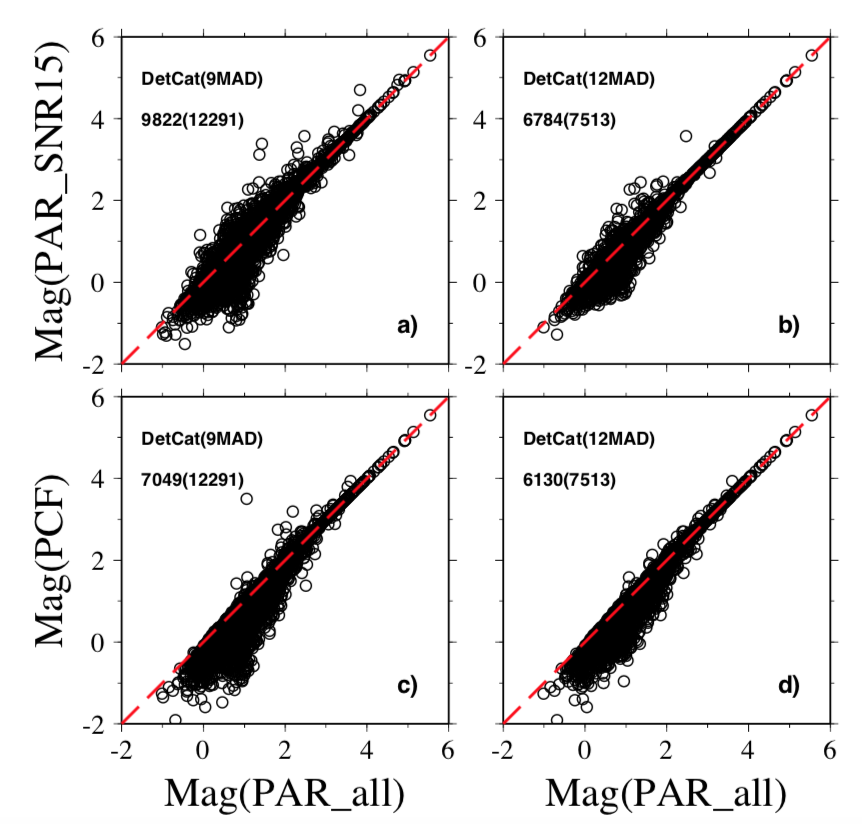


Figure R5. A comparison of different magnitudes. a) and b) show the magnitudes from peak amplitude ratio (PAR) method: X axis is the median PAR for all available traces, while Y axis only uses trace with SNR above 1.5. c) and d) plot out magnitude from principal component fit (PCF) method and PAR method.

**5. In all map plots the authors should mask out regions where you don’t have templates. The reason I originally questioned the application of this method for such a spatial comparison is because, as the authors state - they expect to detect events from the same location as their templates. Because there are not templates everywhere, by the authors logic, the catalogue cannot be spatially complete. At the least, regions in map-plots where there are no templates should be masked out to avoid coming unsupported conclusions. The authors mention this briefly in lines 290-292, but it would be better placed in sections 4.2 and 4.3.**

Thanks for pointing this out. We agree that the MFT could help improve the completes of the catalog, but is still limited by the input catalog. Only events surrounding the catalog ones could be matched. Hence, in our analysis, we exclude the region without templates in our beta map or seismicity rate analysis (set as zero). In addition, we focus our analysis and discussion in regions with templates. The main goal of this work is to obtain a more complete catalog, and present the complicated triggering behaviors in the North Island following Kaikoura mainshock.

**Reviewer #2:**

**6. "we identify up to ten times more earthquakes than listed in the GeoNet catalog."  
The GeoNet catalog immediately after the mainshock has many miss-located events, resulting to false detections by the template matching. So, it is here better to describe the result about template matching using refined events.**

We agree with the reviewer, and believe that it would cause more confusion by parallel presenting both analyses. Hence, in the updated version, we removed the “primary detection” result, and focused on major findings using refined template events, after adjusting parameters accordingly.

**7. In the short time window (Ta=1) after the Kaikoura mainshock, there seems to be no statistically significant increase in seismicity in North island from Figure 5 (detected events based on the refined templates). Can you still conclude the immediate increase in seismicity to be robust? Figure 7b: At the north of Lake Taupo, it is difficult to identify the increase in seismicity. Why? I can see the clear step in the cumulative number in Figure 5b. I am bit confusing.**

As shown in Figure 5b, an obvious increase of seismicity was found to the north of Lake Taupo, as well as β values above 2 (Figure 6b, similar to old Figure 7b). However, within Lake Taupo, due to the active earthquake swarm starting from 2016/11/09 6pm UTC Time (Figure 5c), it exhibits less prominent increase as the region to the North of Lake Taupo. As reported by Peng et al. [2018] (Figure 8), many instantaneously triggered earthquakes occurred during and immediately following the passing of mainshock surface wave (Figure S1).

Reference:

Li, Z., Peng, Z., Meng, X., Inbal, A., Xie, Y., Hollis, D., Ampuero, J.P., 2015. Matched filter detection of microseismicity in Long Beach with a 5200-station dense array. In 2015 SEG Annual Meeting (pp. 2615-2619). New Orleans, Louisiana.

Meng, X., Yang, H., Peng, Z., 2018. Foreshocks, b value map, and aftershock triggering for the 2011 Mw5.7 Virginia earthquake. J. Geophys. Res.: Solid Earth 123, 5082-5098.

Shelly, D.R, Ellsworth, W.L., Hill, D.P., 2016. Fluid-faulting evolution in high definition: Connecting fault structure and frequency-magnitude variations during the 2014 Long Valley caldera, California, earthquake swarm. J. Geophys. Res.: Solid Earth 121, 1776-1795.

Yao, D., Huang, Y., Peng, Z., Castro, R.R., 2020. Detailed investigation of the foreshock sequence of the 2010 Mw7.2 El Mayor-Cucapah earthquake. J. Geophs. Res.: Solid Earth 124, e2019JB019076.