Dear Prof. Hager,

we wish to thank you for your thoughtful review and for the useful insight you provided. Your comments helped us to focus on some key issues of our paper and to achieve a better understanding of our findings. In the following we respond to the issues you raised, one by one (your original text is shown in italics).

The authors of this paper put forward the novel suggestion that the absence or presence of hydrocarbons at typical reservoir depths in fold and thrust belts may be used to infer whether the faults beneath the folds are able to generate large earthquakes or not. The mechanism that they propose is that earthquakes with M > 5.5 produce shaking that fractures the reservoir seals, allowing hydrocarbons to leak to the surface, resulting in sterile fields. This is an interesting conjecture, relevant to Natural Hazards and Earth Systems, and bringing it forward for discussion is good.

This is all correct except for one major point. The mechanism we advocate as being able to fracture the reservoir seals is not the shaking *per se*: in fact we are convinced of the opposite, i.e. that the shaking alone is unable to cause hydrocarbon leaks. Probably we were not clear enough in the paper and we apologize for that, but we believe that what causes such leaks is the actual slip on faults underlying the reservoir, including the main seismogenic rupture plane and any significant splay that may occur above it. In our view earthquakes of M 5.5+ are large enough to 1) guarantee that the causative fault slipped by at least a few cm during the mainshock, and 2) cause sizable dislocation along all faults lying over a considerable thickness of the upper crust (e.g. from 8 to 3 km). Both these conditions increase the likelihood that the earthquake will create open gaps in the cap-rock through which the gas may escape and be lost in the atmosphere.
Hence, to summarize, our key concept for explaining the gas leaks is not "fault-induced shaking" but rather "fault-induced finite dislocation".

Their suggestion is developed in the context of the 2012 Emilia-Romagna earthquake sequence, a sequence of great contemporary interest. However, because of the slow rate of convergence (1 – 3 mm/yr) across the region, the expected recurrence interval for earthquakes on the relevant structures is very long, with only 4 M > 5.5 earthquakes having occurred over the past five centuries. Of these, only the two events in 2012 were located instrumentally, with fault rupture areas relatively well constrained.

The number of earthquakes that occurred in our study area is very limited and in fact incompatible (too small) with the expected moment release budget even assuming a convergence rate of 1 mm/yr. As pointed out by previous investigators, the "seismic coupling" (or earthquake efficiency) of major faults of the Po Plain might be 20% or less, something that is seen across many other compressional belts worldwide.

By limiting their study to this region, the authors are forced to deal with a poor statistical sample.

We used over 400 wells that were analyzed one by one to gather their depth and to allow them to be assigned to one of our categories. We maintain that this is not a small sample for Seismology, a science were dealing with samples of only a few units or tens of units is very common. We will be glad to expand the sample once our suggestion is accepted by the community.

The authors make clear that the structures beneath areas that generate large earthquakes and earthquakes that may not are similar, with folds overlying deeper thrusts in both cases. Since development of these structures is associated with large motions on faults, it would be interesting, but not essential, to discuss why, with comparable materials and geology, some faults slip in earthquakes and others could slip aseismically.

The nature of the rocks being folded beneath the Po Plain and their structural setting is indeed not homogeneous. In their "Three-dimensional segmentation and different rupture behavior during the 2012 Emilia seismic sequence (Northern Italy)" (doi:10.1016/j.tecto.2014.05.006) Bonini et al. analyzed in detail the pattern of coseismic slip associated with the 20-29 May 2012, Emilia earthquakes. The contend that "...that seismogenic ruptures were confined in the Mesozoic carbonates and were stopped by lithological changes and/or mechanical complexities of the fault planes, both along dip and along strike. Our findings highlight that along the active structures of the Po Plain slip tends to be seismogenic where faults are located in Mesozoic carbonate rocks....". Because Mesozoic carbonate rocks are not always encountered at the typical seismogenic depth of Po Plain faults (3-10 km), these results imply that many of such faults have limited or no seismogenic potential. In our view this implies that the overlying reservoirs (if any) are safer than those overlying faults cutting through Mesozoic rocks.

Specific Comments
My major specific comment is that the statistical treatment in support of the novel suggestion is not very convincing. To acquire the data used to try to test the hypothesis that earthquakes result in sterile reservoirs, the authors use a large, publically available data base about hydrocarbon (oil and gas) wells. They group the wells into three categories: a) positively sterile; b) positively productive, and c) wells that encountered oil or gas that either was not produced or was produced from shallow depths – called "ambiguous."
From the standpoint of the main suggestion of the paper – that earthquakes cause reservoirs to leak, the lack of a clear explanation of why wells in category c) should be separated into a separate category from b) and given such a subjective name as "ambiguous" is a concern. It seems to me that the question of whether when oil or gas is present it is produced or not is not related to whether the reservoir leaks – if the hydrocarbon is there, it seems that the reservoir has not leaked. I do not understand why category c) data are excluded from the statistical test.

Based on your observations we have re-analyzed all wells that we included in the "ambiguous" category. We first subdivided this category into two sub-groups including a) exploration boreholes which revealed a gas/oil reservoir, but for which there is no evidence concerning whether or not they went into production, and b) gas reservoirs shallower than 500 m. As for sub-group a), we agree with your observation ("*if the hydrocarbon is there, it seems that the reservoir has not leaked"*). But since the available information does not allow to assess how much gas was found in those wells, i.e. if they can be considered intact reservoirs), we have decided to make two different simulations; the first considering all the ambiguous wells of this sub-group as productive, the second considering them all sterile (see below). As for sub-group b), since we contend that in the seismotectonic context of the Po Plain a typical M 5.5+ earthquake may cause sizable dislocation over faults lying between 10 and 3 km depth, we decided to disregard shallow reservoirs as they are supposed to be insensitive to what happens at seismogenic depth.

However, since the locations of the historical ISS are somewhat uncertain, and including the
category c) data would mainly affect the conclusions about the 1624 ISS, I do not think that including category c) data would completely refute the suggestion.
Another problem with the statistical analysis is that, in my opinion, the application of binomial test is not appropriate. The binomial test is based on the assumption that the distribution being tested is random. It could be argued that the distribution of category a) wells is approximately random. However, category b) wells are clearly clustered in space, as can be seen in Figure 1. The reason for this clustering is that new wells are preferentially drilled close to producing wells – once a sweet spot is found, it is highly likely that adjacent areas will also be productive. One way around this problem would be to divide the region into equal area blocks. That is, test the area containing producing wells, not the number of wells. The area might be chosen to be comparable to the source area of a M = 5, or, alternatively, with a characteristic dimension comparable to the average well spacing .
In addition, limiting the area tested to the interior of the rectangle defining each ISS seems problematic. The largest ground shaking tends to be near the edges of the fault patch, but not confined to its interior.

We gladly acknowledged your suggestion and performed a spatial analysis aimed to determine whether our findings are statistically significant or instead if they cannot be distinguished from randomness.
We tested the results on our selected Individual Seismogenic Sources using a Monte-Carlo simulation. At each run our simulation sampled at random the content of four boxes having the average size of the typical Emilia-Romagna seismogenic faults, about 10 by 5 km. We then compared the result of 10.000 random samples over the study area. To account for the uncertainty caused by the ambiguous wells we ran the test twice, once assuming that the ambiguous wells were all productive and once assuming they were all sterile.
The values to be compared with are the two scores deriving from considering either productive or sterile the only ambiguous well falling within a seismic

source, that is for the sum of the four zones, 19 sterile and 1 productive or 18 sterile and two productive. The two combinations never occurred over our 10.000 simulations.

Interestingly we noted two distinct behaviors, that gave us a better insight on our hypothesis:

1) the distribution of the number of productive wells inside the fault boxes decays slower for higher numbers, confirming Prof. Hager's suggestion that productive wells tend to be more clustered (many productive wells enter a box that intercepts a productive field;

2) the probability of having a large number of sterile wells and no or few productive wells inside the fault boxes is lower than the probability of having a large number of sterile wells and some or many productive wells. This is probably due to the fact that a larger number of sterile wells can be found surrounding the more productive areas; most likely they result from the attempt to test the boundaries of the reservoir. Moreover, it is unlikely that many productive wells are drilled close one to another others, unless a seismic survey has returned a pattern similar to nearby productive reservoir. This means that the sterile tectonic traps are similar to the productive tectonic traps, but the fact that one is seismically active and the other is not, makes the difference at the base of our hypothesis.

This analysis and the relevant diagrams and maps can be included in a revised version of the manuscript.

Although almost all of the paper is devoted to investigations of both oil and gas reservoirs, and most of these, including Cavone, produce far more oil than gas, the paper ends with the conclusion that only depleted gas reservoirs, not depleted oil reservoirs, should be used for underground gas storage. Whether this recommendation is true or not, it does not seem to have support in the preceding text of the paper. Similarly, the specification of "gas reservoirs" in the title is misleading because the bulk of the analysis involves oil reservoirs.

This sentence arises from a misunderstanding, probably caused by the quality of our English of by poor wording. The bulk of our analysis concerns gas reservoirs, about 97% vs about 3% of oil-andgas fields. We contend that the mechanism we envision may explain the lack of gas, but does not necessarily affect the integrity of oil fields for obvious differences in viscosity and volatility between gas and oil.

As noted above, because the rate of occurrence of earthquakes in the region investigated is relatively low, the statistics is quite challenging.

We addressed the issue of the low earthquake activity rate and of its statistical significance in a previous answer.

Perhaps publication of this paper will stimulate further investigation in other regions with more earthquakes. Along those lines, McGarr (BSSA, 1991) examined three earthquakes that occurred beneath producing oil fields (Coalinga

1983, Kettleman N. Dome 1985, Whittier Narrows 1987). So there seems to be evidence that, at least in California, earthquakes and reservoirs coincide.

It would indeed be extremely interesting to test our hypothesis in other regions were hydrocarbon reservoirs lie above large seismogenic faults. The earthquakes investigated by Art McGarr, however, do not make a good case for this test. The 1983 Coalinga and 1985 Kettleman Hills earthquakes occurred next to large oil-hydrocarbon fields. In addition to that, the reservoir is relatively shallow whereas the seismogenic source is somewhat deeper than our Emilia sources, probably between 5 and 10 Km depth (e.g. Stein and Ekstrom, JGR 1992; doi: 10.1029/91JB02847). Similarly, the Whittier Narrows earthquake ruptured between 12 and 17 Km depth (Lin and Stein, JGR 1989; doi: 10.1029/JB094iB07p09614) beneath a shallow oil field. There are probably better candidates for testing our hypothesis in other parts of California as well as in Turkmenistan, in southern Iran and in China.

In the abstract, it seems misleading to say that the 2012 earthquake source regions were "surrounded" by productive wells.
Figure 1 shows almost a complete 180° arc north of the mainshock epicenter with no producing wells, as well as a smaller but substantial gap to the south.

Figure 1 caption – the description of the 1624 earthquake has typos ". . . and 1624, and 19 March. . ."

These two remarks will be addressed in the future version of the paper.