# DK11757: Reply to referees

December 7, 2015

We thanks the referees for the comments/suggestions, which we think have helped to improve the quality and clarity of the paper. In the following there are our answers to the various points raised by the referees.

## 1 Referee A

The paper "First low frequency all-sky search for continuous gravitational wave signals" is a large collaboration paper of the LIGO/Virgo Scientific Collaboration, presenting the results of a continuous wave search in the VSR2 and VSR4 science runs. The analysis was a fully blind search (no prior knowledge about candidates, either in frequency or direction), using a hierarchical approach with a combination of coherent and incoherent techniques, all of which have been discussed well in the literature. A fully coherent search is impossible for this kind of blind search, so at present the presented analysis is the best that can be done. No signals were found in the data, and upper-limits were placed on the presence of continuous waves that are astrophysically somewhat interesting. The presented analysis will be applied to the data of Advanced LIGO and Virgo, with LIGO already operational, which makes this presented work especially relevant.

I found the presentation of this paper comprehensive, the science is great, and it is obvious that a lot of care was taken when preparing this manuscript. I therefore recommend this paper to be published as soon as possible. I do not have many recommendations for this paper that need to be addressed:

1. It feels like the paper is partially written by a committee that subsequently tried to combine the paper. This is especially apparent from the figures, which have similar, but not the same markup and color. Note that Fig. 16 does not have the color 'red', so please adjust the caption.
   **ANSWER: Right. We have tried to make figures more homogeneous.**

2. About the outlier at 43.30 Hz. From the paper, it is not clear to me whether the outlier analysis is manual labor, or whether this is also done automatically. Since the peakmap in Fig. 16 is so different between detectors, surely this does not need to be done by eye? A note to explain might be in place.

**ANSWER: Yes, currently it is done by eye. In perspective, it will important to automatize this step as much as possible.**

## 2 Referee B

The authors have performed a low-frequency (20-128Hz) all-sky search for continuous gravitational waves using data from the second and the fourth Virgo Science Runs. This is the first time this kind of search is conducted on the frequency band (20-50Hz). A novel approach of the "FrequencyHough transform" was applied in the two-stage (both incoherent) hierarchical search. They have examined data conditions and noise characteristics including line noise artifacts carefully. Although no gravitational wave was found, they have set upperlimits on strain amplitudes for the first time in 20-50Hz and with improvements of up to a factor of 2 in 50-80 Hz.

Overall, this paper is written well, provided enough explanations on the method applied in this paper, apart from the several points listed below. Namely, I am not sure if the upper limit is on a strain amplitude, or on a number of the pairs of in the two runs or a kind of a mixture of both, due to their way of combining ten 0.1 Hz sub-bands (See the points 17-19 below).
**ANSWER: the upper limit is on the signal strain, as explained in the first paragraph of Sec. X ("strain" now explicitly added in the paper text): we determine the signal strain amplitude such that 90% of the injected signals, with that strain amplitude, are detected and are more significant than the candidate found in the real analysis, using the same procedure we use to analyze data.**

While I possibly misunderstand the method, but general readers may also misunderstand it. Please revisit the points listed below and give more detailed explanations. After reading a revised manuscript, I would see the paper fit for publication.

1. Abstract:
*This is the first all-sky search for continuous gravitational waves conducted at frequencies below 50 Hz.* I have not done a complete survey, but I think that the method of pulsar timing has been applied for all-sky search for continuous gravitational waves below 50 Hz.
**ANSWER: Right. We have now specified it is on data of ground-based interferometric detectors.**

2. Page 6:
*"On the other hand, based on the electromagnetically observed pulsar population and the results of population synthesis modelling, e.g. [8], [9], we expect that a substantial fraction of the galactic neutron star population emits gravitational waves at frequencies below ∼100 Hz."*: The ATNF catalogue shows that the spin frequencies of the electromagnetically known pulsars follow a bi-modal distribution. And there is a valley between the two modes from 10 Hz (or $f_{gw} = 20$ Hz) to 100Hz. See also, e.g. Figure 3

of "Binary and Millisecond Pulsars", Duncan R. Lorimer, living reviews in relativity (2008). So, I am not sure if electromagnetically observed pulsar population indicates 20Hz - 128Hz is "potentially promising" compared to other bands (amptiudes proportional to (smaller) frequency squared and the smaller number of (known) pulsars). If I misunderstand something (selection bias?), it may be very nice if the authors comment on why the famous bi-modal distribution in the $P - \dot{P}$ diagram does not actually reflect the true distribution.
**ANSWER: In fact we refer to the population of neutron star spinning at frequency larger than about 10Hz, corresponding to a signal frequency greater than about 20Hz. Below 20 Hz detector sensitivity is strongly degraded and it is not worthwhile making the analysis. We have tried to explain this in the text.**

3. Page 6:
Why is the frequency region below 20 Hz not analyzed ?
**ANSWER: see previous answer.**

4. Page 7, below Eq. (8):
The authors say *"Smaller relativistic effects, namely the Einstein delay and the Shapiro delay are not relevant for the search described in this paper, due to the use of short length FFTs, and are therfore neglected."*, but in the page 13, it is said that *"where the Doppler effect, the spin-down and the Einstein delay for a source, having the same parameters as the candidate, have been corrected."*. Which statement is correct? Please correct/clarify either or both. (Perhaps, it makes the manuscript clearer to replace "in this paper" by "the incoherent steps"?)
**ANSWER: Right. In fact the Einstein delay is not considered in the incoherent step of the search, while it is taken into account in the candidate followup. We have tried to clarify this at the end of Sec. II.**

5. Page 7:
*"In the all-sky search described in this paper we need to take into account only the first spin-down (s = 1) parameter (see Sec. IV)."*: Higher order time-derivatives may be necessary for young pulsars like the Crab pulsar. Do the authors actually mean to say that computational cost demands them to consider the first time-derivative only? If so, please clarify it.
**ANSWER: Yes, it is a consequence of the parameter choice we have done: the order of spin-down we need to consider depends basically on the FFT length and the minimum star "age" we want to search for (see Eq. 31 of [7]). We have done "reasonable" choices for this search. With more extreme choices, e.g. a significantly shorter minimum age, we would have explored a larger parameter space, at the cost of a higher computational load. This is explained in the third paragraph in page 8: "We have searched approximately over the spin-down range etc.".**

6. Page 7:
*"The low-frequency sensitivity of VSR4 was significantly better, up to a factor of 2, than that of previous Virgo runs,"*: It might sound strange to researchers outside of the GW community reading "factor 2" is "significant".
**ANSWER: Given that the GW strain h is proportional to 1/r, a sensitivity improvement of 2 corresponds to an increase of a factor of 8 in the accessible volume of space (assuming a homogenous source distribution).**

7. Page 8:
*"The FFTs are interlaced by half and windowed with a Tukey window with a width parameter $\alpha = 0.5$."*: Please define $\alpha$ or provide a reference in which $\alpha$ is defined.
**ANSWER: Reference added.**

8. Page 8:
*"$N_{sd} = 16$ for VSR2 with a resolution of $\dot{f} = 7.63 \times 10^{-12}$ Hz/s, and $N_{sd} = 9$ for VSR4 with a resolution of $\dot{f} = 1.5 \times 10^{-11}$ Hz/s."*: But Table II said $N_{sd} = 15, 8$. Which are correct?
**ANSWER: Right. table corrected.**

9. Page 9:
"$2K_{\dot{f}} = 12$" and "$\hat{K}_{sky} = 5$": For a sky position I see a hat ( $\hat{}$ ) while in other variables I do not. Is there any intention? (Reading [7], I guess $\hat{K}_{\rm sky}$ is the over-resolution factor for the refined sky-grid, while $K_{\rm sky}$ is the one for the coarse sky-grid. But please make the notaions self-explanatory, do not ask readers to dig into past papers.)
**ANSWER: Right, hat in $\hat{K_{sky}}$ removed.**

10. Page 9:
*"The full set of jobs was run on the European Grid Infrastructure (`http://www.egi.eu/`)."*: Several search parameters in the analysis are determined by the computational burden of the analysis. It is very much helpful to readers if the authors mention how large the computational cost was specifically for the current analysis.
**ANSWER: This analysis computational load was about 25,000 CPU*hours (most of which due to the incoherent step). We added this information. (In practice it was more because of the few re-run we made after bug findings, algorithm improvements and so on)**

11. Page 10:
*"we have chosen a time resolution $\Delta t_H = 12$ hours and a frequency resolution $\Delta f_H = 0.01$ Hz."*: It seems there is no need to define $\Delta f_H$ and $\Delta t_H$ (Are they used anywhere other than here?). Besides, $\Delta \dot{f}$ denotes the range of $\dot{f}$ while $\delta \dot{f}$ denotes the frequency time derivative bin size. I

expect a similar relation between $\delta f_H$ and $\Delta f_H$ which is not the case here.
**ANSWER: Right. We have removed the two variable names.**

12. Page 12:
*"we have that that the number of candidates selected per sky patch ..."*: One of the two "that"s is unnecessary?
**ANSWER: Ok.**

13. Page 12:
*"Once the most significant candidate in each sub-band has been selected, an empirically established exclusion region of* $\pm 4$ *frequency bins around it is imposed."*: The authors say (it seems to me) a tail of a candidate ($\pm 4$ frequency bins) is determined empirically. What actually is done? Please cite a reference (if any).
**ANSWER: This means we do not select any further candidate which is within that range from the loudest candidate in each sub-band. In this way we reduce the probability to select more candidates which could be due to the same disturbance. The second loudest candidate is selected only if it is far enough from it. We have tried to explain this better and added a reference to the method paper.**

14. Page 12:
*"...or a particularly strong HI."*: It is not helpful to use the abbreviated word "HI". It is not frequently used in the paper.
**ANSWER: Ok.**

15. Page 12:
*"The choice* $d_{coin} = 2$*, based on a study of software-injected signals, allows us to reduce the false alarm probability and is robust enough against the fact that true signals can be found with slightly different parameters in the two datasets."*: Please cite a reference of the signal injection study that the authors mention.
**ANSWER: In fact this has been a study done purposely for this paper. We have not added more details because the paper is already rather rich of technical details and we would like to avoid to weigh it down too much. Of course, if the referee strongly believes we have to add more technical details we can do.**

16. Page 13: Is it possible that a signal generates two events that can be detected in VSR2 and VSR4 respectively and that lie in two consecutive 0.1 Hz frequency bands, and the method adopted here misses coincidence between them? In other words, how does the pipeline deal with a signal which lies at the edge of a 0.1 Hz band?
**ANSWER: No, this cannot happen. The division in 0.1Hz bands is done at the level of the candidate selection for each run. The coincidences among candidates are done after clusterization, as explained in Sec. VII, in which no division in sub-bands is done.**

17. Page 13:
*"At this point, among the ten candidates in ten consecutive 0.1 Hz bands we chose the most significant one, ending up with 108 candidates, one per 1 Hz band, which will be subject to a follow-up procedure, as described in the next section."*: In what sense the selected one (for each 1 Hz band) out of 10 candidates from ten 0.1 Hz bands is most significant? Do you select one "most significant event" based on smallness of its rank product value? Please clarify it.
**ANSWER: Yes, that's right. Added a sentence.**

18. Page 13:
Related to the above question 17, suppose there are two consecutive 0.1 Hz frequency bands. One is very noisy band in the both runs. I guess that we would obtain a larger number of coincidences $N$ between the runs out of such a noisier band. The other band is less noisier in the both runs, and we would obtain a smaller number of coincidences $N'$. Then the loudest candidate from the noisier band has a smaller rank product $1/N^2 < 1/(N')^2$ even if the less noisier band actually contains a true GW signal that has the smallest rank product $(1/(N')^2)$ in its 0.1Hz band. Does the method adopted in this paper safely detect the GW signal in this hypothetical case when combining these two consecutive frequency bands?
**See next ANSWER.**

19. Page 13:
Related to the above question 17, suppose a similar situation as in the question 18. Suppose the Hough amplitudes of the two loudest candidates (one from VSR2 and the other from VSR4) in the less noisier band are much larger than the ones in the noisier band. Do you label the pair in the noisier band more "significant" based on the smallness of the rank product? Do you put an "upperlimit" on $h_0$ on the 1 Hz band containing these two 0.1 bands based on the smallest rank product pair which actually has not-the-biggest (within that 1 Hz band) Hough amplitudes?
**ANSWER (also to previous question): The reviewer's reasoning would be exact if no data cleaning step was applied. In fact, we do cleanings, especially in the time-frequency domain (as described in Sec. V). As a consequence in a very disturbed 0.1Hz band many, or most, or even all the frequency bins could be vetoed. Then, that 0.1 Hz band could simply not produce any candidate or, more likely, produce just a smaller number of candidates with respect to "average" bands. As a consequence the most significant candidate in that 0.1Hz band likely will not be the most significant over the 1 Hz band. This is what we observe. Obviously there can be exceptions to this. A way to reduce further their effect would be that of simply selecting one candidate every 0.1Hz, without making the further selection over 1Hz. This would imply a number of candidates to followup**

**ten time larger with a relevant increase of the computational load, and this is what we plan to do in the analysis of Advanced detector data. We have added a short sentence on this.**

20. Page 13:
*"The follow-up procedure consists of several steps applied to each of the 108 selected candidates. First, a fully coherent step, using the candidate parameters, is done separately for the two runs. This is based on the same analysis pipeline used for targeted searches [1], [2], and the output is a down-sampled (at 1 Hz) time series where the Doppler effect, the spin-down and the Einstein delay for a source, having the same parameters as the candidate, have been corrected. "*: This paragraph (or the word "fully-coherent") somewhat confused me. Does the authors perform a full-coherent follow-up? But the following paragraph and the abstract says they have used Frequency Houghmap which is an incoherent method. If "a fully coherent step" was applied after the first incoherent step, then why the authors then used less sensitive incoherent method again?
**ANSWER: As said in the second paragraph of Sec. VIII, what we do is to apply a coherent step using candidate parameters with the aim of partially correcting the Doppler and spin-down effect in such a way to be able to increase the FFT length before making a new Hough transform stage. We have tried to explain this better adding a sentence in the first paragraph.**

21. Page 13:
*"From these data, a new set of longer FFTs is computed, from chunks of data of duration 81,920 seconds, which is ten times longer than in the initial step of the analysis."*: From which data? Is it *"a down-sampled (at 1 Hz) time series where the Doppler effect, the spin-down and the Einstein delay for a source, having the same parameters as the candidate, have been corrected"*?
**ANSWER: Yes.**

22. Page 14:
*"In fact, as descibed in Appendix , ..."*: Perhaps, Appendix B? And "descibed" $\rightarrow$ "described"?
**ANSWER: Right, ok.**

23. Page 16:
*"In other words, for each 1-Hz band we generate 100 simulated signals in the time domain, with unitary amplitude and having frequency and the other parameters drawn from a uniform distribution ..."*: Are those 100 fake signals injected at the same time? If so, what is the probability that there are more than two fake signals in a $\pm 0.2$ Hz band mentioned in *"For each set of 100 injections an analysis is done using the all-sky search code over a reduced parameter space around each injection, consisting of a frequency band of $\pm 0.2$ Hz around the injection frequency,..."*? Would

those fake signals other than the one in attention affect estimate of noise floor? If yes, to what extent they affect upperlimits?
**See next ANSWER.**

24. Page 16:
Related to the question 23, if 100 fake signals are injected at the same time, what is the probability that this procedure misidentifies a coincident pair? Would this effect make the pipeline overestimate its signal detection efficiency and underestimate upperlimits?
**ANSWER (for this and the previous question): Yes, the signals are injected at the same time. In principle yes a candidate of one injection could be very near to another injection, altough this is not very likely. However, to avoid this possible effect: a) the signals are generated with a minimum frequency difference between any two of them of about 0.005Hz, which corresponds to about 0.005*8192$\sim$40 frequency bins; b) any two injections having a distance in frequency smaller than 0.4 Hz must have positions in the sky such that their "9-points search region" do not overlap. Because an injection is considered as detected if the pair of its most significative coincident candidates is within d=2 from it, we are guaranteed that candidate mis-identification cannot happen. We have added this information in the paper. We do not use any estimate of the noise floor, but simply count how many injections out of 100 are detected.**

25. *"The gaps are the result of the various cleaning steps applied to the data, and the injections that, for a given amplitude, have frequency overlapping with a gap will not able, in most cases, to produce detectable candidates because the amplitude in the Hough map near the signal frequency is reduced by the low noise contribution."*: It sounds counter-intuitive to me that lower the noise floor is, harder we detect a fake signal injected there. Could you explain the reason in more details?
**ANSWER: An injected signal corresponds to a set of peaks in the peakmap. The number of "signal peaks" is significantly reduced if they occur in a disturbed band, because the peaks (both noise and signal, obviously) are vetoed by one of the cleaning procedures. Then it will be difficult for that signal to produce a high number count in the Hough map.**

26. Page 21
*"They correspond, respectively, to candidates number 69, 25 and 52 in Tab. . Note that the reported parameter values are those after the follow-up step and then are sligthly different from those in Tab. which, on the contrary, have been computed before the follow-up."*: Perhaps, Tab. V? And "sligthly" $\rightarrow$ slightly?
**ANSWER: Right.**

27. Page 22
*"In fact, despite all the cleaning procedures and criteria used to select candidates, some instrumental disturbance can be still present into the data."*: I agree with this statement, but I disagree that this statement is written here as it is. I imagine that you would find it difficult to convince a person that you find a true GW signal (suppose you do) who says that it is just an instrumental line noise artifact because "In fact, despite all the cleaning procedures and criteria used to select candidates, some instrumental disturbance can be still present into the data."
**ANSWER: In fact, this is why we need to do a deeper follow-up of the most interesting candidates. We have added a sentence on this.**