Dear editors,

we thank both the referees for the positive judgement of our work and in the following we reply to their recommendations and comments.

Response to report of Referee A

1. Introduction, first paragraph: “higher order” -> “higher-order”

   - **R**: Done.

2. Introduction, seventh paragraph: “i.e.” -> “i.e.,”

   - **R**: We have done this replacement for every occurrence in the manuscript.

3. Sec. IIA, after Eq. (5): “The phase of the signal follows directly...” This is a bit clunky and could probably be expressed more clearly in an equation.

   - **R**: We added the following reference in the text: PRD 70, 082001 (2004).

4. Sec. IIB, after Eq. (10): “where m represents the nominal single dimension mismatch value.” Could you explain this a little more? What is the significance of m? You give a numerical value below [after Eq. (12)], but you don’t really explain what it means.

   - **R**: In general, the term mismatch is used to indicate the fractional loss in signal-to-noise ratio between the parameters of a certain signal and those of the closest search template. Given the comment of the referee, we added in the text after Eq. (10) two references for further details.

5. Sec. IIIC, first paragraph: “i.e.” -> “i.e.,”

   - **R**: We have done this replacement for every occurrence in the manuscript.

6. Sec. IIIC, second paragraph: “given by Eqs. (12) and (13), respectively and...” -> ”given by Eqs. (12) and (13), respectively,...”

   - **R**: Done.

7. Sec. IVE, last paragraph: “of these two injected signals respectively.” -> “of these two injected signals, respectively.”

   - **R**: Done.

8. Sec. VC, last paragraph: It might be nice to mention how far away we need to be able to look in order to see a significant population of pulsars.

   - **R**: We are looking for an unusual pulsar: the bulk of the population of known pulsars has rotation frequencies lower than 20 Hz, a significant fraction of them are in binary systems and the measured spindowns imply a maximum gravitational wave emission that makes them too weak to be detected, at least for the ones for which the distance is known. Therefore we state distances for some reference cases but do not speculate about the population.

9. Sec. VII, sixth paragraph: missing a space, “is described in[55].” -> “is described in [55].”

   - **R**: Done.

Response to report of Referee B

1. Page 6. Eq. (6): Can the effects due to proper motions of pulsars, such as the Shklovskii effect, be neglected? In the conclusion section, it is written that “the corresponding maximum reach is roughly 4 kpc, assuming epsilon $\sim 10^{-4}$”. Then it is less than 4-40 pc for conventional neutron stars which is so close that a proper motion may be important.

   - **R**: The referee is correct, and indeed proper motion might cause us to mis-identify the parameters of the star and possibly even affect detection probabilities. However, this will happen only if the star is very near to us. A rough estimate of the distance below which proper motion effects become important is as follows: if we take a neutron star with a velocity of 1000 km/s, and a maximum detector velocity $v/c \sim 10^{-4}$, then the distance below which we need to start worrying about proper motion is $d < (f \times T_{\rm obs} \times T_{\rm seg}) \times (\pi/\sqrt{12 \cdot 0.3}) \times (50\ \mathrm{m/s})$. Here $T_{\rm obs}$ is the full observation (about 1 year) time and $T_{\rm seg}$ is the maximum time spanned in each segment (25 hours). This is obtained by saying that the star must not move by more than one fine sky pixel (given by Eq. (13) in the manuscript, with the parameter $\wp = 0.5$) over the full observation time. For $T_{\rm obs}$=1 year and $T_{\rm seg} = 25$ hours, this leads to d $<$ 1.6 pc at 200 Hz and 7.8 pc at 1000 Hz. On the other hand, a neutron star closer than 7.8 pc, with an ellipticity of 10^{-6}, would have a large signal amplitude and would have been detected by the PowerFlux search mentioned in the paper, which employs shorter coherent time baselines and therefore is more tolerant of source parameter drifts.

On page 6., just before Eq. (6), we added the following footnote: "Proper motion of the source can safely be neglected for distances greater than 10 pc."

2. Page 6. Below Eq. (8): What is the "wave-frame"?

   - **R**: Below Eq. (8) we added the following clarification: "The wave-frame is a right-handed Cartesian coordinate system based on the direction of propagation of the gravitational wave. Its z-axis is along the direction of propagation, and its x-and y-axes are along the principal directions of polarization of the wave."

3. Page 7, Eq. (9-13): Why no correlation among two sky position angles, f, and $\dot{f}$ is considered?

   - **R**: There are correlations between the two sky angles and these are reflected in the sky-grid which has cells of approximately constant surface area in the sky. In general, there are also correlations between the two position angles and f, $\dot{f}$, and these are different on different segments (through the effect of the orbital motion of the Earth) so the grids that could exploit such correlations would differ from segment to segment. This would have complicated the incoherent combination of the different segments and in particular it would have made it hard to use the computational techniques that have made the Hough number count sums so fast to compute. Based on these considerations, for the S5R3-R5-R6 runs, we decided to trade computing power for simplicity and set up grids that were conservative in the sense that they sampled more points than strictly necessary. Ways to exploit such correlations were subsequently proposed ("Phys.Rev.Lett.103:181102,2009", "Phys.Rev.D82:042002,2010" and "Phys.Rev.D75:023004,2007; Erratum-ibid.D75:069901,2007") and different techniques to combine the different segments were used in the following runs.

     We have added some references that may help the reader who is interested in the technical details. In particular, before Eq. (9) we added "D. M. . Whitbeck, Ph.D. thesis, The Pennsylvania State University (2006)" and "Phys.Rev., D82, 042002 (2010)". After Eq. (10) we added "Phys.Rev.D75:023004,2007; Erratum-ibid.D75:069901,2007".

4. Page 7. Eq. (12): What is $\mathcal{R}$ in this equation ? Which difference between R3 and R5 makes the author introduce this parameter?

- **R**: It is simply a numerical fudge factor determined from Monte-Carlo studies. After the S5R3 run, Monte-Carlo studies for the S5R5 run suggested that the size of the sky-patch in Eq. (11) could be made bigger without significant loss of sensitivity. We hence decided to spend the resulting computing power on other parts of the parameter space, i.e., searching for a larger spindown interval.

5. Page 8. Eq. (22): p is defined as "the probability that a parameter space pixel is selected in the absence of a signal". So it might vary when noise properties vary. Then cannot p differ from a segment to a segment?

   - **R**: p is a fixed number, defined assuming Gaussian noise. It indicates the false-alarm probability (so in the absence of a signal) that $2\mathcal{F}$ exceeds the particular threshold set in our search (i.e., $2\mathcal{F}_{th} = 5.2$). If the noise were Gaussian and stationary, $2\mathcal{F}$ would follow a central χ^2 distribution with 4 degrees of freedom, and hence p = 1-chi2cdf(5.2,4) = (1+$\mathcal{F}_{th}$)*exp(-$\mathcal{F}_{th}$) = 0.267 (see for example Ref. [19] in the manuscript). When the noise is non stationary and/or not Gaussian, this threshold corresponds to a different false-alarm probability, and such false-alarm probability may well vary from segment to segment. This would alter the distribution of the number count with respect to the gaussian-stationary noise case. However, in Eq. (21) we define the critical ratio CR in terms of the gaussian-noise false-alarm probability corresponding to the $2\mathcal{F}$ threshold of 5.2. This means that the CR of our results might not always follow the same distribution. However, this does not affect the reliability of the results of our search because we do not use the theoretical distributions in our analysis other than to gauge that there are no obvious mistakes and that the bulk of our results follows the expected distributions (see for example the black curves in Fig. 3).

     We slightly changed the text in the manuscript, right after Eq.(21) in the following way: "...that measures the significance of n_c as the deviation from the expected value $\bar{n}_{\mathrm{c}}$, measured in units of the noise fluctuations σ in Gaussian stationary noise in the absence of any signal. In these circumstances, $\bar{n}_{\mathrm{c}} = N_{\mathrm{seg}}p$ and p is the probability that a χ^2 variable with 4 degrees of freedom (as $2\mathcal{F}$ in absence of a signal) exceeds the threshold of 5.2."

6. Page 10. Eq. (23) and below: Δ may differ from a candidate to a candidate. There are 10^{11} candidates. The probability of two invalid results could agree with each other accidentally is 2×10^{-8}. Then I would expect 2000 candidates which mistakenly pass through the validation veto and should not be marked as "candidate". Perhaps my question is: Is this "the probability that two invalid

..” is for two WUs (or two result files), or two candidates in two distinct result files?

- **R**: This probability refers to two result files.

7. Page 10. Eq. (25) below: What is “sky-grid files”?

   - **R**: We have added the following sentence after Eq. (25), at the end of the paragraph: “A sky-grid file is a two-column ASCII file which specifies the positions in the sky (right ascension and declination) of the coarse grid templates to be searched for every frequency and spindown value in a given 10 Hz band.”

8. Page 10. Eq. (25) below: Each WU file contains only 20 mHz data. A host machine may download “sky-grid files” (whatever it is) for only that 20 mHz frequency band. Why should the “sky-grids” be constant over 10 Hz, while such a wide band data is irrelevant for any host machine? Making “sky-grids” constant over 10 Hz makes it easier for a host machine to download “sky-grid files” for 20 mHz data?

   - **R**: In principle a different grid over the sky should be used for every 20 mHz of searched data. However, grids associated with neighbouring frequency bands would not be very different one from the other. This can be seen from Fig. 2. So, in order to not have to deal with so many (60 000 =1 200 Hz/20 mHz) different sky-grid files, we decided to search the same points in the sky every 10 Hz, and used for each 10 Hz band the grid that is suitable for the highest frequency in that band, hence slightly over-covering the sky for all frequencies in the band. So we traded a bit of computing power for simplicity.

9. Page 11. footnote 2: It is nice if this footnote explains why $\dot{f}$ can be positive and spindown age does not make much sense in that case.

   - **R**: We added the following sentence to the footnote: “Moreover, the notion of a spindown-age does not make sense for cases when $\dot{f} > 0$. This might happen either due to accretion, or apparent spinup from acceleration in the vicinity of a globular cluster core, or because of proper motion leading to an apparently positive measured value of $\dot{f}$.”

10. Page 11. footnote 3:It is difficult (for me) to read $\frac{\hat{f}}{c}$ written using such a tiny script.

    - **R**: Improved.

11. What are the standard deviations of τ_H^1 and τ_F^1 ?

    - **R**: It is actually difficult to now retrieve the variations of these numbers for the S5R3-R5-R6 runs. However, the measured overall run time now typically varies by about 30% (at the 1-sigma level). This is a good indication of the accuracy of the predicted run time for the above coefficients as well. We added the following sentence after Eq. (31): "The typical run time is found to vary by about 30% among different machines (at the 1-sigma level)."

12. Section IV.: Why the authors first selected 100 most significant candidates and then removed known instrumental noise artifacts? If first remove and then select 100, we get higher chance of detection?

    - **R**: Removing candidates that originate from data contaminated by known instrumental disturbances should indeed be the first step in the post-processing. In later post-processing runs the scheme is indeed this, and the whole post-processing is in general rationalized and made more efficient: for example, the WU result file size has been decreased by a factor of 3. The S5R3-R5-R6 were the first really large production runs that we made. In these runs we implemented the line removal as a second step, after having selected the top 100 candidates because: 1) when we started the post-processing the information on the line artifacts was not available, 2) in order to make the new data set manageable we wanted to save a much smaller fraction of the reorganized data, 3) because of the way the original result files were stored, the computational task of gathering together the results from 0.5 Hz bands was rather time consuming and we did not want to repeat it.

13. Page 13, subsection C. Section IV: What is the false dismissal probability for this F-statistic consistency veto? Is it zero?

    - **R**: It is virtually zero, evaluated through Monte-Carlo studies.

14. Page 17, subsection F. Section IV, last para: What is the chance probability that all the three candidates are actually due to gravitational wave pulsars but the NOMAD search dismisses all of them?

    - **R**: As stated in the manuscript, based on Monte-Carlo studies, the false-dismissal probability of the NOMAD follow-up is less than 10% for each candidate. Thus, the probability that NOMAD would dismiss all three of them is no more than 10^{-3}.

15. Page 20, Eq. (43): Is $h_0^{90\%}$ proportional to $1/N_{seg}^{1/4}$?

- **R**: No, please refer to Sec. VB of the manuscript and Ref. "PRD 85, 084010 (2012)" therein, and also to our answer to item 18 below for further details.

16. Page 20, Eq. (43) below, last para in Sec. V.B. : Does Eq. (43) incorporates the uncertainties of both the Monte-Carlo and the calibration? It seems to me that the variation of Eq. (43) is just due to the variation of the noise, S_h, and nothing to do with the Monte-Carlo. Or the Monte-Carlo uncertainties affects $h_0^{90\%}$ through $\rho^{90\%}$? Then could you explain how?

    - **R**: Equation (43) does not incorporate Monte-Carlo or calibration uncertainties and the variations are attributed to small changes in the noise floor.

17. Table III: It is nice if $\ddot{f}$'s are shown in this table.

    - **R**: No second order spindown values were considered for the hardware-injected signals of Table III.

18. Conclusion: Taking the lower threshold earns better sensitivity of factor 2.5 than the previous search. Then, the other 0.5 is due to taking the longer data? Having 1 year data instead of two month data contribute a factor of $(12/2)^{1/4} = 1.6$ better sensitivity, assuming an incoherent method. But $1.6 \times 2.5 = 4$, not 3.

    - **R**: To predict the sensitivity of a search, as we have done in our manuscript, we have to resort to numerical methods tailored to the specific thresholds and set-up of the search. However, we follow the referee's attempt to make a ball-park estimate based on simple scaling laws, so that we can point to a misunderstanding about the 2.5 sensitivity improvement, which we have addressed with a modification to the manuscript. A relative sensitivity factor might look like $\sqrt{\frac{40}{30}} \times (121/28)^{0.25} \sim 1.6$, from the fact that S5R1 used 28 segments with 30 hours of data, while S5R5-R6 used 121 segments each with 40 hours of data. Assuming that this factor is not too wrong for this search, the only thing that we could conclude is that the remaining gap between the two results is due to the effect of the lower $\mathcal{F}$-statistic threshold and using a better method (coincidences are after all a "poor-man's Hough transform"). As we state in the manuscript, a factor 2.5 would be expected between two S5R5-R6-style searches differing only in the thresholds applied to the $2\mathcal{F}$-values. Such 2.5 is not the right number to use when comparing the sensitivities of the S5R1 and the S5R5-R6 searches, which are different searches. We emphasize that a proper estimate of such a number (which we have not performed) would entail a numerical study in order to disentangle the dependance of the number of segments from the effect of lowering the threshold. Indeed, these two are coupled through the binomial false-alarm probability, that depends on the threshold through the single-trial-probability, and depends directly on the total number of segments, since these are total number of binomial trials.

We rephrased slightly the first paragraph of the conclusions as follows: "If we had used the much higher threshold of 25 on $2\mathcal{F}$, as in [24], our sensitivity would have been a factor of ~ 2.5 worse than our final upper limits[1], thereby undoing almost all of the factor of 3 improvement mentioned above".

19. Conclusion: If the Powerflux method results in a similar sensitivity as the current Hough method, then what was the benefit of using the Einstein@Home facility? Could the authors explain how essential the E@H was for this search and for the authors' gravitational wave search efforts to motivate the current and potential Einstein@Home users?

    - **R**: The Powerflux search investigated frequencies up to 800 Hz, whereas this search goes up to 1200 Hz. Because of the scaling of the number of sky points to be searched with the square of the frequency, these additional 400 Hz increase the computational cost by a factor of about 3.5. This means that broad-band frequency searches may simply not be possible with computing facilities other than Einstein@Home. Furthermore, the Powerflux search aims at a fast turn around of results, targeting putative loud signals. In the implementation presented in [26] it was however not quite used in that mode, and utilized substantial computing resources on the Collaboration's dedicated clusters. As the computing needs of other types of searches increase (for example the searches for signals from binary inspirals of high mass systems with spin), such resources may not be available to continuous gravitational wave searches any more. Finally, the Powerflux search is itself less sensitive than the one reported here, in fact it produced marginally less sensitive results while having searched about as twice as much data. Having said all this, it is beneficial that more than a single search for the same class of objects be developed and exercised, with independent software. This allows indeed cross-checking and validation of results.

---

[1] Note that this does not mean that simply lowering the threshold to 5.2 in the S5R1 search would increase the S5R1 sensitivity by a factor 2.5.