The Authors analyzed time-averaged quantum Fisher information (which can be a testimony of multipartite entanglement) for the chaotic models described by Dyson's three circular ensembles. They derive the simple and general asymptotic formula for averaged QFI Eq. (2) (by 'asymptotic' I mean valid for large dimensions of Krylov space), in line with intuition and expectations. Next, using both analytical argument and numerical simulations, they perform a detailed analysis of symmetric N-qubit space for three ensembles taken from ref. [62], showing Heisenberg scaling of QFI (which is indeed a testimony of multipartite entanglement). Finally, they compared the results with different chaotic models and the model with semiclassical correspondence, showing, that the previous statements cannot be extended for the latter models.

The fact, that the random state can have large QFI has been noticed and broadly explored in 2016 in [38], where Haar-random states have been analyzed. The Authors extend it to the case of chaotic dynamics in a correct and methodical way. Therefore, I would say that the paper "provides a link between different research areas" and "opens a new pathway in an existing direction, with clear potential for multi-pronged follow-up work", so it satisfies Scipost criteria.

However, I do have some reservations about the presentation of the results and some of the relationships with the literature that need to be addressed before I recommend publication.

1. (optional, but strongly recommended) One of the main results of the paper is formula Eq. (2), which is meaningful only in the limit of large K. It would be profitable to derive and present the exact lower and upper bound for QFI with the exact constant multiplying the term $Tr[O^2]/K^2$, instead of saying, that it is equality up to the term of order $Tr[O^2]/K^2$. Looking at App.A, it seems that it does not require much additional work, but it would be extremely helpful for the reader to judge, for which K the formula is already a reasonable approximation.

2. In Figure 2. the Authors present the perfect correspondence of the numerical results and analytical predictions for the large N. This is highly expected due to applied theorems. It would be profitable to show also the results for smaller N to see, how the asymptotic formula Eq. (2) becomes valid (i.e., what is the smallest N, to make it reasonable). That would also allow us to see the differences in averaged QFI for different ensembles (which disappear with increasing N).

3. In the introduction the Authors refer to [26-28], ending the paragraph with the conclusion "It thus remains unclear which types of chaotic systems can generate scalable multipartite entanglement, i.e., QFI scaling as N^2 known as the Heisenberg scaling.". All references [26-28] refer to the optimal metrology strategy where both the number of particles N, but **also total time** t are treated as resources. For example, in [27] we read "In this scenario, even if the system might be able to be prepared in a global multipartite entangled state with $F_Q \sim N^2$, the Heisenberg scaling would not be feasible due to the required state preparation time that scales at least linearly with the system size". Therefore, [27] does not exclude the possibility of obtaining $F_Q \sim N^2$; that only said, that it would be not useful for metrology (as it takes a long time, which could be more efficiently spent for performing more repetitions of experiments). Do the Authors claim, that their results stay in consistent, contradict, or are completely unrelated to the ones from [26-28]? Please clarify this relation.

4. Regarding the 'Conclusion' section, I really doubt the potential for application in metrology and I find the comparison with GHZ states extremely dishonest. In the paper, the Author's analyze the averaged value of QFI calculated for pure stated obtainable by chaotic evolution. To use such a state in metrological tasks, we need to have full control of the chaotic Hamiltonian and to know perfectly the time of evolution chaotic evolution, because optimal measurement and the estimator may depend on the exact state. Prepering the states by performing chaotic evolution seems to be an extremely unstable method. So even if it could be possible to perform proof-of-principle experiments, I cannot find an example, when using this theorem would be useful for real metrological problems.

5. Next, regarding GHZ states – it is of course true that any **single** GHZ states achieve high QFI for only one specific axis. The authors claim, that using chaotic systems gives an advantage, as the averaged QFI scales like $N^2/3$ for all axes. But this is an unfair comparison – if one considers the set of three different GHZ states, for them the average QFI will be also $N^2/3$. It would be worth stressing, that what the Authors analyzed is **the average value of QFI for pure states**, not **the value of QFI for the averaged (mixed) stated**. Therefore, an operational understanding of their results corresponds to the situation, where different measurements are performed at different times, with the perfect knowledge of the time of performing measurement. It is not that they have found a single state useful for measuring rotation for all axes – they only show that during chaotic evolution different states are good for measuring rotation around different axes, so on average QFI for all axes is large. But exactly the same effect may be obtained for the usage of different GHZ states. So where is the advantage? (optional) If the Authors are interested in showing the true superiority of random chaotic states over the GHZ states, I recommend analyzing for example noise resistance (as done in [38] for random Haar state).

6. Besides that, I have so minor or technical comments:

(A) Before Eq. (4), the definition of fidelity is written for the very strange form to me. I.e., it trivially simplifies to $\langle\psi_0|W^\dagger(\theta,t)\rho(0)W(\theta,t)|\psi_0\rangle$, or even simpler $|\langle\psi_{chaos}(t)|\exp(i\theta O)|\psi_{chaos}(t)\rangle|^2$ (which is the most common form for pure states). Is the usage of the longer and more complicated form is intentional?

(B) The Figure 2. is hard to read after printing in A4 format. Please take, to make it readable (for example, Fig. 3 and Fig. 4 are perfectly readable). Morevoer, in the inset of panel (h) the red line is missing (while comparing with (b) and (f)) – what is the reason for that?

While all these comments are addressed, I will be glad to recommend the publication.