Why I removed the Hellinger distance

The paper is unreadable when I tried to re-read it.

The issue is not simply restricted to story re-telling.

We did not do our job properly of represeting concepts in a the right mathematical way. The review and revision was sloppy and indequate and should have caught this very basic issues with the process of scientific research. We chose the easiest way out without adequare review and challenge. Looking back it is embarrassing that you and I with all our experience could write such a low quality paper.

A noise parameter like initialization fidelity F_I is a statistical estimate to start with. Our paper is talking about fluctuations in F_I which means we are tallking of the statistics of a statisical estimate. When we say that the pdf is itself varying with time, we mean that any particular pdf of F_I is a realization from a stochastic process. This is statistics of statistics of a statisical estimate. On this we layer on the Hellinger distance analysis and make comments on the Hellinger distance's mean and variance. These mean, variance numbers are dervived statistical properties (Hellinger) of the statistics (variation of pdf) of statistics (pdf) of a statisical estimate (F_I). This is insanity. We need to drastically simplify and revisit the whole scientific story we are telling and what is the simplest and most impactful way of telling the story.

In fact, our scientific story is much cleaner and simpler without resorting to Hellinger distance.

Everything we are trying to say can be said much more simply.

What about non-stationarity? I argue we do not need to bring in non-stationaritity analysis at all. In fact, we did not do any statistical tests for non-stationarity. Looking at Hellinger distance there is absolutely no way of justifying our claim. Stationarity is not measuring like this. In fact, the standard tests yield the results that the time-seris is stationary. But that is not our intended message.

Our previous manuscript had several other issues which I have corrected in this new revision:

* we say stability test but there is no test in our old paper. It is just a vague hand waving argument. We have put in a concrete test now in this version.

* the old paper had a mix of two devices yorktown and toronto. this was not needed and is confusing. in this new version, we confine our attention to just one device - the latest one from IBM at the time of writing - and instead we provide a live link which shows all the 44 devices of IBM in one place.

* Reliability is the right term to use. Reliability analysis uses confidence intervals.

* Hellinger distance is good for digital output but not for continuous. For continuous distributions, it is a very ordinary, opaque metric. The number tells us nothing (say about higher order moments).

* Hellinger is many orders away and is not informative.

* The algorithm for calculating Hellinger is full of judgment and is arbitrary. Why 90 days? Why 2/3 overlap? WHat about the fact that due to this overlap you will never hit a hellinger of 1.0? How many bins? Why? What happends when you change number of bins? etc.

* there is no reason to define delta like that for the four parameters.

* the traditional way of doing this is reliability analysis which uses 95 p.c. confidence intervals to test whether that range is contained within the acceptable tolerance regime.

* Inconsistent analysis. The NMI anaysis does not use Hellinger.

* The strength of our data and analysis is granularity and avoiding averaging. By using the full time-series for spatial stability we unintentionally ended up averaging over time. In temporal analysis, we did not average over the qubits. Similarly, in spatial analysis we should not average over time. Else it is inconsistent. Hence, the remedy is to change this to across-time and point-in-time analysis.