The authors discussed the need for operational measures that can be used in farm management and animal breeding, to improve resilience on a large scale across different farm types and livestock species. However, the main discussion was based on large and well-structured datasets, very common in dairy cattle, which may not be the reality for other breeds and/or livestock species. Also, for some species like chickens and pigs, improving general resilience may be less important than specific resilience, because of the homogeneity of the production environment. Of course, the discussion brought up by the authors is super relevant, because there is an appeal to breed for it, but we do not know the exact measure to use (if there is only one way for all species and/or breeds), as described by the authors, which compromises its practical application. Many resilience studies have used precision livestock farming technologies, such as automatic milking systems, to find a way to achieve resilience (e.g. Poppe et al. (2020)), and these data are the best for it due to the complex and multifactorial nature of resilience. A scoring system, as proposed, based on readily available farm data, might be a good measure because it can account for several disturbances at the same time, and the validation will tell us if the items and weights considered were enough to measure full resilience. Thus, I thank the authors for their contribution to the great global discussion about resilience.

It might be interesting to deepen the discussion about general and specified resilience, like disease resilience. Such discussion can give the audience the dimension of resilience and if only one measure will be needed to get resilient animals for all disturbances.

Another point that I think might be interesting is to make a scheme to draw the case study, including the measurements used to make de ranking, adjustments, validation measure, etc. Thus, it may be easier to understand the entire proposed scheme for measuring resilience.

Below are my comments and minor changes (according to the line) in this manuscript :

Line 36: I suggest to include something like "cannot be uniquely measured and selected." to emphasize that there is no single direct measure of resilience and therefore a multiple-trait selection should cover full resilience.

Line 60: I suggest to replace "puts pressures" with "put pressure".

Line 66: I suggest to replace "It is a characteristic that is important to" with "It is a trait that is important for".

Line 66-68: Cannot a more uniform production be an important point valued by the farmers and other livestock stakeholders? Uniformity of production can also be a consequence of a more resilient animal, especially for meat farmers.

Line 68: What is wrong with existing measures, such as those reported by Berghof et al.? It may be relevant to emphasize the reasons to continue the search for a suitable resilience measure, what is wrong or what is missing in the measures already proposed?

Also, based on the concept from the literature, can resilience be seen and improved in general or according to the type of environmental disturbance? This can lead the audience to think that we cannot have a resilient animal for all types of disturbances based on a single trait selection, it is necessary to select multiple traits to capture full resilience. These points can be better explored further latter in the text.

Line 70: I suggest to replace "industry in putting in place genetic selection" with "industry in implementing genetic selection".

Line 75: I suggest to replace "measuring resilience tackles" with "measuring resilience, tackles"

Lines 78-79: I suggest to replace "level, (3) how" with "level; and (3) how".

Line 115-119: So, robustness is long-term resilience?

Line 127: La Fontaine (1668) citation isn't in the reference section.

Line 140: Are there studies that can provide some evidence of the biological mechanisms behind robustness and resilience, based on the definition you provide? It can also elucidate the potential use of resilience measures to specific types of challenges.

Line 150: I suggest to replace "than of resilience" with "than resilience".

Line 173: Based on the discussion in this section, I wonder if it is possible to measure resilience right after a known environmental disturbance, such as disease or heat stress. Thus, not only long-term consequences and measures can be used to evaluate resilience. Another positive impact of this short-term measure is to capture resilience related to the disturbance that just happened and not the "noise" caused by factors other than resilience.

Line 176-177: I suggest to replace "an animal with good resilience" with "a resilient animal". In addition, is resilience only related to the lifetime or can it also be related to more uniform production?

Line 184-185: I suggest to replace "an animal with good resilience" with "a 'generally' resilient animal" because you are talking of all types of disturbances. I also suggest to replace "an animal with poor resilience" with "a less resilient animal" and replace all good and poor resilience by more and less resilient animals.

Line 194-196: Can genomic selection be an alternative to it? Traits measured late in the life and those measured after death are among the traits benefited by genomic selection.

Line 221: I suggest putting the abbreviation for precision livestock farming, since its abbreviation was only mentioned in the abstract, and PLF is used in line 235. Thinking about PLF technologies, which capture performance and behavior at the individual animal level, which tools can we use to measure aspects from the environment, like temperature and humidity? These environmental measures can be combined with PLF data to reveal potential disturbances.

Line 272: I suggest replace "that any one single measure" with "that a single measure".

Line 289: I suggest to add "for" in the middle of "rationale how".

Line 301-302: I suggest to replace "Despite it not being the only factor affecting it" with "Although it is not the only factor affecting it".

Line 369: What is nadir?

Line 421: Is there any special reason for choosing scheme 'B'?

Line 421: Are all 1800 cows from the same farm?

Line 446-449: Can you show this relationship through graphics? Comparing the best and the worst cow in the ranking on the same graphic.

Line 449: Are there any adjustments (like those made before) to correct the resilience reference measure for the farm and to overcome this issue? You cited locally-relevant economic, breeding, and management contexts; can they be used to make this adjustment? I suppose the 1800 cows don't come from the same farm, that is why you discover it, isn't it?