Dear PCI Editor,

We thank the reviewers for their comments and the time and trouble they have taken with our manuscript. We are delighted that all the reviewers consider that our manuscript brings a valuable contribution to the resilience literature. We have responded to all the comments as described below and adjusted the text accordingly (highlighted changes in yellow in the manuscript). We hope that these changes meet with your approval and allow the paper to be published.

With respect to the Editors comment "PCI requires authors to make the raw data, codes and scripts available. Reading the paper, the origin of the data used in the example (1800 lactations from Table 2) was not clear to me. Could you clarify this?": Given the simplicity of the equation for calculating the resilience scores there is no dedicated script, with the information provided in Table 1 anyone can easily calculate resilience scores. However, in the legend for Table 1 we have included a reference to the Adriaens et al paper 2020 which contains an example of the calculation for an example cow. With respect to the request for raw data we are somewhat perplexed as there are no results, for example showing a link between a resilience proxy and the resilience score, presented in the paper. The data in Table 2 are simple means of descriptive data for cows (no. of lactations, no. of inseminations, 305d MY, etc) that have no intrinsic value as such. We therefore did not add these data.

Yours sincerely,

Nic Friggens, on behalf of all the authors

Reviewer 1

Dear editor,

This is a very well written manuscript that presents and discusses important concepts and applications of animal resilience. This is a great contribution to the literature, but I suggest that the authors clarify some points:

1) They have not addressed well the relationship between productive efficiency and resilience or robustness. For example, in Figure 1 it seems to be assumed that robustness is always linked to lower production levels. This is likely the case for most animals, but there are multiple studies showing that there is still within-population variability. Based on the Figure presented, one might assume that it could be more economically feasible to improve the environmental conditions instead of breeding for more robust animals. I am aware that this is not feasible in some production systems, but my point is: we should not disregard the production level when evaluating robustness and resilience or assume that more resilient/robust animals will always be less productive. In my opinion, more direct indicators of productivity should be taken into account.

AUTH: We agree that there is within-population variability in robustness, that it is not inevitably linked to a lower production level, and that consequently it should not be ignored when evaluating

robustness, and to a lesser extent for animal resilience. We have added the following at L132 (original manuscript numbering) to indicate this but as this is not the main thrust of the paper we have not elaborated too much on this point: “It has been shown that there is significant genetic variation in these traits, even after adjustment for production level (Rostellato et al., 2021), which suggests that selection of the optimal blend of resilience and robustness is feasible.”

2) I question the novelty of the strategy proposed to quantify resilience in dairy cattle. As the authors are likely aware, dairy cattle selection indexes include all the variables suggested in their resilience index and many other indicators of resilience: age at first calving, interval from first to last insemination, calving internal, milk production, health traits (somatic cell count and clinical mastitis, metabolic diseases, reproductive disorders), longevity/stayability, and many others. The weighting approach suggested by the authors might actually be unfair to certain animals evaluated during more unfavourable conditions, which doesnt seem to be taken into account in their weighting strategy. For example, in an extensive production system, it is more likely that cows will be more fertile (less inseminations required) during the seasons with greater pasture quality or lower temperatures. However, not all cows in the herd will be in a breeding stage at the same time. Will their ranking approach take all these factors into account? For example, in L399, they mention an adjustnemtn for "herd average", which is not very representative of all these environmental factors influencing animal performance. A genetic evaluation and selection index would consider all these factors and therefore, it would seem to be a better strategy. In other words, wouldnt it be better to create a selection sub-index for resileince instead of ranking the animals based on their "almost unadjusted" phenotypic records alone? Especially from a breeding perspective, it is not clear to me how this ranking strategy is more benefitial compared to the selection indexes already in place in most countries around the world.

AUTH> The issue here is for what purpose one wishes to quantify resilience? We agree that for genetic evaluation and selection a multi-trait index, including herd-year-season adjustments, as described by the reviewer would be obvious. Indeed, this is mentioned at a number of places in the manuscript, e.g. L205-207. However, for evaluating potential resilience proxies, datasets that contain reliable reference measures in addition to relevant sensor data are typically too small for a genetic evaluation approach. We have added the following to indicate this: “it is analogous to a multi-trait index in genetic evaluation” at L402, we have also indicated that the ranking approach could be extended to include herd factor adjustments. We believe that these changes together with the existing phrase that “Ideally, the weights to be used could be informed by population level analyses that have quantified the hazard ratios for these different factors in epidemiological and survival analyses (e.g. Rostellato et al., 2021)” are sufficient to make the point.

3) In Table 2, how would they take culling reason into account? For instance, I dont think a cow that was culled due to temperament/behavior should be considered less resilient than a cow due with fertility issues.

AUTH> Culling reasons were not explicitly taken into account because this information was not readily available. Even when available it is never easy to define a clear threshold between resilience-related and other causes of culling. We have indicated at L403 that the approach could be refined if additional, reliable, information is available.

4) L325: the authors discuss about the challenges of using disease event records, but no clear solution or potential alternatives are indicated at the end. I suggest adding a closing sentence to this section.

AUTH> Added as suggested: “Whilst keeping in mind the above issues, high quality and consistent disease recording, coupled with good cross-validation, would provide a reference measure against which to evaluate resilience proxies.”

L68: ...herd for a long time...

AUTH> changed as suggested

L86: Please cite some examples of these studies.

AUTH> Cited Berghof et al and Colditz and Hine as these also contain further references

L94: ...from an evolutionary...

AUTH> changed as suggested

L122: I disagree that a "good and stable environment requires NEITHER resilience or robustness...". It is almost impossible to have a production system with no environmental challenges. Maybe this could be rephrased as "a good and stable environment is less depedent on the resilience and robusteness level of the animals...".

AUTH> Modified accordingly to: “A good and stable environment requires very little resilience or robustness”

L160: I would also add, what are the trade-offs between resilience and productive efficiency? Can we simultaneously breed for more resilient and productive efficient animals?

AUTH> We agree with the reviewer that this is an important question *in the context of selecting for* resilience and efficiency. However, at this point in the text the questions relate to *the problems of quantifying* resilience. Thus, we have not included the additional question here

L242: ...have been...

AUTH> changed as suggested

Reviewed by anonymous reviewer, 25 Oct 2021 12:53

Reviewer 2

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.

AUTH> This is indeed an important discussion point, that we have alluded to in the original manuscript (e.g. L278-286), and there are many and varied views on the general vs specific resilience debate. We did not feel that such a discussion could be done full justice within the present manuscript. However, we have added the following at L480 in order to draw attention to the issue: "Clearly, the choice of indicators used, and to some extent the reference measures, will depend upon whether one seeks to develop general resilience measures or to develop measures of resilience to a specific type of challenge."

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.

AUTH> We agree with the reviewer and have accordingly provided a Figure 4 with this scheme.

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.

AUTH> In line with the suggestion, changed to "cannot be directly measured and selected for"

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

AUTH> Changed as suggested

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

AUTH> Changed as suggested

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.

AUTH> We agree that uniformity of production can be a consequence of a more resilient animal but we do not see how this particular sentence implies the contrary. However, as we are reporting farmer impressions from a specific paper we have not changed this sentence.

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?

AUTH> The sentence citing Berghof concerns the opportunities to breed for resilience and does not discuss the indictors that Berghof proposed. Accordingly, we have not changed this sentence. To respond to the reviewers question, nothing is wrong with the measures proposed by Berghof et al. However, they only proposed the measures and did not validate their proposed measures against the accumulated consequences of different levels of resilience.

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.

AUTH> See response to general comments

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

AUTH> changed as suggested

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

AUTH> changed as suggested

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

AUTH> changed as suggested

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

AUTH> For the purposes of clarifying the confusion in the literature between the terms robustness and resilience we believe it is useful to make the distinction between the biological mechanisms that create the emergent property of long-term “stableness” (robustness) from the biological mechanisms that underpin short-term responsiveness to environmental variability (resilience). In the same way, in physiology, one talks about homeostasis and homeorhesis as distinct properties. Seen from this perspective one would say that robustness is not long-term resilience. However, as there is almost certainly some overlap between the biological mechanisms of robustness and the biological mechanisms of resilience we cannot say that they are truly distinct processes. A key example is body fat reserves, these can be mobilised as a short-term response to lack of food, i.e. homeostasis. They are also used ‘strategically’ in support of reproductive investment, i.e. homeorhesis. From this perspective we would say that resilience, whilst being a useful ability in its own right, contributes to higher-level robustness.

Given the need for clarity on the difference between robustness and resilience at this point in the manuscript, no change has been made to the text.

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

AUTH> reference added

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.

AUTH> There are many studies that have examined different physiological mechanisms underpinning robustness and resilience. We have inserted a reference to one of these.

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

AUTH> changed as suggested to “than to the concept of 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.

AUTH> we agree with the reviewer that measuring resilience in conjunction with known environmental disturbances would be useful. Indeed, we have been and are actively involved in experiments to do this (the refs Friggens et al 2016, Sadoul et al. 2015, 2017 are exactly that). However, this does not of itself resolve the issue of whether any particular feature of the time-course of response/recovery that we chose to measure actually confers a resilience benefit to the animal. Even in these planned challenges there is still the need to validate the measures against the ground-truth of the accumulated benefits to the animal. No change made to the text at this point.

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?

AUTH> Changed as suggested. With respect to the second point see the arguments in our response to the comment at L173

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.

AUTH> changed as suggested

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.

AUTH> At this point in the text we specifically refer to monitoring animals in the “management context” so not to the use of historical data (which comes in the following sentence). To make this clearer we changes “management context” to “on-farm management context”. We have also indicated the use in genomic evaluation in the following sentence.

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.

AUTH> Changed as suggested

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

AUTH> We don’t think that the suggested change improves clarity. No change made

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

AUTH> changed as suggested

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”.

AUTH> changed as suggested

Line 369: What is nadir?

AUTH> the lowest point in a trajectory or curve, in this case of body weight.

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

AUTH> Not really, scheme B was developed by the team working with this dataset. No change made.

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

AUTH> Yes, this is now stated in the text

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

AUTH> We do not think this is necessary for clarifying the text at L446-449

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?

AUTH> In the study of Adriaens et al 2020 the resilience scores are calculated and standardized per farm, as it was important to see if general predictors could be found despite differences between farms independent of e.g. farm size / number of cows in the farm (these results are not from the study using 1800 cows). When comparing over farms, this is not relevant. As the point about adjusting for “non-resilience” factors is made in the following sentences no change made to the text.

Reviewer 3

The paper makes a valuable contribution to the rapidly growing body of literature in animal science on resilience of livestock to environmental disturbances. The authors suggest that resilience is a multifaceted latent construct that can be partially revealed through proxy measures. The authors examine the use of “big data” from high frequency records of production animals and their environment to construct a resilience index. The authors demonstrate that the index needs to be ground-truthed against long term performance standards such as longevity and calving interval to establish its operational utility. Their worked example illustrates that a resilience index can have utility for animal management and in breeding for resilience in dairy cows.

Recognizing resilience as a latent construct has important implications for our understanding of other latent constructs that have a prominent place in animal science. These constructs include animal welfare, temperament, homeostasis, homeorhesis, allostasis, and stress, among others. Like resilience, these conceptual constructs of multifaceted hidden states and processes that exist within the animal (viz: welfare, personality, autopoietic balance of non-equilibrium dynamic states and life history transitions) are operationalised through proxy measures of the animal and its environment. As the authors note, these various hidden states and biological processes need not be exclusive. Thus, resilience is not an aspect of the animal that is independent of other hidden states and processes. Data rich studies in animal science have the potential to advance not only operational measures of farm animal resilience but to also foster a broader understanding in evolutionary biology and human psychology of how hidden states can be revealed through the interactions of the animal with its environment.

A central problem the authors address in the paper is the methodology for constructing and weighting animal measurements into a resilience index. Indexes used in animal breeding (eg breeding objectives) and animal welfare (eg the Welfare Quality® project) are underpinned by biological concepts and by the ethical, cultural, aesthetic and economic values that we bring to the valorisation of component traits and measures used to construct an index. A major challenge for animal breeding and animal management/welfare is to develop a framework for identifying and articulating these values, that themselves often lie outside the awareness of the individuals participating in the evaluation. Thus studies on resilience, breeding objectives and animal welfare need to include a component of "values discovery and articulation" so that we can improve our understanding of the rationale and consequences of decisions and actions made (and not made) about animals by humans. Studies on resilience, as exemplified by this paper, have an important contribution to make in this endeavour.

The caption to Figure 1 might be improved by inclusion of words describing the left axis (something like environmental variability) and words describing what the red line in each graph represents (something like variation in an animal measure (eg milk yield) over time?). By my reading of the authors' account, an animal with high robustness would have relatively high performance in a poor environment, and an animal with low resilience would have relatively variable performance even in a relatively stable environment. In addition, a resilient animal would have relatively stable performance in a variable environment. The figure does not appear to portray these scenarios. Was it the intent of the authors to do so? By my reading, the top left and right graphs show animals of high resilience (low variability) with high robustness (left graph) and low robustness (right graph), and the bottom two graphs show animals of low resilience (high variability) with high robustness (left graph) and low robustness (right graph). Is this the interpretation the authors wish the reader to take away from the figure? Is it the intent of the authors to portray the 4-dimensional relationship between environment (with 2 dimensions, quality and stability) and animal performance (with 2 dimensions, robustness and resilience)? Further annotation of the figure may make this relationship clearer.

AUTH> Clearly the original description of Figure 1 was insufficient in the legend, we have updated it as requested to indicate the axes, and the meaning of the red lines. Given that these lines represent different patterns of environmental quality, and not animal performances in said environments, the interpretation of the Figure suggested by the reviewer is not correct. That is entirely our fault for insufficient description of the figure, which we apologise for and have now remedied both in the Figure legend and in the accompanying text.