To the editors and reviewers,

Thank you for the thorough review of the manuscript entitled: 'Electronic Nose Sensor Drift affects Diagnostic Reliability and Accuracy of Disease specific Algorithms", and for the opportunity to write a point-to-point response to the reviewers. We have made multiple adjustments to the manuscript and believe our results have now become more comprehensible to the readers.

Reviewers' comments

Reviewer 2:

The research undertaken in this work is important and up-to-date, so the authors presented interesting research. I am only surprised that a significant signal drift was observed after just 2 days of measurement using the e-nose blowing procedure to maintain the baseline. Please comment on this. Below are my other comments:

We thank the reviewer for their kind remarks on the importance of this topic.

line 60-62, which was the method of correction of the sensor signal drift. The authors did not describe it in the introduction, and this is the main goal of this research.

We agree with the reviewer that our method section should be made more clear. In the current study, we assessed sensor variation within the control group. As it is thought that healthy subjects have similar VOC profiles (with small variations due to medical, gender, BMI, smoking habits etc), we used this group for data normalization. We therefore assessed the mean relative resistance change per sensor per day for measurements within the control group, and corrected sensor output per sensor per day for IBD samples using this data. We have added this information to the methods section of this manuscript.

line 64-65, what exactly is an IBD biomarker? what are the chemicals?

Multiple studies have previously presented significant differences in VOC patterns between IBD patients and controls using a variety of techniques: eNose technology, Field Asymmetric Ion Mobility Spectrometry and Gas Chromatography Ion Mobility Spectrometry.
Most of these techniques focus on the entire spectrum of volatile organic compounds rather than on specific chemicals. Using all the volatile data, neural network analysis enables differentiation between patients with colonic diseases and controls, based on so called 'smell prints'. Advantage of these techniques is they are easy to perform, so use in clinical practice can be done by clinicians without the need for specialized laboratory skilled personnel. In addition, measurements are fast and costs are relatively low compared to chemical-specific techniques.

line136-137, why is the room temperature set to 18 degrees? Higher temperature generates higher concentrations of volatile organic compounds. Please comment on this.

We agree with the reviewer that higher temperatures allow for the generation of higher VOC concentrations. The study pipeline for the Cyranose 320© has been validated using this current workflow specifically to improve usability in clinical settings. For example for the analysis of fecal samples by consultants during a consult at the outpatient clinic, which could spare more (unnecessary) invasive diagnostic procedures. As differences between IBD patients and controls have been presented using this (clinically relevant) work flow, we opted to maintain this operational procedure for the current study.

line 147-152, for the e-nose to work properly, after exposing the sensors to a given smell, return to the original point by passing air over the sensors. It is not stated how long a single analysis takes and how long it takes to blow the e-nose with air for the sensors to return to their original state.

> We purged the sensors with air during 90 seconds after every measurement. This is the duration advised in the manual by the producers (Sensigent). Therefore, we believe this should be long enough for the residual VOCs to leave the sensors. In addition, baseline measurements with filtered air need to be performed every 5 to 6 measurements, and cross validation is performed within the apparatus, to avoid influence of residual VOCs on sensor outcome.

line 218-219, what were the sensors? providing information that sensors 6, 19, 24 are not very correlated does not tell the reader anything.

> The sensors consist of a variety of nanocomposite polymers. Exact sensor characteristics are within intellectual property of Sensigent. As stated by the producer of this apparatus, each sensor has specific conducting characteristics, which result in unique changes in electrical resistance per sensor when different compounds pass the array.
> Link: Cyranose 320 Select QA/QC Applications (sensigent.com)

> The fact that most sensor outcomes are correlated in the current study means that the fecal pattern is not as unique for each sensor as suggested by the producers. We have presented that most sensors (apart from sensor 6, 19 and 24) have similar response to a singular sample.
> We have added some thoughts on this outcome to the discussion section of the manuscript.

line 244-246, it is not clearly described what the signal drift correction consisted of. On the other hand, it is quite strange that the drift was noticeable after 2 days of measurement. Please comment what could be the reason for this situation. And please describe in detail what is the correction of signal drift.

> In our study, the sensors were purged with filtered air after every measurement. This is air tapped from a controlled air container in the hospital. In addition, we used a strict operation procedure in which the samples were prepared by the same person, using the sample gloves, the same thawing manner and all samples were measured in the same room. As we kept all circumstances stable, we believe that only drift of sensors may have affected sensor resistance over days. Nevertheless, we agree with the reviewer that the extent of drift was surprisingly large in this study.

> Even though sensor drift is a well-known disadvantage of eNose analysis, the extent to which this drift affects sensors over days has never been published in real-life cohort data. We do believe that raising awareness among researchers on the major effects sensor drift may have on diagnostic potential of diseases is very important. Should researched be unaware of this when performing clinical trials; samples may not be measurement in random order and this may results in false positive test characteristics caused by bias.

> As described above, we used the control group for data normalization. We assessed the mean relative resistance change per sensor per day for measurements within the control group, and corrected sensor output per sensor per day for IBD samples using this data.

line 272-274, two sentences are to be removed

> Thank you for this comment. We have removed this remaining part of the template.