We would like to thank Reviewer #1 for their very constructive comments which helped to further improve this manuscript. Before we address each "line comment" separately, we will respond to the main two concerns (1. Burial History/Conceptual Model and 2. Sample Size) Rev#1 raised in the "General Comment" section:

GENERAL COMMENTS

1. Burial History/Conceptual Model

We added two new sections to our manuscript about the approach and related hypotheses (Section 2), sedimentology (Section 3), and additional paragraphs and figure in order to address REV#1 comments regarding the missing information on burial history and the conceptual model. There are some misunderstandings which we hope to have clarified in the revised version:

a) While we agree with REF#1 that our manuscript will benefit from a more detailed description of the theoretical model and assumptions to better guide the reader on how to determine reworking, we would like to reiterate that our approach is not designed to estimate absolute ages from fluorescence values. We therefore think that conceptual models assuming linear or exponential changes, as described in Rev#1 are not useful, as they would suggest a potential to estimate absolute ages from fluorescence measurements. At least from the Oligocene onwards, the burial history of site U1356 is, like many other sites in Antarctica, dominated by repeated and partly abrupt retreats and advances of glaciers and therefore too complex to be described with a simple linear or exponential function (explained in detail in section 2 and 3).

b) Many comments made by REV#1 refer to how we compared mean values and how mean value and variance might be affected by the low sample size. There seems to be a misunderstanding as we did not compare the mean values of our fluorescence measurements. We instead avoided working with mean values as this might require a larger sample size. We might have confused REV#1 by using the term "mean" in the original manuscript for the fluorescence values produced by the imaging software. The software measures fluorescence by drawing a contour around the pollen grain and measuring multiple spots from this contour image. The program provides the mean value of the grain (i.e. each red fluorescence pixel value from around the contour of the grain). We clarified this in the revised manuscript and avoided using the term "mean". Finally, we would like to thank REV#1 for the suggestion to use variances as a tool to assess the degree of reworking. Unfortunately, our sample number is too low and variable to further explore this approach. However, we have added this suggestion to our discussion for future research.

2. Sample Size

We fully agree with Rev#1 that a much higher pollen count would be better and allow for a more detailed statistical analysis and discrimination between the

different sediment layers. Unfortunately, the Wilkes Land sediment samples have, like all other comparable Antarctic cores (e.g. ANDRILL or SHADRILL), a low pollen concentration. This is particularly the case for post Eocene samples where the pollen deposition is affected by both glacial sedimentation history and reduced pollen production on land. Measuring 500-1000 pollen grain, as suggested by Rev#1, is therefore unfortunately not possible.

However, our method is designed to work with low pollen counts and we are confident that our approach produces statistically robust results. We therefore disagree with REV#1' s comment that our work is "under-sampled" and sample number (n=30) "barely sufficient". The Mann-Whitney U Test is applicable to all samples sizes and may be used with as few as four measurements in each sample (*Fowler et al. 2009. Practical Statistics for Field Biology. 259 p.*) A methodological "disadvantage" of this test it that it reacts very sensitive to small samples sizes and normally indicates no differences if the sample size tends to be low. The fact that in our study the Mann-Whitney U Test shows statistically significant differences, despite our low sample size, gives us even higher confidence in our results and indirectly demonstrates that our sample size was actually large enough. We also think that grouping all pollen into a single group, in addition to test each single taxon, is indeed useful. With this approach we increased the sample size and, by excluding species variation, we tested one group and one explanatory variable only.

We are unsure about REV#1's comments on changes in mean value and variance as the Mann-Whitney U Test does not compare sample means. The confusion might be caused by the use of "mean values" in our manuscript, which refer to the image-processing software data (see comment above). We rephrased the relevant section to avoid further misunderstanding and also added additional paragraphs to the Method and Discussion section explaining the implications of low samples numbers for this approach as suggested by the reviewer.

LINE COMMENTS
(Line numbers are original line numbers before corrections)

Rev#1: L67 suggest "are subjected to" instead of "confronted"

Response: Done

Rev#1: L70 to follow from the previous point, this needs to be qualified with something along the lines of "if burial histories are the same, fluorescence change could be used as an indicator of age"

Response: Previous sentences have been revised to clarify the fluorescence colour can change with burial over geological timescales. However, the change of fluorescence colour cannot be used as a determination of age.

Rev#1: L83 "each should come with" = "we hypothesise"?

Response: Done

Rev#1: L105 A summary of what is known of the the burial history would be helpful here – is there any constraint or estimate of the amount eroded at eh disconformities – i.e. is there any possibility the Eocene pollen was buried to a greater depth before Oligocene time etc. . . if these sort of effects relate to only 10s of meters of extra burial, this is useful for the reader to know

Response: Following Rev#1 suggestions we added a new section, the Sedimentology (Section 3) of Wilkes Land, detailing the complex burial history, sedimentation rates and glacial influence throughout the core.

Rev#1: L132 modern name for Nothofagus fusca type trees has been changed to Fuscospora

Response: We are aware of the discussion started by Heenan & Smissen (2013) to split *Nothofagus* into four genera (Phytotaxa, 146 (1): 1–31). However, in order to be consistent with previous published palynological research at site U1356, Wilkes Land (e.g. Pross et al. 2012. Nature, 488, 73-77; Contreras e al. 2013. Rev Palaeobot Palyn., 197, 119-142) we prefer to keep the "old" genus name when describing the fossil record (see also discussion in Hill et al. 2015 Australian Systematic Botany, 28, 190–193).

Rev#1: *L147, This seems sensible. So why do you then combine them for your statistics?*

Response: The Pearson's correlation values were initially combined to determine which fluorescence values (red, blue, green, brightness, saturation and intensity) show a strong statistical correlation with age to further assess the fluorescence behavior of taxa. Red fluorescence showed the strongest statistical relationship with age to test against geological ages. Due to variations in the chemical composition of the exine affecting the fluorescence of grains, the red fluorescence statistical relationship was then determined for each taxon. There is a different reason to combine the red fluorescence values for the Mann-Whitney U test. As outlined under Sample Size in the General Comments section, the reason for combining the samples in the Mann-Whitney test was to increase sample size and test for one explanatory variable only.

Rev#1: L157, suggest remove "in situ"… All you can infer is they are "not obviously reworked"… that distinction is critical for this paper!

Response: In order to address Rev#1 concerns we replaced "in situ" in the manuscript by "non-reworked"

Reviewer: L160, I suggest that a clear description of conceptual models of reworking is really important about here – to provide some context and reason for the statistics in the next section…the reasons for wanting to know why correlations against age and significant difference between mean values must be laid out.

Response: Two additional sections detailing the burial history and conceptual model (Section 2 and 3) have been added.

Rev#1: L171, what do you mean "set" the p-value? Is this a threshold you have adopted to accept or reject a hypothesis? If so, at least this should be acknowledged/highlighted in Table 1 – perhaps bold the results with acceptable p-values?

Response: For determining if results are statistically relevant, we used the highest significance level of p-values (0.01), the 99 percentile. We deleted the "set threshold" and revised in Table 1 to bold results with p-values indicating the highest significant correlation (0.01).

Rev#1: L175, the meaning of U-values this test generates should be explained…If this is a threshold score, describe what it is, where it is from and what it means, and make this clear in your Table 2 – including same comments on p-values as above.

Response: Additional sentences have been added to section 4.3 to describe the meaning of U-values and a further explanation of U-values have been included in the Table 2 caption.

Rev#1: L175, Once you get into multiple sequential significance tests of this sort, perhaps describe why some sort of Bonferroni – type correction is not appropriate?

Response: The Bonferroni correction was not applied because this type of correction comes at the cost of increasing the probability of producing a false negative, i.e. reducing the statistical power of the test.

Rev#1: L180, Are these results tabulated?

Response: Following Rev#1 suggestion, we added a table with the ANOSIM results to the Supplementary Information (Table S2). In addition, section 3.3 has been reworded to clarify the ANOSIM results.

Rev#1: L200, could you plot these visual data, to demonstrate there really is an advantage to using the digital data? The ranges you quote seem to overlap about as much as the fluorescence red values? The visual data does not appear in your supplementary data?

Response: The number associated with Yeloff and Hunt (2005) colour chart classification has been added to the supplementary material (Table S1). The ranges listed are the visual colours identified through the observations of each pollen and spore grain measured for fluorescence. Additional sentences have been added to help explain the colour classification chart and the implication of visual fluorescence colours overlapping through time.

Rev#1: L230, The visual fluorescence data are not shown or plotted – how can you demonstrate that then that the digital measurements are better or worse at allowing differentiation of mean values between epochs?

Response: A table with results from visual assessment has been added to the supplementary material (Table S1). Additional sentences have been added to Section 6.1 to demonstrate why digital measurements are an advantage for differentiation of mean red values between epochs. We also showed that the subjective colour comparison of fluorescence alone could not distinguish between Oligocene and Miocene grains. We thereby demonstrated that the digital

measurement does in contrast to the visual assessment, not only produce objective and reproducible data but also more accurate results than the visual assessment. Additional sentences have been added to the discussion and result section to make this clearer.

Rev#1: L240, where are these results shown?

Response: ANOSIM table S2 has been added to the Supplementary Information.

Rev#1: L245, following burial models discussed above – it is really not clear to me how demonstrating a linear relationship as you have done is an indicator or otherwise of reworking. This needs to be described more clearly.

Response: REV#1 is correct: we used the correlation to select the best parameter. We deleted this sentence and rewrote the entire section to clarify.

Rev#1: L252, how? What is your threshold value or test to conclude that the sample or stage has enough in situ pollen for reconstruction?

Response: Our "threshold value" is a statistical significant difference in fluorescence colour between different pollen assemblages/depth. We added an improved and more detailed explanation of our approach in section 2 and the discussion.

Rev#1: L259, "applied" rather than "adhered"?

Response: Done.