We wish to thank the reviewer for his/her comments; our replies to concerns and suggestions for changes are given below (marked in red):

The paper is well structured but could better articulate its aim and conclusions in the introduction and abstract. For instance: “here we report on” seems rather a statement from a intercomparison report than a scientific paper. From the abstract the main issues at stake should be immediately clear. Also, a better link with application of the data should be made. Now it seems that validation of models is the main use, while the rest of the paper does not touch upon model activities anymore.

We have re-written the Abstract as follows:

Although atmospheric ^{222}Radon (^{222}Rn) activity concentration measurements are currently performed world-wide, they are being made by many different laboratories and with fundamentally different measurement principles, so compatibility issues can limit their utility for regional-to-global applications. Consequently, we conducted a European-wide ^{222}Rn/^{222}Rn progeny comparison study in order to evaluate the different measurement systems in use, determine potential systematic biases between them, and estimate correction factors that could be applied to harmonize data for their use as a tracer in atmospheric applications. Two compact portable Heidelberg Radon Monitors (HRM) were moved around to run for at least one month at each of the nine European measurement stations included in this comparison. Linear regressions between parallel data sets were calculated, yielding correction factors relative to the HRM ranging from 0.68 to 1.45. A calibration bias between ANSTO (Australian Nuclear Science and Technology Organisation) two-filter radon monitors and the HRM of ANSTO/HRM = 1.11±0.05 was found. Moreover, for the continental stations using one-filter systems that derive atmospheric ^{222}Rn activity concentrations from measured atmospheric progeny activity concentrations, preliminary ^{214}Po/^{222}Rn disequilibrium values were also estimated. Mean station-specific disequilibrium values between 0.8 at mountain sites (e.g. Schauinsland) and 0.9 at non-mountain sites for sampling heights around 20 to 30 m above ground level were determined. The respective corrections for calibration biases and disequilibrium derived in this study need to be applied to obtain a compatible European atmospheric ^{222}Rn data set for use in quantitative applications, such as regional model intercomparison and validation, or trace gas flux estimates with the Radon-Tracer-Method.

A remark on radon applications is also included in the first sentence of the second paragraph of the introduction.

I have one major point. The disequilibrium remains uncertain. It is claimed that Jacobi and André (1963) provide evidence that equilibrium effects are negligible above 50m a.g.l.. However, this will depend strongly on the atmospheric mixing characteristics. Especially the CBW comparison at 180 / 200 m provides a unique opportunity to sample conditionally based on atmospheric mixing characteristics (e.g. based on the potential temperature gradient along the tower). Under stably stratified conditions one would sample in the free atmosphere at 180 / 200 m, where the equilibrium assumption is safe. For well-mixed conditions this assumption is less certain. Therefore, a strong test would be to separately determine calibration factors for stable and unstable conditions. I do not know how this is related to Porstendörfer (1994), but for CBW unique data for mixing classification are available. Without further analysis I think it is too early to write: “no systematic relation between disequilibrium and meteorological conditions was identified in our data sets” (Page 18).

We agree (and have also stated this in the text) that the disequilibrium between ^{214}Po and ^{222}Rn does not only depend on height above ground but also on atmospheric mixing conditions, particularly close to the ground. We thank the reviewer for his/her suggestion to therefore investigate the differences between the ANSTO and the HRM systems at Cabauw at 180/200m separately for situations with stable and with unstable mixing conditions. The "true" calibration difference between HRM and ANSTO would then be derived from the comparison at stable conditions. For this evaluation we did, however, not use the meteorological parameters (i.e. potential temperature) from the Cabauw tower but directly the ANSTO radon measurements from 20m and 200m, i.e. the ratio of ANSTO-measured ^{222}Rn activity concentrations between 200m and 20m. A ratio close to 1 would then indicate well-mixed (unstable) conditions while ratios much smaller than 1 would be normally expected during stable conditions (e.g., during night).

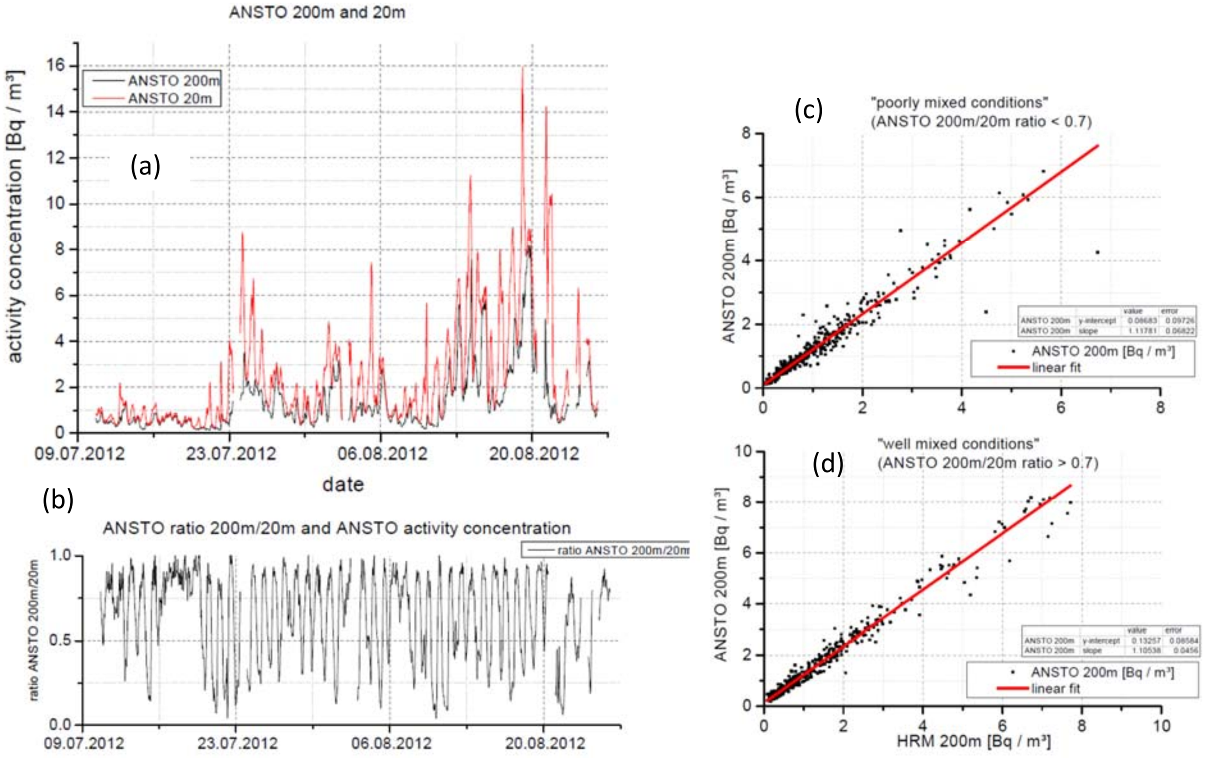

Figure 1: (a) ^{222}Rn activity concentrations at Cabauw tower observed in 200m (black) and in 20m (red) above ground, (b) activity concentration ratio between 200 and 20m (c) correlation between ANSTO-measured ^{222}Rn and HRM-measured ^{214}Po during unstable situations (at activity ratios >0.7) and (d) during stable situations (at activity ratios <0.7).

Figure 1a shows the records of activity concentrations at Cabauw at both heights, while Figure 1b displays the respective activity ratios. As expected during summer, we observe large diurnal variations of the activity concentration ratio, ranging from about 0.1 to 1. We have then selected the situations with 200m/20m ratios above and below 0.7 and plotted the ANSTO-measured ^{222}Rn activities versus the HRM-measured ^{214}Po activities separately for the two different stability situations in Figures 1c and 1d. The slopes for stable and unstable conditions are not significantly different, i.e. ANSTO/HRM = 1.12 +- 0.07 for stable conditions and ANSTO/HRM = 1.11 +- 0.05 for well-mixed conditions. These slopes are also not significantly different from the mean value reported in our manuscript. From this test we think we can now indeed safely state that "no systematic relation between disequilibrium and meteorological conditions was identified in our data sets", at least for measurements around 200m height above ground. We added a respective remark in Section 4.1.

Anyhow, a simple correction for disequilibrium effects based only on height seems a rather crude approach. Concerning its use as "tracer" to validate atmospheric transport and boundary layer mixing, an option could be to simulate 222Rn progeny in models such that disequilibrium effects are modelled instead of a priori corrected for in a data set.

We agree that our approach is crude, however, at most measurement sites ^{222}Rn and other trace gas sampling is above 10m where an uncertainty of the disequilibrium estimate of 30-50% would still be tolerable.

Minor issues:

Abstract: From the abstract it should be clear why a correction is needed. Different measurement principle? Preliminary 214 Po/222 Rn disequilibrium values: this comes out of the blue in the Abstract.

We have re-written the abstract, see above.

Page 2 line 24: it might be nice to mention which instruments are considered to be more accurate.

We added the following sentence: "If properly calibrated, monitors that sample ^{222}Rn directly are principally more accurate than those which sample aerosol-bound ^{222}Rn progeny, because no correction for disequilibrium is needed to estimate atmospheric ^{222}Rn activity concentration."

Page 3, line 4 "It may also occur" unclear what this refers to ... .disequilibrium?

Yes, we clarified this in the revised manuscript.

Page 3, line 17: to maximize the number of 218 Po progeny (T1/2 =3 min) collected ... unclear.

We re-formulated this in the revised manuscript, i.e. "make sure all newly formed progeny are captured on the second filter".

Page 4, line 23: Taking into account the flow rate through the filter, the filter efficiency, and the solid angle of the detector (which depends on the distance of the detector from the filter), enables calculation of the atmospheric 214 Po activity concentration.

Rewrite: e.g. from … the atmospheric 214 Po activity concentration can be calculated, taking into account the flow rate through the filter, the filter efficiency, and the solid angle of the detector (which depends on the distance of the detector from the filter).

Done so, thanks.

Page 8: in both, x and y component, in both the x and y components,

Changed.

Page 9, line 23: "Owing to the station's elevation, it is rarely reaching the atmospheric boundary layer". Unclear. Probably you want to say that the site will normally sample air from the free troposphere, although this likely depends on the time of day and the season.

Yes, we have re-formulated this.

Page 13, line 5: here, suddenly “progeny” changes to “progenies”. I do not know what is correct, but it should be consistent.

Must be “progeny”, we will correct this.

Page 15, line 12: between the two systems; I guess the two ANSTRO systems?

We clarified this, was meant “between the two different systems, ANSTO and HRM”.

Page 15, line 27: As expected ...; please add why this is expected (I guess at 35 m you expect larger disequilibrium effects, but better to articulate this once more).

In fact, we expect the slope (ANSTO/HRM) to be smaller at 35m than at 20m, as is the disequilibrium.

Page 16, line 16: 4.2 Calibration differences; should be 4.1

Yes, thanks for picking this up.

Page 18, line 18: (Capuana, 2016) reference missing in list.

We added the reference: Capuana, C.-A., Calibration of ionization chambers and comparison of two monitors for radon measurement (in German). BA Thesis, Institut für Umweltphysik, Heidelberg University, 2016.