**Author's reply to the comments of Reviewer 1:**

This paper reports on laudable efforts to harmonize DMA measurements of particle size distributions in long-term, ground-based, stationary measurements. It represents the results of European groups to reduce ambiguity in aerosol measurements, and may be considered as a reasonable first start at developing a truly international approach to DMA measurement standardization. That said, the harmonization standards reported here are incomplete, and will ultimately work against efforts to standardize or harmonize DMA measurements in general. The approach is highly prescriptive, which may be appropriate when all measurements are being performed for similar reasons. However, DMAs are used far more broadly than in long-term stationary measurements. Some of the prescriptive elements, such as measuring the size distribution at low relative humidity or time averaging data to report 1 hour average size distribution measurements will, if applied too broadly, confound efforts to understand important aspects of the atmospheric aerosol. Others are convenient approximations that may be incorrect in some important applications of DMAs, e.g., in source characterization studies, or in characterizing aerosols in a number of technological and fundamental studies where the gas is not air. I support the effort that this manuscript represents, but hope that it is a predecessor to a more open consideration of the full range of applications of mobility measurements. In the discussion below, I raise a number of points in the order that they appear in the manuscript. I finish with a number of points that are omitted entirely from the paper.

**Thank you very much for your critical comments. We are completely aware of many other applications of mobility size spectrometers. However, this paper is not meant be give an overall recommendation for all kind of applications. As the title already emphasizes; the recommendations are only valid for long-term atmospheric observations e.g. at GAW- or European super-sites to determine the number size distributions in a comparable manner.**

One critical issue is that, in addition to the references that may be of immediate use according to the proposed prescription, primary references should be cited. For example, on p. 5530, l. 8, Wiedensohler's fit to the Fuchs/ Hoppel & Frick model is cited, but not the primary references. The Wiedensohler fit is a very powerful tool which I also use, but it only applies to dry air under relatively mild conditions. Variations in relative humidity, temperature, pressure, or gas composition may alter the distribution which, throughout the paper, is incorrectly labeled an equilibrium charge distribution.

**The philosophy to determine comparable number size distribution is to measure them at dry conditions and at room temperature as described in the recommendations. The only question left is if another pressure would change the bipolar charge distribution. There is no paper in the literature, which describes such empirical study. However, the mobility of the ion is a function of 1/p. The "Gunn bipolar charge distribution" uses e.g. the ratio mobility of positive to negative ions to calculate the fraction of the different charging states; and this would not lead to another bipolar charge distribution since the pressure dependency is cancelled out. Presently, it is thus common to use the known bipolar charge distribution for different atmospheric pressures.**

It is actually a representation of a steady-state charge distribution which differs dramatically from that which one would predict at thermodynamic equilibrium, especially in the low nanometer size range. The charge distribution is described as an equilibrium distribution at several points in the text, especially on p. 5528. This will suggest to some readers that the Boltzmann equilibrium distribution should be applied, exacerbating a long standing problem in the interpretation of aerosol mobility measurements. The authors should err on the side of

caution in their choice of terminology, especially in a paper that is attempting to improve the quality of data and data reporting.

**Thanks for this comment. Although the Boltzmann bipolar charge distribution is not used in mobility size spectrometers since 20 years anymore, we will mention in the text that the Fuchs and Gunn distribution is meant.**

For CPC counting efficiency curves, only Birmili and Hermann are cited; why not Agarwal and Sem, Stolzenburg and McMurry, and the many others who have provided data, and several of which are cited elsewhere in the paper.

**We will add some more references about CPC counting efficiencies. However, Birmili never published a paper about CPC counting efficiencies.**

For the DMA transfer function, only Birmili is cited. Birmili provides a mathematically convenient form of the transfer function that is not based upon the physics of the instrument. Others have provided important physically based models that can actually be used in a predictive sense, especially Mark Stolzenburg's thesis, whose model has been placed in the journal literature by Hagwood, and by Zhang and Flagan.

**We will add these references to the article as well. However, the data producer has to decide which inversion program is used and how to implement a DMA transfer function.**

Another noteworthy physically-based DMA transfer function model is that of de la Mora whose analysis has motivated much of the work on optimizing DMAs for measurements in the low nanometer size regime.

**You are correct, but this kind of DMA is not build for atmospheric long-term operation, and therefore it is not considered here.**

The focus of this manuscript appears to be on accumulation mode aerosols, but much future work will address the nucleation mode, so the paper should provide guidance for researchers in that field.

**The focus here was indeed on the Aitken and accumulation mode range from 10 to 800nm. Presently, there are only very few instruments in permanent use at observatories, which measure also the nucleation mode range. This measurement requires an additional nano-DMA.**

The authors contend on p. 5527 that no comprehensive intercomparison of custom-built mobility spectrometers has been published to date. There have, in fact, been many intercomparisons – some led by authors to this paper. The statement may be correct if one takes an exceptionally narrow definition for the word comprehensive, limiting it to studies that include all of the instruments used in some narrowly defined range of studies. This study is, in fact, not comprehensive in that a number of DMAs that are used by groups around the world are not included. There is no need for such claims to be first, or to cover new ground. This study reports a rigorous intercomparison of a wide range of instruments. That would be a sufficient claim for this rapidly changing field.

**I think the reviewer is not correct in this case. There was never such an effort of a network bringing mobility size spectrometers three-times together. Standardizing them and finding a way to have them comparable in the size range larger than 20 nm. However, we will cancel the word comprehensive to avoid a misunderstanding.**

The paper also claims that there have been no comparison of inversion routines. There have also been comparisons of data inversion algorithms, although they may not have been focused entirely on mobility analyzers. Many papers describing new algorithms include comparisons with the prior algorithms. I find little value in the intercomparison of inversion algorithms presented in this paper since no useful information is provided about the algorithms. They are merely black boxes supplied by different manufacturers or used by different groups. The nature of the algorithms is not discussed at all. This is an important issue.

**The reason to do this intercomparison was not to highlight this inversion intercomparison. The reason was to find out if differences between different mobility size spectrometers with different inversion routines produce the same size distribution by providing all of them with the same measured mobility distribution. There may have been intercomparisons before, but never in this extent. We will rephrase the sentence saying that there was "no such comparison study" before.**

In other fields where particle measurements are key, many manufacturers produce instruments that do the inversion internally and only report the inferred size distribution. The reported size distributions are often suspiciously free of apparent noise and remarkably consistent in the shapes of size distributions. Users cannot gain access to the raw data, nor can they examine the details of the inversion algorithms. Blind reliance on the manufacturers to produce quality instruments and data analysis has hampered the advance of the science in these fields. That same marketing philosophy may well enter the aerosol field if care is not taken to ensure that all users of aerosol instruments have full access to the raw data and data analysis methods. The recommendation that the raw data be reported is a step in the direction of ensuring such access, but more is needed. I will return to this point later.

**OK**

A key conclusion of this paper is that DMA measurements should be preformed only at relative humidities below about 40%. For a particular study of the type that motivated this paper, this may make sense, but DMA measurements are made for many reasons. Measurements made at such low humidity may not, for example, be directly comparable to impactor measurements. Nor will they directly relate to optical measurements – at least, not without making assumptions about the hygroscopic behavior of the particles being measured. That said, the authors are certainly correct in stipulating that the relative humidity should be measured and reported, as should all other operating parameters.

**Again, this paper focuses only on comparable long-term observations. This recommendation is not meant for intercomparison using an impactor with another humidity in an intensive field campaign. If somebody wants to do this additionally, it should be not a problem.**

Figure 1 suggests an approach for generating the dry aerosol, and for dealing with a number of other operating details that the harmonization standard recommends, e.g., using a NafionTM drier or a silica based diffusion drier, placement of RH and temperature sensors, flow meters, etc. One point in that figure that I find confusing is the inclusion of a mass flow meter in the recirculating sheath loop when they correctly note in the text that a volumetric flow measurement is preferable since it is volumetric flow rate, not mass flow rate that determines the DMA sizing characteristics.

**It is mentioned in the paper that the flow meter has to be calibrated. In the case of a mass flow meter, the volumetric flow has to be calibrated as function of the mass flow at the specific site.**

Recirculating sheath flow is, in theory, a good idea, but there are problems that are only obliquely addressed in the manuscript. Notably, many of the readily available pumps, and most of the blowers have a potential for leaks which can, as noted in the text, lead to biases in the measurements. The manuscript should note that, if a recirculating flow is to be used, the pump/blower must be sealed and leak testing should be part of the regular maintenance schedule for the instrument.

**This is clear; we will check the wording in the text.**

The authors do note that a critical orifice can be used to control the exhaust and sheath flows, but the figure suggests that it be used in a recycle loop. Most use of critical orifices for flow control is in an open loop operating mode which can be just as precise as the recirculating loop. When used in a recirculation loop, the temperature and absolute pressure at the entrance to the critical orifice must be known since the mass flow rate through the orifice is proportional to the gas density of the entering gas. If I interpret Fig. 1 correctly (the location of the orifice is ambiguous), this would require the addition of temperature and absolute pressure transducers between the pump and the orifice. Perhaps some discussion of that mode of operation should be included, along with recommendations about ensuring that the operation meets the desired standards.

**The temperature and pressure is known from the DMA pressure and the T-measurement in the excess flow. However, since there is always a calibrated mass flow meter or a capillary, the flow is known.**

Perhaps this is not a problem in Europe, but it is a problem in the USA: reference to ISO standards is of limited use to the research community since these are not accessible without paying an unreasonably large fee. ISO standards documents are not generally available in university libraries, so researchers may not have access to them. Any aspect of the cited ISO standards that is required as part of this harmonization standard should be reproduced in the paper or cited as an open literature reference.

**The formulas and constants taken from the ISO 15900 are given in the appendix.**

Section 4.1: Inversion routines. This section addresses a critical, and often overlooked part of DMA measurements, but provides remarkably little detail, and that is provided only for the noncommercial versions. The lack of transparent data analysis methods leaves many questions about the measurements. The discussion in this section provides a broad-brush treatment, but will be of little use to a new researcher trying to learn how to analyze their data properly. The comparison provided looks promising, but was it the result of a blind comparison, or were the inversions tuned to reproduce the desired results? I would argue that, like the details of the measurement, the details of the inversion are critical to the evaluation of any data and should be open.

**There was no tuning of the data at all. All owners of the comparison got the same mobility distribution incl. TSI and Grimm, and they provided the resulting size distribution.**

To this end, I would suggest that part of the harmonization should be a requirement that all data inversion algorithms be open source, readily accessible to all users and readers, and fully documented and validated. Furthermore, software version and all data analysis parameters need to be reported. Lacking this, the only way that data can be compared is to reanalyze the raw data, even if the data analysis routines are comparable under some circumstances. This will hold commercial instrument makers and users of many custom-built instruments to a higher standard than that to which they have been accustomed, but many artifacts have been generated with poorly documented and, often, inaccessible, data analysis

software. Too often, even the group that produced a data set cannot reproduce the data analysis some time later. The CPC counting efficiency curves and charging probabilities are critical components of the measurements. This is only part of the CPC characterization that should be reported. The counting efficiency curve is an adequate characterization of CPC performance only for DMA operation in the slow stepping mode. Proper interpretation of measurements from scanning mode operation (SMPS) or fast stepping mode operation (the so-called sequential method of GRIMM) require knowledge of the time response of the detector since particle counts that nominally are attributed to one particle mobility will, in actuality, result from that channel and several previous ones.

**I fully understand the concerns of the reviewer, but I am afraid this requirement is beyond what we can provide so far. We are at the beginning of a process which also includes sustainable capacity building. This means also to keep the expertise at each data provider. Our philosophy is also to compare instruments in intercomparison or on-site workshops as part of the quality assurance.**

Sizing performance can be validated using authentic size standards such as PSL. The approach proposed makes sense, including the emphasis on relatively large particles. However, some instruments may be optimized for size ranges that make the specification of an arbitrary size standard such as 200 nm PSL inapplicable to some instruments. Rather than a prescriptive statement that imposes a particular size standard, the paper should be descriptive in the approach. It may remain prescriptive in terms of documentation and the reporting of all relevant parameters, operating conditions, and data.

**Again, we focused on mobility size spectrometers for the size range 10-800 nm. The decision to use 200 nm is practical. 200 nm latex particles are produced by a nebulizer. So, if we don't want to have an interference with the salt peak below 100 nm, the particles have to be larger, but still to be small enough to have a sufficient number concentration.**

Number concentration standards remain the weak link in DMA measurements so the discussion of the particle number concentration is welcome. The need for a reference instrument is critical. The approach outlined for harmonizing these models will not ensure that the data are correct, but will ensure that the biases and errors are consistent. This does not solve the problem, it only postpones addressing it.

**Exactly**

The discussion of particle losses is reasonable for the most part. Losses in the DMA misses a critical point: at the transition from high voltage to grounded inlet or outlet, losses due to electrophoretic migration augment those due to diffusion (KOUSAKA et al. J Chem. Eng. Japan (1986) 19: 401-407; Zhang and Flagan. J Aerosol Sci. (1996) 27: 1179-1200). This is neglected in the discussion of losses in the DMA. Some, but not all, instrument designs take these additional losses into account.

**The reviewer is not correct at this point. All empirical loss functions of the DMAs suggested here include losses due to the transition from high voltage to grounded inlet or outlet automatically.**

To the list of data that should be included in the archival data records, I suggest some additions. First, the raw data that are recorded should be truly raw data, not data that has been processed through any kind of inversion software. Ideally, the absolute counts recorded in each channel, the flow rate through the detection region of the CPC, and the counting time interval would be recorded. This enables one to assess the role of counting statistics in the reported noise.

**I understand this point, but this we agreed on to avoid this. Especially, since some DMPS programs use a variable counting time for different size depending on the concentration. Therefore, we agreed on the number concentration in each bin. This is a decision we would not like to change. The advantage of this harmonized "raw data" is that everybody can recalculate the number size distribution without knowing any settings of the individual system.**

Reported CPC concentrations are less useful since they mask key parameters needed for proper data inversion. Basically, the recorded raw data should be sufficient to allow reanalysis of the data from scratch, thereby enabling validation of whatever data analysis/inversion routines are employed. In stepping mode operation, the delay time between voltage steps and the start of counting should be recorded since finite detector response time may lead to data analysis biases that can be corrected by data deconvolution. The response time of the plumbing (and electronics in the case of electrometer detection) between the DMA outlet and the detection point within the CPC or other detector may also lead to biases, so these data are also needed for proper data analysis. CPC dead time corrections and methods, and firmware and software versions for all instruments and data acquisition and instrument components should be included in the metadata.

**Everything what is useful should be in the metadata, I agree. Since we recommend frequent latex particle scans, the delay time should not be a problem.**

References to archival literature that provide such data may be appropriate.
Where multiple instruments are used in parallel, redundant channel data should not be removed from the raw data. I further question the recommendation that redundant channel data be removed from Level-1 EBAS data. This may mask problems with the data.

**Since level-1 are final and combined size distribution (inverted at once), there should not be redundant channels.**

Specific points: 1.Check all references and reference citations. At least one reference (Dahmann et al., 2001) is missing from the bibliography.

**Ok**

2.p. 5536, l. 9: "Due to the relative small polydisperse aerosol ..." is unclear. What is small?

**Will be rephrased**

3.p. 5538, l. 28: suggest changing "200 nm PSL size, which was" to "200 nm PSL size; this was"

**ok**

4.p. 5539, l. 6: spectrometers (plural)

**ok**

5.p. 5541, l. 10: delete "again." This is a different range than the previous shaded band.

**Ok**

6.p. 5541, l. 20: references missing

**We will check this**

7.p. 5543, l. 10: Delete section title 4.4.1, and combine next sentence with the previous paragraph. This is a continuation of the previous train of thought. Put the paragraph break before “A more widely ...”

**We will check this.**

8.p. 5544, l. 6: I suggest indicating that a manufacturer’s calibration curve “can be applied with caution.”

**OK**

9.p. 5545, l. 1: I suggest adding after parameters: “not process through any for of data analysis or inversion software.” This would ensure that raw data would be available for reanalysis. Partially inverted data may not be recoverable.

**We will check this**

10.p. 5549, l. 23: The suggestion that “Generally, the delay time is correct of up-and down-scans show the same result” can lead to misinterpretation of data if fast scanning or stepping is performed with a slow response detector. This comparison could be correct if performed on data that has been inverted using the full knowledge of the transient response of all aspects of the measurement. See Collins et al. (Aerosol Sci. Tech. 36: 1-9, 2002).

**We will give an additional explanation**