esac

European Space Astronomy Centre
P.O. Box 78
28691 Villanueva de la Cañada
Madrid
Spain
T +34 91 8131 100
F +34 91 8131 139
www.esa.int

DOCUMENT

XMM-Newton Calibration Requirements

Prepared by	**Norbert Schartel**
Reference	**XMM-PS-RQ-002**
Issue	**1**
Revision	**0**
Date of Issue	**019/06/2015**
Status	**Approved**
Document Type	**RS**
Distribution	

APPROVAL

Title XMM-Newton Calibration Requirements	
Issue 1	**Revision** 0
Author Norbert Schartel	**Date** 19/06/2015
Approved by XMM-Newton Users' Group	**Date 22/05/2015**

CHANGE LOG

Reason for change	Issue	Revision	Date

CHANGE RECORD

Issue 1	**Revision** 0		
Reason for change	**Date**	**Pages**	**Paragraph(s)**

Table of contents:

1 INTRODUCTION 4
2 METHOD 5
3 RESULTS OF THE SURVEY 7
4 SPECIFIC CALIBRATION ISSUES 10
4.1 Calibration of large observatories 10
4.2 Why have calibration requirements tightened? 10
4.3 Contamination and aging 11
4.3.1 Contamination evaluation 11
4.4 Absolute flux calibration 12
5 DETAILED CALIBRATION REQUIREMENTS 14
5.1 Scientific practice and detailed calibration requirements 14
5.2 Scientific reasoning for detailed calibration requirements 14
5.2.1 Astrometry: 15
5.2.2 Timing: 16
5.2.3 Absolute energy/wavelength reconstruction: 17
5.2.4 Line Spread Function and Response: 18
5.2.5 Effective Area and Flux: 19
5.2.6 Point Spread Function and Vignetting: 20
5.3 Detailed Calibration requirements 22
5.4 Comparison with pre-launch requirements 24
5.4.1 Astrometry: 24
5.4.2 Timing: 24
5.4.3 Absolute energy reconstruction: 24
5.4.4 Line Spread Function and Response: 24
5.4.5 Effective Area and Flux 24
5.4.6 Point Spread Function and vignetting 25
5.4.7 Summary pre-launch requirements and conclusions 25

1 INTRODUCTION

The calibration requirements of XMM-Newton are defined in the System Calibration Document (XMM-PS-GM-02 Issue 1) which was established in 1996, - about four years before the spacecraft was launched. In December 2014 XMM-Newton will be operational for 15 years. Since launch the calibration has continuously improved by the dedicated efforts by the scientists of the instrument teams led by the instrument PIs and the scientists at the SOC. However, these efforts are partly counterbalanced by changing instrument behaviour due to aging and tightening of the requirements due to the evolution of scientific questions and increasing amount of data available for scientific research.

In 2013 the Mission Manager, F. Jansen, explained the status of the calibration requirements to the XMM-Newton Users Group (UG) and asked for revised requirements. The UG felt the need for establishing a working group to update the original mission calibration requirements, taking into account current science goals and asked the Project Scientist to organize this revision.

The calibration requirements were revised by a working group consisting of eight scientists V. Braito, C. Done, L. Oskinova, A. Papitto, G. Pratt, N. Schartel, S. Vaughan and Y.-Y. Zhang. In the following the applied method, the results of the analysis and the drawn specific calibration requirements for XMM-Newton are described.

2 METHOD

The review of the calibration requirements was based on

- the current calibration and its documentation,
- a survey into the opinions of the most affected users
- input and comments from XMM-Newton SOC calibration scientists and experts (R. González-Riestra, M. Guainazzi, P. Rodriguez-Pascual, M. Santos-Lleo, M. Smith, M. Stuhlinger and A. Talavera-Iniesta)

The current calibration of XMM-Newton is described in the Calibration Status Document for each instrument, the Current Calibration File (CCF) release notes, the calibration documentation, the SAS release notes and the SAS documentation. The documentation is regularly updated reflecting changes. All relevant documentation is provided at the home page of the SOC (http://xmm2.esac.esa.int/external/xmm_sw_cal/calib/index.shtml and http://xmm.esac.esa.int/sas/current/sasversionchanges.shtml)

For the survey into the opinions of most affected users we contacted 49 Principal Investigators (PIs) of large programs performed during AO8 to AO12, 20 PIs of targets with high photon statistics, e.g. performed triggered observations and RGS-prime observations, and 3 scientists, either members of the UG or recommended by members of the UG. Given these selection criteria, scientists working in all scientific categories were asked to comment.

Although 72 scientists are only a small part of the XMM-Newton user community their replies are highly relevant for the current study. On one side the contacted scientists recently analysed data which corresponds to about 2 years of XMM-Newton observations. On the other side, significant fractions of the analysed datasets show very high photon statistics such that calibration systematics become significantly more important than for observations with low photon statistics.

Within the survey the scientists were asked to answer three questions:

1. Which calibration requirements are currently important for your work (and short example)?
2. Do you see calibration requirements which need to be improved (and what would scientifically be the merit)?
3. Would you be available to participate in a working group to establish revised calibration requirements for XMM-Newton?

33 of the contacted scientists answered. Given previous experiences with users' feedback, the PS assumes that the majority of scientists who have not answered are not facing significant obstacles with respect to the provided calibration. Given the received responses the PS asked seven scientists to join the review group. The scientists were selected such that their expertise reflect the different scientific disciplines as well as current calibration

challenges, e.g. stars (1), neutron stars (1), binary stars (2), active galactic nuclei (2) and clusters of galaxies (2).

The working group agreed on the following evaluation criteria for calibration requirements:

- Scientific Importance
  - Allows unique science (not be possible with any other mission)?
  - Limiting factor for certain kind of research?
  - How many observations are affected?
  - Can the required accuracy be defined?
- Organizational Questions
  - Must a problem be addressed by improving calibration or this problem is not really a task for improving calibration?
  - Is there a workaround (e.g. background from archive, dedicated calibration observations, observations with other missions etc.)?
- Technical Questions
  - Hard / instrumental limit?
  - Does a unique calibration solution exist e.g. do we have a standard candle?

The analysis work focused on the requirements specified in the answers in response to the survey. Established requirements which were not listed in the survey answers, e.g. absolute timing, were considered to be fulfilled and included after critical revision. The working group gained an overview of the current status of the calibration based on the available documentation and discussion with the calibration scientists R. Gonzalez-Riestra, M. Guainazzi, M. Smith, M. Stuhlinger and A. Talavera Iniesta. Discussions with the calibration scientists were fundamental for the working group to understand the on-going calibration work, the expected performance and the limitation of the instruments. The working group made simulations to understand how physical calibration requirements translate into scientific calibration requirements, e.g. how does a relative effective area deviation translate into a power-law slope or temperature. Finally, the working group discussed if there are practical ways to achieve theoretically possible goals, e.g. standard candle in X-ray.

3 RESULTS OF THE SURVEY

Given the nature of the survey and the relative general formulation of the questions the received input was naturally inhomogeneous. Most scientists described issues where they were running into problems with the analysis of large data sets and suggested improvements of calibration, SAS and analysis threads and providing further meta-data. As it was expected part of the comments referred to previous versions of the calibration and SAS. However, overall the input was of extremely high standard. The answers also demonstrate that the scientists overcome problems if a workaround exists, e.g. observations with other missions with different/better performance or by eye identification of spurious sources.

If we look into issues raised by three or more replies in the survey, four issues clearly stand out and the first three are considered as calibration requirements:

1. Discrepancies between the RGS, MOS and pn (effective area) calibration.
2. Calibration of pn timing and burst mode.
3. Absolute (effective area) calibration of EPIC.
4. Background determination for extended sources

Comments to the calibration requirements:

1. The present discrepancies between RGS, MOS, and pn affect all observations with high count statistics. Even deep field observations are affected by these discrepancies. The unique capability of XMM-Newton compared to other X-ray observatories is its high effective area especially at the iron Kα line. But this advantage can only be fully used when all EPIC instruments can be combined to do a joint analysis, and this can only be achieved with a reliable and consistent calibration. Thus, sufficiently good calibration is requested to allow the extraction of unique science, like reverberation mapping of the direct environment of the black holes, phase resolved spectral analysis of neutron stars, etc. Most probably due to imperfect calibration, at present the majority of papers on the high statistic sources discuss only data from either RGS, MOS or pn to avoid a discrepancy in results that may affect the referee process. Thus the calibration discrepancy hampers the benefits of gain in statistics. MOS, and especially RGS, have a much better energy resolution than pn, providing crucial constraints if all instruments could be simultaneously fitted. The XMM-Newton project stresses its capability of simultaneous observations with six instruments as a key feature of the mission. If this characteristic cannot be utilized for the observations with high statistics, then this severely impact the science output of the mission.
2. Galactic black holes and neutron stars in outbursts are amongst the brightest X-ray sources in the sky. At present XMM-Newton EPIC-pn is the only instrument with large effective area which can observe these bursts as all other missions/instruments strongly suffer pile-up at much lower flux levels. At the time of writing this document the main obstacle for spectral analysis of EPIC timing observation is the X-ray flux dependent energy recovery and structure in the PSF introducing artificial spectral features. Given the complex time characteristics and differences between different outbursts in combination with the improved pn-timing observing mode,

also in future major scientific results have to be expected from bursting Galactic sources.

3. The effective area uncertainty and uncertainties in its energy dependence are a limiting factor for progress in many areas, e.g. cosmological studies based on clusters of galaxies, studies utilizing simultaneous observations of XMM-Newton and high energy missions, neutron star radius determination, etc. The case of cosmological studies with clusters serves to illustrate the point. Cluster cosmology relies on an estimate of the total mass, which, assuming hydrostatic equilibrium, can be obtained from X-ray observation by using spatially resolved spectroscopy. This approach heavily relies on very precise flux and temperature measurements. There are currently discrepancies between the Planck CMB cosmology results and those obtained from Planck SZ number counts that use XMM-Newton measurements to derive the total mass from X-ray observations. If we link the systematic error of the absolute effective area to the differences in flux and (especially) temperature determination between MOS, pn and the Chandra instruments, then 50% of the discrepancy between the cosmological parameters obtained from clusters as compared to the CMB could simply be explained by a systematic calibration uncertainty alone. This difference is very crucial as Planck CMB observes at the highest redshifts ($z = 1089 \pm 0.1$) whereas XMM-Newton observations of SZ clusters determine the cosmological parameters locally ($z < 1.7$). The determination of cosmological parameters with X-ray measurements itself is extremely important as its error-ellipse is nearly orthogonal to the error ellipse of cosmic microwave background and supernovae based determinations. The cluster temperature measurements are extremely sensitive to the energy dependence of the effective area (continuum slope reconstruction) and differences of up to 25% (at 10 keV) in temperature may simply be caused by differences in calibration. This has a profound effect on systematics in cluster mass reconstruction. Similarly, the use of joint observations of XMM-Newton and high energy missions critically depends on the ability to extrapolate the XMM-Newton determined continuum towards higher energy range. In November 2013, the XMM-Newton OTAC allocated, for the second time, more than 10% of the A and B priority time to observations joint with NuSTAR. This illustrates how huge the scientific potential is. For instance, measurements of the Compton hump allow one to disentangle different interpretations of the origin of the hard X-ray emission in Galactic as well as supermassive black holes and to map the hot coronae via reverberation. On the other hand, the high effective area of the XMM-Newton EPIC cameras at 6 keV in combination with their spectral resolution allows one to constrain models and provide interpretations that would not be possible from NuSTAR observations alone.

These main calibration issues extracted from the answers in the response to the survey into the opinion of the most affected users are consistent with the UG recommendations.

4. The fourth point raised by a significant part of the scientists is the background determination for extended sources observed with MOS and pn. The background is found to be variable in time and spatially inhomogeneous. However, as EPIC

background determination is successfully performed by several teams who work with samples of clusters of galaxies or large areas in nearby galaxies we have not classified this as a calibration requirement. From the project point of view, the on-going improvements of the ESAS SAS can hopefully provide the required functionality.

In the following we comment shortly on issues raised which are not yet solved but which have not been converted into calibration requirements, either. The current PSF leads to spurious sources near bright sources. However, there are very few cases where the primary scientific goal is affected and eye-identification or statistical analysis provides workarounds. A measured PSF library should in principle be available as it was the input for the recent 2-dimensional PSF parameterization. Improvements of the vignetting calibrations are on-going. Background templates for pn timing-mode observations were generated within a Ph.D. work and hopefully these templates can be validated and made publicly accessible for all users. RGS hot pixel identification is a continuous task that is performed on a regular basis. Currently, scientists at SRON are working on the analysis of extended sources observed with RGS improving the background subtraction and developing optimal extraction methods. Hopefully the gained experiences can be made publicly available in a simple usable form.

4 SPECIFIC CALIBRATION ISSUES

4.1 Calibration of large observatories

The evolution of the scientific program of XMM-Newton being a large observatory like e.g. HST or Chandra is fundamentally different to situations typically found in missions that have a focused scientific goal and subsequent design. As illustrated below, the increase of the number of XMM-Newton observations leads to tighter constraints on the calibration requirements in order to allow scientific progress. In contrast, for many specialized observing facilities once established calibration can be frozen.

4.2 Why have calibration requirements tightened?

The X-ray community gathered together at ESAC from the 4th to the 6th of June 2007 for the workshop "XMM-Newton: The next decade" and discussed about the scientific perspectives for the XMM-Newton mission. As a result of the workshop discussions, the project and the Users' Group encouraged the community to submit large (≥300 ks) and, later, even very large proposals for observing programs (≥1 Ms). The fraction of submitted and accepted large programs was steadily increasing and saturated at the level of about 45% of the observing time since 2010. The photon statistics accumulated in many of the large and very large programs effectively tightens the calibration budget and leads to the situation in which calibration systematics limit the scientific interpretation and progress. This may be illustrated with three examples that follow:

a) There are now many AGN observed for 300 to 800 ks exposure time. Most of them are among the X-ray brightest of their class, which basically meant an increase of the signal-to-noise ratio of the data by a factor 3 with respect to the previous non-large program observations. Such observations do not allow fitting the MOS, pn and RGS data simultaneously as the calibration systematics of up to 20% may become dominate. Therefore only a part of the achievable science can be obtained.
b) A very large program with 3 Ms observing time was performed on a deep field (CDFS) in order to study AGN evolution. Again the scientific interpretation is limited by the calibration systematics (pn versus MOSs, vignetting correction) leading to systematic differences in source fluxes of 20%.
c) Several samples of clusters of galaxies have been observed as large programs with the aim to calibrate the scaling relations. Most of the individual observations are relatively short (~25 ks), such that the uncertainties on the spectral parameters derived from different XMM-Newton instruments overlap (although even here, discrepancies are apparent both between individual XMM-Newton instruments, and between XMM-Newton and other missions). Cluster scaling relations attempt to link

a single observable quantity (e.g., the gas mass, derived from the density, or the overall temperature of the intra-cluster medium) to the underlying total mass. Due to the heavy reliance on precise flux and temperature measurements, an energy-dependent bias in the effective area calibration leads directly to systematic mass (temperature) dependent biases in the mass-observable scaling relations. These biases can appear in slope, normalization, and scatter. They are degenerate both with each other and with the resulting cosmological constraints. Thus reducing any energy dependent bias in the effective area calibration is the key to the effective use of clusters of galaxies for high-precision cosmology in a way that is competitive to other cosmological probes.

4.3 Contamination and aging

The changes in the effective area of RGS and MOS are understood as consequence of accumulation of a layer of absorbing material. Given the obtained measurements the accumulation of absorbing material up to the current epoch appears to be nearly linear with time for MOS where MOS2 shows a larger slope than MOS1. In RGS the accumulation is described with an exponential time dependency and so far no difference between RGS 1 and RGS 2 is statistically required. It is unknown whether the accumulation of these absorption layers is really a steady process or it occurs in (small) steps. Consequently, the effective areas in the ranges affected by the absorbing material suffer from the uncertainty in the parameterisation of the time dependency. As it is not know how these trends continue (e.g. going into saturation?) both instruments require further monitoring nd calibration.

The MOS redistribution changes with time and shows spatial variations. There is currently no evidence for a time dependency of the pn redistribution other than the expected degradation of the charge transfer efficiency (CTE) which leads to a degradation of the energy resolution.

Given the age of the spacecraft in combination with the originally expected lifetime all aspects impacting the calibration may be expected to show aging effects. The calibration must be prepared for currently stable properties to start showing time dependencies, which require corrections in order to maintain the current calibration level.

4.3.1 Contamination evaluation

The details of the RGS calibration are described in de Vries et al. (accepted for publication in A&A, 2014, arXiv1410.5251D). The current RGS effective area calibration was established by J. Kaastra in 2005/2006 and implemented in 2006 (Pollock, XMM-CCF-REL-216). Whereas most features of the effective area could either be established from ground pre-flight measurements or explained with physical mechanisms which could be verified through measurements, e.g. oxygen edge, magnesium and fluorine edges, the overall effective area could not be established based on first principle considerations. For the effective area calibration it was assumed that the effective area based on first principles of RGS 2 is correct for the wavelength range from 10 to 25Å. The absolute calibration of the 10 to 25 Å range was established in comparison to Beppo-Sax LECS measurements of the Crab as standard candle. The effective area from 25 to 38 Å was determined through fitting power-law models to BL-Lac sources in the 10 to 25 Å range and assuming that this model is valid over the entire wavelength range. The difference between the extrapolation and the measurements were then expressed through 12th order Chebychev polynomials (de Vries et al., 2014, arXiv1410.5251D, accepted for publication in A&A). Monitoring of RX J1865.5-3754 and the Vela pulsar, both assumed to be stable X-ray sources, showed that the effective area is permanently decreasing at short wavelengths. The decrease can be modelled assuming a layer of absorbing hydrocarbons which increases with time (Pollock, 2007, XMM-CCF-REL-228; Pollock, 2010, XMM-CCF-REL-262, González-Riestra, 2014, XMM-CCF-REL-314).

As the carbon absorption edge is below the energy range of RGS the observed spectra do not allow to prove the interpretation of a growing layer of hydrocarbons. Also the exponential growth of the thickness of the absorption layer, plausible within an scenario that assumes out-gassing of hydrocarbons from the carbon fibres-structure of the telescope tubes, does not prove the absorption layer hypothesis itself, as exponential functions describe processes where the change is proportional to the amount (dN ~ N). The RGS calibration source may be covered by an additional layer of absorbing material and therefore cannot be used to ultimately decide on the issue (de Vries et al., 2015, A&A 573, A128). The contamination interpretation would be much more self-consistent if the effective area from 25 to 38 Å could have been established from first principles plus increasing layer of absorbing material only without requiring a first adjustment based on the extrapolation of BL-Lac spectra. Therefore, the hypothetical absorption scenario cannot be really discriminated from a material degeneration scenario showing an exponential dependency in energy and time. However, the modelling of the stable sources RX J1865.5-3754 and the Vela pulsar, used to measure the build-up of the contamination layer, show that the exponential form, expressed as carbon absorbing, is a valid parameterization of the changing effective area. The RGS-pn rectification (Pollock & González-Riestra, 2010, XMM-SOC-CAL-TN-0089), based on a large number of BL-Lac spectra modelled with pn and RGS simultaneously, provides an independent verification of the parameterization. After all these efforts calibration differences between EPIC and RGS still remain and have to be resolved.

4.4 Absolute flux calibration

EPIC pn, MOS and RGS show significant differences in the absolute flux calibration and the same holds if in addition Chandra or Suzaku measurements are considered. The systematic comparison of the different calibrations by the International Astronomical Consortium for High Energy Calibration (IACHEC) has significantly improved the individual calibrations of instruments, but it is not obvious if an absolute calibration can be established in this way.

The systematic comparison between X-ray measurements, SZ and gravitational lensing might have the potential to establish certain clusters of galaxies as standard candles. eROSITA plans to use a huge number of galaxy cluster measurements for a self-calibration, which may again have the potential to establish a better relative or absolute flux calibration. Should efforts lead to a better calibration or the establishment of standard candles for the X-ray regime during the lifetime of the mission, then these advances should be propagated into the XMM-Newton calibration such that the large repository of archived observations can be optimally scientifically used.

5 DETAILED CALIBRATION REQUIREMENTS

5.1 Scientific practice and detailed calibration requirements

The primary goal of XMM-Newton observations is the spectral energy distribution (SED) and temporal characterization of the primary target. Due to the low resolution of the X-ray instruments, specifically the EPIC cameras and to a much lower extent the RGS, the SED is expressed through physical and mathematical models which are compared with the obtained data. The scientific user of XMM-Newton data is interested in the most accurate reconstruction of the SED, where accurate continuum reconstructions as well as small features above and below the continuum are important. Through an accurate description of such features astrophysical progress is to be expected. For the reconstruction of the SED several components are required as listed in the detailed requirements below (RM, PSF, energy reconstruction, vignetting). In general the different components cannot be calibrated independently as they are degenerate to different degrees. The scientific interest is on the global reconstruction accuracy and, in principle, the individual (calibration) components are not so important.

It is cumbersome for every scientist to become familiar with a specific instrument and to perform state of the arte analysis. After this knowledge is established, scientists may analyse further sources without questioning the established method and accuracy. In addition, scientists often perform analysis work following descriptions from the literature. In general this is done without checking the documentation for changes. Many scientists of the community (blindly) expect that the established calibration (of a space missions) improves rather than decreases of the calibration accuracy, e.g. a scientist who has identified X-ray counterparts accepting a certain distance between the optical position and the X-ray position, will in general apply the same distance criteria for later analyse of other data sets without re-addressing the calibration documentation. Therefore, for the scientific return of the mission established calibration accuracies must be maintained and aging effects must be properly monitored and calibrated.

5.2 Scientific reasoning for detailed calibration requirements

In order to work out the detailed calibration requirements we summarize the level of accuracy currently reached and try to identify typical examples of important research topics to work out the requirement for the future, i.e. we are trying to predict what calibration requirements will be important for future science. Whereas the future science itself can not be predicted, it is possible to identify objects-classes and typical observing strategies, which most likely will also in future play a fundamental role for the scientific success of the XMM-Newton mission.

5.2.1 Astrometry:

After introducing the time-dependent boresight [XMM-CCF-REL-286, XMM-CCF-REL-315] the XMM-Newton pipeline/SAS data processing achieves an absolute astrometry accuracy ≤1.2" (1σ) for EPIC images and ≤1.4" (1σ) for OM images. These accuracy upper-limits were obtained through a comparison of the mean difference between the determined positions of point sources within an EPIC or OM images and the positions of infrared and optical counterparts which were taken from 2MASS, SDSS DR9 and USNO B1. They correspond to the shift applied during field rectification as applied in 3XMM-DR4 (Watson, M. G. et al., 2009, A&A 493, 339). All pipeline processed data up to mid 2014 were considered. The obtained upper limits consist of possible systematic errors, propagated statistical error of the individual point sources determinations and possible impact through wrongly identified counterparts.

We made a further evaluation of the position accuracy through cross-correlating positions of active galactic nuclei from the catalogue of Véron-Cetty & Véron (2010, A&A 518, 10) versus the EPIC positions provided by the 3XMM-DR4 (Watson, M. G. et al., 2009, A&A 493, 339). Assuming correct optical positions, we found 1σ = 1.65" and 2σ = 3.39" for the final positions after field rectification. The corresponding number before field rectification are 1σ = 2.16" and 2σ = 3.80". We can identify the difference between the numbers obtained before and after field rectification as an upper limit for a systematic calibration error, Δsys < 0.5". The error for the final position (i.e. after field rectification) is dominated by the statistical error of the source position determination. Therefore, a possible systematic error, corresponds to the 1/10 of the pixel-size in pn (FWHM is not determined for a radius below the pixel size) and about 1/10 of the FWHM of the point spread function of the X-ray telescopes.

The dominating contribution to the absolute position uncertainty is the FWHM in combination with the statistical error. The systematic error is about 1/3 of the statistical error (1σ). The achieved accuracy of the astrometry is sufficient to identify with a very high level of confidence extragalactic X-ray populations, e.g. AGN or clusters of galaxies, or SNR and high-mass X-ray binaries in nearby galaxies like in LMC, SMC (Sturm et al., 2013, A&A 558, 3) or M33 (Pietsch et al., 2004, A&A 426, 11). The position accuracy is insufficient to identify optical counterparts in extragalactic fields like (optical) dark GRBs or deep field observations. The scientific community generally uses Chandra observations to address astrometric questions wherever XMM-Newton accuracy is insufficient, e.g. Chandra surveyed systematically distant galaxies, whereas XMM-Newton mapped the nearby galaxies (LMC, SMC and M33) and the deepest XMM-Newton deep-field (Comastri et al., 2011, A&A 526, 9) was selected such that it coincides with one of the Chandra deep fields, which allows to take advantage of the established accurate X-ray positions for counterpart identification.

As the systematic position uncertainty contributes only ¼ to the total position uncertainty of individual sources further calibration/reduction of systematic uncertainty will not allow a significant increase of the number of biunique counterpart identifications. Chandra observations are generally used to mitigate the issue for scientifically important sources. However, as the established astrometry contains time dependent corrections, it is required that calibration maintains the established accuracy also in future. In the detailed requirements the astrometric accuracy is required for EPIC/OM images, i.e. the shifts of the field rectifications. This allows taking the pipeline products directly for test and verification.

5.2.2 Timing:

Considering the read-out time, the effective areas and the pile-up limit, pn shows superior performance in comparison to MOS with respect of any practical time characterization. Consequently, basically all publications of timing studies are using pn observations. The calibration of the pn timing and burst modes is established and verified twice per year through observations of the Crab. The timing calibration of the pn-image modes is established with the calibration of the timing mode modulo the readout time. After the regeneration of the time correlation files in November 2009 and the recalibration of the on-board oscillator (Saxton & Freyberg, 2013, XMM-CCF-REL-298) for non-linear degradation, the following calibration could be established: relative timing ($\Delta P/P$, pn timing): $< 4 \times 10^{-8}$, relative timing ($\Delta P/P$, pn burst): $\leq 1 \times 10^{-8}$, absolute timing: $\leq 140\mu s$. (The quoted numbers here reflect the dynamic range of the results of the Crab measurements including outliers as provided in Fig. 15 of Guainazzi, 2014 XMM-SOC-CAL-TN-0018.)

The relative timing accuracy is a prerequisite for phase-resolved spectroscopy. The duration of the pixel-accumulation/read-out time limits the frequency, which translates in objects/object-classes, for which phase-resolved spectroscopy can be performed. Examples, for phase-resolved spectroscopy are Geminga (Caraveo et al., 2003, Science 305, 376), PSR B0656+14 and PSR B1055-52 (De Luca et al., 2005, 2005, ApJ 623, 1051) or SGR 0418+5729 (Tiengo et al., 2013, Nature 500, 312). Future research intends to use pn burst/timing observations for reverberation mapping of Galactic black holes requiring the full usage on the limits of the cameras time resolution. The relative timing together with the absolute timing accuracy are limiting the ability of finding phase-coherent solutions for spin down measurements which translated in magnetic field determination. A wrong absolute timing calibration leads to wrong magnetic field determinations and consequent publications of wrong scientific results. Due to the combination of the various performance features pn has superior position with respect to the instruments on board of XMM-Newton but also with respect to other X-ray missions. Recent examples are the determination of the (lowest) magnetic field of central compact objects (PSR J1852+0040 (Halpern & Gotthelf, 2010, ApJ 709, 436; 1E1207.4,5209 (Halpern & Gotthelf, 2011, ApJ 733, L28) and of low-magnetic field magnetars (SGR 0418+5729 (Rea et al., 2010, Science 330, 944; Rea et al., 2013, ApJ 770, 65) or Swift J1822.3-1606 (Rea et al., 2012, ApJ 754, 27). It should be noted that both object-classes were not known some six years ago and that for both the magnetic field can only be determined based on X-ray measurements. There is a good chance that also in the next decade new compact objects or object classes will be detected and XMM-Newton will be essential for phase-resolved spectroscopy and determination of (low) magnetic fields

5.2.3 Absolute energy/wavelength reconstruction:

With the alignment of the wavelength scales of the RGS1 and RGS2 and correction for the sun-angle dependency (González-Riestra, 2013, XMM-CCF-REL-297) RGS reaches a wavelength accuracy of ±6 m Å (1σ) in the first and ±5 m Å in the second order (González-Riestra, 2015, XMM-SOC-CAL-TN-0030).

The improvement of the pn energy reconstruction is on-going for the last years. Main drivers are aging effects of the camera which required more sophisticated correction algorithms, but also the increase of available data, e.g. timing mode observations, allow a better assessment of the established calibration and its limitations. After the recalibration of the energy reconstruction for the timing mode (Guainazzi, 2014, XMM-CCF-REL-312 and 2013, XMM-CCF-REL-306), the latest CTI corrections (Smith, 2014, XMM-CCF-REL-323) and the new treatment of double events the following accuracy is reached for the boresight location: EPIC pn image (full-frame) for single events: -3±5 eV(1σ) for Al Kα at 1.5 keV and -2±7 eV(1σ) for Mn Kα at 5.9 keV, EPIC pn image (full-frame) for double events: -6±7 eV(1σ) for Al Kα at 1.5 keV and 2±11 eV(1σ) for Mn Kα at 5.9 keV and EPIC pn timing (single and double events): ±20 eV (1σ). Further improvements are expected with the on-going EPIC pn burst mode calibration and the introduction of the quiescent background correction for the image modes. We therefore expect to reach: EPIC pn image (single events) ±5 eV(1σ), EPIC pn image (double events) ±10 eV(1σ), EPIC pn timing / burst (single and double events) Energy ±20 eV (1σ) for the line energies as defined above. Given the latest update of the MOC CTI and gain (Stuhlinger, 2014, XMM-CCF-REL-317 and XMM-CCF-REL-318) the calibration achieves the following accuracy: EPIC MOS1 image: 2±2 eV(1σ) for Al Kα at 1.5 keV and 3±4 eV(1σ) for Mn Kα at 5.9 keV, EPIC MOS2 image: 2±2 eV(1σ) for Al Kα at 1.5 keV and 2±4 eV(1σ) for Mn Kα at 5.9 keV. During periods with strong solar flares the energy reconstruction accuracy is hampered for both EPIC cameras. In addition the MOS energy reconstruction is affected during eclipse periods. The wavelength reconstruction accuracy for the OM grisms for on-axis sources is as follows for the visible grism (on-axis) the wavelength accuracy of ≤7.5 Å (1σ) and for the UV grism ≤2.0 Å (1σ) (Talavera, 2011 XMM-SOC-CAL-TN-019 and personal communication). In the following we refer to energy reconstruction for all instruments including the RGS and the OM grisms.

The strictest requirement on the RGS energy reconstruction comes from stellar spectroscopy (for an overview see Güdel & Nazé, 2009, A&Arv 17, 309) and especially studies of plasma motions and kinetics (for an overview see Güdel & Nazé, 2010, SSRv 157, 211). In the latter the absolute energy reconstruction (in combination with the line width) is a limiting factor in the X-ray regime. The absolute energy reconstruction is of fundamental importance to characterize outflows in AGN (e.g. Kaastra et al., 2010, A&A 524, 37) and to compare X-ray absorption systems with UV absorption systems (e.g. Kaastra et al., 2014, Science 345, 64). The study of Mkn 509 allowed placing severe constraints on the wind launching area and its physical mechanism. For EPIC the ultra-fast outflows are a main finding (see Tombesi et al., 2012, MNRAS 422, L1 and references therein). Here the scientific importance is the estimate of the total mechanical energy output and its relation to the feedback constraining galaxy and cluster of galaxies evolution. An example of the OM grism usage can be found in Di Gesu et al. (2013, A&A 556, 94).

The accuracy of the energy reconstruction is a major limitation for the spectral reconstruction and background subtraction which may lead to artificial features, e.g. Díaz-Trigo et al. (2013, Nature 504, 260; 2014, arXiv1409.3406). The energy reconstruction is essential for the absolute flux determination and there the impact is strongly model dependent. For example, for a power-law spectrum with $\Gamma=2$ a shift of 15 eV in the energy reconstruction, which corresponds to the current 1σ accuracy for EPIC pn double events, leads to up to 3.1% difference in total flux for the 0.1 -10 keV energy band and a 10 eV shift, which corresponds to the current 1σ accuracy for EPIC pn single events, still leads to a 2.0% difference in the total flux estimate. We will comment on the importance and scientific potential of absolute flux reconstruction under point 6 in Sect. 5.3.

5.2.4 Line Spread Function and Response:

After updating the RGS line spread function (Pollock, 2011, XMM-CCF-REL-275, following an investigation by T. Raassen) the line spread function accuracy for on-axis point source is ΔFWHM/FWHM $\leq$10%, i.e. we can resolve lines if they are broader than 10% of the FWHM at a given wavelength. The response calibration accuracy for an on-axis source for pn singles is Δ FWHM $\leq$ 1eV for 1.0 keV (922/1022 eV) and Δ FWHM $\leq$ 7eV for 6.4 keV in 2001 and Δ FWHM $\leq$ 21eV for 6.4 keV in 2014, where the error of the latter measurement reflects limited photon statistics. For MOSs the response calibration accuracy is Δ FWHM $\leq$ 1.5 eV for 1.5 keV and Δ FWHM $\leq$ 3 eV for 5.9 keV. From highly absorbed sources the redistribution tail can be estimated to <1%. The EPIC response is subject to degeneration with detector age. The MOS response is regularly updated (Sembay et al, 2013, XMM-CCF-REL-208) whereas a time-dependent pn response was introduced in 2014.

A recent scientifically important result based on the line widths determination with RGS is the detection of velocity broadening of the intra-cluster, intra-group and even interstellar gas in the order to 400 km/s (Sanders & Fabian, 2013, MNRAS 429, 2727). The origin of hot halos of galaxies, the structure formation processes in clusters of galaxies, which should be observable in the outskirts of clusters and AGN feedback, leading to velocity broadening, turbulence and sloshing, are important research topics for the coming years. A further outstanding recent RGS result was the detection broadening of O (and Ne) lines of an ejecta knot in the Puppis A supernova remnant (Katsuda et al., 2013, ApJ 768, 182). The broadening is obtained to σ < 0.9 eV, indicating an upper limit of an oxygen temperature of 30 keV. Although such studies are a challenge for RGS resolution, it is important to maintain the calibration accuracy for progress in these areas. For EPIC the response is essential for the spectral reconstruction, where even bursting Galactic sources with >10^6 counts are analyzed, challenging the calibration, especially the response tail. The broadening of the Fe K_α will continue to be an important research topic. Observed mismatches and astro-physically unreasonable variability have led to the (on-going) introduction of time-dependent pn responses, which underlines the requirement to maintain the achieved accuracy.

5.2.5 *Effective Area and Flux:*

The accuracy of the effective area calibration of the X-ray instruments of XMM-Newton can only be estimated as no absolute standard or intrinsic calibration source is available. In order to evaluate the accuracy of the effective area calibration, fluxes and spectra of the different instruments are regularly compared (Stuhlinger et al., 2010, XMM-SOC-CAL-TN-0052). As all high energy missions lack an absolute calibration standard a consortium of scientists from different high energy missions (IACHEC) is trying to improve on the cross-calibration between the various instruments of the different satellites, e.g. Kettula et al. (2013, A&A 552, A47), Ishida et al. (2011, PASJ 63S, 657), Tsujimoto et al. (2011, A&A 525, A25), Weisskopf et al. (2010, ApJ 713, 912), Nevalainen et al. (2010, A&A 523, A22) and Sembay et al. (2010 AIPC 1248, 593S), see also Stuhlinger et al., XMM-SOC-CAL-TN-0052. The difference in the effective area between the EPIC instruments is of the order of 10%. For spectra with good signal-to-noise ratio, this difference is too high to analyse simultaneously the three EPIC spectra of a given source. The improvement of the effective area calibration and especially the reduction of the differences between the EPIC effective areas was the primary calibration goal during the last few years and is still on-going. The accuracy of the calibration of the RGS effective area with respect to the pn effective area is on the order of 5% (Pollock & Gonzalez-Riestra, 2010, XMM-COC-CAL-TN-0089). The rectification method developed for processing of RGS data in SAS provides a method to simultaneously analyse RGS and pn data, although currently not applicable due to calibration evolution. In comparison to Chandra instruments differences of up to 20% are observed (e.g. in Stuhlinger et al., 2010, XMM-SOC-CAL-TN-0052, B XMM-Newton/Chandra cross-calibration). Narrow features can in general be well-identified with the EPIC instruments indicating accuracy in the order of 1% for regions where the effective area is smooth. The accuracy of the effective area across an order-of-magnitude wide energy range (E with respect to 10 × E) is impacted by differences described above and therefore may be up to 10%. Comparing the different instruments we can estimate the absolute flux reconstruction to be of the order of 15% where pn shows the lowest value of all X-ray instruments (e.g. in Stuhlinger et al., 2010, XMM-SOC-CAL-TN-0052, §10 IACHEC cross-calibration). The flux reconstruction of OM is in a significantly better shape as standard stars are available. For image observations flux can be reconstructed better than 7% and the grism spectra allow a flux reconstruction better than 10% in regions with no overlapping orders.

As described above the effective area calibration is the main concern identified in the survey into the opinion of the users. A significant part of the large and very large programs aims to get high quality spectra and for most of these data pn and MOS cannot be used together due to the calibration differences, i.e. some 40% of the accumulated photons cannot be used. As example we might look to the (two) 400-ks datasets obtained for 1H0707-495. Although the scientific interpretation is controversially discussed between different groups, e.g. Kara et al. (2013, MNRAS 428, 2795), Dauser et al. (2012, MNRAS 422, 1914), Zoghbi et al. (2010, MNRAS 401, 2419), Miller et al. (2010, MNRAS 408 1928M) and Fabian et al. (2009, Nature 459, 540), the MOS data are basically ignored due to the calibration differences. It is essential that the full amount of data, pn and MOS can be fully used simultaneously, allowing highest statistics of spectral and/or temporal studies. Also during the next decade XMM-Newton will have the highest effective area of all X-ray observatories allowing studies not possible with any other instrument. Detection of emission lines down to 3% level, e.g. Diaz-Trigo et al. (2013, Nature, 504, 260) allows cutting edge science and must be preserved. The relative effective area across an order-of-magnitude wide energy band (E with respect to 10 × E, in the following we refer to this as relative effective area) and the absolute effective area are most important for several research lines. Here in some sense the requirements on the calibration accuracy have tightened in comparison to the pre-launch expectations. The relative effective area is of fundamental importance to determine the temperature of hot plasma emission which is required to use cluster of galaxies as cosmological probe. For example, as discussed by the Planck Collaboration (arXiv:1303.5080) approximately half of the difference in the measurements of σ_8 obtained from cosmic microwave background (CMB) and from galaxy cluster analyses can be explained by temperature calibration uncertainties. The relative effective area is also fundamental for the extrapolation of models to higher energies which it important in order to make optimal use of simultaneously observations of XMM-Newton and NuSTAR. Currently, and most likely during the following years, some 10% of XMM-Newton observing time is performed simultaneously with NuSTAR observations. For example, such observations allow distinguishing absorption from reflection scenarios which is most important to study the physics close to neutron stars, Galactic and supermassive black holes. We can assume that in EPIC can we surely identify and analyse relativistic broad lines if they account for more than 10% relative to the (extrapolated) continuum (e.g. Bhattacharyya & Strohmayer, 2007, ApJ 664, L103; Madej & Jonker, 2011, MNRAS 412, L113, Miller, et al. 2002, ApJ 570, L69, Wilms et al., 2001, MNRAS 328, L2769, Miller et at., 2004, ApJ 606, L1319, Risaliti et al., 2009 ApJ 696, 160). In order to utilize joint XMM-Newton/NuSTAR observations to distinguish between different scenarios (e.g. absorption versus reflection) we have to ensure that the extrapolated calibration uncertainties are below the expected (physical) effect. Assuming a power-law continuum with Γ=1.8 we have to constrain the accuracy of the relative calibration to <2% and the absolute effective area calibration to <5% in order to identify reflection down to 10% with respect to the continuum flux (assuming NuSTAR is calibrated to comparable accuracy). An absolute effective area calibration accuracy of <5% would constrain the absolute flux determination for clusters of galaxies (<10 arc min) to <10% assuming <5% calibration accuracy of the vignetting. A further scientific topic requiring best possible calibration of the effective area is the study of emission spectra of neutron stars and their use to constrain their equation of state, e.g. Webb et al., 2007, ApJ 671, 727.

5.2.6 Point Spread Function and Vignetting:

With the 2-D parameterization of the EPIC Point Spread Function (PSF) (Read et al. (2011, A&A, 534, 34), Read & Saxton, XMM-CCF-REL-0280) and subsequent improvements (Guainazzi et al., XMM-CAL-SRN-0313) the accuracy of the point spread function model is assumed to be better than 2% (Guainazzi et al., XMM-SOC-CAL-TN-0018). With the calibration of the position of the telescope axis (Lumb et al., 2003, ExA 15, 89 (2004, astro-ph/0403647) and Kirsch, XMM-CCF-REL-156) the differences in flux for off axis sources (vignetting) for each camera are in the order of 5%. However, larger differences (up to 10%-13% excess flux) among the EPIC cameras have been unveiled by a study of a large sample of 2XMM serendipitous sources (Mateos et al., 2009, A&A, 496, 879). Investigation of the causes of these discrepancies is on-going (Guainazzi et al., XMM-SOC-CAL-TN-0018).

The 2-D parameterization of the EPIC PSF has significantly reduced the number of spurious detections near bright sources which supported establishing the third 3XMM-DR4 XMM-Newton Serendipitous Source Catalogue. However, for challenging tasks, e.g. removal of spurious detection in the Galactic centre region in order to measure weak extended emission (XMM-Newton observing program 69464) careful visual inspection and manual cleaning is still required. The main scientific challenge for the PSF is the correction for brightest, piled-up sources, like bursting Galactic black holes (e.g. Ng et al., 2010, A&A 522, 96, example for the potential impact: Díaz-Trigo et al., 2013, Nature 504, 260). Here it is of utmost importance that spectra established after excluding regions which contain piled-up events, show exactly the same spectral shape and spectral features.

The accuracy of the vignetting calibration constrains usage of off-axis sources and analysis of extended sources. The scientific return of XMM-Newton will significantly increase if off-axis sources can be rigorously compared for variability studies. Here sources commonly expected to be constant are of highest interest. An example might be the B star Tau Sco where incorrect indication of variability led to a monitoring program with Suzaku, stressing the importance of vignetting but also of inter-mission calibration. Prominent supernova remnants like Cas A have an age of 300 years, i.e. within its technically feasible lifetime, XMM-Newton might reasonably monitor the source for some 10% of its age and as such constant emission might not be expected. For very extended sources, like clusters of galaxies, the vignetting is very important.

5.3 Detailed Calibration requirements

The detailed calibration requirements are defined as close as possible to the actual calibration work and instrument monitoring. By doing this the fulfilment of the calibration requirement can be monitored with minimal workload.

1. Astrometry:

   a. Absolute astrometry (EPIC image): ≤1.2'' (1σ).
   b. Absolute astrometry (OM image): ≤1.4'' (1σ).

2. Timing:

   a. Relative Timing (ΔP/P, pn timing): $< 4 \times 10^{-8}$
   b. Relative Timing (ΔP/P, pn burst): $\leq 1 \times 10^{-8}$
   c. Absolute Timing: ≤ 140µs

3. Absolute wavelength/energy reconstruction:

   a. RGS (Order 1): Wavelength: ±6 m Å (1σ).
   b. RGS (Order 2): Wavelength: ±5 m Å (1σ).
   c. EPIC pn image (single events): Energy ±5 eV(1σ).
   d. EPIC pn image (double events): Energy ±10 eV(1σ).
   e. EPIC pn timing/ burst (single and double events): Energy ±20 eV (1σ).
   f. EPIC MOS image (combined): Energy ±7 eV (1σ).
   g. OM Visible Grism (on-axis): Wavelength ≤7.5 Å (1σ).
   h. OM UV Grism (on-axis): Wavelength ≤2.0 Å (1σ).

   The energy reconstruction should not show any systematic off-set for any of the instruments and observing modes. For RGS, EPIC-pn timing and OM Grism the requirements are for on-axis point sources whereas for the EPIC image modes the requirement is defined for the full field of view.

4. Line Spread Function and Response

   a. RGS Line Spread Function (on-axis point source): ΔFWHM/FWHM ≤10%
   b. EPIC response matrix (on-axis source, single events, pn resolution at 922/1022 eV): ΔFWHM ≤ 1.5eV
   c. EPIC response matrix (on-axis source, single events, pn resolution at 6.4 keV): ΔFWHM ≤ 10eV
   d. EPIC response matrix (on-axis source, MOS resolution at 1.5 keV): ΔFWHM ≤ 1.5eV
   e. EPIC response matrix (on-axis source, MOS resolution at 5.9 keV): ΔFWHM ≤3eV
   f. EPIC response matrix (on-axis source, tail) <1%

5. Flux and Effective Area

    a. Relative effective area I: The relative effective area between pn, MOS1, MOS2, RGS1 and RGS2 should be unity (σ = 1%), i.e. a fit on a continuum source should provide the same continuum characterizing parameter and normalization for every instrument.
    b. Relative effective area II: The relative effective area of pn, MOS1 and MOS2 with respect to narrow features per instrument: <1% for the energy range from 1.0 keV to 7keV.
    c. Relative effective area III: The relative effective area of EPIC between a given energy E and 10 x E: ≤ 2%
    d. Absolute effective area of X-ray instruments: ≤5%
    e. Absolute flux reconstruction for an extended source (r ≤ 10 min): <10%
    f. Absolute flux reconstruction of pn should be constant over the entire lifetime of the mission: <3%
    g. OM flux reconstruction (image, filter): <7%
    h. OM flux reconstruction (grism, no order contamination): < 10%
    i. OM absolute flux reconstruction should be constant over the entire lifetime of the mission: <3%

6. Point Spread Function and Vignetting

    a. Encircled Energy Fraction (for radii < 2 arcmin): ≤2%
    b. EPIC vignetting: <5%

5.4 Comparison with pre-launch requirements

5.4.1 *Astrometry:*

The pre-launch requirement asks for a calibration accuracy of 1.0" for the image location which is assumed to be "re-centred" using previously known sources. From the comparison with AGN from the catalogue of Véron-Cetty & Véron (2010, A&A 518, 10) we estimate the systematic calibration error, $\Delta sys < 0.5''$.

5.4.2 *Timing:*

The pre-launch requirement asks for a calibration accuracy of $\leq 2\times10^{-5}$s for image mode and <0.1% for timing mode. For pn the performance significantly exceeds these requirements, whereas the MOS timing mode was never finally calibrated.

5.4.3 *Absolute energy reconstruction:*

The pre-launch requirement for the RGS wavelength scale was <5 mÅ which is basically achieved for the second order and slightly exceeded for the 1st order (±6 m Å). The pre-launch requirement for the EPIC was <5eV for the energy range from 100 eV to 15 keV and ≤3eV for an energy range between E and ≤3E. The first requirement is fulfilled for MOS for Al Kα at 1.5 keV whereas for Mn Kα at 5.9 keV the requirement is not fulfilled (3±4). Currently, for the pn the requirement is not fulfilled. For the pn timing mode an accuracy of ±20 eV seems to be the best achievable. For the pn image mode a reconstruction accuracy of ±5eV is expected to be achievable after implementing new corrections, in particular using the quiescent background dependent gain and possibly the Cu-K fluorescence emission.

5.4.4 *Line Spread Function and Response:*

The pre-launch requirement for the calibration accuracy for the RGS line spread function (LSP) was 0.01 Å at 15 Å assuming a LSP of 0.04 Å, i.e. to characterize the LSP with a 25% error. The current calibration characterizes the LSP with a 10% error, i.e. the calibration accuracy is significantly better than the pre-launch requirement. The EPIC response calibration requirement was ΔFWHM = 5 eV for 1 keV and 6 keV. This requirement is fulfilled for MOS and is regularly updated reflecting the changing camera behaviour. For pn a regular update reflecting changing instrument response is under implementation.

5.4.5 *Effective Area and Flux*

The pre-launch requirements for the calibration accuracy for the effective area (EPIC at 0.5, 1.0 and 6 keV, RGS at 10 Å, OM grism from 160 nm to 240 nm) were ≤ 10%. The relative effective accuracy requirement (EPIC between E and 10 × E, full energy band for RGS) was required to be better than 3%. However, these tabulated values (Erd et al., 1996, XMM-PS-GM-02, table 2.4.1, 2.4.2 and 2.4.3) fall short of the scientific requirements defined in the 2.1.3, 2.1.7). In 2.1.7 in XMM-PS-GM-02 "There will also need to be careful cross-calibration of the EPIC MOS and PN CCD detectors, so that the full gasp of XMM can be utilized by addition of data from all 3 focal plane cameras". This requirement is now reflected in 5.3.5a and is currently not fulfilled. In 2.1.3 in XMM-PS-GM-02 "In order to combine the data from these 3 missions (comment from the author: XMM-Newton, AXAF (now Chandra) and Astro-E (now Suzaku) (and possibly SAX, Spectrum-X, ASCA, ROSAT, etc.) relative errors (mission to mission) of ≤ 10% will be required". Currently, these differences are up to 20%. In order to have ≤ 10% relative errors (mission to mission) the individual calibration accuracy must be better that the required ≤ 10%. In 2.1.4 in XMM-PS-GM-02 "thus the required accuracy on the effective area at all energies should be at least 2-3% throughout the passband for both EPIC and RGS". The scientific reasoning for the requirement is optically-thin plasma diagnostics which affects the temperatures of clusters. Comparing the different EPIC instruments, we can constrain the relative effective area accuracy to be 10%. The OM grism are <10% and as such the pre-launch requirement is in first order fulfilled.

5.4.6 Point Spread Function and vignetting

The pre-launch requirement for the point spread function was 1% accuracy from the peak to the FWHM and 0.05% of the peak level in the PSF wings. The unique position of XMM-Newton for cluster observations was well recognised before launch and the subsequent consequences for the vignetting calibration. In 2.1.4 in XMM-PS-GM-02 states with respect to the vignetting accuracy "However, the exposures may then be prohibitively long for 1-2% relative error". The accuracy of the point spread function model is assumed to be better than 2% (Guainazzi et al., XMM-SOC-CAL-TN-0018) and the differences in flux for off axis sources (vignetting) for each camera are of the order of 5%. However, larger differences (up to 10%-13% excess flux) among the EPIC cameras have been unveiled by a study of a large sample of 2XMM serendipitous sources (Mateos et al., 2009, A&A, 496, 879).

5.4.7 Summary pre-launch requirements and conclusions

Overall the pre-launch requirements for the calibration accuracy are very similar to the requirements formulated here, especially, if the detailed discussion in XMM-PS-GM-02 is considered. This goes along with the results of the survey into the opinion of the users. Where the pre-launch calibration requirements are fulfilled, no comments are received. The pn timing and burst mode were not addressed in the pre-launch requirements and as such are new requirements. On the other side the effective area / flux reconstruction for the X-ray instruments was never established at the level requested by the pre-launch requirements and this is reflected in the concern of the users and is a limiting factor for future research. An important issue is maintaining the achieved accuracy level. The energy reconstruction of pn double-events is a good example. From the first calibration onwards pn double-events were energy calibrated in the same way as single events and differences to optimal corrections were considered acceptable. Due to aging of the camera the energy calibration of double-events differs more and more from the energy-calibration of single events leading now to incorrect scientific results. From SAS 14 onwards the double events have been corrected separately from the pn-single events, which improves the pn double-event energy reconstruction over the whole mission implying the need to recalibrate the pn detector response. Maintaining the achieved calibration accuracy of a certain parameter with respect of aging-caused instrument degradation often causes the need to recalibrate the parameter over the whole mission lifetime and the need to recalibrate other related parameters.