The Greenwich Photo-heliographic Results (1874 – 1976): Procedures for checking and correcting the sunspot digital datasets

D. M. Willis, R. Henwood, M. N. Wild, H. E. Coffey
W. F. Denig, E. H. Erwin & D. V. Hoyt

Published version information

Citation: DM Willis et al. "The Greenwich Photo-heliographic Results (1874–1976): Procedures for checking and correcting the sunspot digital datasets." Solar Physics, vol. 288, no. 1 (2013): 141-156.

doi: 10.1007/s11207-013-0312-x

The final publication is available at Springer via https://doi.org/10.1007/s11207-013-0312-x

The Greenwich Photo-heliographic Results (1874 – 1976): Procedures for Checking and Correcting the Sunspot Digital Datasets

D. M. Willis • R. Henwood • M. N. Wild • H. E. Coffey • W. F. Denig • E. H. Erwin • D. V. Hoyt

D. M. Willis (🖂)
Space Physics Division, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0QX, UK, and Centre for Fusion, Space and Astrophysics, Department of Physics, University of Warwick, Coventry CV4 7AL, UK
e-mail: david.willis@stfc.ac.uk

R. Henwood · M. N. Wild
UK Solar System Data Centre, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0QX, UK
e-mail: rjhenwood@yahoo.co.uk
e-mail: matthew.wild@stfc.ac.uk

H. E. Coffey · W. F. Denig · E. H. Erwin
NOAA National Geophysical Data Center, 325 Broadway, Boulder, CO 80305, USA
e-mail: hcoffeycolorado@gmail.com
e-mail: william.denig@noaa.gov
e-mail: erwin_ed@hotmail.com

D. V. Hoyt
108 Buffalo Run Trail, Berkeley Springs, WV 25411, USA
e-mail: dhoytwv@gmail.com

Short title: Greenwich Photo-heliographic Results: Corrections

Version dated: 12 April 2013
(File: SOLA2127R1.doc)

Abstract Attention is drawn to the existence of errors in the original digital dataset containing sunspot data extracted from certain sections of the printed *Greenwich Photo-heliographic Results (GPR) 1874 – 1976*. Calculating the polar coordinates from the heliographic coordinates and comparing them with the recorded polar coordinates reveals that there are both isolated and systematic errors in the original sunspot digital dataset, particularly during the early years (1874 – 1914). It should be noted that most of these errors are present in the compiled sunspot digital dataset and not in the original printed copies of the *Greenwich Photo-heliographic Results*. Surprisingly, many of the errors in the digitised positions of sunspot groups are apparently in the measured polar coordinates, not the derived heliographic coordinates. The mathematical equations that are used to convert between heliographic and polar coordinate systems are formulated and then used to calculate revised (digitised) polar coordinates for sunspot groups, on the assumption that the heliographic coordinates of every sunspot group are correct. The additional complication of requiring accurate solar ephemerides in order to solve the mathematical equations is discussed in detail. It is shown that the isolated and systematic errors, which are prevalent in the sunspot digital dataset during the early years, disappear if revised polar coordinates are used instead. A comprehensive procedure for checking the original sunspot digital dataset is formulated in an appendix.

Key words Greenwich photo-heliographic results · Positions and areas of sunspots and faculae · Reconstructed solar images · Isolated and systematic errors · Mathematical equations · Solar ephemerides · Preliminary corrections · Generalised checking procedures

1. Introduction

As noted in the preceding companion paper by Willis et al. (2013), there are typographical, systematic and isolated errors in the different datasets containing the printed and digital versions of the *Greenwich Photo-heliographic Results, 1874 – 1976*. The preceding paper (hereafter termed "Paper 1") provides a succinct summary of all the essential background information on the Greenwich Photo-heliographic Results (1874 – 1976), published initially by the Royal Observatory, Greenwich and subsequently by the Royal Greenwich Observatory (RGO), which is required to understand detailed discussions of particular errors. Moreover, the preceding paper also outlines a preliminary strategy for correcting the RGO sunspot digital datasets (see Section 6 of Paper 1). The purpose of the present paper (hereafter termed "Paper 2") is to extend this preliminary strategy by providing a more formal and precise explanation of the procedures that have been developed to check and correct the sunspot digital datasets.

The present paper originated in an attempt to provide a precise specification for implementing the corrections necessary to remove known errors in the original sunspot digital dataset stored online at the NOAA National Geophysical Data Center (NGDC), Boulder, Colorado (http://www.ngdc.noaa.gov/stp/solar/greenwich.html; use the 'Sunspot Regions' link) and at the UK Solar System Data Centre (UKSSDC) at the Rutherford Appleton Laboratory (http://www.ukssdc.ac.uk/wdcc1/greenwich). This dataset is conveniently abbreviated to RGO–SDD, as discussed in Section 5 of Paper 1. In particular, this specification relates primarily to the correction of the errors that become apparent when the positions of sunspot groups are plotted in polar coordinates rather than heliographic coordinates, as illustrated in Figure 1. This figure was created by using a visualisation tool, developed as a simple extension to an earlier technique for reconstructing solar images from the information included in the original sunspot digital dataset (Willis, Davda, and

Stephenson, 1996). It should be noted that most of these errors occur only in the original RGO sunspot digital dataset (RGO–SDD) and not in the printed RGO observations, bulletins and annals (RGO–POBA). The much smaller number of typographical errors that occur in the printed RGO publications is considered separately in the following companion paper (hereafter termed "Paper 3") by Erwin et al. (2013). It should also be noted that many of the errors in the sunspot digital dataset currently stored at the NGDC and the UKSSDC have been corrected in the version of this dataset stored online at the Marshall Space Flight Center (http://solarscience.msfc.nasa.gov/greenwch.shtml), although the corrections to this latter dataset have not been fully documented.

The present specification for implementing corrections is based largely on the solution of the mathematical equations that relate polar coordinates to heliographic coordinates. The solution of the mathematical equations requires prior knowledge of the solar ephemerides (P_0, B_0 and L_0) and the semi-diameter of the Sun (S). It is possible to extract solar ephemerides for many, but not all, years in the interval 1874 – 1955 from the more recent sunspot and faculae digital dataset (http://www.ngdc.noaa.gov/stp/solar/greenwich.html; use the 'Solar White Light Faculae' link). This dataset is conveniently abbreviated to RGO–S&FDD, as also discussed in Section 5 of Paper 1. Otherwise, solar ephemerides (particularly missing values) must be obtained from the appropriate Ephemeris for Physical Observations of the Sun (i.e., the *Nautical Almanac* or the *Astronomical Ephemeris*) or generated using a suitable algorithm (e.g., Meeus, 1991). Similarly, values of the semi-diameter of the Sun must be taken from either the *Nautical Almanac* or *Astronomical Ephemeris* or, alternatively, generated using a suitable computer program.

2. Derivation of the Polar Coordinates from the Heliographic Coordinates

There is some "redundancy" in the RGO–SDD, in the sense that the positions of sunspots are given in both polar coordinates and heliographic coordinates (for further information, see Section 6 of Paper 1). Making use of this fact, erroneous polar coordinates can be corrected by recalculating them from the heliographic coordinates.

2.1 Inversion of the Heliographic Coordinates

Generally, positions on the solar disk are measured in polar coordinates and then converted into heliographic coordinates. Let r be the measured distance of the position (P) of a sunspot from the centre (C) of the Sun's apparent disk, χ the position angle of the sunspot from the Sun's axis (measured anti-clockwise from the apparent position (N_0) of the North Pole (N_P) of the Sun; see Figure 2(b)), R the measured radius (CD) of the Sun on the photograph and S the semi-diameter of the Sun in arc (see Figure 2(a)). Let ρ and ρ' be the angular distances of the spot from the centre of the apparent disk, as viewed from the Sun's centre (C) and from the Earth (⊕) respectively (see Figure 2(a)). Moreover, let B_0 and ϕ be the heliographic latitudes, and L_0 and λ the heliographic longitudes, of the Earth and the spot respectively. Then the heliographic latitude and longitude of the position (P) of the sunspot can be determined by the following equations (see e.g., De La Rue, Stewart, and Loewy, 1869; Royal Greenwich Observatory, 1975; and Dezső, Gerlei, and Kovács, 1987):

$$\rho' = (r / R) S , \tag{1}$$

$$\sin (\rho + \rho') = r / R , \tag{2}$$

$$\sin \phi = \sin B_0 \cos \rho + \cos B_0 \sin \rho \cos \chi , \tag{3}$$

$$\sin (L_0 - \lambda) = \sin \chi \sin \rho \sec \phi . \tag{4}$$

When deriving these formulae, the smallness of the angles ρ' and S ($\rho' < S \leq 0.272°$) was utilised.

Recall that the polar angle χ in these equations is measured from the apparent position (N_0) of the solar North Pole. However, it is the direction of the Earth's and not the Sun's axis on the solar disk that is known. Thus another parameter (P_0) is required, namely the position angle of the north end (N_0) of the Sun's axis measured eastwards (anticlockwise) from the geocentric north point (N) on the Sun's disk (see Figure 2(b)).

Next, the inverse relations between the heliographic and polar coordinates are determined. As can be seen from the spherical triangle shown in Figure 2(b), the following equation also holds:

$$\cos\rho = \sin B_0 \sin\phi + \cos B_0 \cos\phi \cos(L_0 - \lambda). \tag{5}$$

This equation has been presented explicitly by Győri (1989; personal communication, 2008).

Using Equation (5), it can be shown from Equations (3) and (4) that the following three trigonometric relations hold for χ:

$$\sin\chi = \frac{\cos\phi \sin(L_0 - \lambda)}{\sin\rho}, \tag{6}$$

$$\cos\chi = \frac{\cos B_0 \sin\phi - \sin B_0 \cos\phi \cos(L_0 - \lambda)}{\sin\rho}, \tag{7}$$

$$\tan\chi = \frac{\sin(L_0 - \lambda)}{\cos B_0 \tan\phi - \sin B_0 \cos(L_0 - \lambda)}. \tag{8}$$

These equations can be solved for χ if the heliographic coordinates ϕ, λ, B_0 and L_0 are all known. Equation (8) can be solved immediately to give two values of χ. However, two of the three Equations (6) – (8) must be solved to find the correct value of χ (i.e., the value of χ in the correct quadrant). First, Equation (5) must be solved to find the value of ρ, which can be substituted into

Equations (6) and (7). For example, using this value of ρ, Equation (6) can then be solved to give two further values of χ. Only one value of χ from the pairs of solutions of Equations (6) and (8) is in common. This is the required value of χ.

Equations (1) and (2) can be combined to give the following equation for r/R:

$$\sin[(r/R)S+\rho]=r/R\,. \tag{9}$$

To be completely rigorous, Equation (9) must be solved for r/R using an iterative procedure such as Newton's method, However, if it is believed that only the position angle (χ) is incorrect, the problem is essentially solved once the correct value of χ has been determined. Nevertheless, the archived value of r/R can be substituted into equation (1) to find ρ', which can then be substituted into Equation (2) to verify that this equation actually holds.

2.2 Solar Ephemerides

As can be observed readily from the equations presented in the previous sub-section, values for the solar ephemerides (P_0, B_0, L_0, and S) must be known in order to solve them. These quantities are intimately related to the rotation and the revolution of Earth. Their values depend on the time of the observation.

For simplicity, in this initial investigation, the solar ephemerides (P_0, B_0, and L_0) have been obtained from the algorithm published by Meeus (1991) and the semi-diameter of the Sun in arc (S) has been obtained from the JPL Horizons On-Line Ephemeris System (JPL, 2005: http://ssd.jpl.nasa.gov/?horizons). However, as a check on the accuracy of the solar ephemerides derived from the algorithm of Meeus (1991), the solar coordinate transformations have also been performed with software developed at the Rutherford Appleton Laboratory (RAL) by P. T. Wallace, in connection with pointing systems for ground-based solar telescopes.

In this latter development, the adopted ephemeris models are taken from Seidelmann (1992) and a minor inaccuracy perpetuated since the work of Carrington has been corrected. Specifically, the North Pole of the Sun defined by Carrington (1863) and used canonically thereafter is ambiguous, in the sense that the relevant equations appear to define a pole that is moving with respect to the stars (P. T. Wallace, private communication, 2010). This minor inaccuracy arose because Carrington used ecliptic coordinates but neglected planetary precession — the slow (47″ per century) change in orientation of the ecliptic plane caused by planetary perturbations of the Earth's orbit. In addition, allowance is made in the RAL software for the geographic coordinates and altitudes of the appropriate solar observatories. The values of the geographic coordinates and altitudes used in this investigation are essentially those given in Table 1 of Paper 1. However, it should be noted that any revision of the dates, times or observatory codes in the current version of the sunspot and faculae digital dataset (RGO–S&FDD) would inevitably necessitate further calculations in the case of the algorithm developed by P. T. Wallace. Conversely, the algorithm of Meeus (1991) is based on a simpler geocentric system of coordinates, which neglects the observer's position. This simpler system of coordinates is the one that was used originally by RGO staff to convert from polar coordinates to heliographic coordinates, when measuring sunspot areas on the solar photographs and then expressing these areas in millionths of the visible solar hemisphere.

2.3 Calculated Polar Coordinates

The histogram presented in Figure 3 shows the differences between calculated and recorded values of r/R for small discrepancies $\Delta(r/R)$ lying in the restricted range $-0.01 \leq \Delta(r/R) \leq +0.01$ (i.e., for discrepancies in the radial distance from the centre of the Sun with a magnitude less than or equal to one hundredth of the radius of the Sun). The calculated values of r / R are obtained from the reverse solution of the mathematical equations, assuming that the heliographic coordinates are correct. The recorded values of r / R are obtained from the original sunspot digital dataset (RGO–SDD). These small discrepancies have been calculated using solar ephemerides obtained from both the geocentric algorithm published by J. Meeus (un-shaded histogram) and the topocentric algorithm derived by P. T. Wallace (shaded histogram).

As a convenient computational simplification in this initial investigation, the annual variation in the angular size of the Sun (*S*), used in conjunction with the solar ephemerides (P_0, B_0 and L_0) derived from the algorithm published by Meeus (1991), is taken from the JPL Horizons System (JPL, 2005). Although, a typical annual variation has been taken for the semi-diameter of the Sun (*S*), it has been found that the solution of the mathematical equations is relatively insensitive to year-to-year differences in the variation of *S* with calendar date, as obtained from other sources (e.g., the *Nautical Almanac*). Moreover, the relative insensitivity of the results to these effects can be inferred indirectly from Figure 3, since the algorithm developed by P. T. Wallace allows for all known variations in P_0, B_0, L_0, and *S*.

Similarly, the histogram presented in Figure 4 shows the differences between calculated and recorded values of χ for small discrepancies $\Delta\chi$ lying in the restricted range $-1.0 \leq \Delta\chi \leq +1.0$ (i.e., for discrepancies in the position angle measured anticlockwise from the North Pole of the Sun's axis with a magnitude less than or equal to 1 degree). Once again, the calculated values of χ are obtained from the reverse solution of the mathematical equations (assuming that the heliographic coordinates are correct) and the recorded values of χ are obtained from the original sunspot digital dataset (RGO–SDD).

It is clear from the histograms presented in Figures 3 and 4 that the discrepancies between the calculated and recorded values of r/R and χ both show a pronounced peak at very small discrepancies, whichever algorithm is used to calculate solar ephemerides. However, the algorithm published by Meeus (1991) is marginally better in the sense that it yields slightly smaller discrepancies between calculated and recorded values of both r/R and χ. This result is consistent with the fact that the algorithm of Meeus follows more closely the procedure adopted by RGO staff when measuring sunspot positions and areas on the solar photographs and then converting sunspot positions and areas to heliographic coordinates, which resulted in the recorded data in the RGO–SSD.

Exactly the same procedures have been performed on the alternative version of the original sunspot digital dataset stored at the Marshall Space Flight Center (http://solarscience.msfc.nasa.gov/greenwch.shtml). Accordingly, Figure 5 shows the differences between calculated and recorded values of r/R for small discrepancies $\Delta(r/R)$ lying in the restricted range $-0.01 \leq \Delta(r/R) \leq +0.01$ and Figure 6 shows the differences between calculated and recorded values of χ for small discrepancies $\Delta\chi$ lying in the restricted range $-1.0 \leq \Delta\chi \leq +1.0$. The calculated values of r/R and χ are again obtained from the reverse solution of the mathematical equations, assuming that the heliographic coordinates are correct. Comparing Figures 5 and 6 with Figures 3 and 4, respectively, it is clear that there is a broader peak near zero difference for the MSFC dataset than for the NGDC (and UKSSDC) dataset. Moreover, there is a small offset of about 0.002 in the difference between the calculated and recorded values of r / R in the case of the MSFC dataset. Further research is required to determine the origin of this offset.

The differences between calculated and recorded values of r / R and χ , presented in Figures 3 and 4, have been evaluated for just those values that are common to both the original sunspot digital dataset (RGO–SDD) and the sunspot and faculae digital dataset (RGO–S&FDD). This restriction is necessary because the solar ephemerides for the topocentric algorithm developed by P. T. Wallace also involve the geographic coordinates and altitude of each contributing solar observatory. Currently, the observatory code that identifies the contributing solar observatory is only available from the latter dataset (RGO–S&FDD), as explained in Section 5 and Appendix B of Paper 1. For consistency, however, the same restriction has been imposed in Figures 5 and 6, which show the corresponding differences between calculated and recorded values of r / R and χ for the alternative sunspot digital dataset stored at the Marshall Space Flight Center, using just the solar ephemerides for the geocentric algorithm published by J. Meeus.

3. Summary and Conclusions

The primary purpose of this paper is to draw attention to the existence of errors in the original digital dataset containing sunspot data extracted from certain sections of the printed *Greenwich Photo-heliographic Results, 1874 – 1976*. It must be emphasised again that most of these errors occur only in the original RGO sunspot digital dataset and not in the printed RGO observations, bulletins and annals. Moreover, it is not clear how these particular errors were introduced when the original sunspot digital dataset was compiled.

It has been found that there are both isolated and systematic errors in the RGO–SDD. Surprisingly, many of the isolated errors in the digitised positions of sunspot groups are apparently in the measured polar coordinates, not the derived heliographic coordinates. Similarly, the systematic errors found in the annual accumulations of sunspot positions are in the measured polar coordinates not the derived heliographic coordinates (see Figure 1).

It has also been found that the different determinations of the solar ephemerides have only a slight influence on the polar coordinates determined from the heliographic coordinates of the sunspots (see Figures 3 and 4).

For convenience, the abbreviation RGO–SDD–R is used to refer to the revision (R) of the original RGO sunspot digital dataset (RGO–SDD); this revised dataset contains the original heliographic coordinates and the calculated, or revised, polar coordinates. The revised dataset can be used to revise Figure 1, which shows the systematic errors that arise if sunspot positions are plotted using the original polar coordinates. Figure 7 reproduces Figure 1, with the revised reconstructed solar images — using the revised sunspot digital dataset (RGO–SDD–R) — appended at the bottom of the figure (the third frames). Reconstructed solar images are presented in the following coordinate systems: original polar coordinates (top); (original) heliographic coordinates (middle); and revised polar coordinates (bottom). It is clear that the revised solar images plotted in revised polar coordinates do not exhibit the systematic errors that occur for the original solar images plotted in original polar coordinates. Moreover, the revised solar images plotted in revised polar coordinates are consistent with the solar images plotted in (original) heliographic coordinates.

As noted in the Introduction to this paper (and Section 4 of Paper 1), many of the errors in the original sunspot digital dataset stored online at the National Geophysical Data Center (NGDC) and the UK Solar System Data Centre (UKSSDC) have been corrected in the version of this digital dataset stored online at the Marshall Space Flight Center (MSFC), although the various changes to this latter dataset have not been fully documented. The procedure for calculating revised polar coordinates from the heliographic coordinates has also been applied to the MSFC dataset. Figures 5 and 6 present the histograms that correspond to those presented in Figures 3 and 4, using just the simpler algorithm for the required solar ephemerides (J. Meeus). It is clear that there is a broader peak near zero difference for the MSFC dataset (Figures 5 and 6) than for the NGDC (and UKSSDC) dataset (Figures 3 and 4). Moreover, there is a small offset of about 0.002 in the difference between the calculated and recorded values of r/R in the case of the MSFC dataset (Figure 5). Further research is required to determine the origin of this offset.

Figures 3 and 4 indicate that the majority of the recalculated values of r/R and χ differ only slightly from the recorded values of these quantities. Figure 7 shows that the smaller, but still significant, number of isolated and systematic errors in the original sunspot digital dataset (RGO–SDD), which produce the "invisible wings" of the histograms shown in Figures 3 and 4 (i.e., the smaller number of large discrepancies that occur within the unrestricted ranges $-1 \leq \Delta(r/R) \leq +1$ and $-360 \leq \Delta\chi \leq +360$), are corrected by recalculating the polar coordinates from the heliographic coordinates.

Finally, a comprehensive procedure for checking the original sunspot digital dataset (RGO–SDD) is formulated in an appendix to this paper, although not all of the tests outlined in the appendix have yet been fully implemented. However, a subset of the tests formulated in this appendix has been implemented in the following companion paper (Paper 3) by Erwin et al. (2013), which considers the correction of errors in the sunspot and faculae digital dataset (RGO–S&FDD) and thereby the correction of typographical errors in the RGO printed observations, bulletins and annals (RGO–POBA).

Acknowledgements The authors thank J. H. Allen, L. Győri, R. A. Harrison, F. R. Stephenson, and P. T. Wallace for much valuable advice, assistance and encouragement during the preparation of this paper. They are also greatly indebted to a referee for suggesting substantial improvements to the presentation of the material included in the paper.

Appendix A: Formal Procedure for Detecting and Correcting Errors in the GPR

A.1 Aim of the Appendix

A comprehensive and detailed procedure for checking the original sunspot digital dataset (which is introduced and defined in Sections 3 and 5 of Paper 1) is formulated in this appendix. Although the various checks apply specifically to the original sunspot digital dataset (RGO–SDD), the methodology could easily be adapted to apply similar checks to the more recent sunspot and faculae digital dataset (RGO–S&FDD). A preliminary procedure for checking the sunspot and faculae digital dataset (RGO–S&FDD), developed at the NOAA National Geophysical Data Center (NGDC), has already been used to identify a number of typographical errors in the RGO printed observations, bulletins and annals (RGO–POBA). This preliminary procedure is defined and discussed in Section 2 of Paper 3. Section 3 of this third paper includes a table presenting the typographical errors in the RGO–POBA for the interval 1874 – 1917, which have been identified in a preliminary investigation.

A.2 'Unit Testing' to validate that the GPR is Free from Errors

The present approach to validating that RGO–SDD is free from errors is to construct a series of simple, self-contained tests that this dataset must pass in order to be classified as "error free". This approach has been adopted from the discipline of software engineering, where it is typically applied to functional code. For software, a 'unit' describes the smallest testable part of an application. Ideally, tests should be independent.

In applying the unit-testing methodology to this dataset (RGO–SDD), particular care has been taken to ensure good coverage. By considering the data from different abstractions (logical, physical), exploiting redundancy (polar coordinates and heliographic coordinates) and the duplication of information in different datasets (RGO–SDD, RGO–S&FDD, RGO–POBA), coverage is maximised. It is important to gain good coverage, since no acceptance tests for the dataset as a whole are available.

For the RGO–SDD dataset, a list of unit tests is provided in the tables below. The procedure for developing a unit test is as follows. Identify an assertion that must hold for every element or group of elements in the dataset (Column: Assertion). Identify a method to test if this assertion holds for a given element or group of elements (Column: Test). Identify a correction or remedy to perform on elements or groups that fail the test (Column: Failure). Record any relevant additional information (Column: Notes).

Tests that are based on different conceptual views of the database are collected together into groups. These abstractions include: Group I: File Format (line endings must be consistent); Group II: Data Format (all lines must be the same length); Group III: Published Errata (check that all published errata have been implemented); Group IV: Observational Method (observation times must be consistent with the known method of observation); Group V: Observational Data (recorded values must be within physically acceptable bounds); Group VI: Physical Properties (sunspot observations must be in accordance with physical understanding of the Sun); Group VII: Redundant Data (redundant values must be self-consistent, as discussed in Section 6 of Paper 1). Group VIII: Completeness Checks (confirm that the results of any manual checks performed by individual scientists have been incorporated). This last test should provide a sampling of the different types of error and successful completion of this test should then confirm that the previous tests provide good coverage.

By iteratively testing and correcting the given dataset it is possible to arrive at a final dataset that is error free — where an ‘error’ is defined as a failure to pass one or more tests. This unit testing stresses simple ‘atomic tests’, each of which has one of two possible outcomes — success or failure. An intermediate dataset is created between each successive test. These intermediate datasets can be recorded as ‘versions’ to provide an exact list of changes introduced and, if necessary, allow for additional subsequent changes to the dataset in an arbitrary manner.

A.3 Test Group I: File Format

This test may be defined as a pre-processing test for the consistency of the dataset. This test can be applied with little or no knowledge of the physical properties that are represented by the data.

Table 1 File Format

No.	Assertion	Test	Failure	Notes
1	All entries in the dataset end with ‘\n’	‘File’	Use ‘dos2unix’.	This test may not be necessary — provided that one format is adhered to throughout.

A.4 Test Group II: Data Format

Interpolated values are indicated in the sunspot digital dataset (RGO–SDD) by a ‘special’ time format after the decimal point in the decimal part of a day (.__0). This time format is not specified in the metadata, which currently accompanies the sunspot data. In any future revision of the following test, the complication of interpolated values will be addressed more rigorously but until then such values should be excluded from the dataset, as recommended in Section 6 of Paper 1.

Table 2 Data Format

No.	Assertion	Test	Failure	Notes
1	Interpolated observations should be excluded from the dataset.	Identify observations that have a decimal part of the day in the form: ‘.__0’.	Delete such entries.	Interpolated values are less reliable and will be addressed in a future test.

2	Every entry in the dataset complies with the format string as described in http://www.ukssdc.ac.uk/wdcc1/greenwich/grnwich.fmt	Boolean check.	Look up in RGO book.	

Note: The underscore symbol (_) is used to denote a blank space.

A.5 Test Group III: Published Errata

Errata have been published by the RGO, as noted in Section 5 of Paper 1. These errata have been included in the sunspot digital dataset (RGO–SDD) but have not yet been fully incorporated in the sunspot and faculae digital dataset (RGO–S&FDD). It is also necessary to signify (with a special character) those values that have been adjusted or included because of these published errata. Then, if subsequent reference is made to the books with respect to later tests, it is known whether to examine the original records or the corresponding errata.

Table 3 Published Errata

No.	Assertion	Test	Failure	Notes
1	The errata in the RGO books must be applied to the original dataset.	Perform checks on each of the published errata against the digital dataset.	Correct the dataset.	All subsequent tests that require checking against the RGO books are dependent on this test.

A.6 Test Group IV: Observational Method

This set of tests exploits a priori knowledge of the data collection method. This knowledge includes explicit and implicit aims of the observers as well as physical restrictions on the method of observation.

Table 4 Observational Method

No.	Assertion	Test	Failure	Notes
1	Only one observation is available for each day.	Check that all groups recorded on a given day occur at the same time.	Check errors against the books	

2	An observation is available on every day for which a dated and timed solar photograph was taken.	Generate a list of all days for which a dated and timed photograph was taken and check they all occur once in RGO–SDD.	Check against the RGO books. If books have an entry, but RGO–SDD does not, add this to RGO–SDD. Set sunspot areas to zero if book entry is just a non-zero faculae area.	Currently, RGO–SDD excludes days when only faculae were visible. This test adds all cases where only faculae were visible and the time of the observation is known.
3	Time (UT) is Greenwich Civil Time reckoned from midnight and January 01 is taken to be Day 1.	Look at times in RGO books and compare with times in the dataset.	Add 1.5 to the time of each day in the interval 1874 – 1884 and add 1.0 to the time of each day in the interval 1885 – 1955.	Greenwich Mean Solar Time originally reckoned from Greenwich Mean Noon on January 01 (up to 1884 December 31). Also January 01 was taken to be Day 0 in the interval 1874 – 1955.
4	Each sunspot group is allocated a unique group number.	Check each day for duplicate sunspot group numbers.	Check errors against RGO books.	

Notes: The Assertion in Test No. 1 is true for RGO–SDD but false for RGO–S&FDD because the latter dataset includes measurements from two photographs taken on the same day during the early years (see Section 6 of Paper 1). The Assertion in Test 2 is necessary because a date and time is recorded in RGO–POBA and hence RGO–S&FDD if faculae but not sunspots were visible on the solar photograph, whereas there is no entry in RGO–SDD if no sunspots were visible on the solar photograph (see Section 6 of Paper 1). It might be preferable to include all days in RGO–SDD, although (regrettably) no time is given if neither sunspots nor faculae were visible, and (obviously) no time is available if no photograph was taken. Moreover, during the early years, the entry "No spots or faculae" in the "Measures of positions and areas of sun spots and faculae" sections of RGO–POBA apparently corresponds variously to "(0, 0, 0)" (i.e., zero umbral, whole-spot and faculae areas) and "No photograph" in the "Total projected areas of sun spots and faculae for each day of the year" sections of RGO–POBA. Further research is required to resolve this inconsistency, which may involve incomplete implementation of the published errata.

A.7 Test Group V: Observational Data

This set of tests exploits a priori knowledge of the data collection method. These tests are intended to validate recorded values.

Table 5 Observational Data

No.	Assertion	Test	Failure	Notes
1	Sunspot group must be observed on the solar disk: $r / R \leq 1.000$ (Section 5).	Boolean check: $r / R > 1.000$	Check against RGO books.	
2	Sunspot group angular coordinates must be	The digit after the decimal point of an	Convert automatically.	Angles originally expressed in degrees and

	recorded in decimal degrees.	angular coordinate should be evenly distributed between 0 and 9.		minutes (up to 1881 December 21).
3	Sunspot group must satisfy: $0 \le \chi \le 360°$ (Section 5).	Boolean check.	Check against RGO books.	
4	Sunspot group must satisfy: $-90° \le \phi \le +90°$ (Section 5)	Boolean check.	Check against RGO books.	
5	Sunspot group must satisfy: $0° \le \lambda \le 360°$ (Section 5).	Boolean check.	Check against RGO books.	
6	Sunspot group must satisfy: $-90° \le \psi \le +90°$ (Footnote)	Boolean check.	Check against RGO books.	
7	Observed (projected) whole-spot area cannot be less than observed umbral area	Boolean check.	Check against RGO books.	The most probable explanation for this will be a typographic error.
8	Corrected whole-spot area cannot be less than corrected umbral area.	Boolean check	Check against RGO books.	The most probable explanation for this will be a typographic error.
9	Half the observed umbral area must be less than or equal to the corrected umbral area.	Boolean check.	Check against RGO books.	Foreshortening; area of hemisphere is twice area of disk.
10	Half the observed whole-spot area must be less than or equal to the corrected whole-spot area.	Boolean check.	Check against RGO books.	Foreshortening; area of hemisphere is twice area of disk.

Note: The variable ψ (Test No.6) denotes angular distance (longitude) of the sunspot group from the central solar meridian.

A.8 Test Group VI: Physical Properties

This set of tests exploits a priori knowledge of the physical properties of the Sun.

Table 6 Physical Properties

No.	Assertion	Test	Failure	Notes
1	The Sun must be above the horizon at the observing station in order to make a valid observation at a given	Using stations recorded in the sunspot and faculae dataset (see also Table 1 of Paper 1), identify when the Sun would be visible and check that the observation is during this	Check against the RGO books.	

	date and time.	period.		
2	Sunspot groups are known to move on the surface of the Sun within certain tolerances.	Measure the incremental movement between consecutive sunspot group positions for the variables: (a) r / R ; (b) χ ; (c) ϕ ; (d) λ ; (e) ψ ; (f) size, and alert if it is too great.	Check against the RGO books.	Need to define a tolerance.

A.9 Test Group VII: Redundant Data

This set of tests searches for numerical inconsistencies using "redundant" information in the original sunspot digital dataset (RGO–SDD), as outlined in Section 6 of Paper 1. Some of these tests use ephemeris values, which poses problems already documented (see Sections 5 and 6 of this paper).

Table 7 Redundant Data

No.	Assertion	Test	Failure	Notes
1	Observed whole-spot area should be consistent with corrected whole-spot area.	Use mathematics provided in the RGO books.	Check RGO books.	Within a specified tolerance.
2	Observed umbral area should be consistent with corrected umbral area.	Use mathematics provided in the RGO books.	Check RGO books.	Within a specified tolerance.
3	Converting from heliographic coordinates to polar coordinates should be consistent.	Assuming the heliographic coordinates (ϕ, λ) are correct, check that the calculated polar coordinates ($r / R, \chi$) are within given tolerances for observations of a given sunspot.	If only the polar angle is incorrect, accept the polar coordinates of the sunspot obtained using the mathematical equations. Otherwise check against the RGO books.	This test captures "gross" errors (e.g., N v S) (Section 4). Need to define tolerances.
4	Converting from polar coordinates to heliographic coordinates should be consistent.	Assuming the polar coordinates ($r / R, \chi$) are correct, check that the calculated heliographic coordinates (ϕ, λ) and (ψ) are within specified tolerances for observations	Check against the RGO books.	Need to define tolerances.

		of a given sunspot.		
5	Central meridian distance (ψ) should be consistent with heliographic longitude (λ).	Use a solar ephemeris model to convert between the two angular measurements.	Check against the RGO books.	Need to define a tolerance.

A.10 Test Group VIII: Completeness Check

This test is intended to check that all known errors have been corrected. For example, it is important to include the results of any manual checks that may have been performed by individual scientists.

Table 8 Completeness Check

No	Assertion	Test	Failure	Notes
1	Previous tests cover all currently known errors.	Check that any errors identified in various manual checks have been corrected.	Investigate further.	

References

Carrington, R. C.: 1863, *Observations of the Spots on the Sun from November 9, 1853 to March 24, 1861, Made at Redhill*, Williams and Norgate, London.

De La Rue, W., Stewart, B., Loewy, B.: 1869, Researches on solar physics: heliographical positions and areas of sun-spots observed with the Kew photoheliograph during the years 1862 and 1863, *Phil. Trans.*, **159**, 1–110.

Dezső, L., Gerlei, O., Kovács, Á.: 1987, *Debrecen Photoheliographic Results for the Year 1977*, Heliophysical Observatory of the Hungarian Academy of Sciences, Debrecen.

Erwin, E. H., Coffey, H. E., Denig, W. F., Willis, D. M., Henwood, R., Wild M. N.: 2013, The Greenwich Photo-heliographic Results (1874–1976): initial corrections to the printed publications, *Solar Phys*., in press

Győri, L.: 1989, *Problems in determining heliographic coordinates* (in Hungarian), Dr. Thesis, Kossuth Lajos University, Debrecen, Hungary.

JPL: 2005, JPL Horizons On-Line Ephemeris System. http://ssd.jpl.nasa.gov/?horizons.

Meeus, J. H.: 1991, *Astronomical Algorithms*, Willmann-Bell, Richmond, Virginia (Second Edition, 1998).

Royal Greenwich Observatory: 1975, *Royal Observatory Annals Number 11* (Photoheliographic Results 1967, Royal Greenwich Observatory, Herstmonceux), Science Research Council, Sumfield and Day Limited, Eastbourne, East Sussex.

Seidelmann, P. K. (Ed): 1992, *Explanatory Supplement to the Astronomical Almanac: a Revision to the Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac*, University Science Books, Mill Valley, California.

Willis, D. M., Davda, V. N., Stephenson, F. R.: 1996, Comparison between oriental and occidental sunspot observations, *Q. J. Roy. Astron. Soc.*, **37**, 189–229.

Willis, D. M., Coffey, H. E., Henwood, R., Erwin, E. H., Hoyt, D. V., Wild, M. N., Denig, W. F.: 2013, The Greenwich Photo-heliographic Results (1874–1976): summary of the observations, applications, datasets, definitions and errors, *Solar Phys*., in press.

Legends for the Figures

Figure 1 The annual accumulations of daily sunspot positions and areas are shown for the following selected years: (a) 1874; (b) 1883; (c) 1887; (d) 1911; and (e) 1914. In this figure the reconstructed solar images for each day of the year are superposed, starting the superposition on January 01 and progressing chronologically to December 31. The upper frames show the apparently anomalous annual accumulations of sunspots plotted in polar coordinates and the lower frames show the apparently normal annual accumulations of sunspots plotted in heliographic coordinates.

Figure 2 The coordinate systems used in the formulation of the mathematical equations required to convert from polar coordinates to heliographic coordinates and vice versa. (a) The figure on the left defines the angles ρ , ρ' and S , which appear in Equations (1) and (2). (b) In this illustration, the curved lines represent great circles on the solar hemisphere and the dotted lines lie in the plane of the observed solar disk. The figure on the right defines explicitly the angles P_0 , χ and λ , and it also defines implicitly the angles $\phi\,(P_S C P_E)$, $B_0\,(C_S C C_E)$ and $L_0\,(C_E C \lambda_0)$, where λ_0 signifies $\lambda = 0$; these six angles appear in Equations (3) – (5). Figure 2 is an adaptation of Figures 8 and 9 in the report by Dezső, Gerlei, and Kovács (1987).

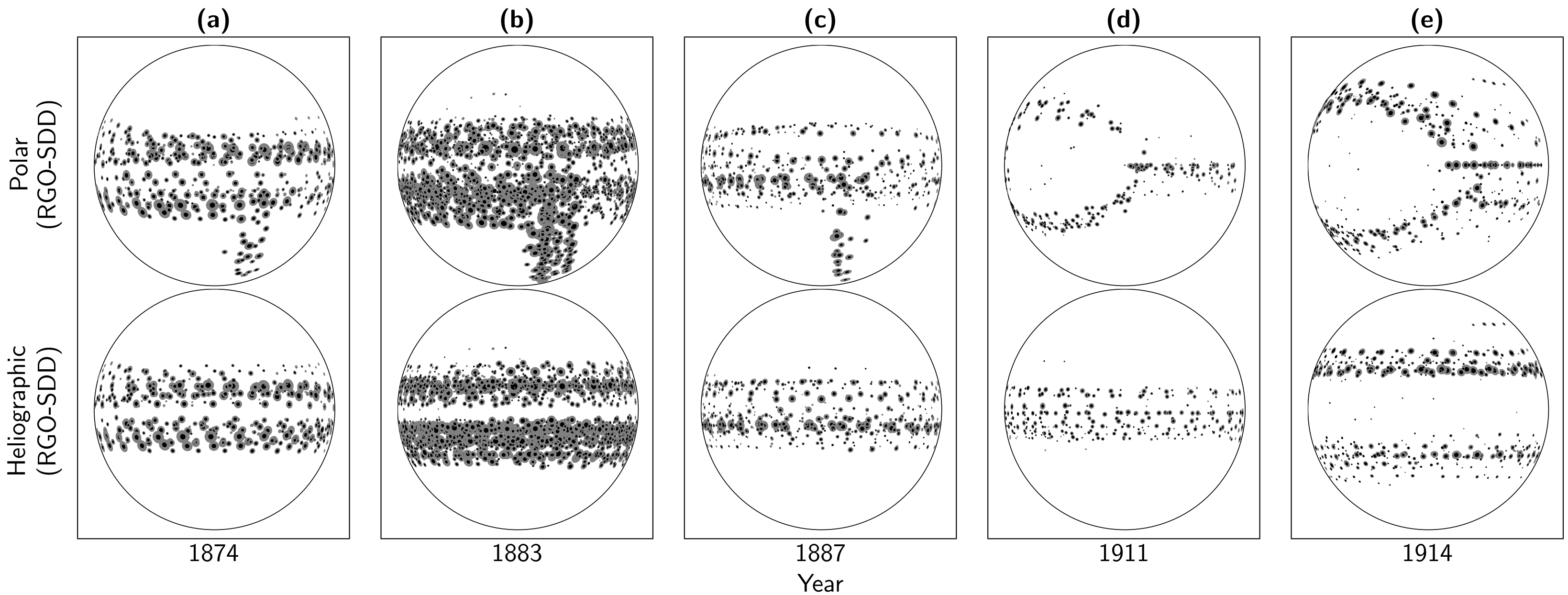
Figure 3 Differences between calculated and recorded values of the polar coordinate r/R (radial distance) for small discrepancies $\Delta(r/R)$ lying in the restricted range $-0.01 \le \Delta(r/R) \le +0.01$. The calculated values of r/R are obtained from the reverse solution of the appropriate mathematical equations (see text), assuming the original heliographic coordinates are correct: the recorded values of r/R are obtained from the original sunspot digital dataset (RGO–SDD). The differences have been calculated using solar ephemerides obtained from both the geocentric algorithm of J. Meeus (un-shaded histogram) and the topocentric algorithm of P. T. Wallace (shaded histogram).

Figure 4 Same as Fig. 3 but showing the differences between calculated and recorded values of the polar coordinate χ (position angle) for small discrepancies $\Delta\chi$ lying in the restricted range $-1.0 \leq \Delta\chi \leq +1.0$.

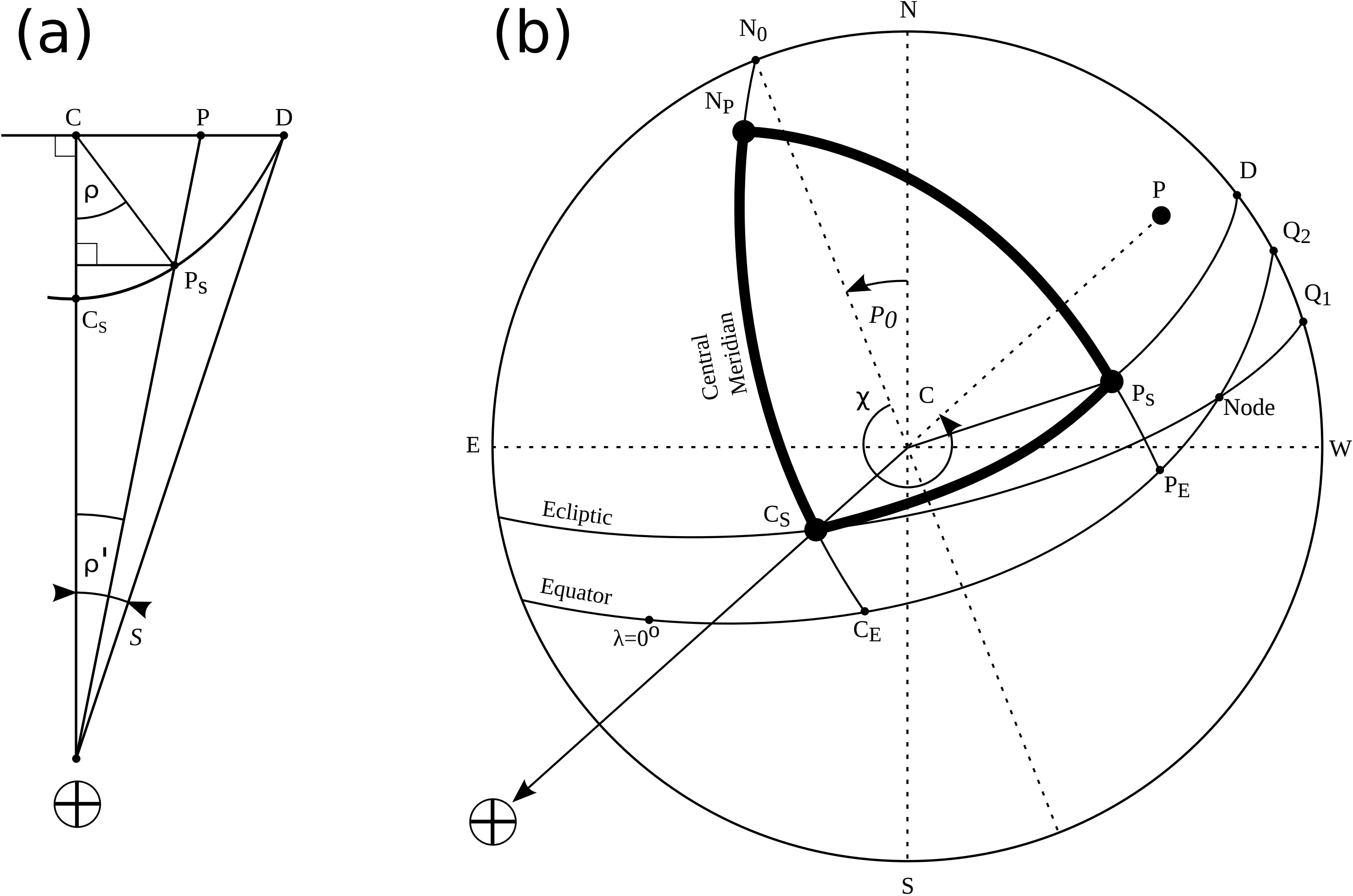
Figure 5 The procedures used to produce Figure 3 have been repeated using the version of the original sunspot dataset stored at the Marshall Space Flight Center (MSFC). However, the differences between calculated and recorded values of the polar coordinate r/R (radial distance) for small discrepancies $\Delta(r/R)$ lying in the restricted range $-0.01 \leq \Delta(r/R) \leq +0.01$ have been evaluated using just the geocentric algorithm for deriving solar ephemerides. It should be noted that there is a small offset of about 0.002 in the difference between the calculated and recorded values of r/R in the case of the MSFC dataset, which requires further investigation.

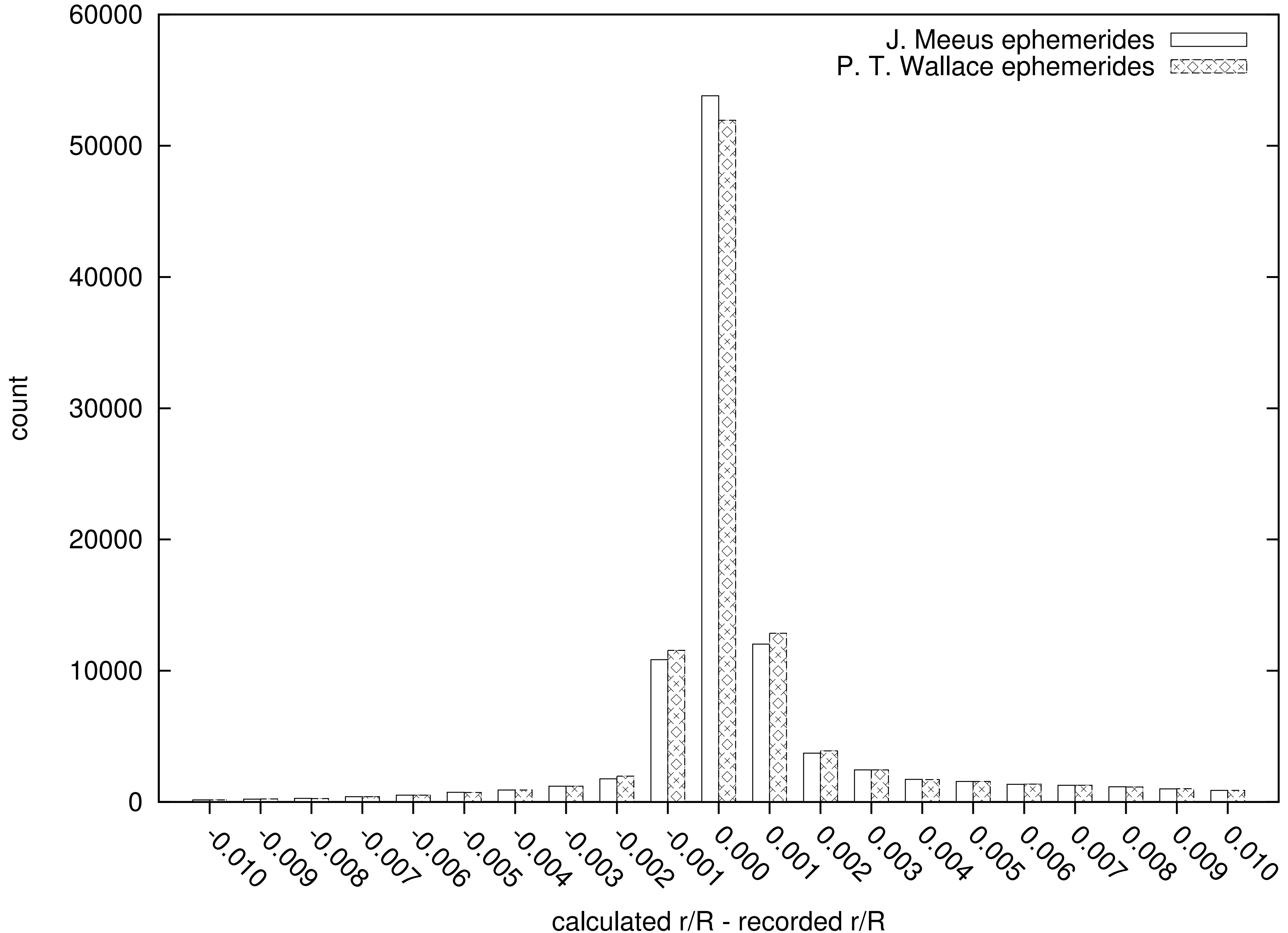
Figure 6 The procedures used to produce Figure 4 have been repeated using the version of the original sunspot dataset stored at the Marshall Space Flight Center (MSFC). However, the differences between calculated and recorded values of the polar coordinate χ (position angle) for small discrepancies $\Delta\chi$ lying in the restricted range $-1.0 \leq \Delta\chi \leq +1.0$ have again been evaluated using just the geocentric algorithm for deriving solar ephemerides.

Figure 7 The amended version of Figure 1 with the annual accumulations of daily sunspot positions and areas plotted in revised polar coordinates added as the lower (bottom) frames of the figure. For all five years, the annual accumulations of daily reconstructed solar images are shown in original polar coordinates in the top frames, in (original) heliographic coordinates in the middle frames, and in revised polar coordinates in the bottom frames. The apparently anomalous annual accumulations of sunspots plotted in original polar coordinates (top frames) disappear if these annual accumulations are plotted in revised polar coordinates (bottom frames).

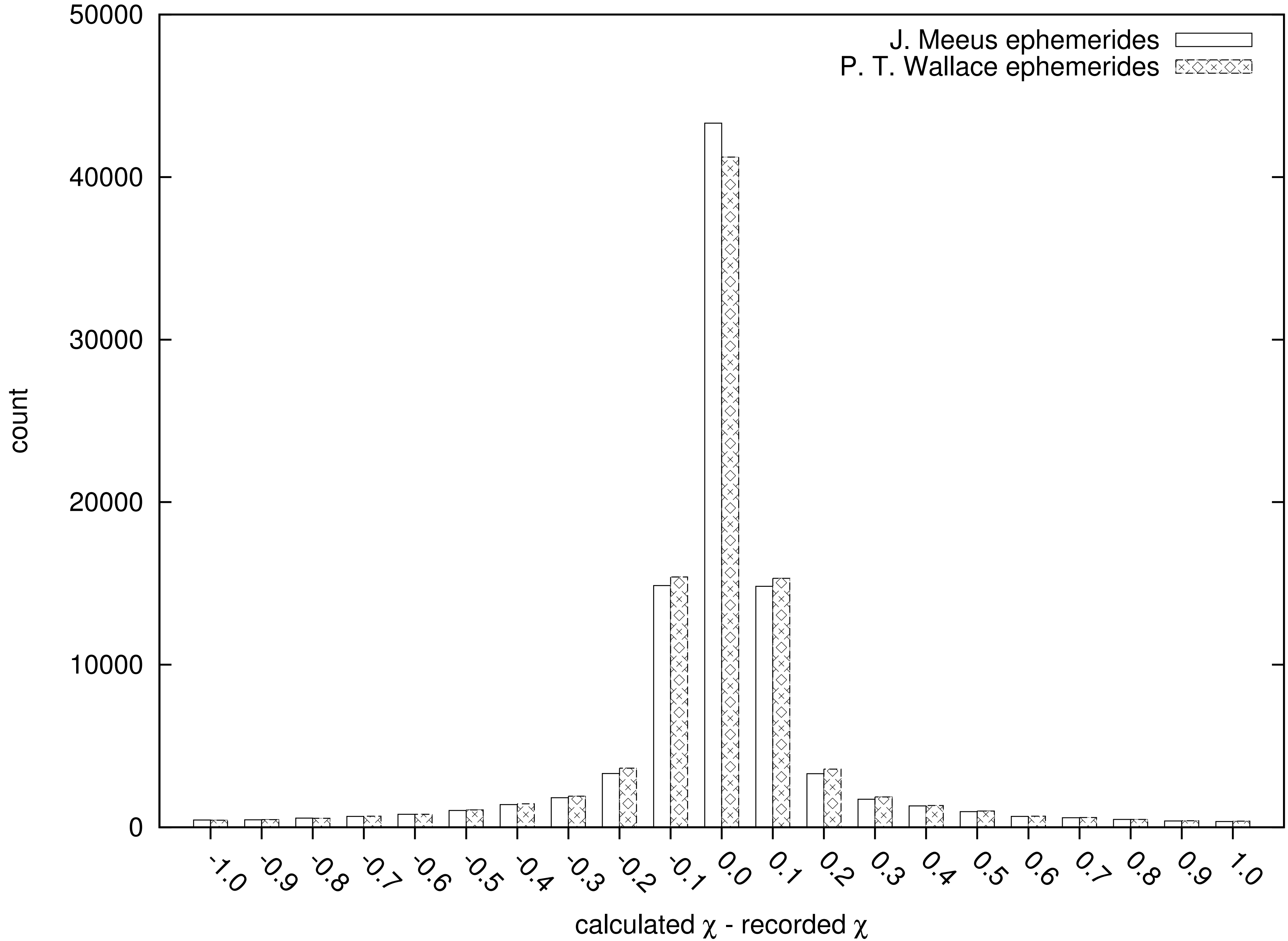


(a)
(b)
(c)
(d)
(e)
Polar
(RGO-SDD)
Heliographic
(RGO-SDD)
1874
1883
1887
1911
1914
Year

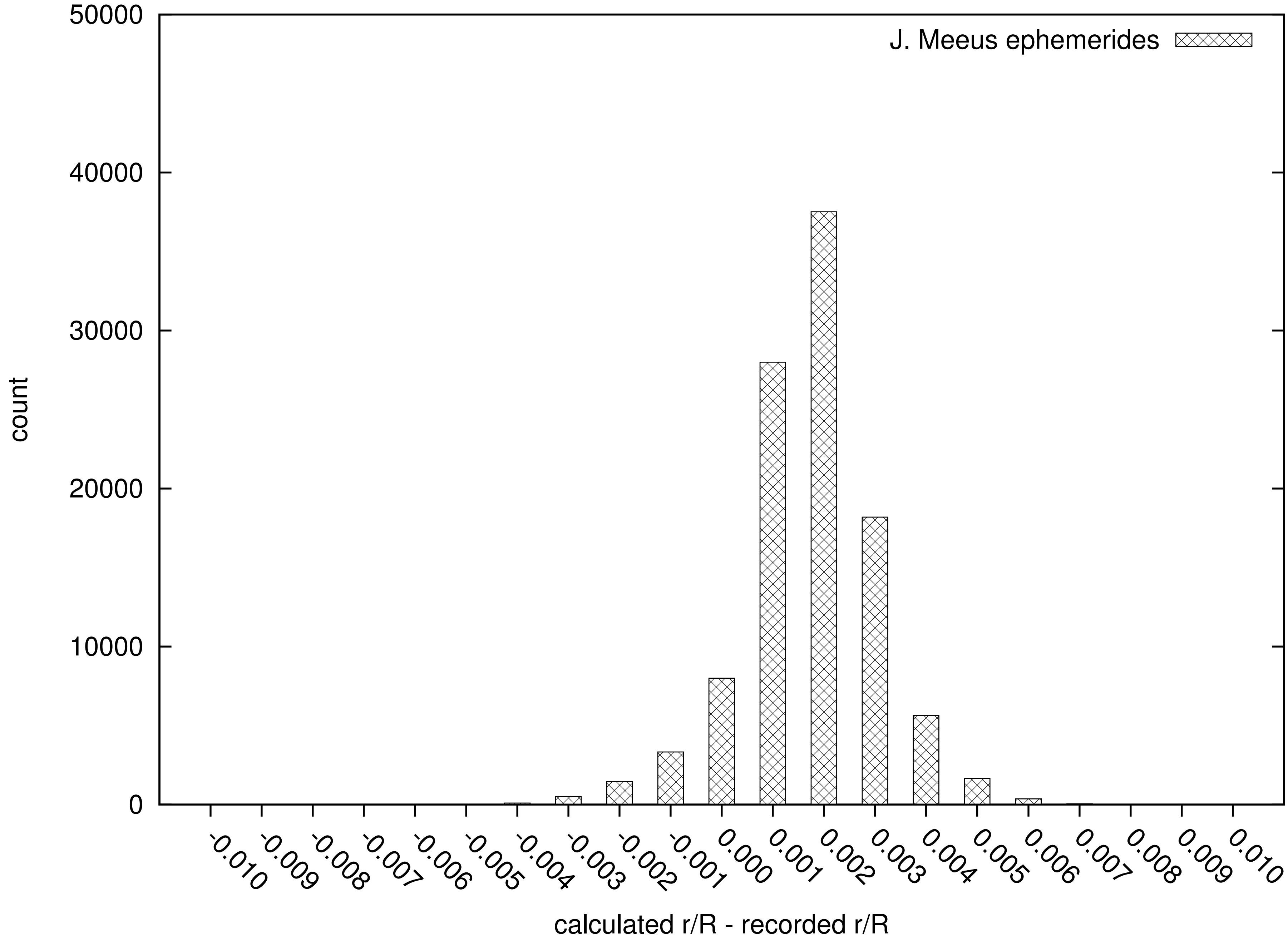

(a)
C
P
D
ρ
P_S
C_S
ρ'
S
(b)
N
N_0
N_P
D
P
Q_2
Q_1
P0
Central
Meridian
χ
C
P_S
Node
E
W
P_E
Ecliptic
C_S
Equator
λ=0°
C_E
S



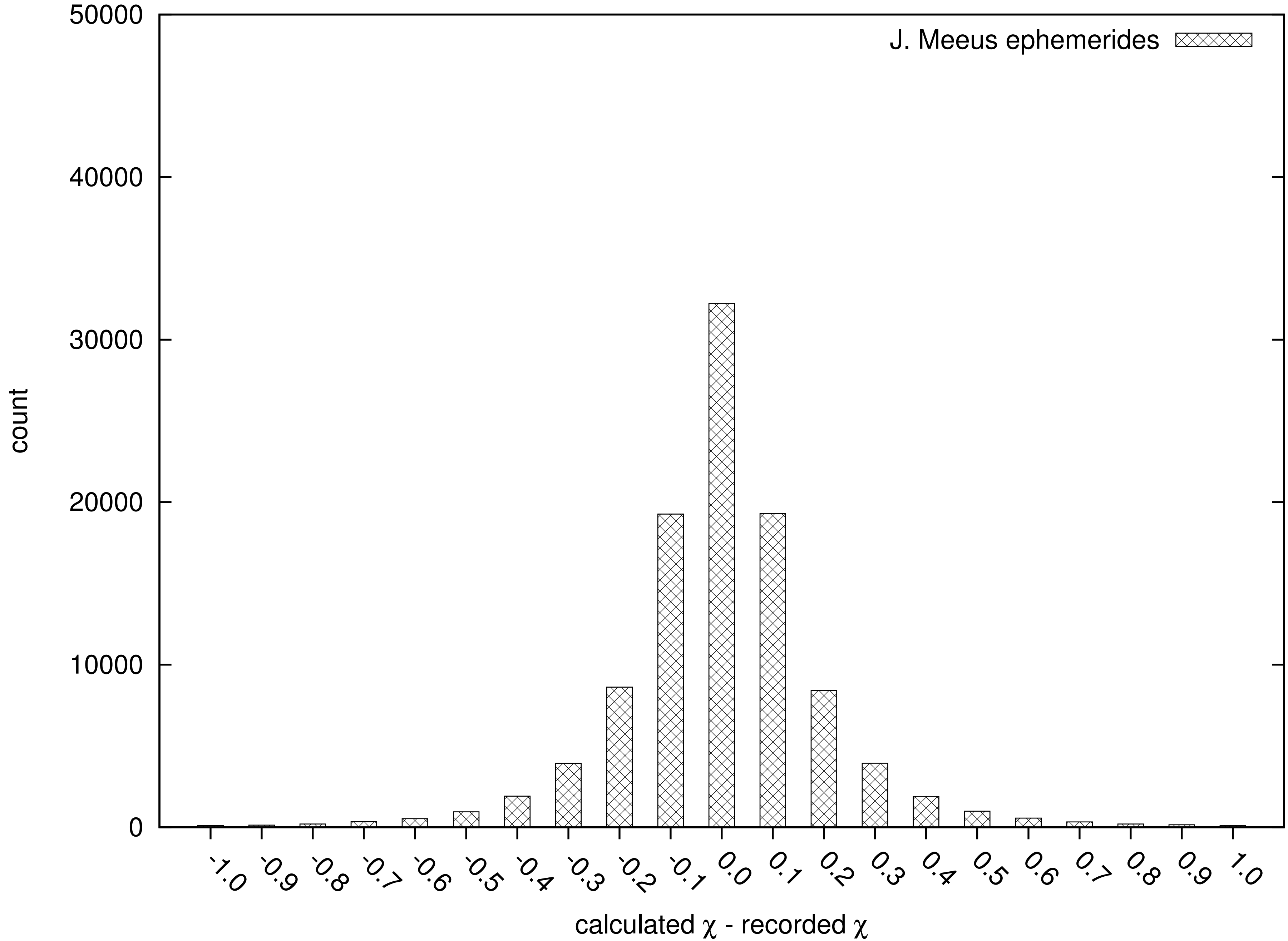
60000
50000
40000
30000
20000
10000
0
count
J. Meeus ephemerides
P. T. Wallace ephemerides
-0.010
-0.009
-0.008
-0.007
-0.006
-0.005
-0.004
-0.003
-0.002
-0.001
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.009
0.010
calculated r/R - recorded r/R



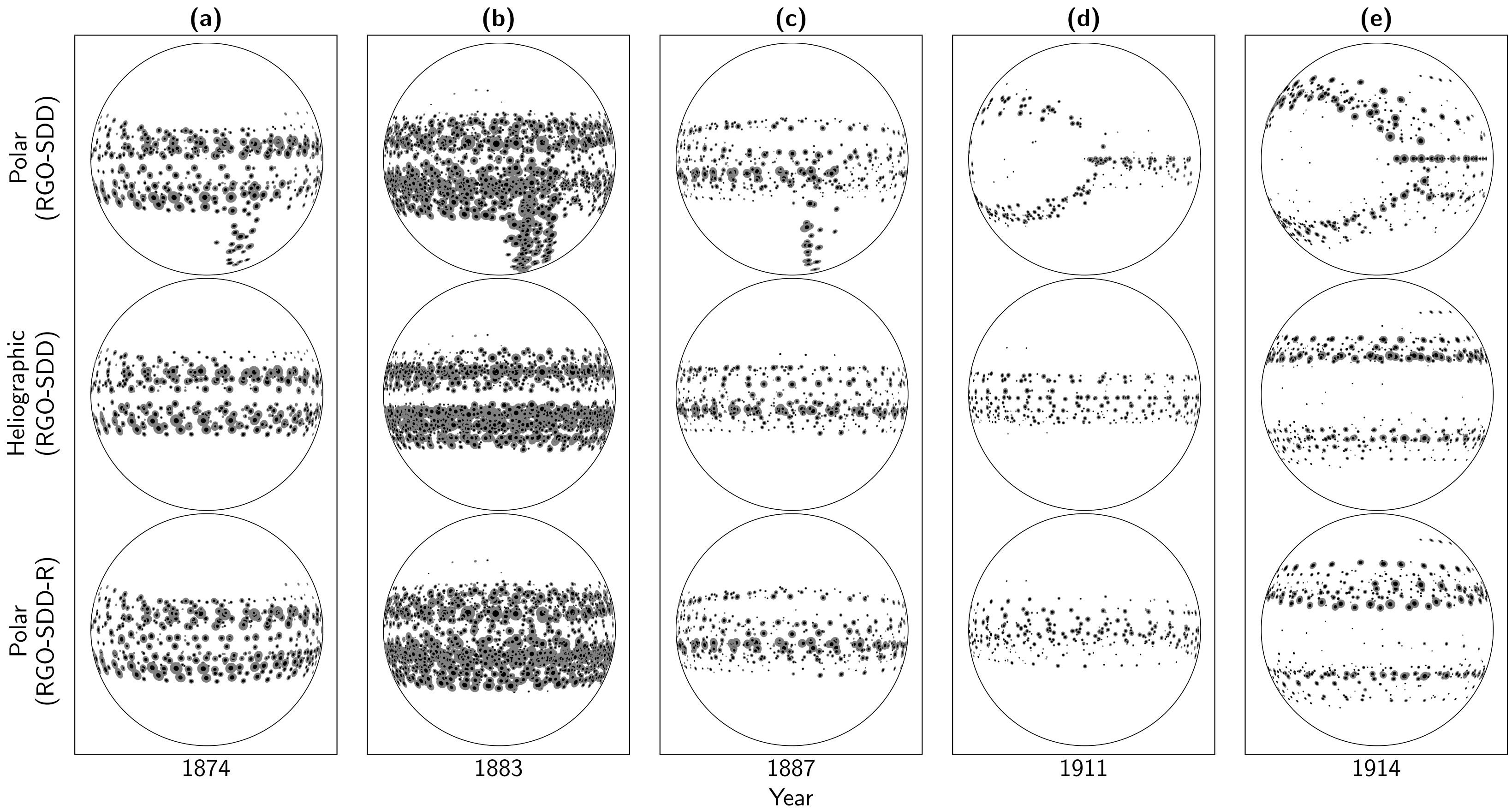
J. Meeus ephemerides
P. T. Wallace ephemerides
count
50000
40000
30000
20000
10000
0
-1.0
-0.9
-0.8
-0.7
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
calculated χ - recorded χ



J. Meeus ephemerides
count
50000
40000
30000
20000
10000
0
-0.010
-0.009
-0.008
-0.007
-0.006
-0.005
-0.004
-0.003
-0.002
-0.001
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.009
0.010
calculated r/R - recorded r/R



J. Meeus ephemerides
count
50000
40000
30000
20000
10000
0
-1.0
-0.9
-0.8
-0.7
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
calculated χ - recorded χ



(a)
(b)
(c)
(d)
(e)
Polar (RGO-SDD)
Heliographic (RGO-SDD)
Polar (RGO-SDD-R)
1874
1883
1887
1911
1914
Year