1

Introduction

Astrometry is positional astronomy. It encompasses all that is necessary to provide the positions and motions of celestial bodies. This includes observational techniques, instrumentation, processing and analysis of observational data, positions and motions of the bodies, reference frames, and the resulting astronomical phenomena.

The practical side of astrometry is complemented by a number of theoretical aspects, which relate the observations to laws of physics and to the distribution of matter, or celestial bodies, in space. Among the most important are celestial mechanics, optics, theory of time and space references (particularly with regards to general relativity), astrophysics, and statistical inference theory. These scientific domains all contribute to the reduction procedures which transform the observed raw data acquired by the instruments into quantities that are useable for the physical interpretation of the observed phenomena. The goals of this book are to present the theoretical bases of astrometry and the main features of the reduction procedures, as well as to give examples of their application.

Astrometry is fundamental to, and the basis for, all other fields of astronomy. At minimal accuracy levels the pointing of telescopes depends on astrometry. The cycle of days, the calendar, religious cycles and holidays are based on astrometry. Navigation and guidance systems are based on astrometry, previously for nautical purposes and now primarily for space navigation.

Astronomy and astrophysics are also strongly dependent on astrometry. The dynamics in the Solar System and in other gravitationally linked celestial bodies, stellar motions as obtained from the determination of proper motions, masses of double and multiple star systems, distances of stars in the vicinity of the Sun, apparent dimensions and shapes of planets, stars and other celestial objects, are all derived from astrometric observations.

Improvements in accuracies of observational techniques have been accompanied by increases in knowledge and improvements in theories. This process should continue in the future.

A very touchy but essential feature of astrometry is the expression of the uncertainties of the observations. It is customary, and actually very important, to distinguish precision, which describes the capability of an instrument to reproduce the observations, and accuracy, which represents how much these results deviate from what are thought to be the true values (which are by nature not attainable). The precision is improved by increasing the power, the resolution, or the sensitivity of the instrument. The accuracy is improved by measuring, or modeling, all effects that possibly can bias the results. This is done during the reduction process, even though it may mean additional measurements (temperature or humidity of air, optical distortions of the instrument, etc.). A given instrument's observations of stars may have a precision of 20 milliarcseconds (mas), but when compared with a catalog from many instruments, one may infer that its accuracy is 50 mas (assuming, however, some evaluation of the accuracy of the reference catalog itself). Evaluating the accuracy of observations is one of the most difficult and fundamental tasks of astrometry.

1.1 Classical astrometry

Astrometry began with the first naked-eye observations of the sky. These observations led to the discovery of *wanderers* (planets) among the stars, knowledge of the cycles of days and years, predictions of eclipses, the concept of almanacs and calendars, and the catalogs of observations. The naked eye limited both the faintness of objects and the accuracies of the positions observed. The observations of phenomena, such as eclipses and occultations, along with the location on Earth of the observer are the most accurate observations from ancient times.

From the beginning, astrometric observations were the basis for theories and discoveries. In 150 BC, Hipparchus discovered precession, the long period (26 000 year) motion of the Earth's pole with respect to the fixed stars, from observations with a precision of 1200″ (Dick, 1997). This was possible due to observations of Spica from 160 years earlier. Hipparchus may have prepared the first star catalog, but certainly Ptolemy's (AD 150) star catalog, as in his *Almagest*, is the first passed down from ancient times. Ptolemy developed theoretical explanations of the motions of the wanderers around the Earth. In 1496 Copernicus concluded that the wanderers moved around the Sun, not the Earth.

The Islamic culture produced the star catalogs of Al-Sufi (960), using armillary spheres, and Ulugh Beg (1430), who used a huge sextant, of precision of the order of 5–10′. Tycho Brahe (1546–1601) constructed improved instruments using sights and scales on quadrant- and sextant-type instruments. His observations, at a precision of 15–35″, had a wide impact on astronomy. However, even these improved observations were unable to measure stellar parallax, the angle through which a star seems to be displaced due to the orbital motion of the Earth, and this raised questions about the Copernican theory, since the parallactic displacement due to the motion of the Earth around the Sun was not detected. On the other side, the observations of the planets, in particular Mars, which was the only planet whose observations were of the necessary precision, led Kepler to develop his laws of motion, which in turn led to Newton's Universal Law of Gravity. The invention of the telescope permitted Galileo to discover the satellites of Jupiter and make astrometric observations of those objects, the first astrometric observations made with a telescope.

The application of astrometry for navigation and time-keeping led to the establishment of national observatories in Paris (1667), Greenwich (1675), Berlin (1701), and St Petersburg (1725). Telescopes were mounted on instruments similar to those used by Tycho Brahe. At the Royal Observatory at Greenwich, John Flamsteed produced the *Historiae Coelestis* (1725), the first great star catalog based on telescopic observations. In 1718 Edmond Halley, who would succeed Flamsteed as Astronomer Royal, showed that the bright stars, Aldebaran, Sirius, and Arcturus, were displaced by many minutes of arc from their positions in antiquity. Thus, the star positions are not fixed, but they move, perpendicular to the line of sight, by stellar proper motions. In 1728 James Bradley, Halley's successor, discovered a stellar aberration of 30″ due to the Earth's orbital motion. Aberration is the displacement of the angular position of an object due to the finite speed of light, in combination with the motion of the observer and the observed object. In 1748 Bradley detected the periodic motion of the Earth's celestial pole with respect to the stars due to nutation, an effect which can amount to 18″. William Herschel, who discovered Uranus in 1781, detected the motion of the Sun toward the constellation Hercules from an analysis of stellar proper motions.

Improvements in the precision of astrometric observations were sufficient that in 1838–40 Fredrich Wilhelm Bessel, Wilhelm Struve, and Thomas Henderson independently announced observations of a few tenths of an arcsecond precision showing that star positions shifted as the observer changed position in the Earth's orbit. Thus, the effect of parallax was observed, the truth of the Copernican theory solidified, and a method of directly measuring accurate stellar distances was established. In 1844 at Pulkovo, Bessel, from positional observations of 0.″7 and proper motions of 0.″5/year precision, announced unseen stellar companions, which were producing variations in proper motions of Procyon and Sirius. Direct telescope observations confirmed the companions later in the century. In 1885 Chandler detected, from stellar measurements, the motion of the Earth's axis of rotation with respect to the Earth's crust. This is called polar motion and amounts to 0.″5/year.

For most of the twentieth century, the primary instrument for large-angle astrometry was the transit circle. The transit instrument, which could observe only right ascensions, was first used by Ole Roemer in Copenhagen in 1689. The transit circle, which has a circle to measure declination, was succesfully introduced by Troughton in 1806 for Groombridge. By the mid nineteenth century most observatories had transit circles and for 150 years they were the prime instruments of large-angle astrometry. The observations from transit circles were the bases for the positions and proper motions of the fundamental catalogs through the FK5, which provided the astronomical reference frames.

The transit circle is a specialized telescope that can be moved only along the meridian, a North–South arc passing through the observer's zenith. This restriction of motion and the rigid mounting provides the stability necessary for precise observations. Observations are made in the East–West coordinate (right ascension) by timing when the star crosses, or transits, the meridian due to the Earth's rotation, while the North–South coordinate (declination) is measured on a very finely divided circle.

While the transit circle was good for accurate measurements of individual stars on the hour circle of the instrument all over the sky, the photographic plate could observe many stars in a small field of view. David Gill began a photographic survey of the southern sky from Cape Observatory at the end of the nineteenth century. The plates were measured and the positions analyzed by J. C. Kapteyn. The 1887 International Astrophotographic Congress in Paris coordinated the first international photographic survey of the sky, the *Carte du Ciel*, and the resulting *Astrographic Catalogue* (AC). This sky survey reached about 12th magnitude, while subsequent photographic surveys using Schmidt telescopes have gone as deep as 21st magnitude. Schmidt telescopes achieve wide fields of view by a correcting glass at the center of curvature and a concave focal plane, which the plate must be bent to form.

Long-focus refractors were being used at the end of the nineteenth century in small fields for observations of double stars, asteroids and satellites. The advantage of long-focus astrometry is that it provides a large scale on the plate, proportional to the focal length. One gets better precision in positions of star images, namely a few hundreds of arcseconds. A large number of stars are thought to be in multiple star systems. In some cases the components can be detected individually; others can be detected photometrically by changes in magnitude, some by astometry due to their nonlinear proper motions, some by periodicities in their radial velocities and others by multiple spectral characteristics (Chapter 12). At the beginning of the twentieth century Frank Schlesinger at Yerkes, Allegheny and Yale Observatories began the application of photography on long-focus refractors for the determination of parallaxes (Section 11.2).

1.2 New astrometry

The last decades of the twentieth century have experienced a revolution in astrometry. This was because of the charge coupled device (CCD) as a detector replacing the photographic plate and measuring machine, optical interferometers able to observe at milliarcsecond levels, space astrometry pioneered by the Hipparcos project, and radio astrometry by Very Long Baseline Interferometers (VLBI, Section 2.4.2) producing a celestial reference frame of extragalactic sources at milliarcsecond accuracy levels.

The CCD is a silicon chip that converts photons into electrons. The CCD directly provides a digital output, eliminating the need for the measuring machine. It is a small-size detector compared with the photographic plate, but it is extremely efficient, more than 50%, dependent on the filter used, compared with 1% for photographic plates (see Sections 2.1 and 14.3). Also, the data can be processed to eliminate backgrounds, subtract bright stars, etc. The small pixel sizes and the point spread function across many pixels provides the means of measuring the centroids of star positions very accurately. As a result, parallaxes of faint stars with respect to the mean parallax of background stars (relative parallaxes) can be determined with accuracies of better than 0.5 milliarcseconds (mas); and relative positions with 20 mas accuracies are possible from ground-based observations.

Albert A. Michelson, in the late nineteenth century, developed interferometry, which uses the interference of light waves for its measurement. The first stellar interferometer was not built until 1920 and only in the 1970s was it used for astrometry, other than double-star observations. Optical interferometers (Section 2.3) are currently being developed for astrometric positional measurements at the mas level, double-star measurements, and imaging of individual stars.

The Hubble Space Telescope (Section 2.5.2) is capable of astrometric observations with the Wide Field Planetary Camera, and the Fine Guidance Sensors. Double stars, parallaxes, proper motions in clusters, planetary satellite positions, and planetary nebulae motions have been measured (Section 2.4). The Hipparcos satellite (Section 2.5.1) was developed as an astrometric satellite to determine parallaxes. It has determined the positions and proper motions of 120 000 stars to accuracies of 1 mas and 1.1 mas/year for stars of 9th magnitude and brighter and parallaxes for 60 000 stars to 1 mas accuracies.

There are now plans for interferometer and astrometric satellite missions that could reach microarcsecond accuracies for stars as faint as 20th magnitude. The Space Interferometer Mission (SIM) and the Global Astrometric Instrument for Astrophysics (GAIA) are two such missions under study (Section 2.6).

In parallel with these developments in optical astrometry, the field of radio interferometric techniques (Section 2.4.2) applied to astrometry was developed primarily for geodesy and Earth orientation purposes, but the accuracies and the wealth of data have had a significant impact on astrometry via the recently adopted International Celestial Reference System (ICRS).

Observations at a few microarcsecond precision imply that, in parallel, one must model the light path between the source and the observer with a similar accuracy (Chapter 6). This means, in particular, that one uses general relativity to describe the space-time (Chapter 5). Another consequence is that sophisticated mathematical methods must be used for the treatment of astrometric data (Chapter 4).

1.3 Time

From early times people kept track of time based on the apparent diurnal motion of the Sun. This is called apparent solar time and undergoes seasonal variations, because of the obliquity of the ecliptic and the eccentricity of the Earth's orbit. This is the time from sundials. Clocks and chronometers are not subject to the variations of the Earth's orbit, so mean solar time was introduced based on a fictitious mean Sun with a constant rate of motion and Earth rotation. It was known that the diurnal rotation with respect to the stars was different from that with respect to the Sun. Sidereal time is defined by the apparent diurnal motion of the catalog equinox, and, hence, is a measure of the rotation of the Earth with respect to the stars, rather than the Sun. The rotation of the Earth with respect to the Sun is called Universal Time (UT), and is related to sidereal time (now, stellar angle), by means of a numerical formula (see Section 10.5).

For ephemerides of Solar System bodies' motions, the tabular times are the values of a uniform measure of time. Discrepancies between observed and computed positions were most evident for the Moon, because its motion is both complicated and rapid. Adams (1853) showed that the observed secular acceleration of the mean motion of the Moon could not be produced by gravitational perturbations. Ferrel (1864) and Delaunay (1865) showed that the tides exert a retarding action on the rotation of the Earth, accompanied by a variation of the orbital velocity of the Moon in accordance with the conservation of momentum. Newcomb (1878) considered the possibility of irregular variations of the Earth's rotation as an explanation of lunar residuals, but he could not find collaboration from planetary data (Newcomb, 1912). De Sitter (1927) and Spencer Jones (1939) correlated irregularities of the motions of the inner planets with those of the Moon, thus proving the irregularity of the rotation of the Earth.

Danjon (1929) recognized that mean solar time did not satisfy the need for a uniform time scale, and suggested a time scale based on the Newtonian laws of planetary motion. Unaware of Danjon's paper, Clemence (1948) proposed a uniform fundamental standard of time which was adopted by the International Astronomical Union (IAU) in 1952 and designated as *Ephemeris Time* (ET). The tropical year of 1900.0, the period of one complete revolution of the longitude of the Sun with respect to the dynamical equinox, was the basis for the definition of Ephemeris Time.

Until 1960, the unit of time, the second, was defined as a specific fraction (1/86 400) of the mean solar day. With the adoption of ephemeris time, the second was defined in the Système International (SI) of units as a specific fraction of the tropical year. This definition was more accurate, but its determination was less precise, because the motions of the Sun and Moon being slow, one could not transform the observations of their positions, made to a few tenths of an arcsecond, into time measurements of better than a few hundredths of a second. So, in 1967, a new definition of the second was adopted, following a proposal by Markowitz *et al.* (1958), in terms of a cesium beam frequency standard. Since then, it has been the fundamental unit of time. Whenever another unit is used, it has to be specified (e.g. ET second, etc.).

The second is defined as the duration of 9 192 631 770 periods of radiation corresponding to the transition between two hyperfine levels of the ground state of the cesium-133 atom. International Atomic Time (TAI) is a practical time scale determined by the Bureau International des Poids et Mesures from time and frequency standards worldwide and conforms as closely as possible to the definition of the second (Section 5.5.1). The second is the unit that is the most accurately and precisely realized at present, and thus, with the conventional value of the speed of light, is the basis for the definition of the meter. Coordinated Universal Time (UTC) is based on TAI and kept within 0.9 s of Universal Time (UT1) by the introduction of leap seconds as needed. UTC is the standard time scale defined for the zero meridian and used with appropriate time differences for time zones all over the Earth (Section 10.5.3).

Ephemeris Time was difficult to determine and its definition was not consistent with the theory of relativity. In 1976 and 1991 the IAU adopted resolutions clarifying the relationships between space-time coordinates. Terrestrial Time (TT), previously named Terrestrial Dynamical Time (TDT), is the time reference for apparent geocentric ephemerides, has a unit of measurement so that it agrees with the SI second on the geoid, and is offset from TAI by +32.184 s. Geocentric Coordinate Time (TCG) is the coordinate time appropriate for a coordinate system with its spatial origin at the center of mass of the Earth (Section 5.5.1). TCG differs from TT by a scaling factor. Barycentric Coordinate Time (TCB, section 5.5.2) is the coordinate time appropriate for a coordinate system with its spatial origin at the barycenter of the Solar System. The relationship between TCB and TCG involves a full four-dimensional transformation involving the position and velocity of the center of mass of the Earth, a scaling factor, and periodic terms. Thus, the resulting space-time coordinates are consistent with the theory of relativity.

The time scale Barycentric Dynamical Time (TDB) was introduced in 1976 to be a time scale for the barycenter of the Solar System such that it differed from TDT only by periodic terms. This results in a definitional ambiguity dependent upon the length of time being considered. Also, the definition results in different values of some constants, such as the unit of length, to be used with the two time scales. These time scales can be precisely determined from atomic clocks and the theory of relativity. They do not consider the variability of the rotation of the Earth as needed for actual observations.

1.4 Earth orientation and reference frames

During the first three quarters of the twentieth century, the rotation of the Earth and polar motion were determined by observations with visual or photographic zenith tubes, transit circles, and astrolabes. Then the techniques of lunar and satellite laser ranging (Section 2.4.3), radio interferometry (Section 2.4.2), and the Global Positioning System (Section 2.4.1) became available and much better accuracy could be achieved. The International Earth Rotation Service (IERS) was founded to coordinate these activities and to provide international values for polar motion and other Earth rotation parameters.

In determining Earth orientation parameters by means of Very Long Baseline Interferometry (VLBI) many observations of radio sources have been made. The accuracies of the positions are about a milliarcsecond and the sources are very distant, so they have no appreciable proper motions. Thus, it became apparent that an accurate, fixed, stellar reference frame could be developed, improving the preceding FK5 reference frame.

A list of stable, point-like radio sources was needed. The ties between these generally optically faint sources and optical and dynamical reference frames were necessary. Such a list of sources was prepared and investigated for structural motions so that an International Celestial Reference Frame (ICRF) could be designated (Section 7.2). Optically this reference frame is represented by the *Hipparcos Catalogue* (Section 11.6.1). The reference frame is independent of the dynamical reference frame, defined by the equator and the ecliptic planes which are dependent on the Solar System bodies. The origin of the ICRF has been selected to be as close as possible to the FK5 origin and the dynamical equinox of J2000.0. However, in principle, the ICRF is based on observations from the Earth and the kinematics of the Earth's equator. A new Celestial Intermediate Pole (CIP), and Celestial Ephemeris Origin (CEO) were adopted in 2000 by the IAU. Solar System observations can be made on this reference frame and the equinox and ecliptic determined on that system.

The Hipparcos satellite has provided milliarcsecond accuracies for optical sources. VLBI is determining radio sources at sub-milliarcsecond levels and detecting motions of 25 microarcseconds per year. Optical interferometers in space are planned for microarcsecond accuracies. Thus, astrometry must be developed fundamentally on an extragalactic reference frame to microarcsecond accuracies.

2

New observational techniques

What characterizes the new observational techniques is, of course, not only that they appeared recently – say in the past couple of decades – but also that they strive to reach one or two orders of magnitude better accuracy than the classical techniques. The unit for describing the new astrometric capabilities is no longer one tenth of a second of arc, but one thousandth of a second of arc, a subunit that we shall designate throughout this book by the abbreviation mas (milliarcsecond). Plans are underway to reach microarcsecond (μas) accuracies. Such a gain in precision impacts directly on all aspects of astronomy and, particularly, on the reduction procedures. One may divide these new techniques into three major groups: interferometry, time-measuring techniques, and space astrometry.

2.1 New detectors

Many astrometric techniques take advantage of the development of much more sensitive new detectors. Among them, one must mention the following.

- *The CCD detector*. The CCD (for charge coupled device) has become a major tool in astronomy since the 1970s (Monet, 1988). It consists of a semiconductor device, where a photoelectric effect takes place when light reaches it, producing an electronic image. This image is preserved by arrays of small positive electrodes, which attract photoelectrons and keep them in a similar array of potential wells, providing the possibility of long exposure times by adding photoelectrons in the same pack (Kovalevsky, 2002). Once the exposure is over, one shifts, by a periodic change of the potential of electrodes, the electronic image along the array. It is then recovered pack by pack at the edge of the array, digitized, and registered in a computer. There, software can analyse it and produce images, determine relative positions or magnitudes of the objects, or deduce any other information that is needed. It is also possible to synchronize the transfer speed with the image shift