61. Discovering variability with Gaia

HUNDREDS OF THOUSANDS of stars are, today, known to be variable. Some of this variability is truly 'intrinsic', being due to the luminosity of the star itself changing with time. In contrast, some stars classified as variable are 'extrinsic', meaning that the apparent variability occurs because something affects the stellar light on the way from the star itself to us on Earth.

Even to summarise the different *types* of variability would be lengthy: for example, `wiki/variable_star` lists around 50 types of intrinsic variables, and around 20 types of extrinsic variables. Since I want to focus here on how these are detected and classified with Gaia, let me give just a selective flavour of these variability types.

INTRINSIC VARIABLES include a huge range of pulsating star types. They change in brightness as they cyclically expand, and then contract, on periods ranging from days to weeks or months. The expansion phase is due to the blocking of the outflow of energy by gas with a high opacity, with expansion eventually halting as the density decreases, then reversing due to gravity.

In this class are the Cepheids, the RR Lyrae stars, along with Delta Scuti, SX Phoenicis, Mira, Beta Cephei, and Gamma Doradus variables. Their physics is reasonably well understood, and they vary over very different timescales, with very different amplitudes (extending up to many magnitudes), and sometimes along with multiple frequency (harmonic) components.

There are other intrinsic variables: eruptive variables show irregular or semi-regular brightness variations caused by material lost from, or accreted onto, the star. And there are 'flare' stares, along with cataclysmic or explosive variable stars including novae, dwarf novae, all the way up to the most explosive supernovae.

At much lower variability amplitudes, even our Sun (and along with it other main sequence stars) is variable. Our Sun varies over the 11-year solar cycle due to the long-term behaviour of its magnetic field. And at even lower amplitudes, the Sun oscillates in a large number of modes (with periods around 5 minutes), driven stochastically by turbulent convection in its outer layers.

EXTRINSIC VARIABLES include rotating stars which appear to vary due to surface sunspots rotating into and out of the line of sight, and other rotating stars which appear to be variable due to their underlying ellipsoidal shape. The huge class of eclipsing binaries vary because of a companion orbiting binary star which can regularly eclipse the light of its companion. Stars with orbiting planets may also show small brightness variations if their planets pass between Earth and the star, and many thousands have been discovered by NASA's Kepler satellite mission and others. Gravitational lensing provides another, rarer, form of extrinsic variability.

Detecting, classifying, and characterising such stellar variability, whether intrinsic or extrinsic, provides a vast and rapidly growing panorama for an almost endless range of studies of the physics and environment of the stars in our Galaxy and beyond.

Historically, variable stars were first discovered through naked-eye observations of individual stars, and more recently through very large-scale sky surveys (such as Pan-STARRS, ASAS, and many others) which return to the same parts of the sky again and again, and at different cadences, to detect and categorise variability on all sorts of time scales, and at all sorts of amplitudes. The most ambitious of these, under development today, is the 8.4-m Vera Rubin telescope (or LSST), which will photograph the available sky every few nights.

BECAUSE OF ITS sky scanning above the atmosphere, Gaia is contributing enormously. Its strength is not so much the *number* of observations of a given star over its possible 10-year mission, but rather its all-sky visibility, and its unprecedented photometric *accuracy*. This is in part due to its detection system, but in particular its location in space, above the Earth's atmosphere.

Gaia should also detect a reasonable number of exoplanet transits and microlensing events, although asteroseismic oscillations will lie below its detection threshold. But even within the arena of the classical pulsating stars, eruptive variables, and the ubiquitous eclipsing binary stars, Gaia's contribution is unprecedented.

Gaia scans the sky in a systematic but non-uniform 'revolving' manner, resulting in around 70 focal plane passages for an average star over 5 years; more at intermediate ecliptic latitudes, with fewer crossings at higher and lower latitudes. As the stars cross the focal plane, the images are effectively integrated, over the 9 successive astrometric CCDs, over about 40 s.

On-ground calibration then leads to (broad G-band) photometric errors, for some two billion stars, of around 0.3 mmag (for $G < 13$), 1 mmag (at $G = 17$), and 6 mmag (at $G = 20$), along with lower accuracy in two (blue and red) photometric channels (van Leeuwen et al., 2017).

Of data releases to date, Gaia DR2 (made public in April 2018) was based on observations between July 2014 and May 2016, while Gaia EDR3 (released in December 2020), included all observations to May 2017.

Through the work of the processing teams, 550 737 stars were classified as variable in Gaia DR2. While no new variability updates were provided with Gaia EDR3, many more variable stars are expected to be identified with the full Gaia DR3 release in mid-2022, and beyond.

Within the Gaia Data Analysis and Processing Consortium (DPAC), Coordination Unit 7 (CU7), led by Laurent Eyer from the Geneva Observatory, has the responsibility of detecting variable objects, classifying them, and deriving characteristic parameters for specific variability classes.

CU7 interfaces with other Coordination Units, each with specific (and often closely related) responsibilities. Thus, CU5 provides the calibrated photometric data, CU4 processes binary stars, CU6 processes the spectra and derives radial velocities, and CU8 extracts astrophysical parameters from all the available Gaia data.

In parallel to the variability processing by CU7, a Cambridge-led 'alerts team' analyses the Gaia data in near-real time to detect transient phenomena, such as supernovae, that benefit from rapid follow-up by the scientific community (Hodgkin et al., 2021).

The main activities of the CU7 processing are, sequentially, data import into a dedicated database, cross-matching of sources, execution of the various variability detection and classification modules, followed by export of the data to the other Coordination Units. These activities are performed iteratively, adding more observations, identifying errors, and improving the results.

Source-by-source processing allows for a parallel use of the computing cluster, in Geneva, that is gradually being expanded as the mission and data quantity progresses. In 2017, it was composed of 57 quad core nodes, totalling 456 simultaneous processes, each process having 4 GB of RAM. The incoming data from the Main Data Base of the Data Processing Centre at ESAC, Madrid, is stored in partitioned tables. The database is hosted on a single node with 40 cores and 256 GB of RAM.

The various steps in the processing of such a large and complex data set are many and varied (e.g. Eyer et al., 2017; Holl et al., 2018; Eyer et al., 2019). For example, observations tagged with the satellite on-board time are converted to the standard 'Julian epoch' time scale, and then transformed to the solar system barycentre.

The statistical processes of general variability detection aim to separate 'constant' stars from those showing any type of variability, while handling challenges such as the different number of source measurements, and the dependence of measurement error on star brightness.

The task of 'special variability detection' specifically targets short time-scale or small-amplitude periodicity, as well as variability induced by planetary transits and solar-like magnetic activity (e.g. Roelens et al., 2017).

The 'characterisation task' analyses the resulting time series. For periodic objects, variability is characterised using classical Fourier decomposition, with the process comprising various period-search methods.

The 'classification module' uses automated supervised statistical classifiers to assign probabilities that any given star is a member of one of a number of pre-defined variability types. Classification using well-established machine-learning techniques is performed using three different classifiers, known as Gaussian mixtures, Bayesian networks, and random forests.

Before Gaia, the number of known variables stars was around 150 000. DR2, the latest Gaia data release providing the results of the global variability analysis (Holl et al., 2018), contains 550 737 variable sources, comprising 228 904 RR Lyrae stars and 11 438 Cepheids (Clementini et al., 2019; Rimoldini et al., 2019), 151 761 long-period variables (Mowlavi et al., 2018), 147 535 stars showing rotation modulation, 8882 Delta Scuti and SX Phoenicis stars (Rimoldini et al., 2019), and 3018 short-timescale variables (Roelens et al., 2018).

As Holl et al. (2018) conclude, the DR2 release represents just a subset of the data processed to date, and future releases will include many more variable sources. Nonetheless, DR2 already shows the very high quality of the data, and Gaia's great promise for variability studies.

Scientific application of these results is only just starting. As one example, and at the broadest level, Eyer et al. (2019) have located each of these variables in a Galactic diagram of colour versus absolute magnitude, focusing on pulsating, eruptive, and cataclysmic variables, as well as on stars that show apparent variability due to stellar rotation and binary eclipses.

They could characterise the locations of the different variable star classes, variable object fractions, and typical variability amplitudes throughout the diagram. This demonstrates distinct regions in which variable stars occur, anticipating new insights into variability phenomena, and a greater understanding of stellar physics.