I study star formation using multiwavelength observations (X-ray, optical, infrared, and submillimetre), employing methods from statistics and machine learning to obtain astrophysical results. I use my empirical results to test predictions of the leading theoretical star-formation models (e.g., the global hierarchical collapse scenario versus the gravoturbulent scenario). These comparisons shed light on fundamental open questions in this field such as: 'What controls the rate of star formation?', 'Which star-formation events produce gravitationally-bound star clusters?,' and 'How do stars assemble their masses?' A few highlights from my research include:

- My highly cited study of the expansion of young clusters was the first to demonstrate that most stellar groups form dynamically hot (Kuhn et al. 2019; ∼230 citations).
- I generated one of the largest catalogues of disc-bearing young stars ($>10^5$ objects). Using this sample, we found that many Sagittarius Arm star-forming regions lie along a giant molecular filament with a discrepant orientation relative to the arm (Figure 1), indicating the need to revise the traditional model of the Milky Way's spiral arms (Kuhn et al. 2021a, b; collectively ∼50 citations).
- I constructed a mixture-model algorithm to detect clusters that deviate from Gaussian distributions while accounting for field-of-view boundaries. This resulted in an influential paper on subclustering of young stars (Kuhn et al. 2014; ∼120 citations) and an invited chapter in the statistical reference book *Handbook of Mixture Analysis*, published by Chapman & Hall/CRC.

Much of my research (including the above results) involves significant statistics or data science components. Starting while a PhD student, I have been involved with the professional statistics community in my work modelling the spatial distributions of young stars. I am an active member of the Cosmostatistics-Initiative, where I lead the SPICY project (described in the following sections). In the SPICY collaboration, we experiment with data-driven strategies for characterising young stars and their behaviours, e.g., statistical classifiers, deep-learning based imputation, hierarchical Bayesian models, spatial point pattern summary statistics, unsupervised image classifiers, interpretable machine learning, and light-curve outlier detection. These projects not only lead to scientifically impactful outputs, such as the ongoing string of SPICY publications, but have also fostered interdisciplinary collaboration between astronomy, statistics, and computer science.

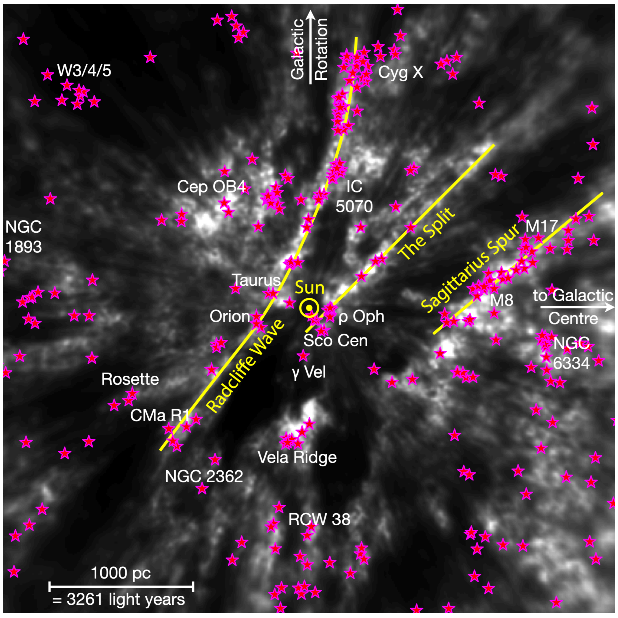

Figure 1: This is the most accurate map, to date, of star-forming regions (magenta stars) in the 2-kpc neighbourhood of the Sun. The 3D positions of the regions were inferred via Bayesian hierarchical modelling of Gaia astrometry for young stellar objects (Kuhn et al. 2021b; 2023 in preparation). These regions are plotted on the Vergely et al. (2022) dust map (greyscale background image), showing strong correlation between Galactic molecular clouds and star-forming regions. We discovered the "Sagittarius Spur" (marked in yellow), one of three large kpc-scale filaments that collectively explain most of the nearest star-forming regions.

Further details about my research areas are provided below.

1 Massive Star-Forming Regions

I am particularly interested in massive star-forming regions since they dominate the star-formation rate of our Galaxy and serve as analogues for the environment in which our Solar System formed. My PhD was based on an archival study of X-ray and infrared observations of ∼20 massive star-forming regions: 'The Massive Young Star-forming Complex Study in Infrared and X-Ray' (MYStIX). My work in this project (Kuhn et al. 2010, 2013a, b) provides a view into dense, embedded pre-main-sequence clusters at the centers of these regions. I am keenly interested in the spatial distributions of stars in these regions, since stellar subclustering provides insight into star-formation scenarios (Kuhn et al. 2014, 2015a, b). I am the PI of an accepted JWST/NIRSpec programme with the Extreme Ultraviolet Environments (XUE) Collaboration to obtain the largest spectroscopic sample of highly-ionized protoplanetary discs in Tr 14.

2 Kinematics of Young Stellar Clusters and Associations

I published the first Gaia-based study to deeply investigate the kinematics of partially embedded young stellar clusters (Kuhn et al. 2019a). In this study, I showed that (1) most (>75%) groups of young stars are expanding with typical mean expansion velocities of ∼0.5 km s^{-1} and outwardly increasing velocity gradients, (2) most young clusters are not significatly rotating, (3) their stellar velocity dispersions of 1–3 km s^{-1} are supervirial, and (4) subgroups of young stars show no tendency to coalesce into more-massive clusters. These results provide some of the strongest constraints on the early dynamical evolution of young stars. Later studies of mine have investigated various causes of expansion (gas expulsion, tidal disruption, etc.) and refined statistical models used for measuring expansion.

3 The Molecular Cloud–Young Star Connection

Joint kinematic analysis of stars and gas (probed by CO molecular line maps) links the dynamical state of the molecular cloud to the young stars. Figure 2 illustrates such an analysis of the North America/Pelican (= NGC 7000/IC 5070) region (Kuhn et al. 2020a, b). Gaia astrometry was used to select stellar members and divide them into subclusters. I then investigated subcluster properties, including expansion, median ages, and relationships to the molecular cloud. In this complex, I found that the ages and expansion timescales of the subclusters were similar to the free-fall collapse timescales for the associated molecular cloud clumps, suggesting rapid star-formation across this complex. In a subsequent study of the Trifid molecular cloud, I found stellar velocity dispersions correlated with cloud velocity dispersions in different regions of the complex, with evidence for an expanding cluster within an expanding H II bubble (Kuhn et al. 2022).

4 The SPICY Catalogue

The Spitzer/IRAC Candidate YSO (SPICY) catalogue, containing 117,446 sources, arose from a Cosmostatistics-Initiative Residence Programme to identify the young stellar objects among the more than 80 million mid-infrared sources in Spitzer's cryogenic surveys of the Galactic midplane (Kuhn et al. 2021a). Spitzer's IRAC instrument is particularly sensitive to infrared excess from discs around young stars. However, numerous intrinsically red non-YSOs in the Galactic midplane had previously limited use of Spitzer for searching for YSOs in this area. Our tailored statistical approach combined spectral energy distribution fitting and random forest classification with diagnostic statistics (e.g., assessment of variable importance and partial dependence plots) to better interpret the classifications. Furthermore, a variety of statistical tools were employed to better understand these sources, their clustering, their environments, and their variability. Given the large number of star-forming regions covered by SPICY, these sources can serve as the basis for a variety of studies of young stars, such as the JWST/NIRCAM study of H_2 outflows from young stars in NGC 3324 in which I participated (Reiter et al. 2022).

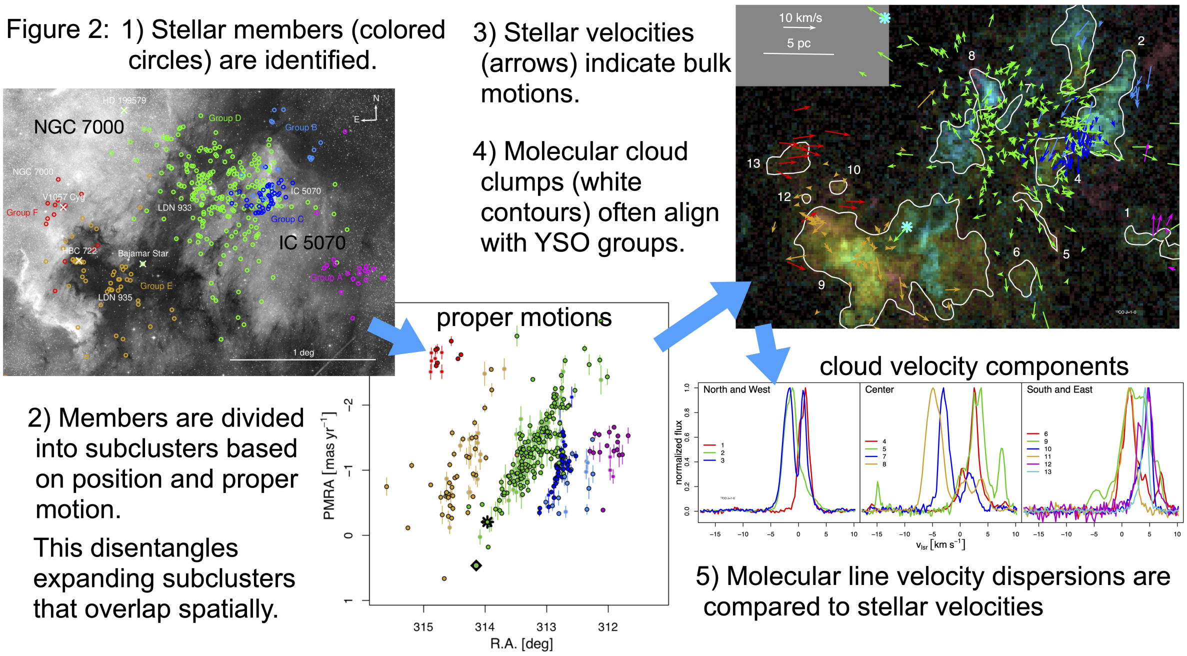

Figure 2: 1) Stellar members (colored circles) are identified.

2) Members are divided into subclusters based on position and proper motion.

This disentangles expanding subclusters that overlap spatially.

3) Stellar velocities (arrows) indicate bulk motions.

4) Molecular cloud clumps (white contours) often align with YSO groups.

5) Molecular line velocity dispersions are compared to stellar velocities

5 The Solar Neighbourhood and Galactic Structure

With the SPICY team, I identified a kpc-long chain of star-forming regions arranged as a distinct structure in position and proper-motion space. We corroborated this structure using extinction maps, maser measurements, and position-velocity CO molecular cloud maps. The structure has a high-pitch angle ($\psi \approx 56°$) relative to the Sagittarius Arm, which, combined with its discrete nature, suggests that it is a 'spur' rather than a Spiral Arm segment as previously thought (cf. Morgan et al. 1953; Taylor & Cordes 1993). This structure, the 'Sagittarius Spur,' along with similar structures known as the 'Radcliffe Wave' (Alves et al. 2020) and the 'Split' (Lallement et al. 2019) are recognised by the community (see Protostars and Planets VII Chapters 2 and 3) as explaining most nearby star-forming regions.

6 Eruptive Outbursts from Young Stars

Large young stellar catalogues, such as SPICY, assist in searches for infrequent events, such as eruptive accretion outbursts. These outbursts deposit significant mass onto stars and may partially explain how stellar mass is assembled (Fischer et al. 2022). However, their rarity means that only a few dozen of each class have been spectroscopically confirmed. Over 200 Gaia Alerts (Hodgkins et al. 2022) have been triggered by SPICY sources, and SPICY is included in the Gaia Alerts' classification process, meaning that this is a promising sample for future investigations.

I studied the V1741 Sgr outburst (= SPICY 71482; Kuhn et al. 2024) with spectroscopic and multi-wavelength monitoring, allowing me to assess the origin of the brightening during this event (talk recording). Differences in the optical and near-infrared spectral features between this source and others of the same class highlight the diverse nature of accretion outbursts (Figure 3). I have also reported a repeating EX Lup-type outburst from SPICY 97589 (Kuhn et al. 2023b). At Caltech, I designed a filter to identify candidate events from the Zwicky Transient Facility.

7 X-ray Emission from Young Stars

My work with young stars began in the X-ray part of the spectrum, where I analysed grating spectra of a B-type type star, demonstrating that the very soft X-ray emission from this source comes from a wind and discovering a pre-main-sequence binary companion (Cohen, Kuhn, et al. 2008). My PhD work on MYStIX (described above) produced, for the time, the largest catalogue of YSOs,

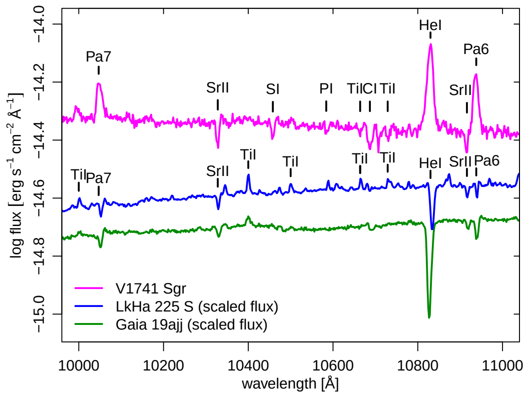

Figure 3. Y-band spectra for 3 EX Lup-type eruptive YSOs, highlighting differences in spectral features. These differences indicate that, even within this class, there is considerable variation in where emission originates within the outbursting YSO (Kuhn et al. 2024)

largely via X-ray selection. In my Chandra study of NGC 6231 (Kuhn et al. 2017a, b), I found that the X-ray 'colour–magnitude' diagram of this lightly obscured young cluster showed distinct loci for massive OB stars and low-mass pre-main-sequence stars. However, anomalously hard X-ray emission from some massive stars indicated the presence of pre-main-sequence companions, allowing multiplicity to be examined in systems where other techniques are not feasible. Finally, In an XMM-Newton study of outbursting FU Ori stars (Kuhn et al. 2019b), I found that these sources have significantly elevated magnetic activity relative to other pre-main-sequence stars.

8 Key References

Cohen, Kuhn, et al. 2008, MNRAS, 386, 1855; Kuhn et al. 2010, ApJ, 725, 2485; Kuhn et al. 2013a ApJS, 209, 27; Kuhn et al. 2013b ApJS, 209, 29; Kuhn et al. 2014 ApJ, 787, 107; Kuhn et al. 2015a ApJ, 802, 60; Kuhn et al. 2015b ApJ, 812, 131; Kuhn et al. 2017a AJ, 154, 87; Kuhn et al. 2017b AJ, 154, 214; Kuhn & Feigelson 2019, in *Handbook of Mixture Analysis*, Chapman & Hall/CRC (New York); Kuhn et al. 2019, ApJ, 870, 32; Kuhn et al. 2020a, ApJ, 899, 128; Kuhn et al. 2020b, RNAAS, 4, 224K Kuhn et al. 2021a, ApJS, 254, 33; Kuhn et al. 2021b, A&A, 651, L10; Kuhn et al. 2022, ApJ, 937, 46; Kuhn et al. 2023, AJ, 165, 3; Kuhn et al. 2024, MNRAS, in press; Kuhn et al. 2023b, RNAAS, 7, 57; Morgan et al. 1953, ApJ, 118, 318; Reiter et al. 2022, MNRAS, 517, 5382; Taylor & Cordes 1993, ApJ, 411, 674; Wright et al. 2023, ASPC, 534, 129 (PPVII Ch. 3); Zucker et al. 2023, ASPC, 534, 43 (PPVII Ch. 2)