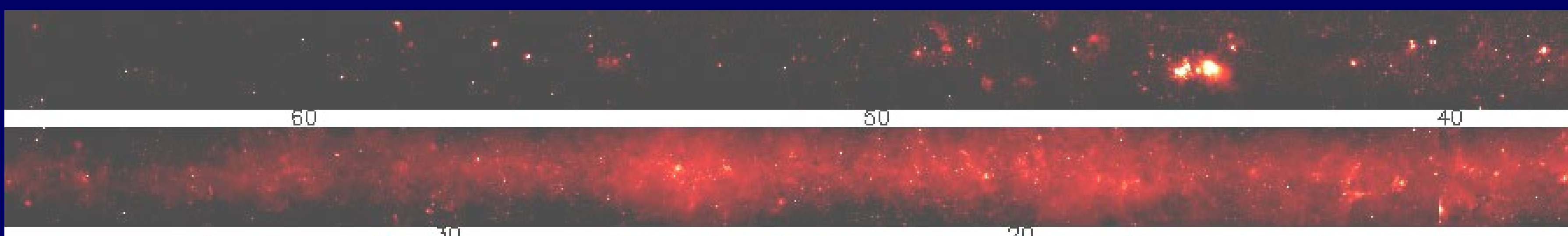
UNIVERSITY OF LEEDS

Analysing spectra from the Red MSX Source Survey

Stuart Lumsden & Jake Ward



RMS Team:
Melvin Hoare, Rene Oudmaijer, Heather Cooper, Nichol Cunningham, Luke Maud (Leeds), James Urquhart, Hugh Wheelwright (MPIfR), Joe Mottram (Leiden), Ben Davies, Toby Moore (Liverpool JMU), Cormac Purcell (Sydney), Michael Burton (UNSW)

The RMS Survey – Finding the Galaxy's Young Massive Stars

UNIVERSITY OF LEEDS

- Solving the problem of how massive stars form...
- Colour selection from MSX & NIR
- Find all the luminous (>10^4 $L_{\odot}$), embedded IR sources – both massive protostars and ultracompact HII regions
- Lumsden et al 2013 (ApJS, 208, 11)
- rms.leeds.ac.uk

GL2591

Gemini JHK

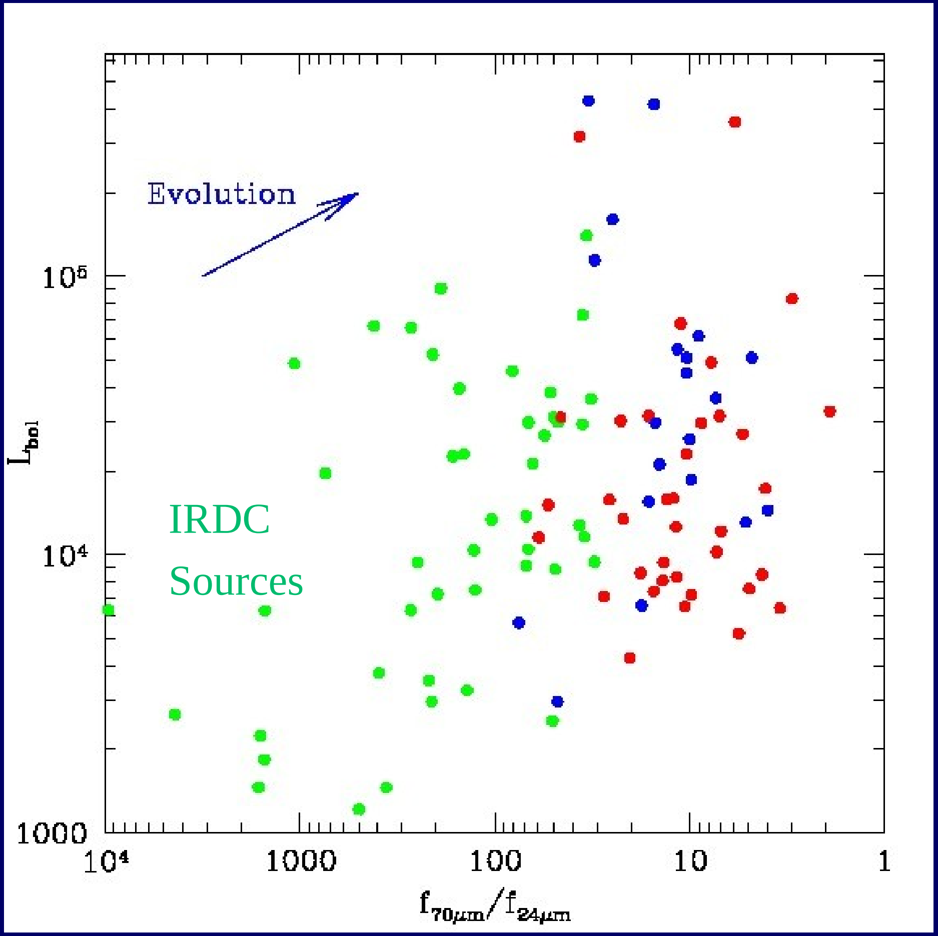
Multi-wavelength Ground-based Follow-up Campaign

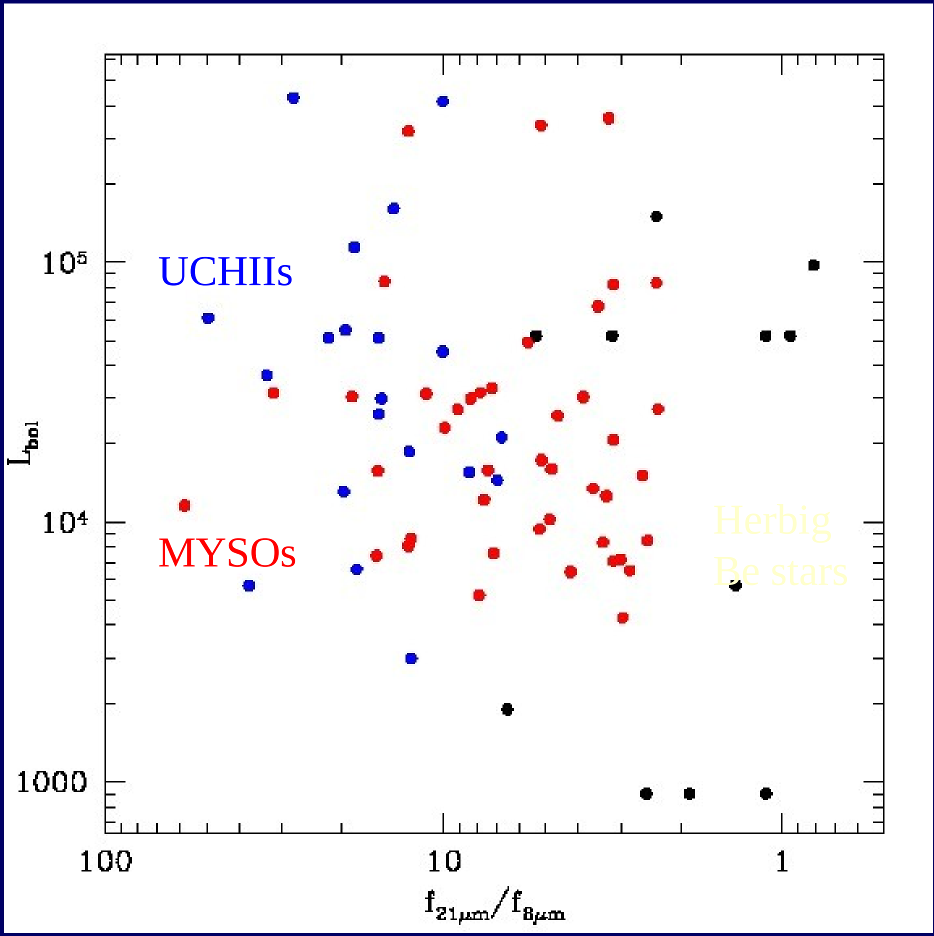
UNIVERSITY OF LEEDS

- Astrometry: radio, GLIMPSE, 2MASS, UKIDSS/VVV, WISE
- Radio continuum (Urquhart et al. 2007, 2009) => HII regions
- Mid-IR (Mottram et al 2007) => dust morphology (MYSOs vs HII regions)
- ^{13}CO & HI (Urquhart et al. 2007, 2008, 2010, 2011) => distances
- Spitzer MIPSGAL and IRAS IGA => SEDs, Luminosities (Mottram et al 2010a, b)
- NIR spectroscopy => characterisation (Cooper et al 2013).
- Additional data always being added as it becomes available (eg UKIDSS/VISTA/WISE/HiGAL)
- Lots of data – heterogeneous, photometry, line parameters etc

Evolutionary Outline

UNIVERSITY OF LEEDS





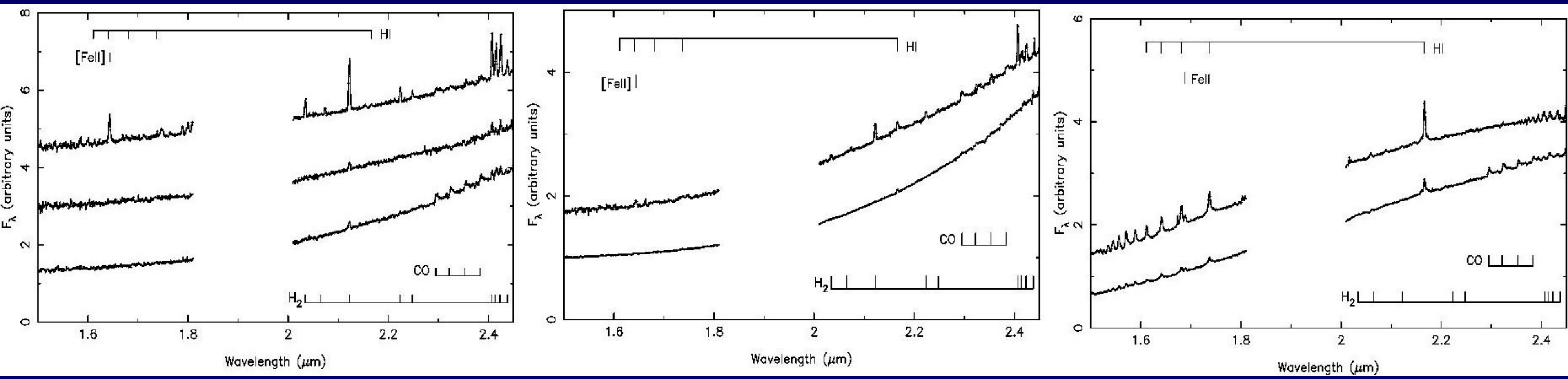
High near infrared extinction to all sources

RMS Spectroscopic Results

UNIVERSITY OF LEEDS

Largest homogeneous MYSO sample ever studied (Cooper et al 2013, MNRAS, 430, 1125). Protostars classified manually as:

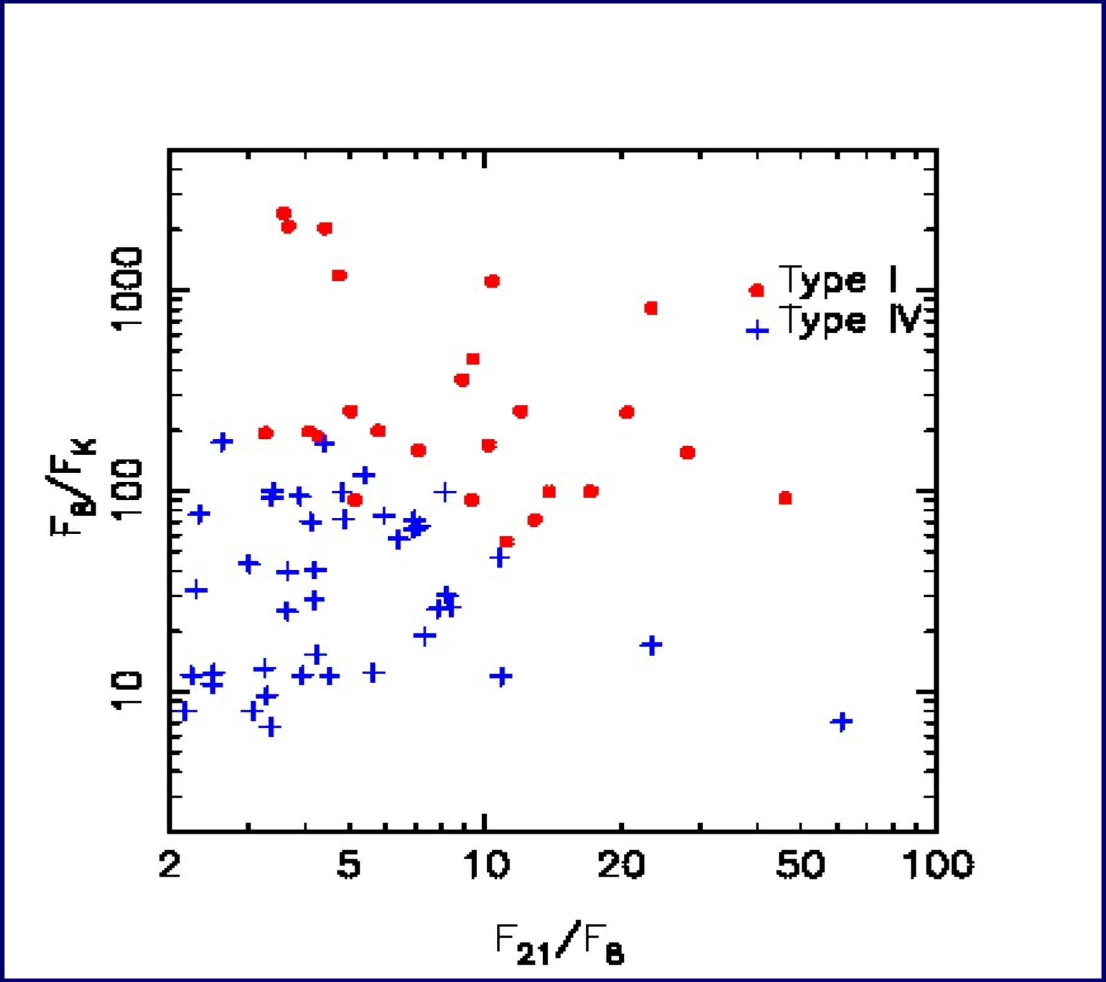
- Type 1 – H_2, no ionised gas (MHD jets?)
- Type 2 – H_2, weak ionised gas (Brγ)
- Type 3 – weak H_2, strong ionised gas (Br series), FeII
- Type 4 – no H_2, strong ionised gas, FeII (radiative winds?)



Spectroscopic Results

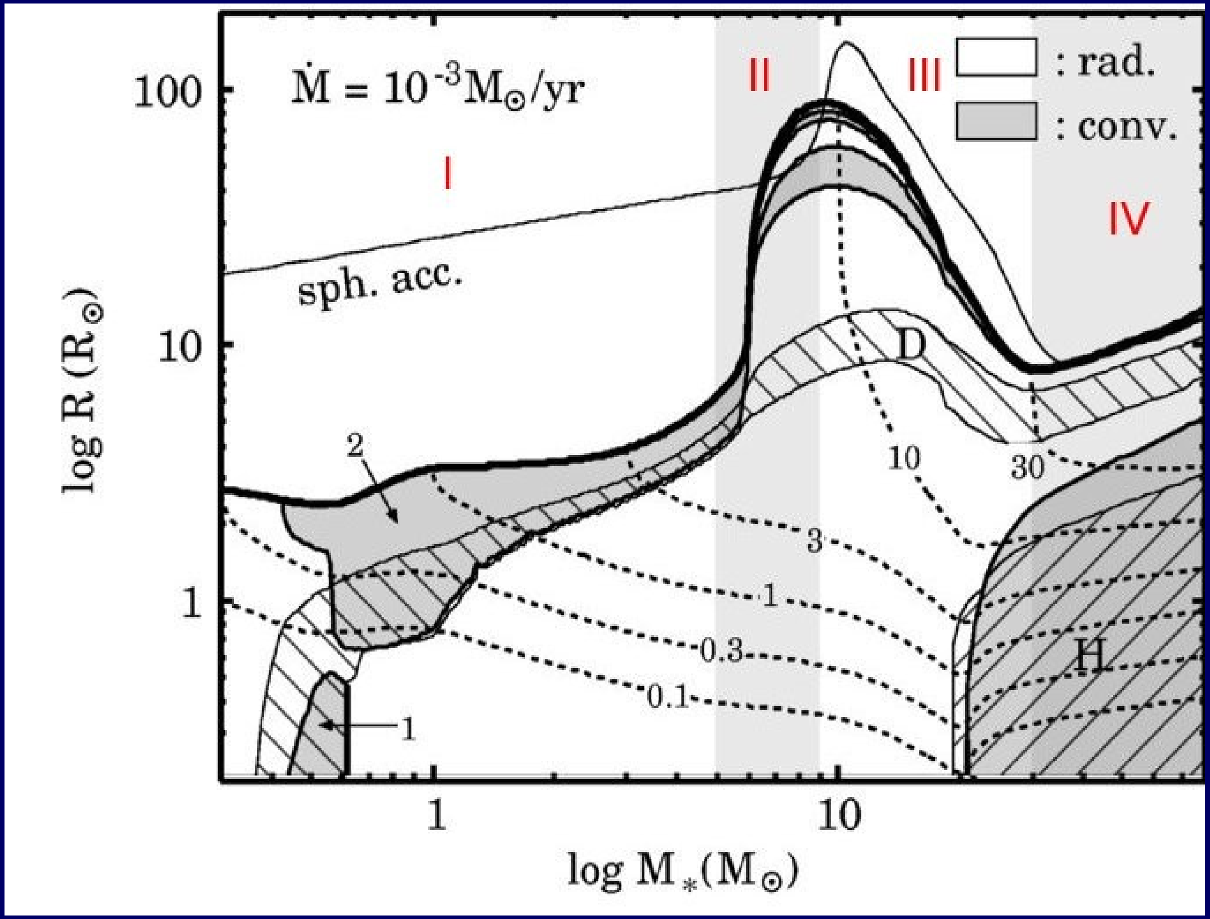
UNIVERSITY OF LEEDS

Colours largely consistent with age as the primary driver of type



Spectroscopic Results

UNIVERSITY OF LEEDS



Models (Hosokawa & Omukai 2009 ApJ 691 823) can explain picture

Automated Classification

UNIVERSITY OF LEEDS

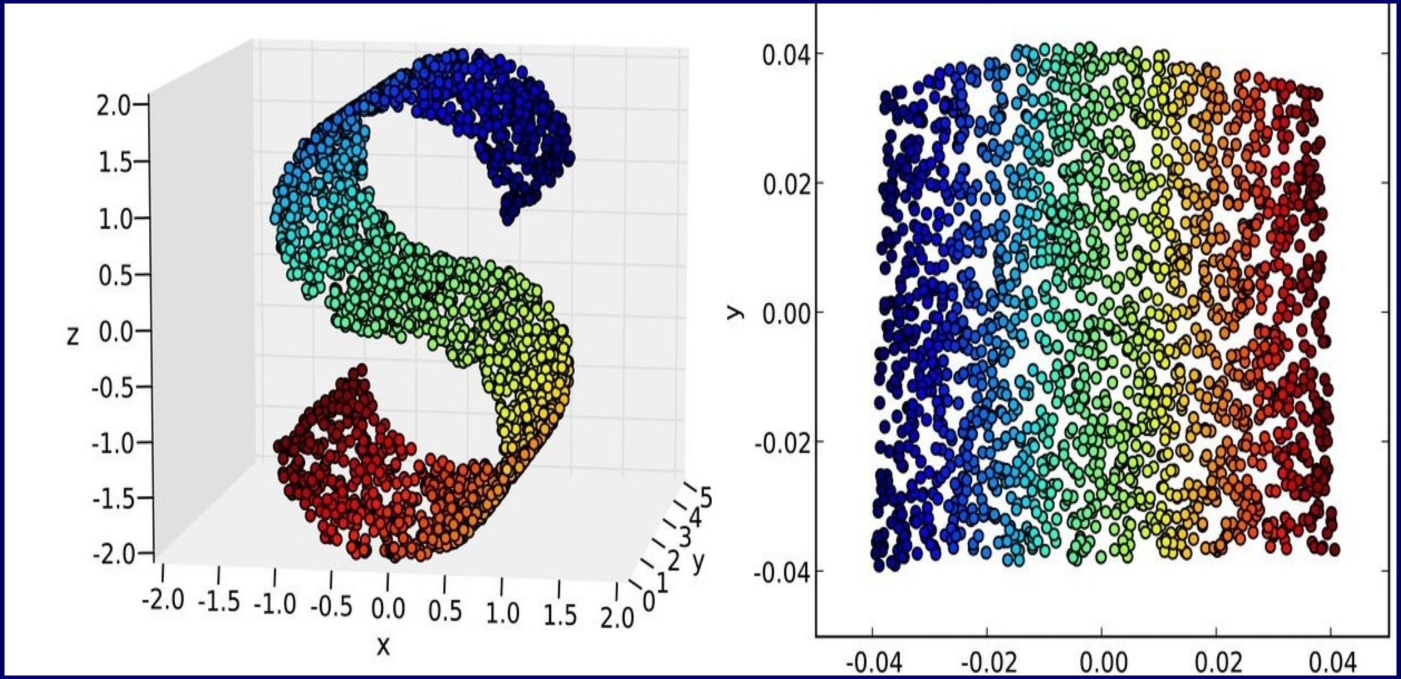
Can we do the same without human input?

- Extinction makes problem non-linear. PCA doesn't give meaningful results
- Locally linear embedding (LLE) used previously by Vanderplas & Connolly (2009, AJ, 138, 1365) on Sloan galaxy spectra, and on stellar spectra by Scott et al (2011, AJ, 142, 203).
- Available in several Python packages – we used MDP (Modular toolkit for Data Processing)
- Dimensionality reduction – LLE maps a lower dimensional manifold on which the data lie. But no easily attributable "eigenvectors" (unlike PCA).

Automated Classification

UNIVERSITY OF LEEDS

Locally linear embedding – mapping of non-linear manifold to lower dimensional space:

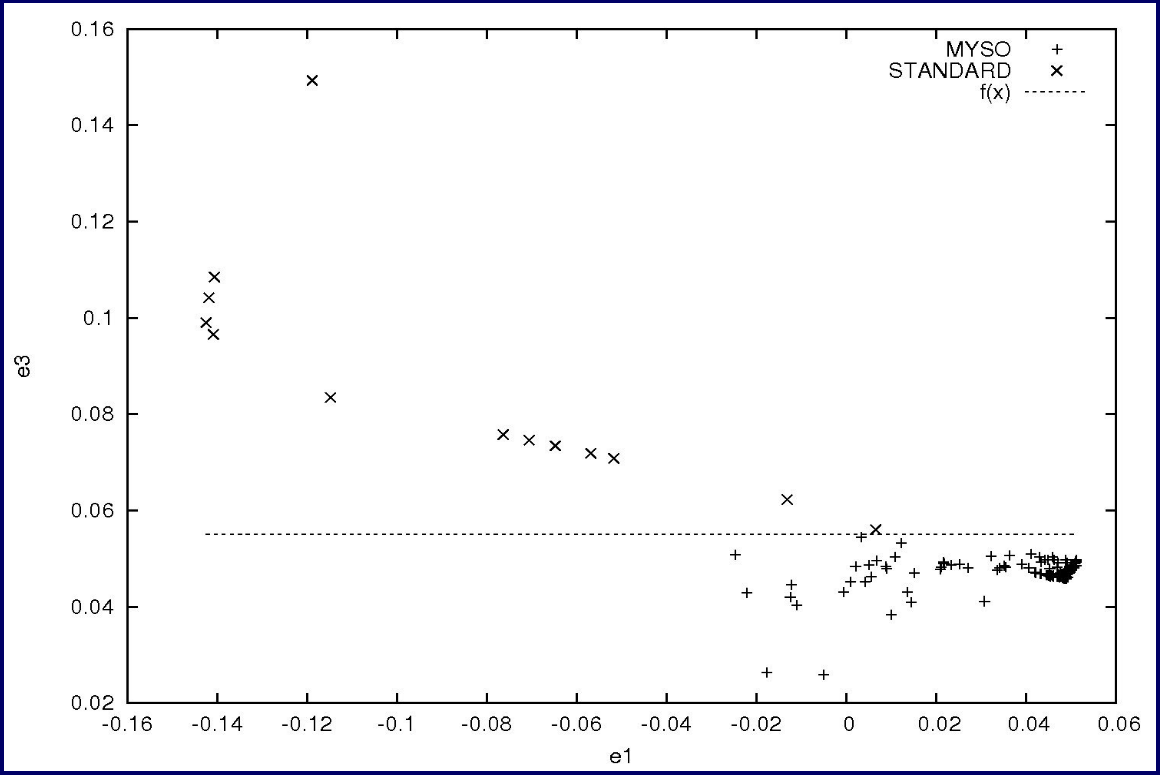


Vanderplas & Connolly

Automated Classification

UNIVERSITY OF LEEDS

Simple test – can we distinguish objects from main sequence A type standards...

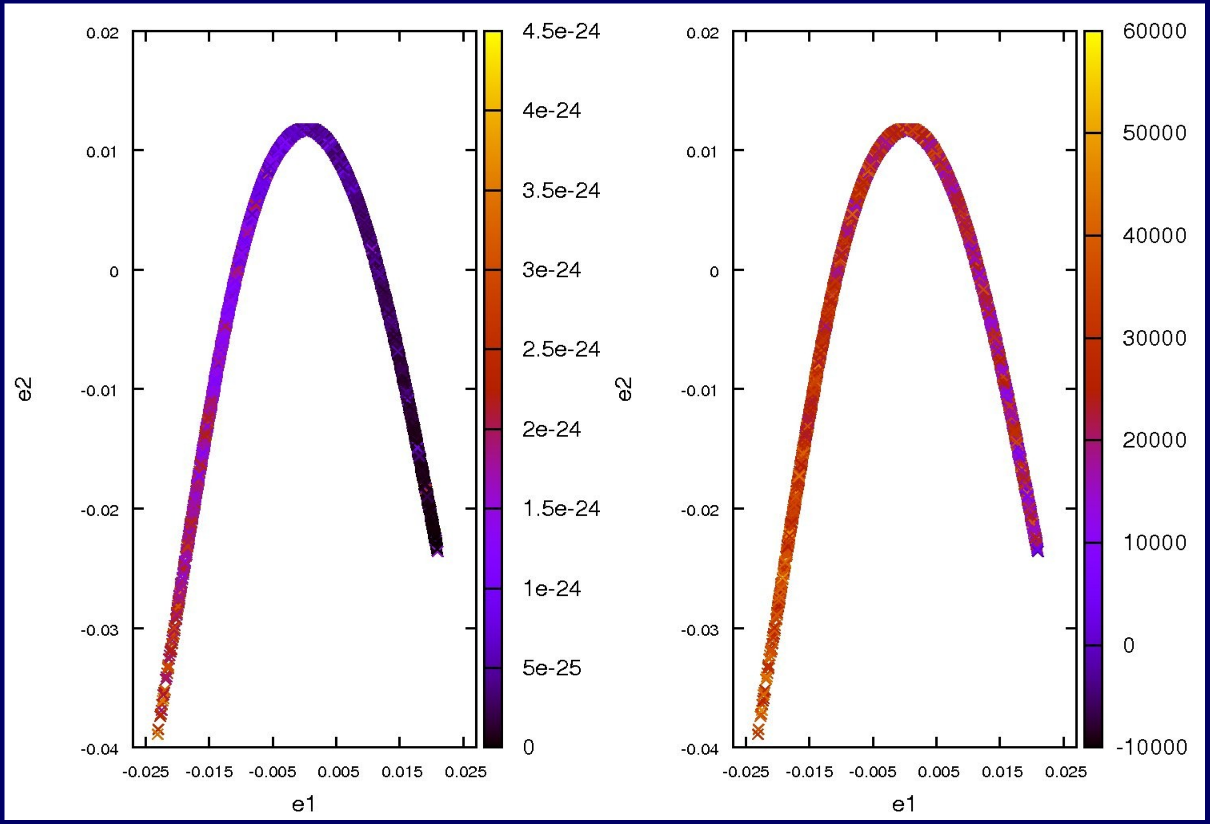


Consistent with trends related to colour (black-body temp).

Automated Classification

UNIVERSITY OF LEEDS

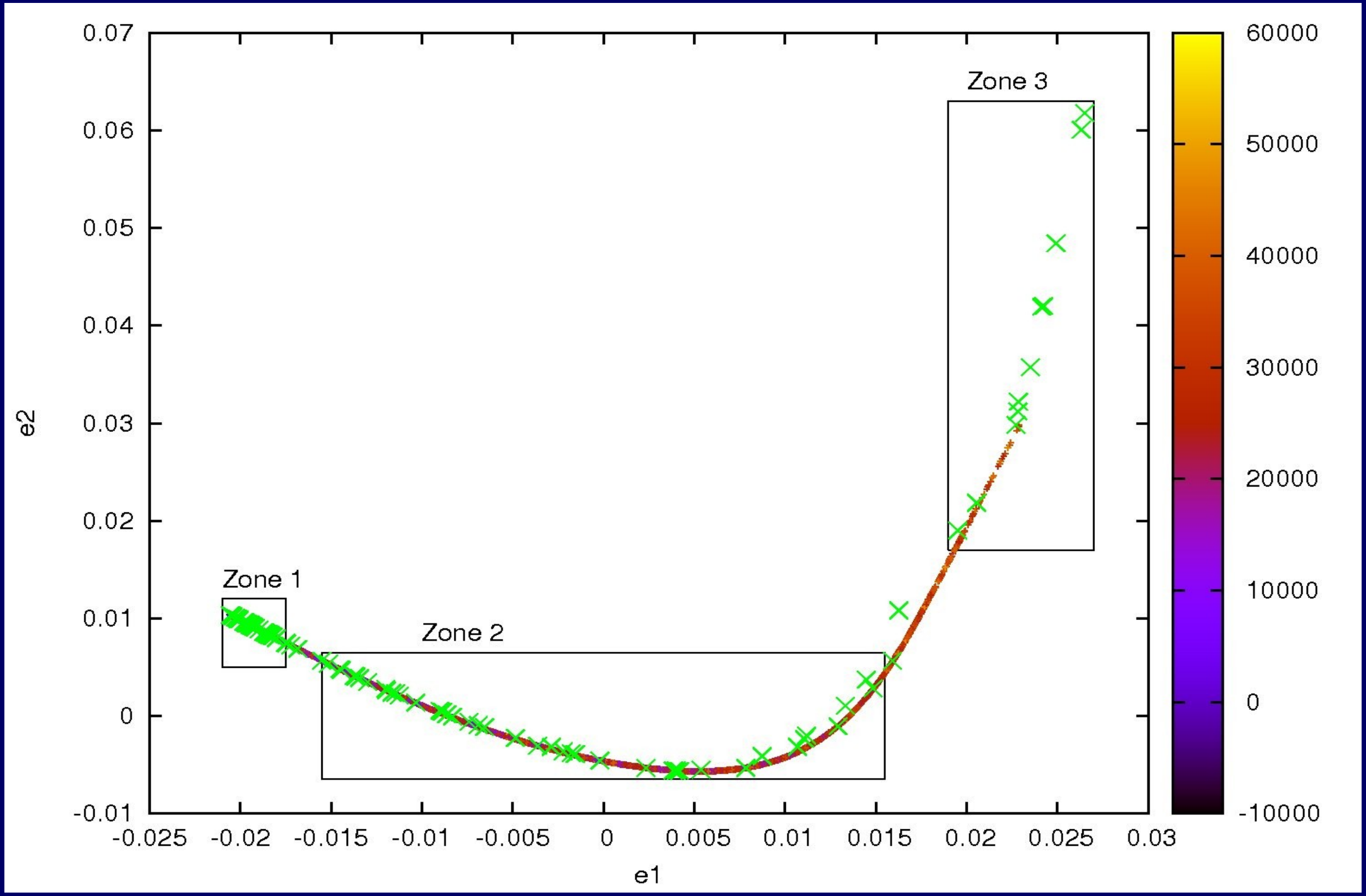
Indeed colour is the dominant term even after correction for extinction e1, e2 components as function of effective temperature from model continuum only spectra – need to remove continuum as well



Automated Classification

UNIVERSITY OF LEEDS

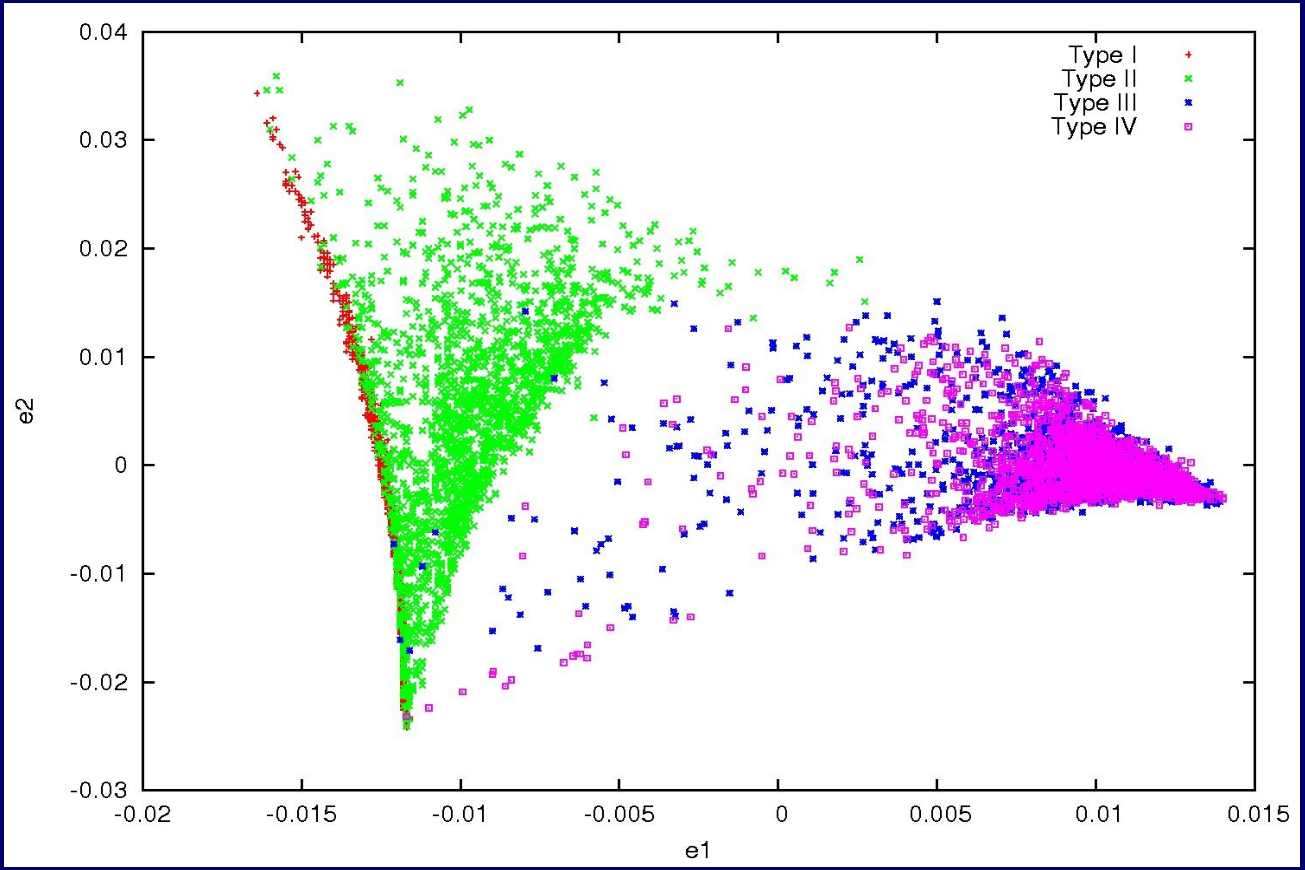
And with real data after extinction correction but without continuum subtraction.



Automated Classification

UNIVERSITY OF LEEDS

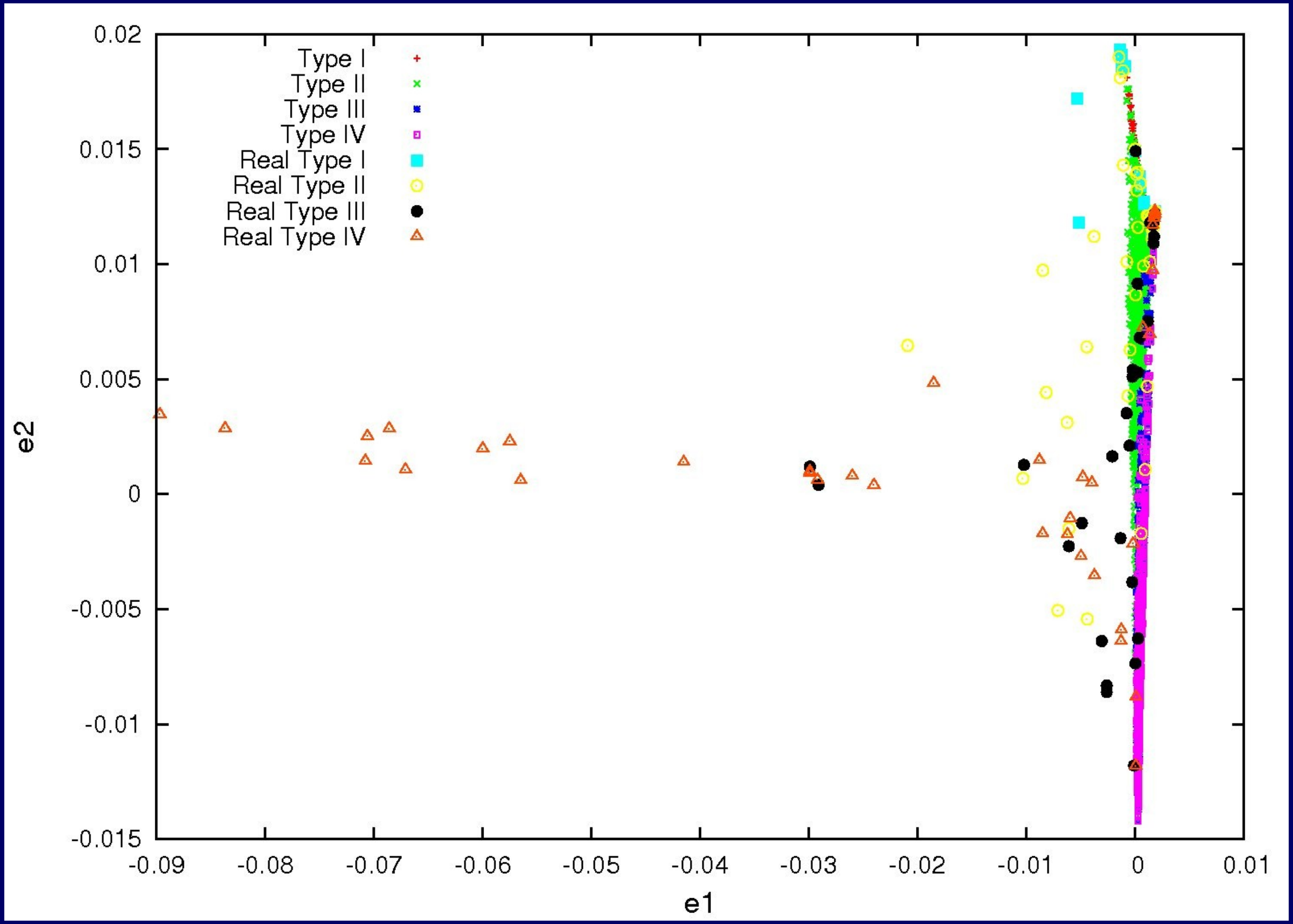
Use model spectral data with lines only to map out the behaviour. Scale lines according to the original manually assessed types. I/II vs III/IV split due to strong H-band lines in latter



Automated Classification

UNIVERSITY OF LEEDS

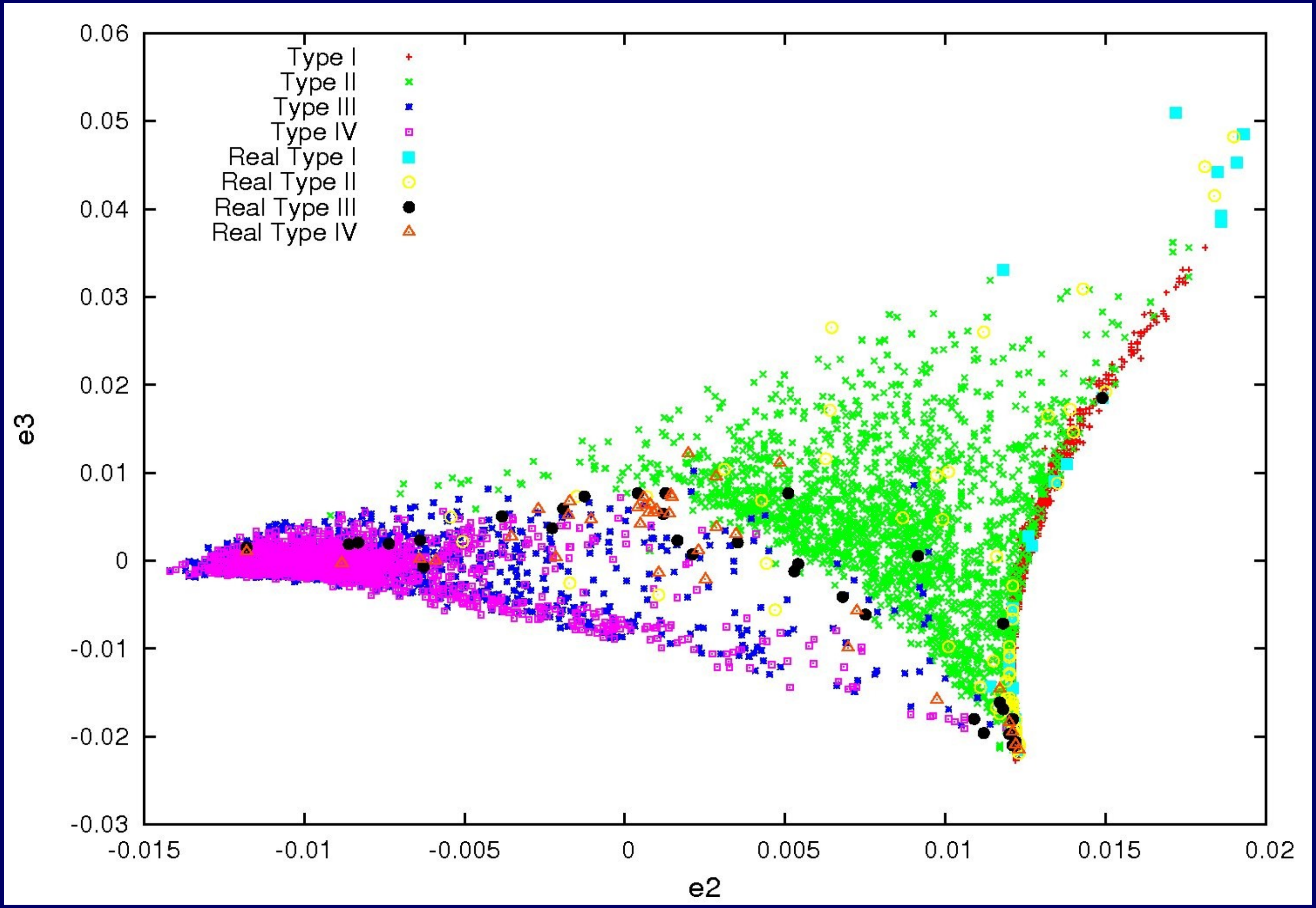
Adding in real data gives greater variance



Automated Classification

UNIVERSITY OF LEEDS

But reasonable agreement between models and real data in e2, e3



Automated Classification

UNIVERSITY OF LEEDS

Use these results to try and recover the types in the original real data.

Gets ~70% “correct”

Only ~3% totally wrong – the other 27% cannot be ascribed solely to a specific type from their location in the e1,e2,e3 plots.

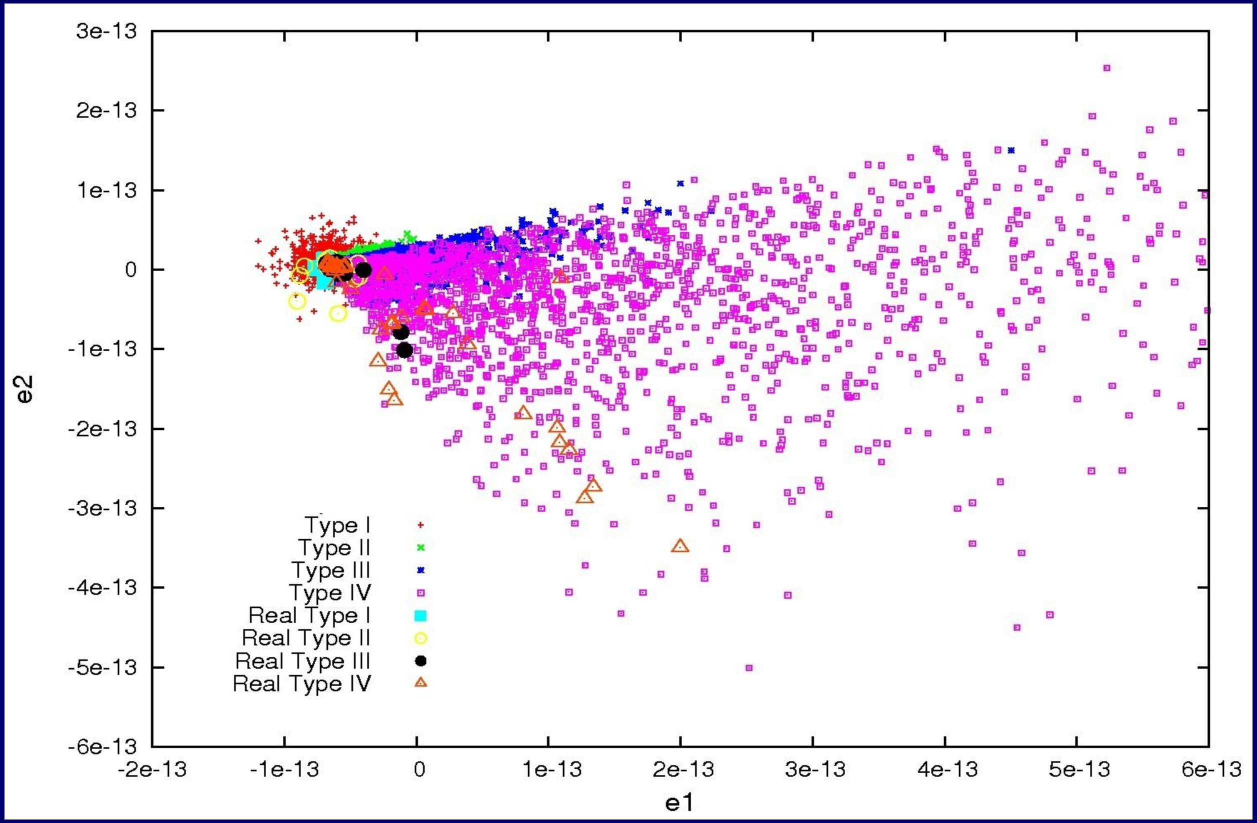
III/IV tend to be mixed in the LLE decomposition. Probably only need the 3 distinct types.

Need better model spectra to help determine trends, and to include full sample of real data (only about 50% here).

Automated Classification

UNIVERSITY OF LEEDS

How well can PCA do?



Much more extreme mixing of most of the types – much less variance.

Conclusions

UNIVERSITY OF LEEDS

LLE can recover much of the same information as we perceive in the data

Needs continuum subtracted data – otherwise the dominant component is the stellar effective temperature/reddening.

Same techniques could be used on mm/sub-mm data where we expect to see chemical evolution in the surrounding molecular core.

Where extinction is significant, LLE is preferable to PCA.