Detection/selection & telemetry: Selection Strategy

RVS-WG VI, MSSL

Mark Cropper

Main Impacts of Data Selection Strategy:

- One of GAIA's major scientific advantages is that there is no formal preselection of targets: there is no human prejudgement about what it should look at…
   - it is a big net, providing resources for the future (including the distant future), outside of the current paradigms
- Obviously there is a practical selection of the data set by
   - instrument sensitivities
   - instrument resolution
   - starmapper algorithms
   - downlink rates
- Much of GAIA science will deal with samples, and it will be critical to the success of the mission that the selection biases are known
- A model will need to be constructed *somehow* which encapsulates the selection function.
- Instrument sensitivities and instrument resolution are given largely by practical system constraints; but there some is some flexibility on starmapper algorithms and (perhaps) downlink rates

Main Impacts of Data Selection Strategy:

- Total GAIA return strongly depends on the combination of astrometric, spectroscopic and photometric data
  ⇒ need to have a coordinated strategy for the data selection in the different instruments
- Not sure if this has been sufficiently considered?
- Issues:
  - different sensitivities of MBP/RVS *c.f.* Astro/BBP
  - different spatial resolutions, ⇒ different sensitivity to crowding
  - different wavebands (*e.g.* RVS is quite red)
  - different selections on different scans (due to different crowding issues)?
  - buffer/telemetry constraints (see later)
- This has large implications for
  - the operation of the automatic classification algorithms, and
  - on the selection functions.

Implications of buffer policy

- It is proposed (PDHE contract) that data are marked with different (2?) levels of priority
- Current buffer policy is to allow data to overflow bulk storage facility if downlink rate is insufficient
- Issues arising from this *laissez-faire* policy:
  - how are priorities assigned to data (by instrument, sky location, magnitude)?
  - how does the overflow loss depend on scan history through the Galactic plane?
  - what are the implications for the selection functions?
  - what are the implications for the subsequent data analysis (eg astrometric solutions, binary star solutions) and the automatic classification algorithms?
- Suggest this data loss policy is much more carefully considered and a more controlled/deterministic scheme designed.

Starmapper Algorithms: Implications for the RVSM

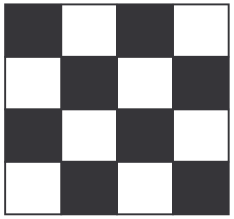
- What bandpass should the RVSM operate in (G band, RVS band)?
- If G band (white light), then may not be reaching maximum sensitivity for RVS for red stars, while faint hotter stars are included below sensitivity threshold suggest RVS band
- If stars are rejected by other instruments (*e.g.* MBP) because of blending, should these be rejected also for RVSM, even though this is less of a problem spectroscopically?
- Extended objects?
- Should there be some communication/coordination between the RVSM and other starmappers?

Downlink Rates: implications for RVS

- Data selection for RVS has included the concept of selecting only those regions around the Calcium lines for stars fainter than some threshold
- This implies a significant loss of information on the bulk of the sample, including a large sample of unusual objects, such as
  - accreting objects,
  - QSOs,
  - systems with winds etc.

  ⇒ need to consider whether this loss is justified, even though it is not “core” RVS science
- Width of those regions around the strong lines to be telemetered needs to be considered carefully to ensure sample is not biased by loss of sources with high radial velocity (some work done already in the WG on this)

Downlink rates: implications for RVS



- An additional scheme proposed to remain within the telemetry allocation is to do chequer-board sampling of the sky
- Clearly it is necessary that there is a coordinated sampling strategy between instruments for such a scheme, and some way of informing instruments that data are to be telemetered/discarded based on sky coordinates
- Issues:
  - it it better to observe certain areas of the sky at many epochs, while others are not at all, or better to average out the number of visits?
  - difference in science requirements: RVS main goal is Galactic kinematics, while Astro may have to provide sets of astrometric measurements, complete to some sample limit
  - sampling will need to cover those parts of the sky that are central to the science goals: *e.g.* it is essential to obtain sufficient sampling of the thick disk if the disk kinematics are a high priority.

Data Selection Issues: Summary

- Downlink rates remain a serious source of degradation to the science that GAIA could produce
- If selections have to be made they should be made in such a way that they
  1. consider the system as a whole (interaction between different instruments, starmappers *etc.*)
  2. remain orientated to GAIA main science goals while retaining as much of the additional science as possible
  3. consider the effect on the selection functions
  4. consider the effect on the classification/data mining/analysis process.
- More questions than answers at this stage: suggest that this needs significant attention

Derivation of Radial Velocities in Crowded Areas

RVS-WG VI, MSSL

Mark Cropper

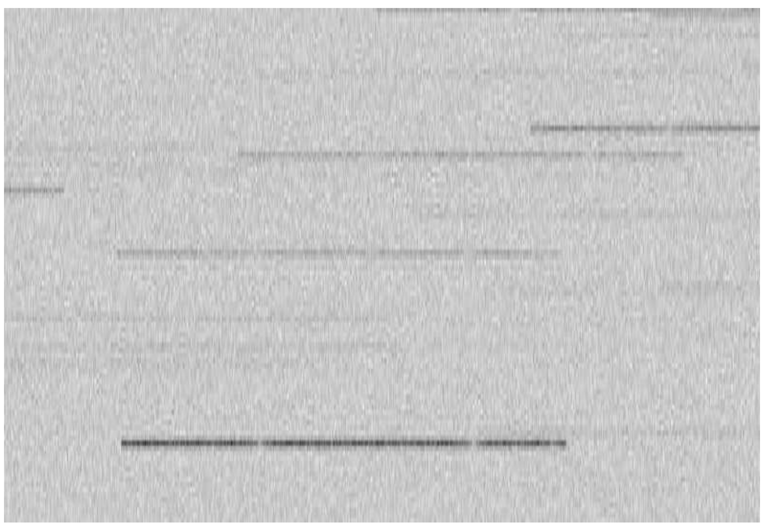
Derivation of Radial Velocities

- Extraction of radial velocities from RVS data is a process requiring a great deal of consideration, prototyping and optimising
- Process needs to be
  - almost absolutely robust in operation, even in very different field densities
  - able to deal with unusual and binary objects
  - clear as to what biases the automated procedure produces
  - as light as possible in terms of resource usage
- Since the implications for instrument performance are significant, understanding this process is essential even at this early stage need to start work in this area directly (some initial work already done by WG)

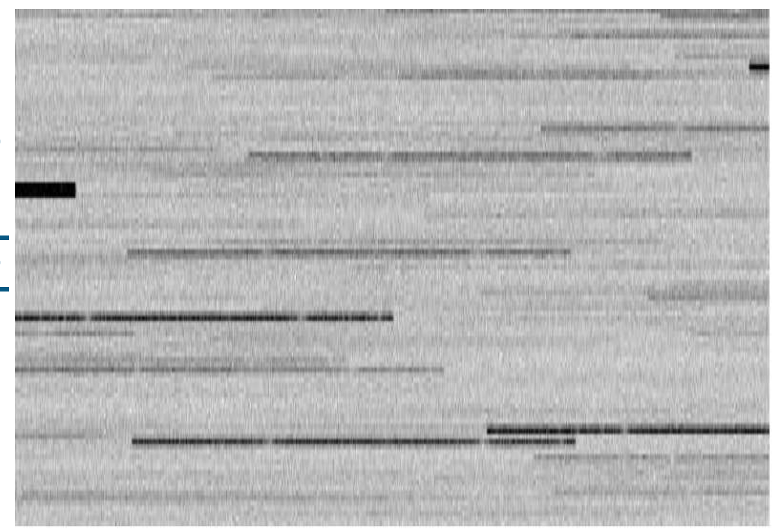
Derivation of Radial Velocities

- Current situation regarding simulations is to
  - generate sample images of sky using MSSL simulator
  - extract the sources
  - cross-correlate with a template
- Results shown at Monte-Rosa conference…
- Image generation:

|b|>30°



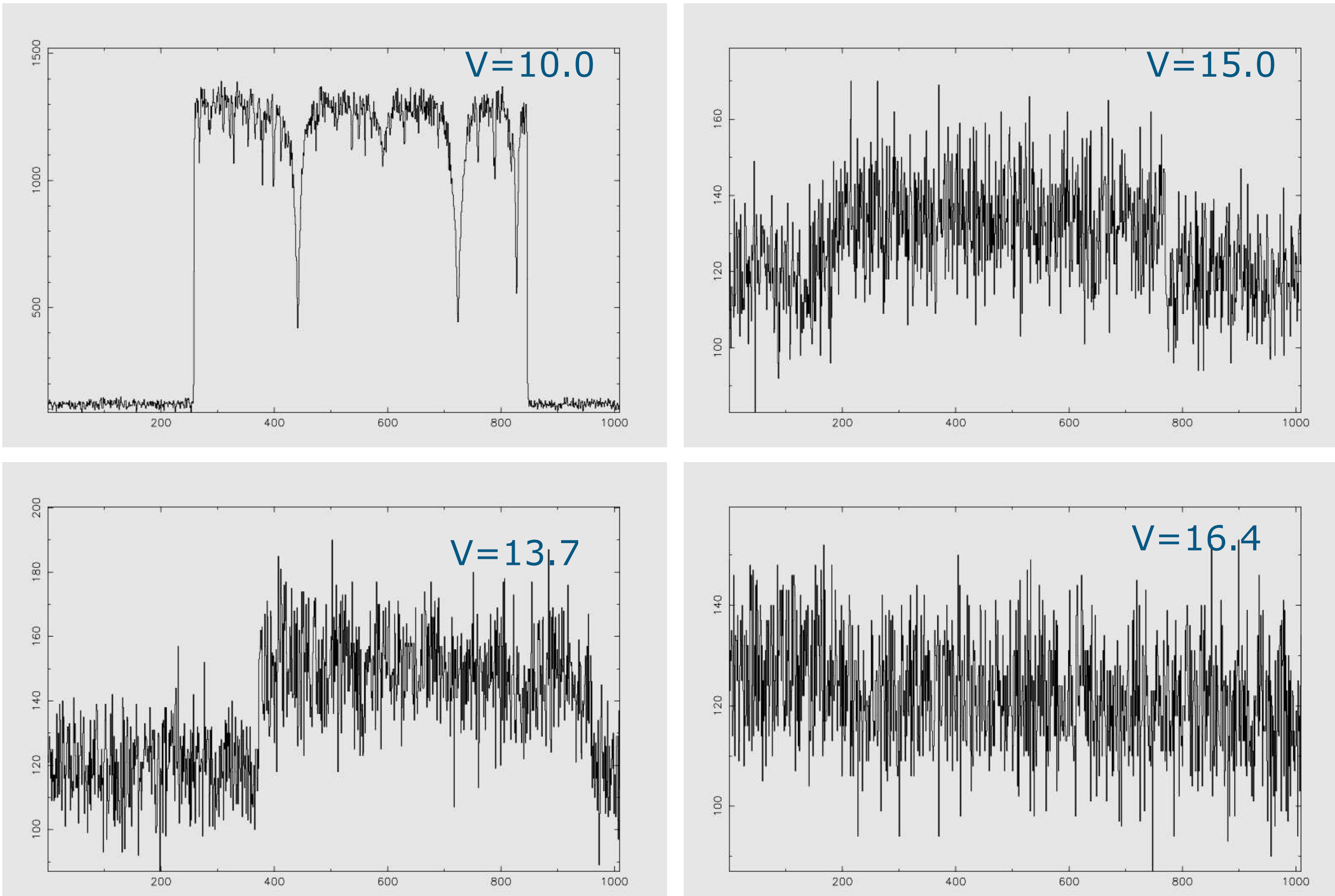
|b|<10°



Simulated RVS images, 1 transit (6 CCDs, 100 sec), detail view scaled between 0 - 300 cts, linear intensity scale, no scan drift, Gaussian PSF.

Derivation of Radial Velocities

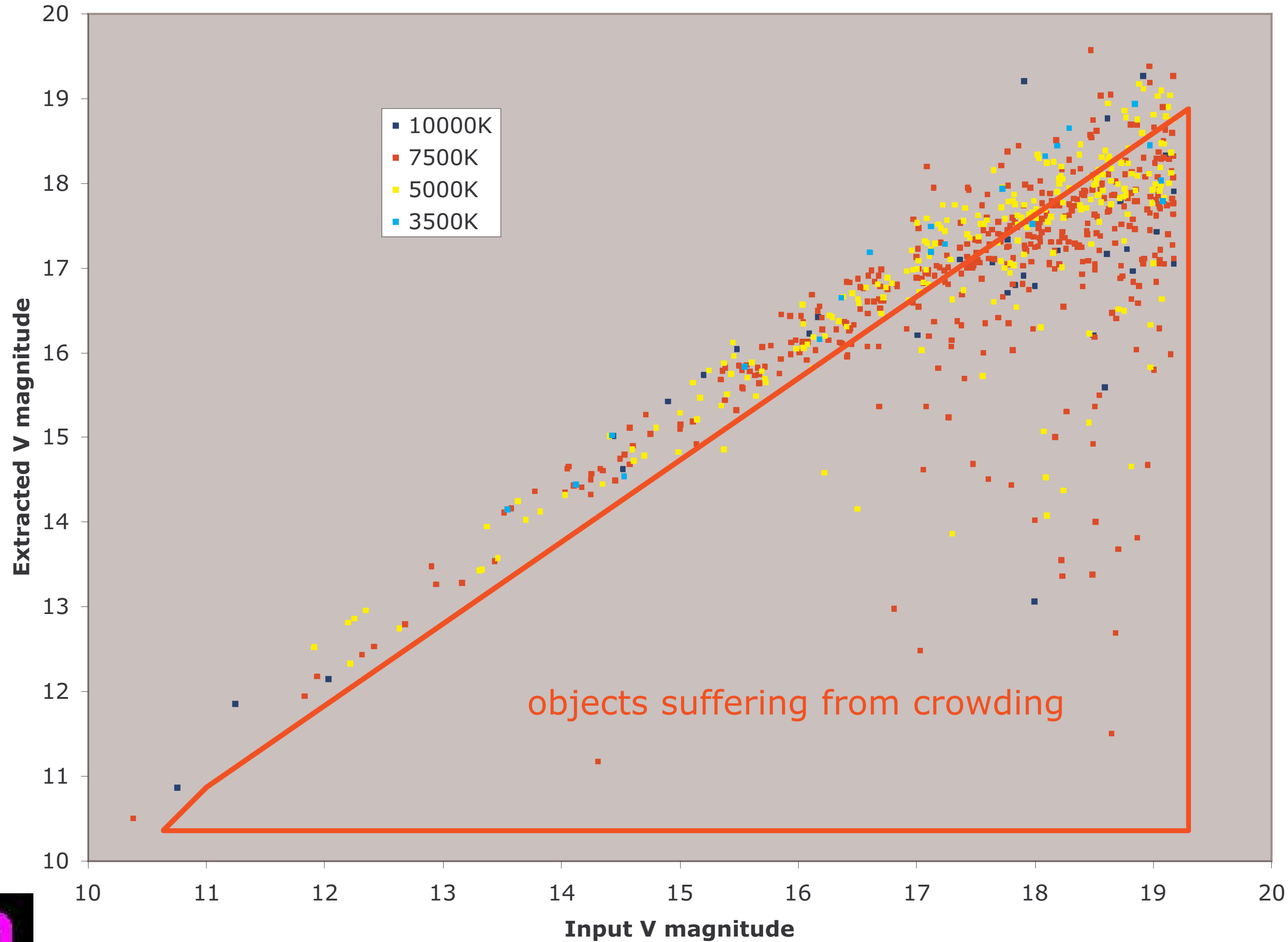
- extraction:



F star, 1 row extracted, Gaussian PSF, 6 CCDs stacked (100sec exp).

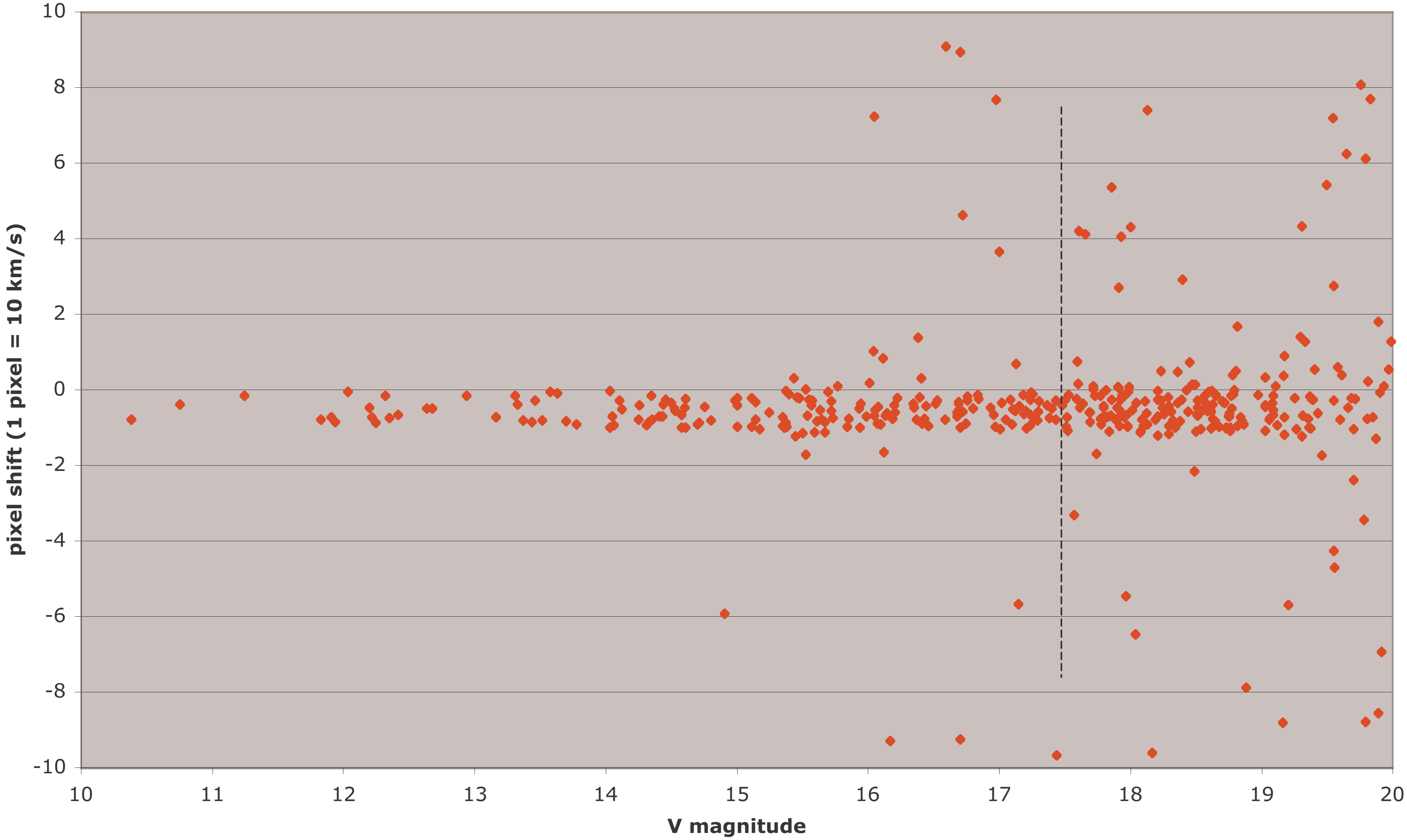
Derivation of Radial Velocities

- extracted vs input magnitudes (high Galactic latitude field):



Derivation of Radial Velocities

- extracted radial velocities vs input magnitudes:



F stars with correct template, high galactic latitude field, 100 sec exp, peak of cross-correlation, 1 pix ~10 km/s

To note from previous simulations

- even at high Galactic latitudes there is a significant amount of crowding (knew this already)
- many sources have derived velocities that are consistent with the template, given the expected errors
- but there are also outliers, resulting from overlapping spectra, even at fairly bright magnitudes, though the number increases as the magnitude decreases
- For some sources, the correct radial velocity is derived even below the magnitude limit of *V*~17.5 – problem is, we wouldn't know which ones are correct from real data
- It is essential we have a robust algorithm for deriving the radial velocities: how can we improve?
- main deficiency in previous simulations was that they did not remove the brighter spectra from the data, so that the fainter spectra were strongly affected

 need to develop a more considered algorithm for the 2-D extraction

A draft algorithm for extraction

1. start here
2. identify the (next) brightest spectrum
3. knowing its position, remove spectrum from image with the appropriate row weighting
4. knowing its spectral type from photometry, identify a suitable template spectrum
5. cross-correlate with the template to obtain the radial velocity
6. resample the template spectrum on rows as appropriate, blur in the spectral direction as appropriate to the RVS resolution and apply derived velocity shift to generate a template image
7. subtract template image from the real image
8. if limiting magnitude reached, stop, else go back to 2

A draft algorithm for extraction

- Eventually at the end we get sky. Now
  1. fit some sort of 2-dimensional model to the sky – could be a plane with a brightness variation in only 1 dimension, while flat in the other
  2. identify regions where there are significant residuals from the fit.
- Call this the first pass.
- Now consider the radial velocity of each spectrum as a dimension in some space. Then crawl around in the space to minimise the χ^2, concentrating on dimensions for which there are significant residuals.
- Could do this via maximum entropy or via some genetic algorithm? Problem of system resources?

A draft algorithm for extraction

- One of the optimisations at each stage would be a recalculation of the background.
- Could envisage a global multi-dimensional fit from the beginning, but probably best to do this after the initial sequential pick through the extraction as above as the fit ought to converge fairly quickly with this starting point.
- Some points:
  1. The routine may find a double peak in the cross-correlation, identifying a double-line binary.
  2. Some (many) stars may not have a suitable template, especially binaries
  3. There may be difficulties with having enough sky if this is not sent down

More on resampling the spectrum into an image

- the background subtraction has to be done carefully – from the ends of the spectra?
- the PSF (point spread function) to broaden the spatial profile is best kept in a 2-d format
- the PSF is not just the theoretical one, but has to include the optics distortion broadening, as well as the scan law.
  - Potentially needs to be generated for each row in the image, and for each CCD if these are kept separate.
- there needs to be a RV shift before re-insertion - derived from the data itself.
- there needs to be a broadening of the template spectra in the spectral direction by the resolution before reinsertion.

Major source of residuals to the fit

- It is not possible to use the data itself rather than the template in the overall cross-correlation.
  Problem: the major source of residuals is likely to be due to an inappropriate template.
- Actually, the radial velocity is only one parameter to be fitted, with the spectral classification another (multi-parameter) dimension.
- So are we able to iterate in this space to reduce the photometric errors at the same time as the radial velocity ones?
- This may be possible to achieve using a PCA (principal components analysis approach) (*eg.* Bailer-Jones). Unusual spectra need more dimensions, so it is easy to flag these up.

Determining astrophysical parameters

- How do we extract the spectrum in a blended case anyway? We have the radial velocities and we have the position but how do we assign particular photons to the one spectrum rather than the other?
- Here we may not have the profiles of each spectrum if it is unusual/binary
   - deblending may be impossible, especially if the spectra are very different in brightness.
- Some progress may be possible if we deal with those spectra that clearly do have a representative template (are these spectra with a minimum number of PCA components)?
- Also, where there is not complete overlap, there is pristine spectrum for part, which could perhaps be used for a photometric or PCA template.
- Alternatively, perhaps we're interested only in the spectral parameters of bright spectra anyway in which case the fainter contaminating spectra can be represented by their templates and subtracted, without too much difficulty?

Next simulations

- Use MSSL RVS simulator (Brindle/Yershov). This includes:
  1. star surface densities and spectral types from ESA-SP(2000)4
  2. PSF based on square pupil without obscuration
  3. scan law (crudely)
  4. parameters for readout noise, background levels, spectrograph throughput *etc.*
- Does *not* include:
  1. CCD cosmetics
  2. CCD fixed pattern sensitivity
  3. Optics distortion
  4. effects of combining the individual CCDs
- Recent improvements include filter response, so out-of-band traces of bright stars can be seen on images (currently being implemented).

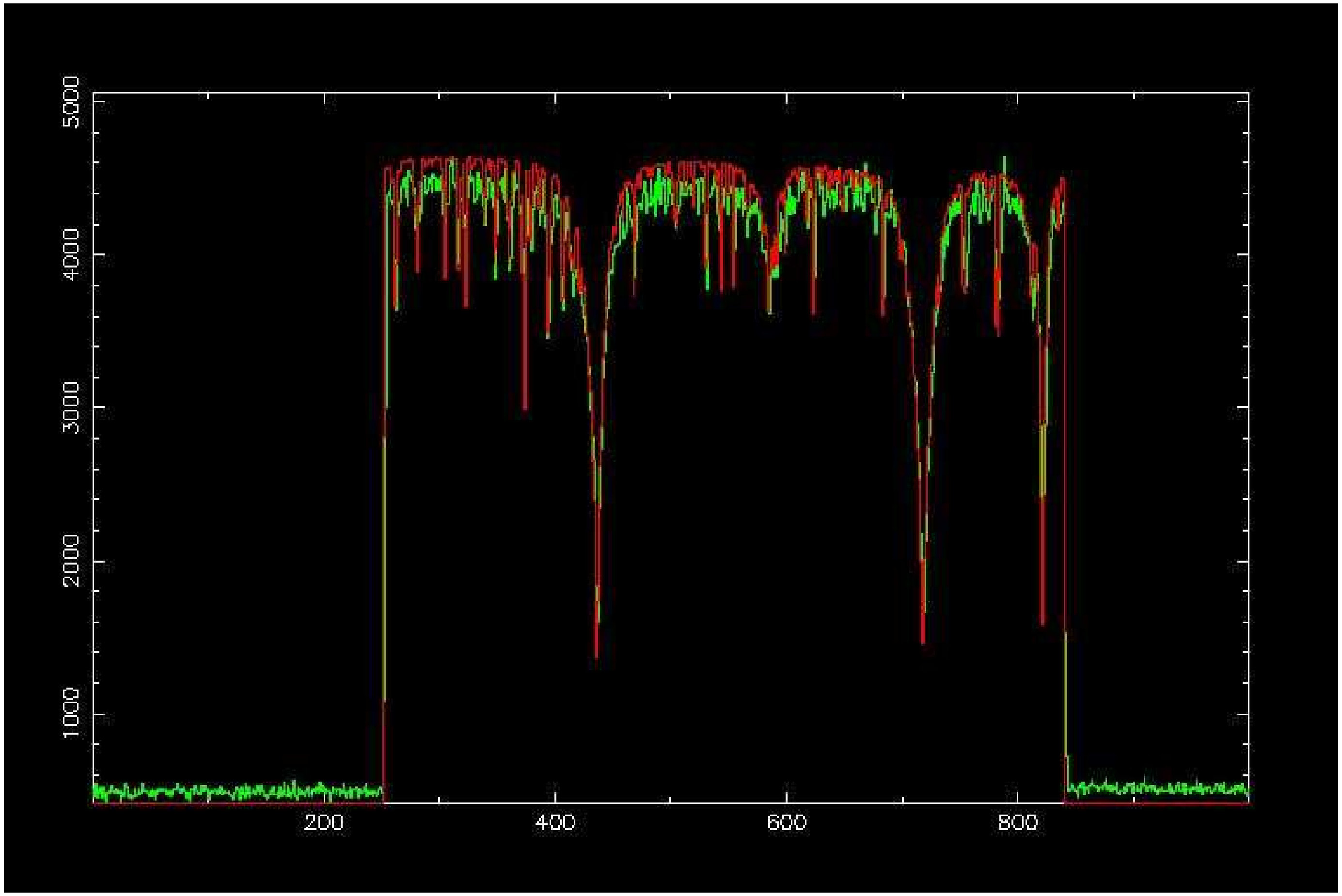
Simulator to be made available on http://www.mssl.ucl.ac.uk/gaia-rvs/ when possible (expect within a month or two)

New extractions

- New extraction procedures have been developed to implement the first part of the extraction algorithm considered earlier.
- Does not include the χ^2 optimisation.
- Also:
  - does not include any radial velocity shifts
  - uses only standard spectral types from Zwitter et al library; assumed known.
  - resamples spectral templates only in spatial direction, no spectral resolution broadening
- Main aim is to see if the number of outliers could be reduced compared to the earlier RV determinations
- Initial results…

New extractions

- Template star generation: green data, red template



New extractions

- Residual image of a V=11 star after a single subtraction

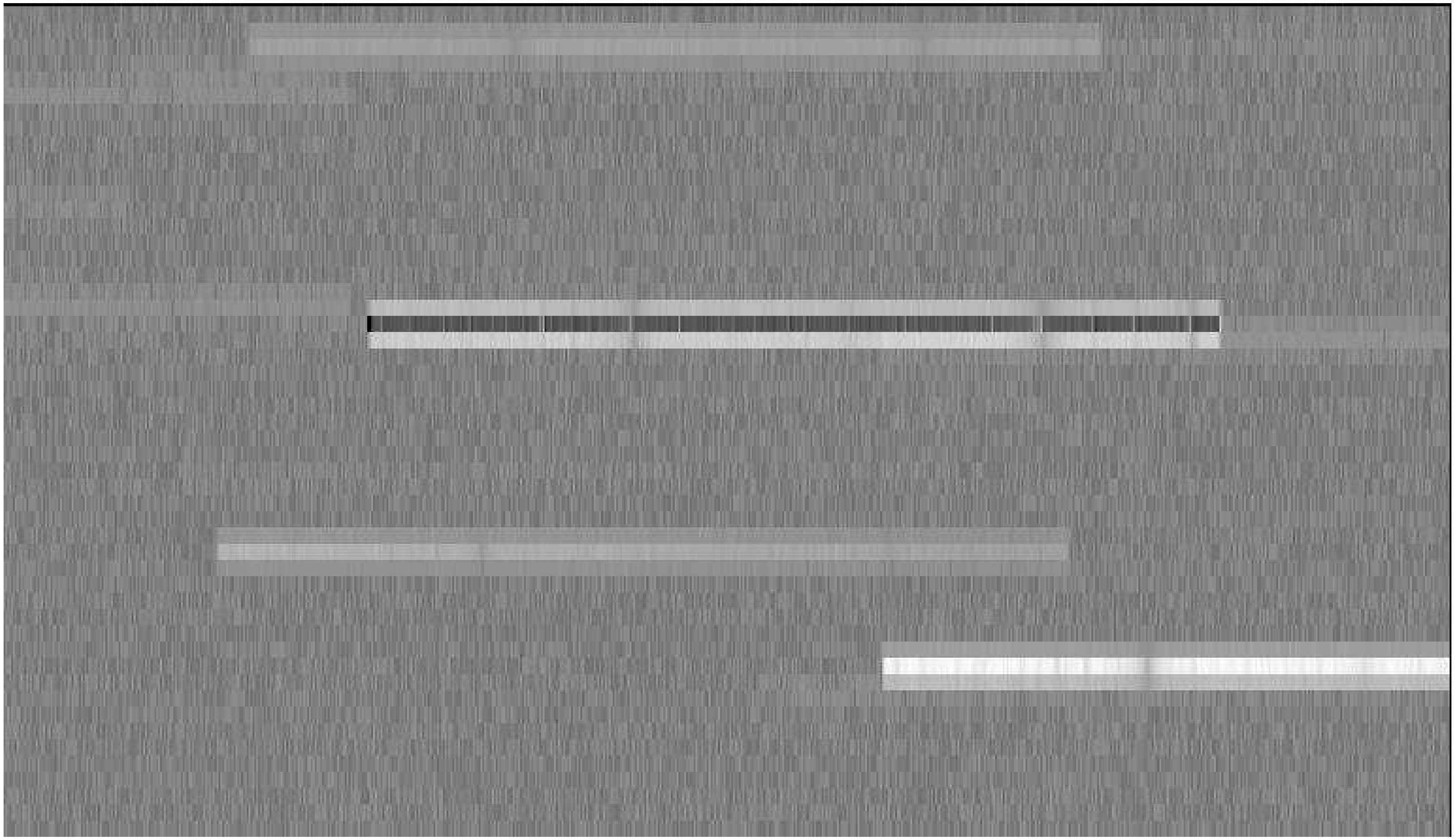
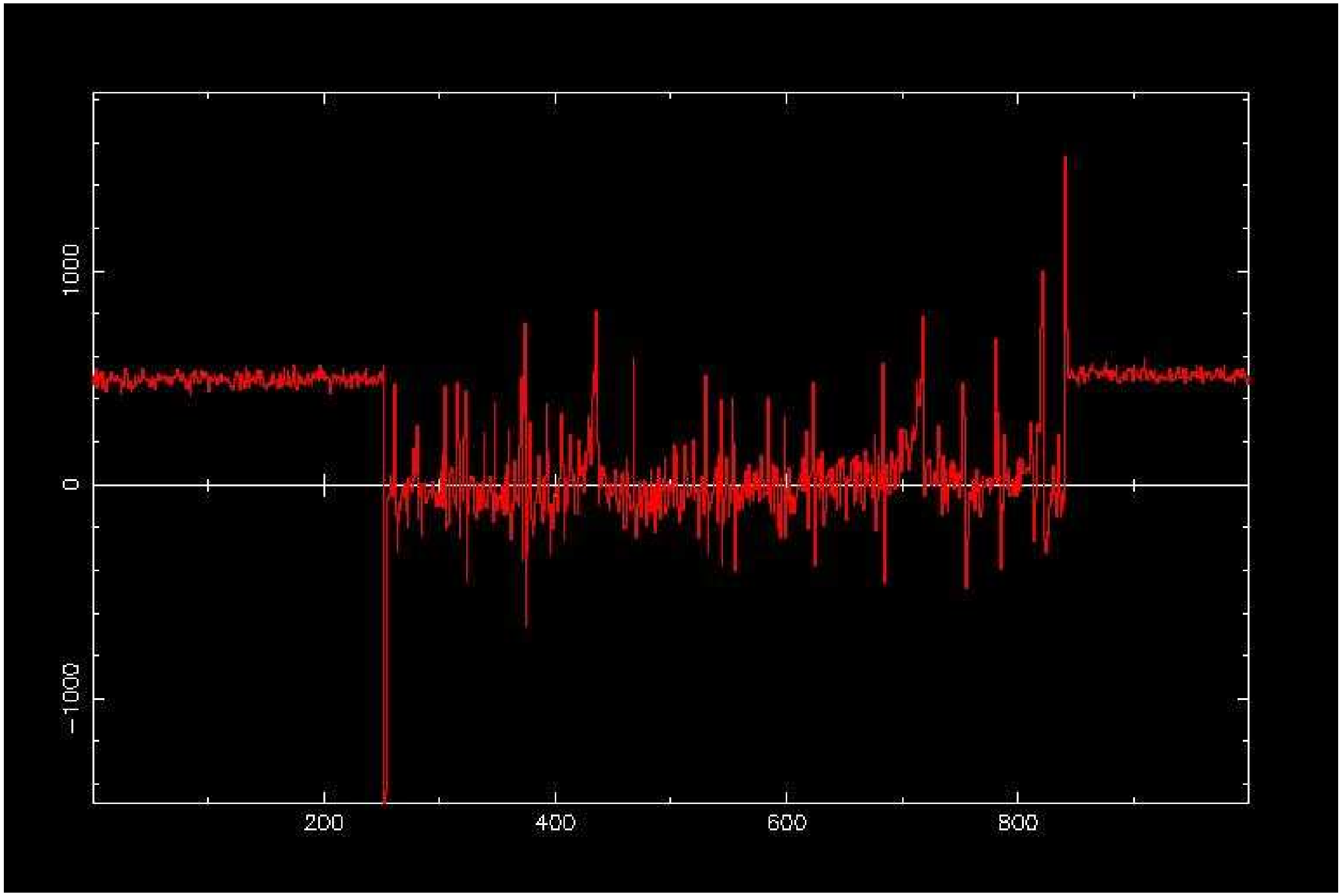


image made up of 4 focal plane scans to increase S/N ratio
|b|>30, image detail

New extractions

- Subtraction (background neglected)



New extractions: summary

- Work in progress
- Illuminates a number of issues that arise in practical implementation
  - issues of dealing with non-integer pixel insertions for subtraction (resampling onto grids *etc.*)
  - even good templates leave some residuals in the subtractions, but many stars will not have good representations as templates
    - it is indeed likely that photometric residuals will remain in the subtracted images (effect of this on the radial velocity determinations?)
- The cross-correlations are still to be done to extract the radial velocities to test the improvement over the older simulations
- Only the beginning: scheme to explore the overall 2-D optimisation still to be designed, and issues such as sky subtraction explored.