Detection and parameter estimation of type II solar radio bursts

Joseph Jenkins[1], Adeline Paiement[2,1], Xianghua Xie[1], Jean Aboudarham[3], Xavier Bonnin[3], Karine Issautier[3]

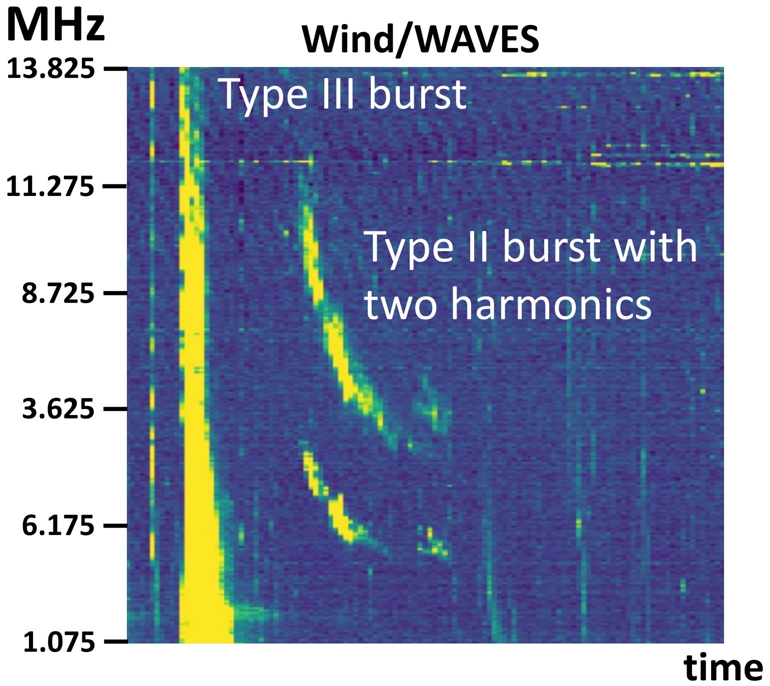
[1] Swansea University, [2] Laboratoire d’Informatique et Systèmes, Université de Toulon,

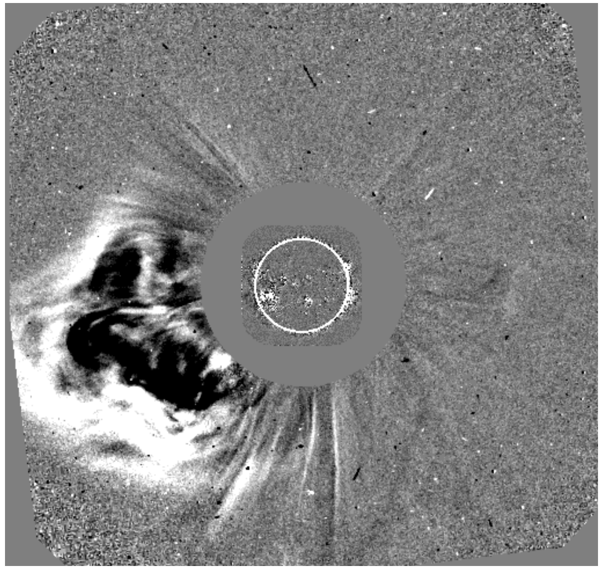
[3] Laboratoire d’Etudes Spatiales et d’Instrumentation en Astrophysique, Observatoire de Paris

Solar radio bursts

- Bursts = signals on radio spectrograms
- Multiple types of bursts
  - Indicative of specific solar behaviours
- Type II bursts:
  - Indicative of shocks
  - CMEs are common sources of shocks
- Type II parameters ↔ CME parameters
  - Start time
  - Duration per frequency
  - Drift rate
  - Intensity
  - Presence of harmonic

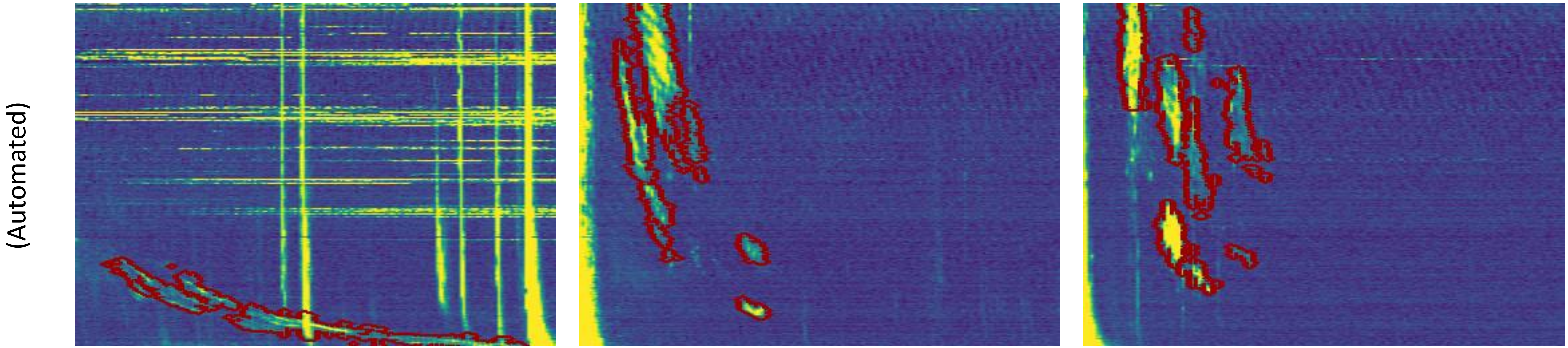
→ How to isolate and characterise the signals of type II bursts?



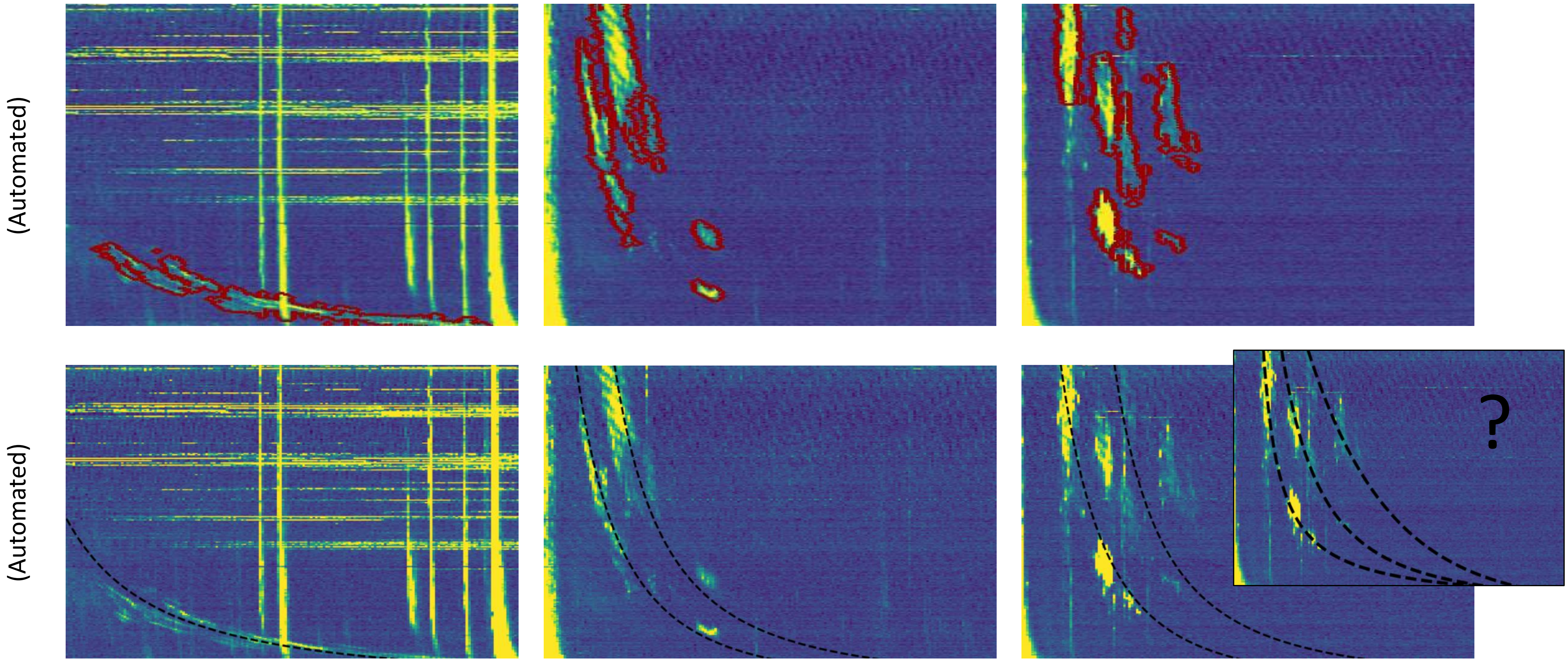


SOHO/LASCO

Aim 1: automated detection and segmentation



Aim 2: automated characterisation



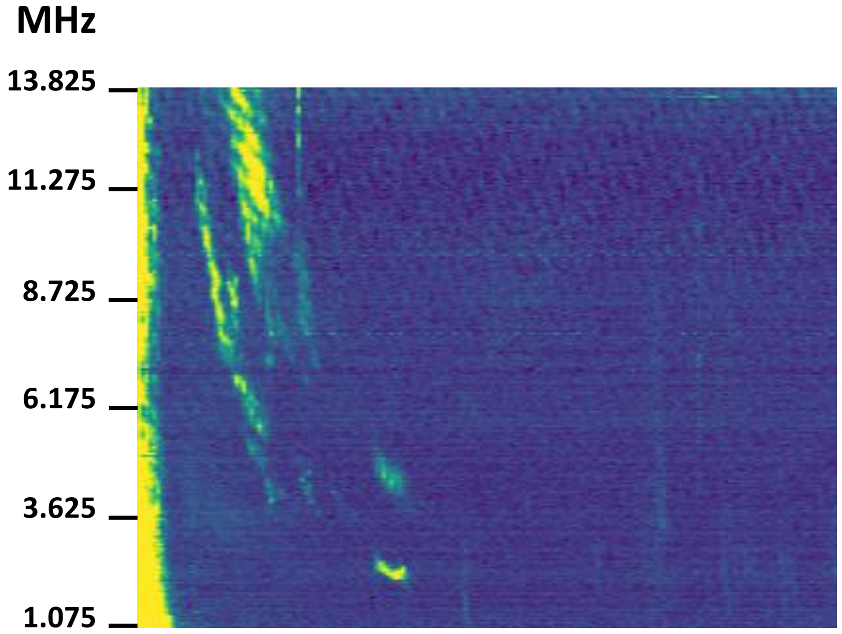
The drift rate challenge

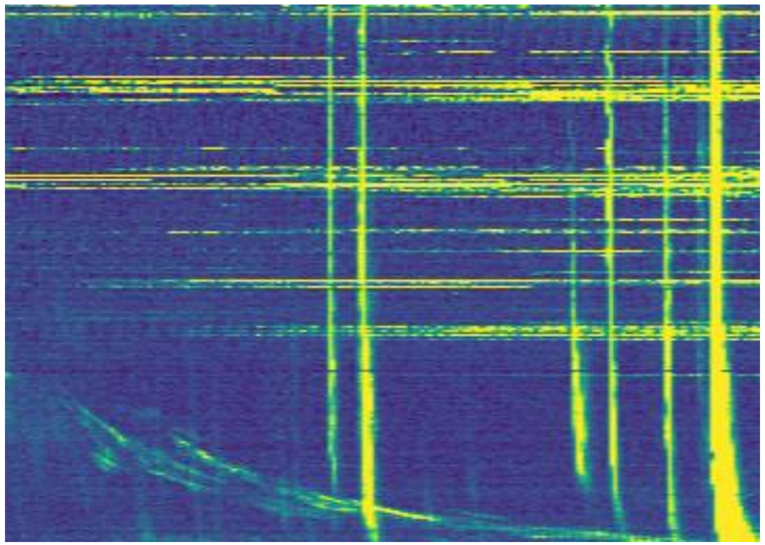
Dependency between **drift rate** and **frequency range**

→ Visual appearance depends on frequency

→ Higher variance in the data

If all bursts looked similar
→ they would be easier to detect!





➢ Can we reduce the variance without losing information?

A simple model for drift rate

Drift rate as a power law of frequency [1]:

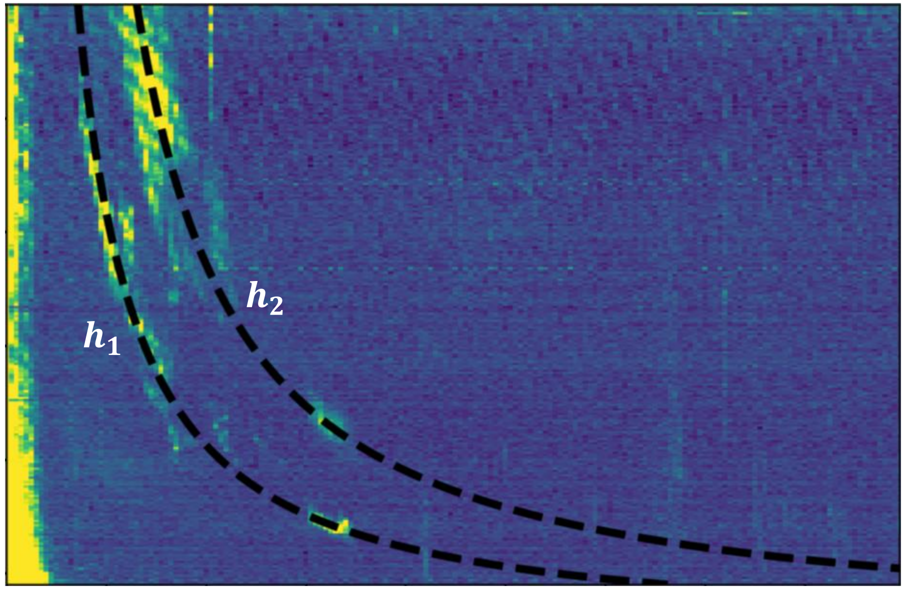
$$-\frac{df}{dt} = af^k$$

How can this knowledge be integrated into a detection system?

"not just better methods but *more physically relevant data*"
[Monica Bobra, this Tuesday(?)]

- Can we make a data representation that:
  - accounts for this physical law
  - is more appropriate for ML algorithms

$a_1 = 1.2 \times 10^{-4}$ $a_2 = 6.9 \times 10^{-5}$
$k_1 = 0.91$ $k_2 = 0.91$

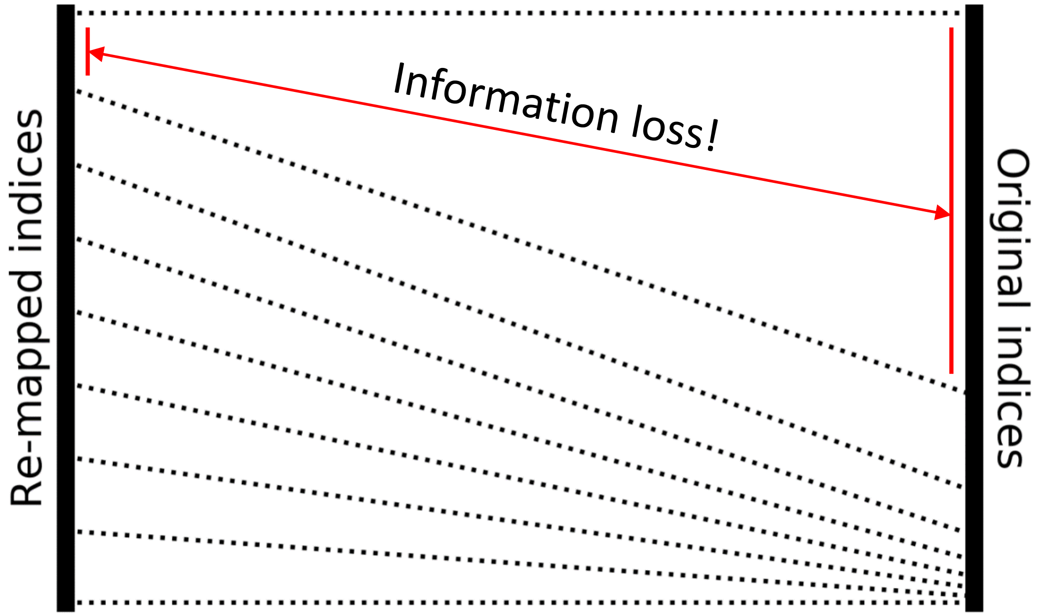


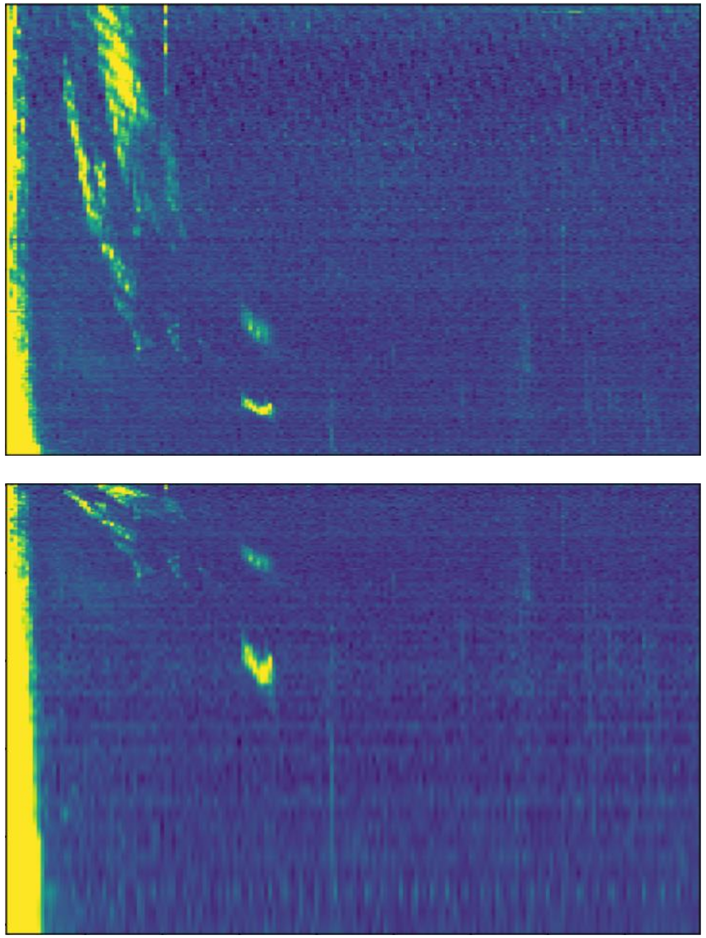
[1] Aguilar-Rodriguez et al.: A Study of the Drift Rate of Type II Radio Bursts at Different Wavelengths. Solar Wind 11/SOHO 16, Connecting Sun and Heliosphere. Vol. 592, 2005

Normalising for drift rate: previous attempt

Re-map frequencies to its inverse so that bursts become almost straight [2, 3]

$$f \rightarrow \frac{1}{f}$$

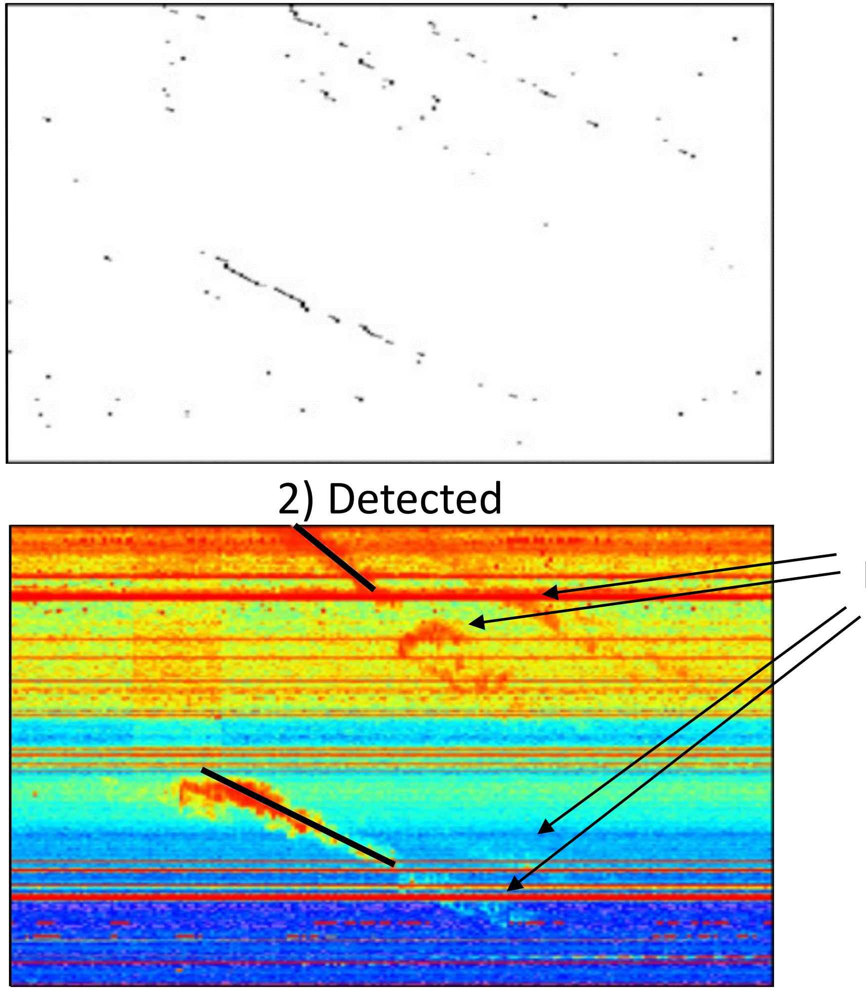




[2] Lobzin et al.: Automatic recognition of coronal type ii radio bursts: the automated radio burst identification system method and first observations. The Astrophysical Journal Letters, 2010
[3] Reiner et al.: A new method for studying remote type II radio emissions from coronal mass ejection-driven shocks. Journal of Geophysical Research: Space Physics 103.A12, 1998

Detection: previous attempt [2]

1. Pre-processing:
   - Contrast normalisation
   - Re-map frequencies
   - Remove weak signals
   - Morphological thinning → skeletons of bursts
2. Detection:
   - Identify straight segments using **Hough transform**



Conclusions:

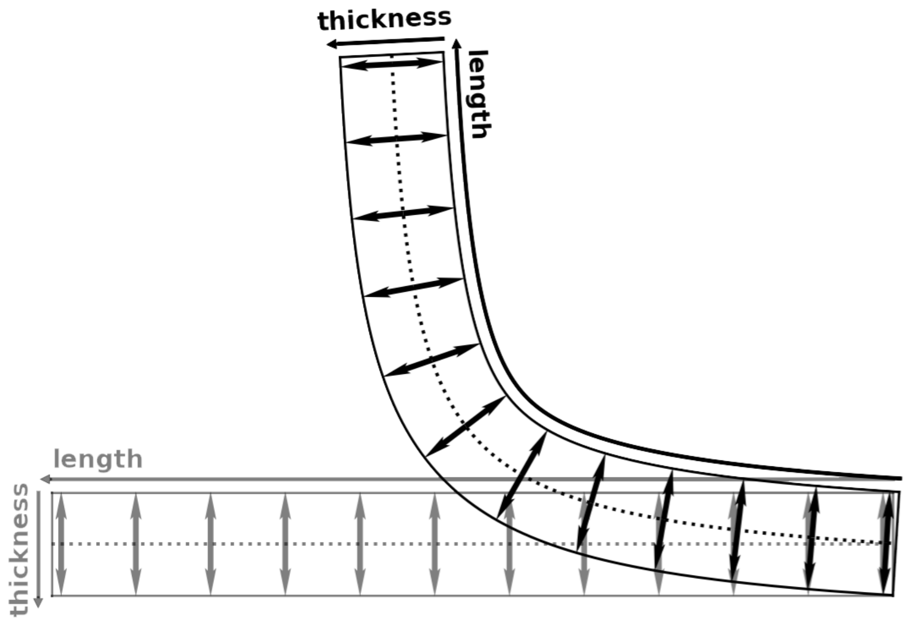
- OK for detection
- Unsuitable for parameter estimation

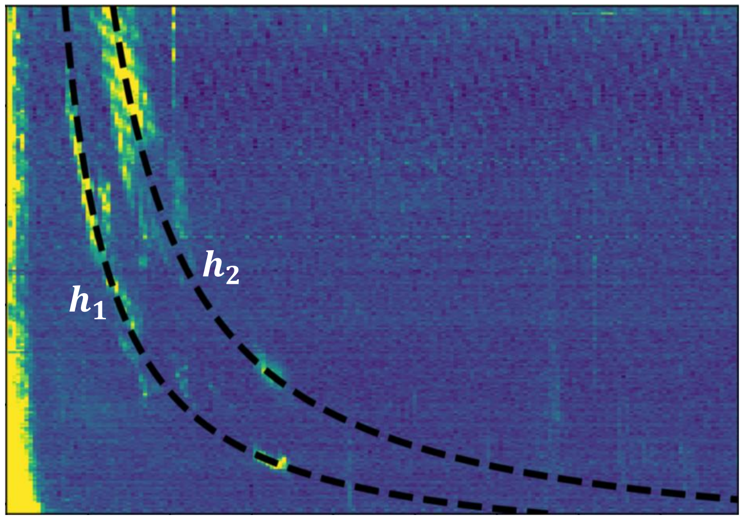
[2] Lobzin et al.: Automatic recognition of coronal type ii radio bursts: the automated radio burst identification system method and first observations. The Astrophysical Journal Letters, 2010

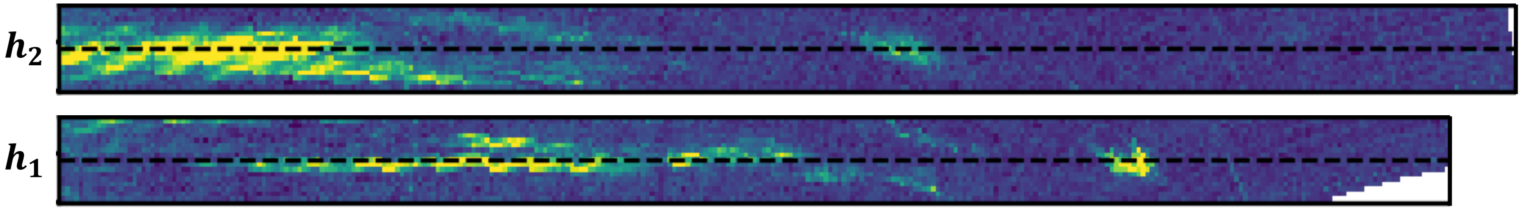
Exploiting the drift model: our solution

Re-ordering the data:

1. Parameterise curve by arc length
2. Consider a thick tube around the curve
3. Sample normals of the curve to straighten the tube

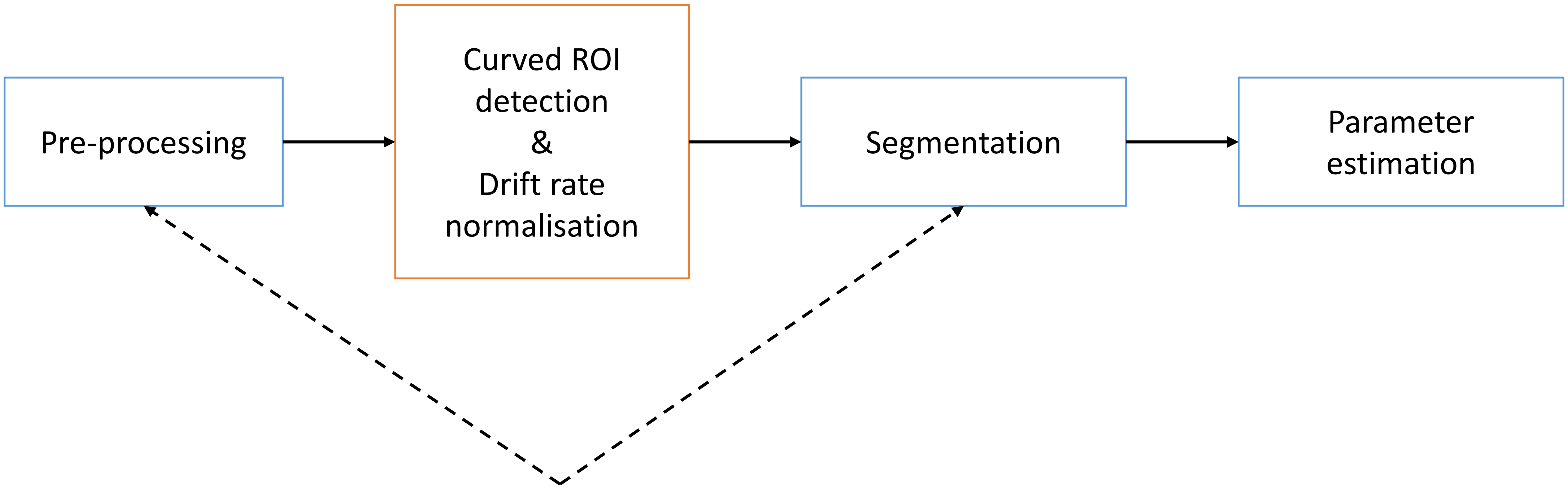






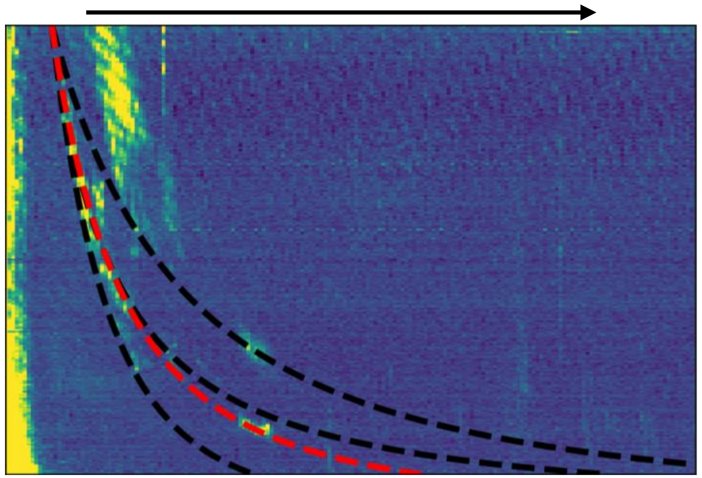
- ➢ No information loss: all frequency & time context is preserved
- ➢ Better representation of the data for ML algorithms

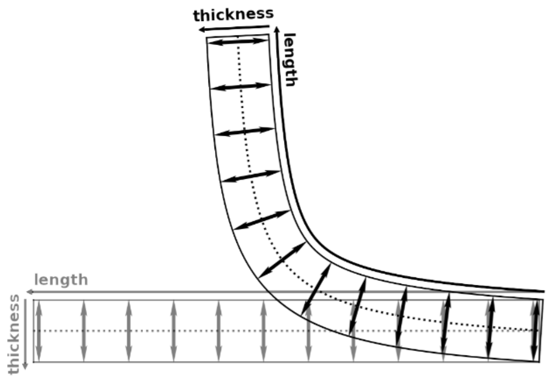
Our full pipeline



- Classical computer vision methods → Simple methods for proof of concept
- Facilitated by the new data representation

ROI detection

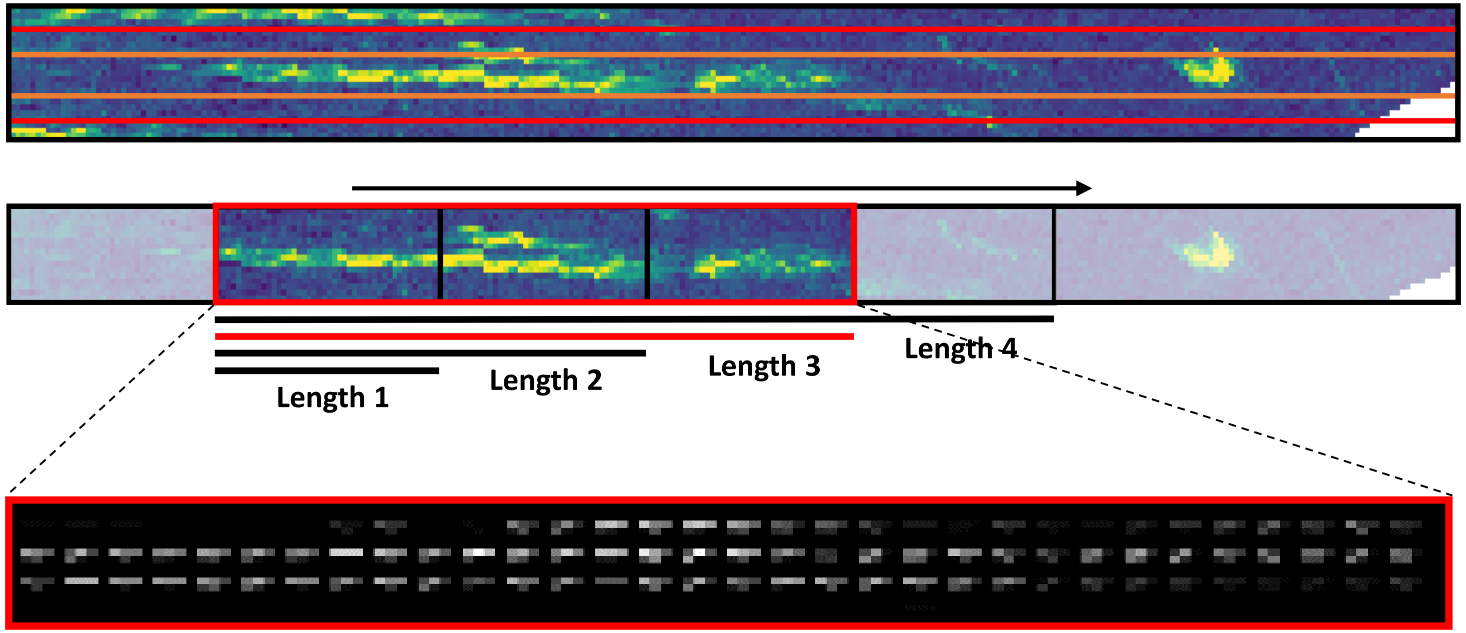




1. Sliding ROI windows:
   - 4 choices of drift rate
   - 3 choices of thickness
   - 4 choices of length

   → 48 ROIs tested at each location

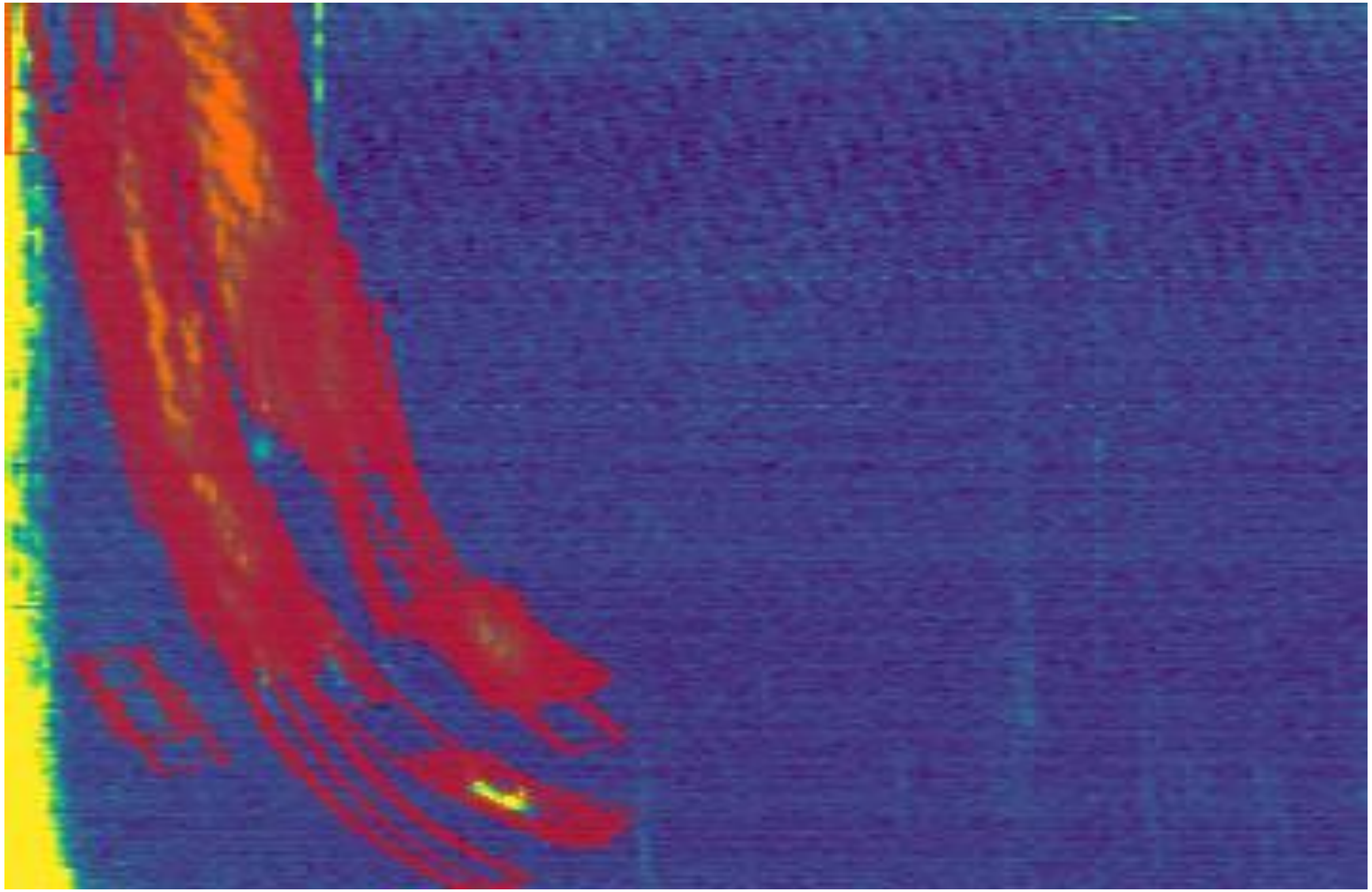
2. **Drift rate normalisation**



3. HOG feature extraction

4. Classification by logistic regression

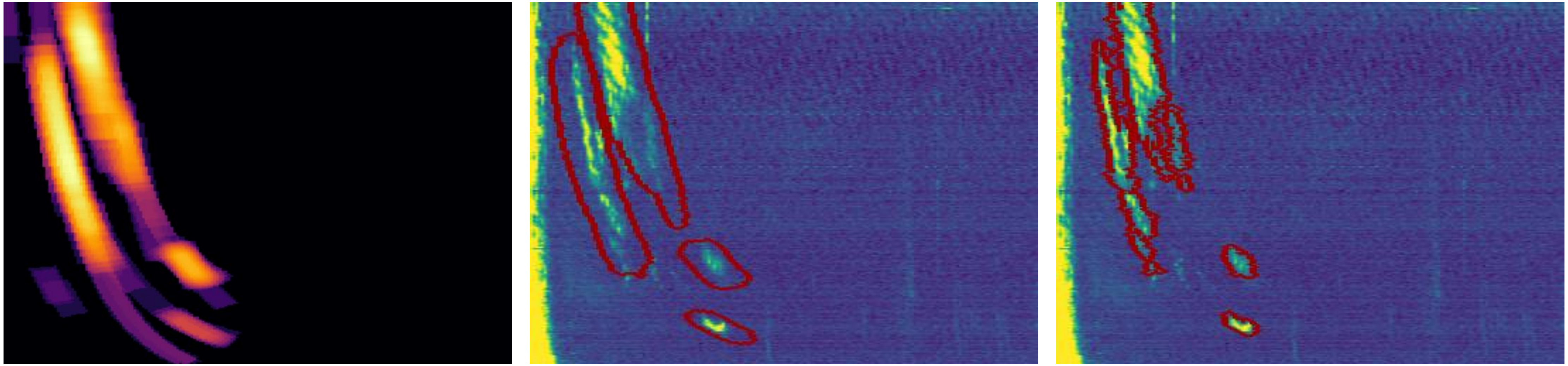
ROI detection



Segmentation

Combine ROI detections:

1. Per pixel **voting** (like in AdaBoost):
    a) # ROI detections → detection confidence
    b) Threshold on confidence
2. Refinement



1a **1b** **2**

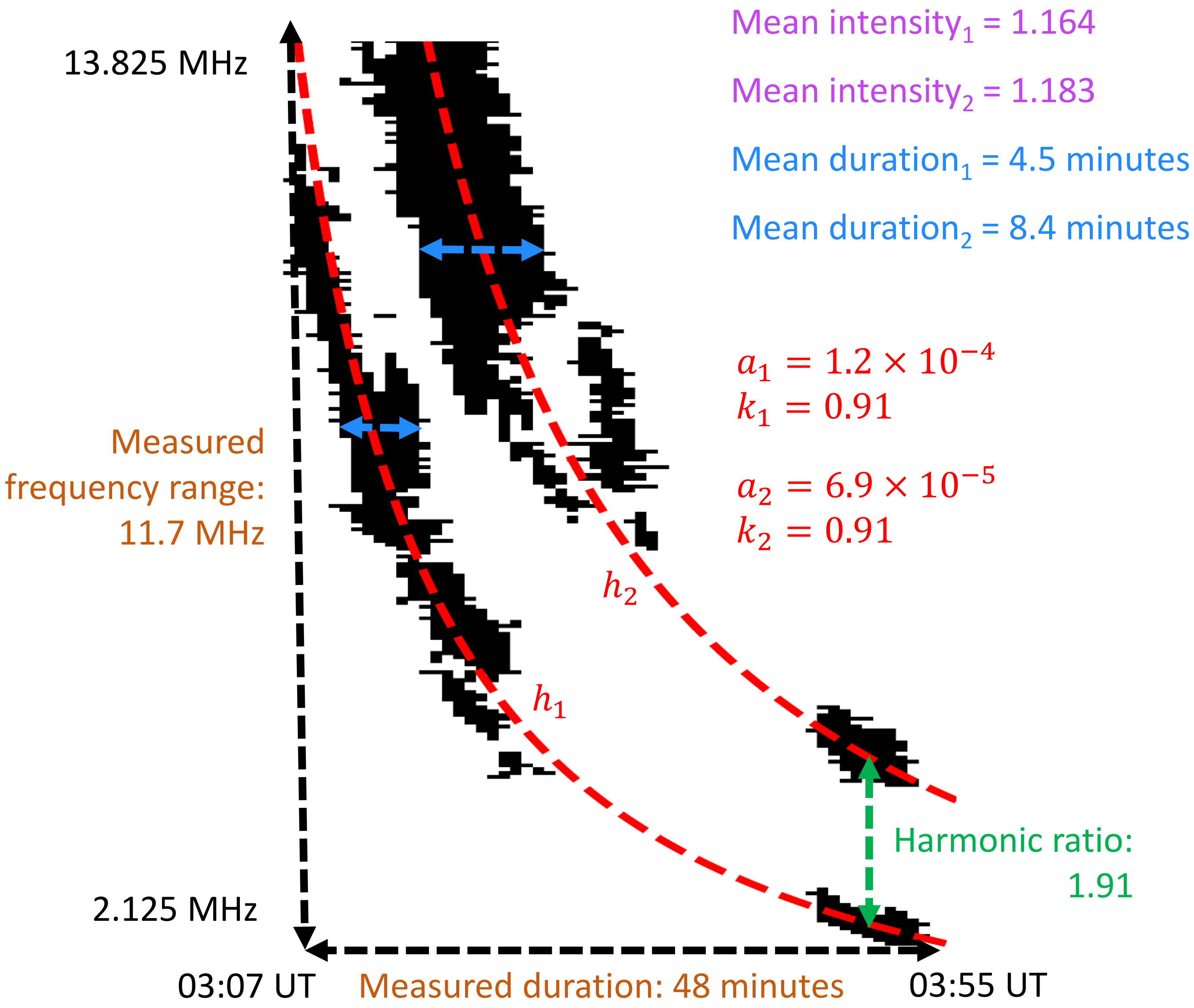
Parameter estimation

Fitted burst model:

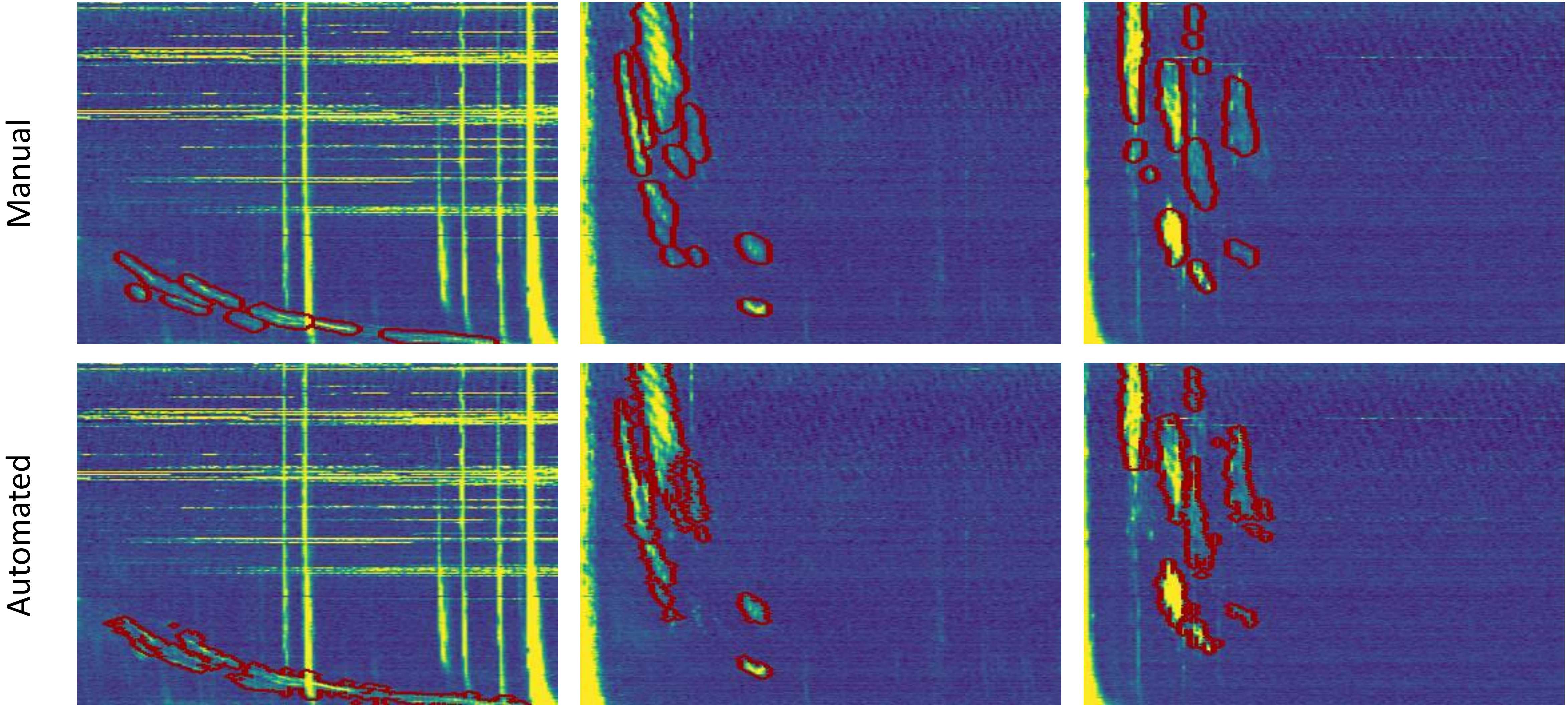
- ✓ Grouping of individual segments
- ✓ Derivation of burst parameters

  Using:

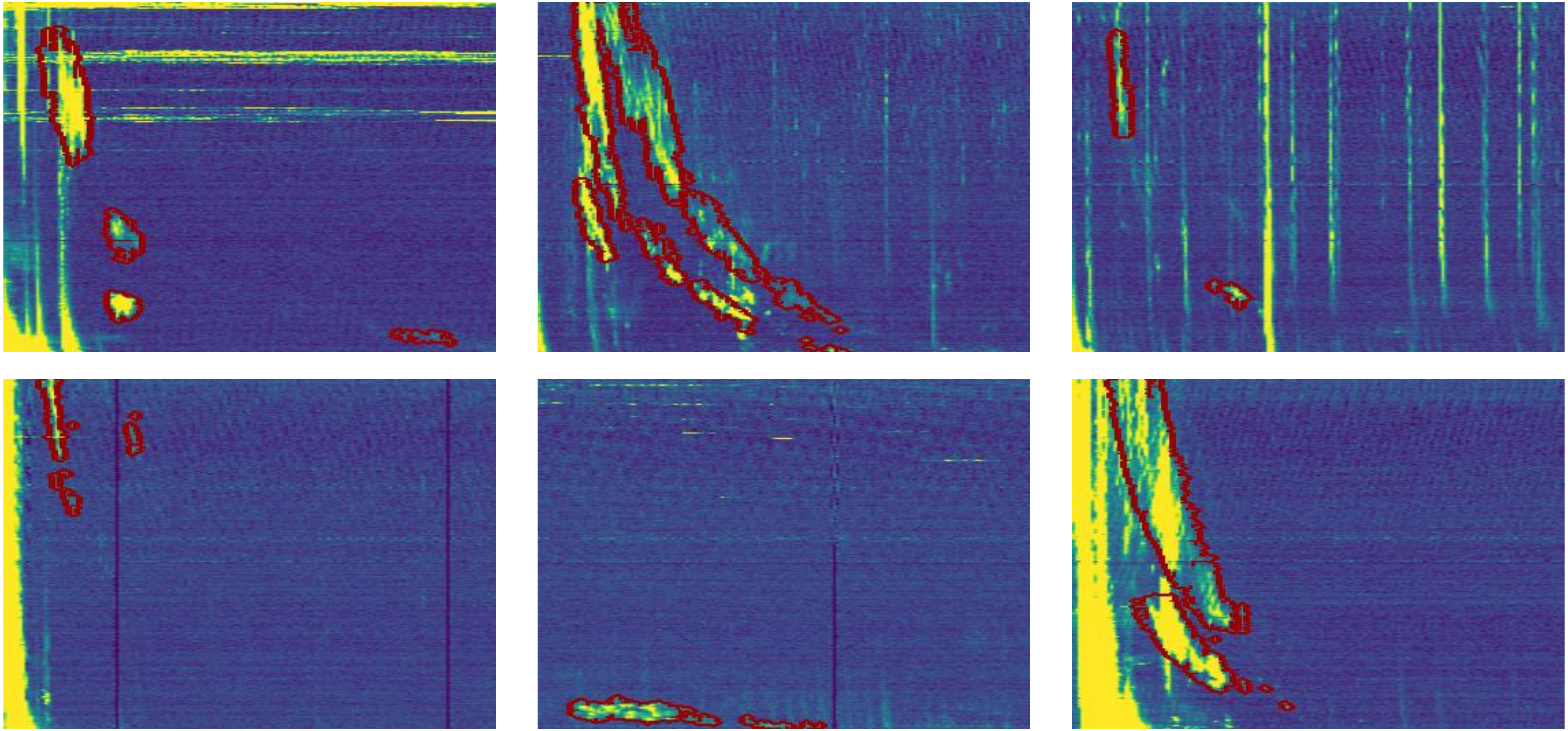
  - Segmentation mask
  - Drift rate model
- ✓ Harmonic classification



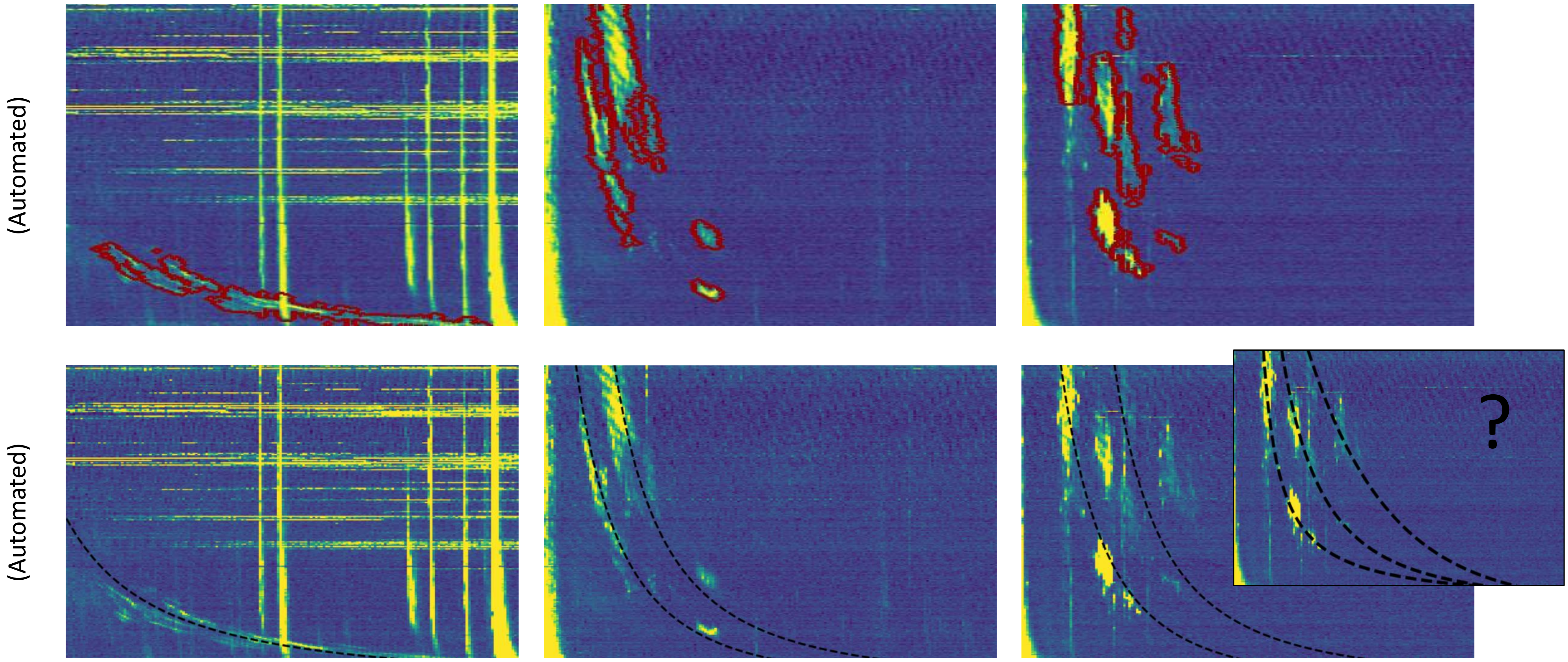
Some qualitative detection and segmentation results…



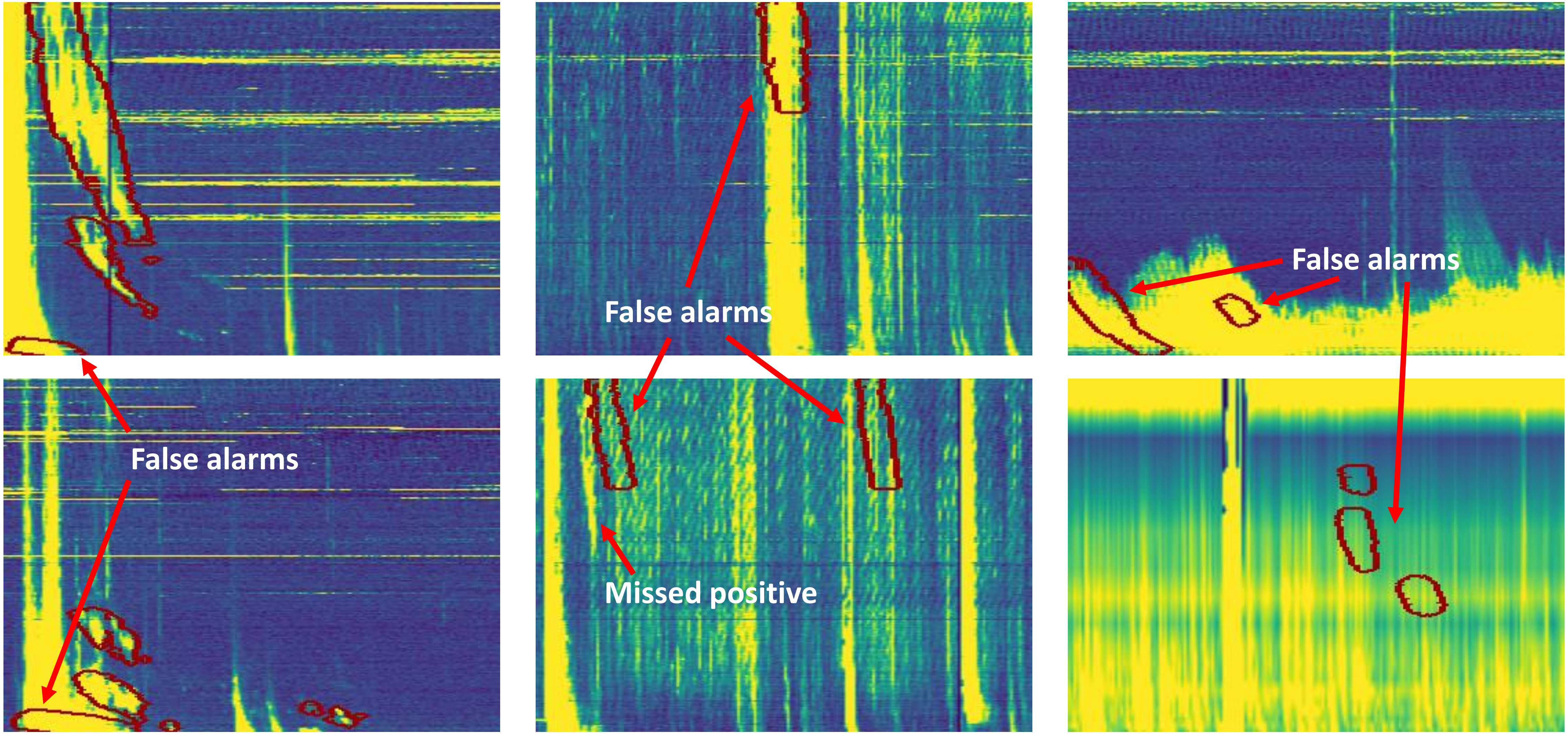
More automated detection and segmentation results…



Some qualitative model fitting



Some false positive detections



Some initial quantitative results

	Hours searched	Number of events	TP	FN	FP	Percentage of events found	Occurrence of FP	Precision	Recall	Segmentation precision (IoU)
[2] Learmonth 25 – 180 MHz 2002 (solar maximum)	510	46	36	10	5	78%	1 every 102 hours	0.88	0.78	-
Proposed WIND/WAVES 1 – 14 MHz 1997-2016	2217	244	202	42	55	83%	1 every 40 hours	0.79	0.83	42%

Experiments still in progress:

- Better grouping of burst parts?
- **Same datasets** (frequency range) for comparison
- **Same** range of solar **activity levels** for comparison
- Quantitative results on parameter estimation

TP: true positives
FP: false positives (false alarms)
FN: false negatives (missed)
Precision = TP / (TP + FP)
Recall = TP / (TP + FN)

[2] Lobzin et al.: Automatic recognition of coronal type ii radio bursts: the automated radio burst identification system method and first observations. The Astrophysical Journal Letters, 2010

Summary

- **Detection** and **segmentation** of type II bursts to enable estimation of **physical parameters**
- New data representation based on physical model of drift rate

  → data is better suited for general computer vision and machine learning methods

  → data from other instruments?
- New annotated dataset (to be released)
- Future work:
  - Type II vs all → type II vs type III vs type IV vs all
  - Group burst parts using machine learning?
  - Specialised models for different solar activity levels? (ongoing)
  - Other instruments / frequency range (e.g. Learmonth)
  - State-of-the-art computer vision and machine learning methods (e.g. deep learning)
  - Too high annotation effort → semi-supervised methods