TOWARDS AN AUTOMATED E-CALLISTO RADIO BURST IDENTIFICATION AND EVENT REPORTING SYSTEM FOR THE SPACE WEATHER COMMUNITY

Javier Bussons Gordo (1); Mario Fernández Ruiz, Manuel Prieto Mateo (2); Christian Monstein (3)

1) Physics Department, Campus Espinardo, Edificio 19. University of Murcia. 30100 Murcia, Spain. (bussons@um.es)
2) Escuela Politécnica Superior, University of Alcalá de Henares, Spain.
3) Istituto Ricerche Solari "Aldo e Cele Daccò", Università della Svizzera Italiana, Locarno, Switzerland.

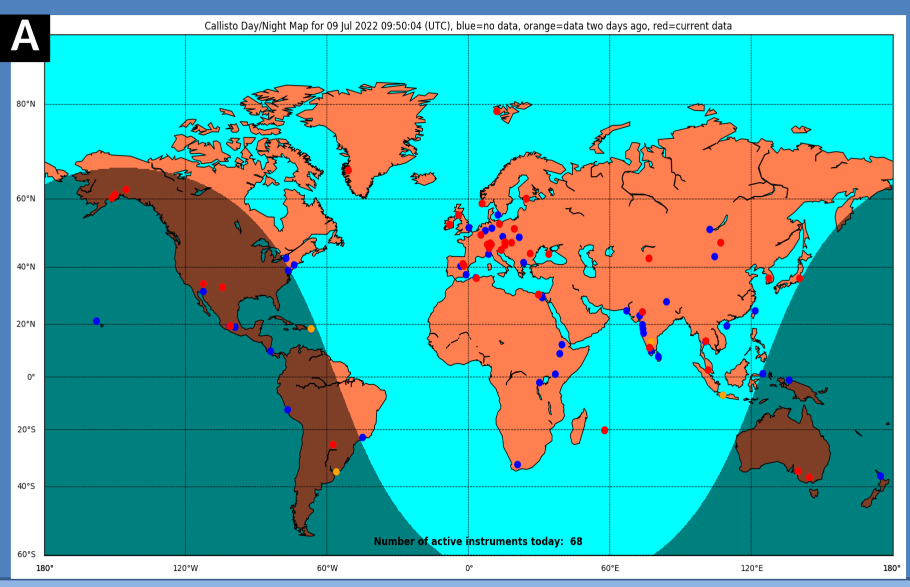
COSPAR 2022
44th SCIENTIFIC ASSEMBLY
16-24 July 2022, Athens, Greece
aMUSEd by the Athenian URANIA

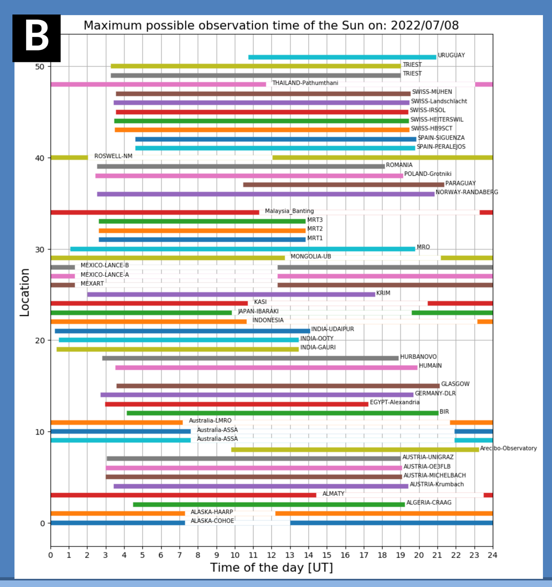
1 Introduction

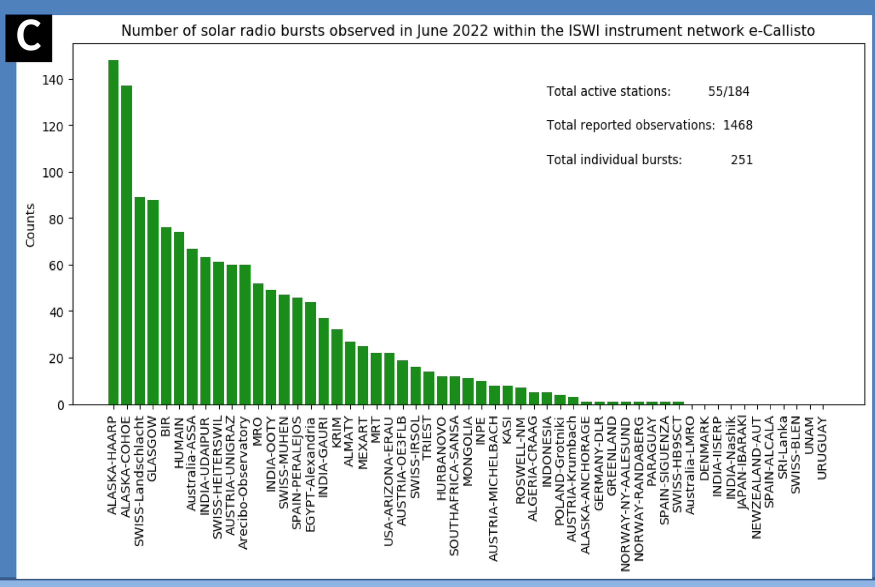
ABSTRACT. Cutting-edge space weather forecasting demands quick user-friendly access to solar observations around the globe and agile cross-match functionality. In this paper, progress on a new quicklook service to be provided by the e-Callisto global network is presented. The new service will include Deep-Neural-Network-based (DNN) solar burst identification taylored to each working e-Callisto station, automatic event reporting and a centralized database with cross-matches between satellite- and ground-based burst catalogs. As an implementation example, missing e-Callisto reports from the past are being produced both in text and scatter-plot formats.

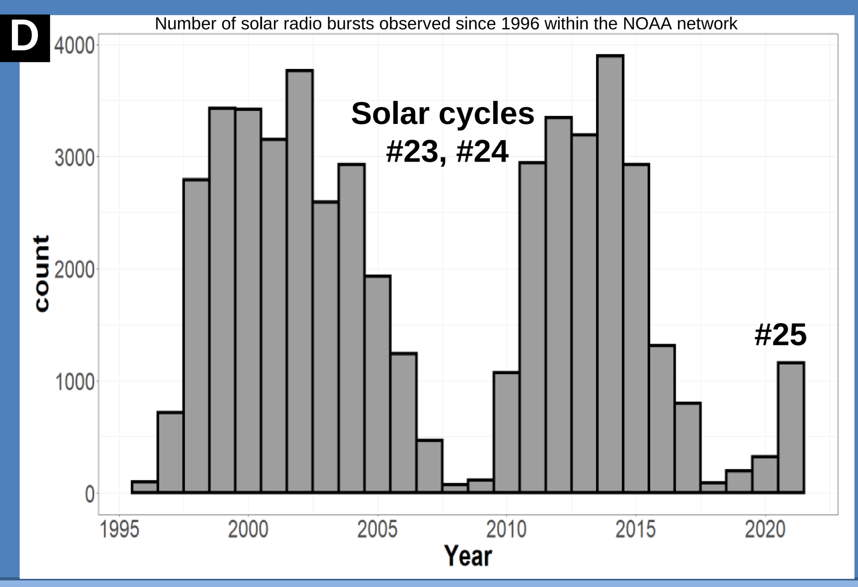
CONTEXT. Solar radio bursts (SRBs) play a key role in the study of particle acceleration and propagation processes at the heart of solar flares and coronal mass ejections:
• Since features like the presence of open magnetic field lines, shock waves or particle confinement are revealed by different types of bursts (types III, II and IV, respectively), good-quality burst identification and discrimination is crucial.
• Since early forecasting of potentially-hazardous Solar Energetic Particle (SEP) events hinges on the empirical association between SEPs and SRBs –the latter reaching us well in advance–, prompt burst reporting is a must.

e-CALLISTO. Event-scene reconstruction involves careful cross-matching of data collected both by space-borne devices, either in situ or via remote sensing, and by Earth-based observatories. Among the latter, the e-Callisto worldwide network (Fig. A) of inexpensive radio spectrometers around the globe (193 as of July 2022, 70 of them providing daily data) offers full-day coverage of the Sun with the redundant event perspective given by typically ten to thirty stations active at any given time (B). Burst detection statistics for e-Callisto stations are shown (Fig. C, June 2022). We are now heading towards solar maximum (see solar activity cycles Nos. 23, 24 and the onset of 25 in Fig. D in terms of bursts detected by NOAA stations since 1996).

A


B


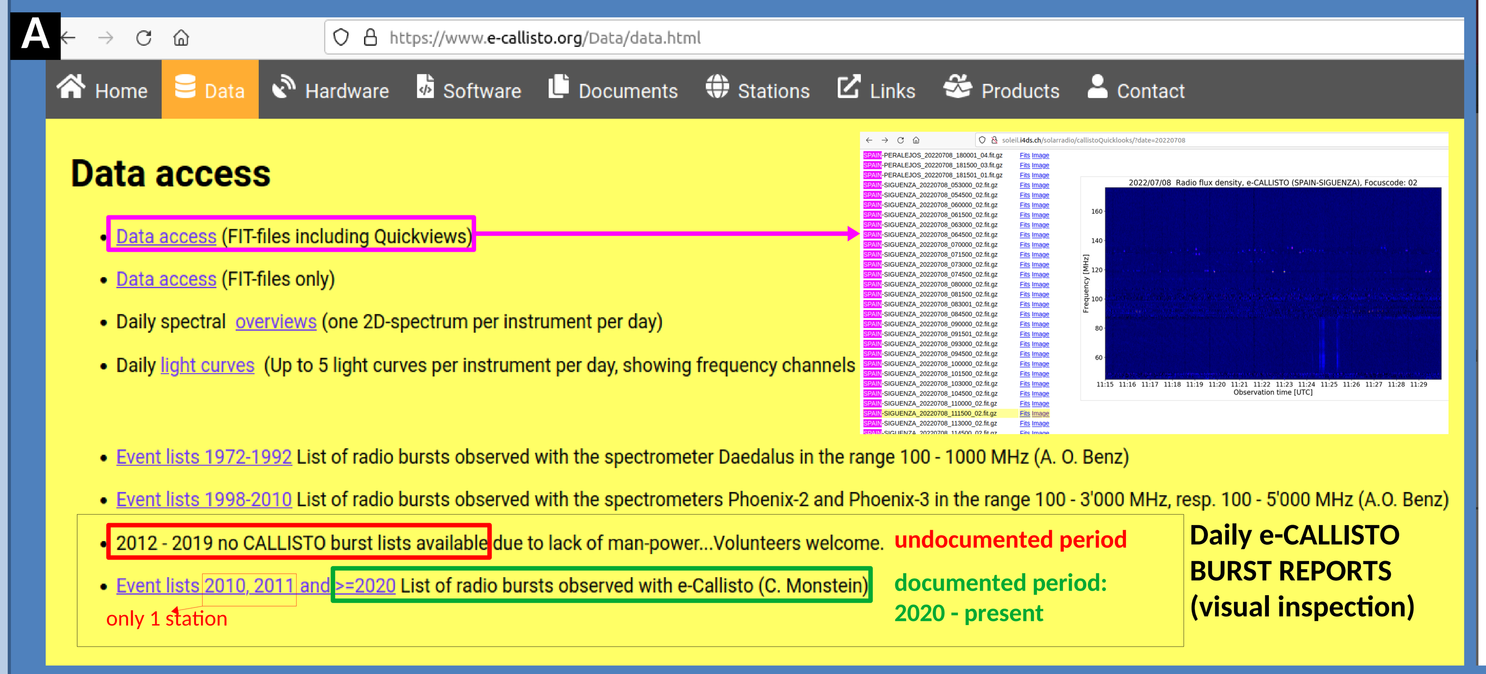
C


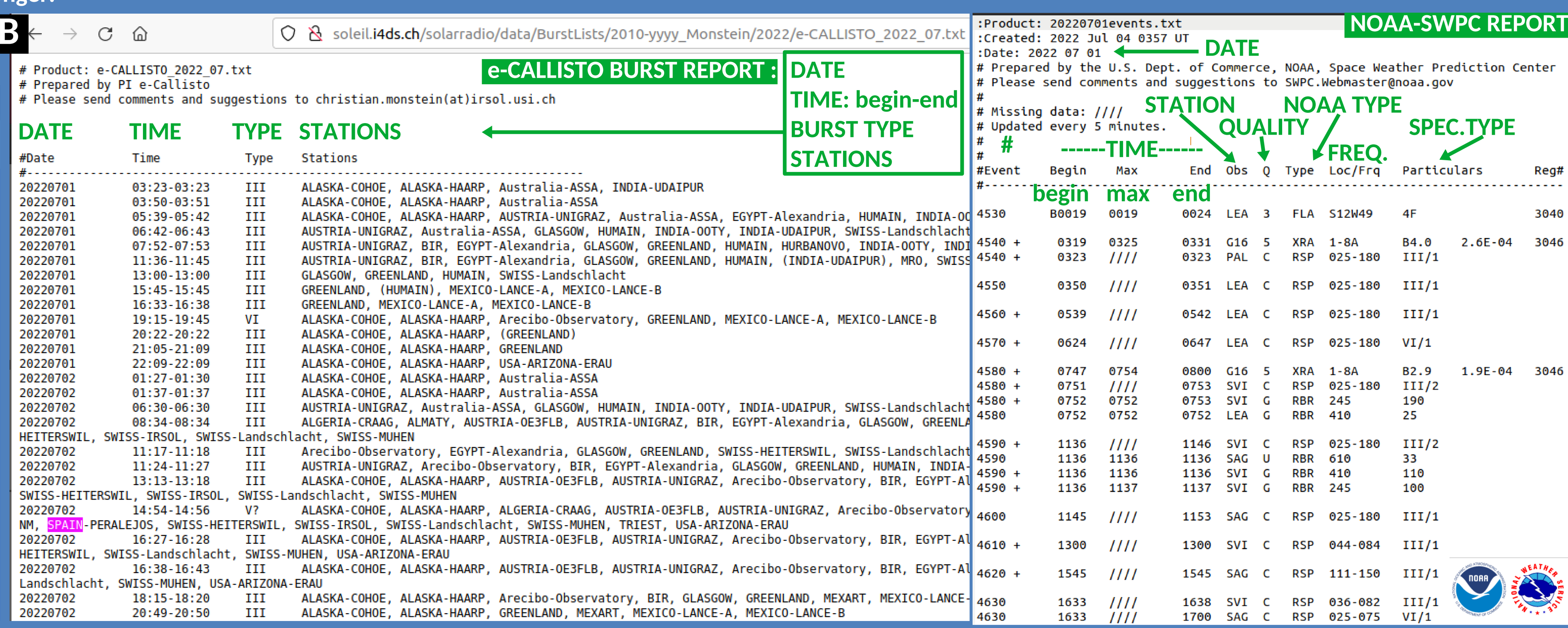
D


2 Data

A) FITS-format dynamic spectra in the 20 – 870 MHz station-dependent tunable frequency range from all e-Callisto stations are centralized at Fachhochschule Nordwestschweiz (FHNW), published on the internet in near-real time and quick-looked by an expert before a daily burst report is issued based on human inspection. This kind of data, with significant spectral and time resolution, provides valuable information about prompt phases of SEP acceleration. In combination with satellite measurements of X-ray and particle flux, it helps reconstructing the sequence of phases in episodic events: flare onset, early Coronal Mass Ejection (CME), mature CME. At long time scales, where short events ride over a continuum of solar activity in an 11-year-long cycle, this wealth of data may become a mine for statistical studies of SRBs: types, duration, grouping, ordered sequences, solar cycle comparison. Yet, in order for the scientific community to benefit from its full potential, the implementation of automatic burst identification and event reporting seems mandatory, as current manual inspection of thousands of files a day cannot be sustained much longer.

B) Excerpts from an e-CALLISTO BURST REPORT (available only since 2020) and a NOAA-SWPC REPORT.

A


B


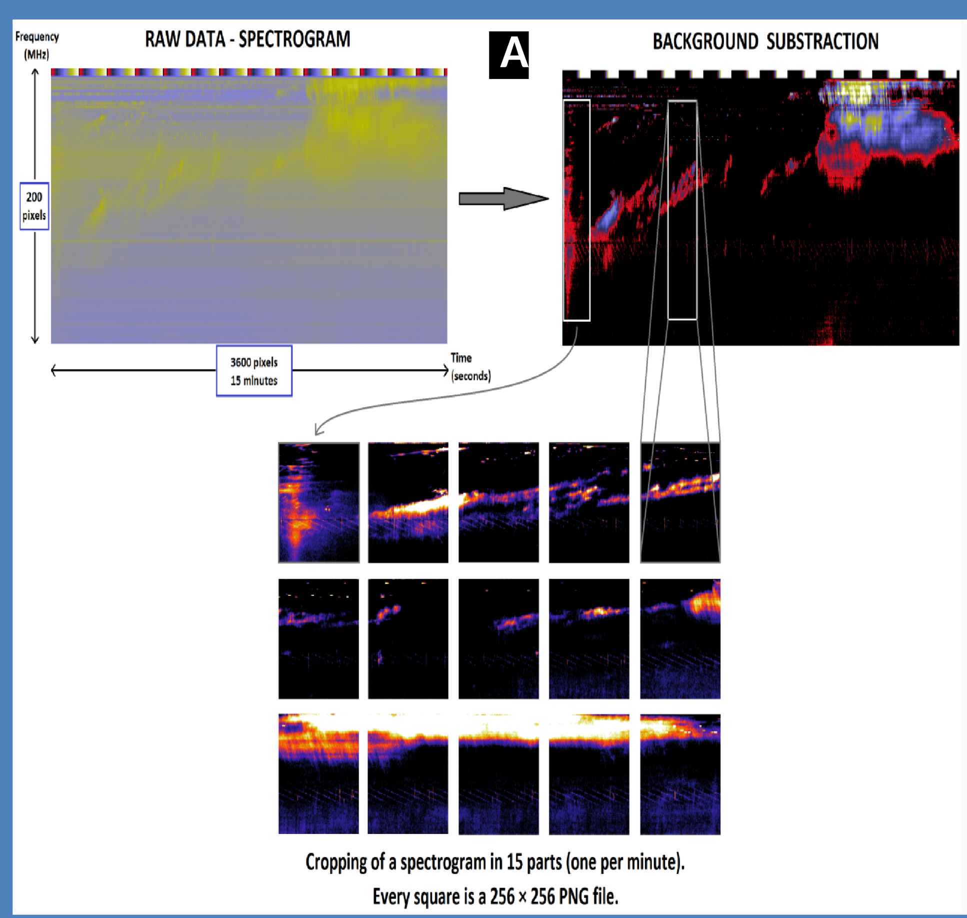
3 Method

DEEP LEARNING. The NVIDIA Deep Learning GPU Training System (DIGITS) is used to train AlexNet, a convolutional neural network, to discriminate burst signals (Yeas) from noise (Nays), both given as input. The network output (prediction) is the probability P(Yea) for each observation to be a burst. Let us follow the stages of the process:

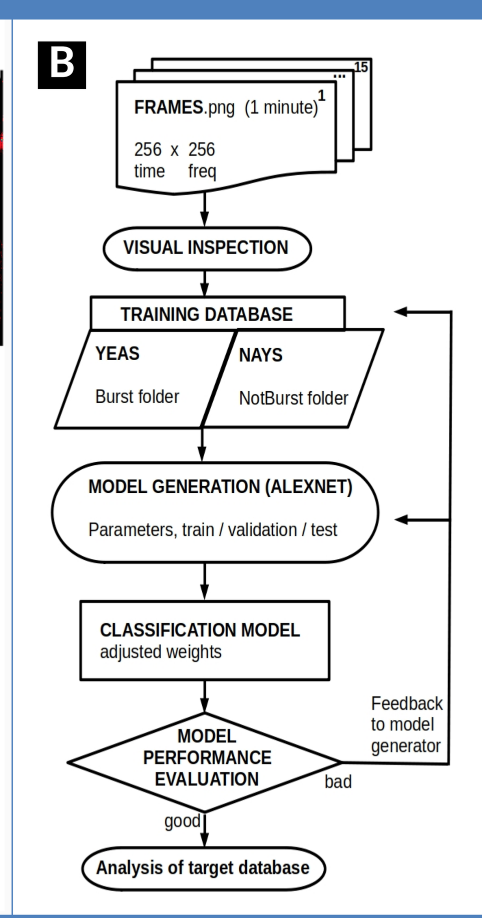
A) PRE-PROCESSING (steps and goals)
1) Background subtraction: the average spectrum over a 15-minute run is subtracted, thereby eliminating those noise sources which are constant during the run. 2) Frequency range selection: avoid observatory-dependent noise-swarmed frequency channels. 3) Split spectrogram into 1-minute frames: enhance burst features and adapt resolution to DNN requirements (256 x 256-pixel PNG format).

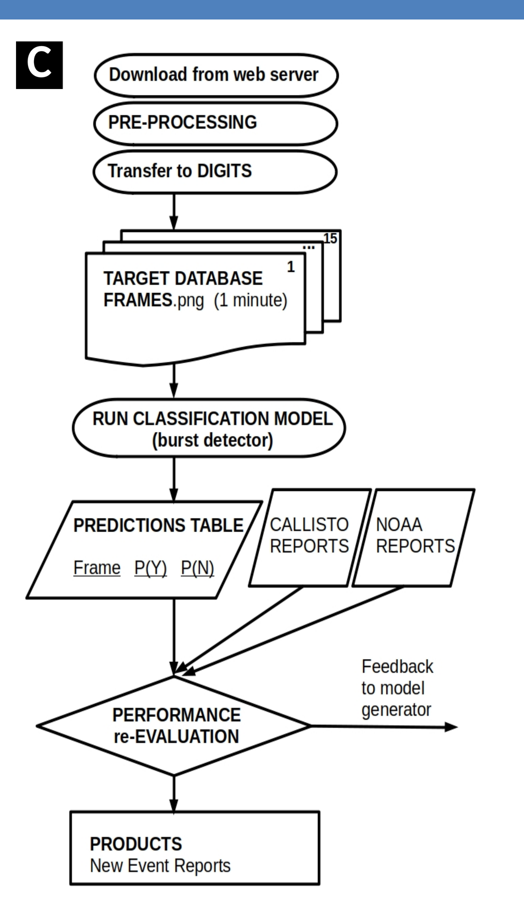
B) TRAINING OF AN IMAGE CLASSIFIER OR MODEL.
1) Build training database: made of two sets of frames, one from documented bursts (YEAS folder) and the other from background data (NAYS folder). 2) Model generation: optimize network parameters (epochs, gradient solver, train/validation/test percentages) and obtain the final set of convolutional weights ("the classifier"). 3) Model performance evaluation: we set as figures of merit the percentages of False Negatives FN (true bursts "missed" by the classifier) and False Positives FP along with the fraction of runs with at least one positive prediction –the subset of runs the user will want to study in more detail. The (ground) truth of a prediction is assessed using Callisto Burst Reports issued by an expert on duty. Lessons learned from careful study of False Negatives and False Positives are incorporated as feedback to improved models; in this process, a number of UNREPORTED BURSTS, unnoticed to the human inspector, are discovered. By default, the threshold P(Yea)>P_thresh for the Yea/Nay decision is set to 50% but optimization of this parameter, currently underway, holds room for further improvement.

C) RUN CLASSIFIER through the TARGET DATABASE to obtain PREDICTIONS [frame, P(Yea)] and final PRODUCTS. Currently, the pipeline generates the following PRODUCTS:
- for documented data (ground truth is know): re-evaluate performance (dynamic feedback to model)
- for undocumented data: NEW EVENT REPORTS (Fig.4B)

A


Cropping of a spectrogram in 15 parts (one per minute).
Every square is a 256 × 256 PNG file.

B


C


4 Results

A) PERFORMANCE EVALUATION of automatic burst identification models for 3 e-Callisto stations (Glasgow, UK; Landschlacht, Switzerland; ASSA, Australia) using one year's worth of data (2021, human inspection reports available). In general we are now capable of filtering out >87% (100%–13%) of the total data flow, while keeping >86% of the bursts, with most, if not all, of the <14% missing bursts being very weak ones. Initially, so-called "single" models, where training datasets contain images only from the target station, were tested; now, for comparison, "hybrid" models have been included in our study, with similar or better results. The table also shows the effect of the probability threshold (P_thresh): imposing a higher threshold (fewer images pass the filter) results in an increase in False Negatives and a decrease in False Positives.

B) AUTOMATIC BURST REPORTS FOR THE UNDOCUMENTED PERIOD (2012-19) are now being produced.
A preliminary example with a month's worth of data from the BIRR Observatory (July 2014) is shown where only one other e-Callisto station (Glasgow) and the NOAA-SWPC report are cross-examined.

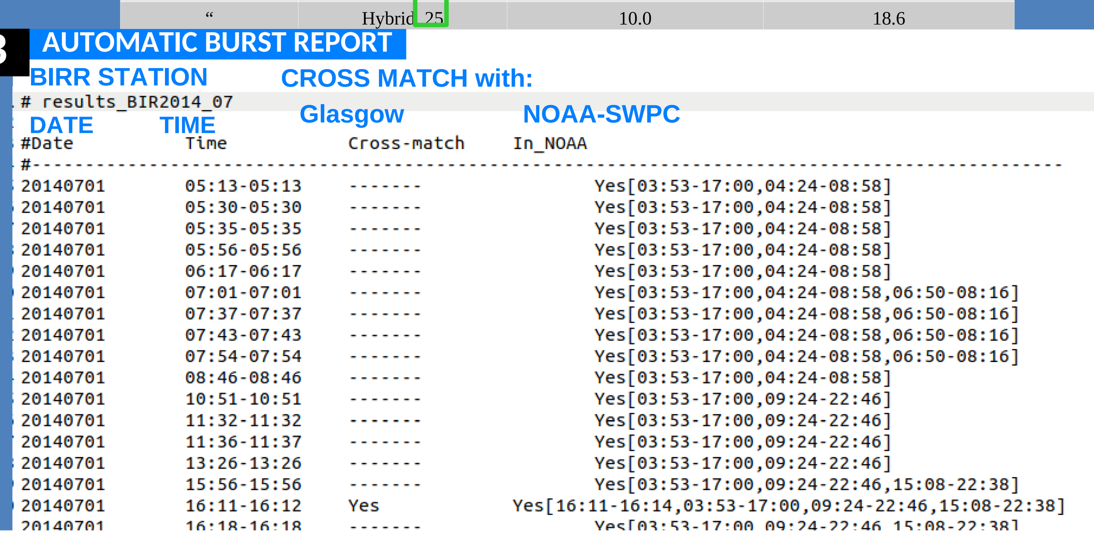
C) AUTOMATIC STATION-vs-TIME SCATTER-PLOT OUTPUT (preliminary): minute-by-minute burst coincidences for 17 e-Callisto stations (ordered by geographic longitude) and the NOAA list (first row, in blue). This information, to be issued in near-real time along with automatic burst alerts, will be useful for both immediate space weather and deferred scientific usage.

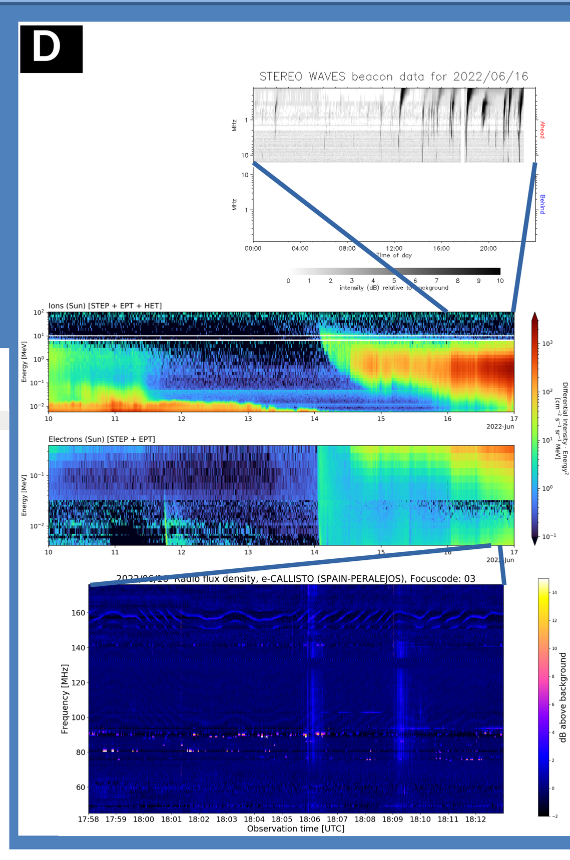
D) Recent example (June 2022) of burst cross match between e-Callisto and space-borne observatories (STEREO/Swaves and recently-launched Solar Orbiter).

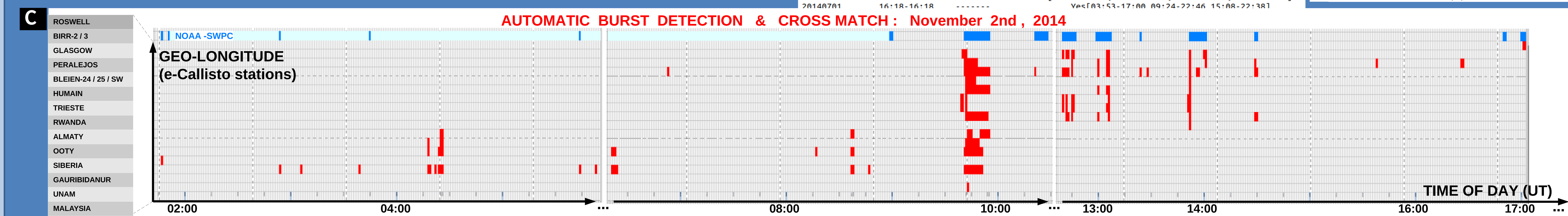
IMMEDIATE FUTURE: A second e-Callisto data center is currently being installed at Casa del Doncel, Sigüenza, Spain, which will mirror the server at FHNW and offer automatic solar radio burst detection, cross-matching and reporting capabilities via an open-access web service for the space weather community.

A

OBSERVATORY	Model Type, P_thresh	FALSE NEGATIVES %	FALSE POSITIVES %
GLASGOW	Single, 50	14.8	10.5
LANDSCHLACHT	Single, 50	13.7	19.5
ASSA-Australia	Single, 50	14.6	13.8
"	Single, 75	19.6	9.8
"	Single, 35	13.1	16.6
GLASGOW	Hybrid, 50	15.1	7.6
ASSA-Australia	Hybrid, 50	11.9	13.1
"	Hybrid, 75	12.9	9.5
"	Hybrid, 25	10.0	18.6

B


D


C


Acknowledgments:
The CELESTINA project is funded by the Junta de Comunidades de Castilla-La Mancha and the European Union (project SBPLY/19/180501/00023). JBG (FISPAC-UM) is supported by Fundación Séneca, Región de Murcia (project 20949/PI/18).
We also thank:
- FachHochschule NordWestschweiz (FHNW), Switzerland, for hosting the e-Callisto server;
- Jorge Alvarado, Paco Chávez and Iñaki Hidalgo (UEX, UCM, Micro-Stress-MAP-CM project Y2018/NMT-4668), for making available the DIGITS environment and for advice with Deep Learning;
- Our site staff, especially Abel Moreno (Peralejos de las Truchas) and Patricia Rodrigo (Casa del Doncel, Sigüenza), for their maintenance work.

UNIVERSIDAD DE MURCIA
Universidad de Alcalá