REINFORCE

REsearch INfrastructures FOR Citizens in Europe

Glitch Hunting

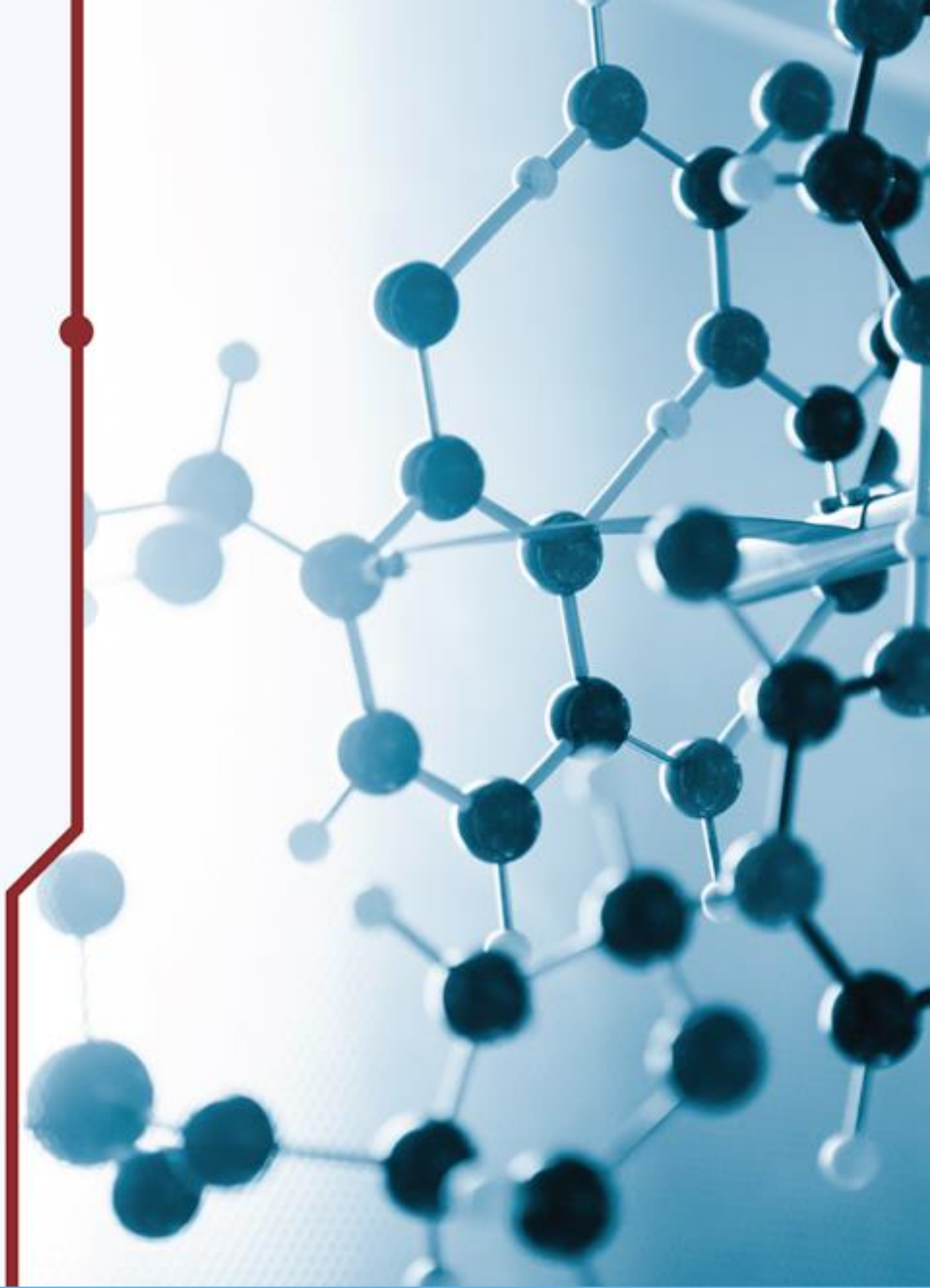
with Citizen Science and Machine Learning

Massimiliano Razzano[1,2], Francesco Di Renzo[3], Francesco Fidecaro[1,2] Gary Hemming[3], Stavros Katsanevas[3]

(1) University of Pisa (2) INFN section of Pisa (3) EGO

On behalf of the GWitchHunters REINFORCE team

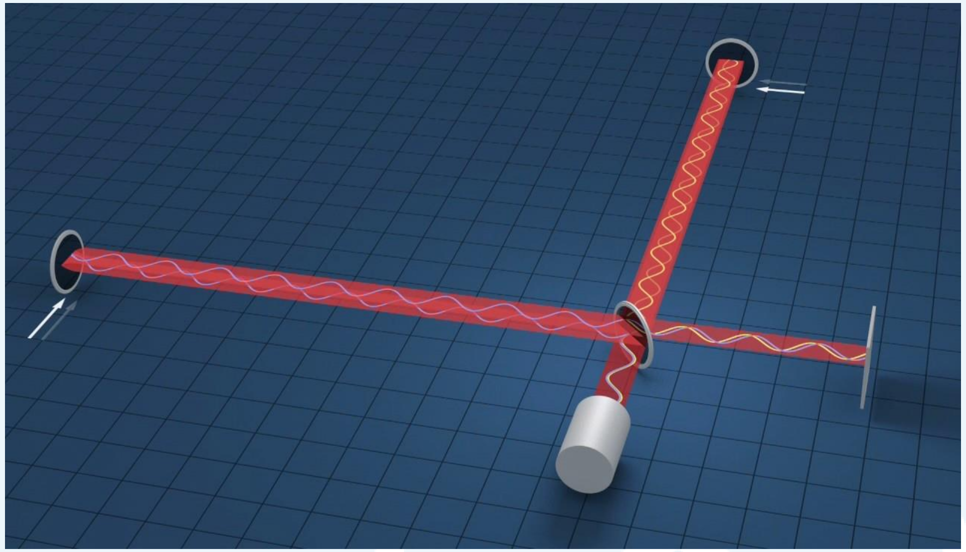
Machine Learning in GW searches
EGO 38 – 30 Sept 2022

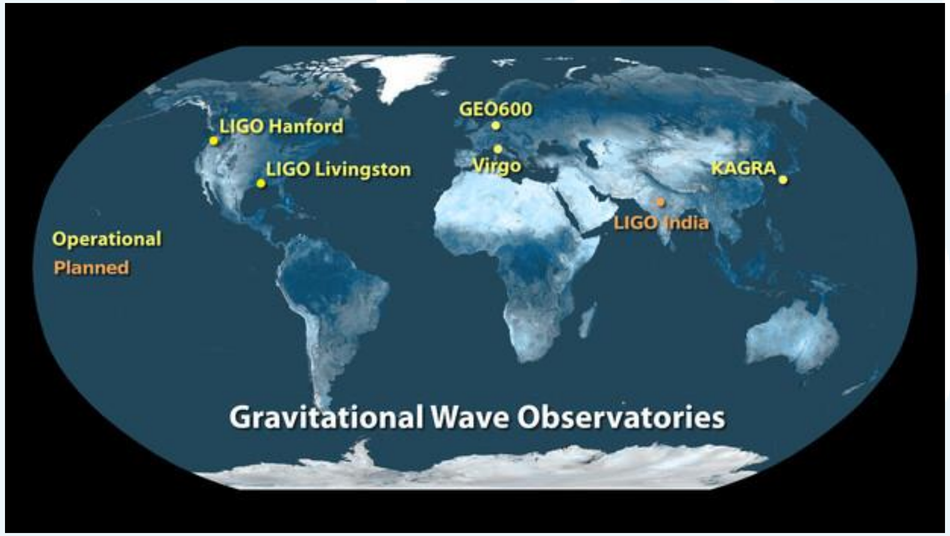


How to detect gravitational waves

- **Extremely tiny signals**
  - Typical GW sources induce a deformation of 10^{-18} m over a length of ~ few km
  - High background noise!

- **Laser interferometers**
  - Exploiting interference between orthogonal laser beams
  - Typical km-long scale + Fabry-Perot cavities
  - Frequency range ~20-20000 Hz
  - Advanced methods to reduce noise
  - Detectors working as a network

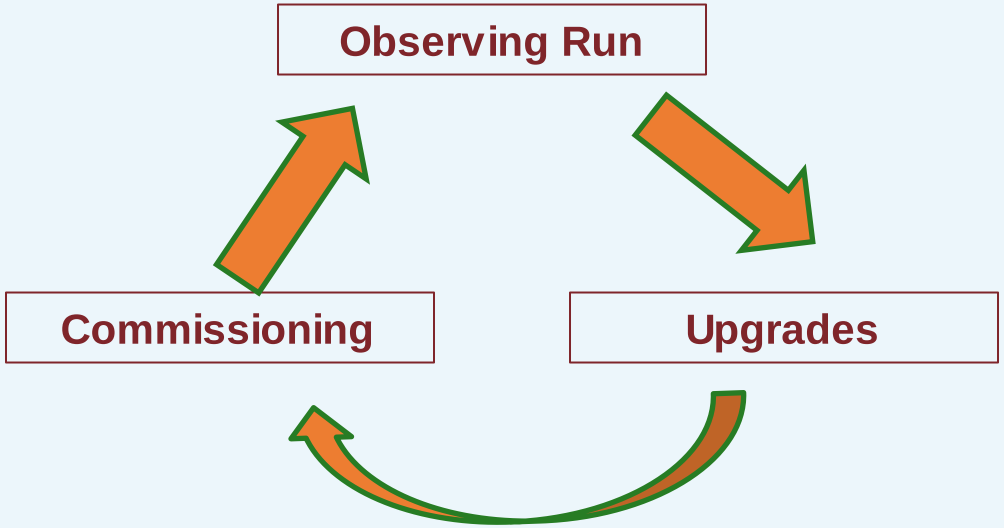




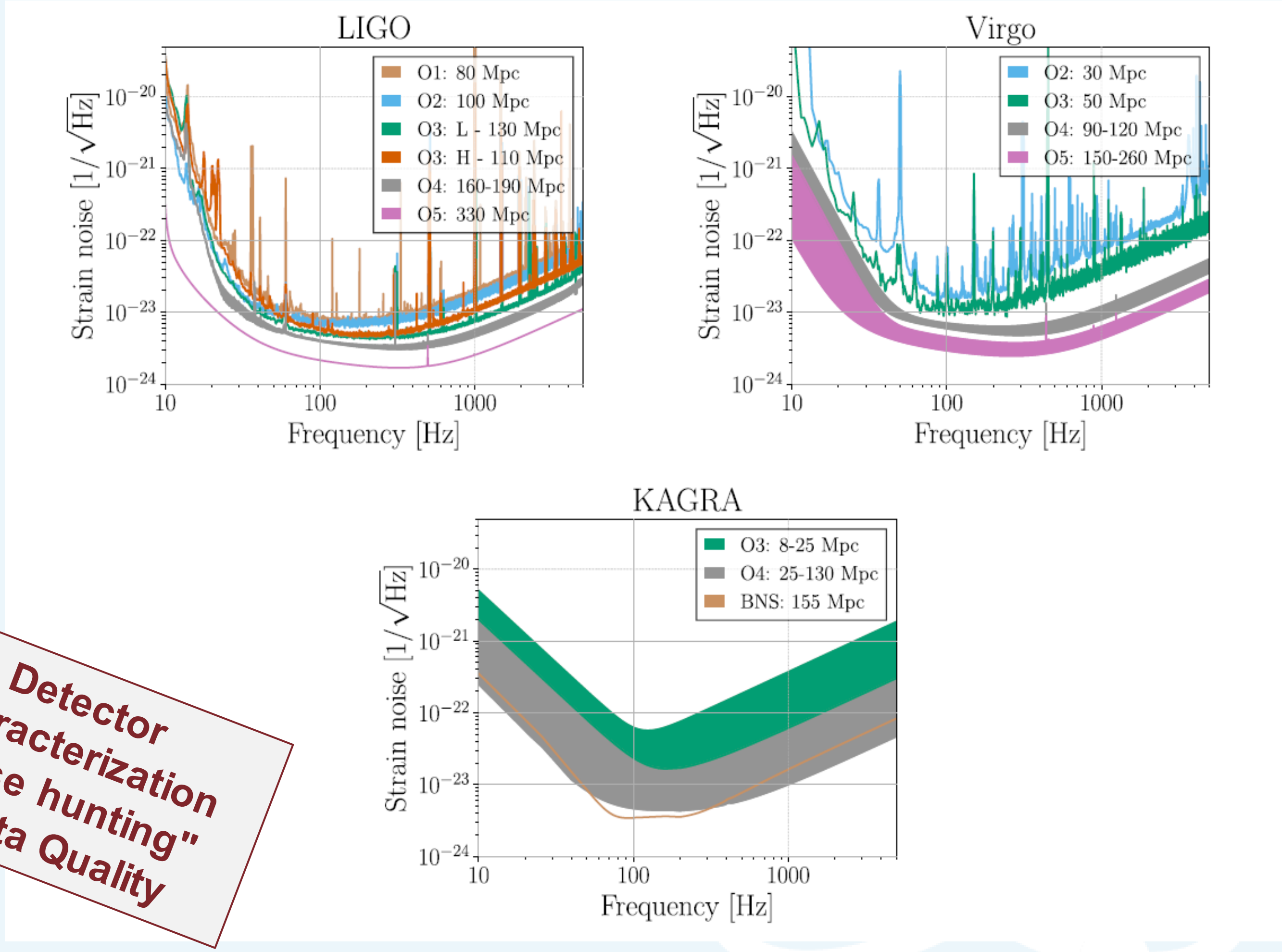
Credits: LIGO/T Pyle; Caltech/MIT/LIGO Lab

Detecting Gravitational Waves

- **Sensitivity varies with frequency: main noise sources**
  - Low frequencies: Newtonian, seismic
  - Mid frequencies: thermal processes
  - High frequencies: quantum noise



Detector Characterization "noise hunting" & Data Quality

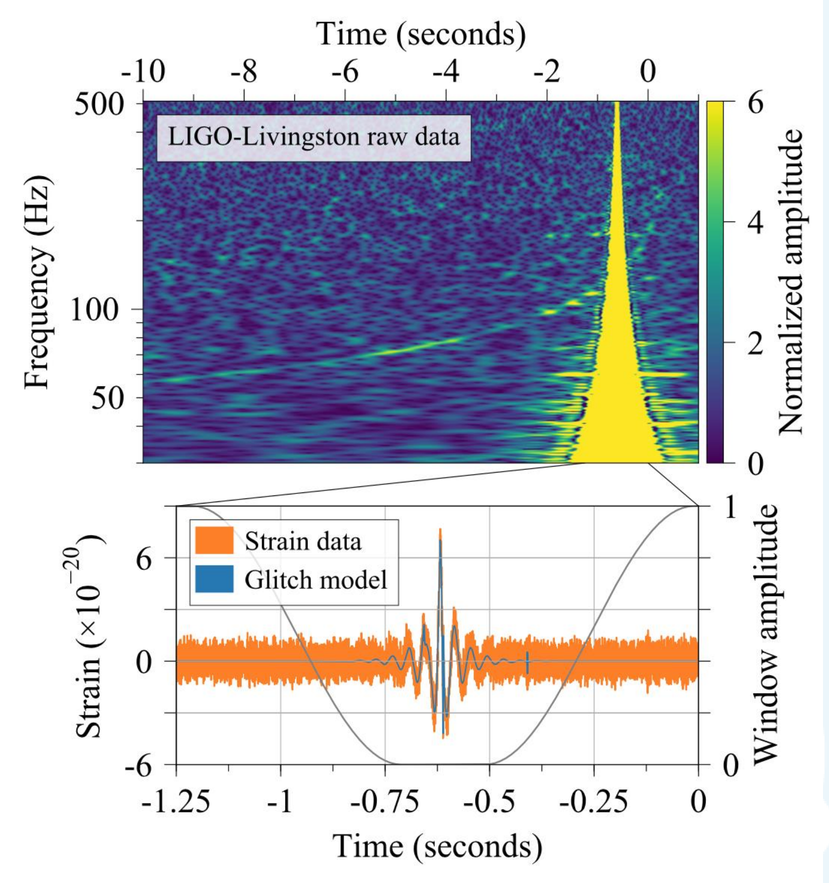


"Observing Scenario" paper
Abbott et al, (LVK collaborations) 2020, LRR, 23, 3

Noise glitches

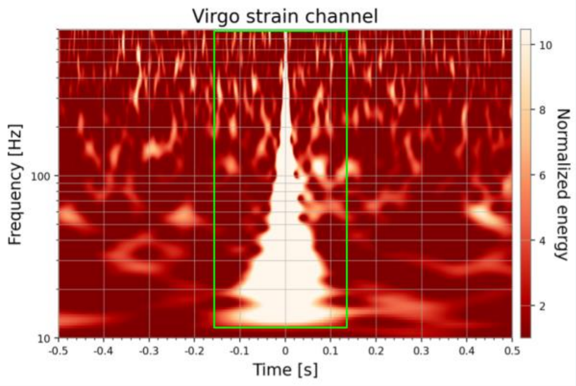
- **Noise is not stationary**
  - Transient events can happen
  - Not related to astrophysical source, but local disturbances
  - Different timescales/frequency ranges
  - Affect data quality, stability and GW detection

- **Noise hunting & characterization is critical**
  - Detect and classify glitches to find their origin and remove them
  - Hardware/software origin
  - Glitches have complex time-frequency morphologies
  - Data from auxiliary sensors important to understand origin
  - Machine learning offers promising approach (e.g. George&Huerta2017, Razzano&Cuoco 2018)

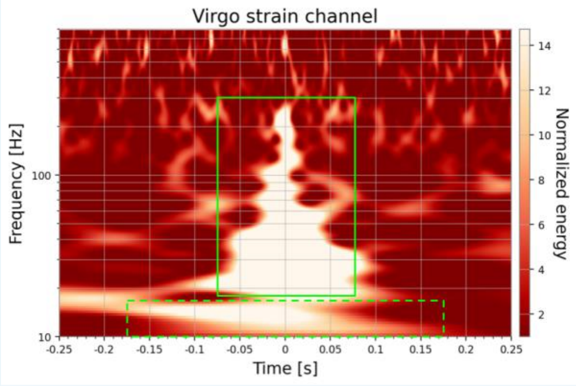


Glitch in LIGO L1 detector during GW170817
Abbott et al 2017

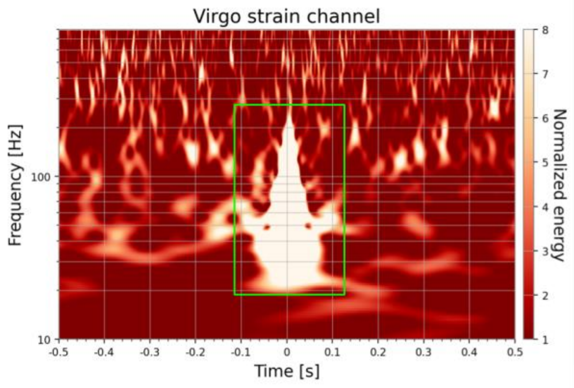
Glitch morphologies sample from Virgo



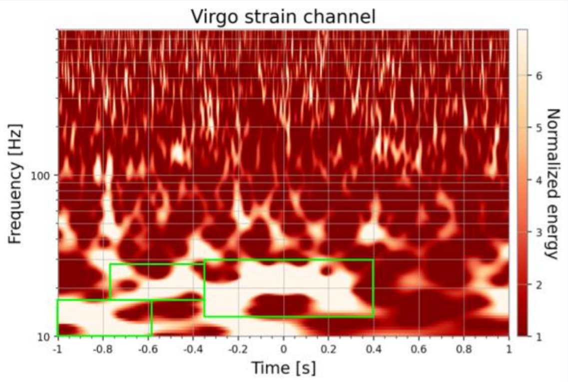
Blip



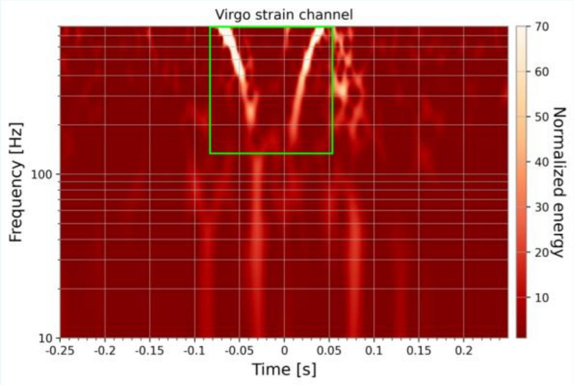
Helix



Koi Fish

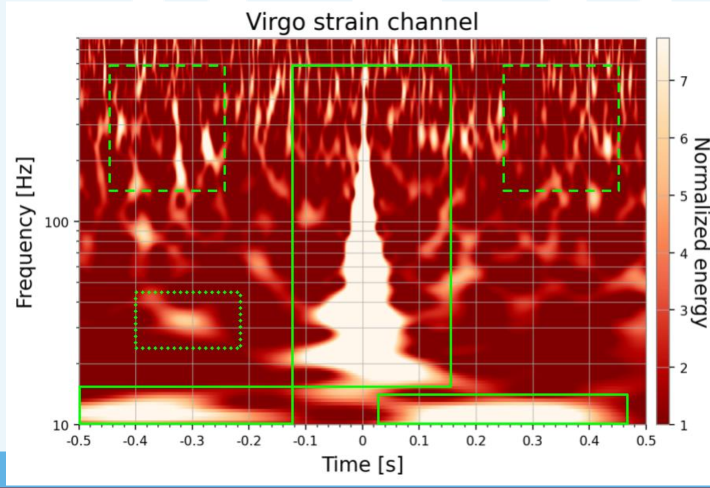


Scattered Light



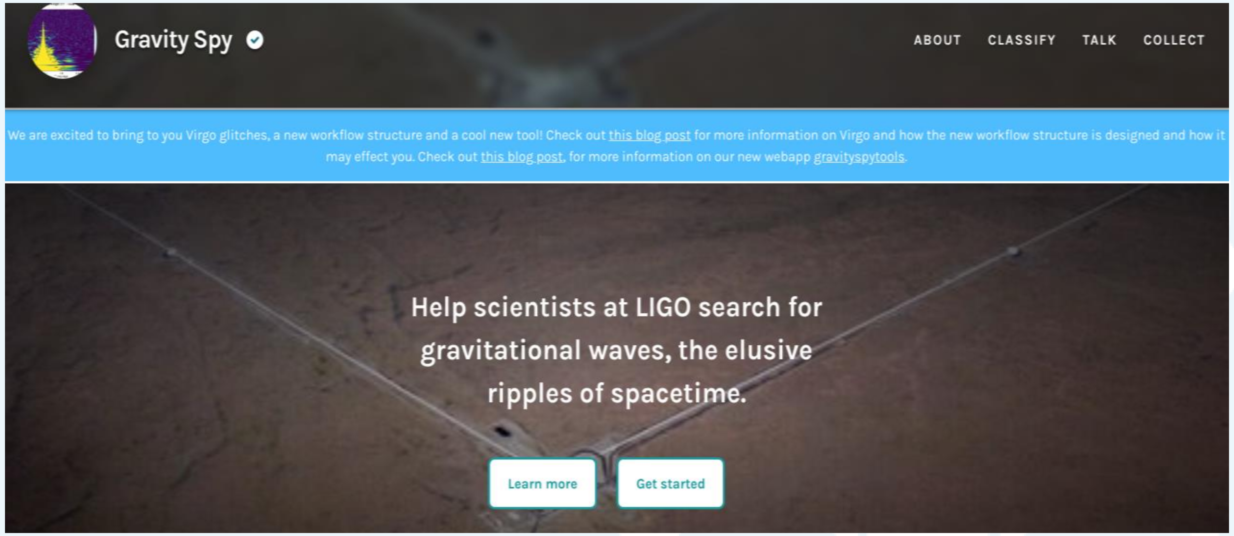
Whistle

Many Glitches



Glitches and citizen science

- **Machine Learning approach**
  - Promising to classify complex time-frequency patterns of glitches (timeseries/images)
  - Large input required to train machine learning models
  - Various approaches (supervised classification, GANs,clustering...)



https://www.zooniverse.org/projects/zooniverse/gravity-spy

- **Citizen scientists can help**
  - Preparing labeled dataset (à la GalaxyZoo)
  - Time-frequency spectrograms (images)
  - Look at glitches & other noise sources
  - and help characterizing them
  - Unveil novel glitch classes
  - Success story: Gravity Spy on Zooniverse (2016) by LIGO team and NSF

The REINFORCE Project

REINFORCE
REsearch INfrastructures FOR Citizens in Europe

- Horizon 2020 SWAFS “Science with and for Society” work program
- Cutting-edge citizen science projects on Frontier Physics Research
- Engage >100k citizens in Large Research Infrastructures in Europe (Virgo, KM3Net, LHC)
- PI S. Katsanevas (EGO)
- Participatory design methodology taking into account the special characteristics of different target groups, their barriers and constraints, their perceptions and biases and their attitudes and knowledge regarding **science**
- Sonification methodologies to increase the senses & inclusion
- Engaging Activities (workshops, schools, etc)
- Started in November 2019

EGO EUROPEAN GRAVITATIONAL OBSERVATORY
ELLINOGERMANIKI AGOGI
UNIVERSITY OF OXFORD

Search for new particles at the LHC DEMONSTRATOR (WP5) Leader
IASA

Gravitational Wave Noise Hunting DEMONSTRATOR (WP3) Leader
UNIVERSITÀ DI PISA

Deep Sea Hunters (WP4) & Interdisciplinary studies with Geoscience and Archaeology (WP6) DEMONSTRATORS Leader
cnrs

Impact Assessment (WP9) Leader
ZSI

Increasing Inclusion (WP7) Leader
CONICET

Sustainability and Exploitation (WP10) Leader
the Lisbon council
think tank for the 21st century

Coordination with other relevant SwafS projects Leader
The Open University

Communication and Dissemination Leader
Trust-IT Services
Communicating ICT to markets

REINFORCE and GW

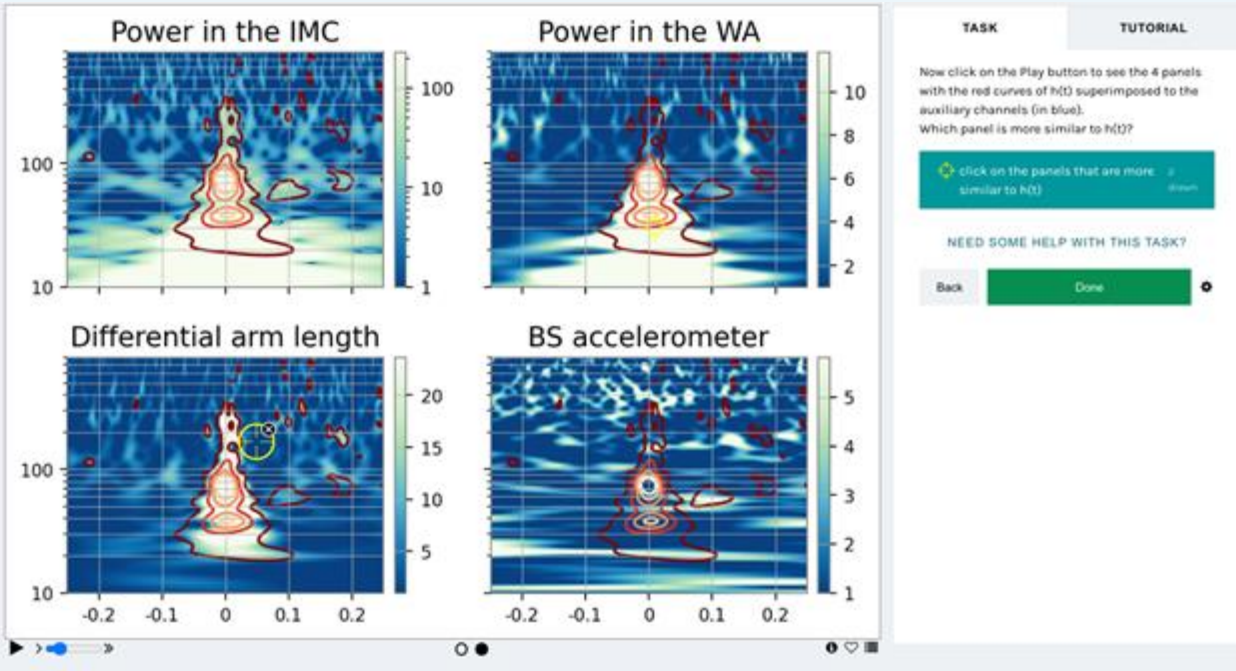
- Citizen science is already supporting gravitational wave science (GravitySpy)
- REINFORCE focus on 4 projects (aka "demonstrators")

https://www.reinforceeu.eu/

Work Package 3
Institutions:
Unipi, EGO,
EA,Conicet,
OU,UXF

- Demonstrators developed on Zooniverse, the leading platform for citizen science
- Our goal: engaging citizens as active participants
- Output used to train machine learning models, but not only that

Our Challenges



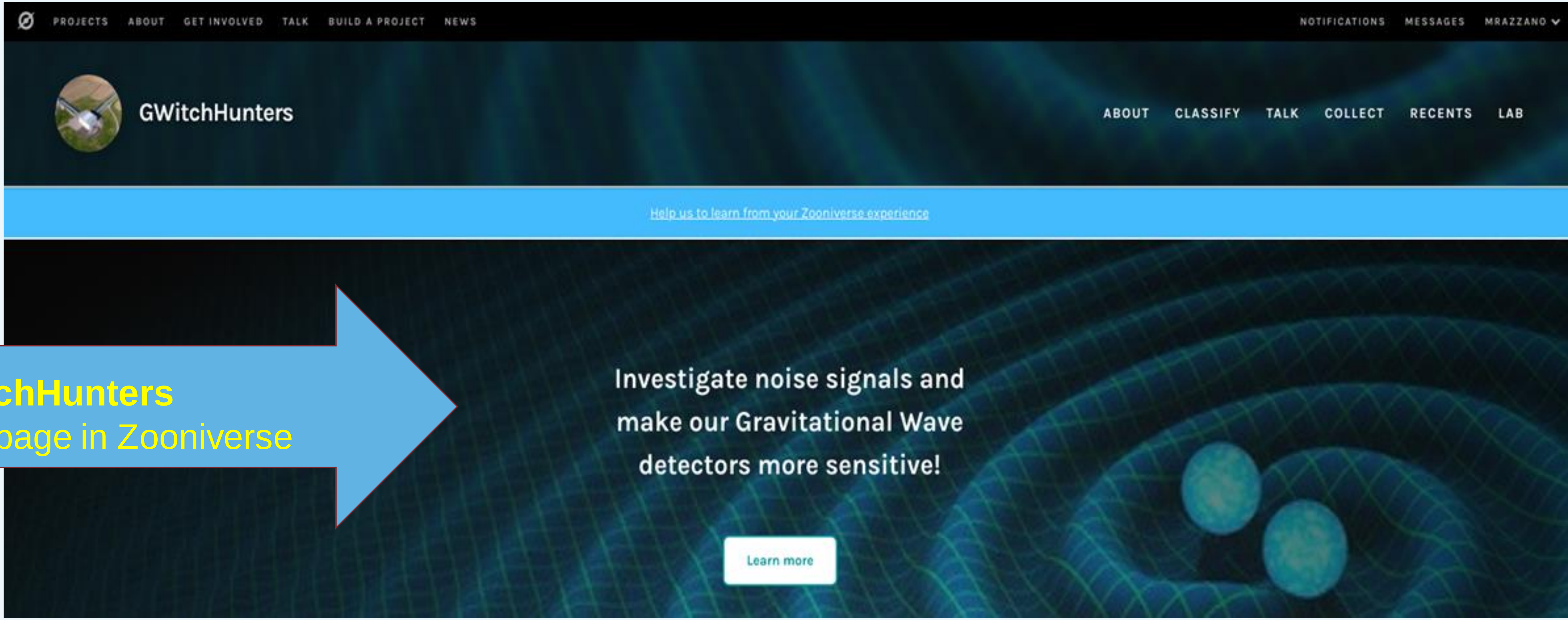
- **Be engaging and innovative**
  - **Attract** GravitySpy fans and more
  - Offer **new challenges** in GW science

- **More data, more fun!**
  - Auxiliary channels offer new insights into **how detectors work**
  - Auxiliary channels are not public. Prepared **ad hoc REINFORCE-EGO MoA**
  - As a result, an **updated & more rich dataset** than first dataset

- **Go beyond glitches**
  - **Flexible workflow**, can accomodate other noise features (lines/slowly-varying noise)

GW demonstrator development

GWitchHunters
homepage in Zooniverse



10

Highlights of GWitchHunters

- Introduce a **new, original** way to power GW research with citizen science
  - *GravitySpy* as a success story
  - Can we expand this approach?

- New frontiers
  - Go beyond classification tasks → **Noise hunting**
  - Not only **glitches**
  - Include signals from sensors in the detector (“**auxiliary channels**”)
  - Run on mobile **devices**

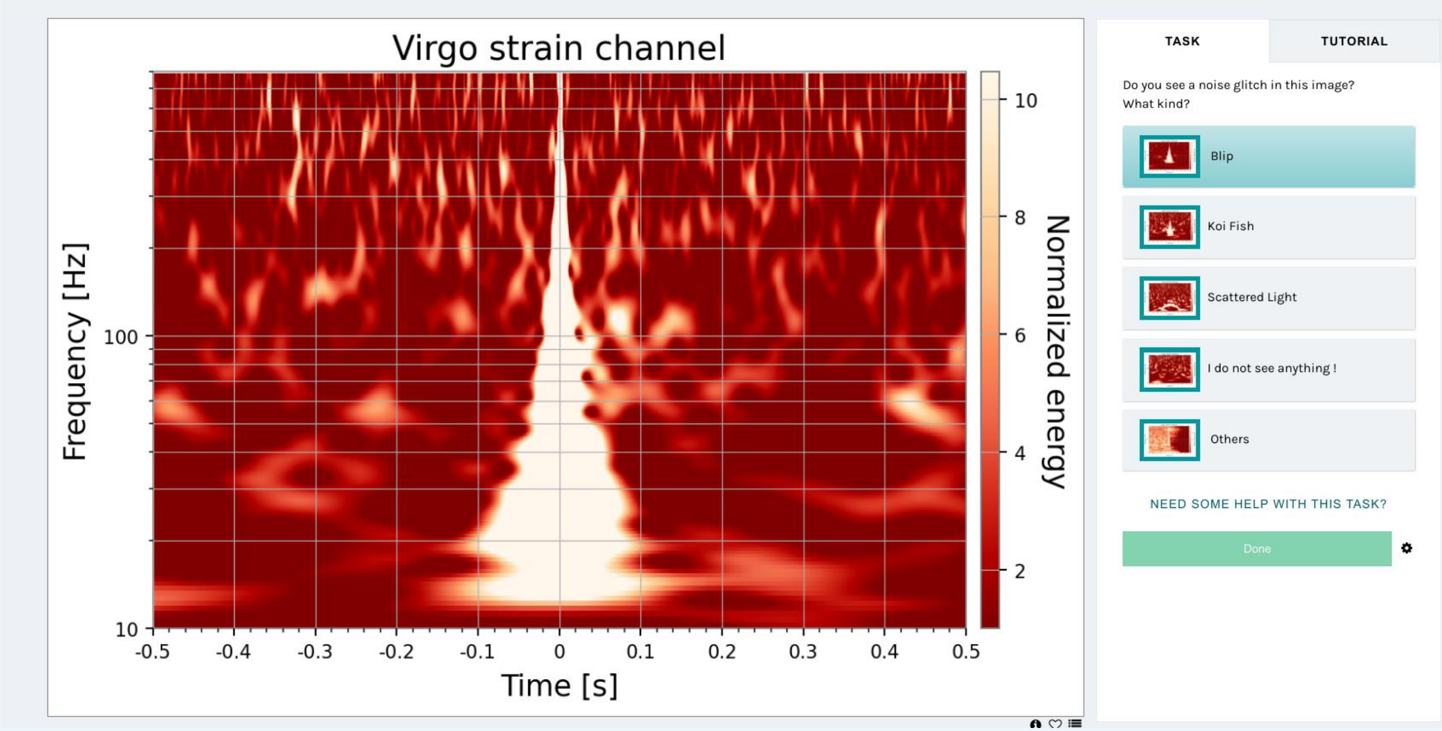
GWitchHunters Levels

Desktop → Playground - What is a glitch? | Level 1 - Catch the noise! | Level 2 - Find them all! | Level 3 - Get the auxiliary channels

Mobile → Mobile Challenge - Noise Profilers | Mobile Challenge - Lasso that glitch!

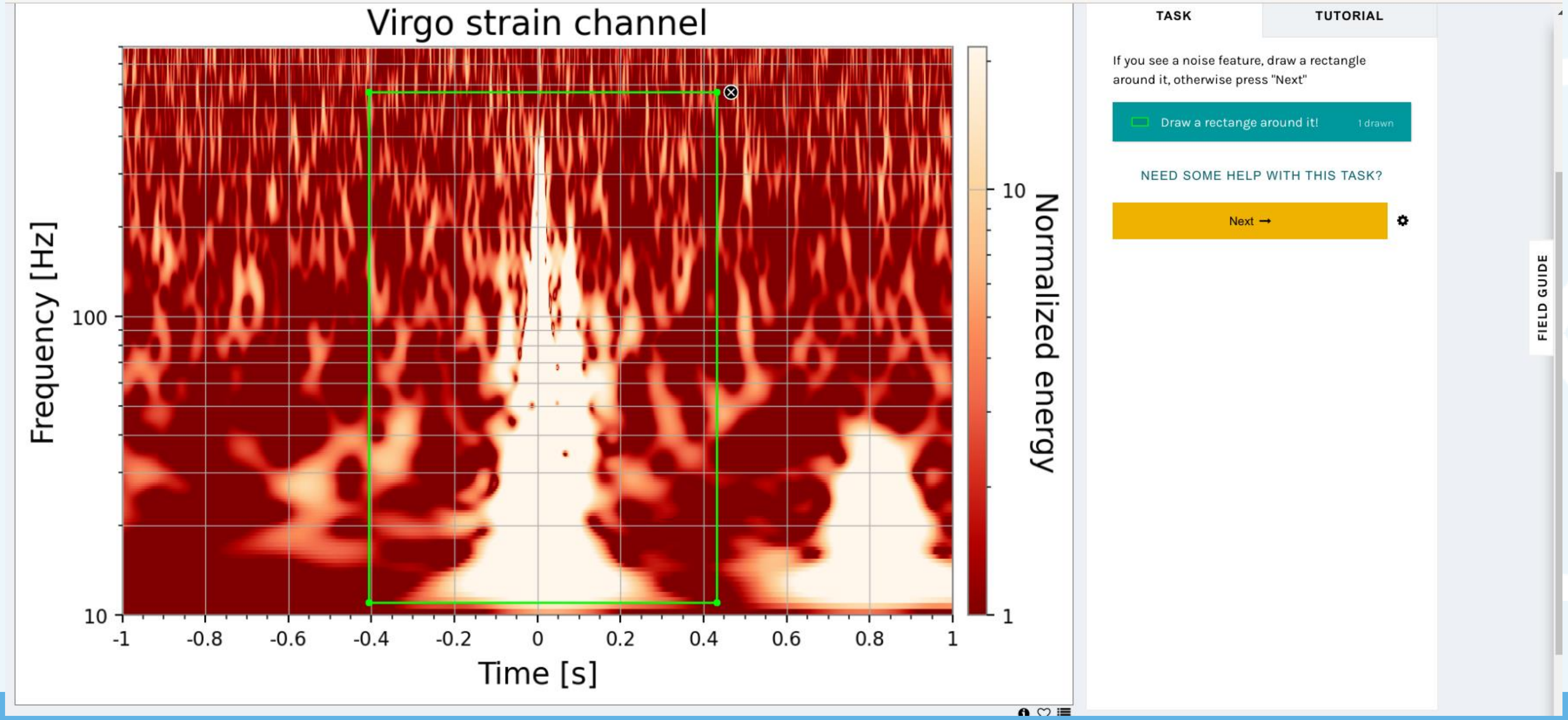
Citizens Tasks - Classification

- **Classify noise features**
  - Similar to GravitySpy
  - Data presented as time-frequency spectrograms
  - Meant mainly to introduce participants to the problem and train them



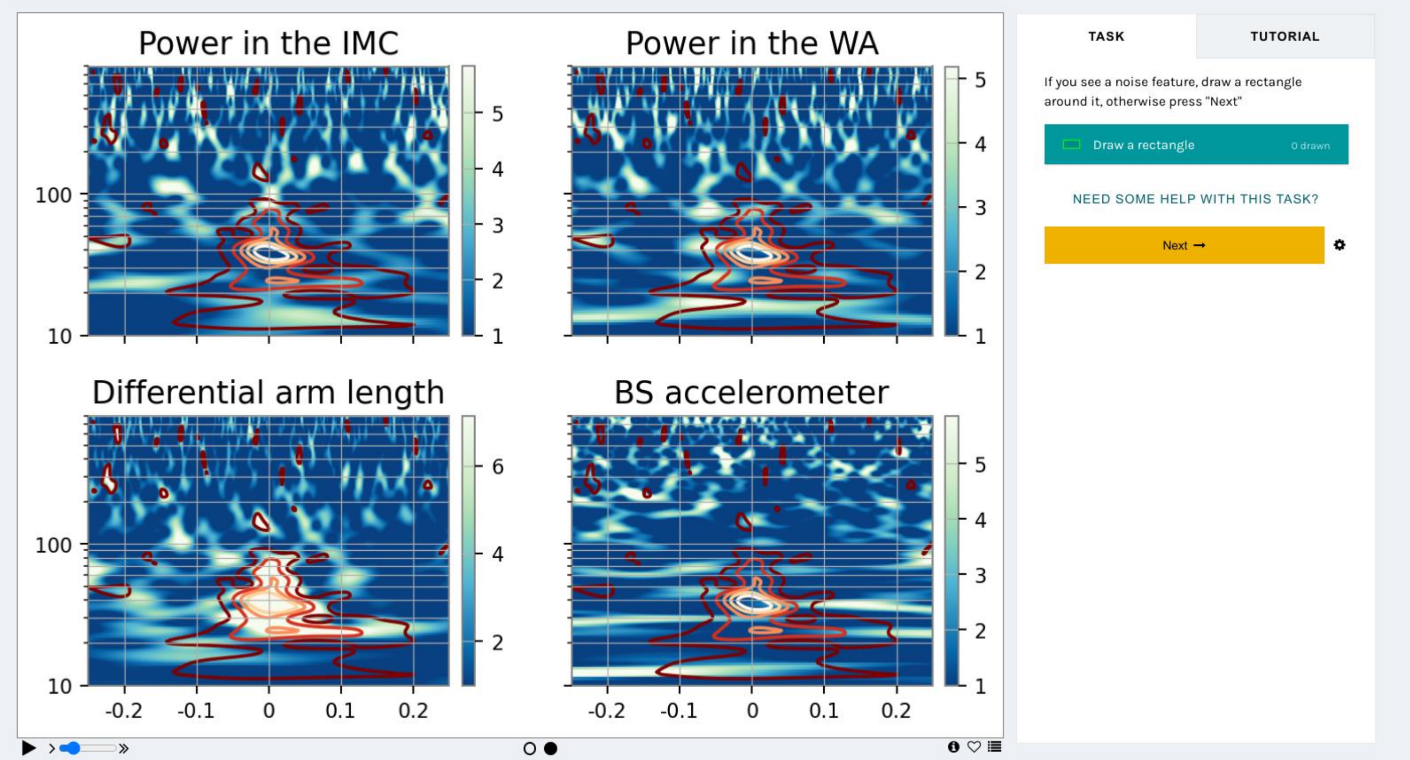
Citizens Tasks - Detection

- **Localize the noise**
  - **Draw rectangles around noise features**
  - Can be extended to non-glitch features



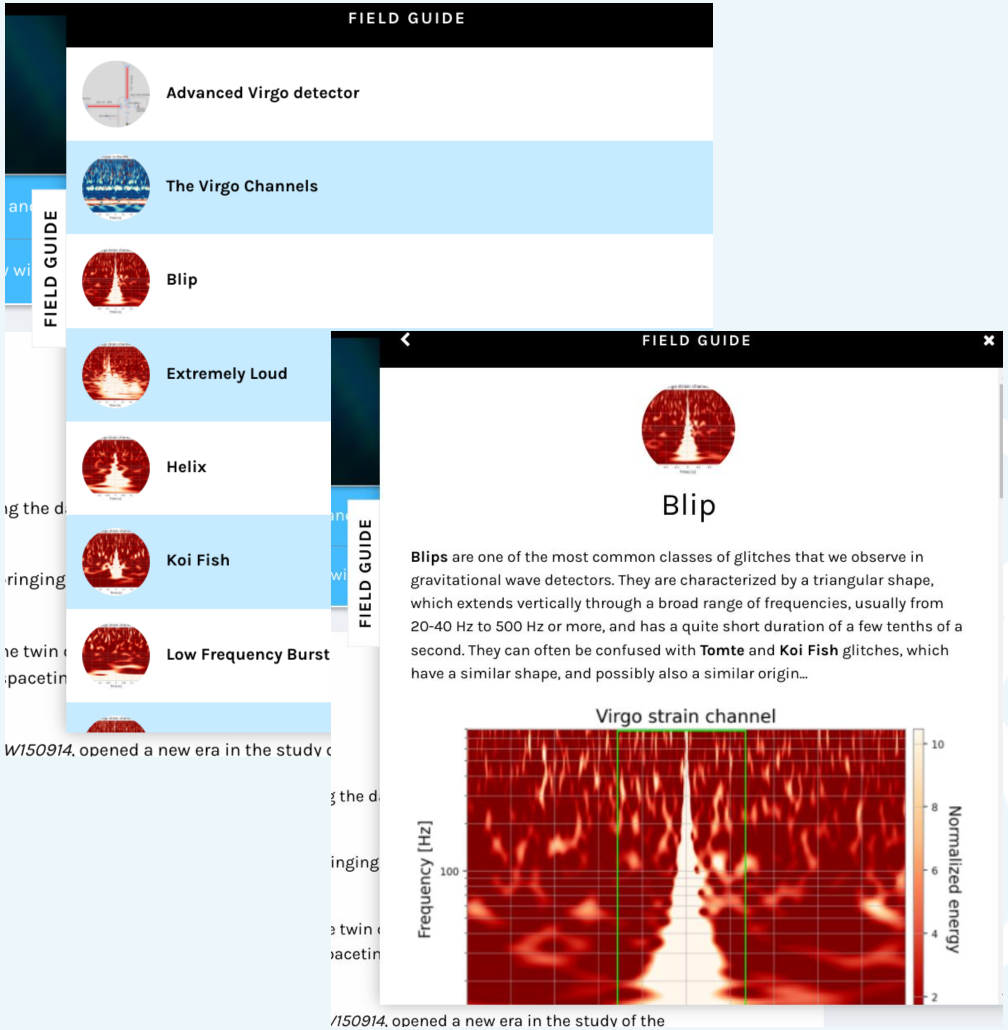
Citizens Tasks – Auxiliary Channels

- **Study data from auxiliary sensors**
  - Find similar morphologies between h(t) and aux
  - Initial set of 8 channels, discuss with Virgo experts on others to add



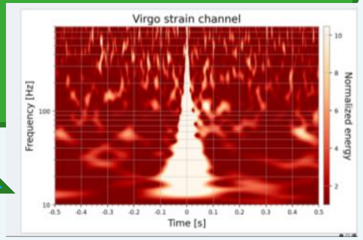
Tutorials and accessibility

- **Introduce the project to non scientists**
  - Detailed tutorials
  - Field Guide on GW detectors, Virgo, glitches
- **Promote engangement**
  - Implement mobile tasks using Zooniverse app
  - Forum to discuss and interact with research team and among citizens
  - Multilanguage support
- **Self-training**
  - Playground level with real-time feedback



GWitchHunters: The big picture

- Glitch triggers
- Noise features



Virgo
Glitch Database

GWitchHunters
(labeling and characterization)

Process & ML training

Classes, aux channels
(i.e., Virgo detchar input)

Where we are

Launched in Nov 2021, very good feedback from the citizens community

Some numbers:

GWITCHHUNTERS STATISTICS

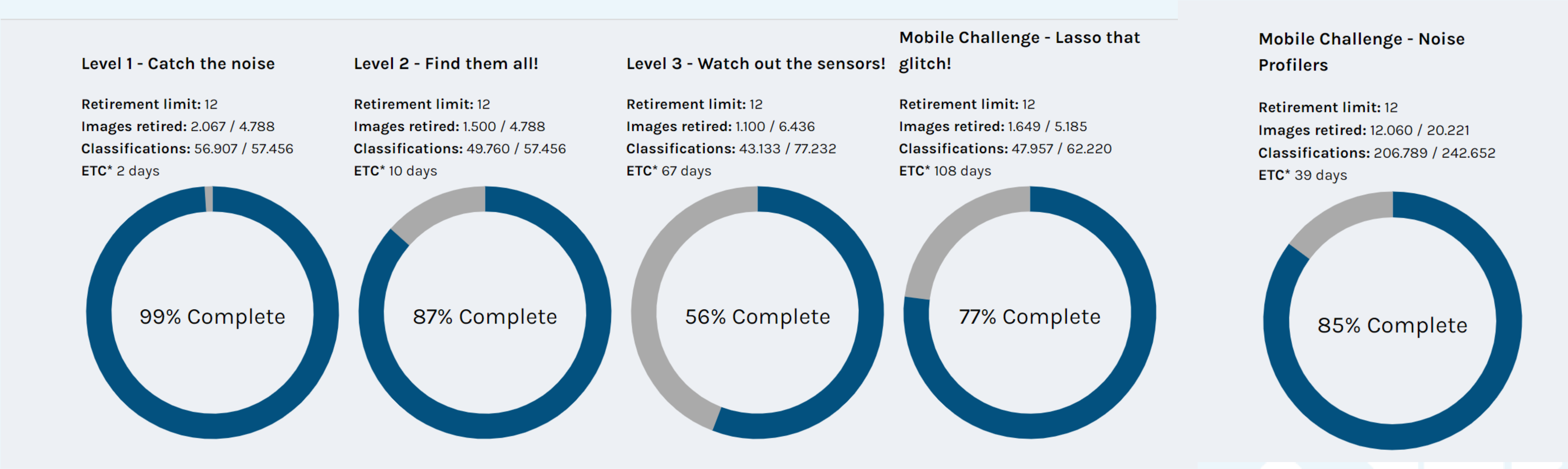
30% Complete

3.328	561.451	41.620	18.376
Volunteers	Classifications	Subjects	Completed Subjects

Mobile Challenges done quickly: first data release completed in mid Feb

Status of the tasks

Some tasks are in a very good stage

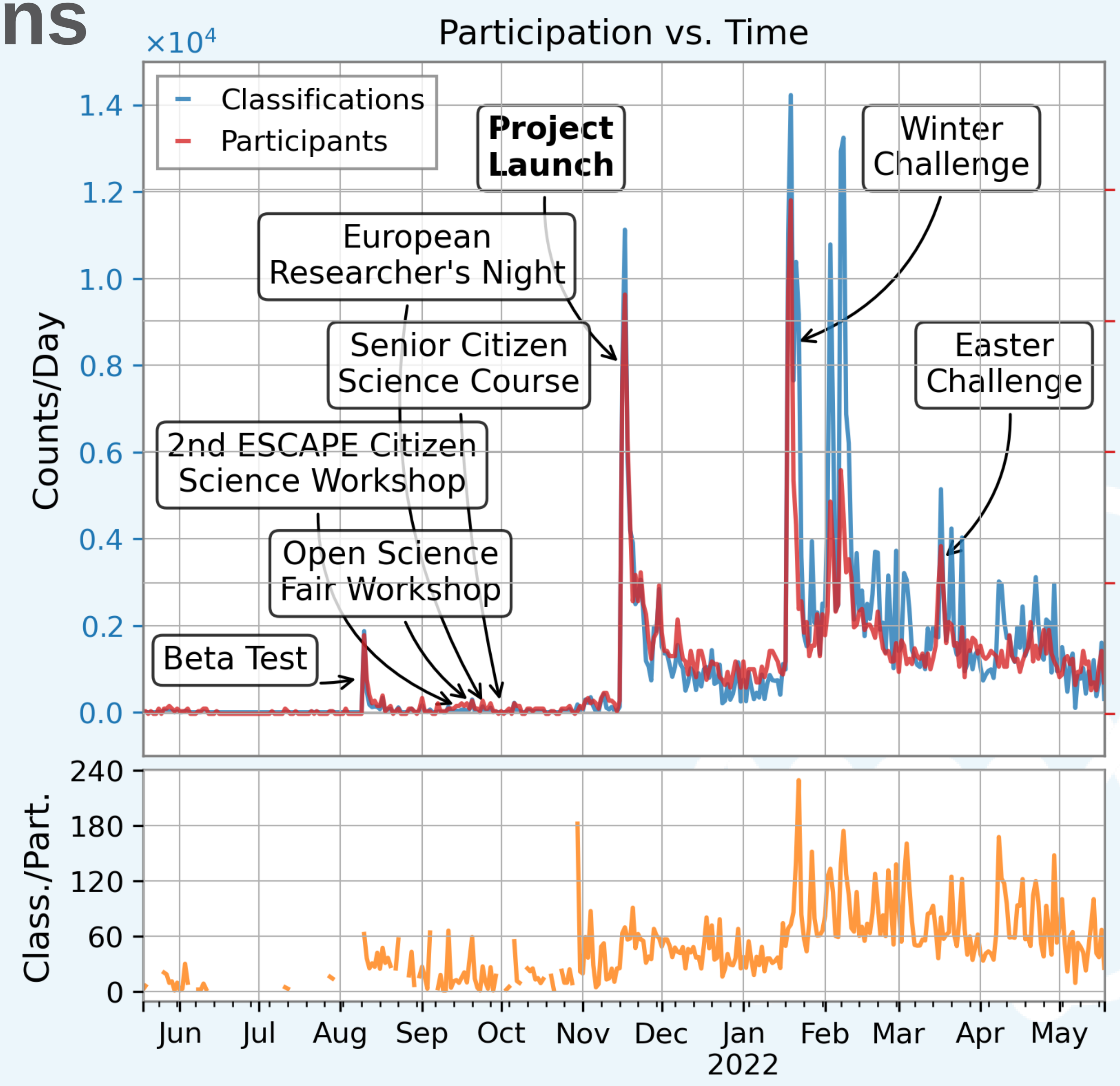


https://www.zooniverse.org/projects/reinforce/gwitchhunters

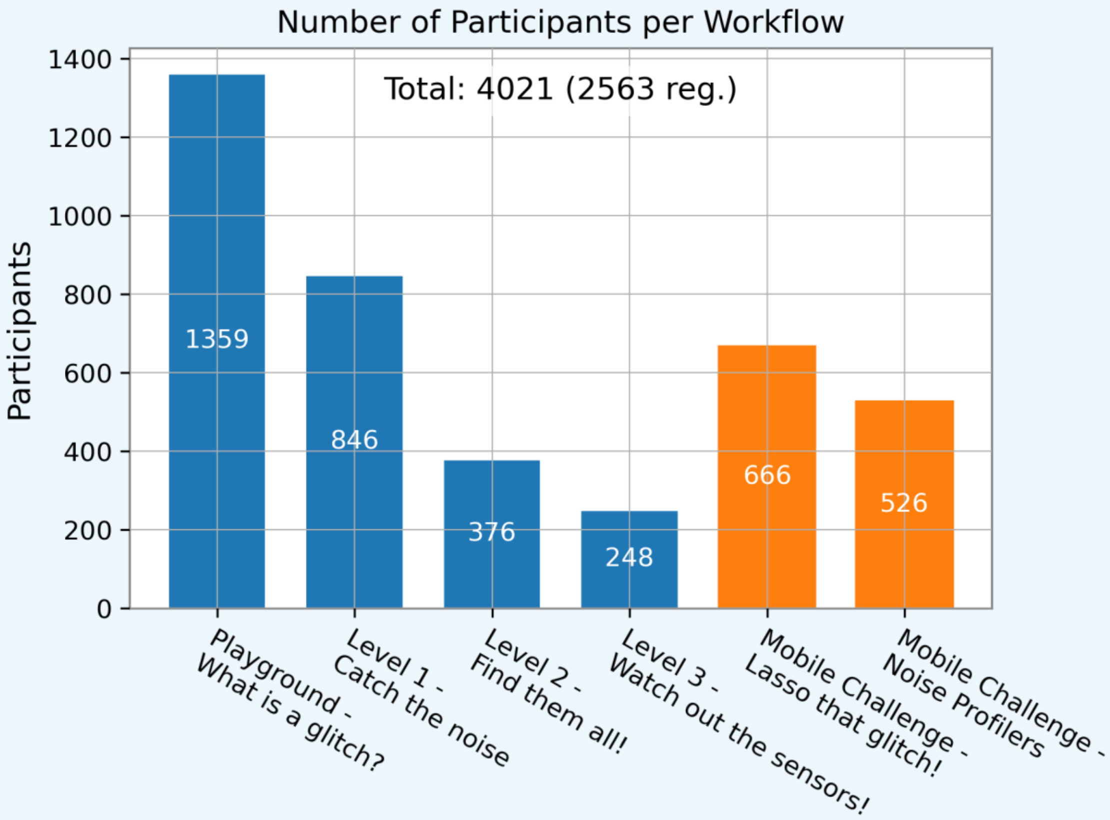
Engaging citizens

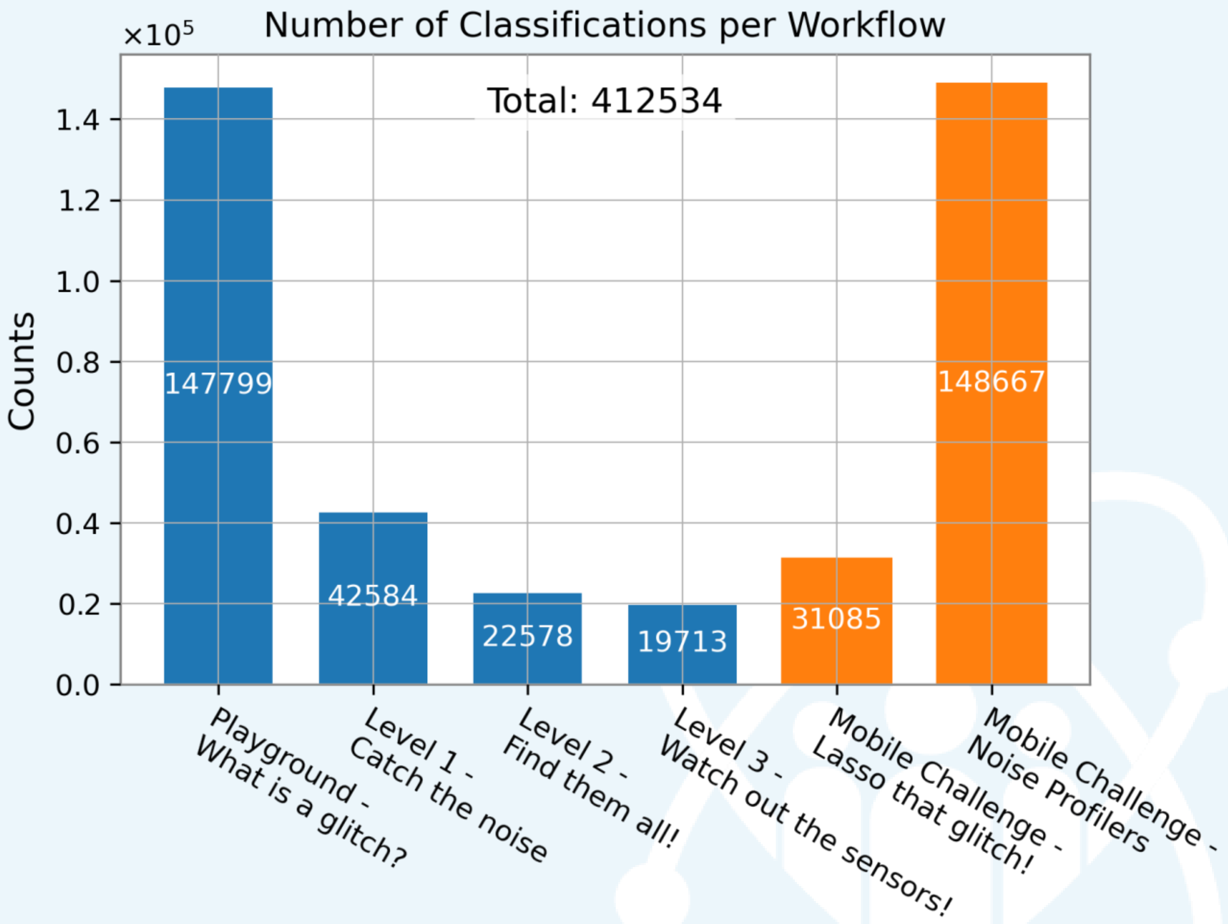
Many initiatives to promote the project

- Winter/Easter challenges
- Dedicated Workshops
- Events
- Courses



Processed data





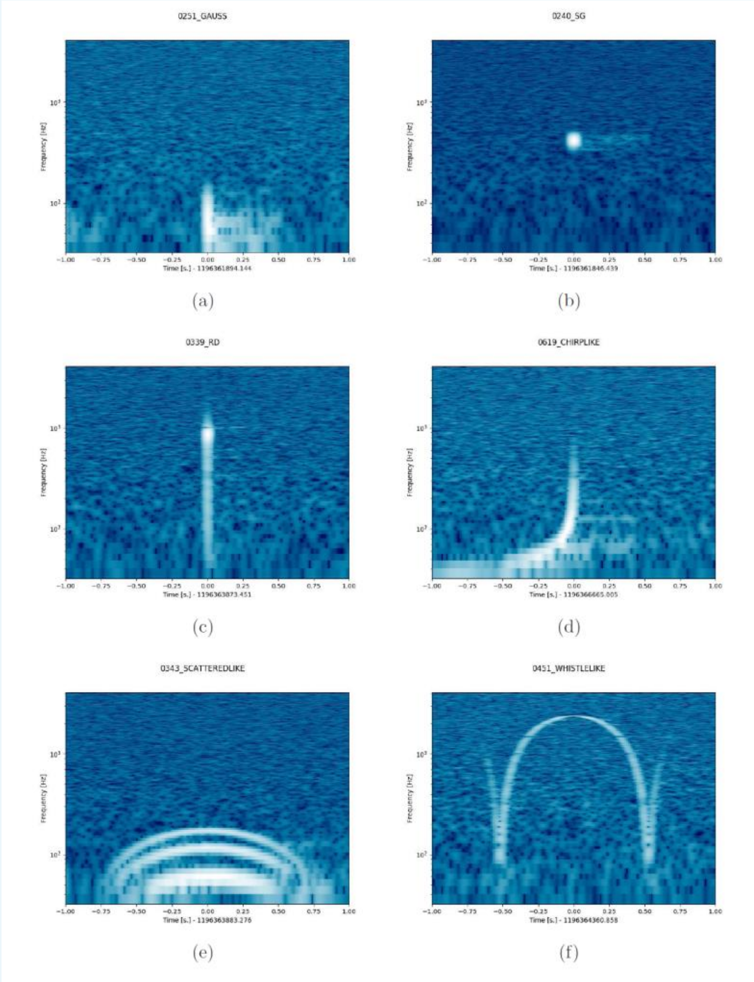
Machine Learning processing

CNN for automatic classification

- First step of a multi-stage pipeline (including other tasks)

Checks and development on simulations

- Experience from previous works
- Deep, multi-layer structure
- Developed in Python + standard libraries (tensorflow,Keras)
- Run on GPUs



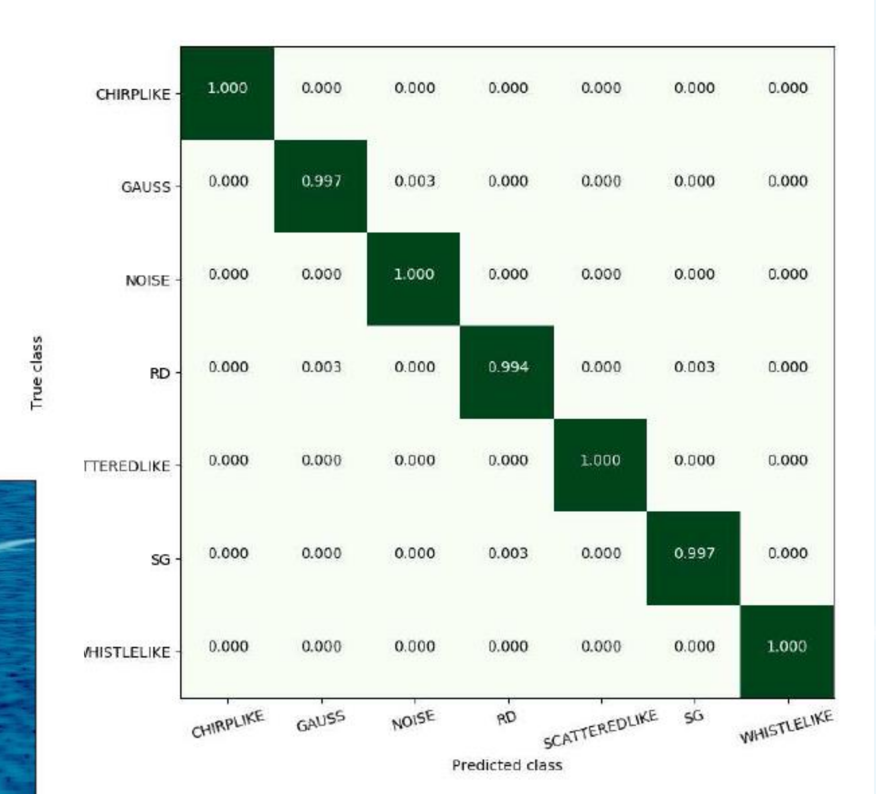
Razzano&Cuoco, 2018

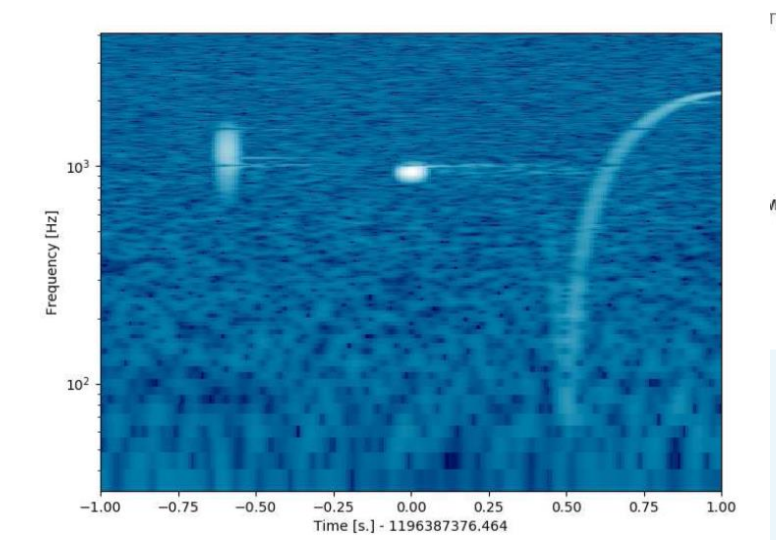
21

Machine Learning processing

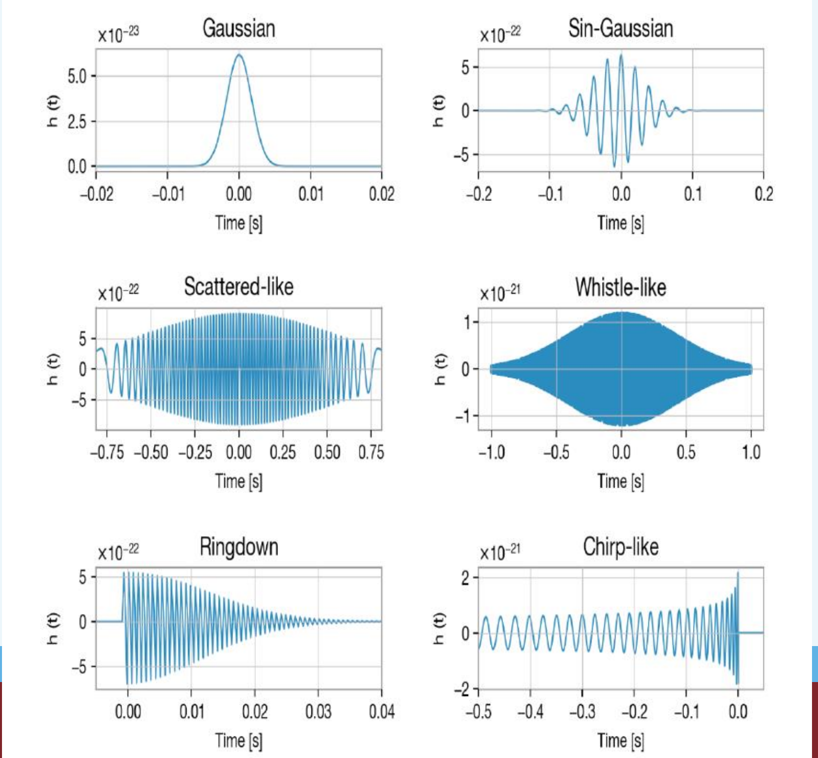
2D vs 1D

- 2D good results, both on simulations and previously labeled real data
- Convolutions robust for multi-glitch situations
- Image-building (little more) time consuming, tried 1D CNN





Razzano&Cuoco, 2018



Talpini&Razzano, 2022

Preparing the analysis: Citizens scientists performance

- Putative class assigned on majority vote
- Results good for some classes, for other requires more data

Confusion Matrix of Level 3 - Watch out the sensors!

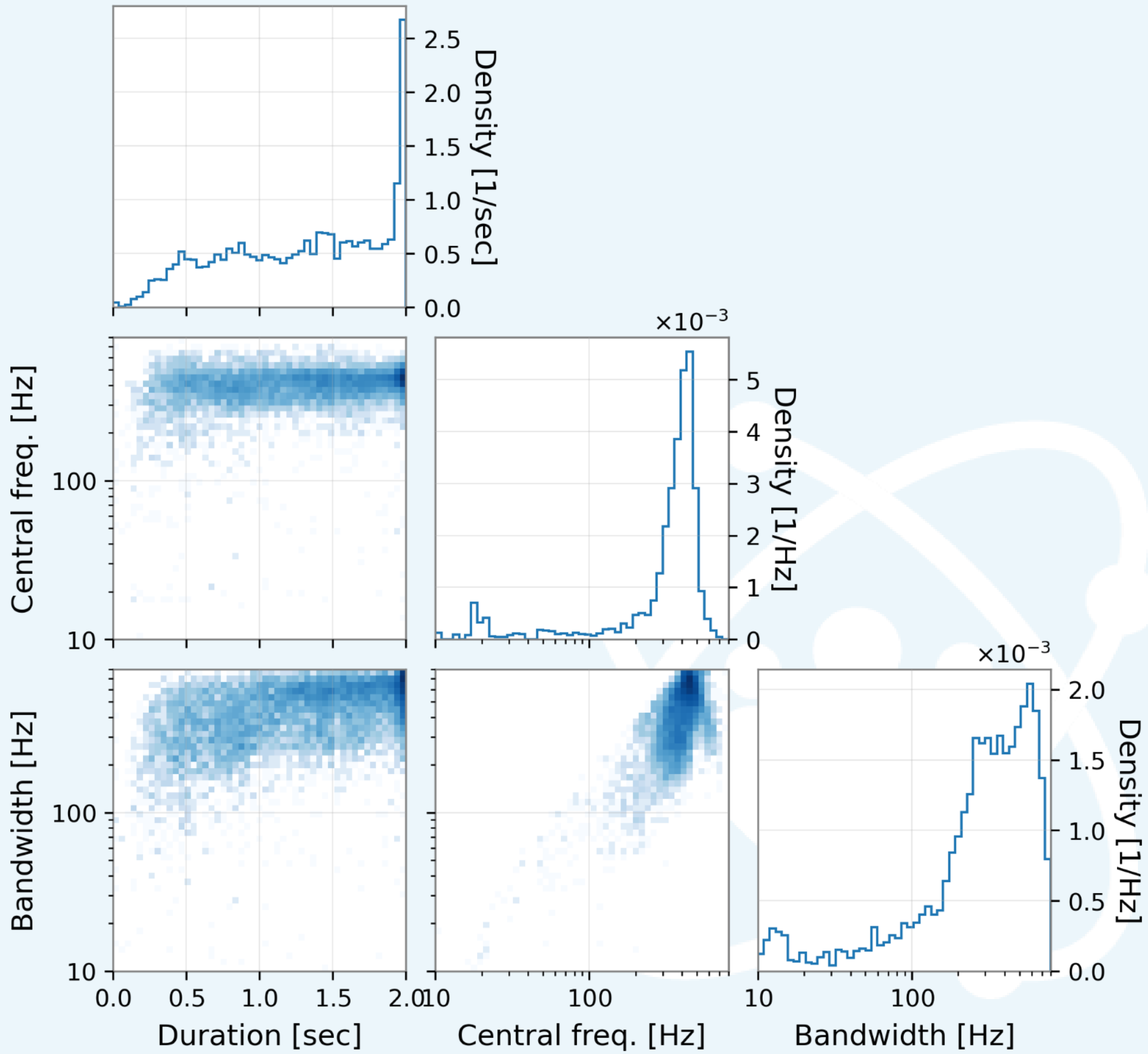
	Others	Blip	Extremely Loud	Helix	Koi Fish	Low Frequency Burst	Scattered Light	Scratchy	Tomte	Violin Mode
Others	56%	6%	4%	4%	2%	3%	14%	6%	3%	2%
Blip	7%	63%	4%	10%	8%	0%	0%	0%	7%	0%
Extremely Loud	6%	12%	65%	3%	12%	0%	0%	0%	1%	0%
Helix	9%	16%	2%	49%	6%	1%	1%	0%	16%	0%
Koi Fish	3%	12%	12%	4%	66%	0%	0%	0%	3%	0%
Low Frequency Burst	14%	0%	3%	1%	0%	61%	20%	0%	0%	1%
Scattered Light	8%	0%	1%	1%	0%	7%	80%	0%	1%	0%
Scratchy	13%	0%	0%	0%	0%	2%	3%	71%	0%	12%
Tomte	5%	6%	1%	15%	5%	1%	2%	0%	66%	0%
Violin Mode	5%	0%	0%	0%	0%	0%	0%	14%	0%	81%

Data from May 18,2021 to Jul 15 2022

Citizens performance (II)

- Useful to find glitch parameter distribution
- Better discrimination among glitch classes

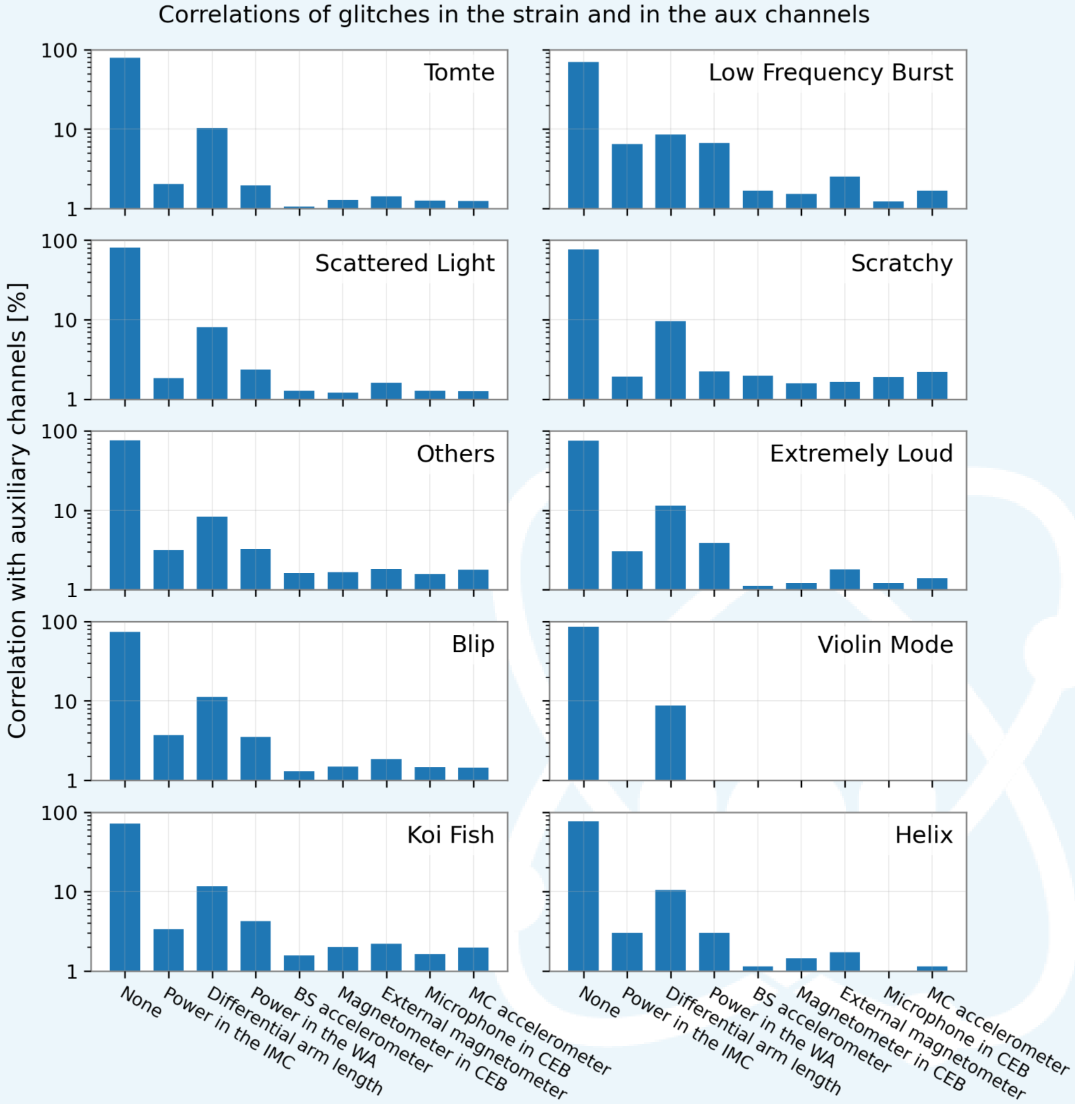
Distributions of Scattered Light parameters



Data from May 18,2021 to Jul 15 2022

Citizens performance (III)

- Only 8 channels, many info already
- DARM correlation is a known fact
- Useful to show how glitch with similar morphology correlate with different aux channels (e.g. Scattered Light and Low Frequency Bursts)



Data from May 18,2021 to Jul 15 2022

The road ahead

- **GWitchHunters project on Zooniverse**
  - Successfully launched: we keep monitoring the data inflow
  - New datasets included, more challenges to come

- **Data Analysis**
  - We are gathering data and new datasets will be prepared
  - Validating and assessing citizens performance (comparison CNN/human)
  - Standard CNN (2D/1D) producing good results
  - Working on automatically processing pipeline
  - Not only classifications, exploring regression and localization

- **Next steps**
  - Add new glitch datasets & new auxiliary channels
  - Offer novel challenges for data exploration & noise hunting

Stay Tuned
(& classify glitches!)

Join the community

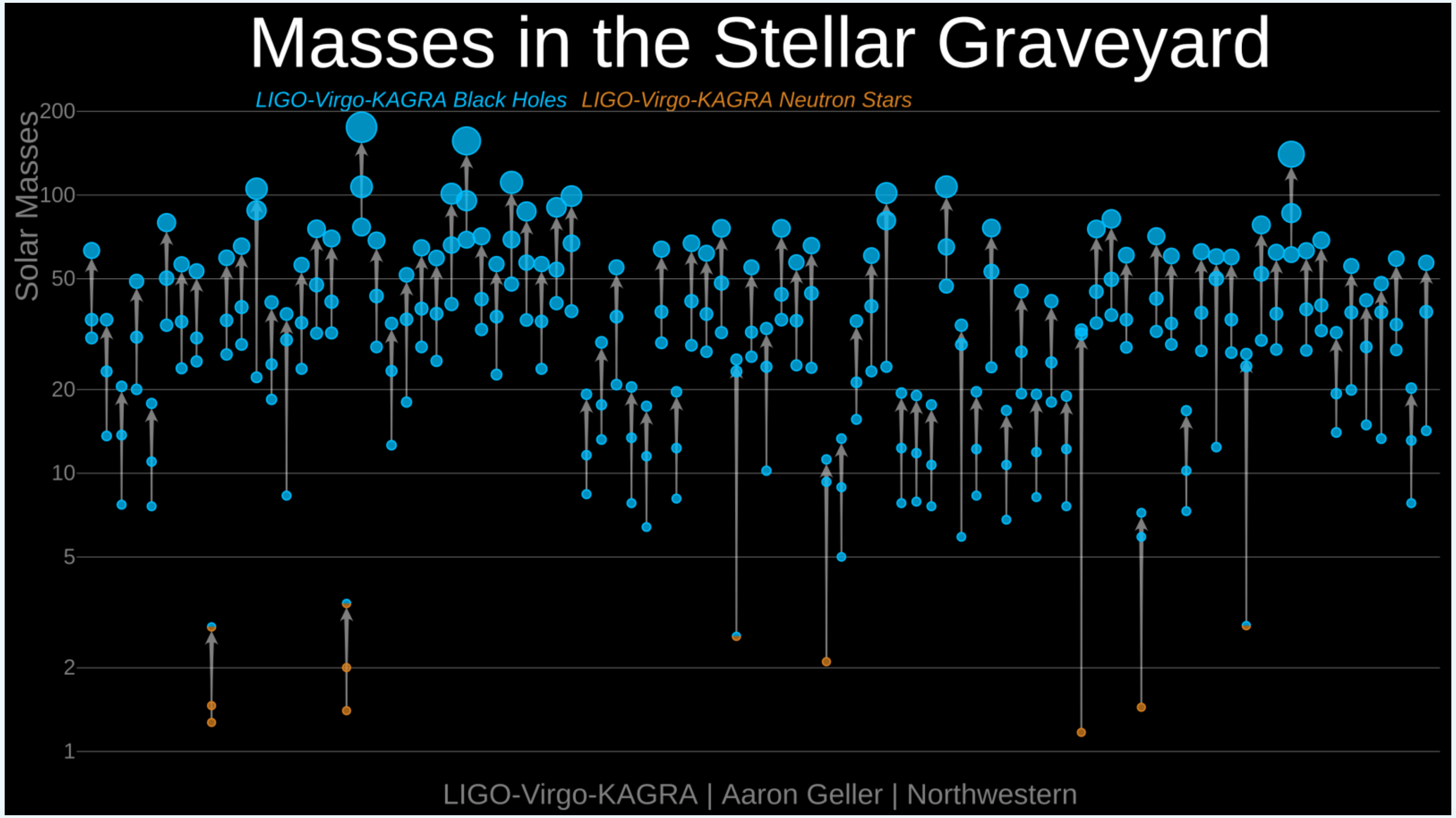
Reinforceeu.eu

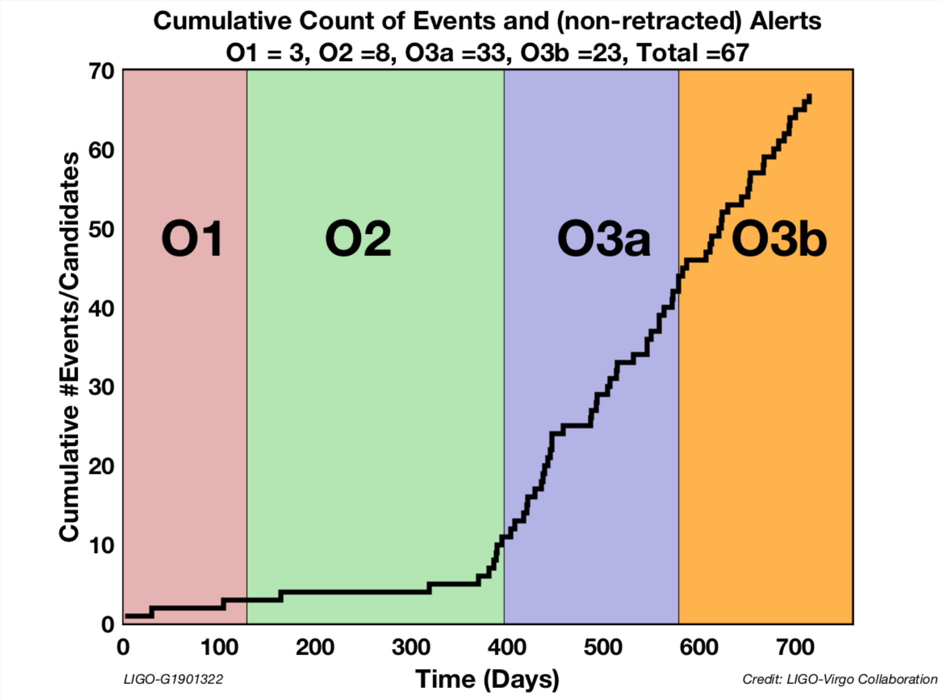
/company/reinforceeu

@ReinforceEU

The era of Gravitational Waves

O4 expected to start in December 2022





GW Transient Catalogs

GWTC-1: Abbott et al 2019, Physical Review X, 9, 3

GWTC-2: Abbott et al 2021, Physical Review X, 11, 2

GWTC-2.1: Abbott et al 2021, arXiv: 2108.01045

GTWC-3: Abbott et al 2022, arXiv:2111.03606

Few words on data

- **Source data**
  - Real data from Advanced Virgo O3 (for the launch, next steps we plan to include LIGO)
  - Main "strain" channel (aka h(t)), also publicly available
  - Auxiliary channels, not public but available thanks to Virgo-EGO-REINFORCE MoA
- **Datasets**
  - Timeseries (+ preprocessing, whitening) to Time-Frequency image maps
  - Transitioning from "pathfinder" to "production" phase: 2 Data Releases so far
    - DR01 (2021, launch) - 2000 glitches from O3a + 8 aux channels
    - DR02 (2022, follow-up) - +4000 glitches from O3b + 8 aux channels