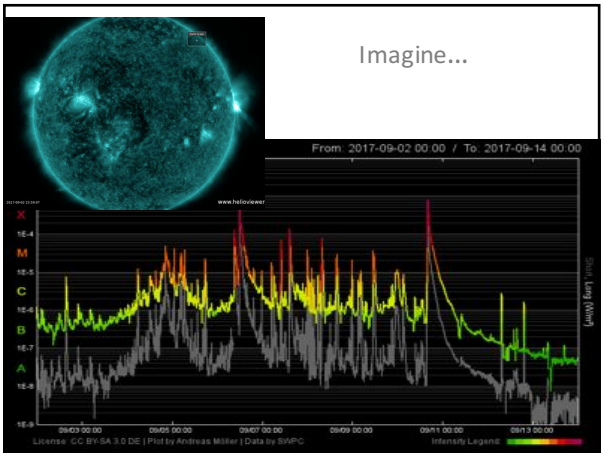
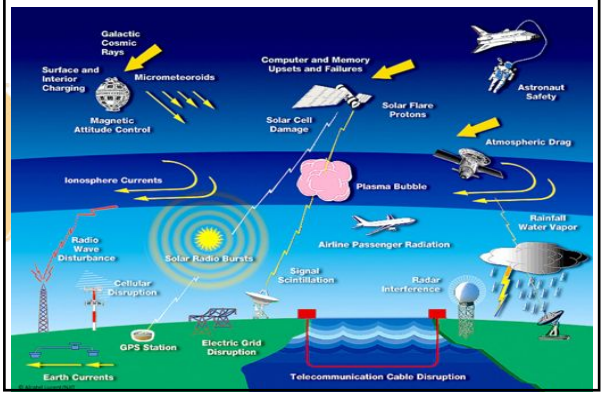
FLARECAST: Learning from the past to predict the future

André Csillaghy for the FLARECAST Team

Imagine…



Why automating solar flare predictions?

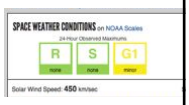


FLARECAST HIGH-LEVEL OBJECTIVES

- Develop a solar flare prediction system based on automatically extracted physical properties of solar active regions, coupled with state-of-the-art solar flare prediction methods and validated using the most appropriate forecast verification measures.
- FLARECAST top-level objectives:
  - To understand the drivers of solar flare activity and improve flare prediction
  - To provide a globally accessible flare prediction service that facilitates expansion
  - To engage with space weather end users and inform policy makers and the public

FLARECAST IN SIMPLE TERMS

- What is the data trying to tell us about flare prediction?
- Input: SDO HMI images
- Output: a flare prediction of the kind:
  - Binary forecasting: Flare or No Flare
  - Probabilistic: 0 <p< 1

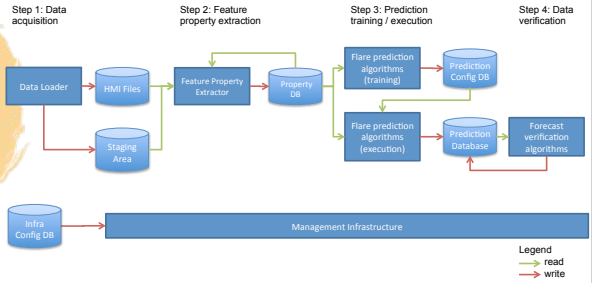


- For the following characteristics
  - Within a flare class (e.g. M1 – M9.9)
  - Within a forecast time window (e.g. 24 hours) & above a threshold
- Steps:
  Feature property extraction → prediction learning → forecast verification
- Support: Infrastructure

THE FLARECAST SCENE

- H2020 Project 2015 – 31.12.2017
- Partners: Academy of Athens (Georgoulis, PI), Northumbria U. (Bloomfield Project Scientist), U. Genova (Piana), CNR (Massone), CNRS (Vilmer), U. Paris Sud (Buchlin), FHNW (me ☺), Met Office (Jackson)
- A diverse group of ~50 scientists and engineers working together
- Mix of expertise in flare prediction (AA, TCD, UN), mathematics (UGE, CNR), Computer Science (UPS, FHNW), and user perspective (Met Office)

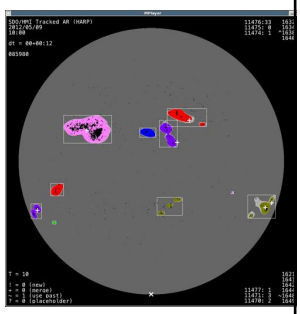
FLARECAST WORKFLOW



FLARECAST INPUT DATA

SDO / HMI data
- LOS magnetograms (hmi.M_720s)
- SHARP vector magnetograms – definitive (hmi.sharp_720s)
- SHARP vector magnetograms – NRT (hmi.sharp_720s_nrt)

SRS active region & flare data (YYYY_events.tar.gz)



Feature Properties

Mix of information already available and newly generated

Link

Property Extraction Algorithms

Data Source	Property Group	Developer	Status
SWPC catalogues			(To do / In progress / Under testing / Delivered)
	Solar Region Summary properties	TCD	Delivered
Details	GOES X-ray events	TCD	Delivered
LOS magnetograms			
	SMART-derived properties (Ahmed et al., 2013)	TCD	In progress
	SMART- delta finder (Padirkattanet et al., 2015)	TCD	To do
Details	Effective connected magnetic field strength (B_{eff}) (Georgoulis & Rust, 2007)	AA	Delivered
Details	Fractal dimension (Georgoulis, 2012)	AA	Delivered
Details	Multi-fractal structure function $s(q)$ inertial range index k (Georgoulis, 2012)	AA	Delivered
Details	Fourier power spectral index (Guerra et. al., 2015)	TCD	Delivered
Details	CWT power spectral index (Hewett et. al., 2008)	TCD	Delivered
Details	Generalized correlation dimension (Georgoulis, 2012)	AA	Delivered
Details	Hölder exponent h (Conlon et al., 2010)	AA	In progress
Details	Hausdorff dimension D(h) (Conlon et al., 2010)	AA	In progress
	WTMM (Conlon et al., 2010)	TCD	Under testing (further investigated in WP6)
Details	Decay index (Zuccarello et al, 2014)	TCD	Delivered
Details	Magnetic polarity inversion line characteristics (Mason & Hoeksema 2010)	TCD	Delivered
Details	3D magnetic null point (Greene 1988)	TCD	Delivered
Details	R (Schrijver 2007)	TCD	Delivered
Details	$^{total}_{abs}$ (Falconer et al. 2008) *	TCD	Delivered
Details	Ising energy (Ahmed et al. 2010)	AA	Delivered
Details	WG_{sh} and S_{sh} (Korsos et al. 2015)	AA	Delivered
Details	Magnetic helicity injection rate proxy (Park et al. 2013)	TCD	Delivered
Vector magnetograms			
Details	SHARP properties (Bobra et al. 2014)	TCD	Delivered
Details	Magnetic helicity injection rate (Berger & Field 1984)	TCD	Delivered
Details	Magnetic energy injection rate (Kusano et al. 2003)	TCD	Delivered
Details	Non-neutralized currents (Georgoulis et al., 2012)	AA	Delivered
Details	Flow field characteristics (Deng et al. 2006; Wang et al. 2014)	TCD	Delivered
	Magnetic bipolar feature characteristics	TCD	Under testing (further investigated in WP6)
Intensity images			
	Flow field characteristics	TCD	Under testing (further investigated in WP6)

PREDICTION ALGORITHMS

- Non ML
  - Linear Discriminant Analysis
  - Bayesian Quantile Regression
- Standard ML
  - Clustering and Regression Analysis
  - Simple Recurrent Neural Networks
- Advanced ML
  - Multi-Layer Perceptron
  - Possibilistic C-Means

Link

- Innovative ML
  - Multi-Task Lasso
  - Poisson Re-Weighted Multi Task Lasso
  - Hybrid Method
  - Simulated Annealing
  - Reccurent Neural Network trained with an evolutionary algorithm
  - Random Forest

FORECAST VALIDATION via statistical assessment

	Forecasted Flare	Forecasted No-Flare
Observed Flare	True Positive	False Negative
Observed No-Flare	False Positive	True Negative

Probability of detection: $POD = \frac{TP}{TP + FN}$

False Alarm Rate $FAR = \frac{FP}{FP + TN}$

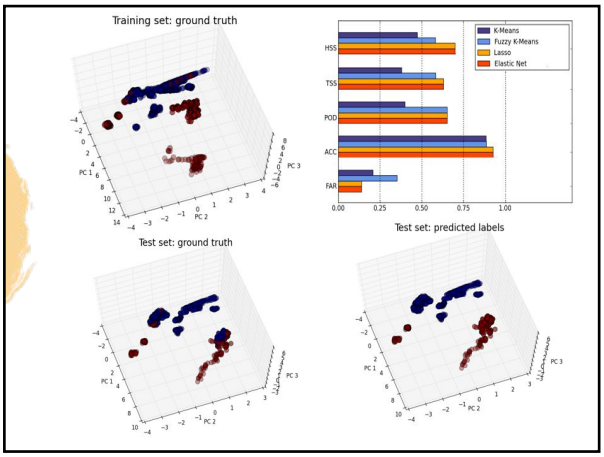
Accuracy $ACC = \frac{TP + TN}{TP + TN + FP + FN}$

Heidke Skill Score: $HSS = \frac{2(TP + TN) - N}{N}$

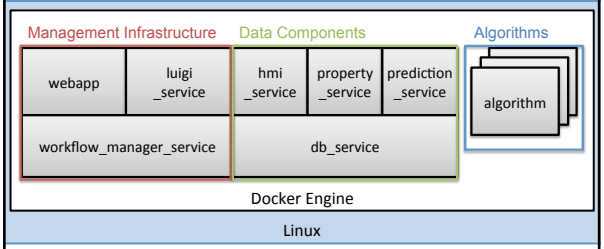
True Skill Statistic (Bloomfield, 2012): TSS = POD-FAR

Preliminary analysis of HMI data

- Consider the following properties extracted with SMART (Higgins et al 2011, Ahmed et al 2013):
  - Minimum value of the magnetic field in the region
  - Maximum value of the magnetic field in the region
  - Variance value of the magnetic field within the region
  - Sum of the magnetic field values within the region
  - Skewness value of the magnetic field within the region
  - Kurtosis value of the magnetic field within the region
  - Area of the region
  - Total positive flux in the region
  - Flux imbalance fraction in the region
  - Total negative flux in the region
  - Total unsigned magnetic flux of the region
  - Neutral-line length in the region
  - Falconer WLss value
  - Schrijver R value (with a lower threshold).
- Training Data: *8 August 2011* are used to train the models (2,623 samples)
- Test Data: first 12 hr of *9 August 9 2011*. (1,474 samples)



SUPPORT: DOCKER CONTAINERS

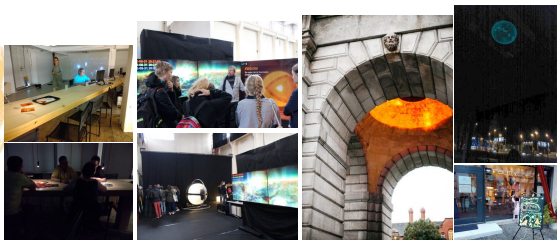


The evolution of science communication

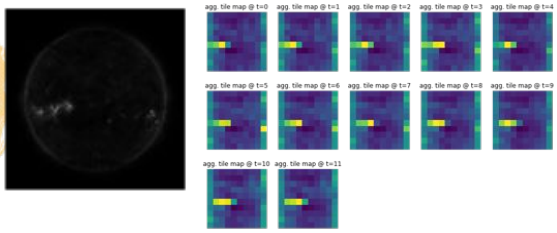
- *Public Understanding of Science*
- *Dialog with the Public*
- *Public Engagement in Science* →
- *Public Participation in Science*
- *to Responsible Research and Innovation RRI.*

http://flarecast.eu/outreach-resources

Experimenting with new communication formats



FURTHER ONGOING WORK: DEEP LEARNING



Thanks to the FLARECAST team !

Aleksandar Torbica, Andre Csillaghy, Anna Maria Massone, Annalisa Perasso, Chloé Guennou, Colin Klauser, Costis Gontikakis, Cristina Campi, D. Shaun Bloomfield, Dario Vischi, David Jackson, Douglas Biesecker, Eric Buchlin, Etienne Pariat, Federica Sciacchitano, Federico Benvenuto, Flavio Müller, Fraser Lott, Frederic Baudin, Graham Barnes, Hanna Sathiapal, Ioannis Kontogiannis, Jonas Lüthi, Jordan Guerra, Kostas Florios, Manolis Georgoulis, Manuel Ramirez Lopez, Marco Soldati, Mark Worsfold, Michele Piana, Neal hurlburt, Nicole Vilmer, Pablo Alingeri, Pascal Demoulin, Pedro Russo, Peter Gallagher, Roman Bolzern, Sabrina Guastavino, Samuel von Stachelski, Silvia Villa, Sophie masson, Sophie Murray, Stefan Müller, Sung-Hong Park, Vangelis Argoudelis, Vittorio Latorre, Véronique Bommier