An Overview of the European Union FLARECAST Project

Where do We Stand and Potential Future Directions of Research

16th European Solar Physics Meeting, September 6 – 10, 2021

FLARECAST: Consortium & Objectives

H2020-PROTEC-2014 RIA; Project No. 640216

Top-level Objectives:

- **Science:** understand the drivers of solar flare activity and improve flare prediction
- **Research-to-Operations (R2O):** provide a globally accessible flare prediction service that facilitates expansion
- **Communication:** engage with SWx users and inform policy makes and the wider public on solar flares and SWx in general

Paper available at JSWSC Agora Section, Topical Issue: Space Weather Research in the Digital Age and Across the Full-Data Lifecycle

https://doi.org/10.1051/swsc/2021023

M. K. **Georgoulis**, D. S. **Bloomfield**, M. **Piana**, A. M. **Massone**, M. **Soldati**, P.T. **Gallagher**, E. **Pariat**, N. **Vilmer**, E. **Buchlin**, F. **Baudin**, A. **Csillaghy**, H. **Sathiapal**, D. R. **Jackson**, P. **Alingery**, F. **Benvenuto**, C. **Campi**, K. **Florios**, C. **Gontikakis**, C. **Guennou**, J. A., **Guerra**, I. **Kontogiannis**, V. **Latorre**, S. A. **Murray**, S.-H. **Park**, S. **von Stachelski**, A. **Torbica**, D. **Vischi** & M. **Worsfold**

FLARECAST Near-Realtime Property Database

FLARECAST

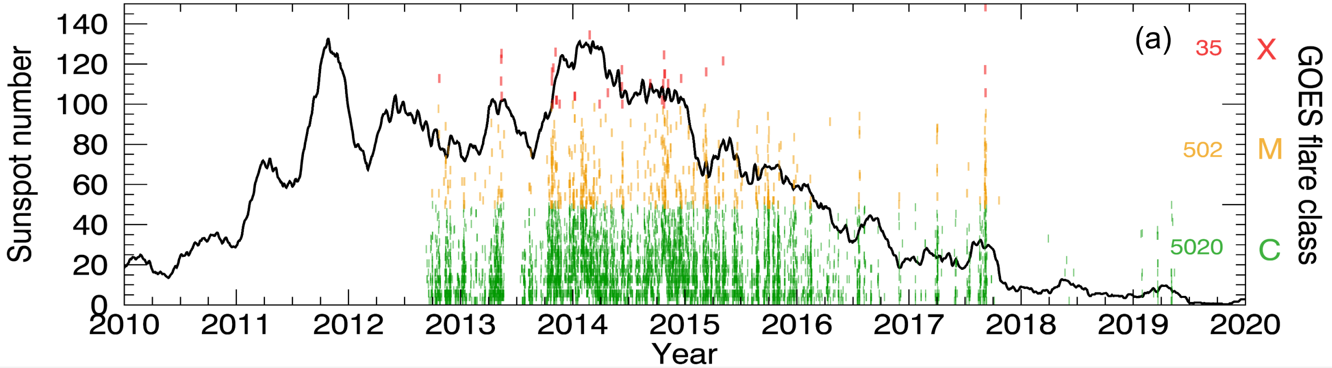
Top science objectives:

- Identify existing preflare properties of solar active regions, treat them as predictors and find those that work best for flare prediction
- Identify promising new predictors
- Implement an explorative research component, expanding into eruptive (i.e., CME-associated) flares

A total of 209 flare predictors

Data Source	Property Group	No. of Predictors	Relevant Predictor	Adapted from	Related references
SWPC	Solar Region Summary (SRS) properties	2	McIntosh and Hale classes;	McCloskey et al. (2016)	McIntosh (1990)
		3	Number, area and longitudinal extend of sunspots		Lee et al. (2012)
Catalogues	GOES soft X-ray flare events[a,b]	4	Flare magnitude, start, peak, and end times		
Surface-normal component (radial and line-of-sight) magnetograms	Effective connected magnetic field strength[c]	1	B_{eff}	Georgoulis and Rust (2007), Georgoulis (2011, 2013)	
	Fractal and Multifractal parameters	1	Fractal dimension	Conlon et al. (2008)	Abramenko et al. (2003), Abramenko (2005), Al-Ghraibah et al. (2015)
		1	Generalized correlation dimension		
		2	Holder exponent; Hausdorff dimension		
		2	Structure function's Inertial range index		
	Fourier and Wavelet power spectral indices	2	Power-law exponent	Hewett et al. (2008), Guerra et al. (2015)	
	Decay index (DI)[d]	8	Mean DI over PIL segments; Height of DI; Ratio of PIL length to DI height	Liu (2008), Zuccarello et al. (2014)	
	Magnetic PIL properties	5	Sum of PIL segments, Longest PIL segment	Mason and Hoeksema (2010)	
		1	R value	Schrijver (2007)	
		1	WL_{sg}	Falconer et al. (2012)	
	3D magnetic null points[d]	6	Number of null points in different height ranges (from 2 to 100 Mm above photosphere)	Haynes and Parnell (2007)	Pontin et al. (2013) Barnes and Leka (2006)
	Ising Energy[c]	6	Original and partitioned Ising energy	Ahmed et al. (2010)	Kontogiannis et al. (2018)
	Magnetic energy and helicity	11	Poynting flux and magnetic helicity flux proxies	Park et al. (2010), Park et al. (2012)	
Full-Vector magnetograms	SHARP properties[e]	100	Horizontal gradient of B components; Shear angle; Unsigned vertical current; higher-order moments of timeseries	Bobra et al. (2014) (validated) Leka and Barnes (2003b, 2007)	
	Magnetic energy and helicity	22	Poynting flux and magnetic helicity flux	Kusano et al. (2002)	Berger and Field (1984), Welsch et al. (2009)
	Non-neutralized Currents	6	Total non-neutralized current	Georgoulis et al. (2012)	Kontogiannis et al. (2017)
	Flows around PIL	22	Speed of diverging/converging/shear flows	Park et al. (2018),	Deng et al. (2006), Wang et al. (2014)
Intensity Images[f]	Magnetic field gradient	3	Total horizontal magnetic gradient	Korsós et al. (2014)	Kontogiannis et al. (2018)

[a] AR coronal information.
[b] The project uses the flare attributes as properties. At least one flare of desired magnitude in the forecast window signifies a positive instance.
[c] Photospheric proxy for coronal information.
[d] Uses PFEs.
[e] Most SHARP parameters correspond to mean values. In the FLARECAST pipeline, other relevant magnetogram-related parameters are also used as predictors; these are maximum and minimum values, median, standard deviation, kurtosis, and skewness.
[f] Used in conjunction with magnetograms to calculate the sum of the magnetic field gradients between all possible opposite-polarity umbrae pairs.



Most of solar cycle 24 covered!

A total of 5,557 flares:

- 5,020 GOES C-class
- 502 GOES M-class
- 35 GOES X-class

FLARECAST Performance Verification Engine

FLARECAST

	Forecast Flare	Forecast No-flare
Observed Flare	TP	FN
Observed No-flare	FP	TN

Both deterministic (YES/NO) and probabilistic metrics in place

Name	Notation	Formula	Range
Accuracy	ACC	$\frac{TP+TN}{N}$	[0, 1]
False alarm ratio	FAR	$\frac{FP}{TP+FP}$	[0, 1]
Bias	BIAS	$\frac{TP+FP}{TP+FN}$	[0, ∞]
Threat score	TS	$\frac{TP}{TP+FN+FP}$	[0, 1]
Equitable threat score	ETS	$\frac{TP-R_{ETS}}{TP+FN+FP-R_{ETS}}$	$[-\frac{1}{3}, 1]$
		Using $R_{ETS}=\frac{(TP+FN)(TP+FP)}{N}$	
Probability of detection	POD	$\frac{TP}{TP+FN}$	[0, 1]
Probability of false detection	POFD	$\frac{FP}{FP+TN}$	[0, 1]
Odds ratio	OR	$\frac{TP\cdot TN}{FN\cdot FP}$	[0, ∞]
Odds ratio skill score	ORSS	$\frac{(TP\cdot TN)-(FN\cdot FP)}{(TP\cdot TN)+(FN\cdot FP)}$	[−1, 1]
Heidke skill score	HSS	$\frac{TP+TN-R_{HSS}}{N-R_{HSS}}$	[−1, 1]
		Using $R_{HSS}=\frac{(TP+FN)(TP+FP)+(TN+FN)(TN+FP)}{N}$	
True skill statistic	TSS	POD − POFD	[−1, 1]
Symmetric extremal dependence index	SEDI	$\frac{\log(POFD)-\log(POD)-\log(1-POFD)+\log(1-POD)}{\log(POFD)+\log(POD)+\log(1-POFD)+\log(1-POD)}$	[−1, 1]
Appleman's discriminant	AD	$\frac{TN-FN}{FP+TN}$ if $(TP+FN)>(FP+TN)$	$[-\frac{FN}{FP}, 1]$
		$\frac{TP-FP}{FN+TP}$ if $(TP+FN)<(FP+TN)$	$[-\frac{FP}{FN}, 1]$

Brier skill score:

$$SS = 1 - \frac{MSE_{forecast}}{MSE_{ref}}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(o_i - p_i)^2$$

$$BSS = 1 - \frac{MSE(\bar{o}, p)}{MSE(\tilde{o}, p)}$$

$$\bar{o} \equiv \{0, 1\}$$

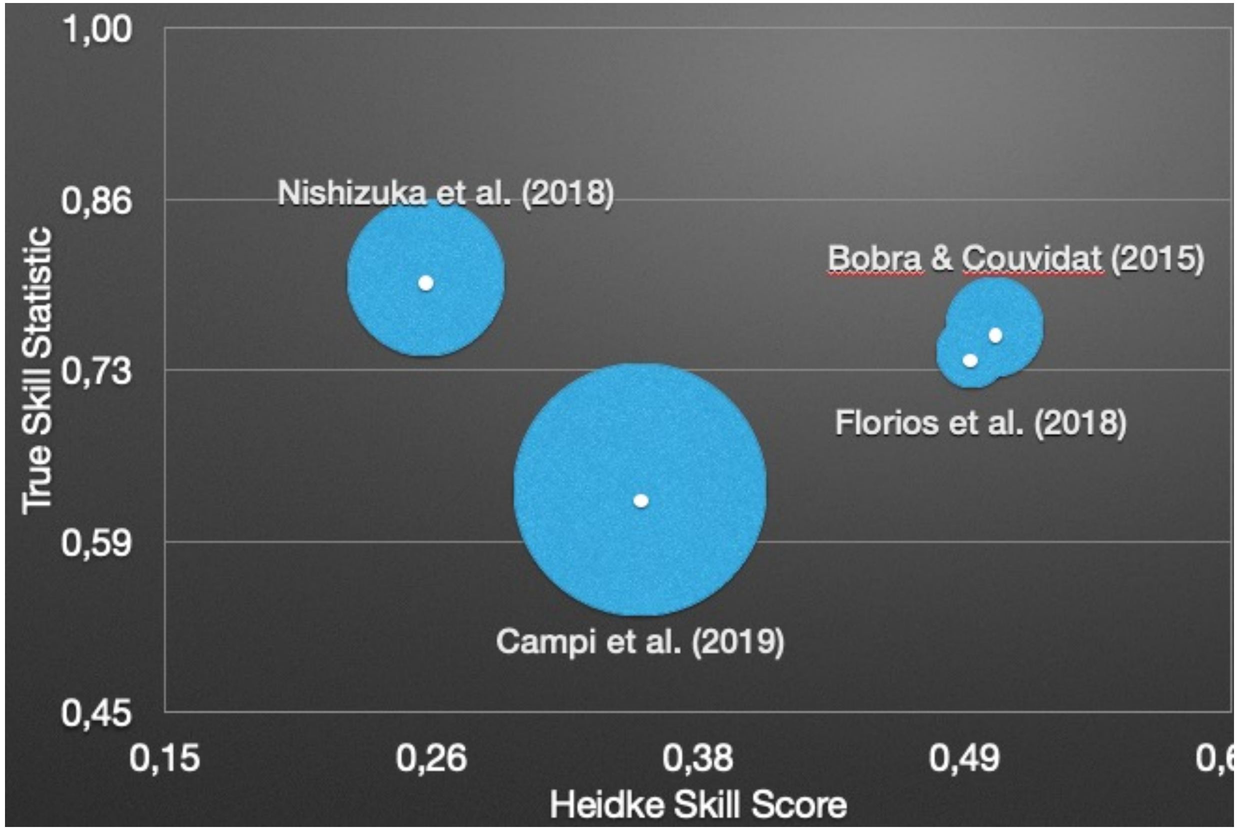
$$\tilde{o} = \frac{1}{N}\sum_{i=1}^{N} o_i$$


```
{
  "services": {
    "hmi": {
      "read": {"host": "hmi", "port": 8001}
    },
    "property": {
      "read": {"host": "cluster-r730-1", "port": 8002},
      "write": {"host": "localhost", "port": 8002}
    },
    "prediction": {
      "read": {"host": "cluster-r730-1", "port": 8004},
      "write": {"host": "localhost", "port": 8004}
    },
    "verification": {
      "read": {"host": "cluster-r730-1", "port": 8006},
      "write": {"host": "localhost", "port": 8006}
    },
    "log_prefix": "FLARECAST-LOG"
  },
  "runtime": {
    "run_id": 1
  },
  "algorithm": {
    "algorithm_name": "HybridLogit",
    "flaring_level": "abovec",
    "forecast_window": 24,
    "forecast_latency": 0,
    "base_issuing_time": 0,
    "property_set": "Blos_all_nofh"
  },
  "dataset": {
    "time_interval": {
      "start_time": "2015-01-01T00:00:00Z",
      "end_time": "2017-12-31T23:59:59Z"
    },
    "probabilities": {
      "rel_dia_stepsize": 0.05,
      "threshold_stepsize": 0.05
    },
    "uncertainties": {
      "calculate": true,
      "sampling_scheme": "bootstrap",
      "sample_percent": 0.9,
      "realizations": 1000
    }
  }
}
```


verification_engine.py

FLARECAST Science

- Replicating, evaluating and re-shaping best practices in training and testing of forecast methods, *at the same time re-assessing stochasticity in flare occurrence*
- Heavily investing in supervised and unsupervised machine learning methods
- Resonating on the use of decision trees (i.e., random forests) and hybrid (i.e., hybrid LASSO) methods
- Finding new and interesting predictors mostly associated to magnetic polarity inversion lines in active regions

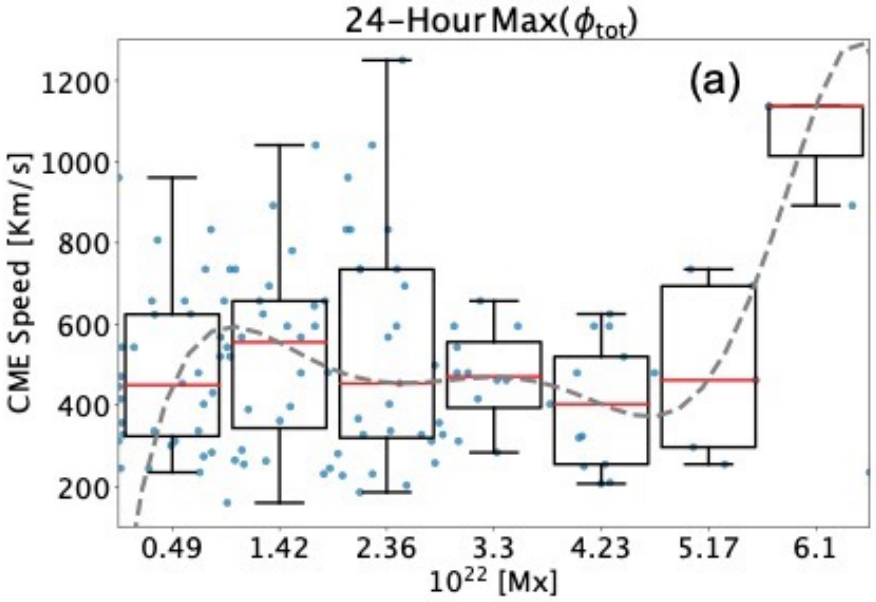


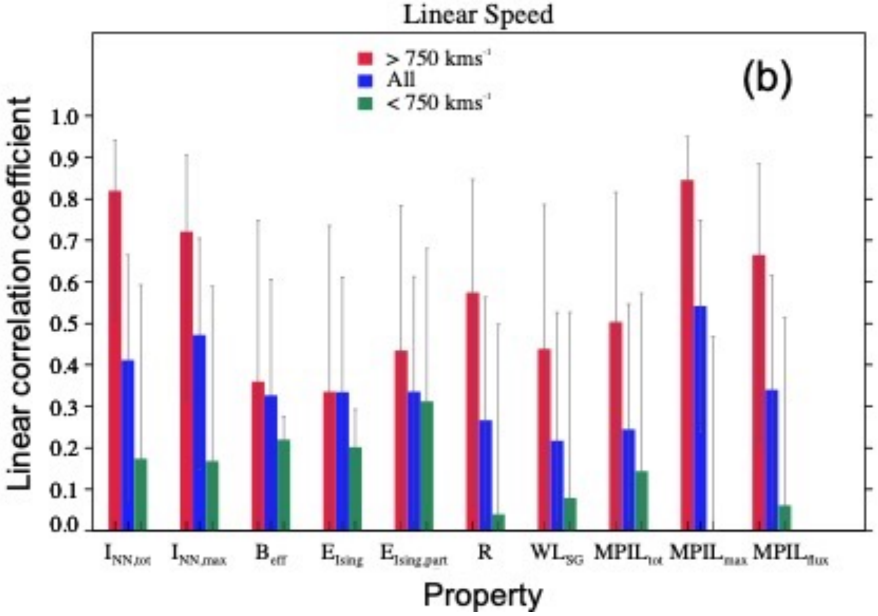
	Test Set-C1+	Test Set C1+	Test Set-M1+	Test Set-M1+
	TSS	HSS	TSS	HSS
HLA	0.58 ± 0.01	0.51 ± 0.01	0.70 ± 0.02	0.31 ± 0.03
RF	0.61 ± 0.01	0.56 ± 0.02	0.71 ± 0.03	0.39 ± 0.02
Florios et al. (2018)	0.60 ± 0.01	0.59 ± 0.01	0.74 ± 0.02	0.49 ± 0.01
Bobra & Couvidat (2015)	⋯	⋯	0.76 ± 0.04	0.52 ± 0.04

Promising Future Possibilities

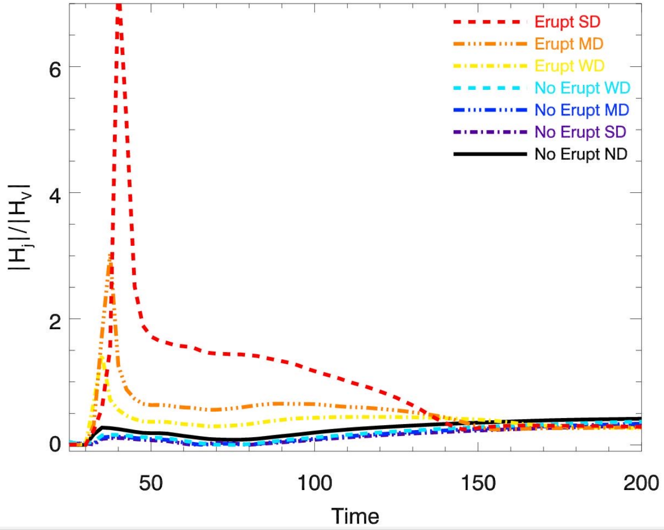
FLARECAST

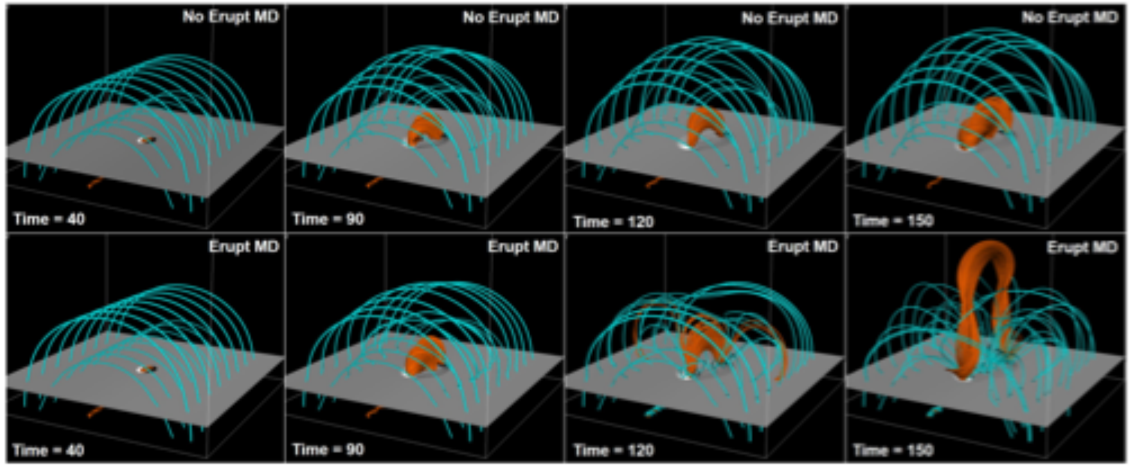
- Correlating active-region properties from the property database to CME speed and other CME characteristics in case of eruptive flares – ***PIL electric currents and lengths showing best performance***





- Using synthetic data and evaluating magnetic helicity terms in 3D to characterize the pre-eruption situation – ***an apparent eruptivity ratio occuring***





FLARECAST Science Crop (partial + full support)

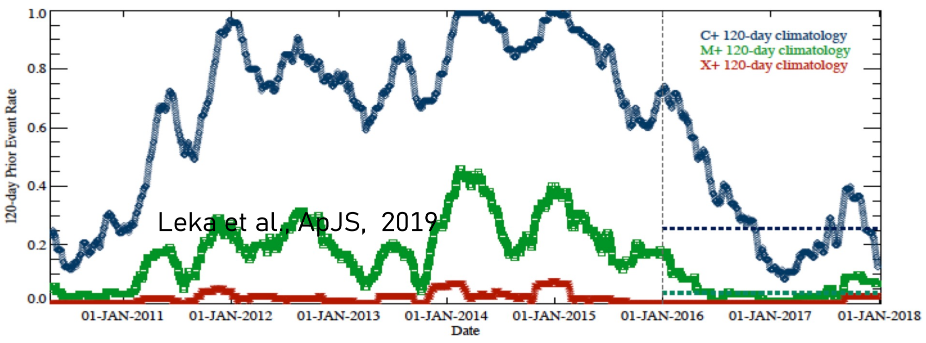
FLARECAST

- Barnes G, Leka KD, Schrijver CJ, Colak T, Qahwaji R, et al. 2016. A comparison of flare forecasting methods. I. Results from the "All-Clear" workshop. *Astrophys J* **829**: 89. DOI.
- Benvenuto F, Piana M, Campi C, Massone AM. 2018. A hybrid supervised/unsupervised machine learning approach to solar flare prediction. *Astrophys J* **853(1)**: 90. DOI.
- Campi C, Benvenuto F, Massone AM, Bloomfield DS, Georgoulis MK, Piana M. 2019. Feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence. *Astrophys J* **883(2)**: 150. DOI.
- Florios K, Kontogiannis I, Park S-H, Guerra JA, Benvenuto F, Bloomfield DS, Georgoulis MK. 2018. Forecasting solar flares using magnetogram-based predictors and machine learning. *Sol Phys* **293(2)**: 28. DOI.
- Gontikakis C, Kontogiannis I, Georgoulis MK, Guennou C, Syntelis P, Park SH, Buchlin E. 2020. Differential emission measure evolution as a precursor of solar flares. *arXiv e-prints*, ArXiV.
- Guennou C, Pariat E, Leake JE, Vilmer N. 2017. Testing predictors of eruptivity using parametric flux emergence simulations. *J Space Weather Space Clim* **7**: A17. DOI.
- Guerra JA, Park SH, Gallagher PT, Kontogiannis I, Georgoulis MK, Bloomfield DS. 2018. Active region photospheric magnetic properties derived from line-of-sight and radial fields. *Sol Phys* **293(1)**: 9. DOI.
- Kontogiannis I, Georgoulis MK, Park S-H, Guerra JA. 2017. Non-neutralized electric currents in solar active regions and flare productivity. *Sol Phys* **292**: 159. DOI.
- Kontogiannis I, Georgoulis MK, Park S-H, Guerra JA. 2018. Testing and improving a set of morphological predictors of flaring activity. *Sol Phys* **293**: 96. DOI.
- Kontogiannis I, Georgoulis MK, Guerra JA, Park S-H, Bloomfield DS. 2019. Which photospheric characteristics are most relevant to active-region coronal mass ejections? *Sol Phys* **294(9)**: 130. DOI.
- Massone AM, Piana M, the FLARECAST Consortium. 2018. Chapter 14 – machine learning for flare forecasting. In: *Machine learning techniques for space weather*, Camporeale E, Wing S, Johnson JR (Eds.), Elsevier Amsterdam, Oxford, Cambridge MA, pp. 355–364. ISBN 978-0-12-811788-0. DOI.
- McCloskey AE, Gallagher PT, Bloomfield DS. 2016. Flaring rates and the evolution of sunspot group McIntosh classifications. *Sol Phys* **291(6)**: 1711–1738. DOI.
- Murray SA, Bingham S, Sharpe M, Jackson DR. 2017. Flare forecasting at the Met Office Space Weather Operations Centre. *Space Weather* **15**: 577–588. DOI.
- Murray SA, Guerra JA, Zucca P, Park S-H, Carley EP, Gallagher PT, Vilmer N, Bothmer V. 2018. Connecting coronal mass ejections to their solar active region sources: Combining results from the HELCATS and FLARECAST projects. *Sol Phys* **293**: 60. DOI.
- Pariat E, Leake JE, Valori G, Linton MG, Zuccarello FP, Dalmasse K. 2017. Relative magnetic helicity as a diagnostic of solar eruptivity. *A&A* **601**: A125. DOI.
- Park S-H, Guerra JA, Gallagher PT, Georgoulis MK, Bloomfield DS. 2018. Photospheric shear flows in solar active regions and their relation to flare occurrence. *Sol Phys* **293**: 114. DOI.
- Sharpe MA, Murray SA. 2017. Verification of space weather forecasts issued by the Met Office Space Weather Operations Centre. *Space Weather* **15(10)**: 1383–1395. DOI.

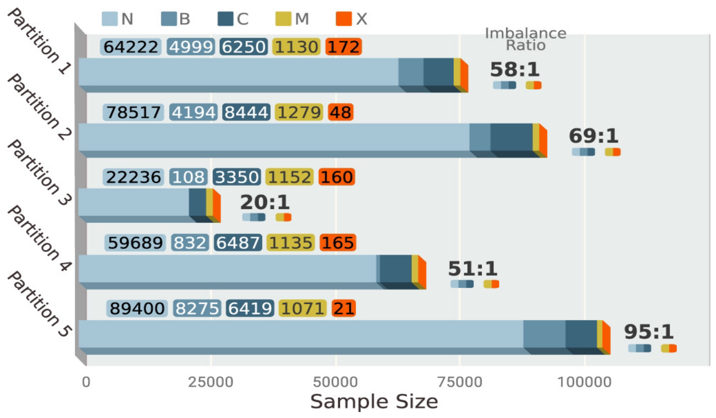
Forecasting Challenges Ahead

FLARECAST

- Varying climatology over a given solar cycle:

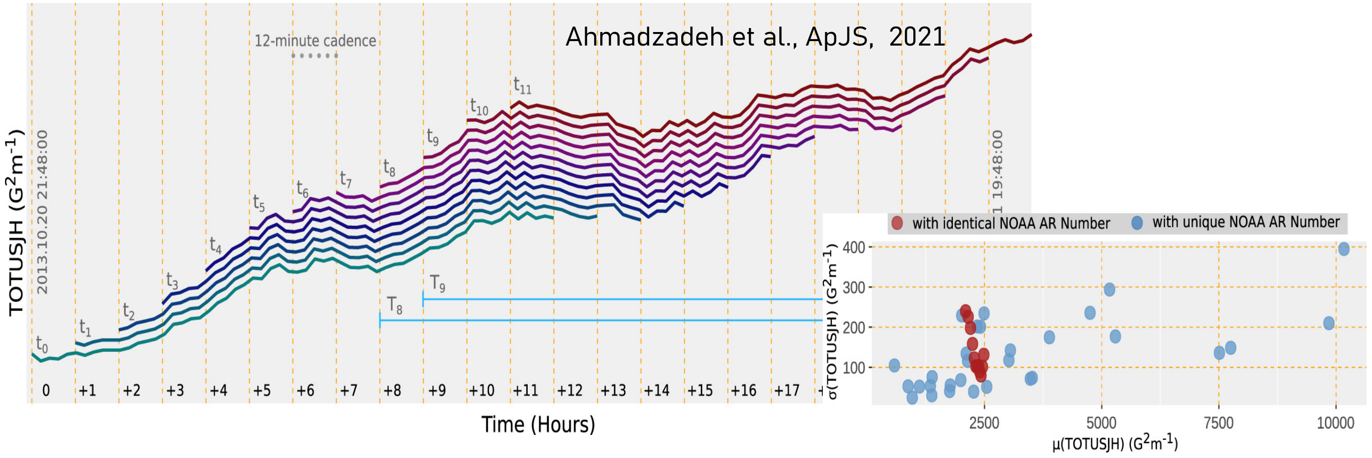


- Class imbalance in flare occurrence:

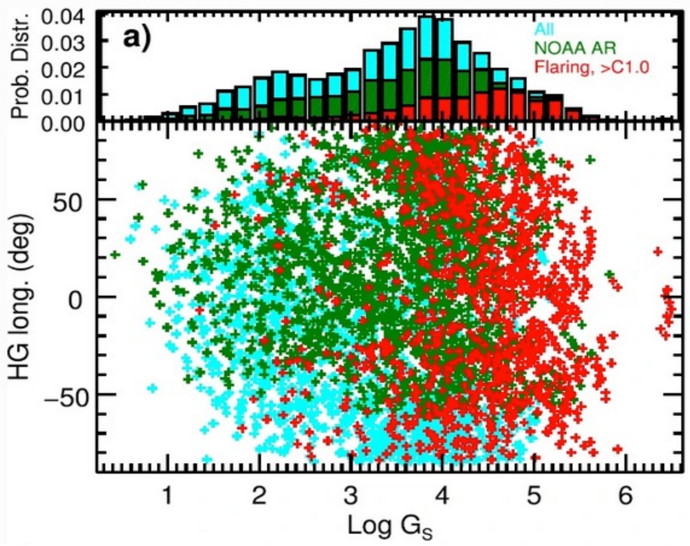


Ahmadzadeh et al., ApJS, 2021

- Temporal coherence in timeseries forecasting:



- Stochasticity:



Kontogiannis et al., SoPh, 2018

Open Access: Data; Codes; Infrastructure

- API access in place
- Two databases:
  - Property DB: https://api.flarecast.eu/property/ui
  - Prediction DB: https://api.flarecast.eu/prediction/ui
- Generic command for Property DB:

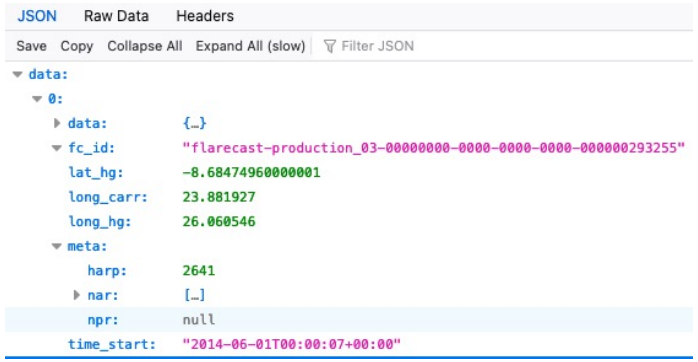
  https://api.flarecast.eu/property/region/production_03/list?cadence=all&exclude_higher_cadences=false&time_start=between(2014-06-01,2014-06-03)&property_type=*&region_fields=*

  (example covering all metadata cadences for a two-day interval [June 1 – 2, 2014])

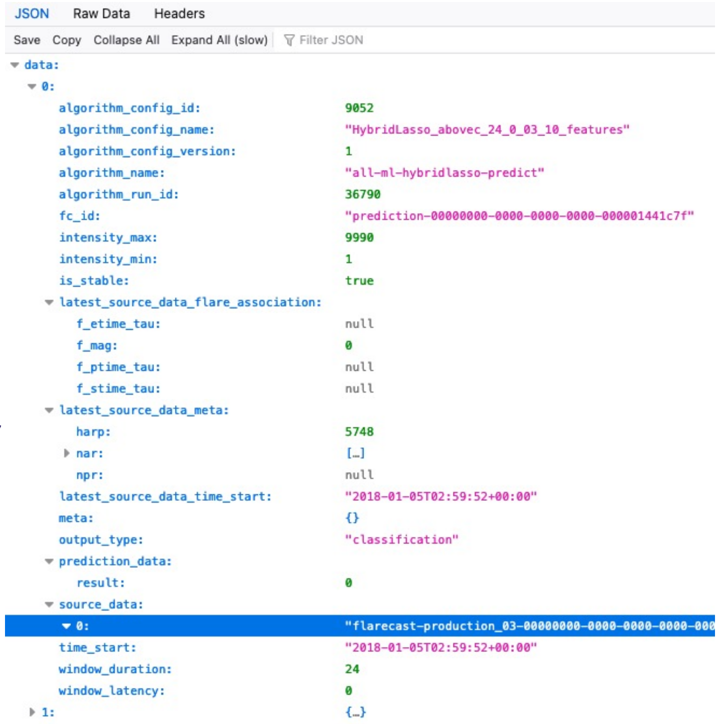
- Generic command for Prediction DB:

  https://api.flarecast.eu/prediction/prediction/list_v2?include_flare_associations=true&algorithm_config_version=latest&prediction_time_start=between(2018-01-01,2018-06-01)

  (example covering all available predictions on a five-month interval [Jan 1 – Jun 1, 2018])



Results in JSON format



Open Access: Data; Codes; Infrastructure

FLARECAST

Infrastructure Installation (on Unix / Linux):

- Dependency: Docker engine : https://www.docker.com
- Download the installer script:

```
curl -o infrastructure.sh
https://dev.flarecast.eu/stash/projects/INFRA/repos/
dev-infra/browse/infrastructure.sh?raw
```

- Assign executional mode to the script:

```
chmod a+x infrastructure.sh
```

- Update to the most recent version:

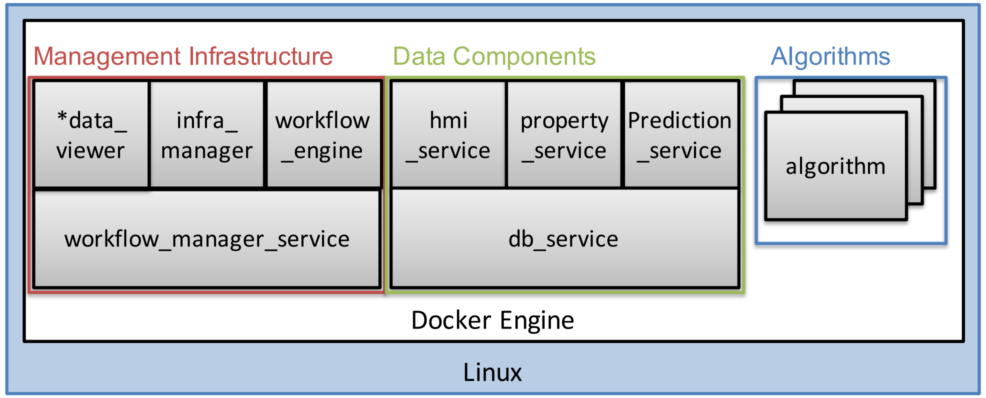
```
./infrastructure.sh update
```

- And, finally, run the infrastructure:

```
./infrastructure.sh run
```

- Repository of codes: https://dev.flarecast.eu/stash/projects/

docker



In accordance with:

https://www.openaire.eu/

OpenAIRE

Conclusions

FLARECAST has generated data and infrastructure that can help avoid error duplication in future SWx forecasting efforts

- ❑ Solar flare occurrence is fundamentally stochastic. Even one of the most data intensive efforts such as FLARECAST could not lift this stochasticity barrier
- ❑ Stochasticity shows up in alternating patterns of best predictors using different machine learning methods, or even different threshold flare classes
- ❑ Training, testing and verification of a given forecast method are of utmost importance, at times more important than the method itself
- ❑ Flare prediction, like many real-world problems, is an interdisciplinary endeavor. A fusion of specialty and expertise is imperative and instrumental

Sincere thanks are due to:

- NASA/SDO and Stanford's JSOC
- NOAA / SWPC
- EU Project Officer & assigned Reviewer
- The admin structure of the nine partner's and the overall Project Manager
- The FLARECAST Steering Committee
- Each and every external code provider
- All participants of the FLARECAST Science & Stakeholder Workshops
- Each and every team member, for their dedication, devotion and hard work

See you in Athens next year!