NASA PARTNER FDL X HELIO

TRILLIUM USA

FDLXHELIO.ORG

A Scientific Cloud Computing Platform for Ingestion and Processing of SDO Data

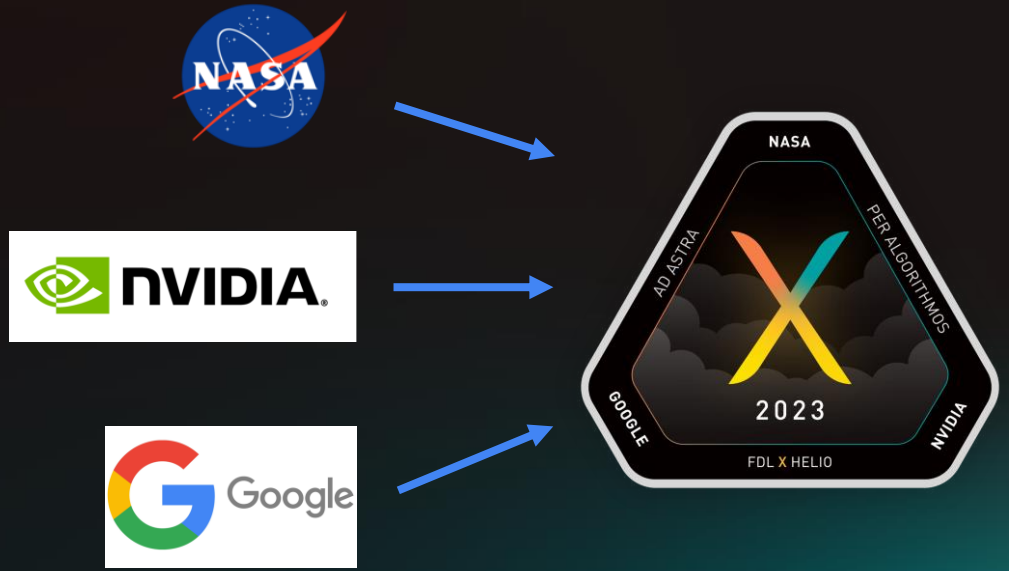
4th Eddy Cross-Disciplinary Symposium
Wednesday, 1st November

Manuel Indaco[1]
Daniel Gass[2] (presenter), William Fawcett[3] , Richard Galvez[4],
Andrés Muñoz-Jaramillo[5], Paul Wright[6]

[1]Auburn University, [2]UCLAN, [3]University of Cambridge, [4]DataTalk AI,
[5]Southwest Research Institute, [6]Dublin Institute for Advanced Studies

Google Cloud NVIDIA

FDL Is a private/public partnership between NASA, Google, and NVIDIA.



Innovative AI solutions of interest to NASA (& everyone else!)

FDL-X combines integrated AI pipelines, machine learning and domain science across heliophysics challenges. Please join us for presentations from all three teams.

Multiscale Geoeffectiveness Forecasting using SHEATH and DAGGER

Vishal Upendran
Tuesday 2:25 PM

Improving thermospheric drag modeling with EUV images: an FDL-X 2023 project

Tom Berger
Wednesday 1:45 PM

AIA is All You Need: SDO MEGS A&B virtualization via Convolutional Deep Learning

Daniel Gass
Tuesday 2:15 PM

A Scientific Cloud Computing Platform for Ingestion and Processing of SDO Data

Manuel Indaco
Wednesday 2:10 PM

AI Inference products, foundation models and multi-domain approaches to NASA Heliophysics.

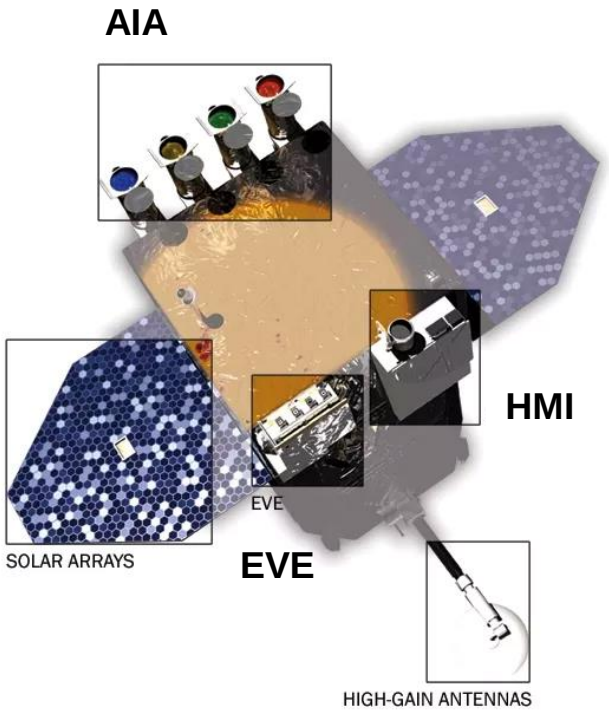
FDL-X
James Parr
Wednesday 2:20 PM

Learn more at FDL.AI

Thank you to our partners

NASA PARTNER FDL X HELIO Google Cloud NVIDIA FDLXHELIO.ORG TRILLIUM USA

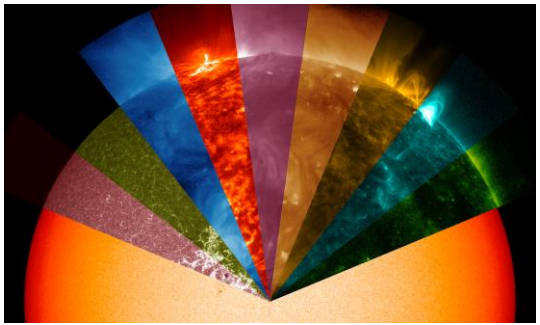
The Solar Dynamics Observatory (SDO)



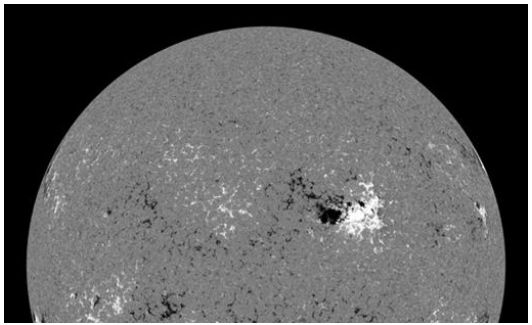
Atmospheric Imaging Assembly (AIA)
4096 x 4096 full-sun images in 10 channels

Helioseismic and Magnetic Imager (HMI)
Effectively image sun's magnetic activity

Extreme ultraviolet Variability Experiment (EVE)
Measures EUV irradiance in select ion ranges

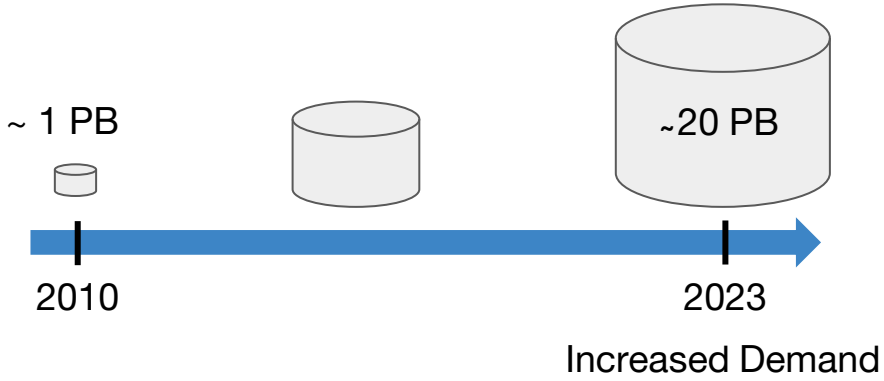


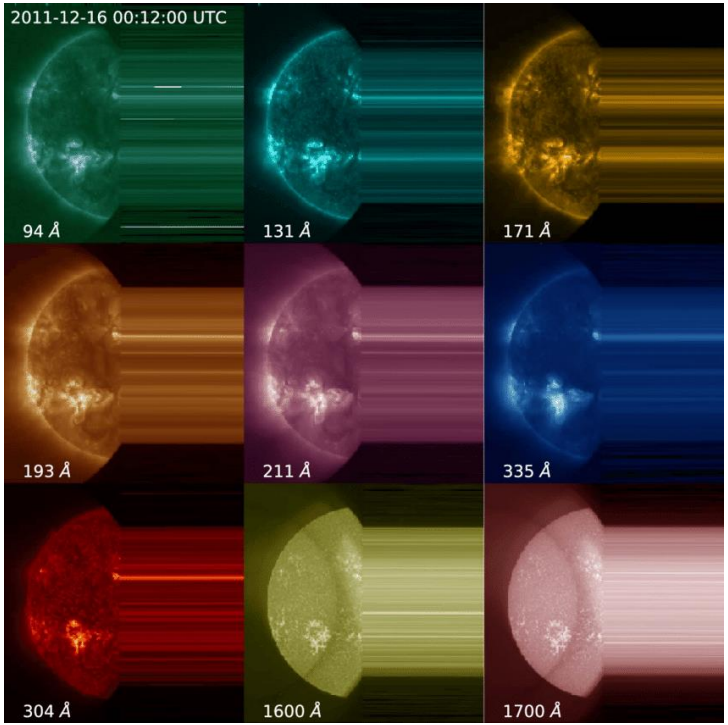
AIA



HMI

SDO Data Infrastructure Issues





- Data infrastructure designed decades ago.
- Data must be curated.
- Need of compute resources to perform large-scale analysis.

Need for an **automated** pipeline to prepare and serve data.

Can we make SDO data more accessible for everyone?

SDO data are hard to get…

- Egress limitations from JSOC
- Many different types of data product to choose from
- Data types at different calibrations / cadences / resolutions
- No compute resources available

The SDO Machine Learning dataset (SDOML)

- The SDOML dataset was introduced in FDL2018 and has had subsequent improvements over the years

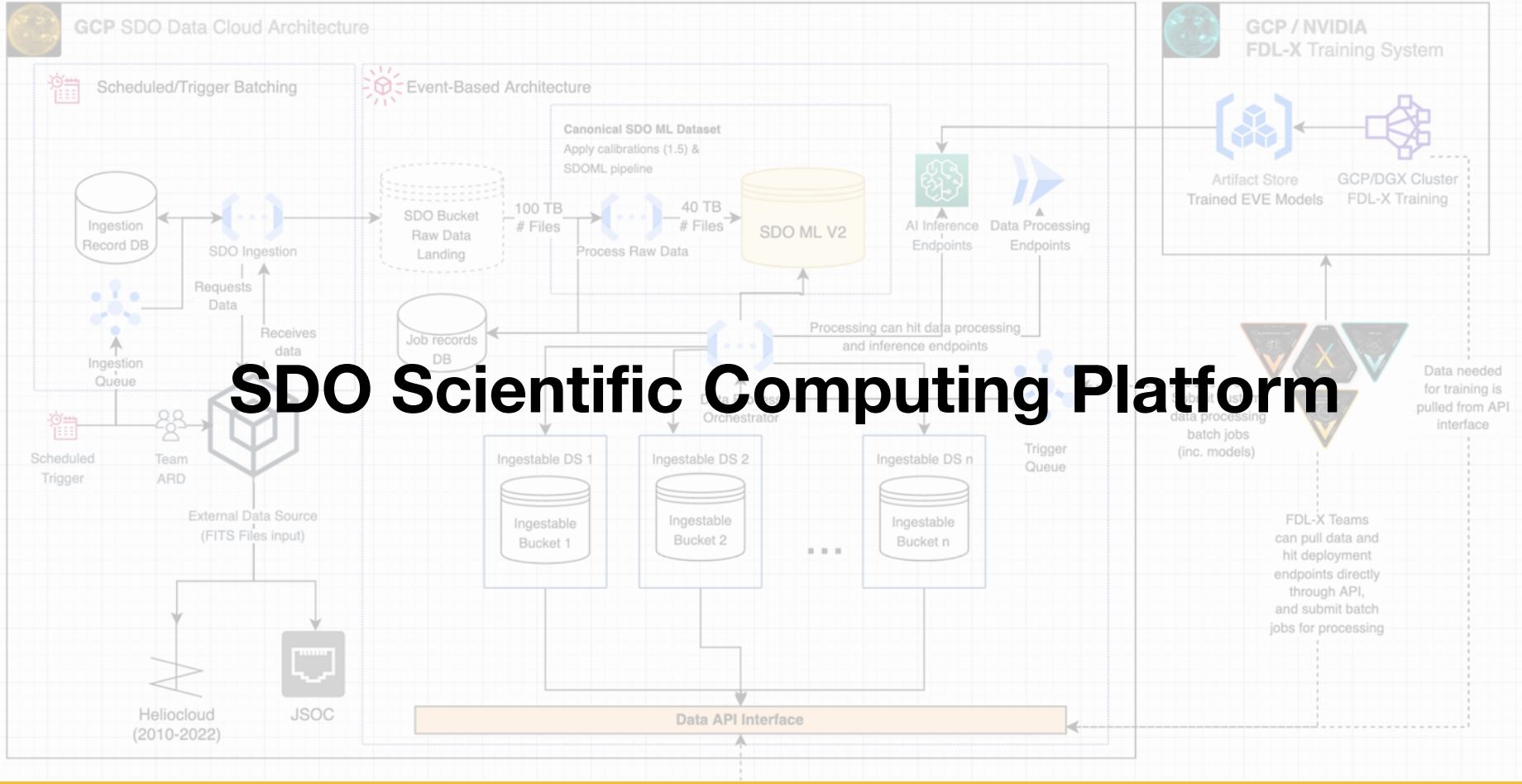
- Some key points are:
  - AIA & HMI images, and ion irradiance from EVE
  - Images corrected for instrument degradation
  - Calibration to “level 1.5”
  - Solar disk position & size harmonized across images
  - 512x512 resolution (easy for ML)

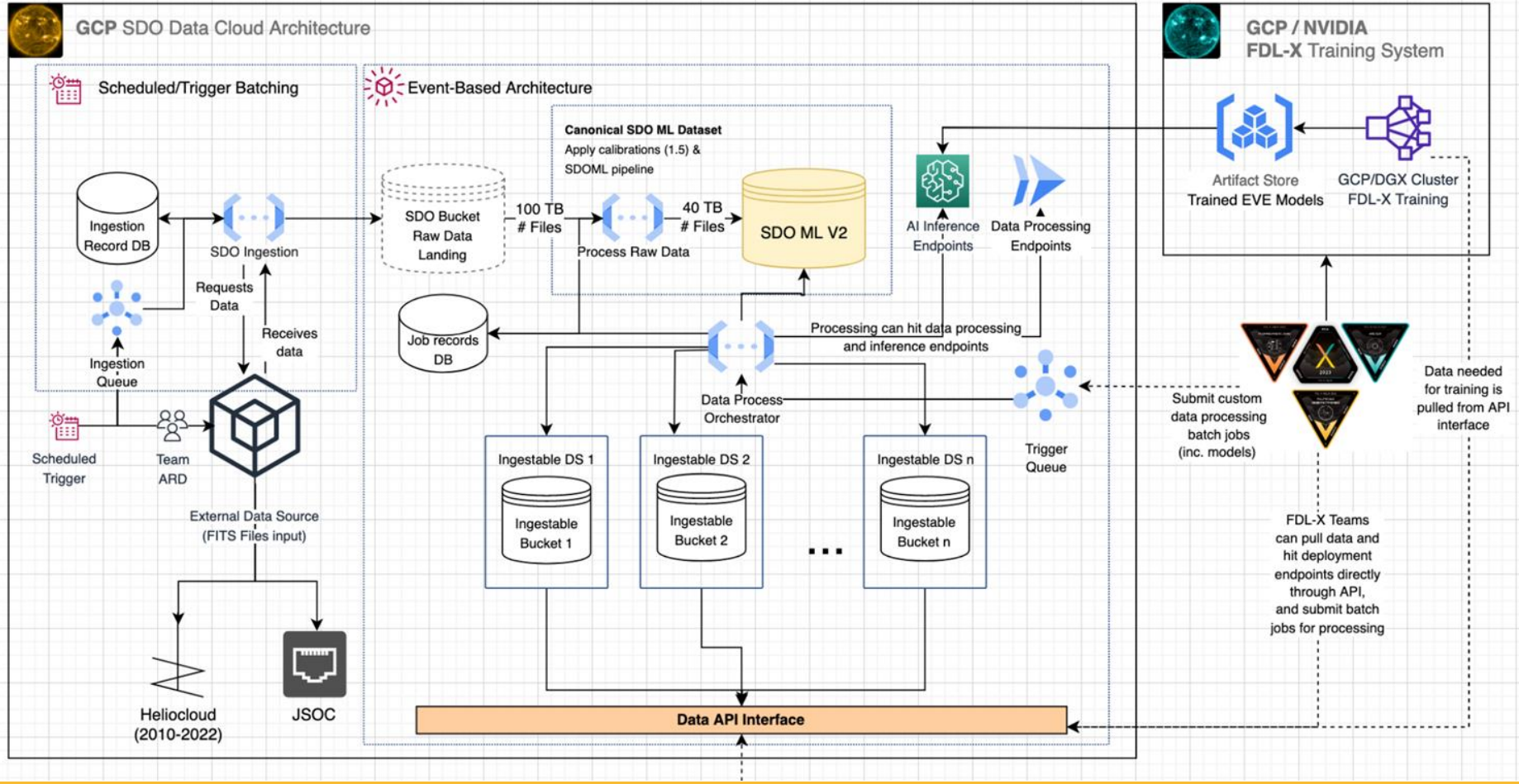
- The dataset is a curated, **machine-learning ready** dataset

- SDOMLv2 **only had SDO data up to 2020**

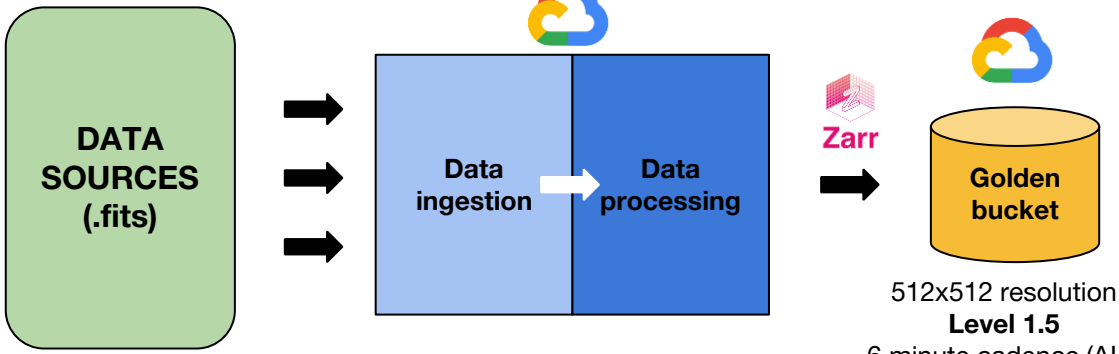
- **Addition of new data not automated**

SDO Scientific Computing Platform




GCP SDO Data Cloud Architecture
Scheduled/Trigger Batching
Event-Based Architecture
GCP / NVIDIA
FDL-X Training System
Canonical SDO ML Dataset
Apply calibrations (1.5) &
SDOML pipeline
Ingestion
Record DB
SDO Ingestion
Requests
Data
Receives
data
Ingestion
Queue
Scheduled
Trigger
Team
ARD
External Data Source
(FITS Files input)
Heliocloud
(2010-2022)
JSOC
SDO Bucket
Raw Data
Landing
100 TB
Files
Process Raw Data
40 TB
Files
SDO ML V2
Job records
DB
AI Inference
Endpoints
Data Processing
Endpoints
Artifact Store
Trained EVE Models
GCP/DGX Cluster
FDL-X Training
Processing can hit data processing
and inference endpoints
Data Process
Orchestrator
Trigger
Queue
Submit custom
data processing
batch jobs
(inc. models)
Data needed
for training is
pulled from API
interface
Ingestable DS 1
Ingestable
Bucket 1
Ingestable DS 2
Ingestable
Bucket 2
Ingestable DS n
Ingestable
Bucket n
FDL-X Teams
can pull data and
hit deployment
endpoints directly
through API,
and submit batch
jobs for processing
Data API Interface

Data from new sources must be processed and aligned with existing SDOMLv2



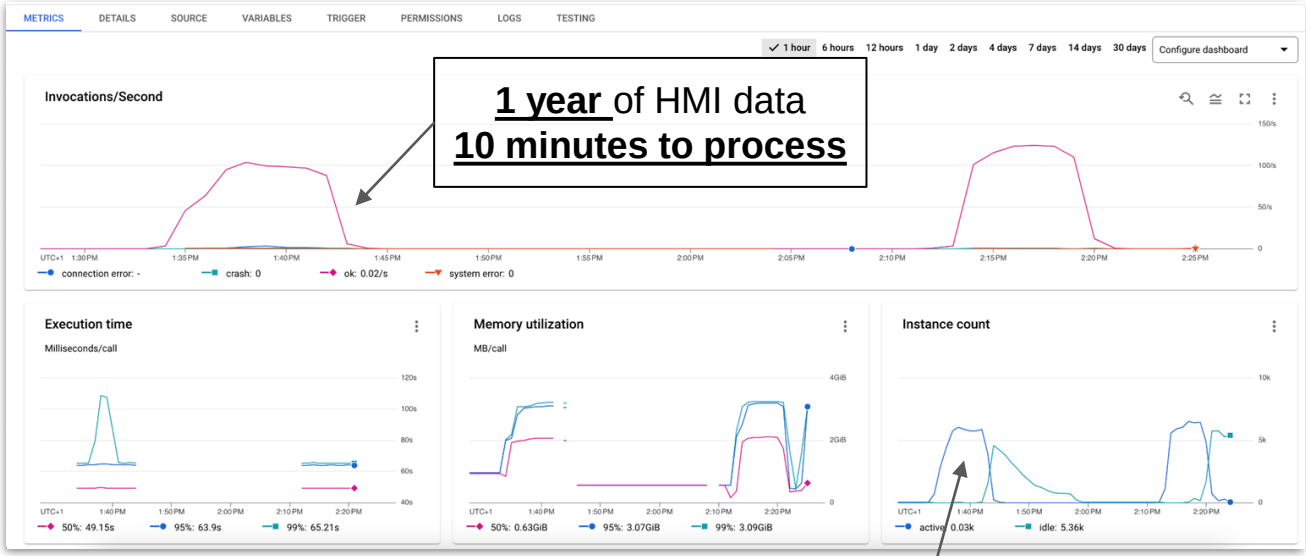
Pipeline in Action: Data Processing

Data processing

Zarr

Golden bucket

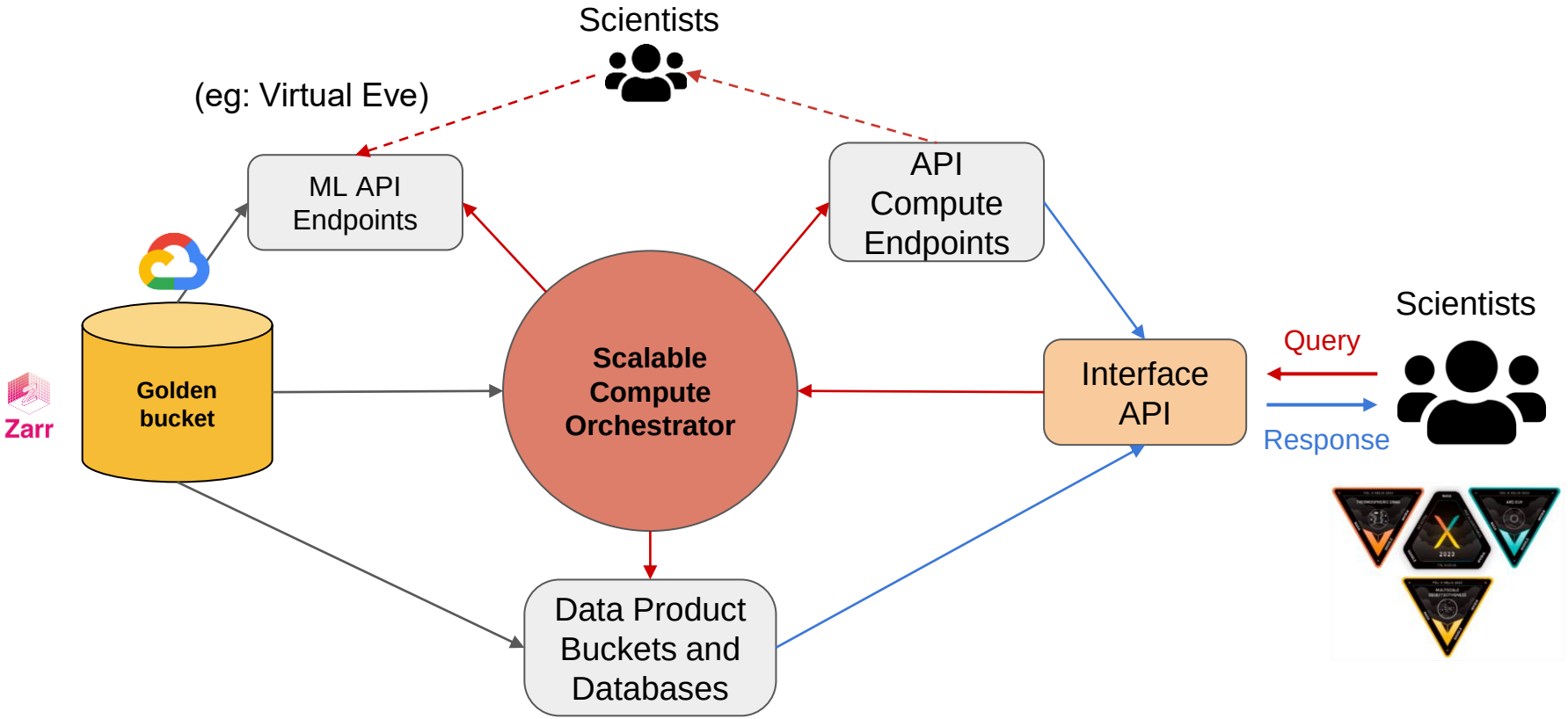
512x512 resolution
Level 1.5
6 minute cadence (AIA)
12 minute cadence (HMI)

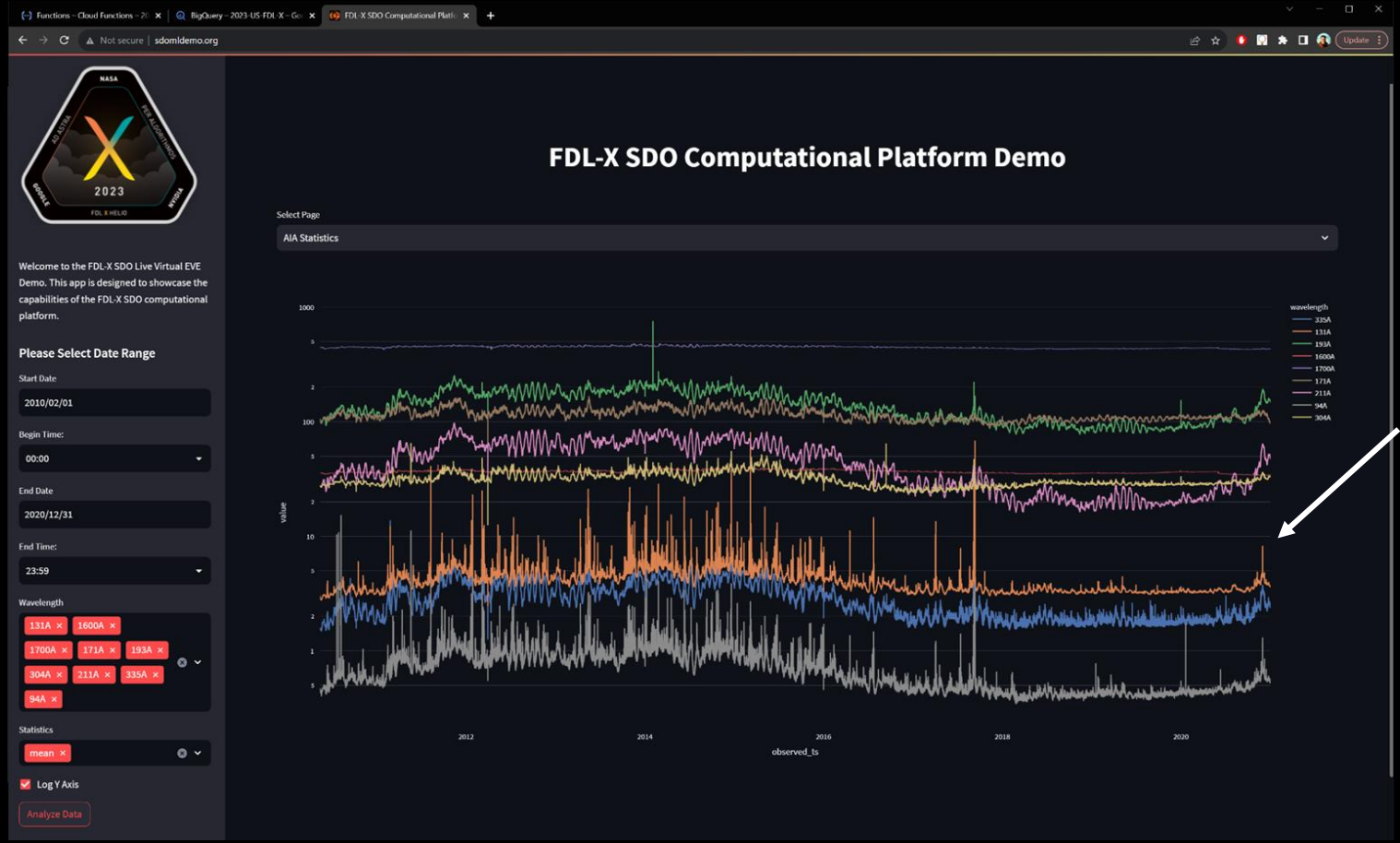


1 year of HMI data
10 minutes to process

6000 concurrent machines

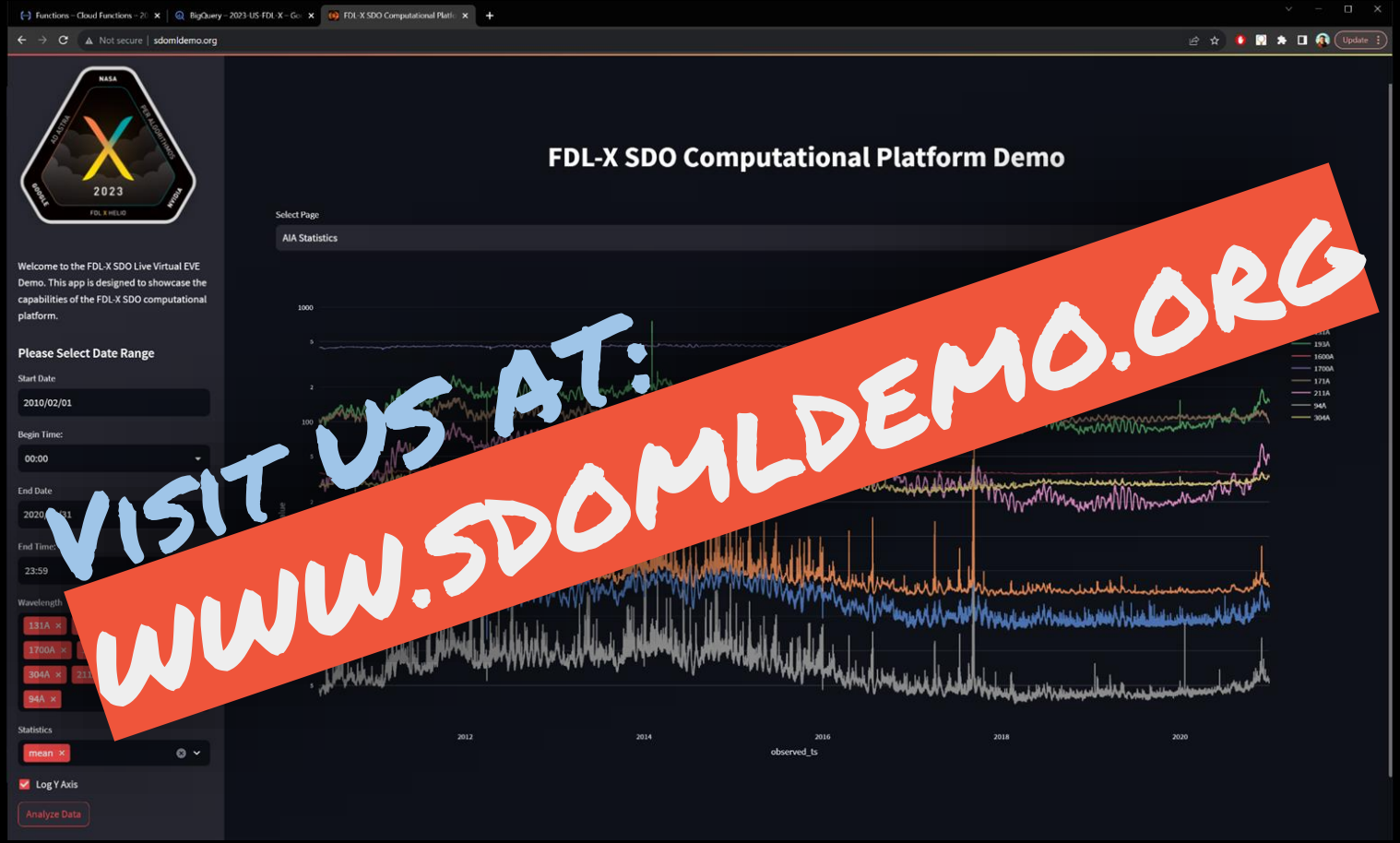
Serverless Computational Platform





sdomldemo.org/

AIA summary statistics


sdomldemo.org
FDL-X SDO Computational Platform Demo
Select Page
AIA Statistics
Welcome to the FDL-X SDO Live Virtual EVE Demo. This app is designed to showcase the capabilities of the FDL-X SDO computational platform.
Please Select Date Range
Start Date
2010/02/01
Begin Time:
00:00
End Date
End Time:
23:59
Wavelength
131A
1700A
304A
94A
Statistics
mean
Log Y Axis
Analyze Data
1000
100
2012
2014
2016
2018
2020
observed_ts
193A
1600A
1700A
171A
211A
94A
304A
VISIT US AT:
WWW.SDOMLDEMO.ORG

Results

SDO Computational Platform
A cloud-based large-scale scalable data ingestion, processing, and ML platform.

SDO ML V2
Analysis ready data for *all* SDO AIA, HMI, and EVE data

Virtual EVE
A deep learning model based on AIA input data providing live proxy EVE solar irradiance measurements to the community; a blueprint for future virtual instruments like it.

Impact

We've built the computational data analysis cloud-based platform that we believe SDO has always needed.

We produced a standardized and easy-to-access suite of SDO data ready for science and machine learning that live-updates.

We've built the necessary tools that bridge the divide between data and compute which will help further unlock the scientific value of the SDO mission (and others).

TRILLIUM USA

FDLXHELIO.ORG

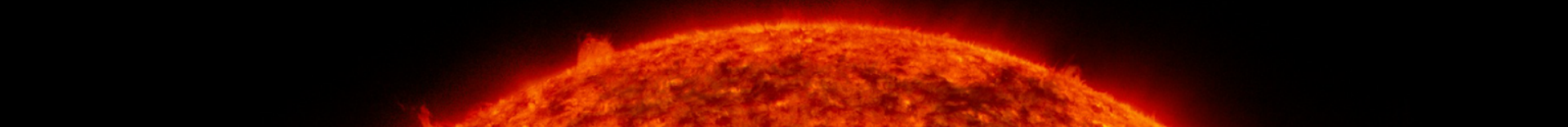
Thank You!

NASA PARTNER

FDL X HELIO

Google Cloud

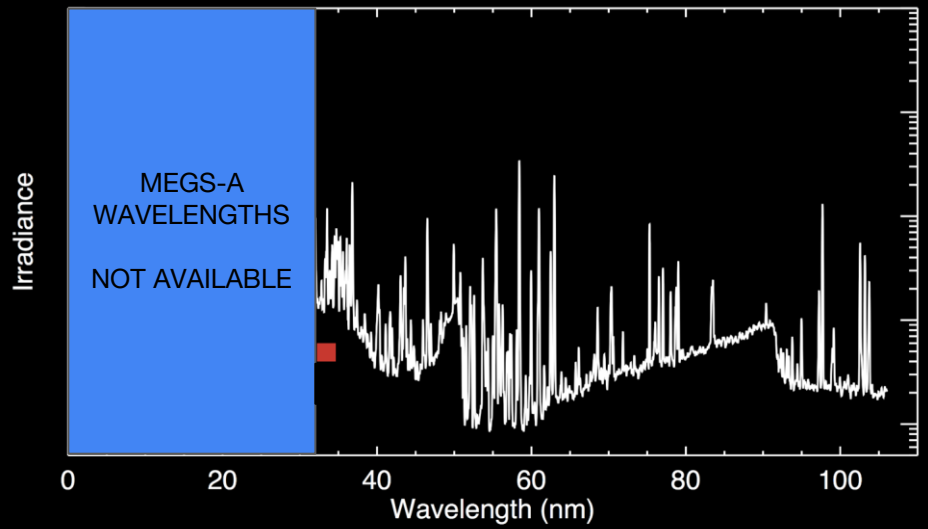
NVIDIA



BACK-UP

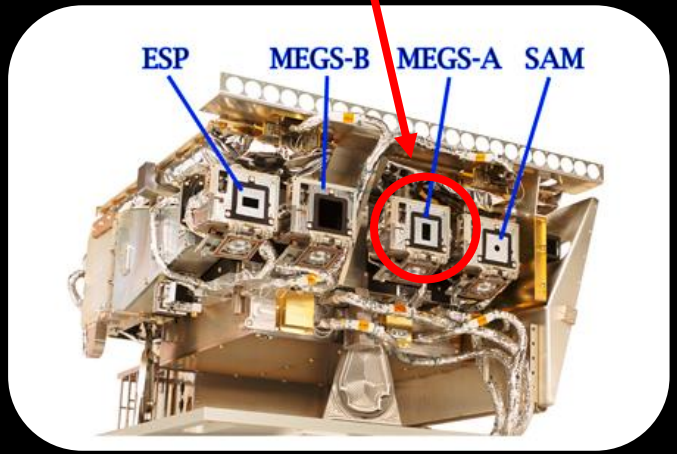
A Case Study: EVE Instrument Virtualization

EVE: How it's going

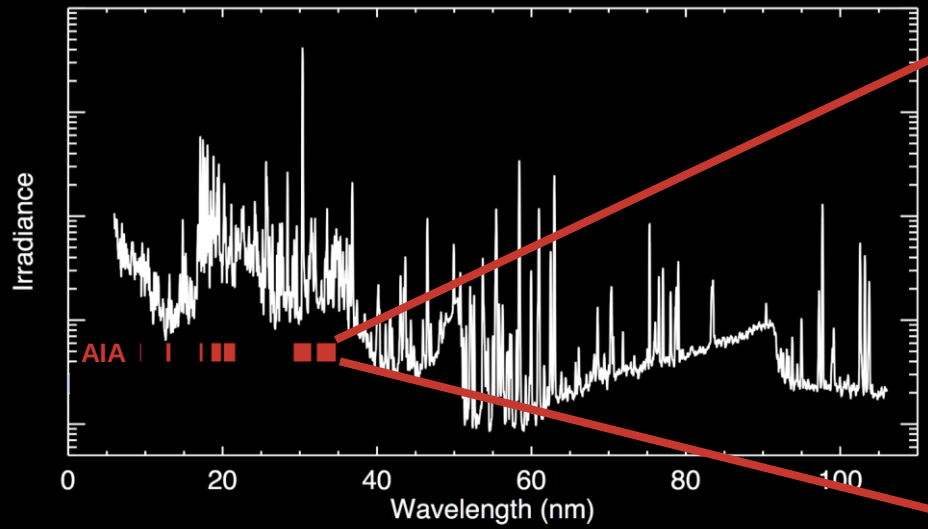


EVE Instrument

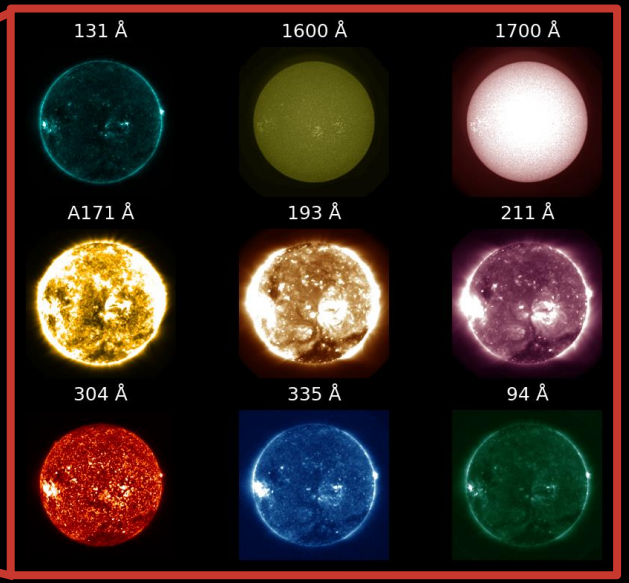
Partly broken since 2014
(Capacitor short in MEGS-A)



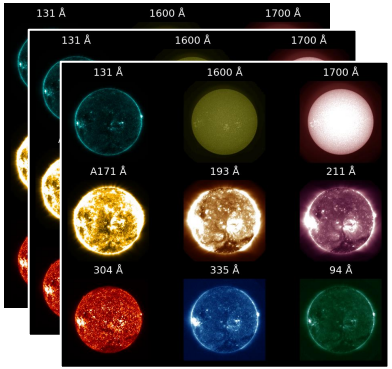
EVE: AIA to the Rescue



Our dataset



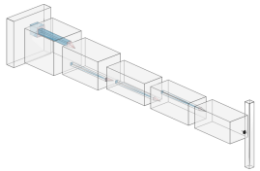
Virtual EVE: Hybrid Model



LINEAR MODEL

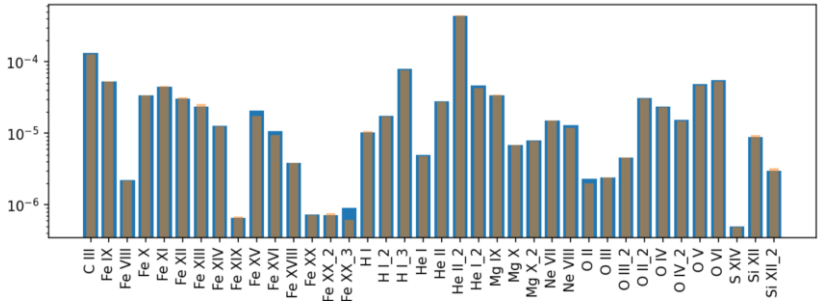
+

CNN MODEL



(EfficientNet backbone)

Irradiance Prediction



	Previous work (FDL 2018-2022)	This work (FDL 2023)
# ions	14	**38**
# wavelength	8	**9**
prediction	historical	**live inference**

Not Secure | sdomldemo.org

Virtual Eve

NASA

AD ASTRA

PER ALGORITHMOS

GOOGLE

NVIDIA

2023

FDL X HELIO

Welcome to the FDL-X SDO Live Virtual EVE Demo. This app is designed to showcase the capabilities of the FDL-X SDO computational platform.

Please Select Date Range

Start Date

2017/09/26

Begin Time:

00:00

End Date

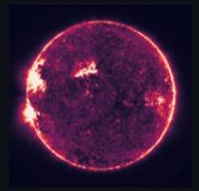
2017/09/27

End Time:

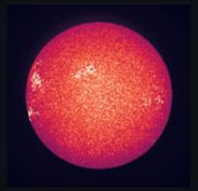
23:59

AIA Data

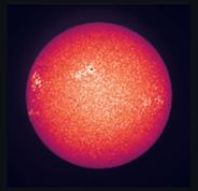
131A



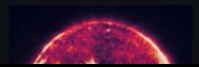
1600A



1700A



171A



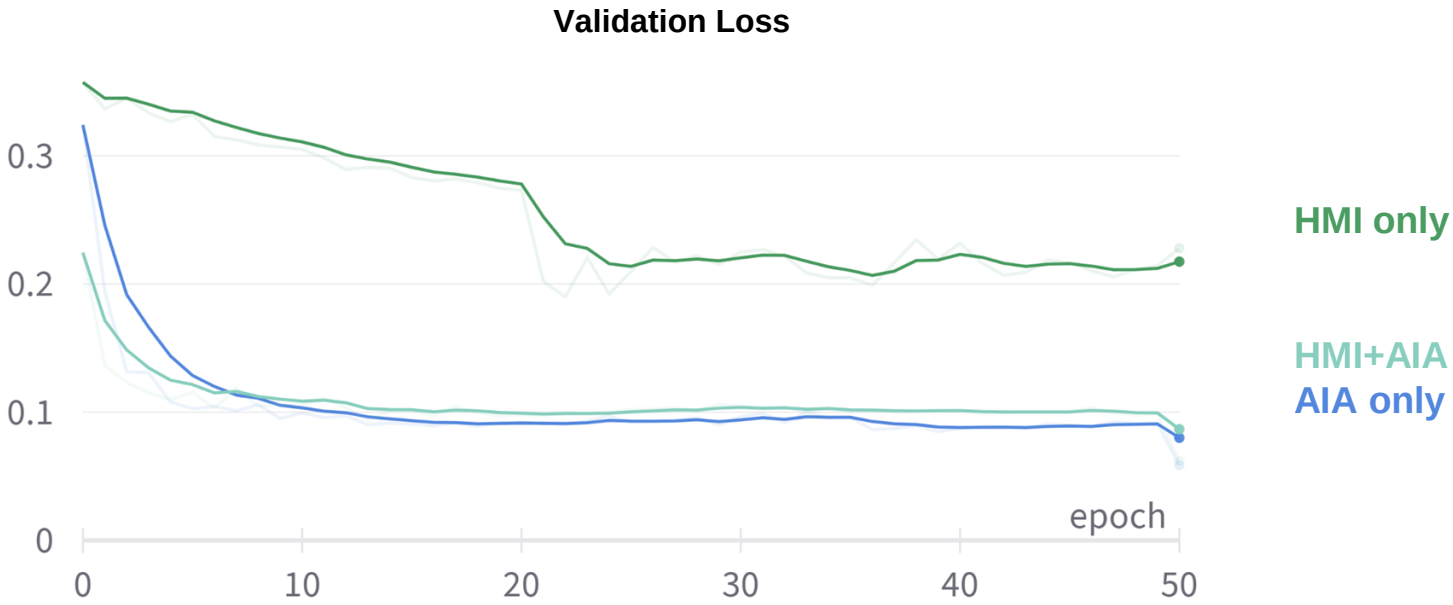
193A

211A

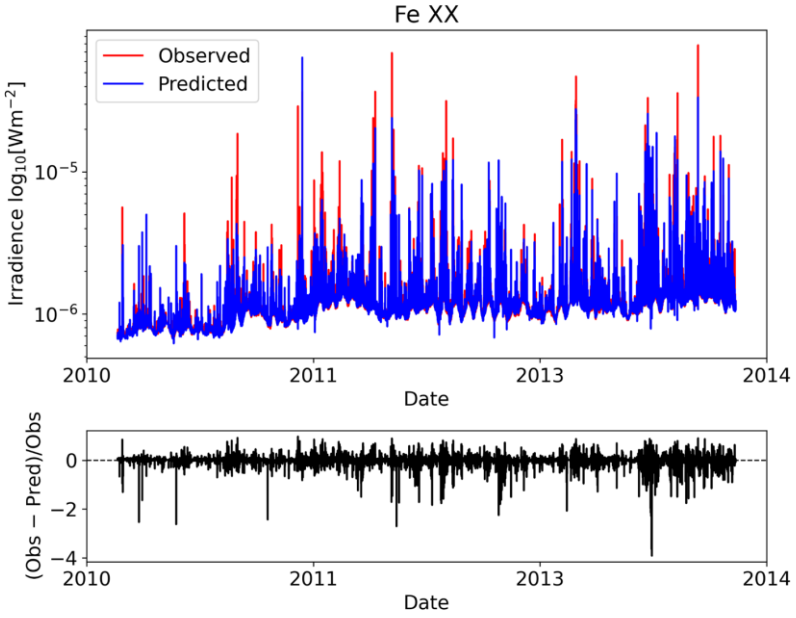
Virtual EVE Irradiance

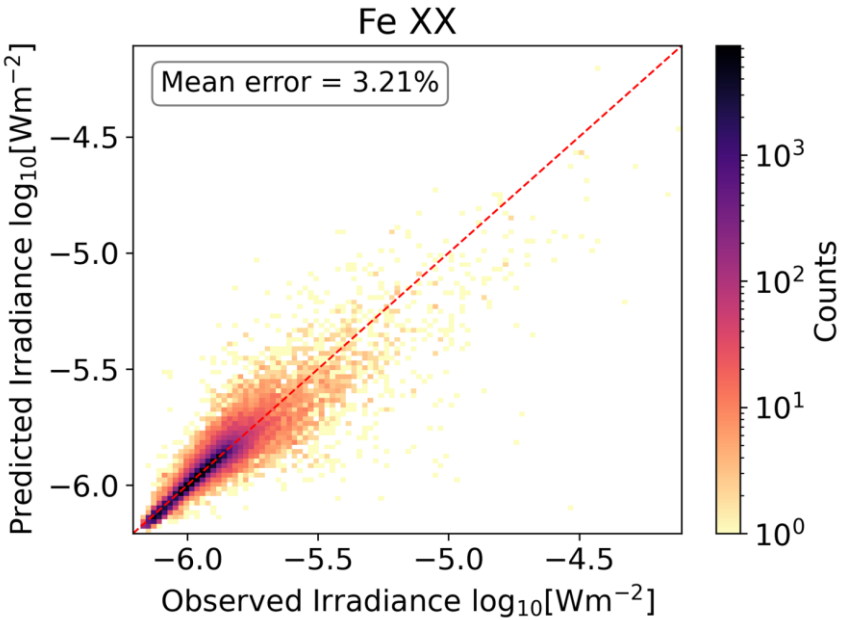
Virtual EVE Results: AIA is All You Need

HMI line-of-sight data ***do not*** improve irradiance prediction quality



Virtual EVE Results: Irradiance predictions





Virtual EVE Results: Irradiance predictions 2020

- Results show irradiance prediction vs observation (MEGS-B), 6 years after training data ends
- Shape and trend looking good, but there's a systematic offset
- Degradation corrections have been applied, but perhaps the degradation correction is slightly off?