

LST1 data analysis & Deep Learning

Thomas Vuillaume, 01/12/2020

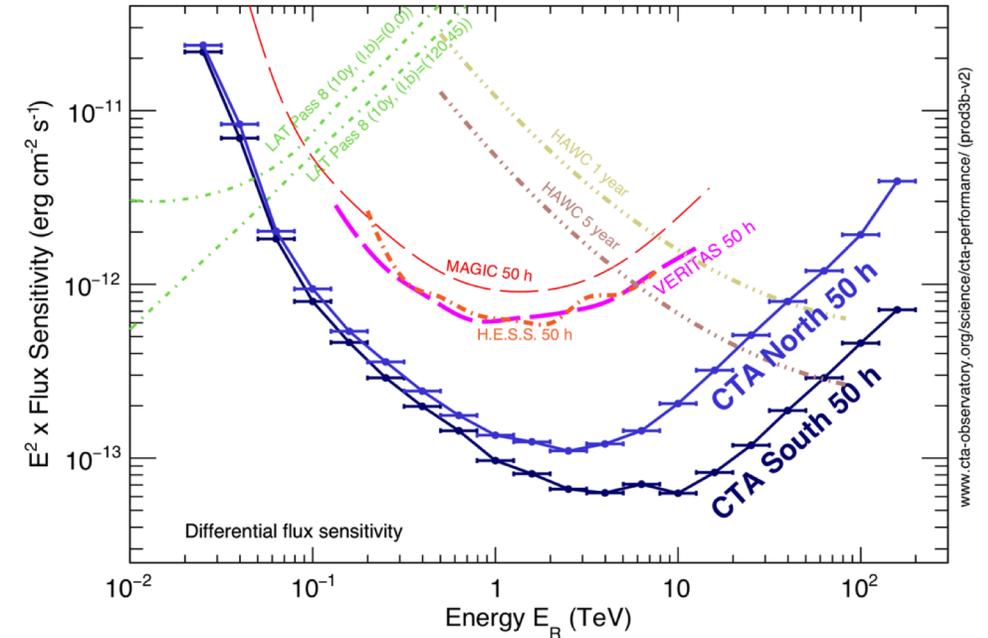
Analyse de données

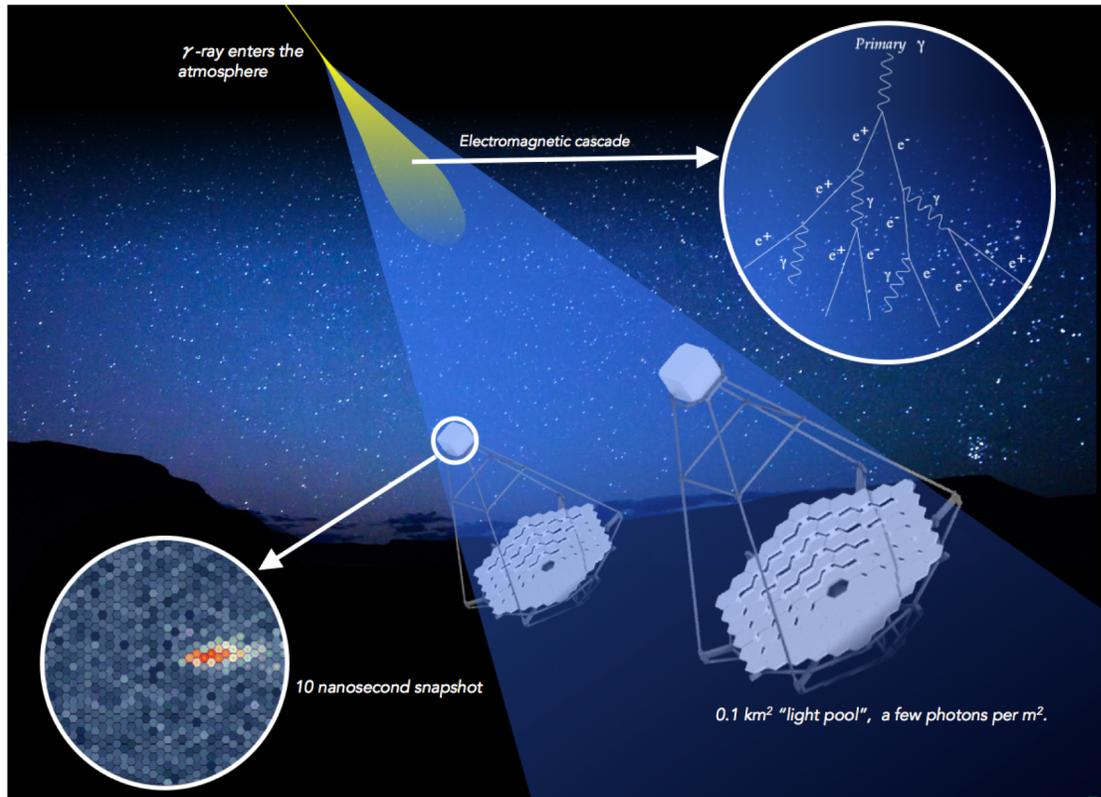
Des simulations aux données réelles du LST1

- 2 sites et **118 télescopes**
- **Sensibilité x10**
- Observatoire interconnecté

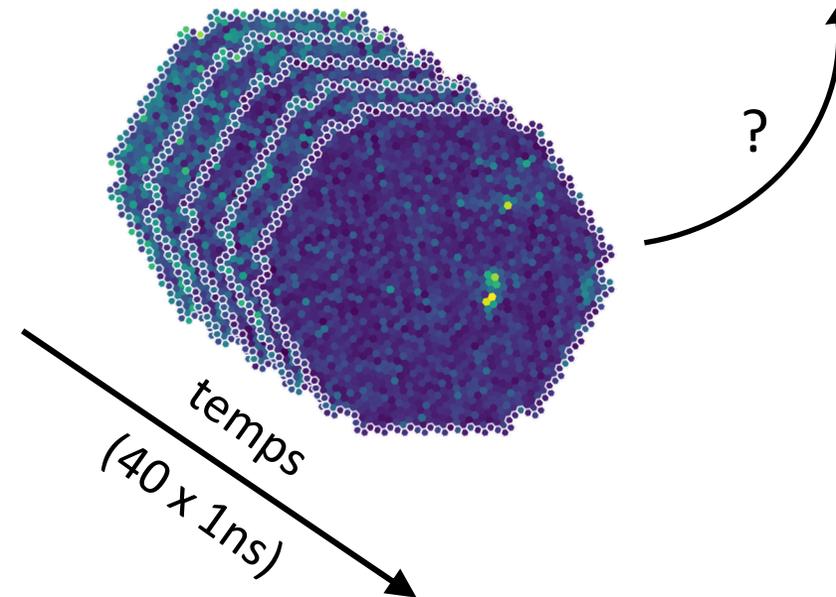
Défis:

- Différentes types de caméras et télescopes
- Flux de données : **5 GB/s/caméra**
- Volume de données: **3 PB/an** après réduction

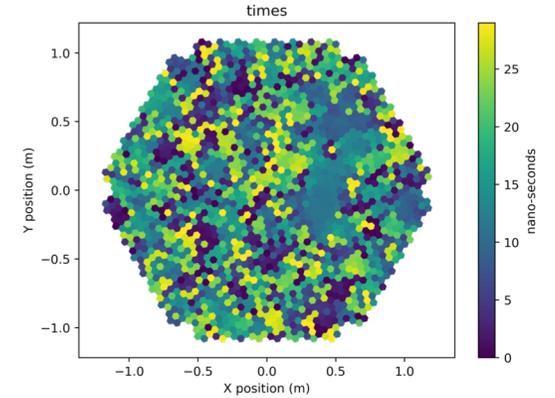
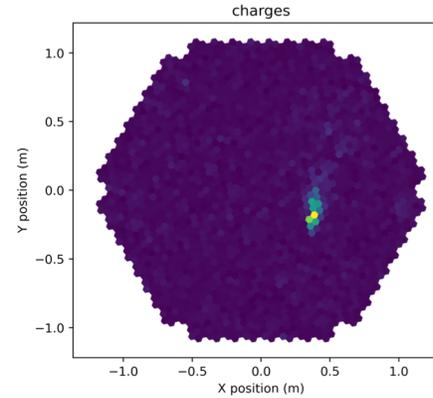
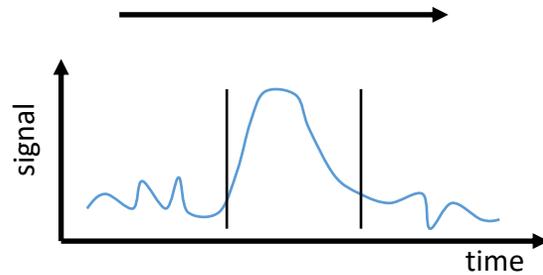
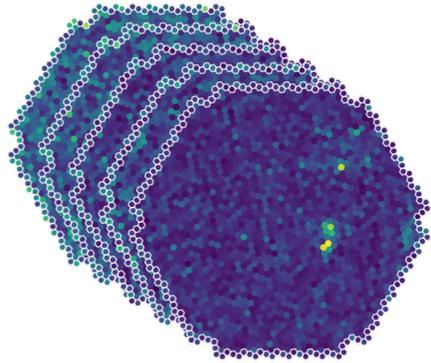




- Évènement gamma
- $E = 0,189 \text{ TeV}$
- Direction = (1,23; 6,22)rad

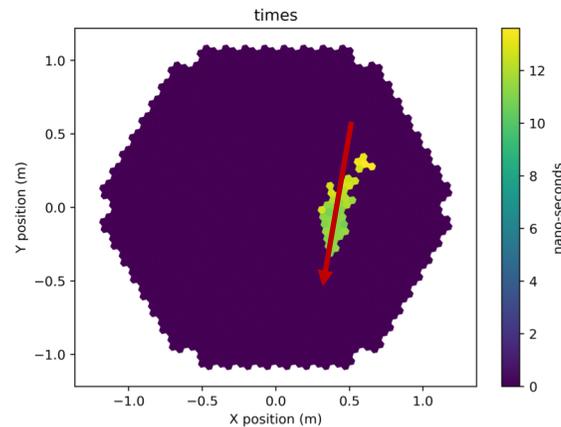
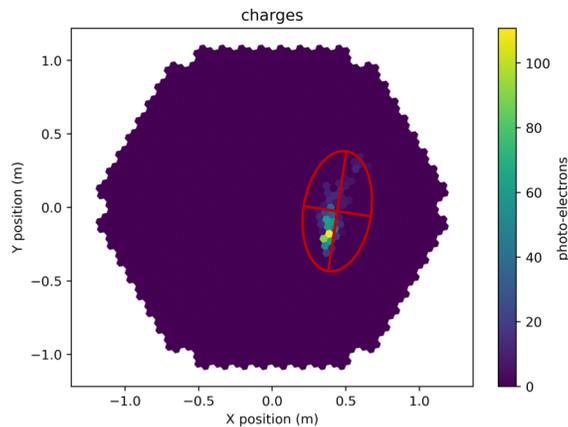


1. Calibration, Intégration

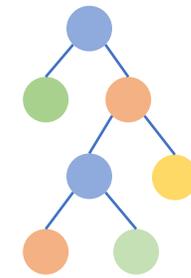


2. Traitement du signal : Nettoyage du bruit, Extraction de paramètres d'image

3. Arbres de décisions + simulations Monte-Carlo → Paramètres physiques



Paramètres d'image



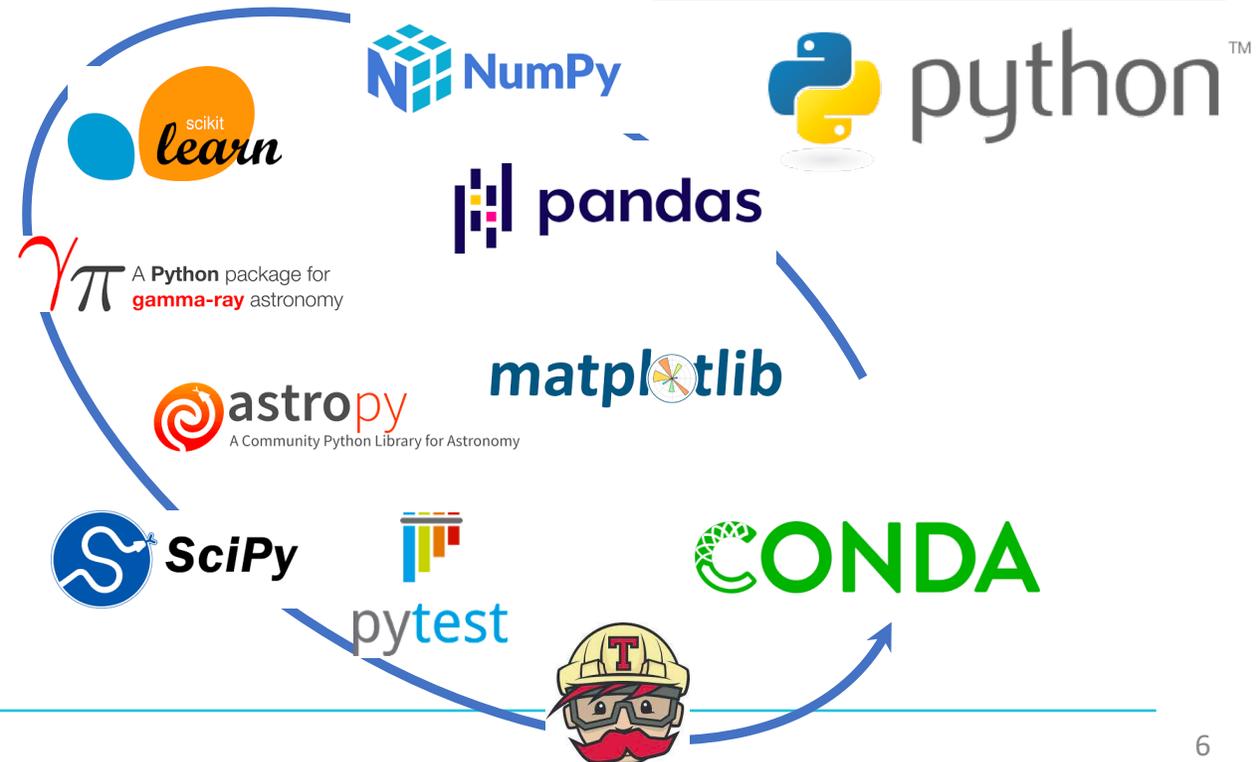
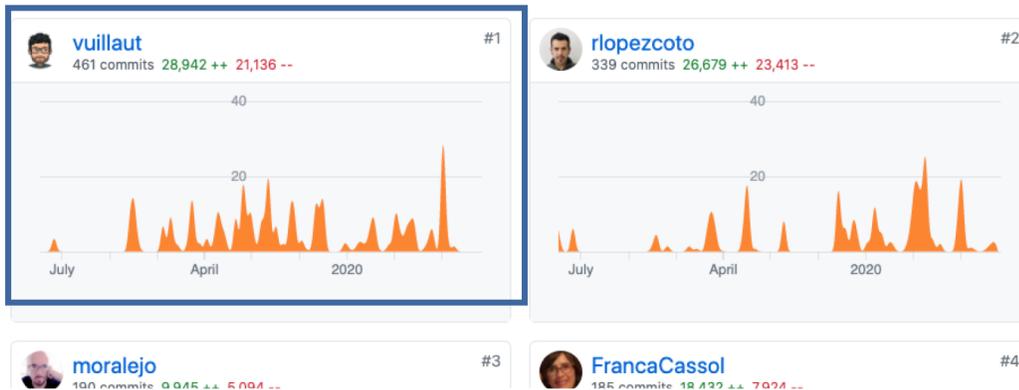
$E = 0,189 \text{ TeV}$

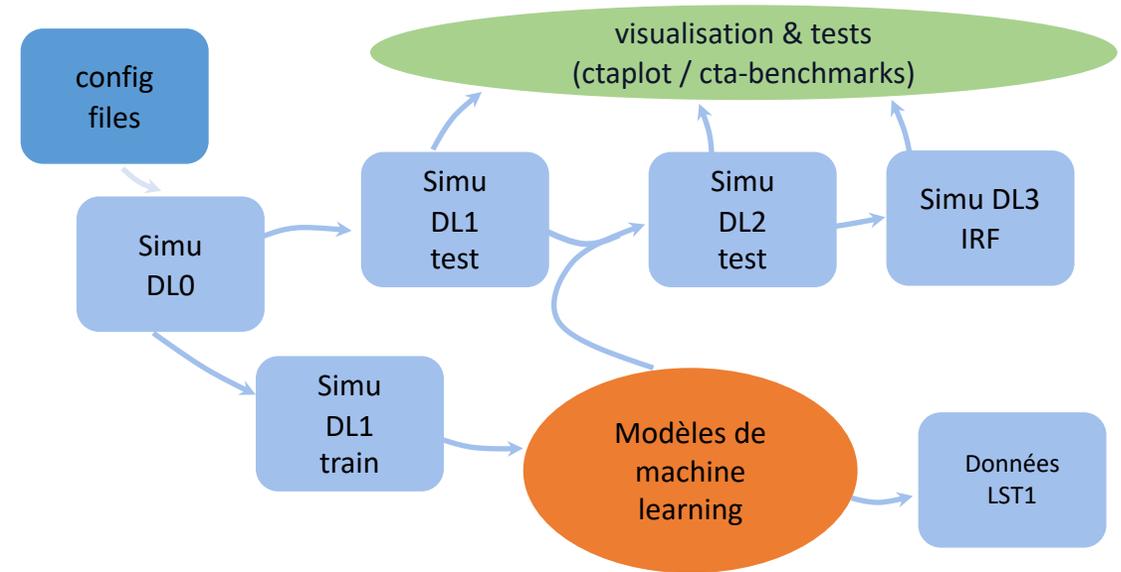
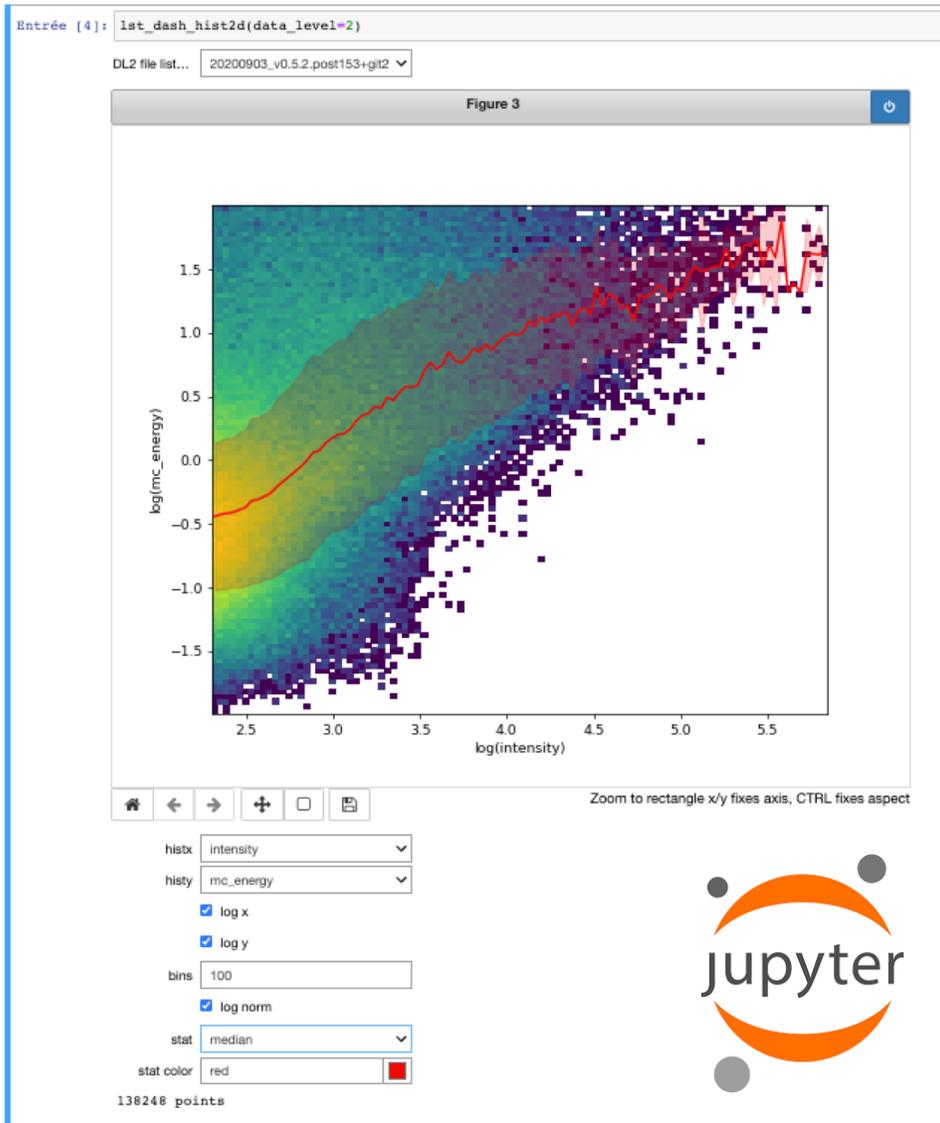
- Développement collaboratif (git) de bibliothèques Python pour l'analyse de données (>200 utilisateurs)
- Importante contribution à *Istchain*
- Revue des contributions (GitHub)



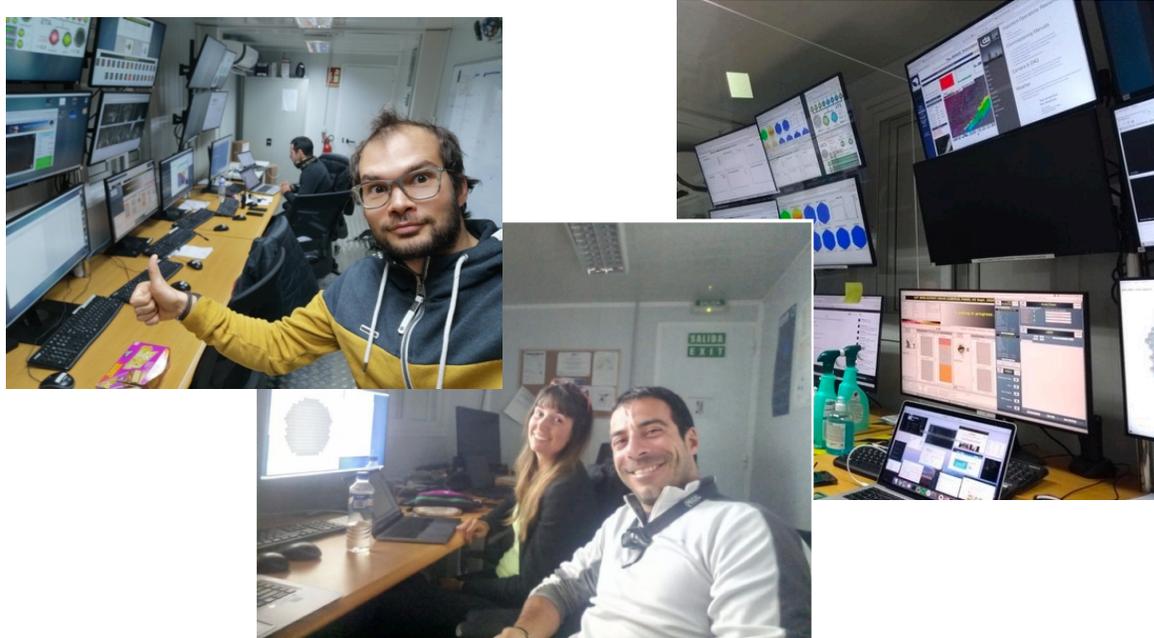
May 20, 2018 – Sep 18, 2020

Contributions to master, excluding merge commits

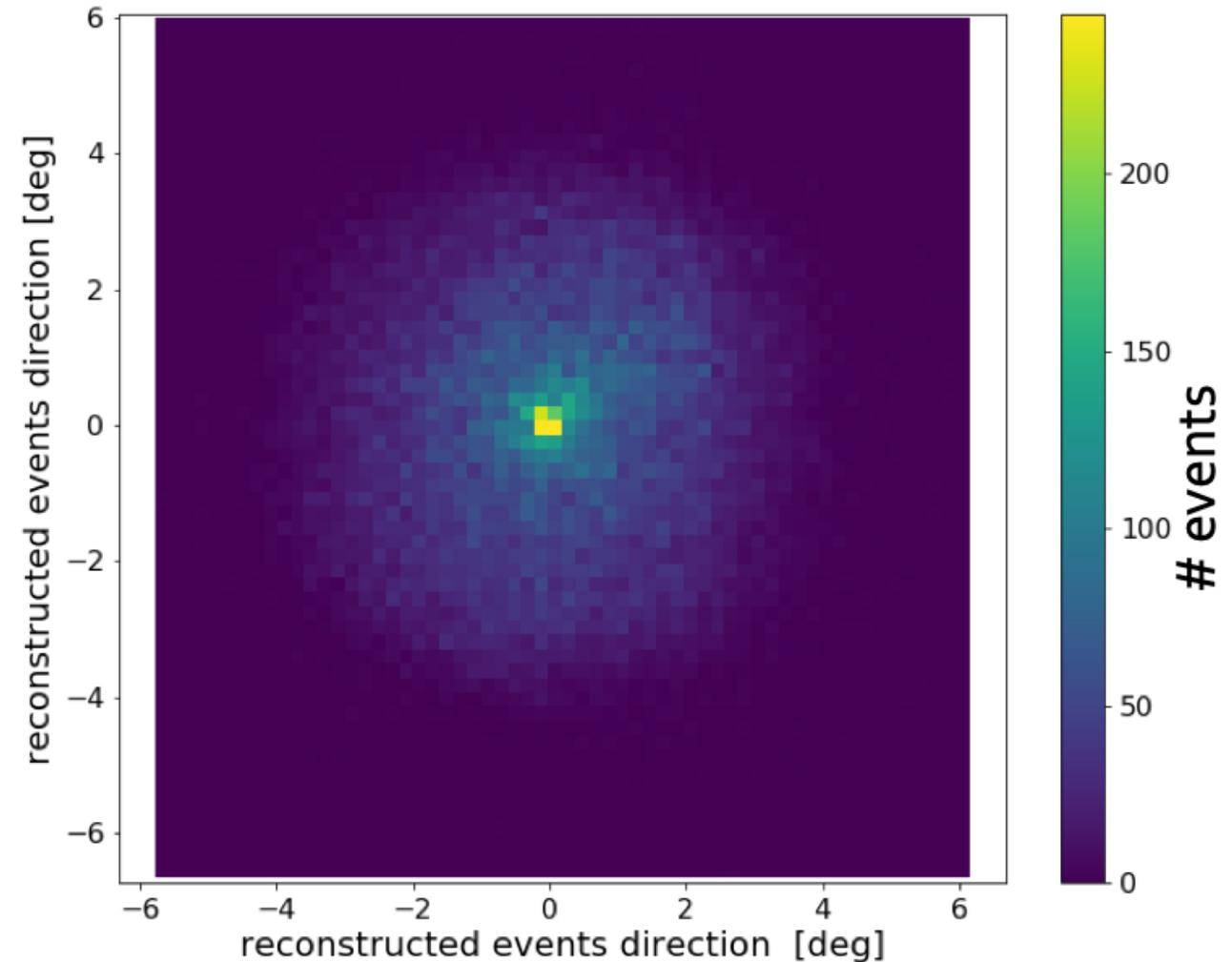




- mise en production sur la **ferme de calcul**
- Entraînement des arbres de décision
- Outils de **visualisation** et de test
 - À distance sur le serveur de calcul
 - Interface graphique (**jupyter + widgets**)



- Mise en **application** de Istchain
- **Shifts** à La Palma
- **Observation et analyse** des premières sources avec le LST1



GammaLearn

Établir des méthodes d'analyse plus performantes



Laboratoire
d'informatique
USMB

- P. Lambert
- A. Benoit

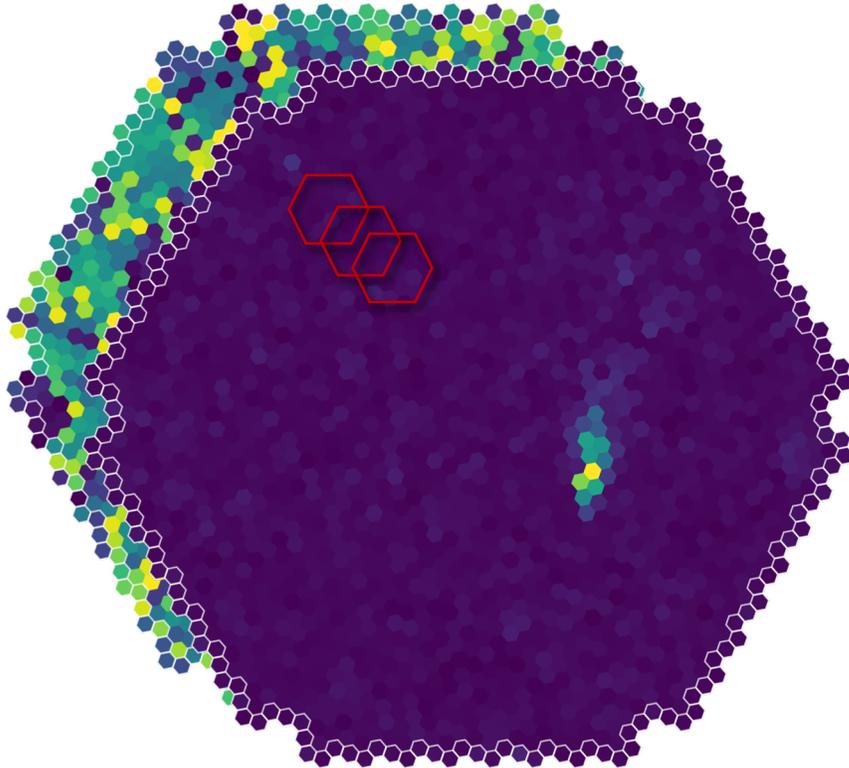


LISTIC



- T. Vuillaume (PI)
- M. Jacquemont (PhD)
- G. Maurin



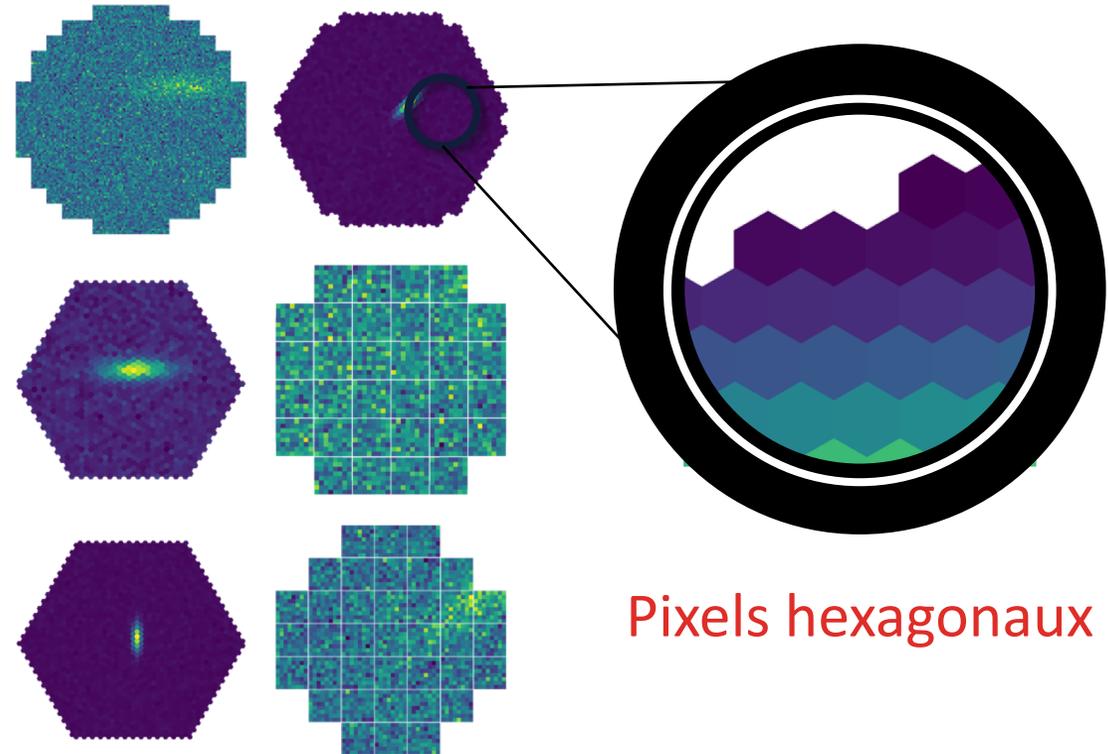


Deep learning

- réseaux de neurones convolutionnels profonds
- pas ou peu de preprocessing
 - garder toute l'information des pixels
 - analyser des images très bruitées
- rapidité d'exécution
 - candidat pour l'analyse online

) ↗ sensibilité

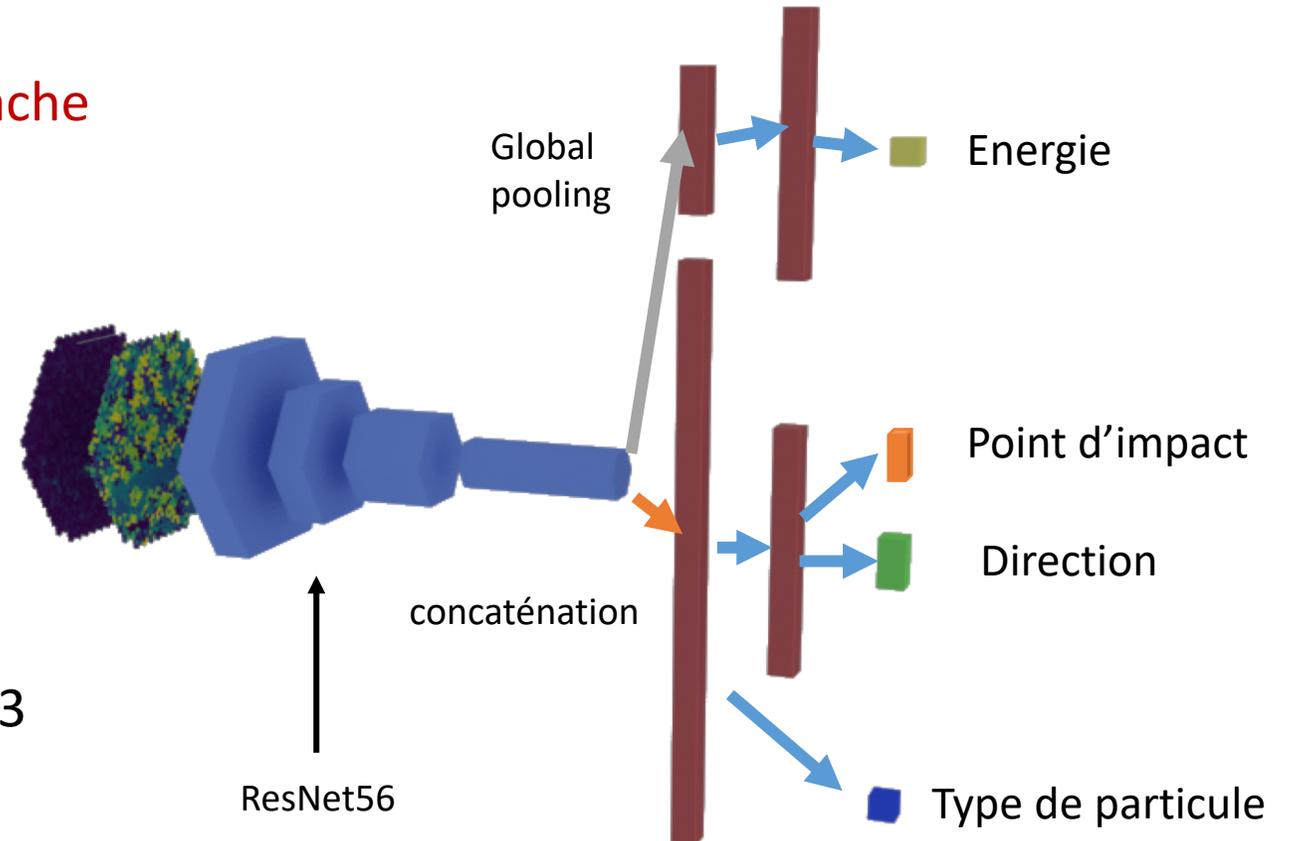
- Framework traditionnels = images standard
 - Méthodes non adaptées
- Développement de **convolutions et pooling indexées**

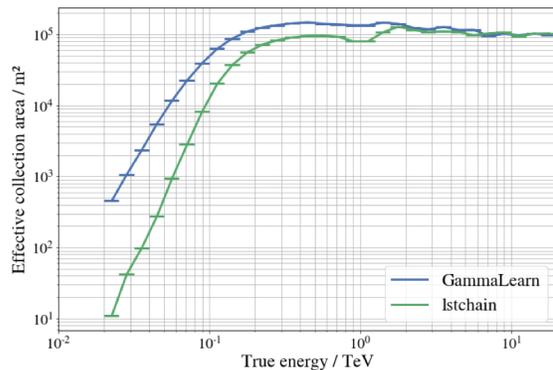
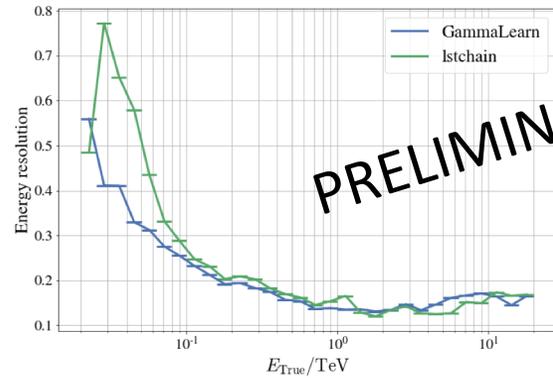
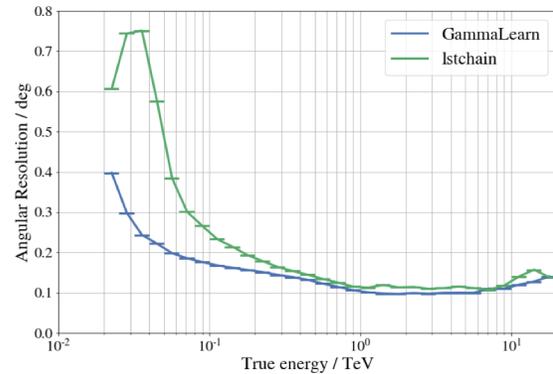


- **Librairie IndexedConv** open-source et **générique** (application à toute forme de capteur) pour **PyTorch**

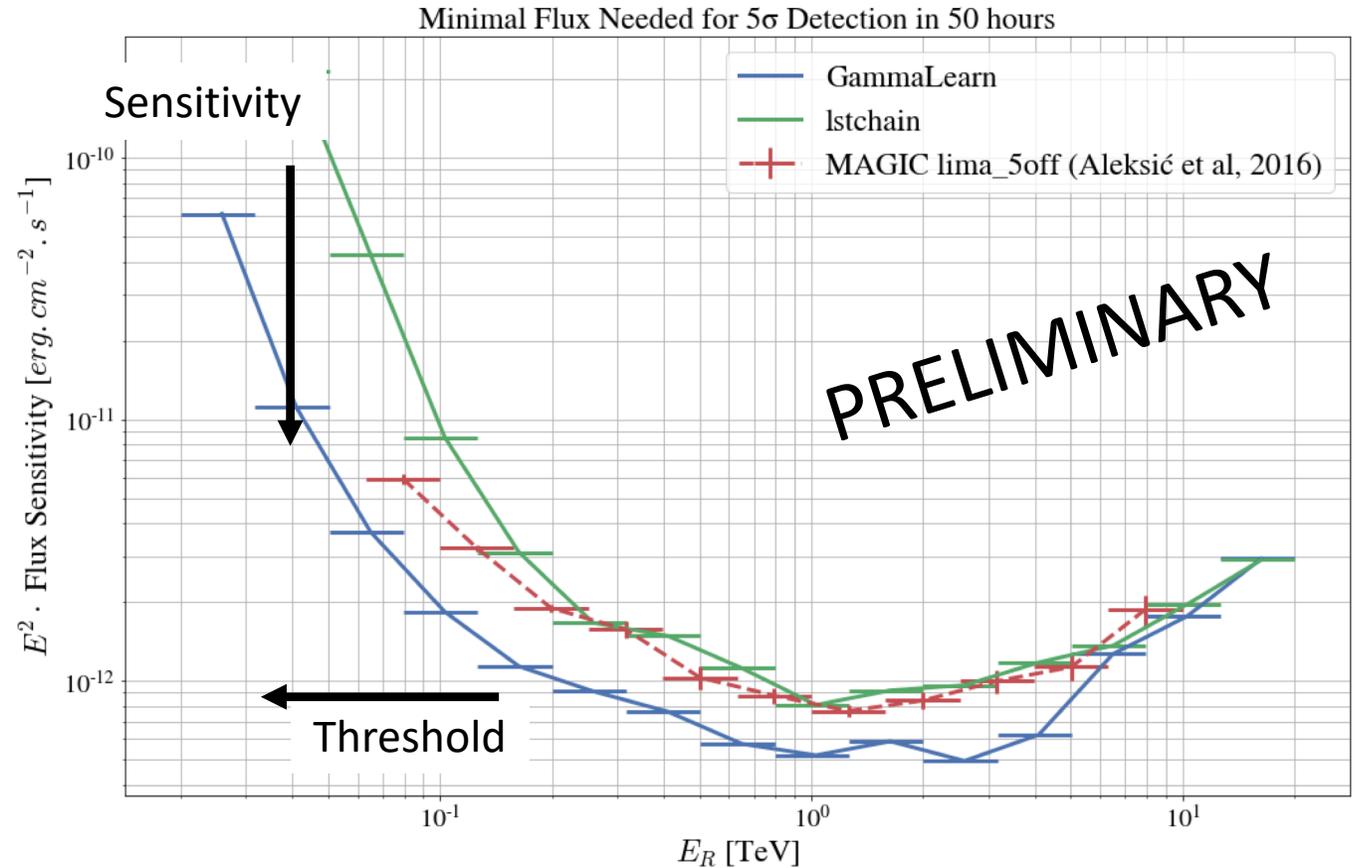


- Première reconstruction complète en deep learning
- Un seul réseau = apprentissage multi-tâche
 - Meilleures performances
 - Évite une partie de la dégénérescence
 - Plus rapide
- Architecture inspirée par la physique du problème
- Entraînement et optimisation sur différentes fermes HTC: MUST, CC-IN2P3 (~1500h), Jean Zay (8000h)





PRELIMINARY



- Extra-galactique, transients...
- Science des LST !

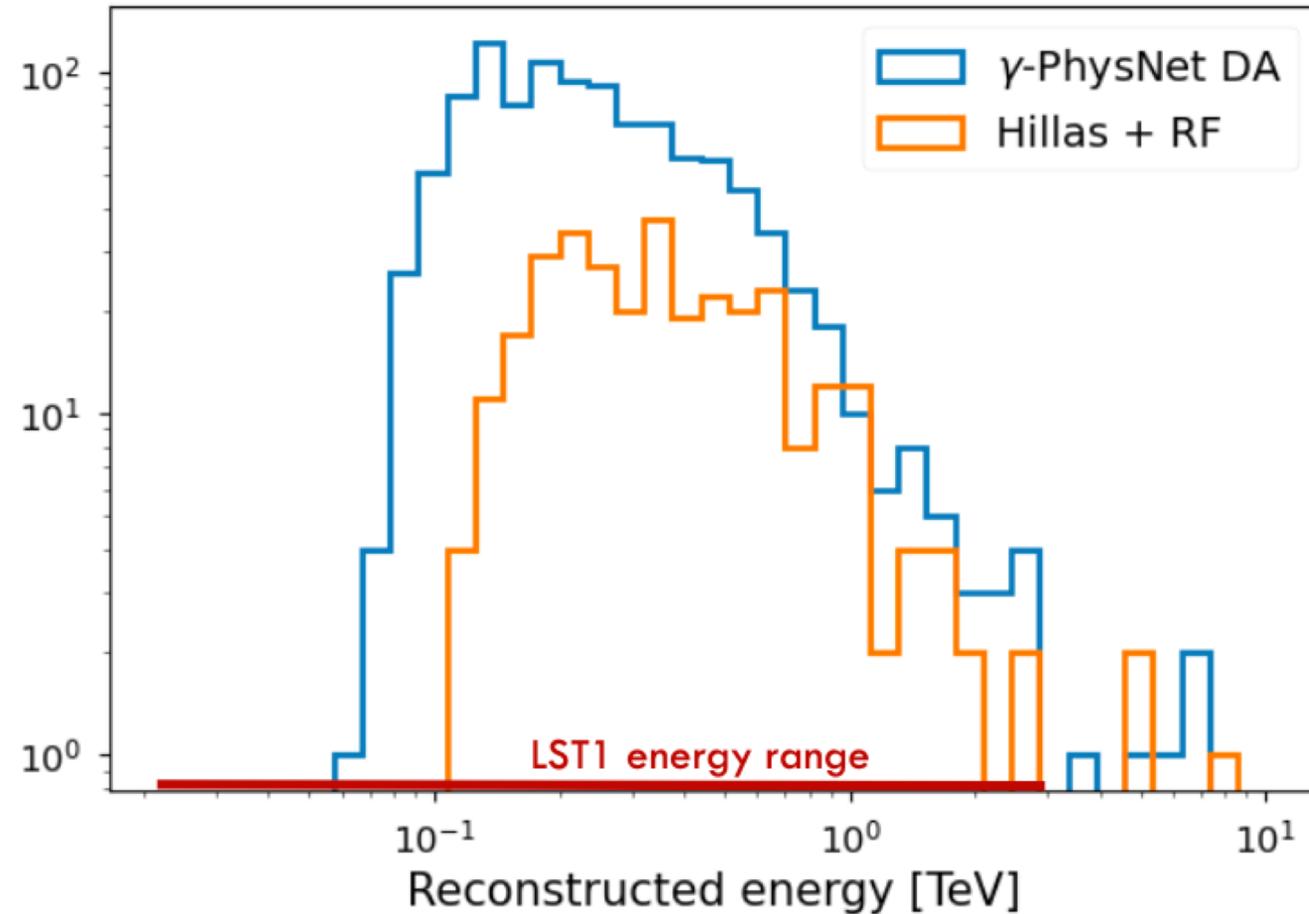
GammaLearn

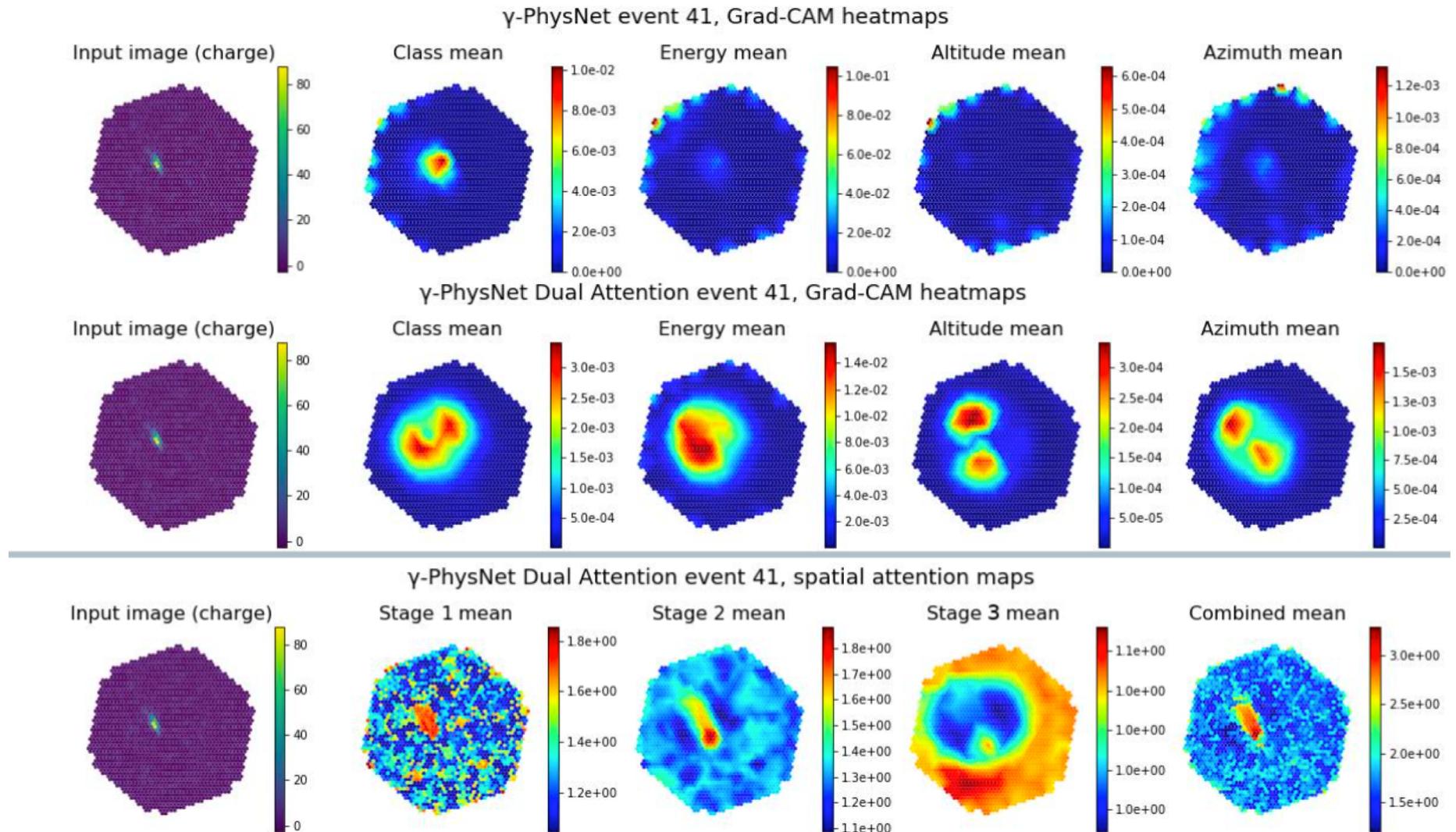
Istchain

	γ -PhysNet DA	Hillas + RF
Excess	608	210
Significance	15.0 σ	10.8 σ

- Premières analyses sur les vraies données confirment le gain observé sur les simulations

Reconstructed energy distribution (low cuts)





Projets européens H2020

ASTERICS & ESCAPE

- Projets européens H2020
 - Amener les grandes infrastructures de recherche en astronomie, astro-particule et physique des particules ensemble **pour relever les défis** communs liés à l'analyse de données et développer des outils communs
 - Promouvoir **l'Open Science**
 - Intégration à l'European Open Science Cloud (EOSC)
- Animation de groupes de travail
- Organisation de workshops (2017, 2018, 2019, 2020, 2021)
- **Organisation de 4 écoles d'été** visant à former les chercheurs et physiciens aux techniques et outils modernes de programmation et d'analyse de données
 - Organisateur principal de l'édition 2020, reportée à 2021 (COVID-19)
- **Responsable** du développement d'un *repository* commun pour les grandes infrastructures de recherche pour le futur d'EOSC



- Importante contribution au développement du pipeline et des outils d'analyse pour la collaboration
 - Maitrise de toute la chaine de production
 - Développement de techniques d'analyse innovantes avec un gain important pour la science
 - Sensibilité équivalente aux méthodes état de l'art en offline
 - Rapidité compatible pour l'analyse online
 - Contribution au développement de la communauté (workshops, enseignement, revue de code...)
-