

# Intelligence artificielle et recherche de lentilles gravitationnelles dans les grands relevés astronomiques

## Application aux données du catalogue Gaia

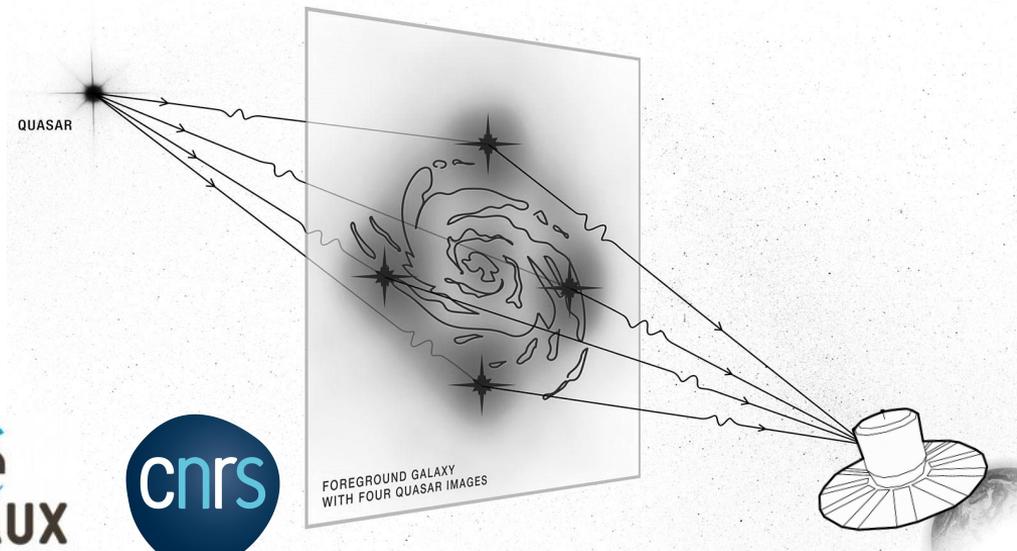


Quentin Petit<sup>1</sup>, Christine Ducourant<sup>1</sup>, Eric Slezak<sup>2</sup>,  
Dominique Sluse<sup>3</sup> & Ludovic Delchambre<sup>3</sup>

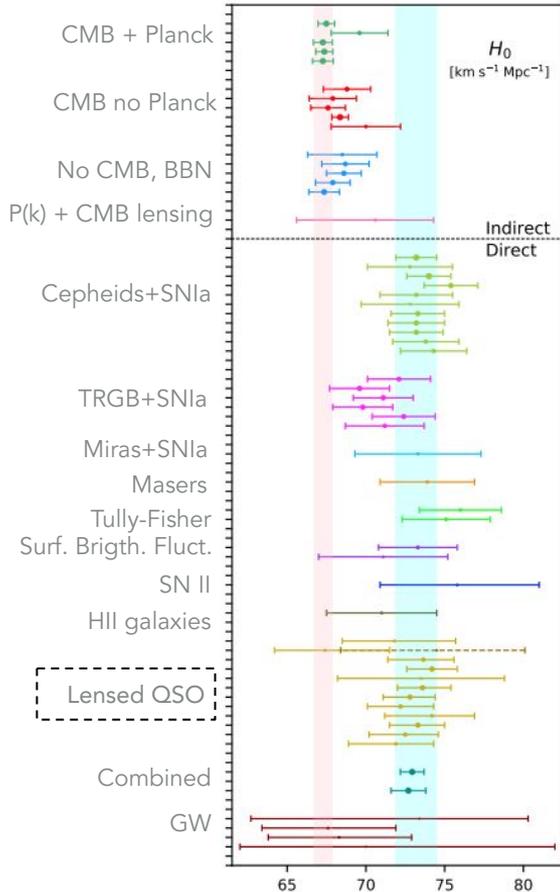
(1) Laboratoire d'Astrophysique de Bordeaux, France

(2) Observatoire de la Côte d'Azur, France

(3) Université de Liège, Belgique



# La tension de Hubble



**$H_0 \sim 67.4 \pm 0.5$   
km s<sup>-1</sup>Mpc<sup>-1</sup>**

**$H_0 \sim 73.2 \pm 1.4$   
km s<sup>-1</sup>Mpc<sup>-1</sup>**



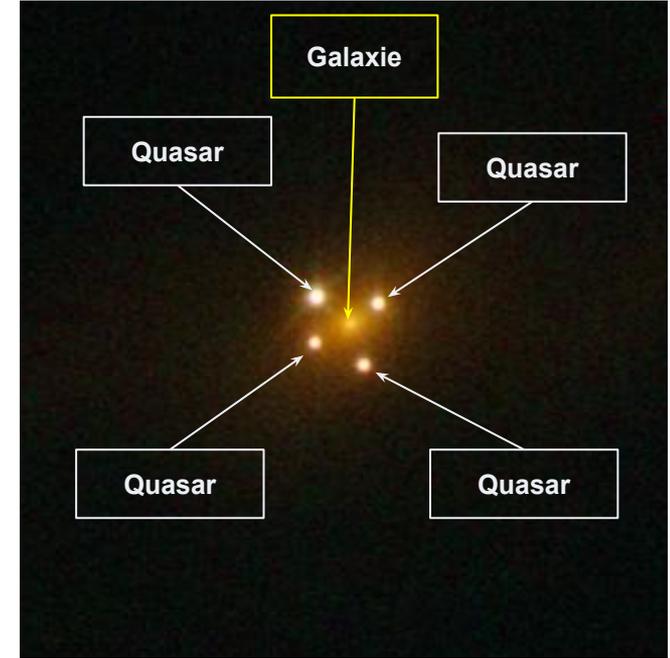
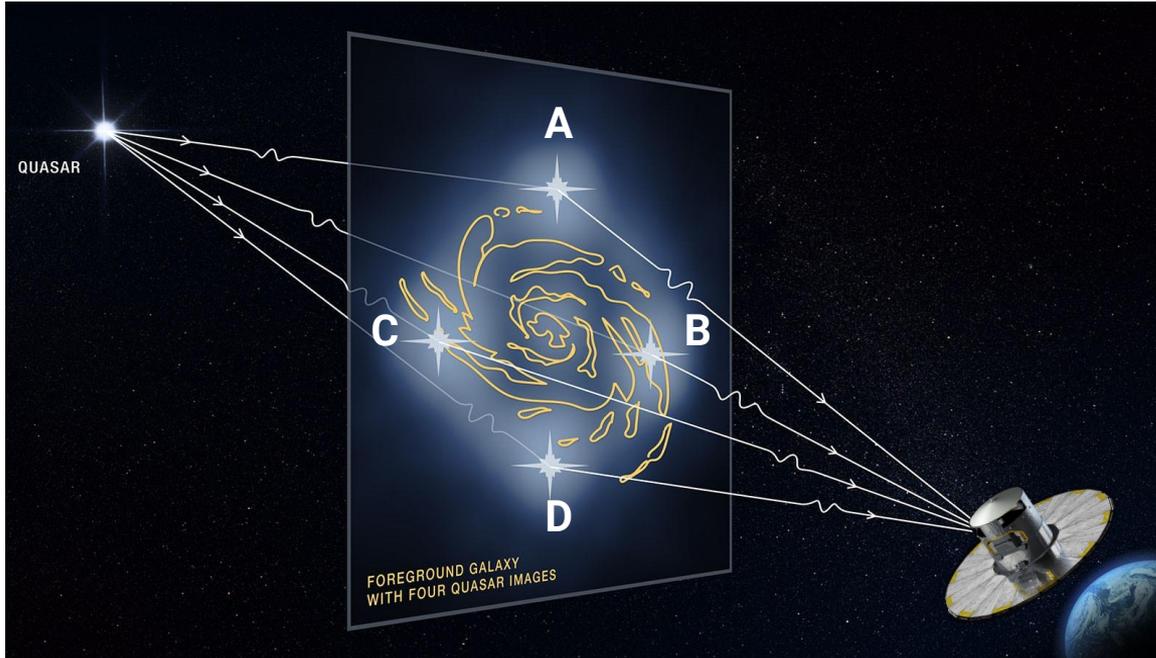
Credit : Impression d'artiste montrant ULAS J1120+0641

QSO = Quasi-Stellar Object / Quasar

Galaxies dont le noyau est une région extrêmement énergétique.

**Comment trouver ces QSO lentillés ?**

# Qu'est-ce qu'une lentille gravitationnelle ?



*Credit : Hubble Space Telescope*

La présence d'une masse déforme l'espace-temps environnant.

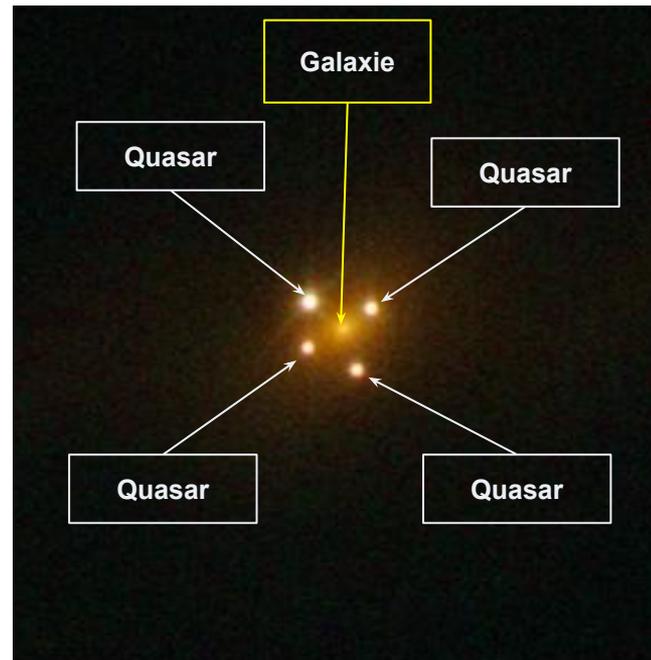
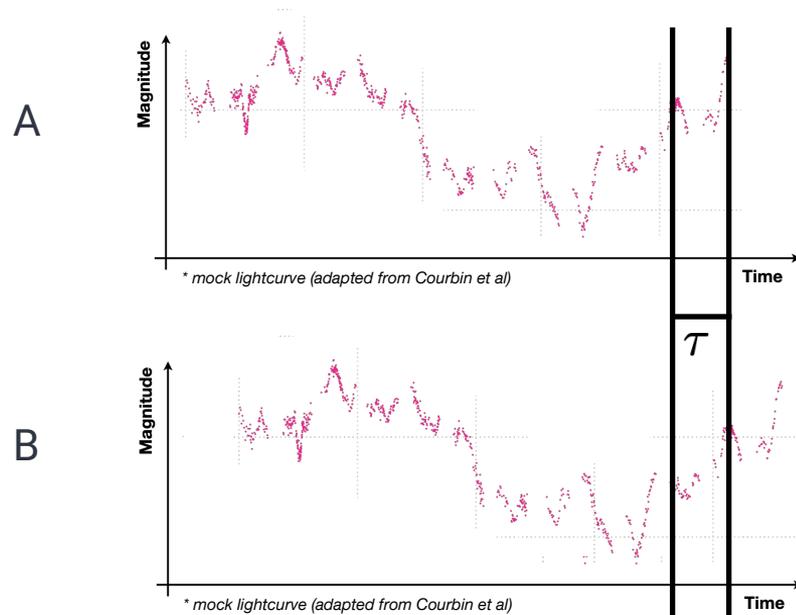
La lumière va suivre la géodésique.

**Un quasar lentillé à quatre images est appelé un quad.**

# Mesurer la constante de Hubble $H_0$

Les quasars sont des sources variables.

Séries photométriques pour mesurer le délai temporel :



Credit : Hubble Space Telescope

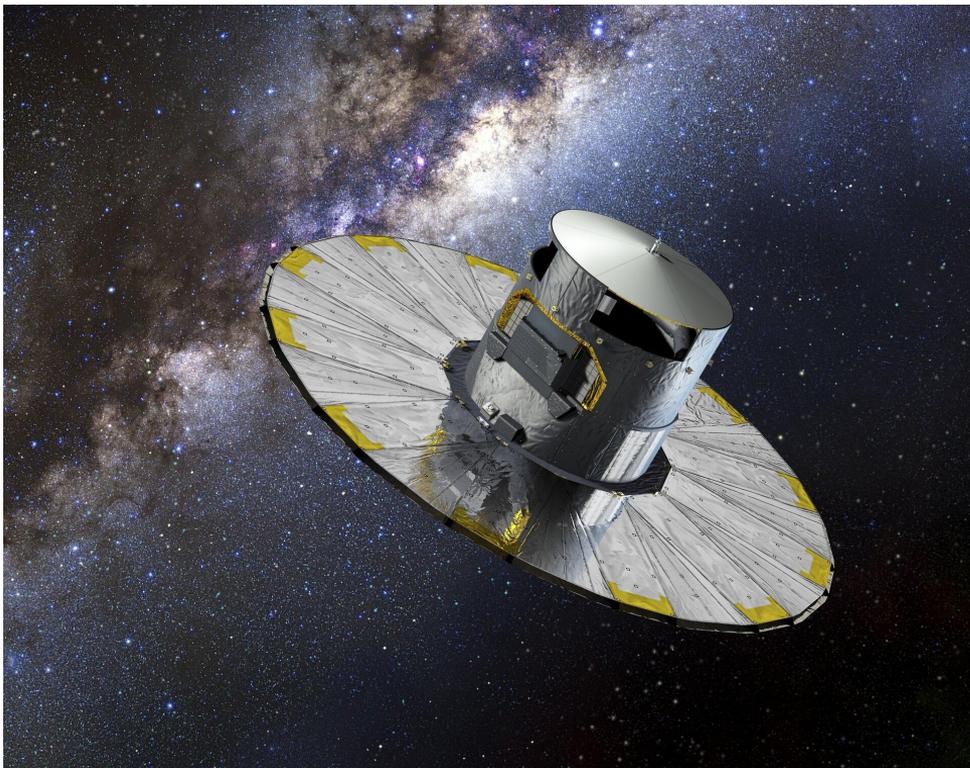
Redshift de la lentille

$$\tau = \frac{(1 + z_L) d_L d_S}{H_0 d_{LS}} \left[ \frac{1}{2} |\vec{\theta} - \vec{\beta}| - \psi_{2D}(\vec{\theta}) \right]$$

Distance Lentille - Source

# Le satellite Gaia : un découvreur de lentilles

---



- Gaia est lancé le 19 décembre 2013 ;
- Gaia est un satellite d'astrométrie avec une résolution exceptionnelle jusqu'à 300 mas.

Le catalogue final devrait contenir :

**~250 quads parmi 2 milliards de sources**

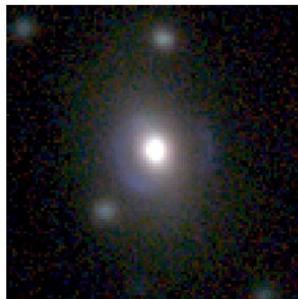
(Finet&Surdej, 2016)



**Comment les trouver ?**

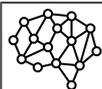
# La recherche de lentilles gravitationnelles à l'aide de l'IA

(Petrillo et al. 2017)



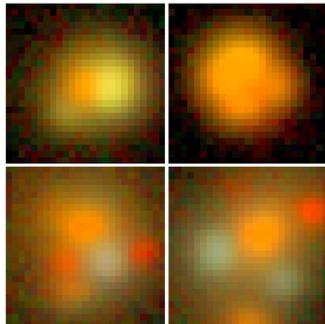
KSL427 (70)

Optimisation pour les images  
Dark Energy Survey (DES) images



Réseau de neurones

(DES Collaboration et al. 2021)



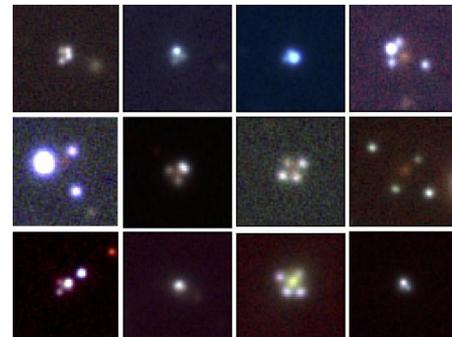
Bright

Dark Energy Survey images



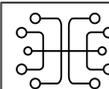
Réseau de neurones

(Delchambre et al. 2019)



Catalogues Gaia DR2

Découverte de 12 nouveaux  
quads.



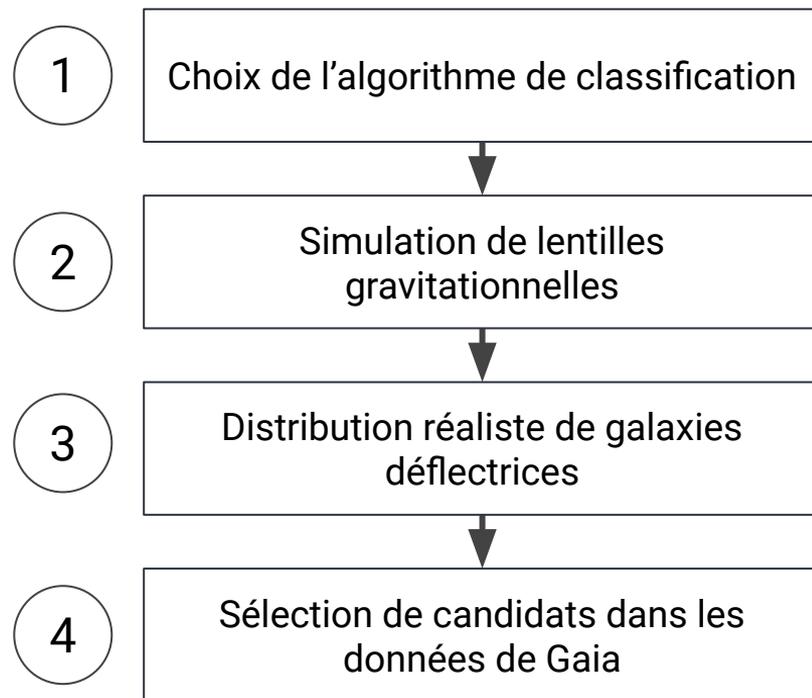
Arbres de décisions

Deux différences majeures avec ces études :

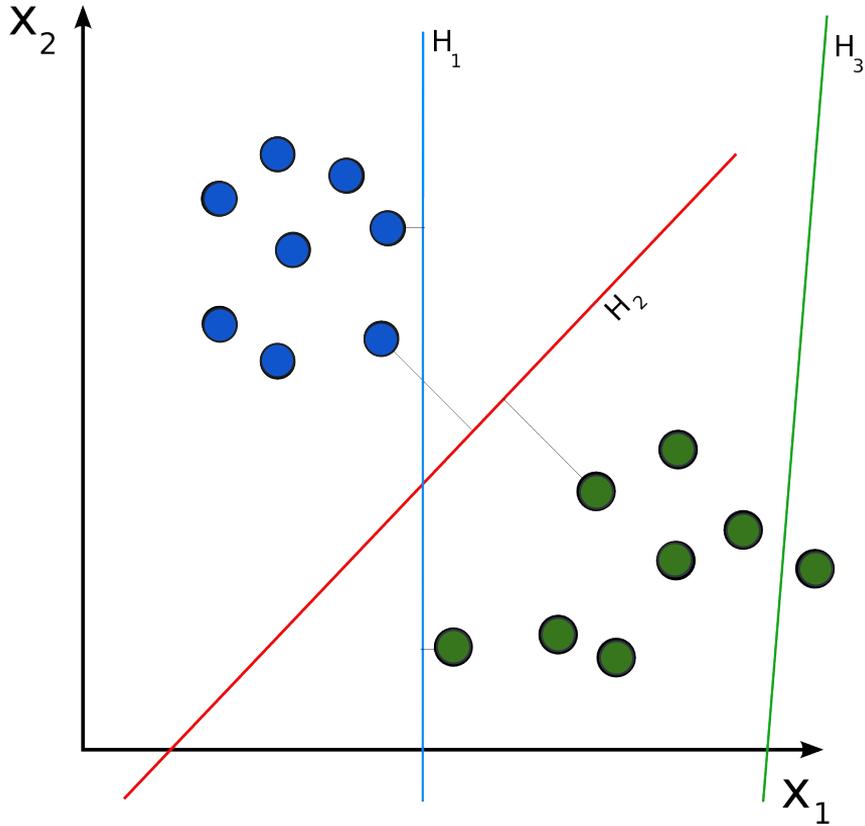
1. Simulations de lentilles réalistes à partir de **distributions non-uniformes** ;
2. Gaia est un satellite d'astrométrie = pas d'images mais une meilleure résolution spatiale.

# Plan

---



# La classification automatique



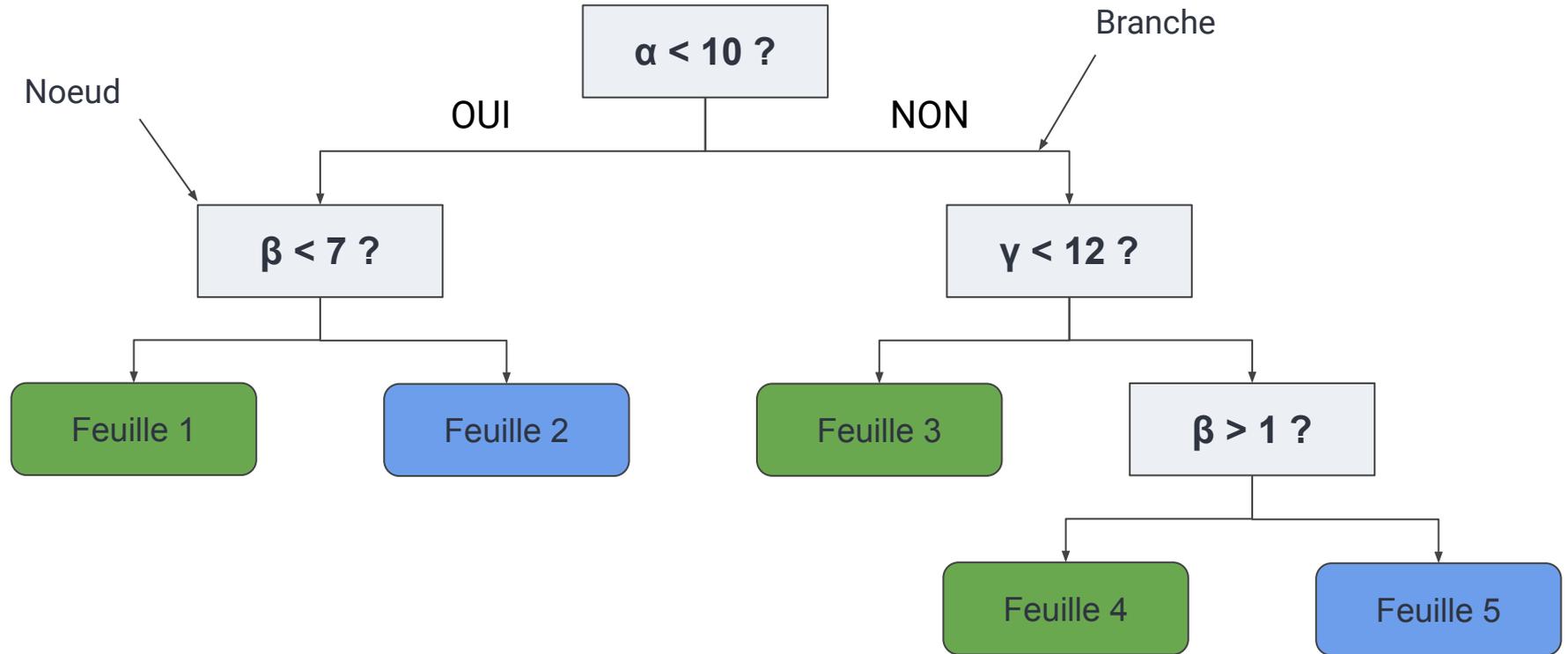
L'apprentissage automatique permet de développer des **modèles prédictifs**.

Une fois entraîné, le modèle permet de faire des prédictions sur des données **similaire** avec lesquelles il n'a jamais été entraîné.

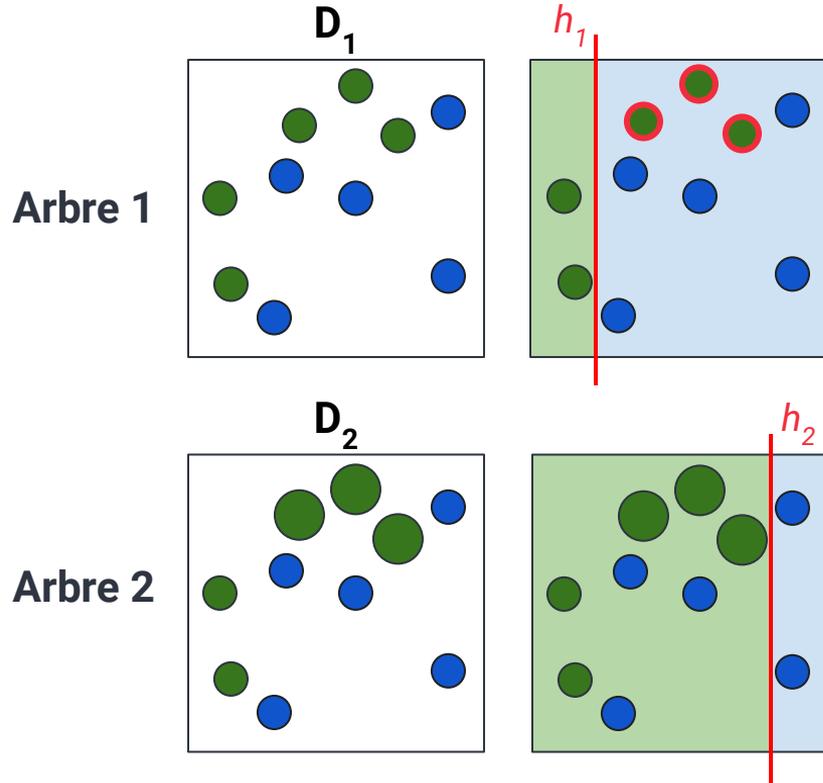
Supervisé

Les données sont préalablement annotées.

# Les arbres de décisions



# Principe du boosting et de la descente de gradient



Le terme "Boosting" fait référence à **une technique d'ensemble** où plusieurs modèles faibles (dans ce cas, des arbres de décision) sont combinés pour former un modèle fort.

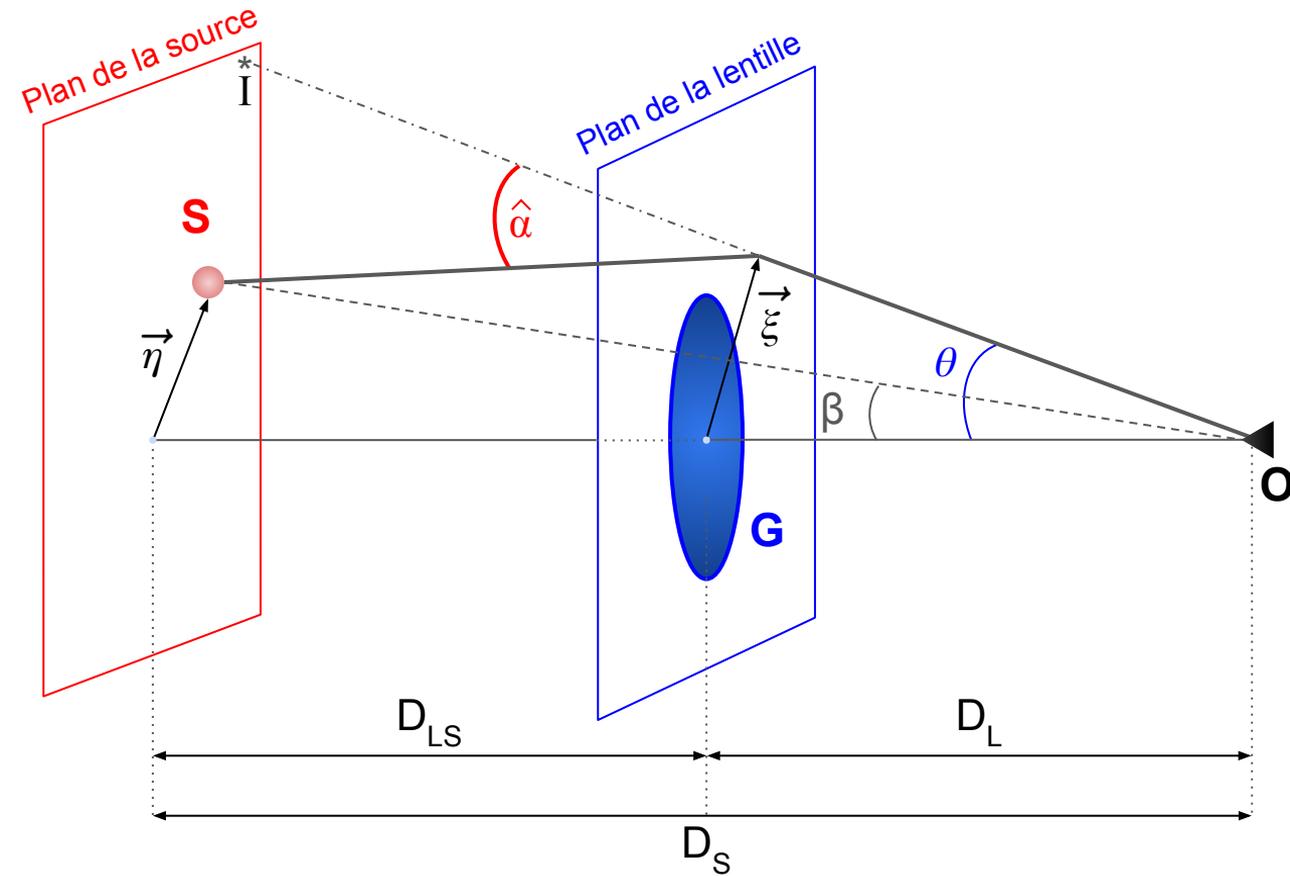
XGBoost utilise une fonction de coût. Il optimise cette fonction de coût en ajustant les prédictions à chaque étape.

**= descente de gradient**



Pour entraîner XGBoost à classer des configurations de quads, il nous faut un catalogue d'apprentissage le plus réaliste possible.

# Géométrie d'une lentille



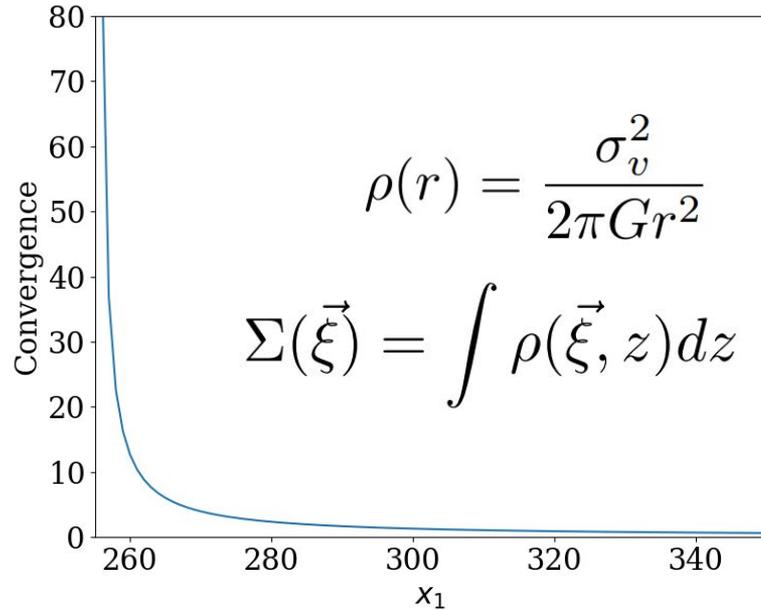
L'équation de la lentille :

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta})$$

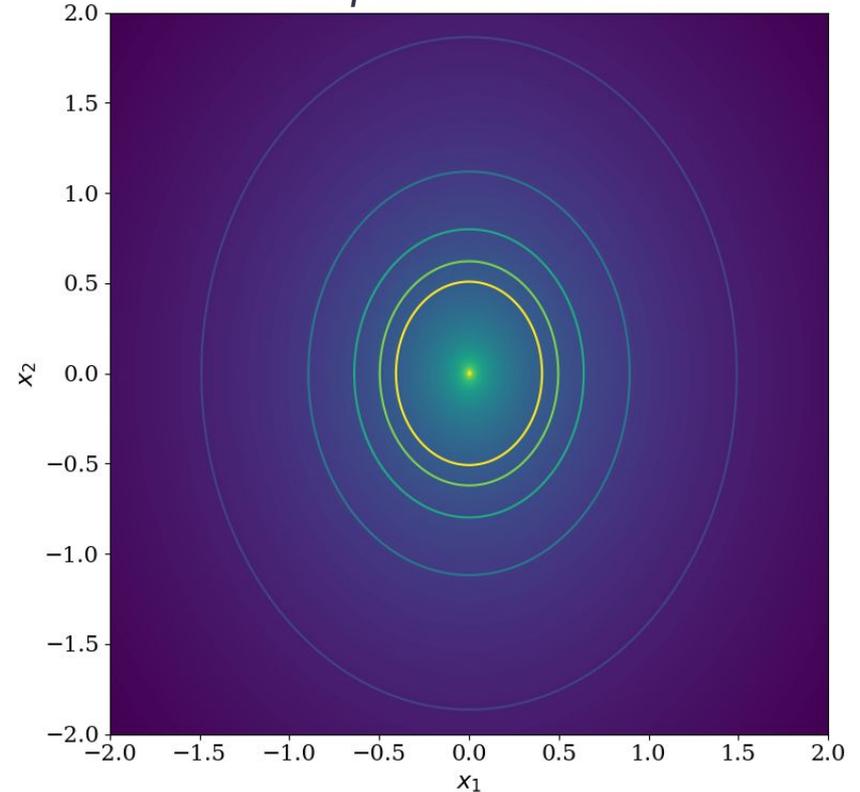
# Singular Isothermal Ellipsoid (SIE)

Distributions de masse des défecteurs :

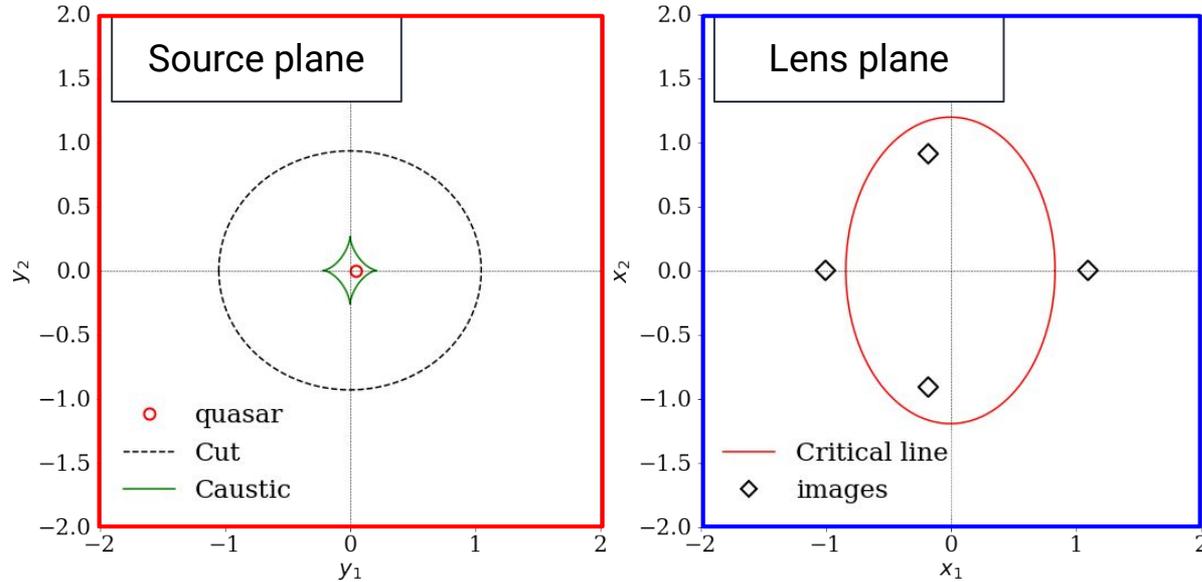
- **SIE (Singular Isothermal Ellipsoid)**



*Carte de la répartition de la masse dans le plan la lentille*



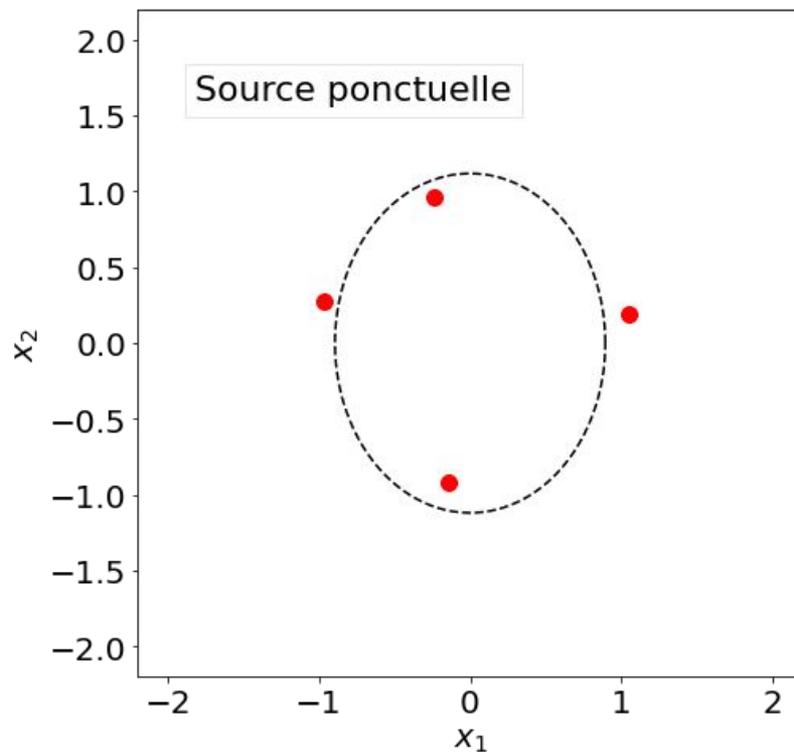
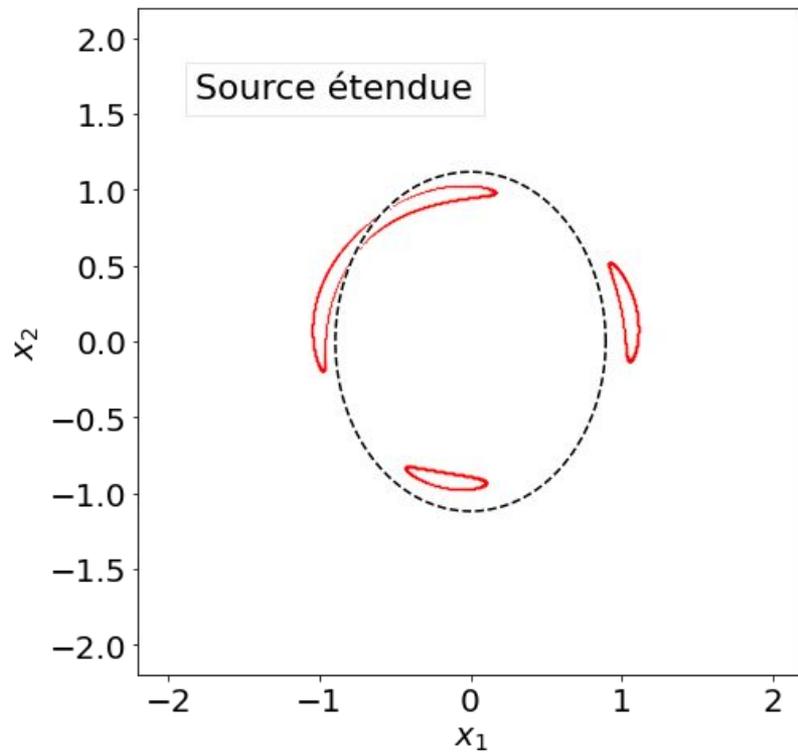
# Calcul de la ligne critique



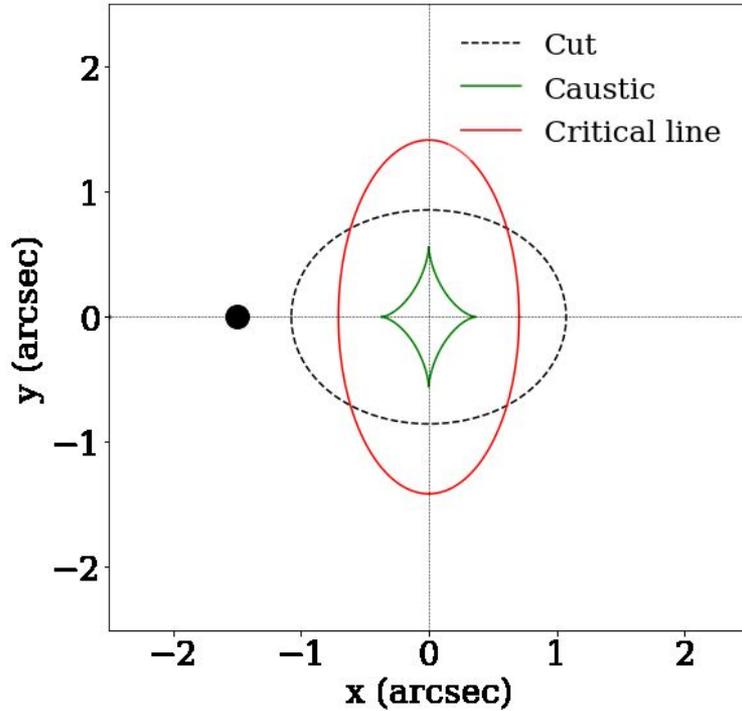
Ligne critique du SIE :

$$\theta_E = \frac{4\pi\sigma_v^2}{c^2} \frac{D_{ls}}{D_s}$$

# Formation des arcs gravitationnels



# Influence de la position de la source



Voici une animation qui montre l'impact de la position de la source ● sur le **nombre d'images** ;

Dans le **cut** on a 2 images ;

Dans la **caustic** on a 4 images ;

Notre source est un **quasar**.

Ce qu'on veut :

- + classer des configurations entre deux classes ; **lentilles gravitationnelles** ou **étoiles**.

Ce qu'il faut :

- + des simulations **réalistes** de lentilles gravitationnelles pour entraîner ce modèle.

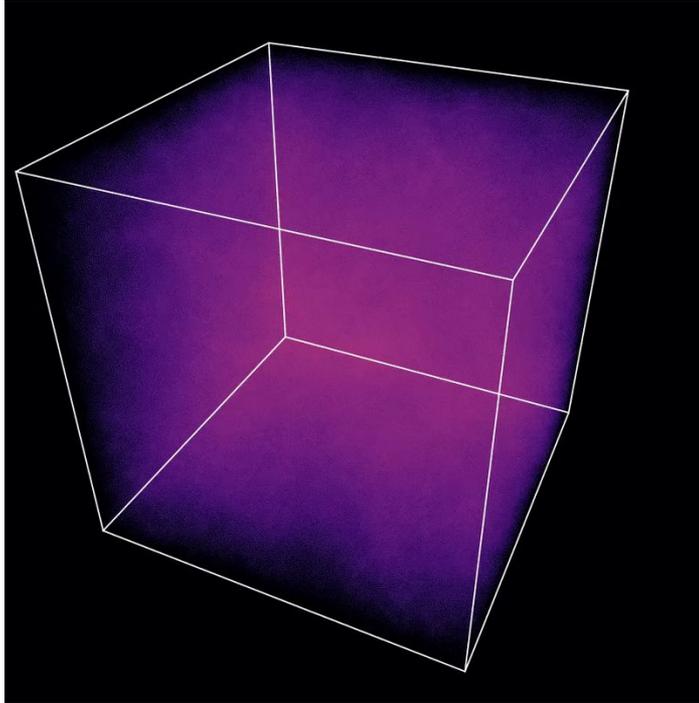
J'ai besoin d'une distribution réaliste pour les paramètres suivants :

- le rapport d'axe ( $b/a$ )
- $\sigma_v$ , la dispersion de vitesse (taille caractéristique)
- redshifts de la galaxie déflectrice et de la source

# Utilisation des simulations EAGLE (Schaye et al. 2015)

EAGLE (Evolution and Assembly of GaLaxies and their Environments)

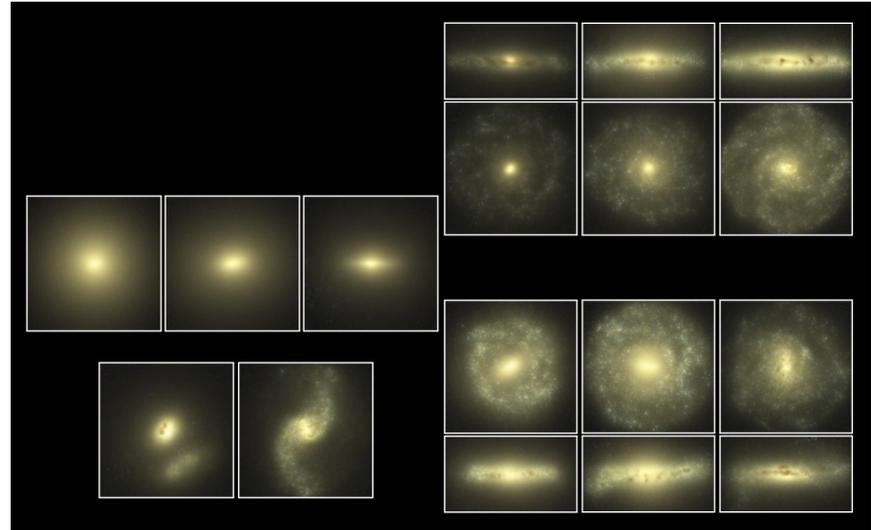
$\Lambda$ CDM =  $\Lambda$ -Cold Dark Matter



Credit : EAGLE team

4 types de particules :

- + Étoiles, gaz, matière noire et trous noirs
- + Les galaxies individuelles sont résolues



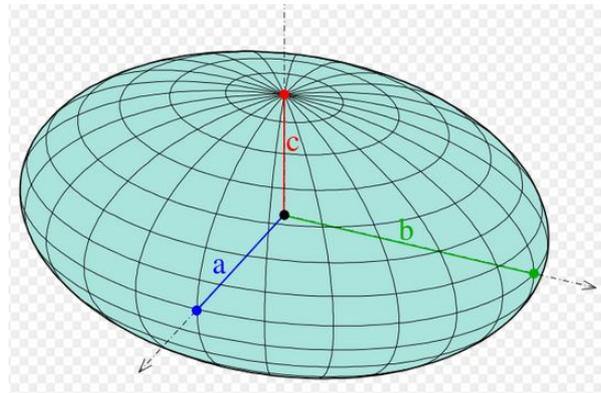
Credit : EAGLE team

## Quels paramètres mesurer ?

---

Pour modéliser chaque halo quelle que soit sa forme, j'ai développé une procédure d'analyse spécifique :

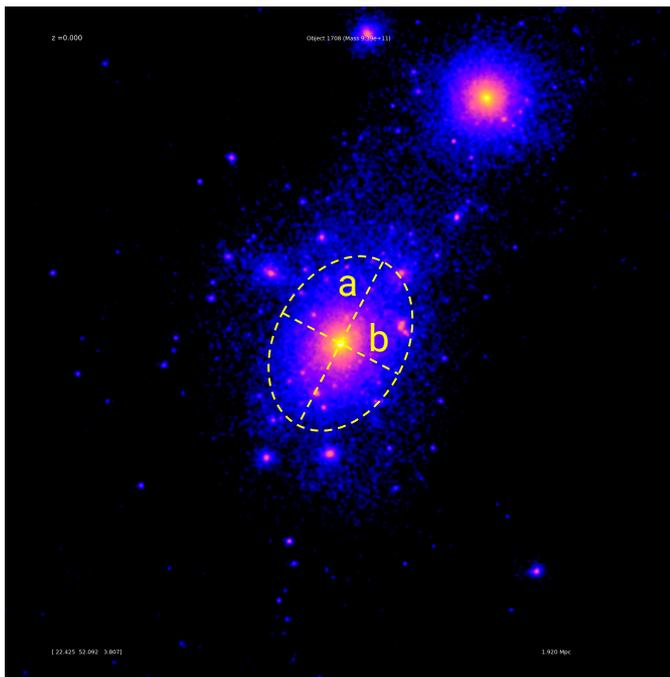
1. Application d'une analyse en composantes principales (**ACP**)
  - a. Rapports d'axe ( $b/a$ ,  $c/a$ ,  $c/a$ )
  - b. Angles d'orientation ( $\varphi$ ,  $\theta$ ,  $\psi$ )
2. Caractéristiques de forme à partir d'un profil de densité radial construit le long des axes principaux
  - a. Profil de densité
  - b. Concentration



+ en 3D

+ **en 2D (projection sur le ciel)**

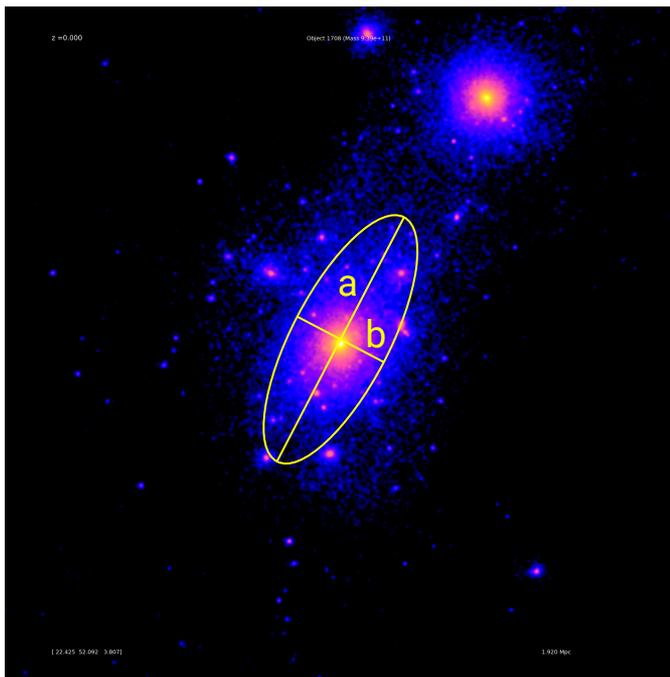
# Comment sont définis les halos dans les simulations EAGLE ?



*Credit : EAGLE team*

1. Trouver le centre avec une méthode itérative
2. Ajustement d'un ellipsoïde (3D et projection 2D)
  - + utilisation d'une Analyse en Composante Principale (ACP) sur le nuage de particules associé à la galaxie

# Comment sont définis les halos dans les simulations EAGLE ?



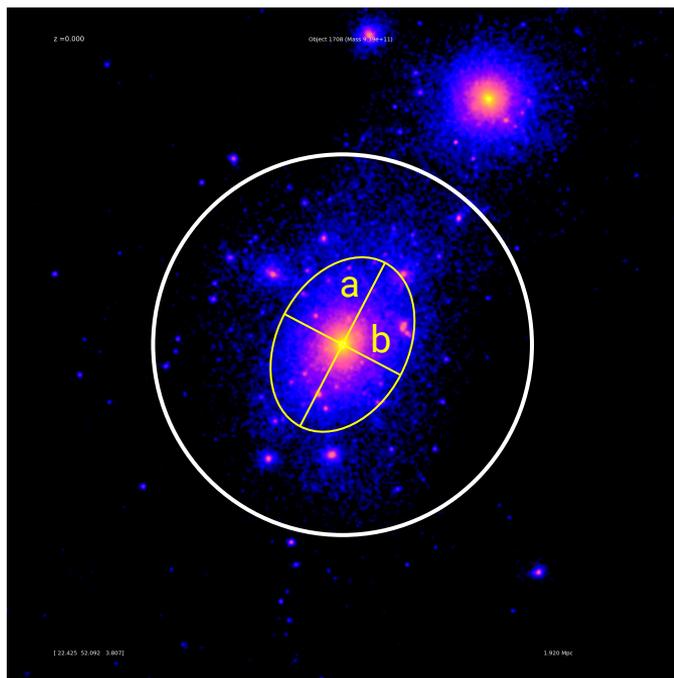
Credit : EAGLE team

1. Trouver le centre avec une méthode itérative
2. Ajustement d'un ellipsoïde (3D et projection 2D)  
+ utilisation d'une Analyse en Composante Principale (ACP) sur le nuage de particules associé à la galaxie



Les sous-halos et les autres structures externes vont grandement influencer l'analyse.

# Comment sont définis les halos dans les simulations EAGLE ?



Credit : EAGLE team

1. Trouver le centre avec une méthode itérative
2. Ajustement d'un ellipsoïde (3D et projection 2D)  
+ utilisation d'une Analyse en Composante Principale (ACP) sur le nuage de particules associé à la galaxie

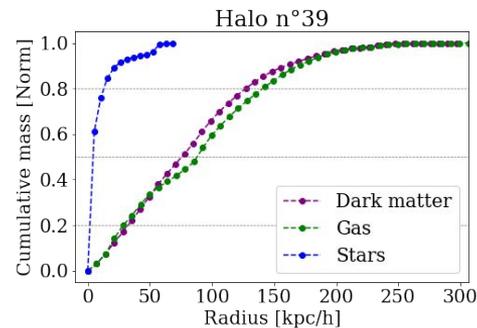
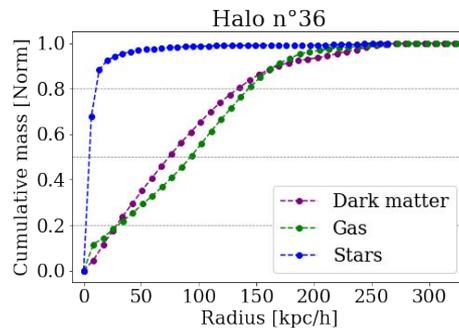
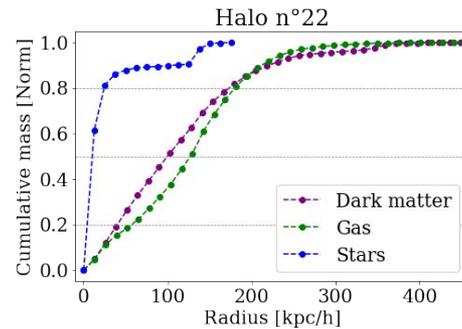
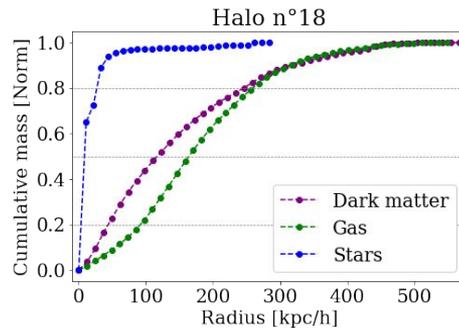
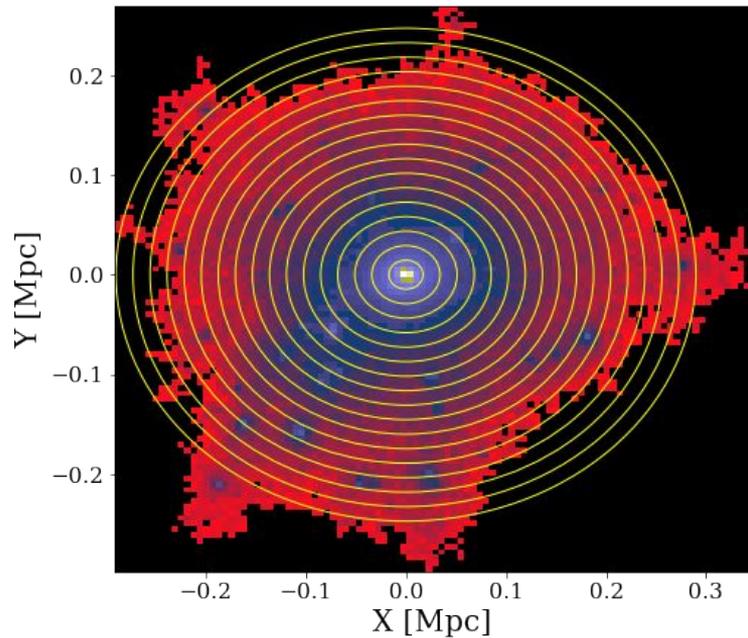


Les sous-halos et les autres structures externes vont grandement influencer l'analyse.



La solution : définir une sphère de travail pour chaque halo individuellement.

# Profils de densité

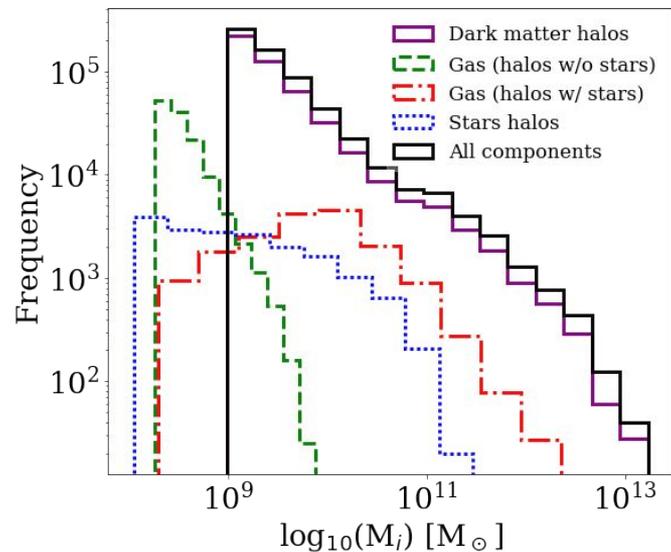
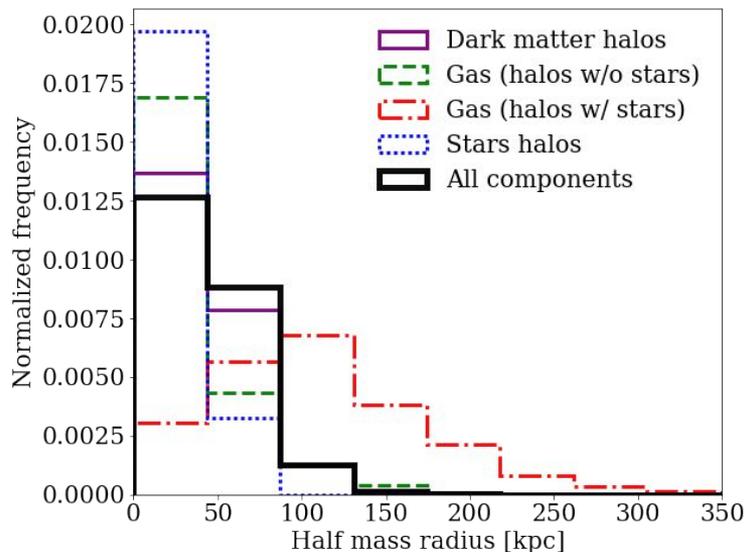


Les changements abruptes de pente sont la signature de la présence de sous-halos.

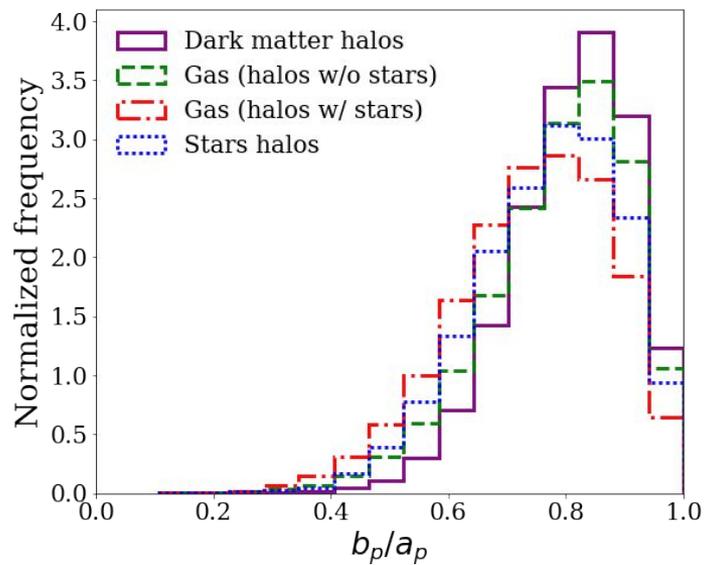
# Résultats de l'étude morphologique des simulations de EAGLE

- 336 540 halos
- redshift de  $z=15$  à  $z=0$
- plus de 10 paramètres mesurés pour chaque halo

De nouvelles distributions de paramètres ;  
Pour chaque type de particules individuellement.

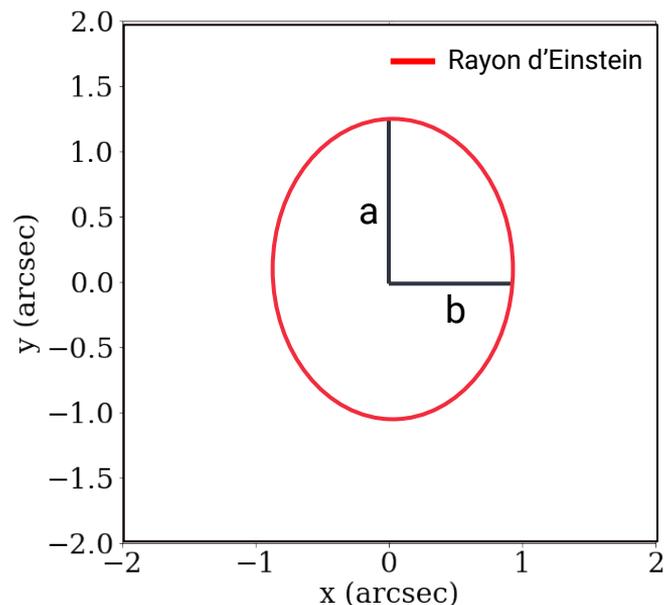


# Distributions des rapports d'axes projetés (2D)



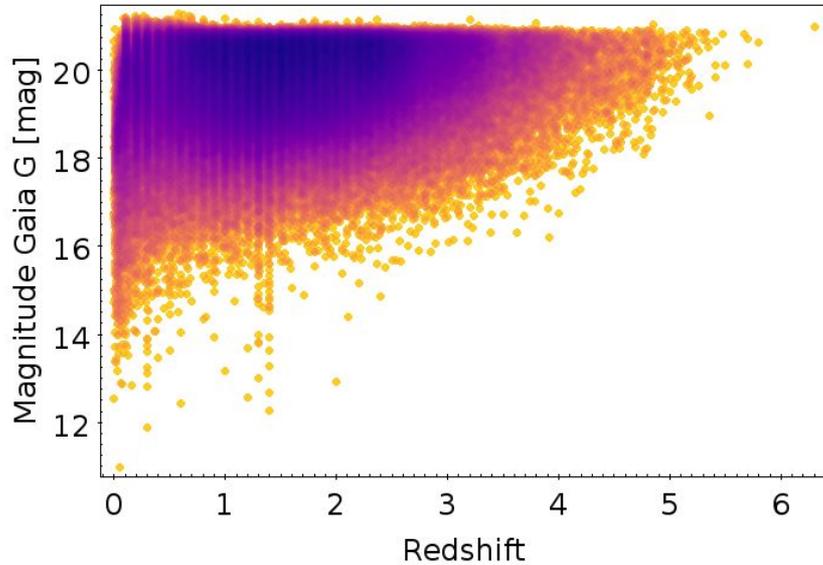
Différent des distributions uniformes initiales envisagées dans les simulations de lentilles gravitationnelles ;

Tous les résultats ont été publiés dans *Petit et al. 2023*.

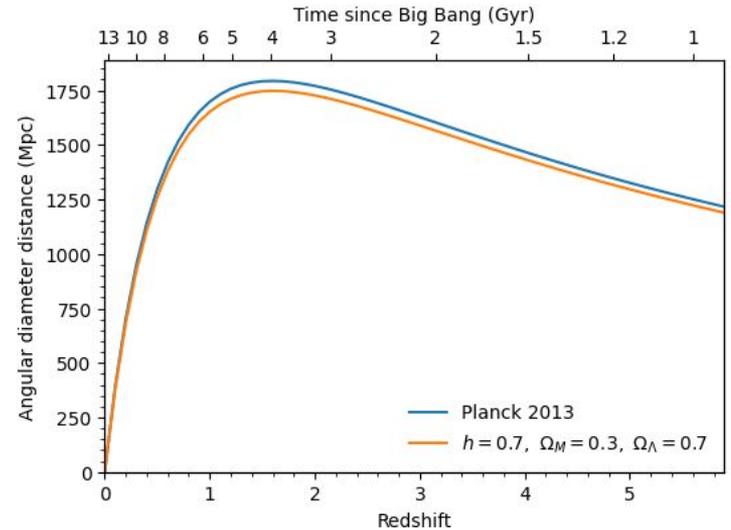


# Sélection de la source d'arrière plan : les quasars

Utilisation du Million Quasars (MILLIQUAS) Catalog, Version 7.10 (15 April 2023)

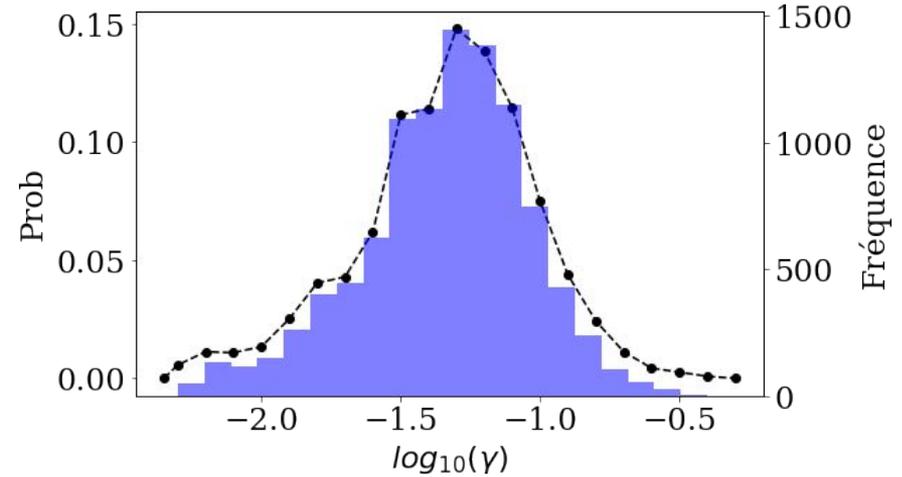
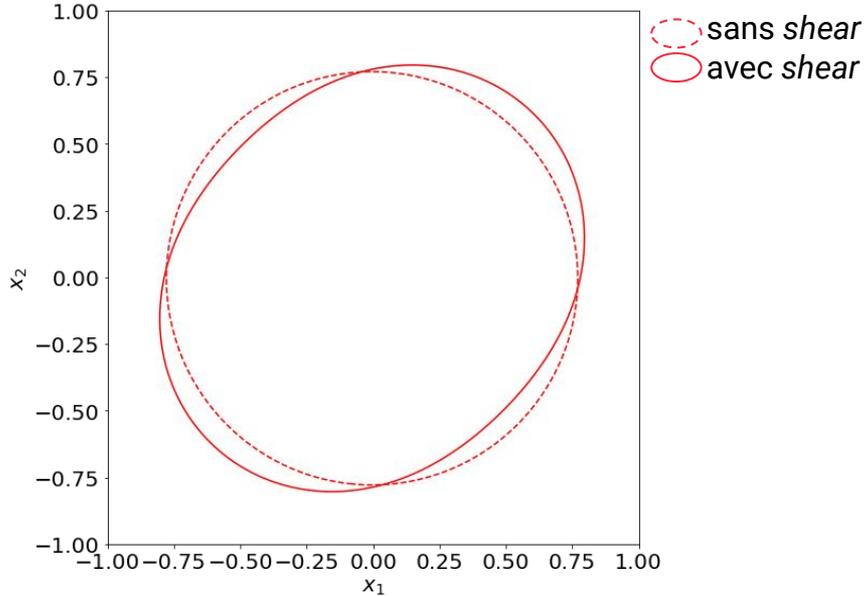


- + 544 981 quasars mesurés par Gaia
- + magnitudes
- + distances



# Ajout du *shear* $\gamma$ externe (cisaillement) et d'un bruit de mesure

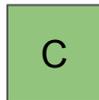
Rayon d'Einstein



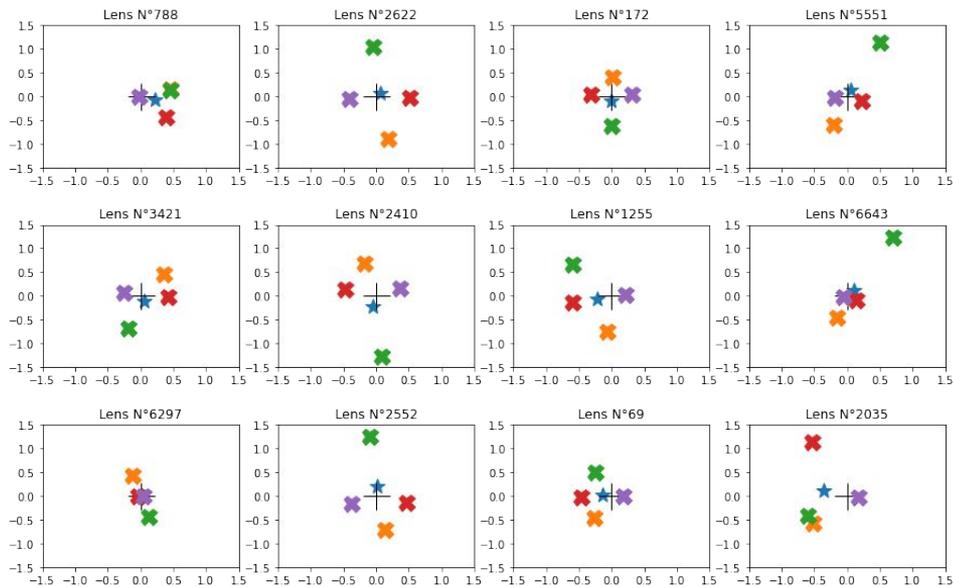
Holder and Schechter (2003) | Semi-Analytic Galaxy Formation - GIF

1. Il est nécessaire d'intégrer un *cisaillement* externe pour tenir compte de la matière présente autour de la lentille.
2. Afin de tenir compte des sources d'incertitudes, j'ai bruité les positions des images du quasar avec des valeurs typiques des erreurs de Gaia.

# Résultats des simulations

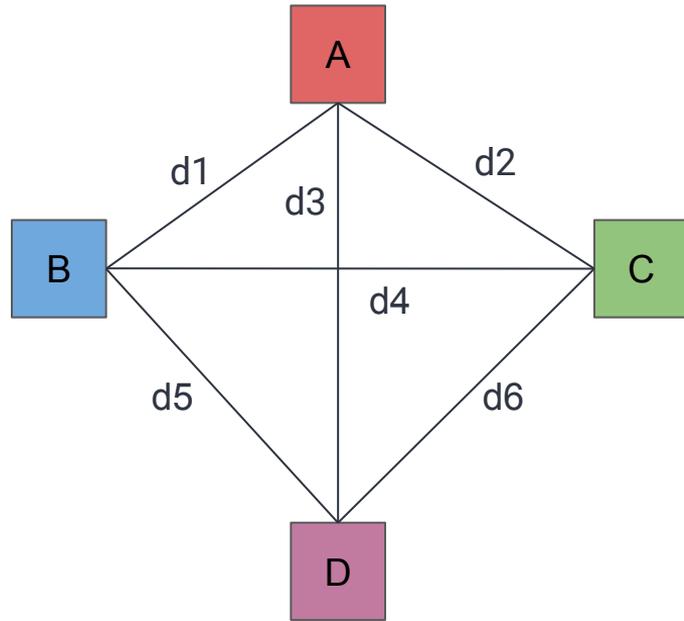


## Sample of GL



# Création des catalogues d'entraînement

---

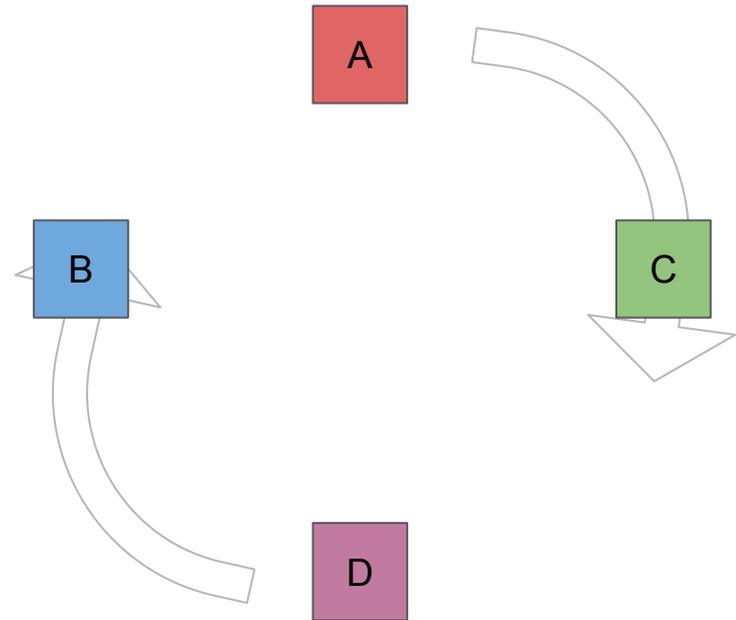
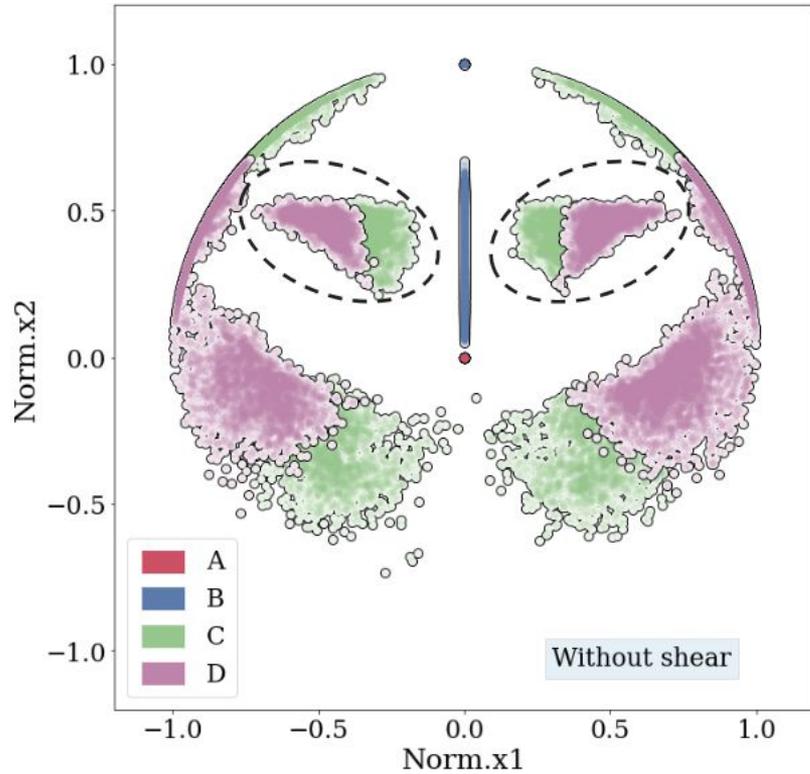


Pour entraîner XGBoost, j'ai besoin d'un catalogue d'entraînement **tabulaire**.

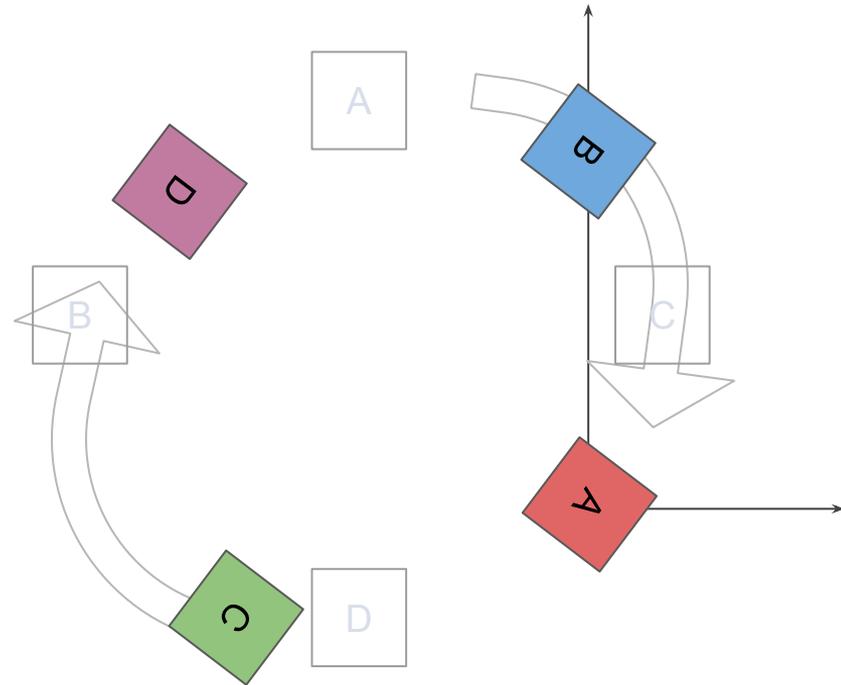
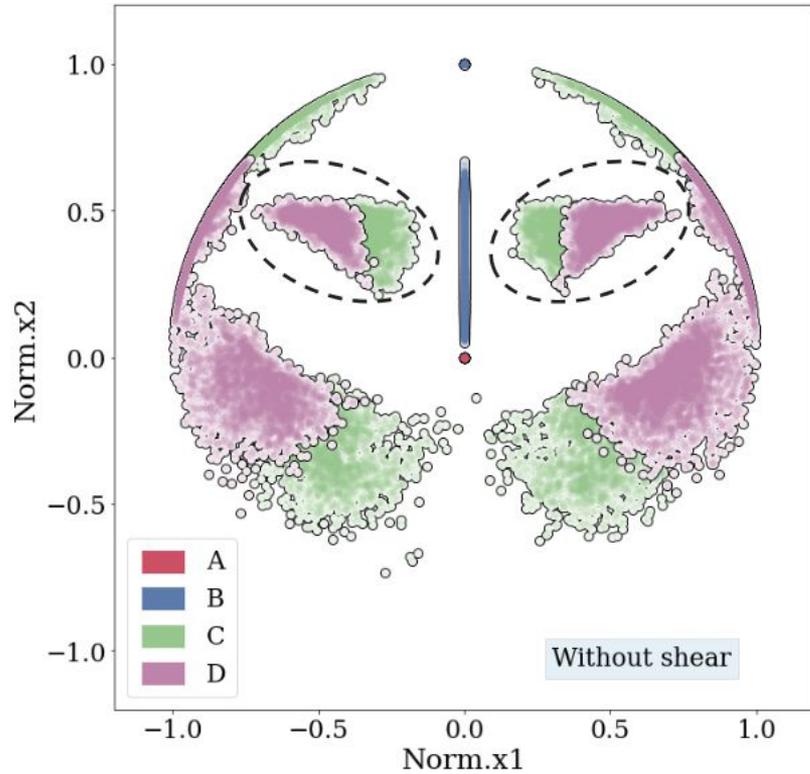
Ensemble des paramètres :

- Positions relatives
- Distances : d1-d6, MaxDist, MinDist
- Distances normalisées : Nd1 - Nd6
- Différences de magnitudes :  $N\mu$
- Angles

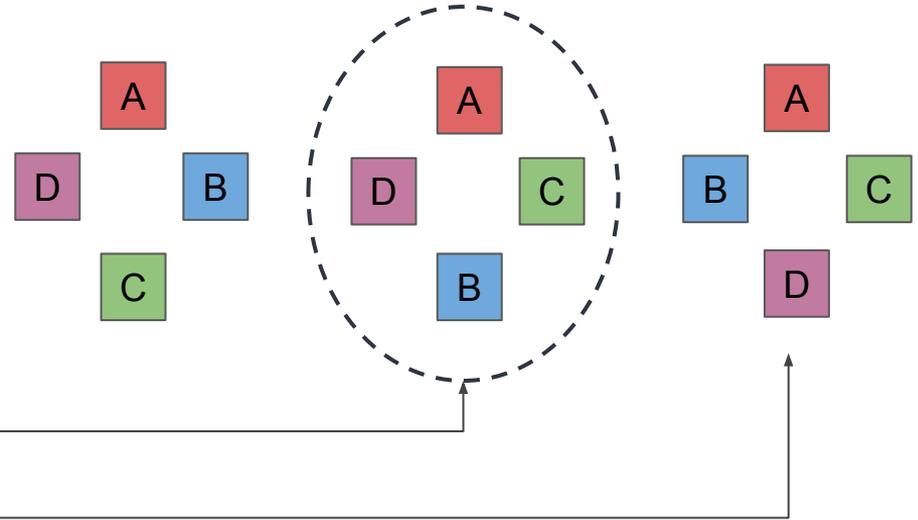
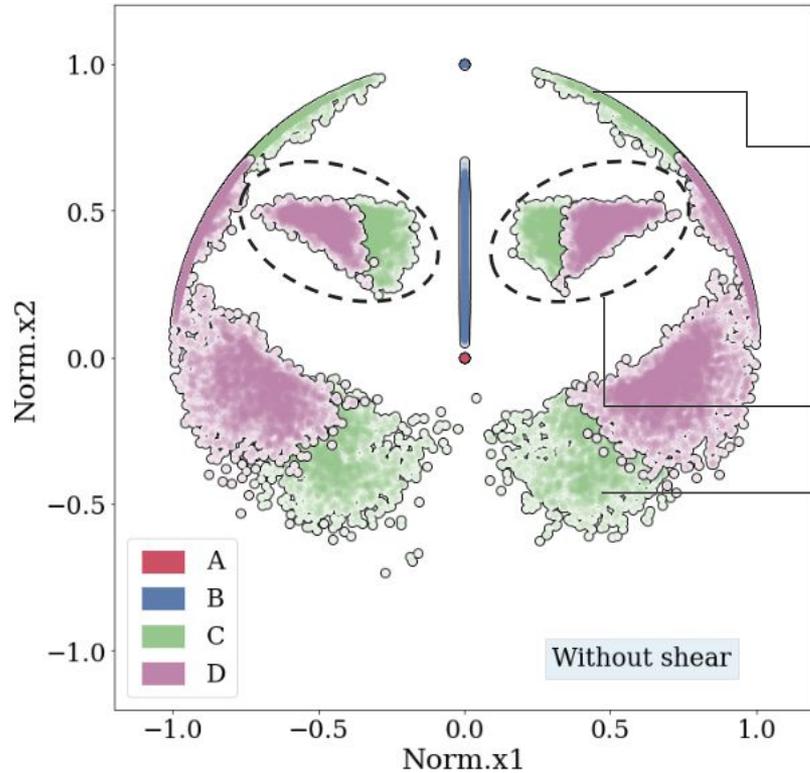
# Ajout de nouveaux paramètres



# Ajout de nouveaux paramètres



# Ajout de nouveaux paramètres



Les simulations de lentilles gravitationnelles constituent la première classe de mon catalogue d'entraînement ;

Quid de la seconde classe ?

# Groupes d'étoiles de Gaia

Catalogue d'étoiles de la Gaia DR2 (Delchambre et al. 2019)

Sélection de 4 étoiles dans un champ de 6 secondes d'arc (= **65 694 groupes d'étoiles**)

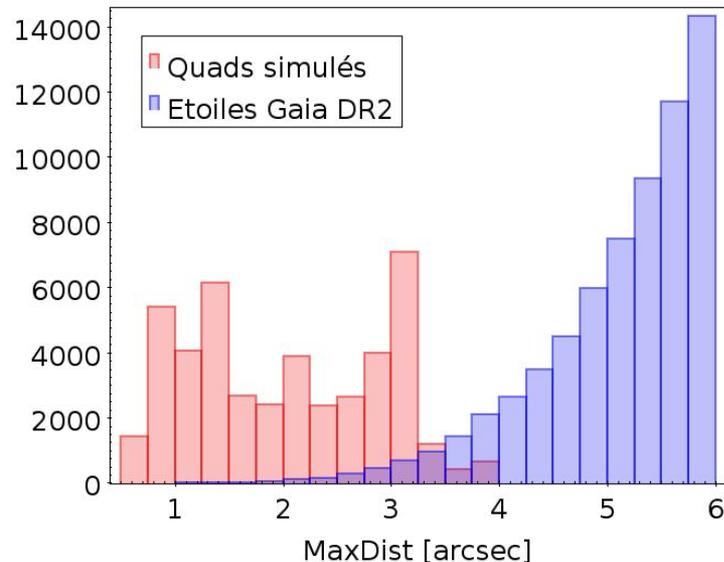
Puis j'applique le même traitement que pour les quads pour obtenir les mêmes paramètres.

## Deux classes (50% / 50%):

- lentilles gravitationnelles réalistes (EAGLE + SIE)
- groupes d'étoiles de Gaia DR2

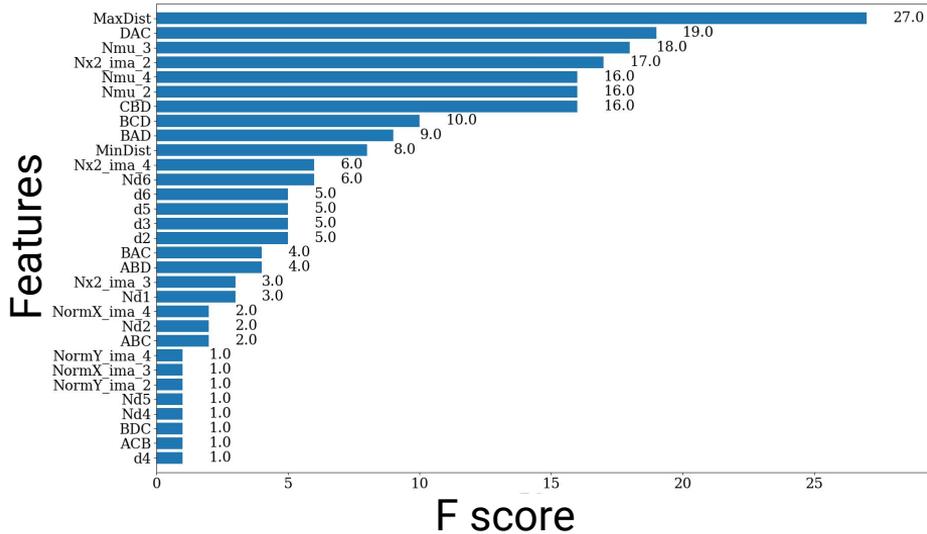
## Deux apprentissages :

- Seulement géométrique
- Géométrique + distances

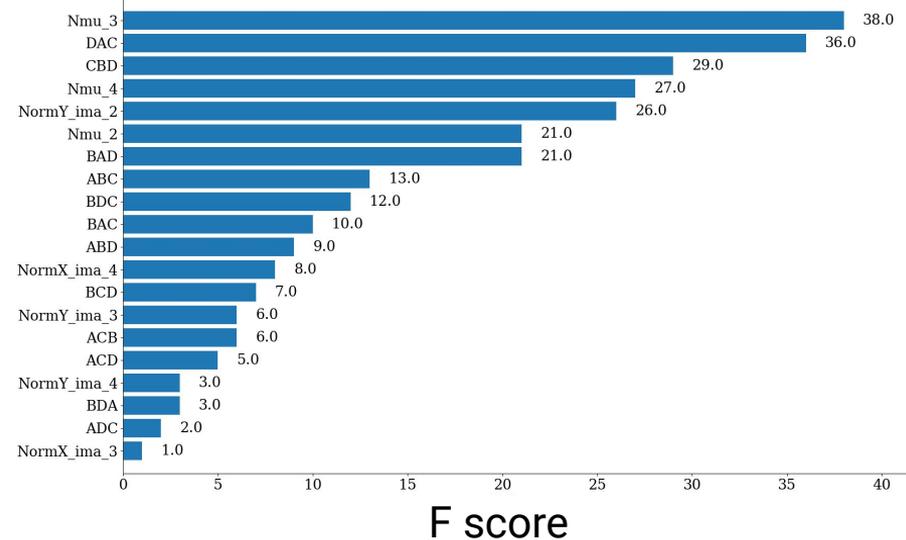


# Importance des paramètres lors de l'apprentissage

## Apprentissage géométrique + distances



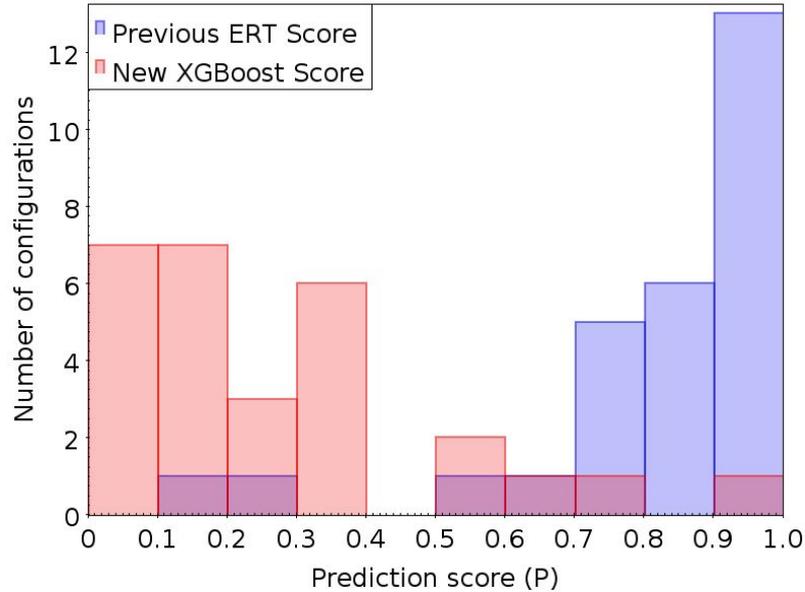
## Apprentissage seulement géométrique



- L'importance des paramètres restent relativement dans le même ordre ;
- La distance joue un rôle important du fait de la différence entre les lentilles gravitationnelles et les groupes d'étoiles du catalogue d'entraînement.

# Validation de l'apprentissage

## Candidats observés faux positifs



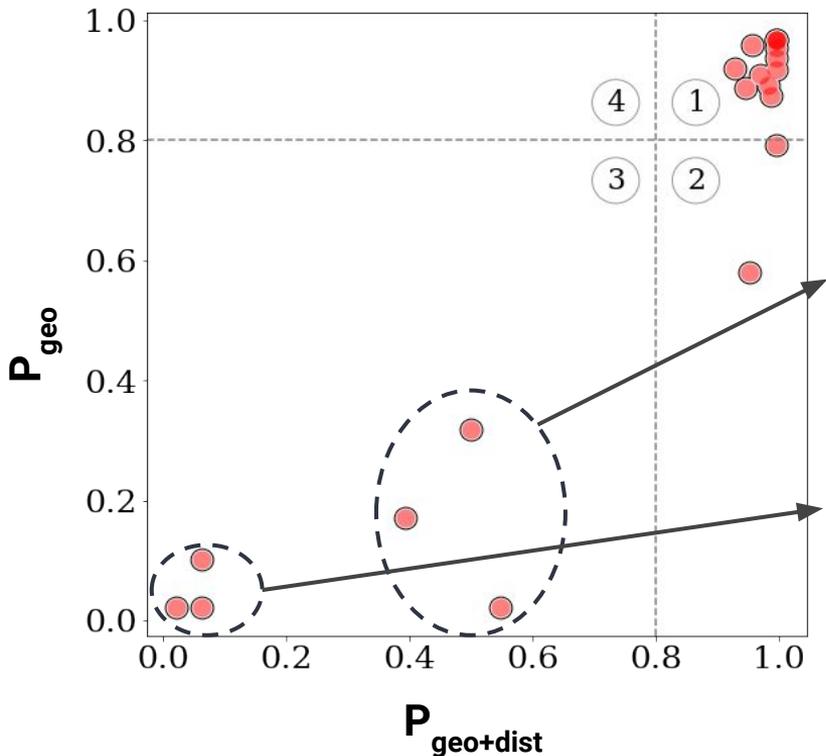
P = Probabilité que la configuration soit une lentille gravitationnelle de quasar à quatre image

ERT = Extremely Randomized Trees (Delchambre et al. 2019)

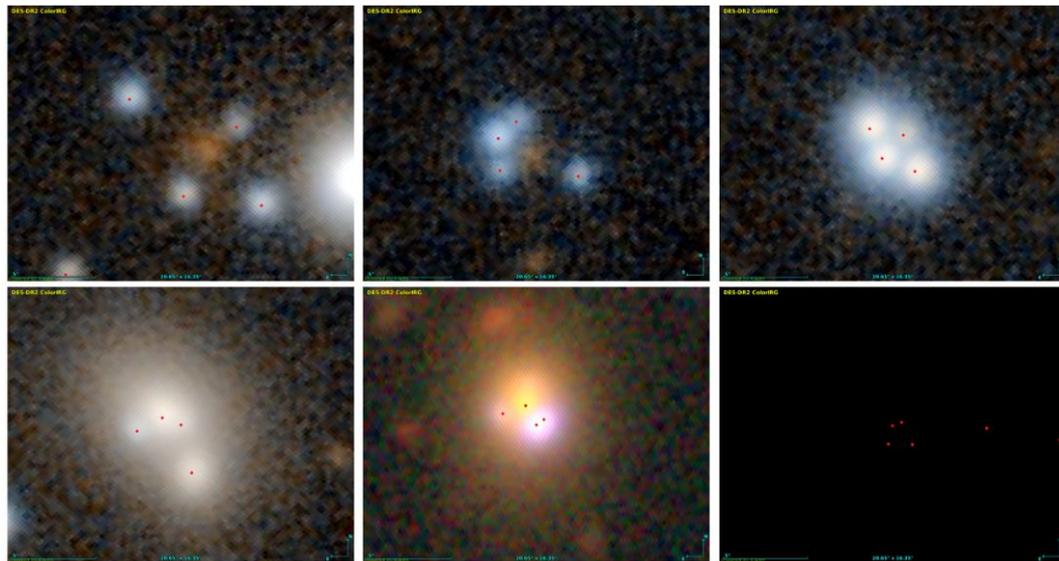


Grande amélioration sur le taux de faux positifs

# Validation sur des quads déjà connus



$P$  = Probabilité que la configuration soit une lentille gravitationnelle



Première limite clairement identifiée : je ne simule pas tous les cas possibles.

# Gaia Focused Product Release (FPR)

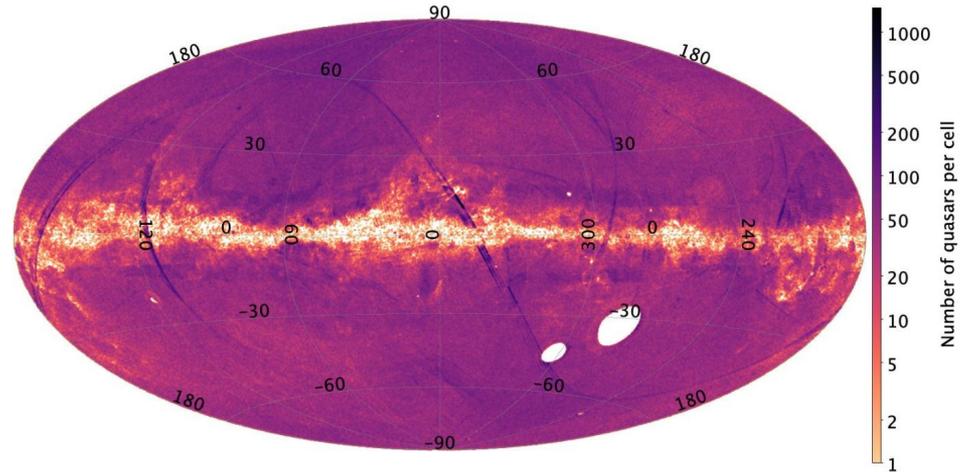
*Gaia Coll. & Krones-Martins et al. 2023*

Gaia FPR produit un catalogue de sources autour de 3.8 M de quasars (6")  
(il y a plus de sources que dans la DR3)

Particularité du travail actuel :

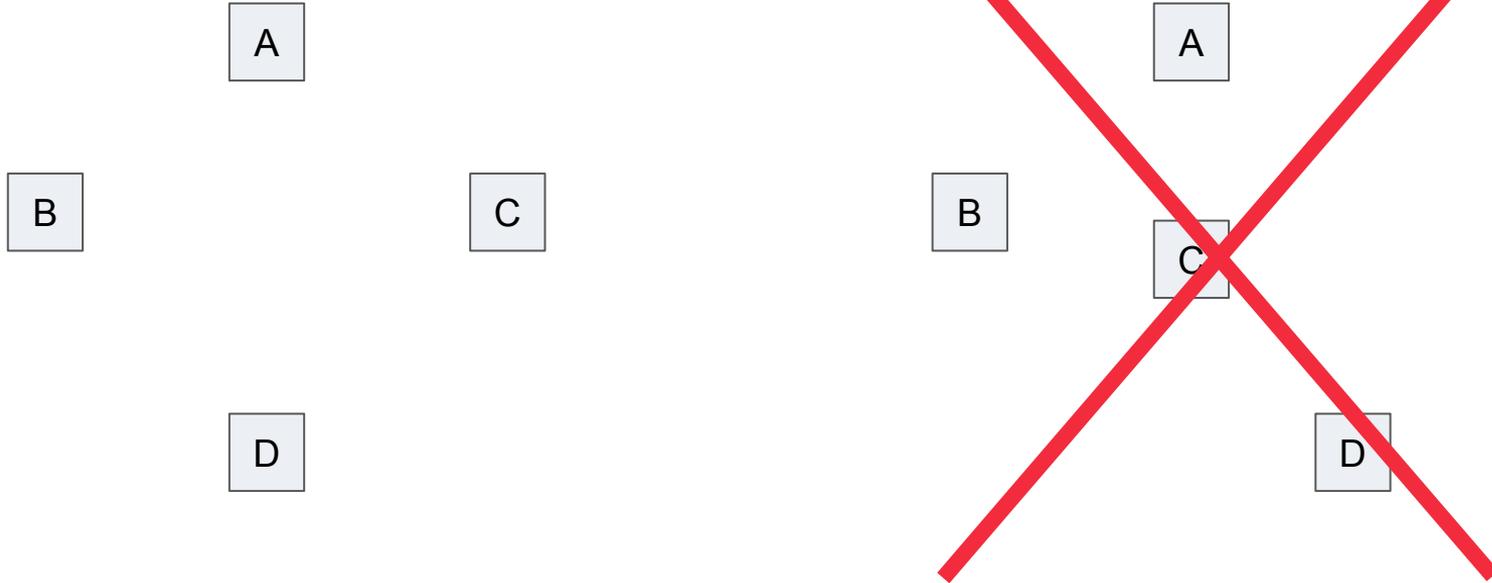
- **uniquement des multiplets à 4 images sont analysés**

➡ 17 610 groupes de 4 sources



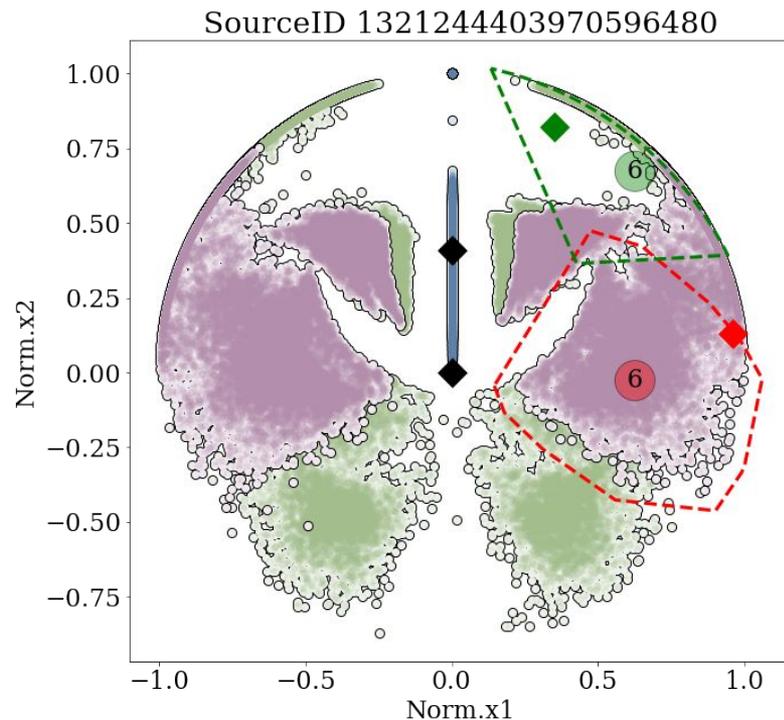
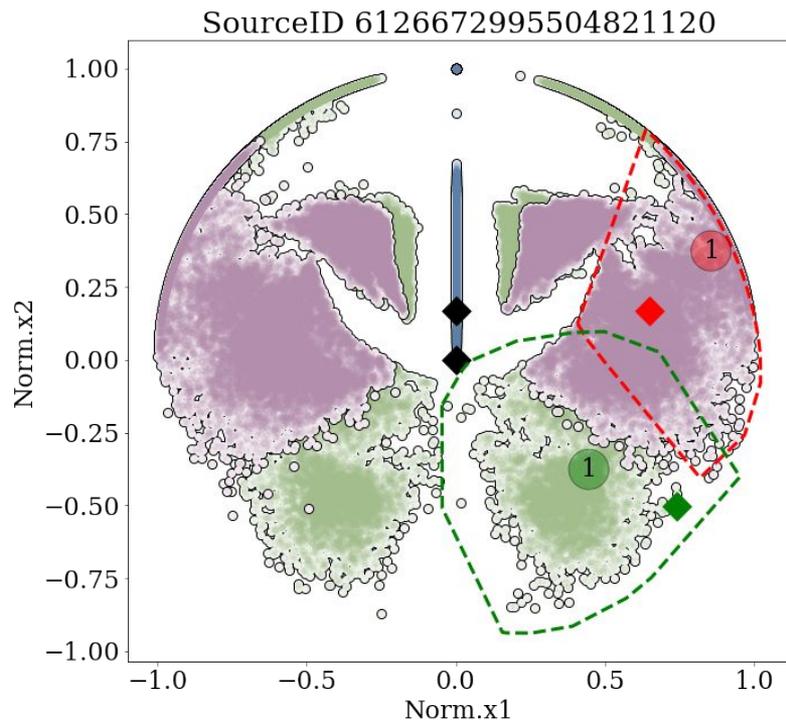
## Présélection dans le catalogue FPR pour diminuer le taux de faux positifs

---



De **17 610** à **14 106** multiplets à quatre sources.

# Présélection dans le catalogue FPR pour diminuer le taux de faux positifs

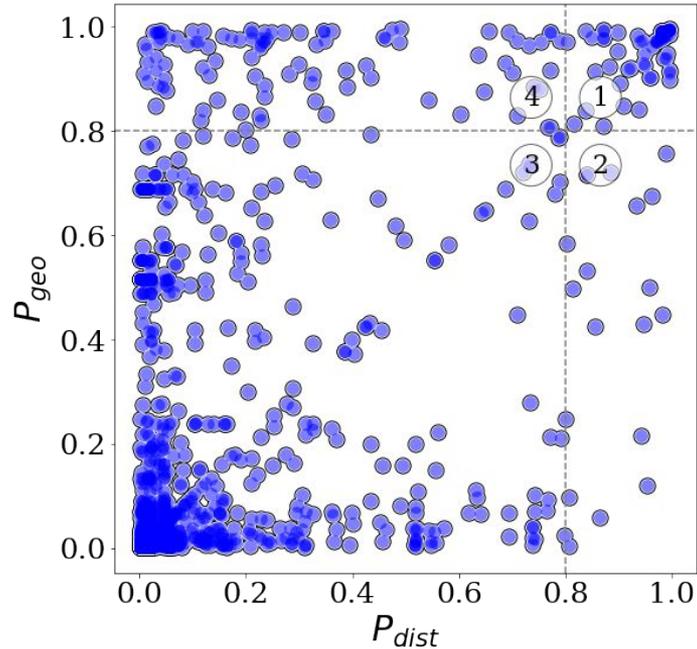


On élimine les quadruplets pour lesquels C et D ne sont pas dans les zones probables.

De **14 106** à **1 893** multiplets à quatre sources.

# Application au catalogue Gaia FPR

Probabilités issus de XGBoost pour ces 1 893 multipléts :



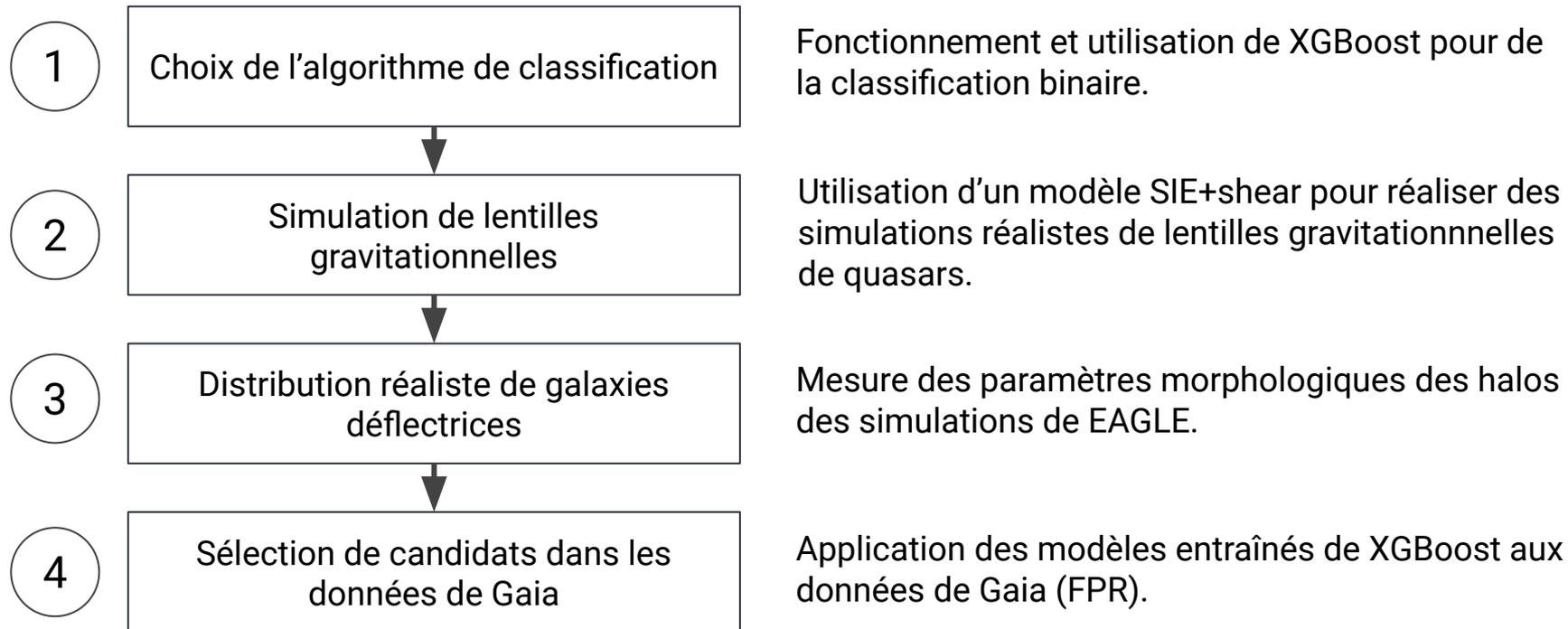
	Cadran 1	Cadran 2	Cadran 3	Cadran 4
Nombre de multipléts	67	19	1708	99



Sélection de candidats avec une méthode complète que j'ai développée.

# Conclusion

---



# Conclusion



# Perspectives futures

---

## **Amélioration des simulations :**

- Prendre en compte les effets de microlentille ;
- Passer à un modèle NSIE ;

## **Prendre en compte plus de groupes de sources :**

- Analyser les multiplets restants à 3 ou 6, 7, 8... composantes ;

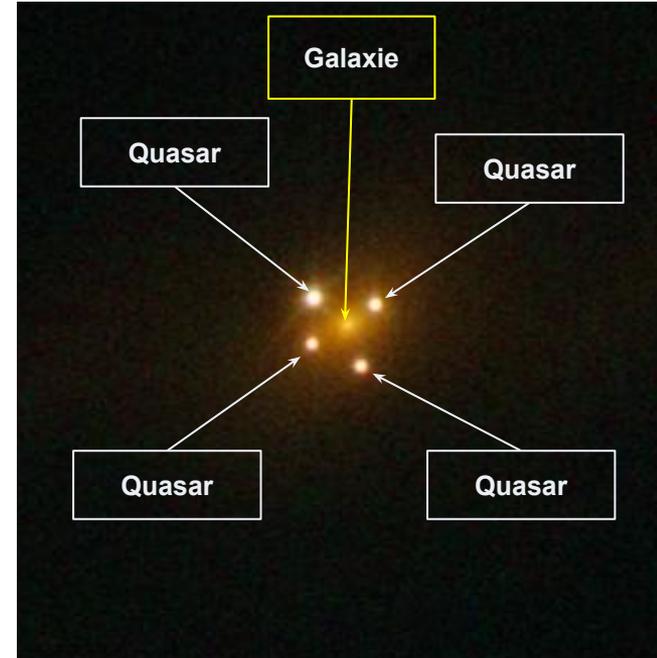
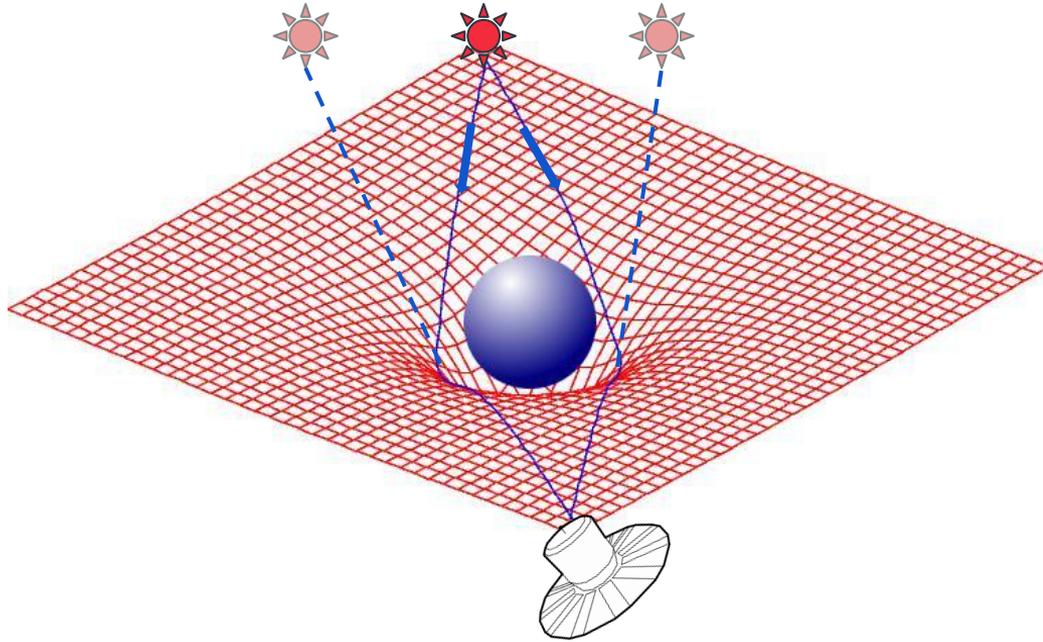
## **Ajouter de filtres avant des observations :**

- Ajout d'un filtre sur les couleurs ;
- Ajout d'un filtre sur le nombre de détections de Gaia.

**Merci pour votre attention !**

---

# Qu'est-ce qu'une lentille gravitationnelle ?



*Credit : Hubble Space Telescope*

La présence d'une masse déforme l'espace-temps environnant.  
La lumière va suivre la géodésique.

**Un quasar lentillé à quatre images est appelé un quad.**

# Les arbres de décisions

---

Pour décider quelle caractéristique utiliser pour diviser les données à un nœud donné :

➡ une mesure d'impureté (entropie, le **critère de Gini** ou l'erreur de classification)

Le critère de Gini (indice de Gini ou impureté de Gini) est calculé comme suit :

$$Gini(D) = 1 - \sum_{i=1}^c (p_i)^2$$



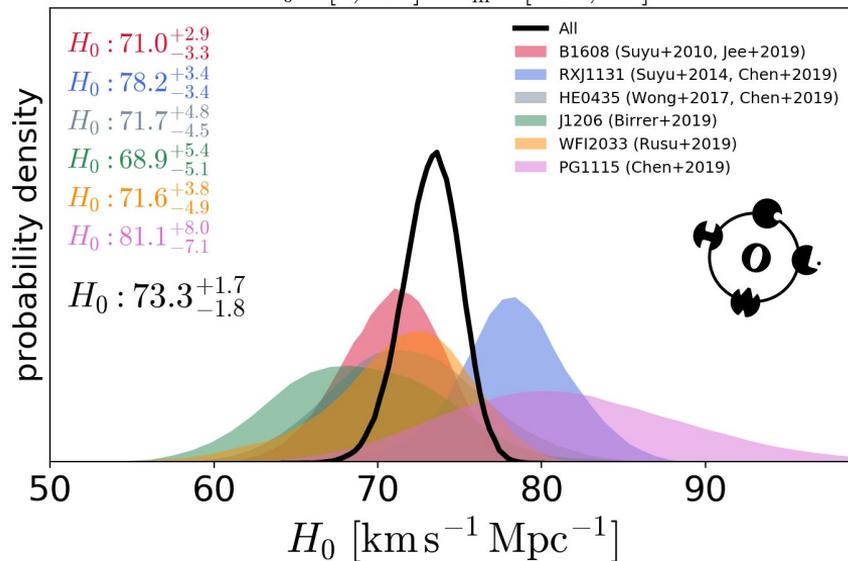
où :

- Gini(D) est l'indice de Gini de l'ensemble de données D.
- c est le nombre de classes ou de catégories cibles.
- $p_i$  est la proportion d'échantillons dans l'ensemble D qui appartiennent à la classe i.

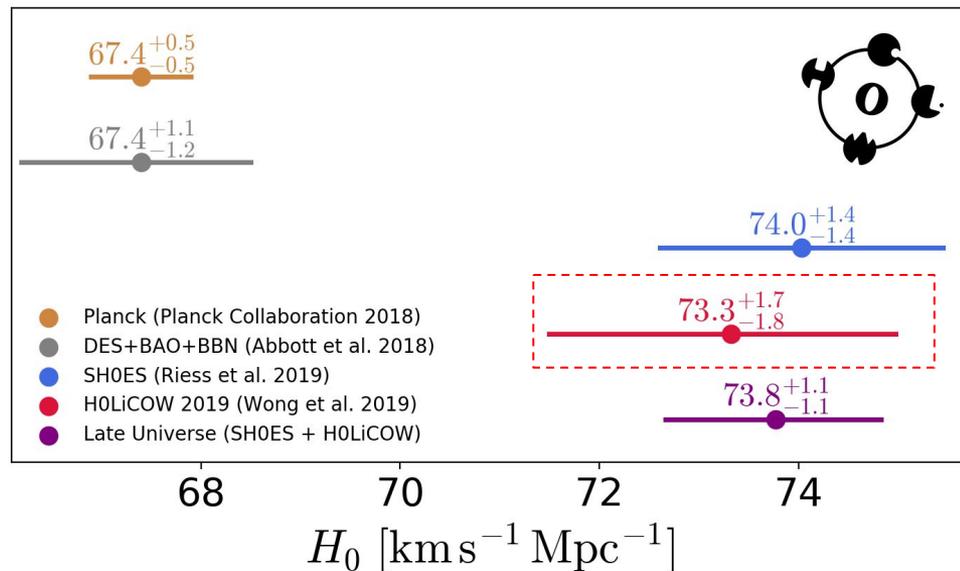
# Détermination de $H_0$

## H0LiCOW ( $H_0$ Lenses in COSMOGRAIL's Wellspring)

$H_0 \in [0, 150]$   $\Omega_m \in [0.05, 0.5]$

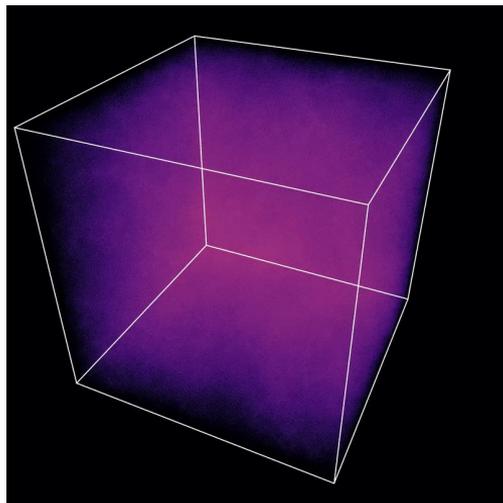


flat  $\Lambda$ CDM



Analyse en aveugle de six quads grâce à une forte lentille gravitationnelle (Wong et al. 2020 - XIII)

# Plusieurs simulations disponibles



Name	$L$ (comoving Mpc)	$N$	$m_g$ ( $M_\odot$ )	$m_{\text{dm}}$ ( $M_\odot$ )
L025N0376	25	$376^3$	$1.81 \times 10^6$	$9.70 \times 10^6$
L025N0752	25	$752^3$	$2.26 \times 10^5$	$1.21 \times 10^6$
L050N0752	50	$752^3$	$1.81 \times 10^6$	$9.70 \times 10^6$
L100N1504	100	$1504^3$	$1.81 \times 10^6$	$9.70 \times 10^6$

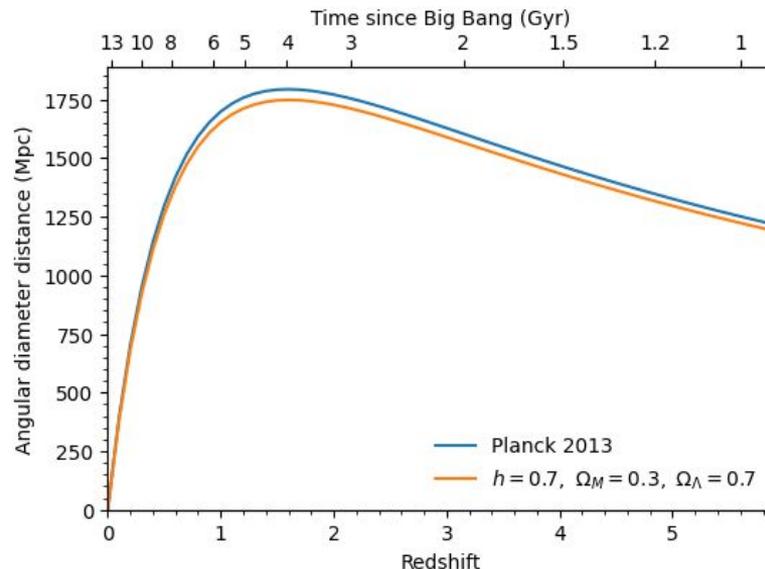
**L** : taille de la simulation

**N** : nombre de particules de matière noire

$m_g$  : masse particules de gaz

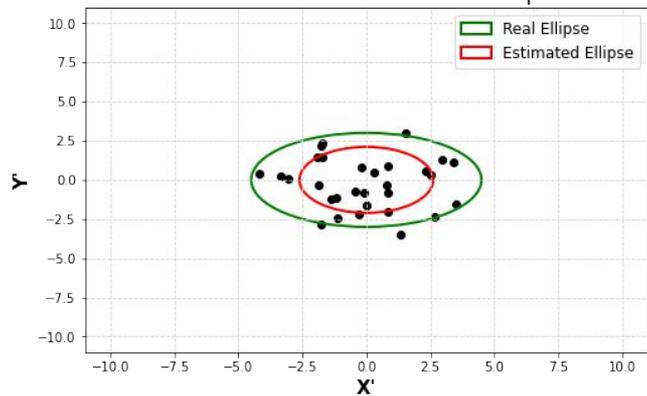
$m_{\text{dm}}$  : mass particules de matière noire

28 snapshots (redshifts |  $z=20$  à  $z=0$ )

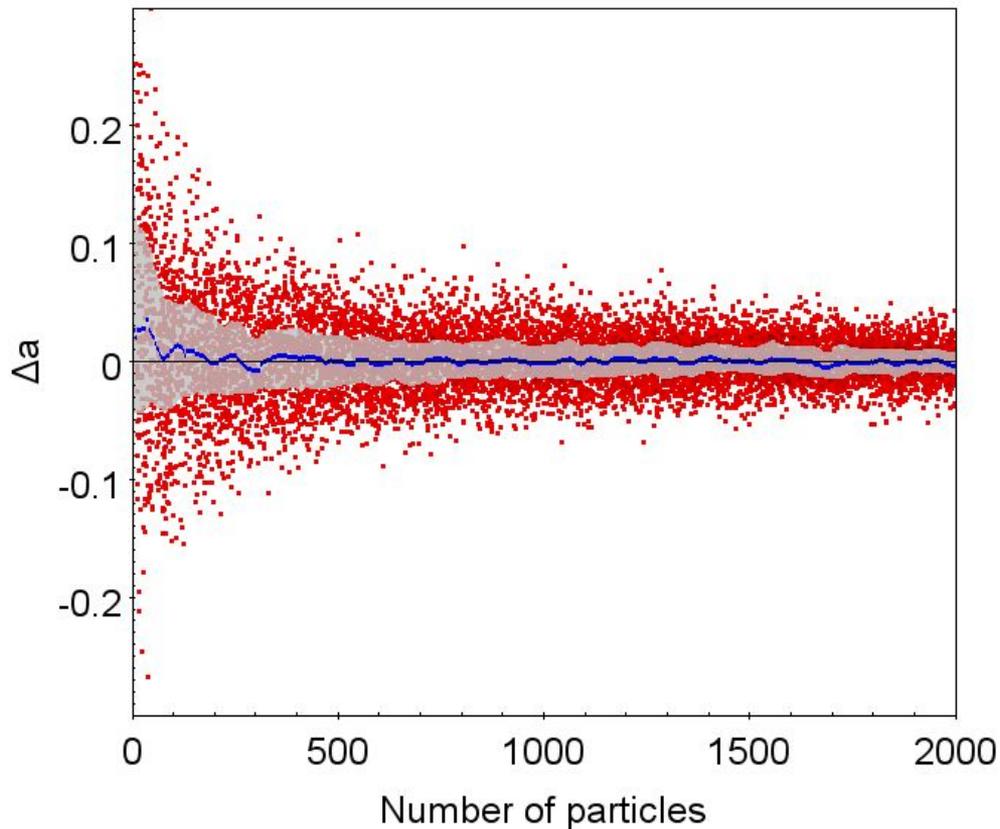
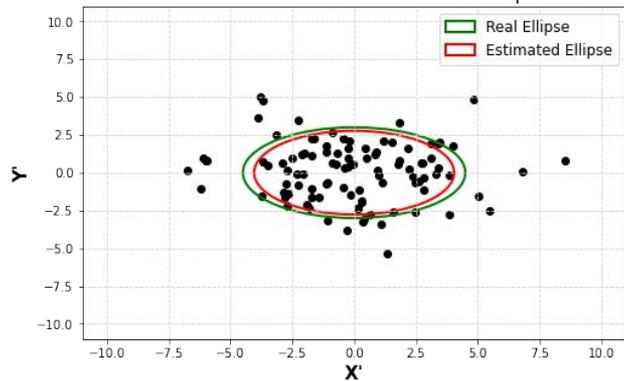


# Estimation des erreurs

Real and estimated on X'Y' for 30 points



Real and estimated on X'Y' for 100 points



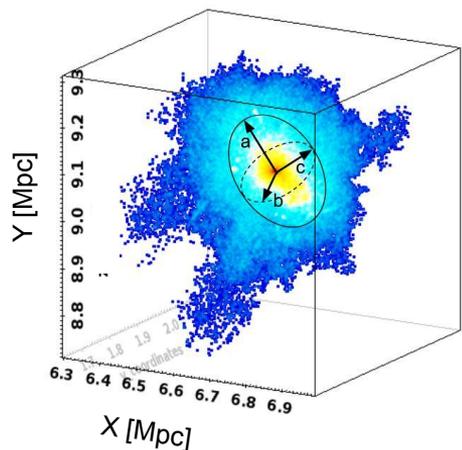
# Simulations de lentilles gravitationnelles

---

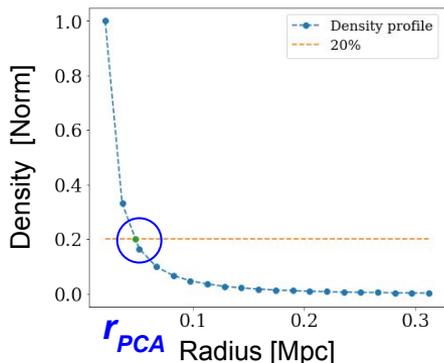
Paramètre	Range
Rapports d'axe	EAGLE b/a projeté
Redshifts	EAGLE pour zL Gaia QSOs pour zS
Angle de position	EAGLE
Rayon d'Einstein	EAGLE halos (Masse totale - half-mass radius)
Shear externe et angle du shear	Déterminer un shear à partir de la simulation EAGLE (Méthode Holder 2022)

# Distribution réaliste de galaxies défectives

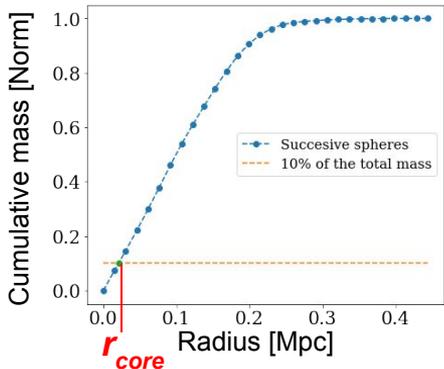
Applying a Principal Components Analysis (PCA) : limit the influence of asymmetrical substructures



3D plot of a typical halo with large scale sub-structures from EAGLE ( $z=0.5$ )



1. Determination of the centre of the halos.
2. The radius ( $r_{PCA}$ ) at which the density **drops by 80%** from the central density.

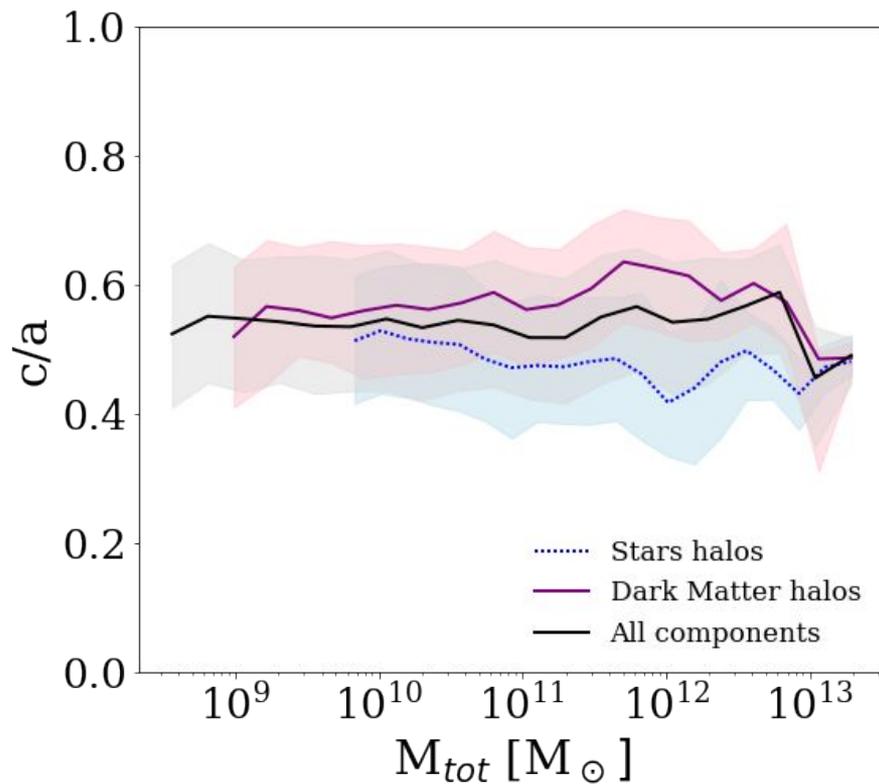
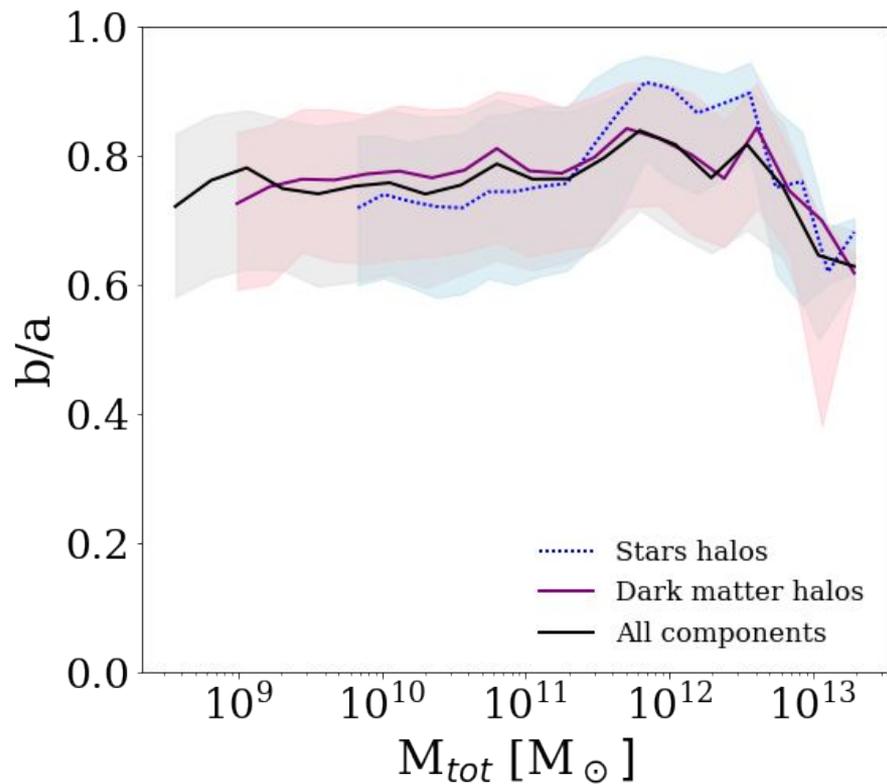


We determine the core radius  $r_{core}$  encompassing 10% of the total (asymptotic) mass.

# Résultats



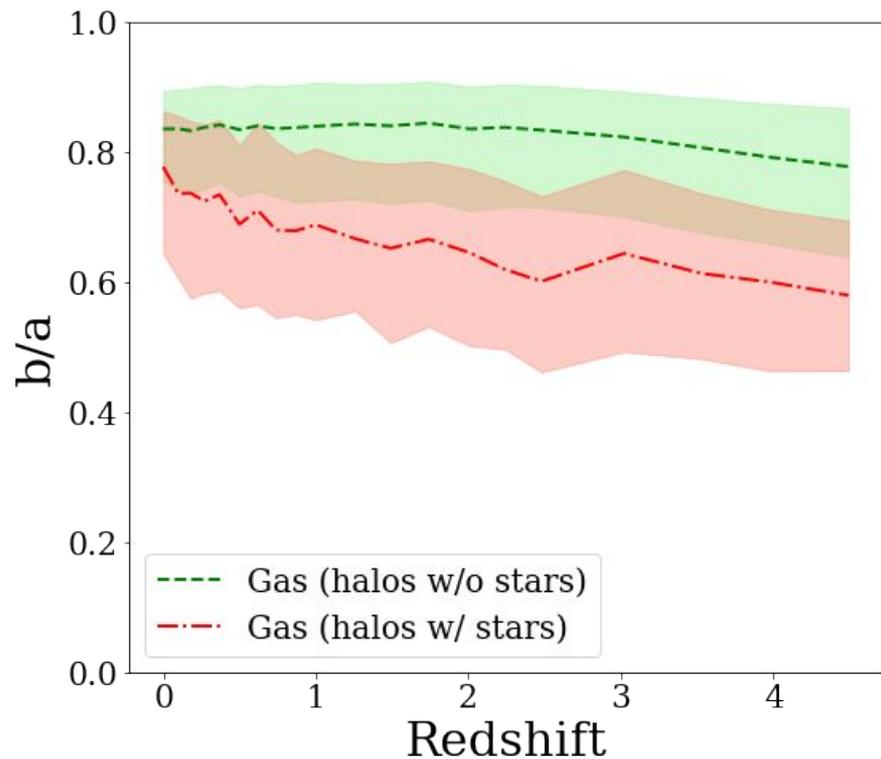
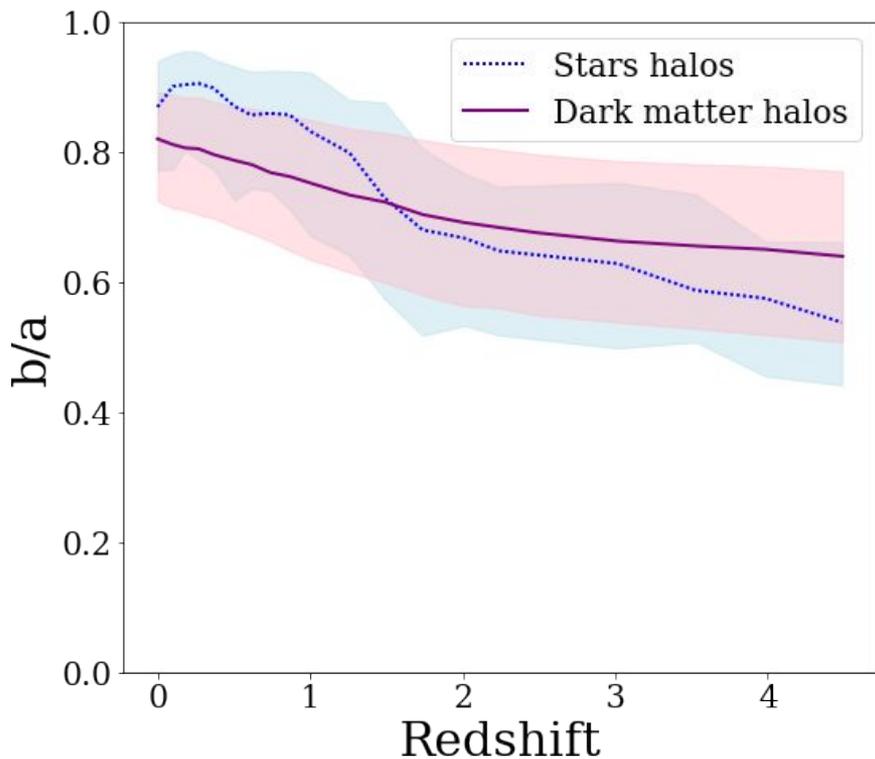
Rapports d'axes des étoiles / matière noire en fonction de la masse totale des halos



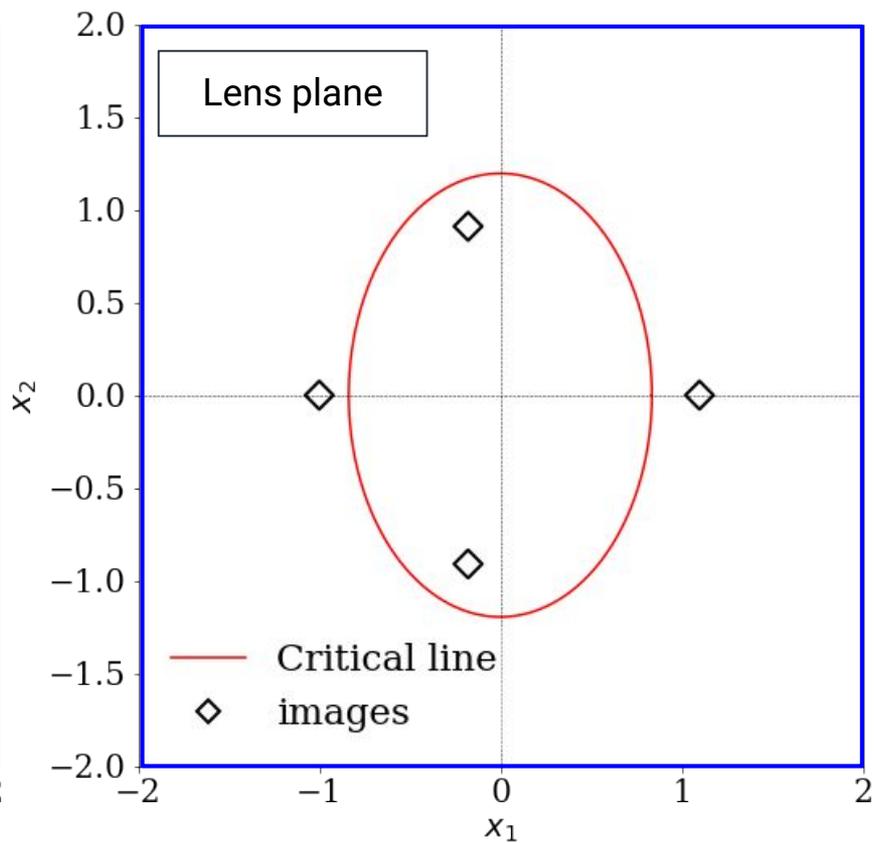
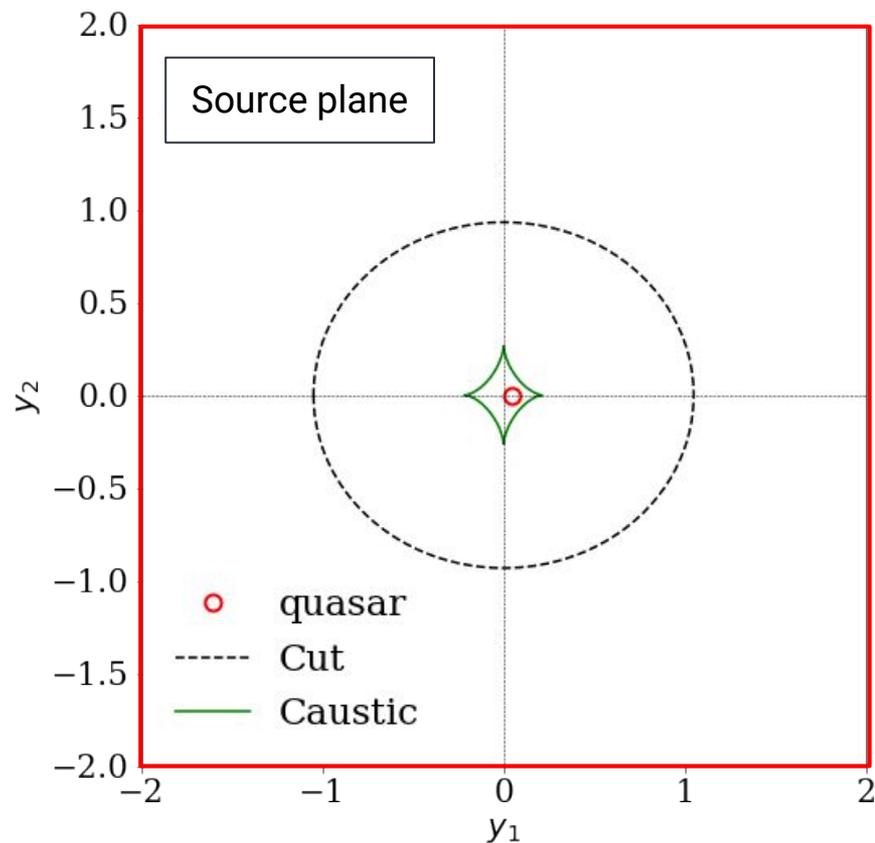
# Résultats



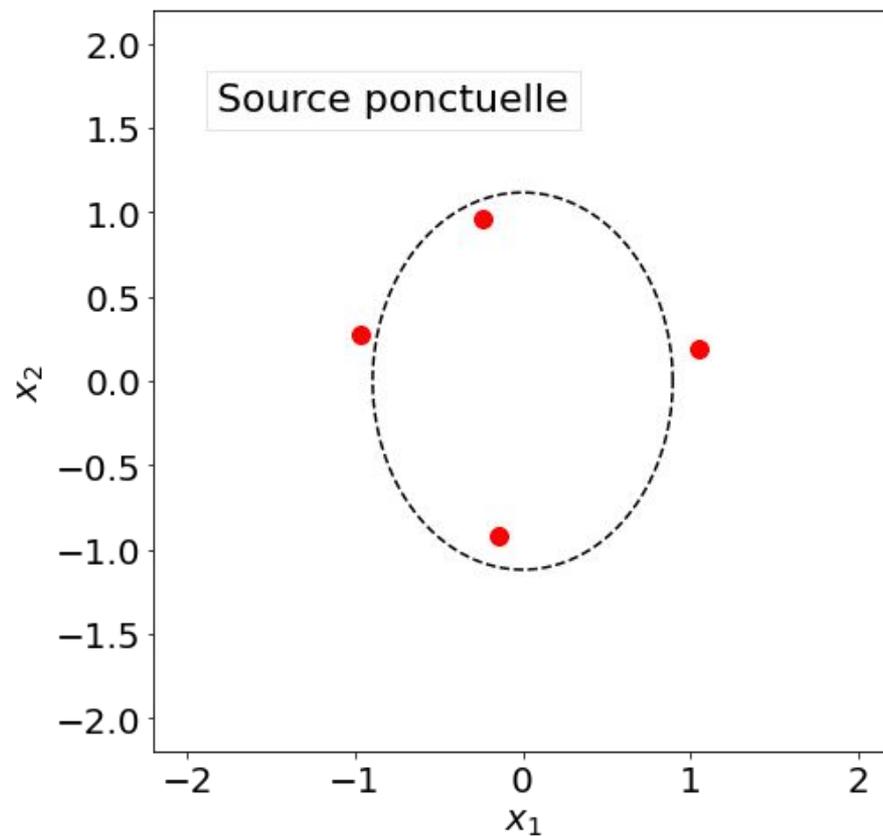
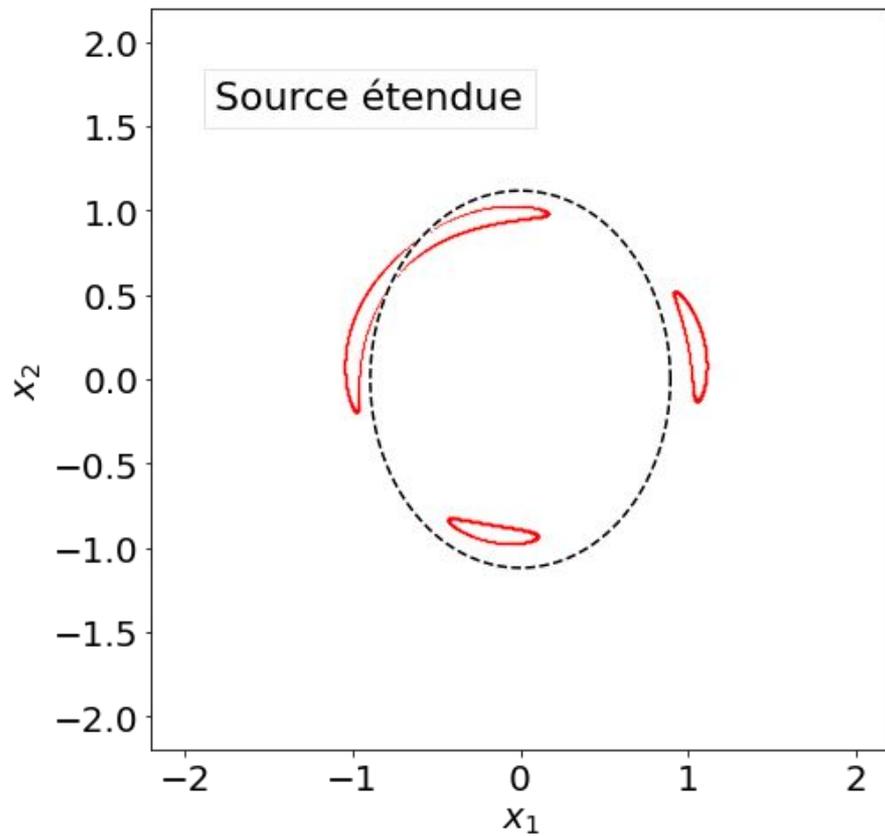
## Rapports d'axes du gaz en fonction de la masse totale des halos



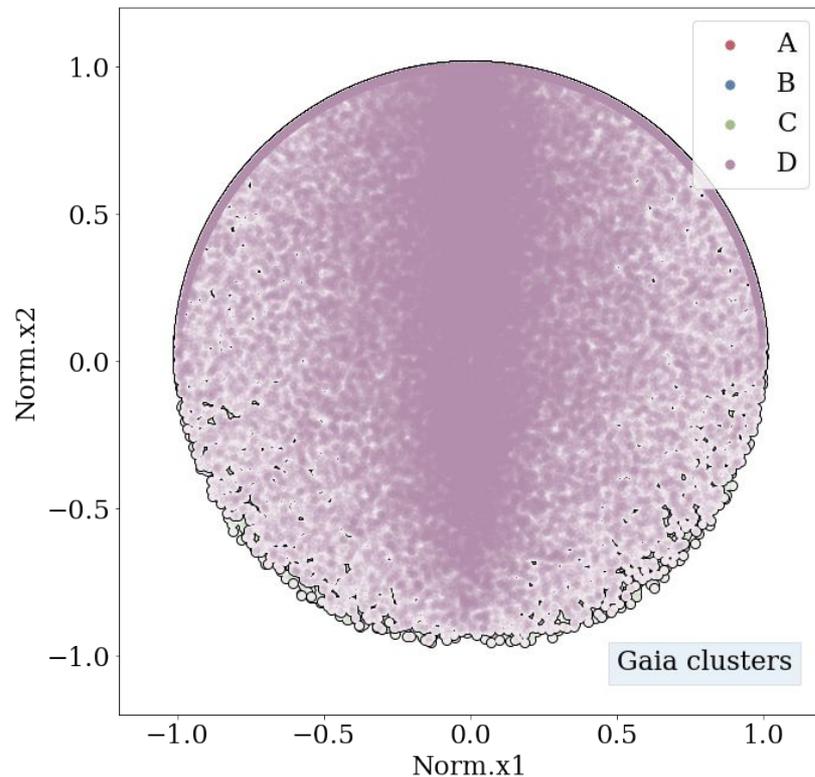
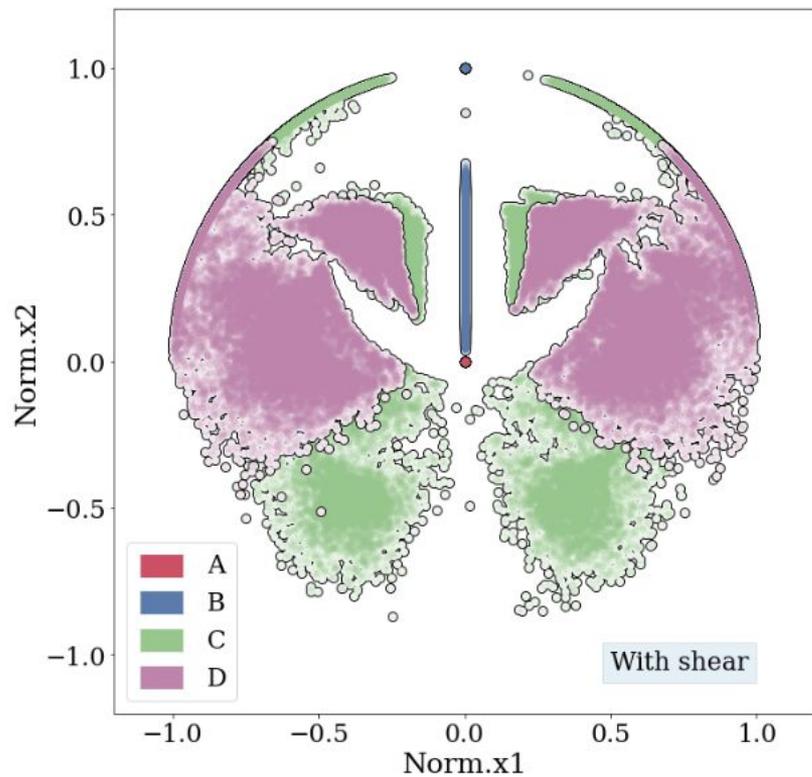
# Simulations de lentilles gravitationnelles



## Partie 2 : Simulation réaliste de quads

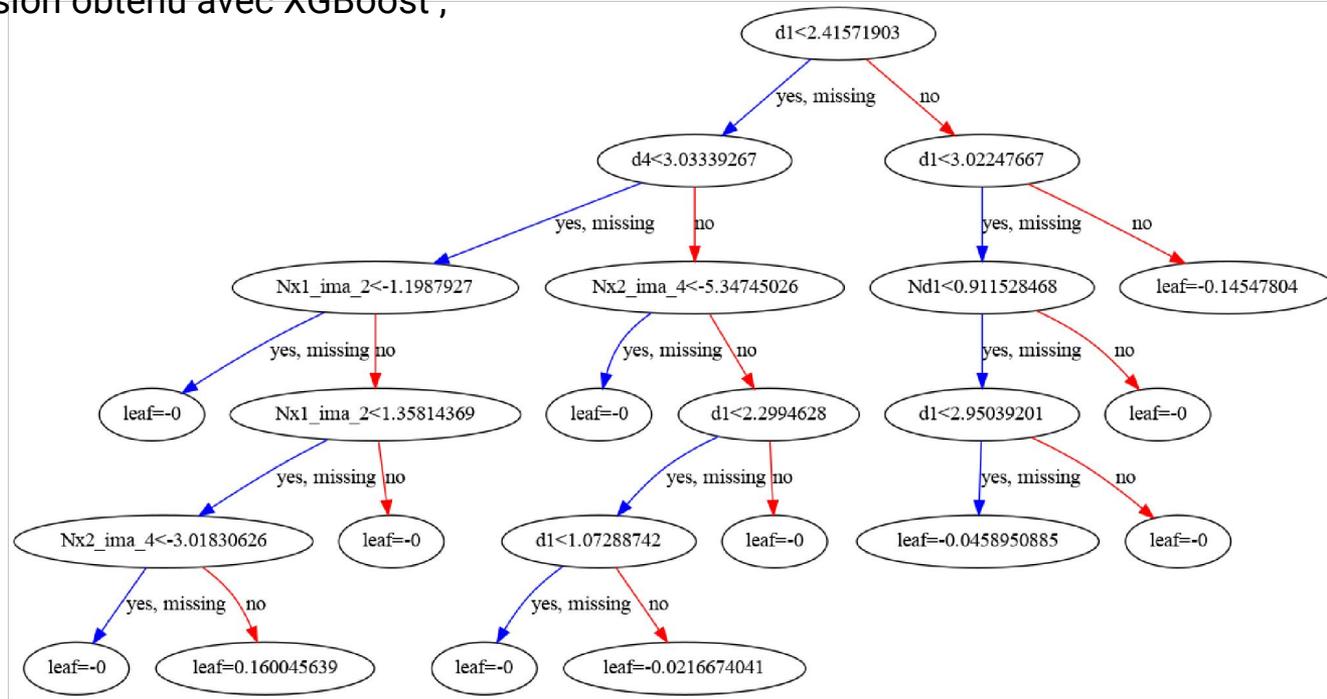


# Différence avec les groupes d'étoiles



# Apprentissage avec XGBoost

Arbre de décision obtenu avec XGBoost ;



## Relation de Faber-Jackson

---

- L'énergie de liaison gravitationnelle d'une distribution de masse :  $U = -\alpha \frac{GM^2}{R}$
- L'énergie cinétique :  $K = \frac{3}{2}M\sigma^2$

D'après le théorème du viriel :  $2K + U = 0$

$$\sigma_v \simeq \sqrt{\frac{GM_{tot}}{R_{\frac{1}{2}}}}$$

Faber-Jackson  
relation