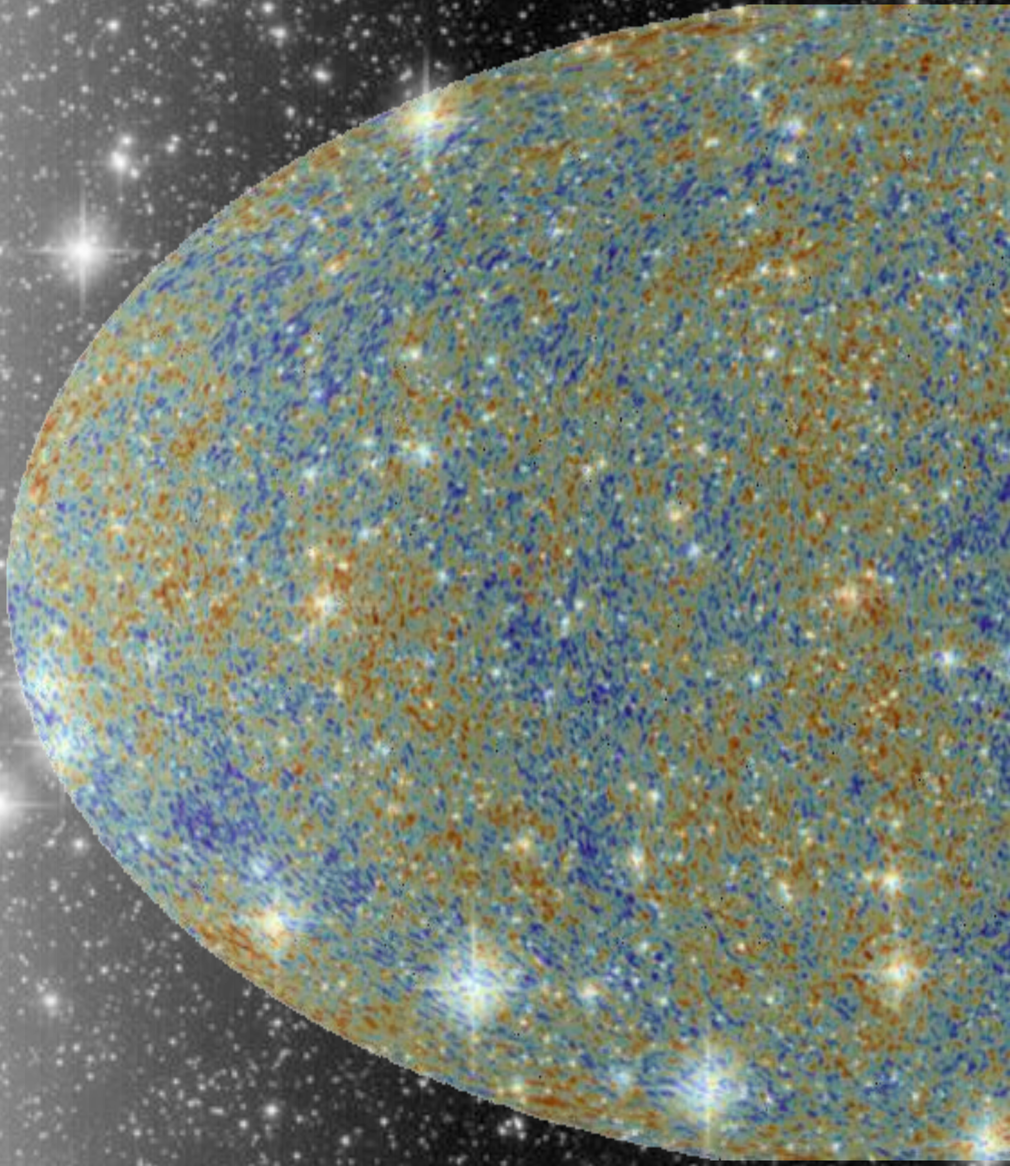
Instituto de Física Teórica UAM-CSIC

Fundación "la Caixa"

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

esa

Interpretable Neural Networks for testing Beyond-ΛCDM scenarios with CMB and LSS data



Indira Ocampo Justiniano

IFT UAM CSIC – Madrid

Les Houches

14th of July. 2025

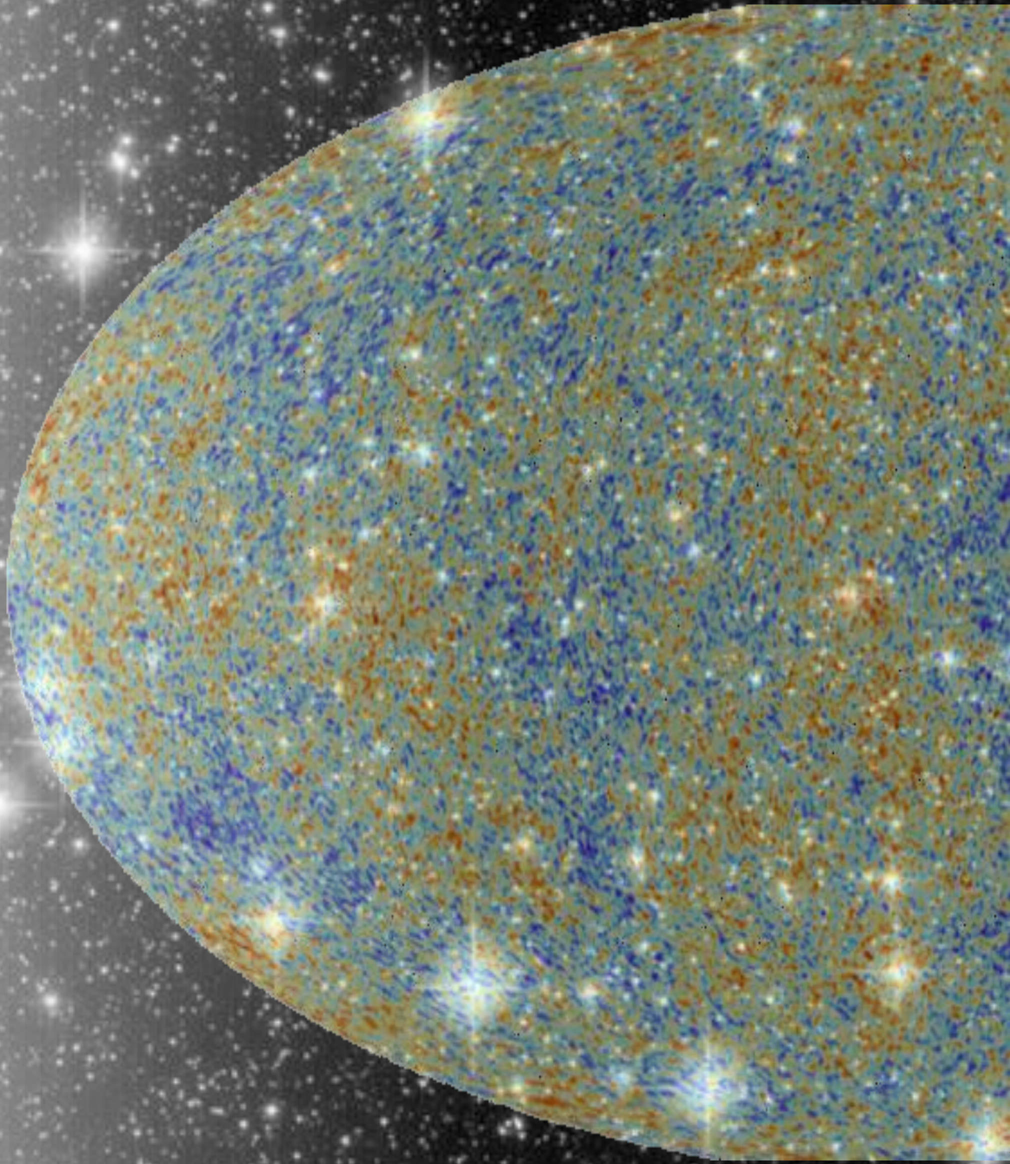
Interpretable Neural Networks for Cosmology

Indira Ocampo Justiniano

IFT UAM CSIC – Madrid

Les Houches

14th of July, 2025

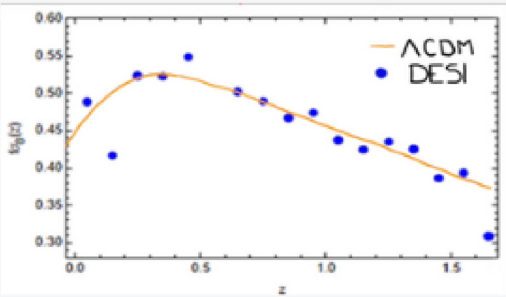


Outline

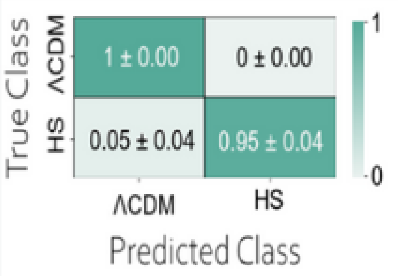
TESTING BEYOND ΛCDM SCENARIOS

CLASSIFY MODELS WITH NNS

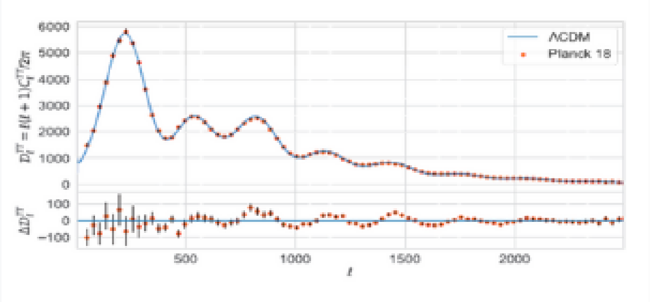
LSS



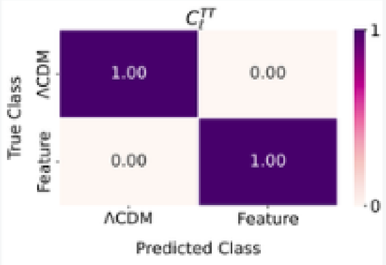
ΛCDM
VS.
HU SAWICKI

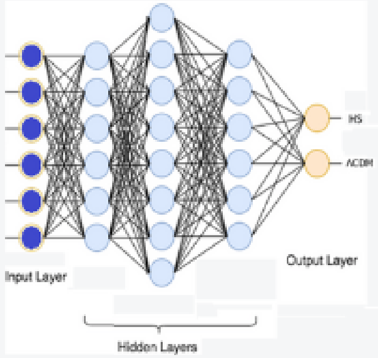


CMB



ΛCDM
VS.
FEATURE TEMPLATE





LIME TOOLS FOR INTERPRETABILITY **SHAP**

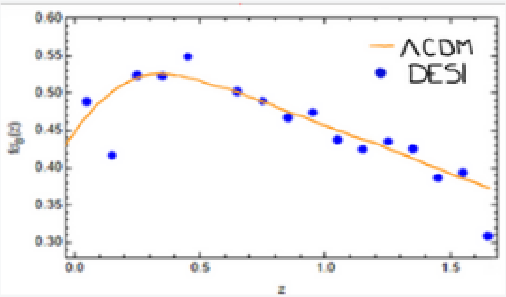
Outline

Part 1.
LSS

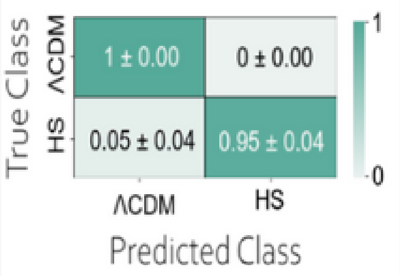
TESTING BEYOND ΛCDM SCENARIOS

CLASSIFY MODELS WITH NNS

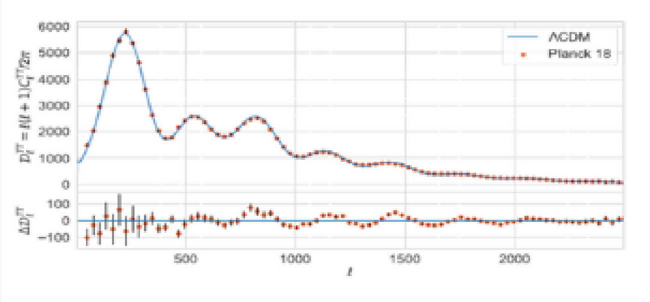
LSS



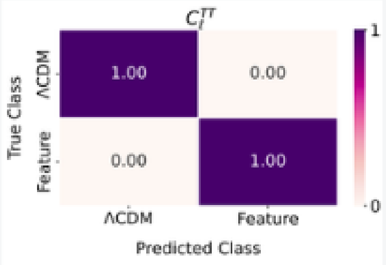
ΛCDM
VS.
HU SAWICKI

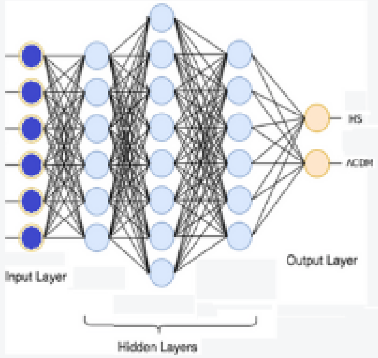


CMB



ΛCDM
VS.
FEATURE TEMPLATE





LIME TOOLS FOR INTERPRETABILITY **SHAP**

PHYSICAL REVIEW LETTERS **134**, 041002 (2025)

Enhancing Cosmological Model Selection with Interpretable Machine Learning

Indira Ocampo,[1,*] George Alestas,[1,†] Savvas Nesseris,[1,‡] and Domenico Sapone[2,§]
[1] *Instituto de Física Teórica UAM-CSIC, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*
[2] *Departamento de Física, FCFM, Universidad de Chile, Santiago, Chile*

(Received 20 June 2024; revised 12 November 2024; accepted 13 January 2025; published 31 January 2025)

We propose a novel approach using neural networks (NNs) to differentiate between cosmological models, and implemented LIME as an interpretability approach to identify the key features influencing our model's decisions. We show the potential of NNs to enhance the extraction of meaningful information from cosmological large-scale structure data, based on current galaxy-clustering survey specifications, for the cosmological constant and cold dark matter (ΛCDM) model and the Hu-Sawicki $f(R)$ model. We find that the NN can successfully distinguish between ΛCDM and the $f(R)$ models, by predicting the correct model with approximately 97% overall accuracy, thus demonstrating that NNs can maximize the potential of current and next generation surveys to probe for deviations from general relativity.

DOI: 10.1103/PhysRevLett.134.041002


George Alestas


Domenico Sapone


Savvas Nesseris

f(R) family – Hu Sawicki model

$$f(R) = R - \frac{2\Lambda}{1 + \left(\frac{b\Lambda}{R}\right)^n}$$

Extension of GR: R → R + f(R)

$$R = 6(\dot{H} + 2H^2).$$

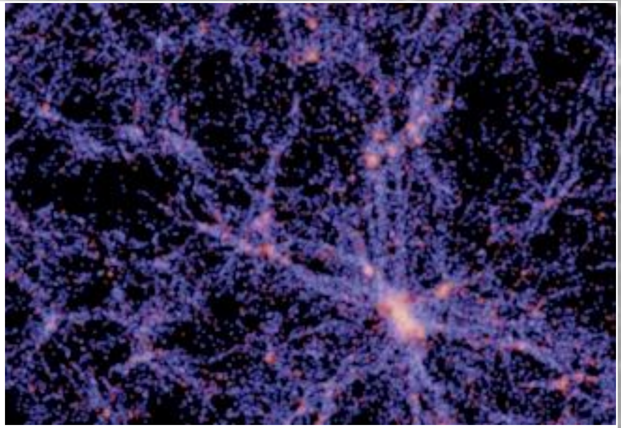
→ Screening mechanisms

For n = 2 → HS passes the Solar system tests.

→ HS model can be considered as a **small perturbation around ΛCDM.**

$$\lim_{b \to 0} f(R) = R - 2\Lambda$$

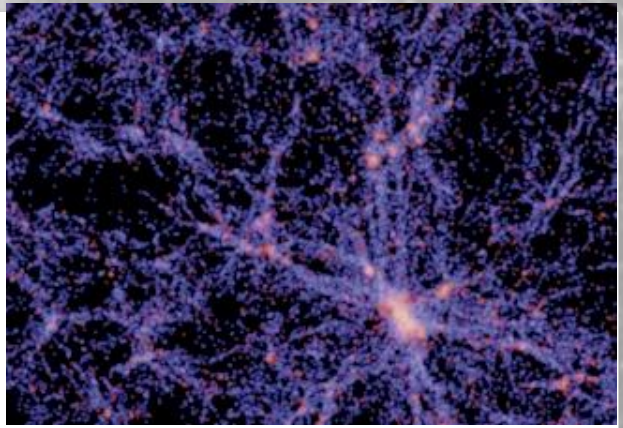
Growth of Matter Perturbations f



- Study **LSS** through perturbation theory:

  density: $\rho = \bar{\rho} + \delta\rho,$ pressure $P = \bar{P} + \delta P$ and $\delta_m \equiv \frac{\delta\rho}{\rho}$

Growth of Matter Perturbations f



- Study **LSS** through perturbation theory:

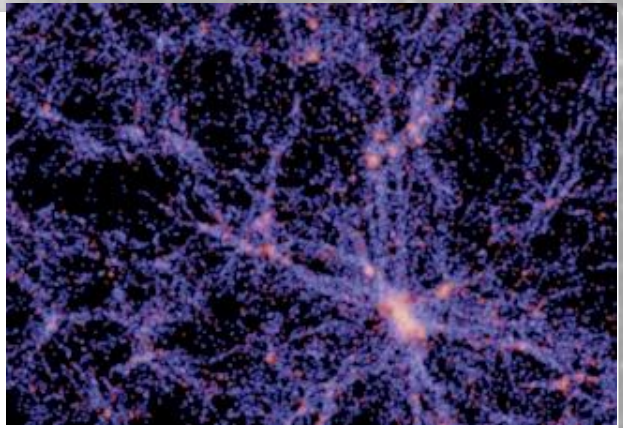
  density: $\rho = \bar{\rho} + \delta\rho,$ pressure $P = \bar{P} + \delta P$ and $\delta_m \equiv \frac{\delta\rho}{\rho}$

  We study the eq:

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G_{\rm eff}\rho\,\delta_m \approx 0$$

(evolution of the matter density perturbations).

Growth of Matter Perturbations f



- Study **LSS** through perturbation theory:

density: $\rho = \bar{\rho} + \delta\rho,$ pressure $P = \bar{P} + \delta P$ and $\delta_m \equiv \frac{\delta\rho}{\rho}$

We study the eq:

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G_{\rm eff}\,\rho\,\delta_m \approx 0$$

(evolution of the matter density perturbations).

With a solution (for ΛCDM, $G_{\rm eff}$ = 1):

$$\delta_{\rm m}(a) = a \cdot {}_2F_1\left(\frac{1}{3}, 1; \frac{11}{6}; a^3\left(1 - \frac{1}{\Omega_{\rm m,0}}\right)\right) \longrightarrow f = \frac{d\ln\delta_m}{d\ln a}$$

The growth $f\sigma_8$

In galaxy surveys we observe the galaxy density fluctuations

$$\delta_g = b\,\delta_m$$

The growth in a bias independent way

$$f\sigma_8(z) \equiv f(z)\sigma_8(z)$$

$$f = \frac{d\ln\delta_m}{d\ln a}$$

The growth $f\sigma_8$

In galaxy surveys we observe the galaxy density fluctuations

$$\delta_g = b\,\delta_m$$

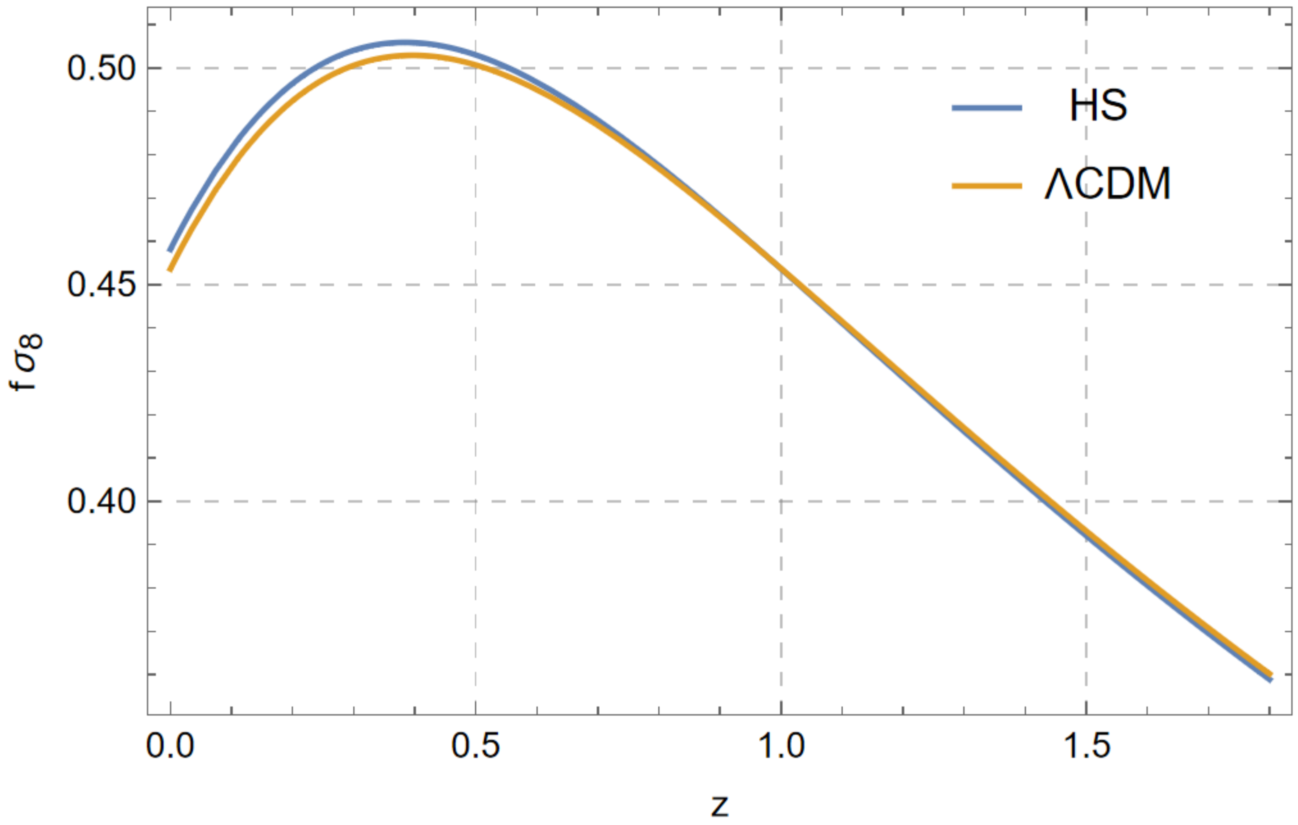
The growth in a bias independent way

$$f\sigma_8(z) \equiv f(z)\sigma_8(z)$$

$$\delta_{\rm m}(a) = a \cdot {}_2F_1\left(\frac{1}{3}, 1; \frac{11}{6}; a^3\left(1 - \frac{1}{\Omega_{\rm m,0}}\right)\right)$$

For ΛCDM, G_{eff} = 1

For f(R) $G_{\rm eff} = \frac{G}{F}\left[\frac{4}{3} - \frac{1}{3}\frac{M^2a^2}{k^2 + M^2a^2}\right]$,



$$f(R) = R - \frac{2\Lambda}{1 + \left(\frac{b\Lambda}{R}\right)^n}$$

Dataset simulation strategy

$f\sigma_8$ values w/ uncertainties. Cosmological parameters varied as:

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G_{\text{eff}}\rho\,\delta_m \approx 0 \qquad \delta_m \equiv \frac{\delta\rho}{\rho}$$

$$f = \frac{d\ln\delta_m}{d\ln a}$$

ΛCDM

$\sigma_8 \in [0.7, 0.9]$

$\Omega_m \in [0.2, 0.4]$

Hu Sawicki - f(R)

$\sigma_8 \in [0.7, 0.9]$, $\Omega_m \in [0.2, 0.4]$

b ∈ [10−5 , 5 × 10−5]

Dataset simulation strategy

$$f(R) = R - \frac{2\Lambda}{1 + \left(\frac{b\Lambda}{R}\right)^n}$$

$f\sigma_8$ values w/ uncertainties.
Cosmological parameters varied as:

ΛCDM

$\sigma_8 \in [0.7, 0.9]$

$\Omega_m \in [0.2, 0.4]$

Hu Sawicki - f(R)

$\sigma_8 \in [0.7, 0.9]$, $\Omega_m \in [0.2, 0.4]$

$b \in [10^{-5}, 5 \times 10^{-5}]$

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G_{\text{eff}}\rho\,\delta_m \approx 0 \qquad \delta_m \equiv \frac{\delta\rho}{\rho}$$

$$f = \frac{d\ln\delta_m}{d\ln a}$$

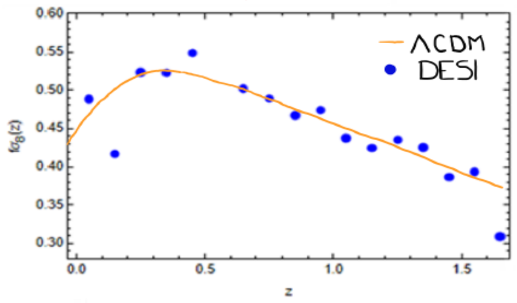
$$f\sigma_8(a) = a\frac{\delta'_{\text{m}}(a)}{\delta_{\text{m}}(1)} \cdot \sigma_{8,0}. \qquad a = \frac{1}{1+z}$$

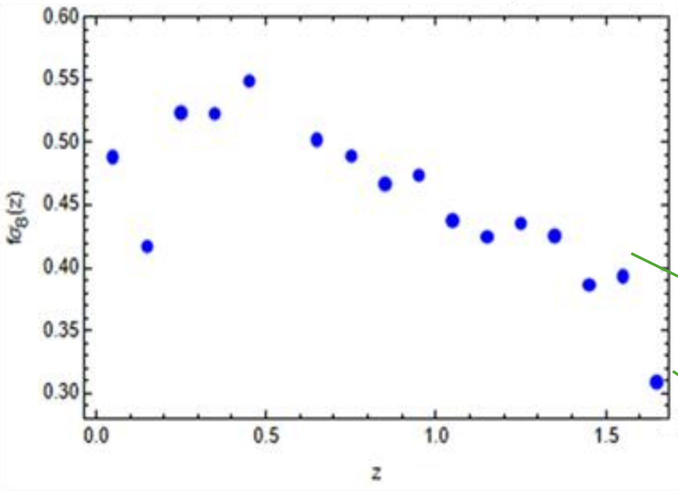
$z \in [0.05, 1.85]$ (16 z-bins)

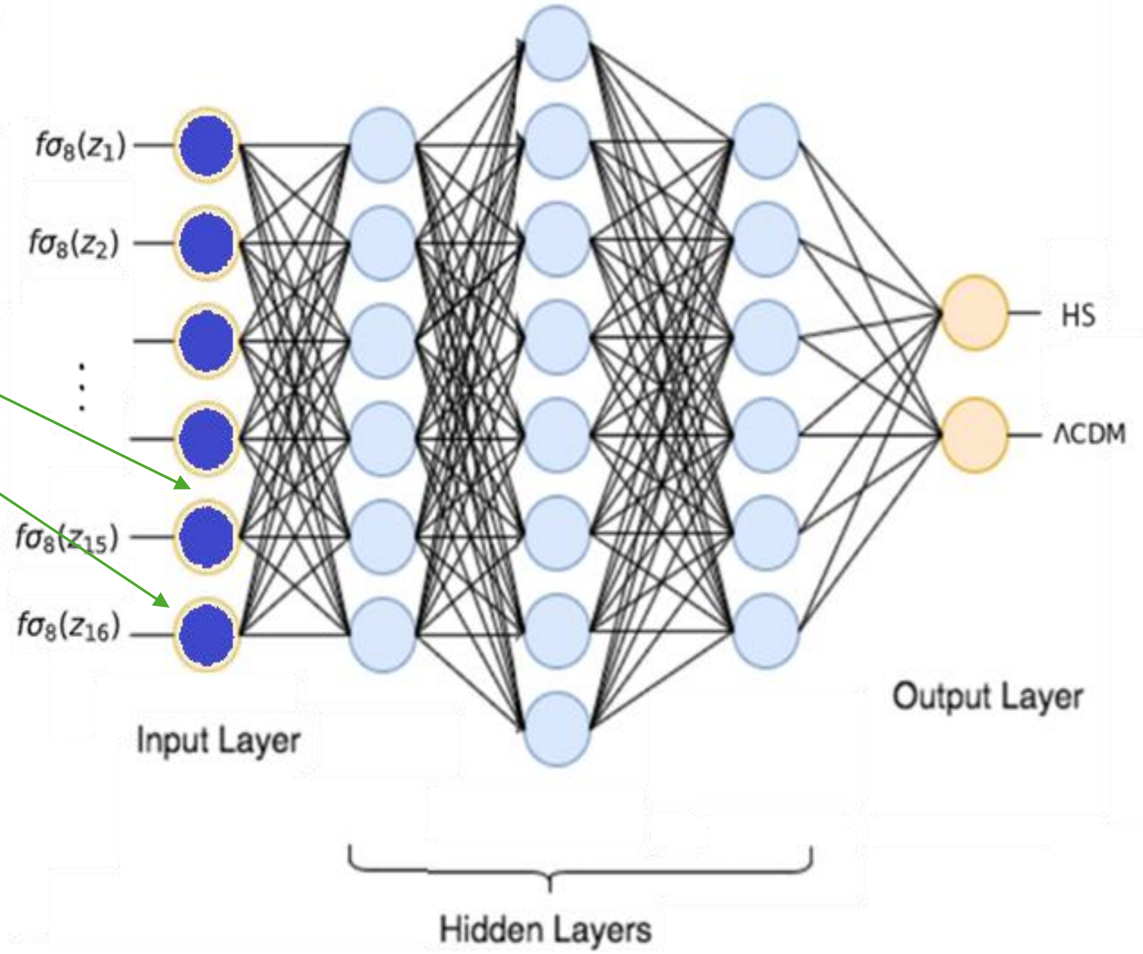
+

$\sigma_{f\sigma_8}(a)$ (DESI-like Cij)

Machine Learning analysis





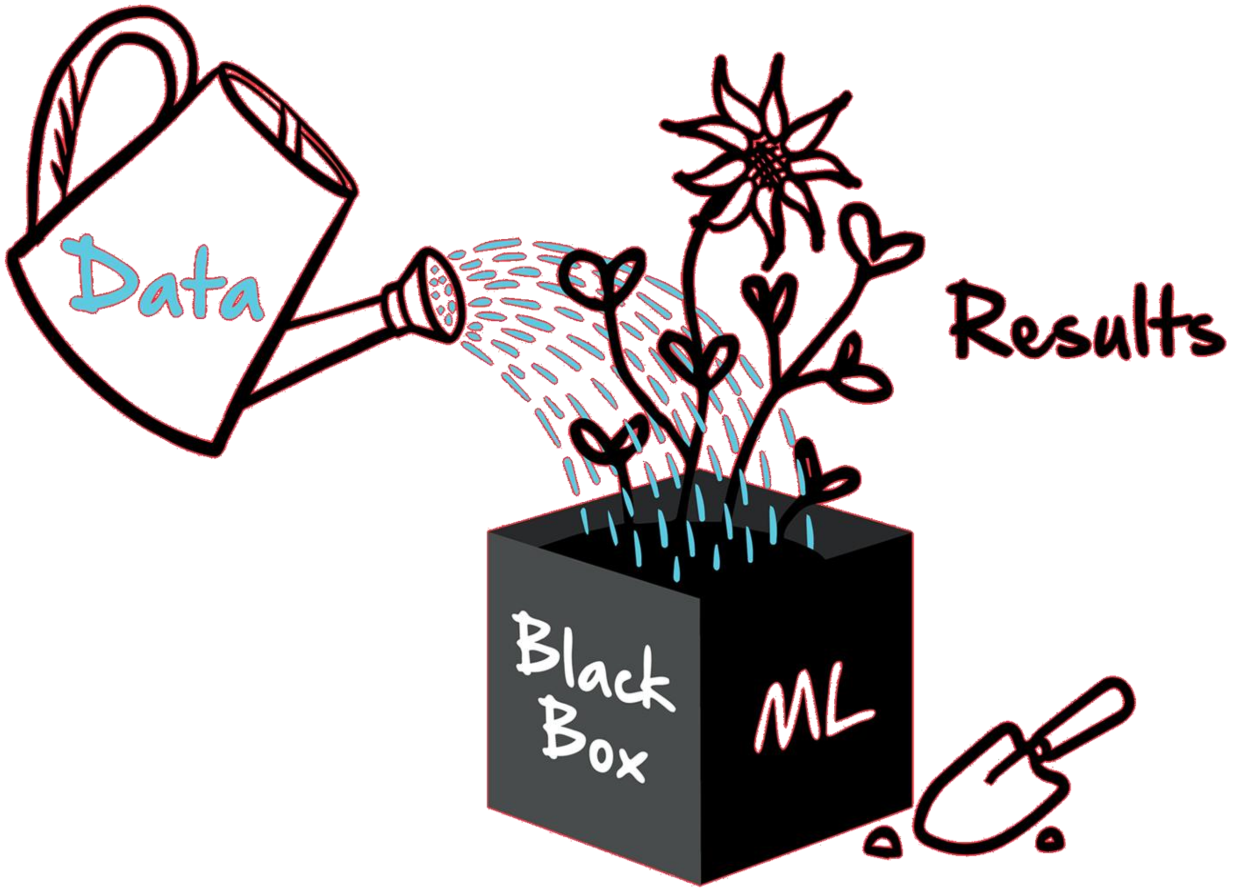


Model independent framework:

Data

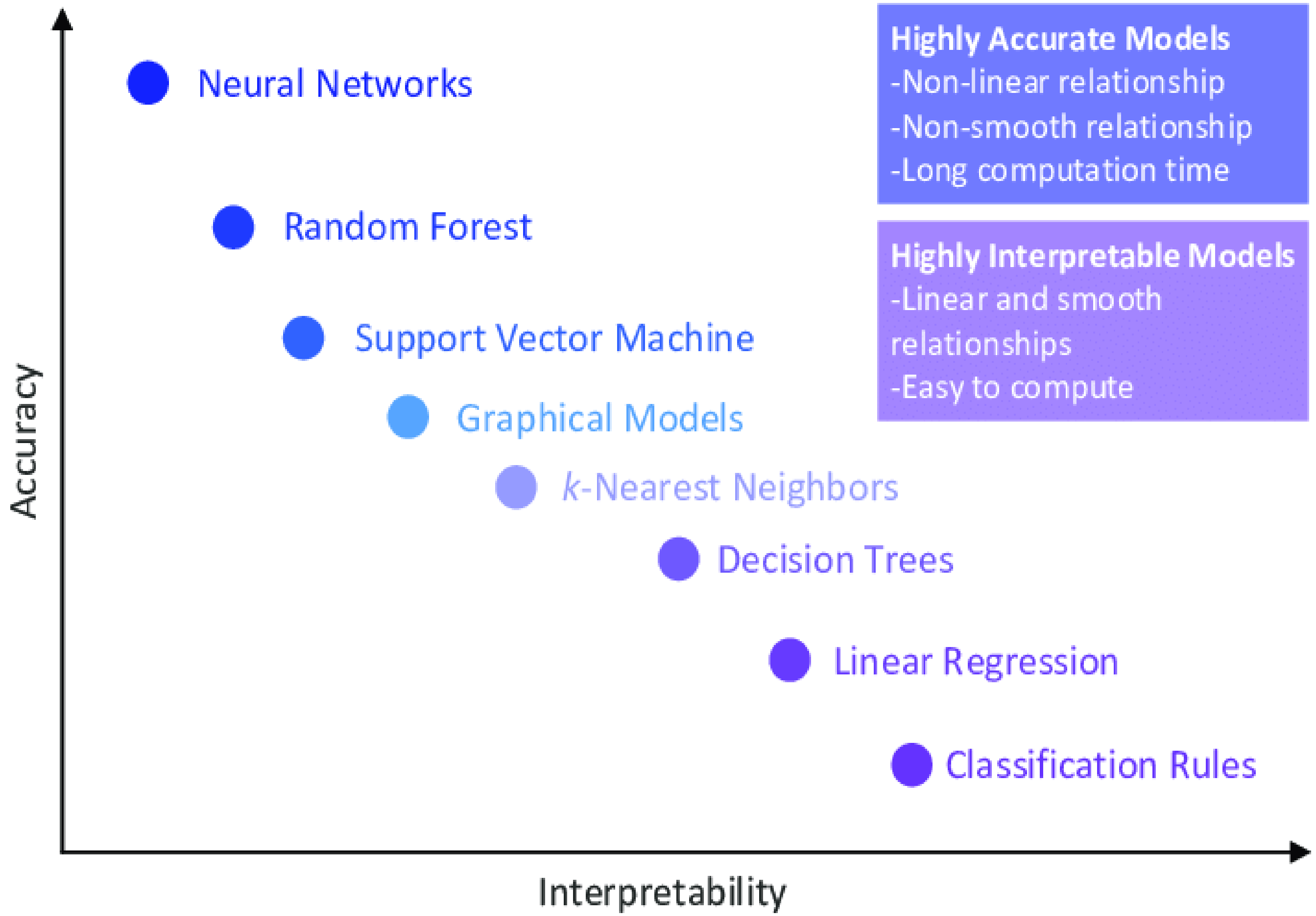
↓

Theory



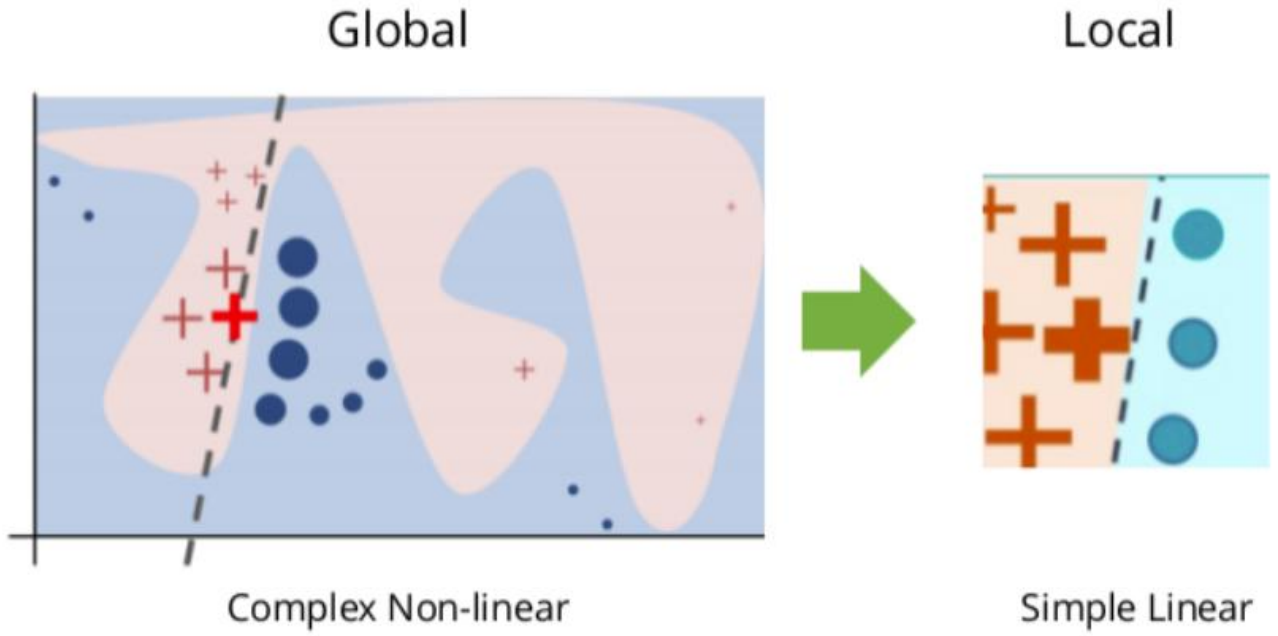
Source: https://simons.berkeley.edu/

Interpretable Machine Learning



Source: An Introduction to Interpretable Machine Learning Models: analyticsvidhya.com/

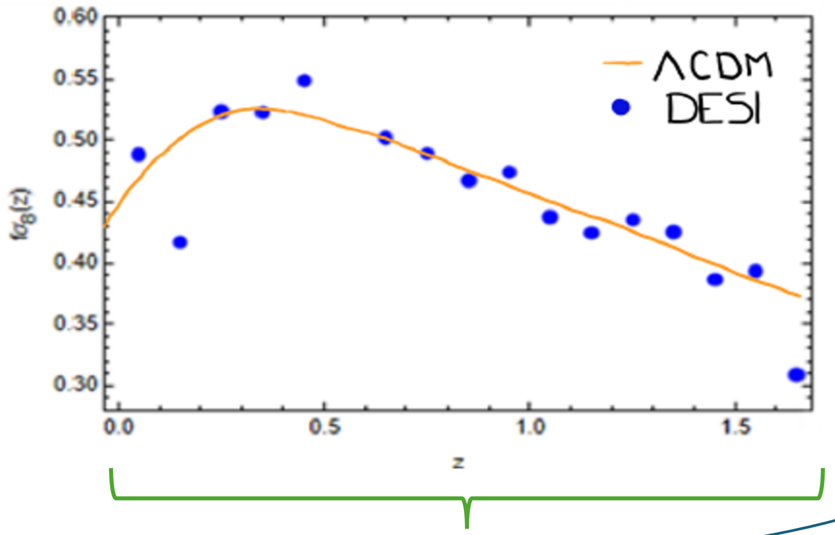
LIME (Local Interpretability Model agnostic Explanations)

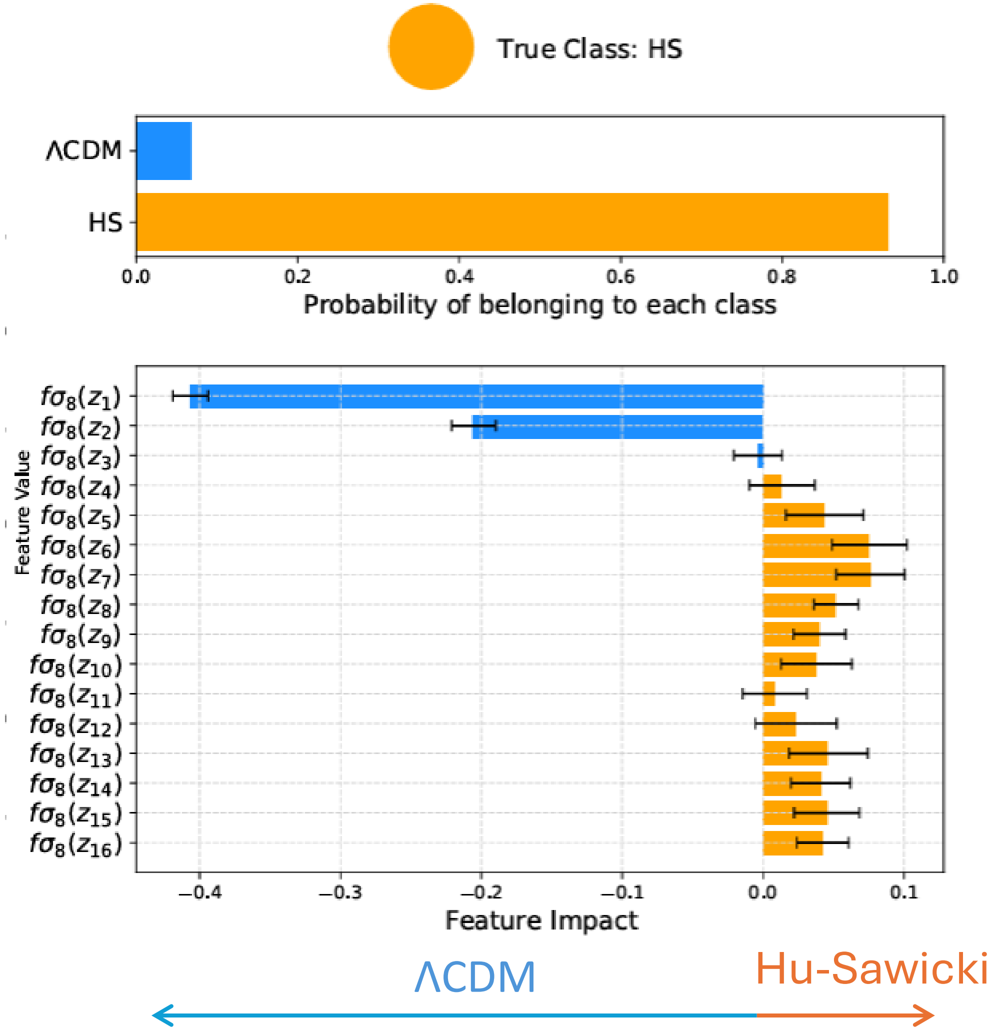


Ribeiro, Singh (2018)

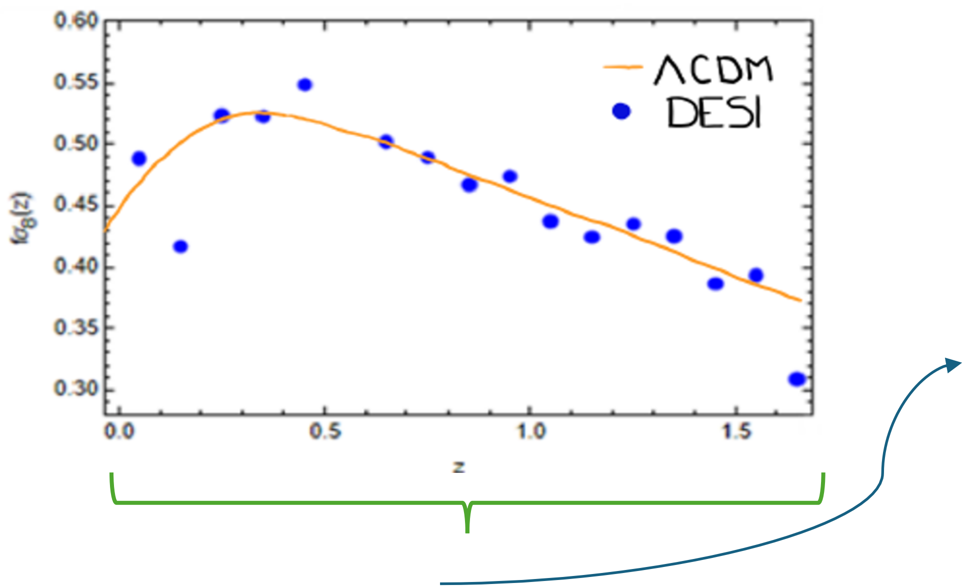
Source: Explainable AI, https://bigdatarepublic.nl/

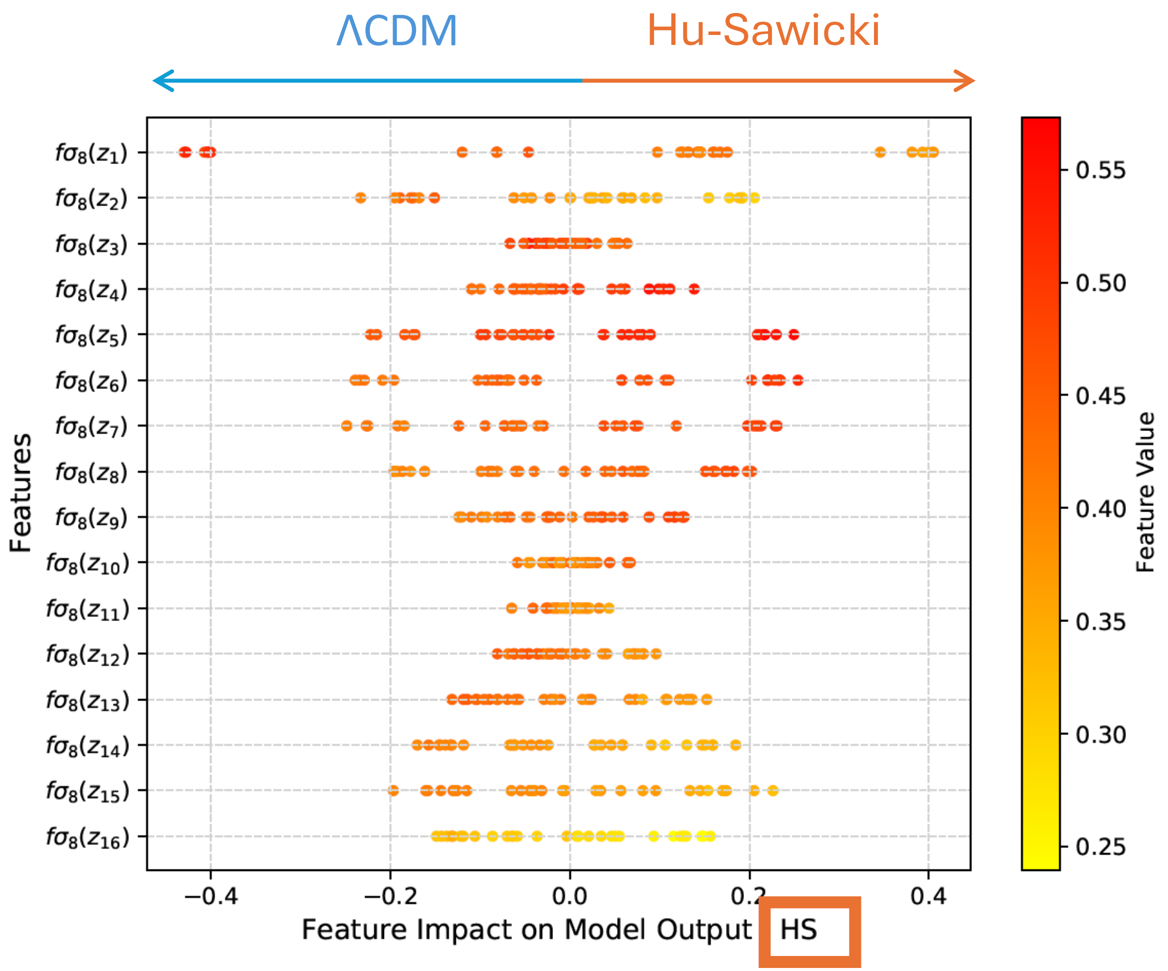
Feature Importance using LIME





Distribution of LIME feature impact

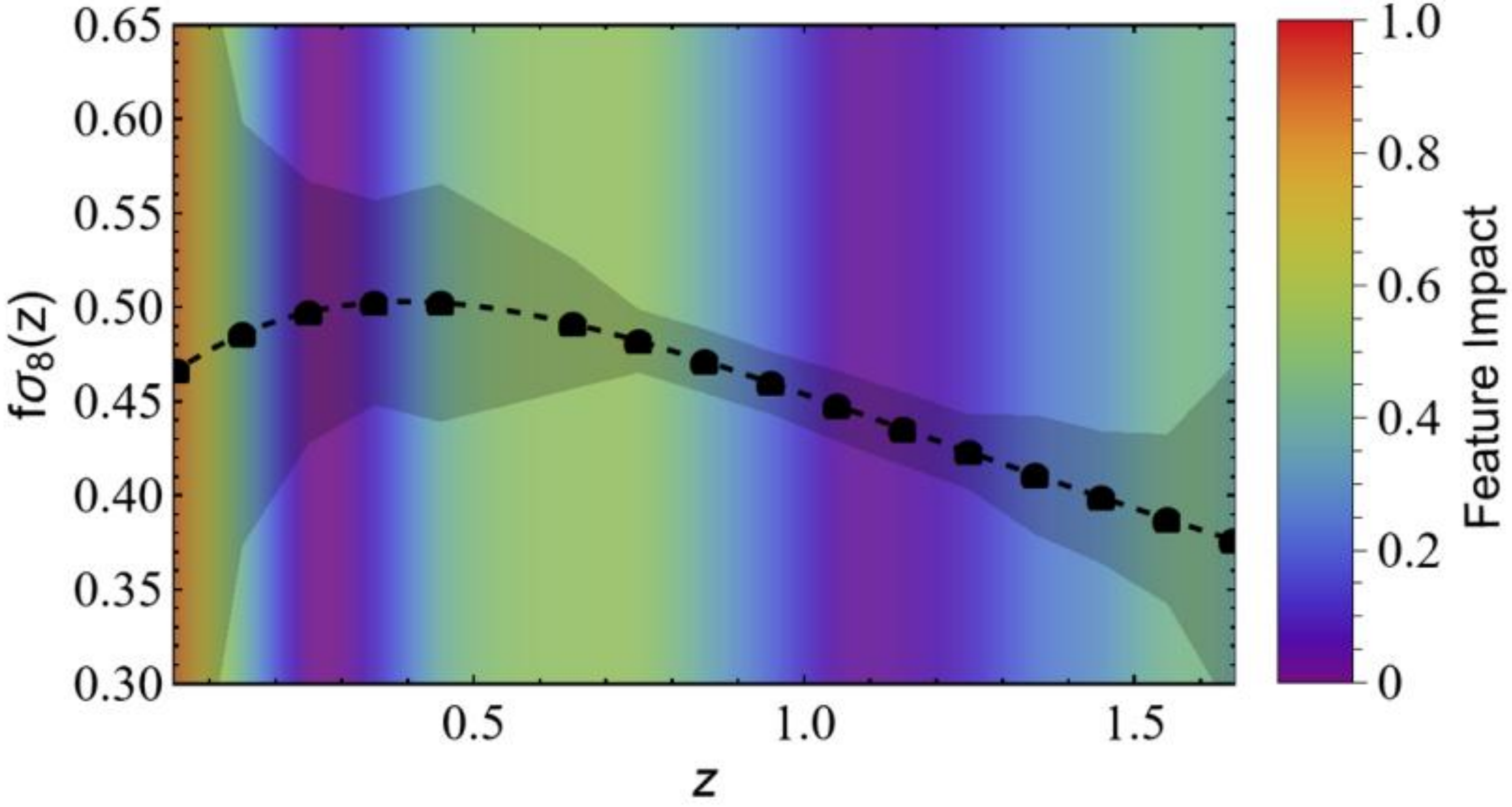




Feature Impact and Redshift for $f\sigma_8$

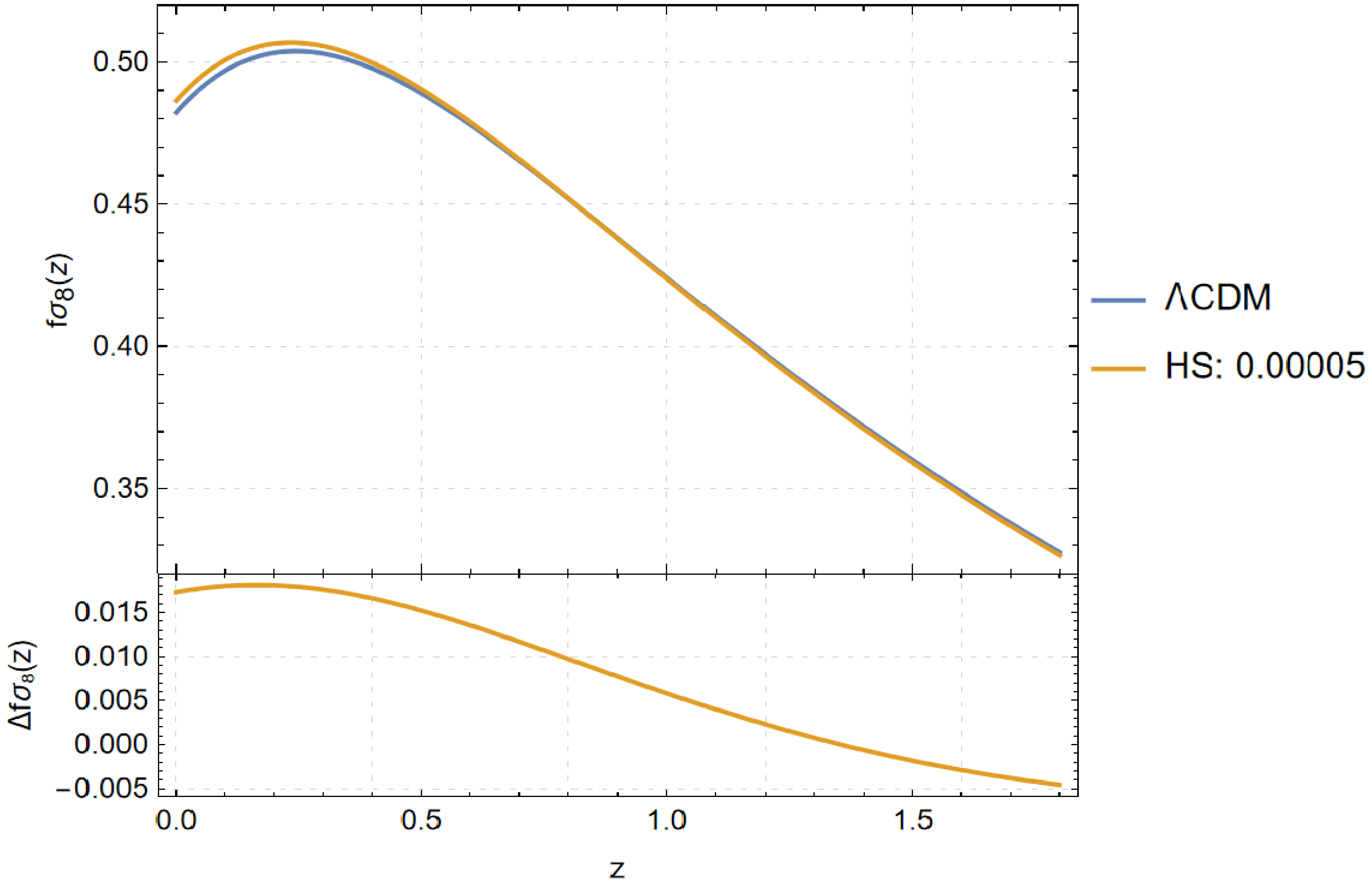
One realization of $f\sigma_8(z)$.

Rainbow color code: “feature impact” of each z-bin according to LIME.



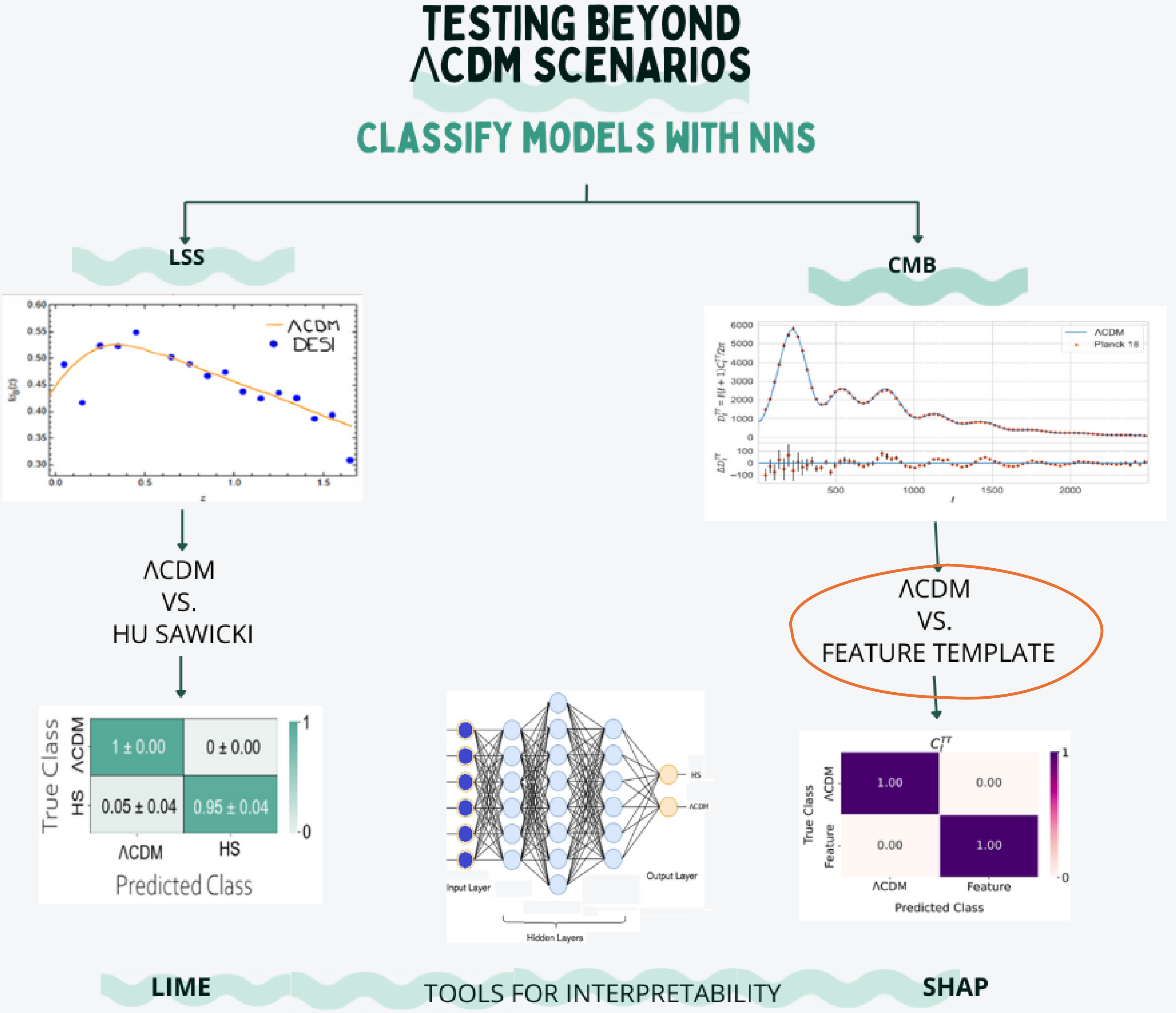
The Hu Sawicki model and $f\sigma_8$

$$f(R) = R - \frac{2\Lambda}{1 + \left(\frac{b\Lambda}{R}\right)^n}$$



Outline

Part 1.
LSS



Part 2.
CMB

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Neural Networks for cosmological model selection and feature importance using Cosmic Microwave Background data

I. Ocampo,[a] G. Cañas-Herrera[b] and S. Nesseris[a]

[a] *Instituto de Física Teórica UAM-CSIC, C/ Nicolás Cabrera 13–15, Cantoblanco, 28049 Madrid, Spain*

[b] *ESTEC — European Space Agency, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands*

E-mail: indira.ocampo@csic.es, Guadalupe.CanasHerrera@esa.int, savvas.nesseris@csic.es

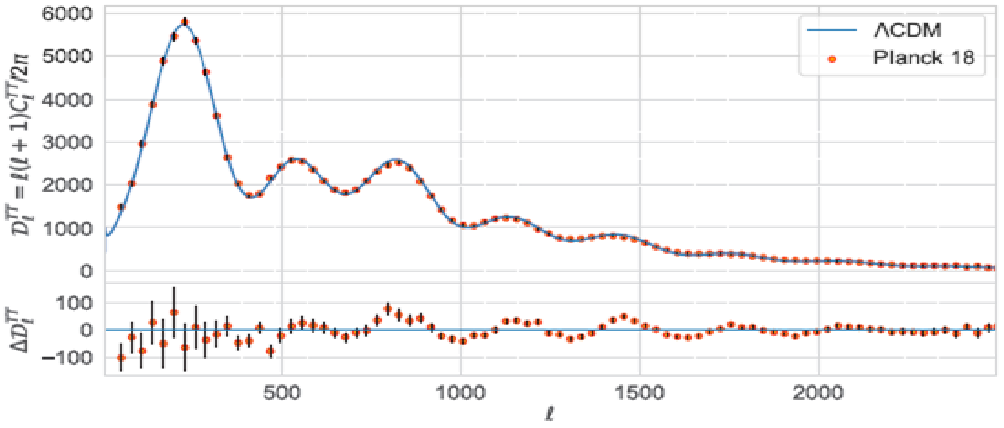
ABSTRACT: The measurements of the temperature and polarisation anisotropies of the Cosmic Microwave Background (CMB) by the ESA Planck mission have strongly supported the current concordance model of cosmology. However, the latest cosmological data release from ESA Planck mission still has a powerful potential to test new data science algorithms and inference techniques. In this paper, we use advanced Machine Learning (ML) algorithms, such as Neural Networks (NNs), to discern among different underlying cosmological models at the angular power spectra level, using both temperature and polarisation Planck 18 data. We test two different models beyond ΛCDM: a modified gravity model: the Hu-Sawicki model, and an alternative inflationary model: a feature-template in the primordial power spectrum. Furthermore, we also implemented an interpretability method based on SHAP values to evaluate the learning process and identify the most relevant elements that drive our architecture to certain outcomes. We find that our NN is able to distinguish between different angular power spectra successfully for both alternative models and ΛCDM. We conclude by explaining how archival scientific data has still a strong potential to test novel data science algorithms that are interesting for the next generation of cosmological experiments.

Guadalupe Cañas-Herrera

Savvas Nesseris

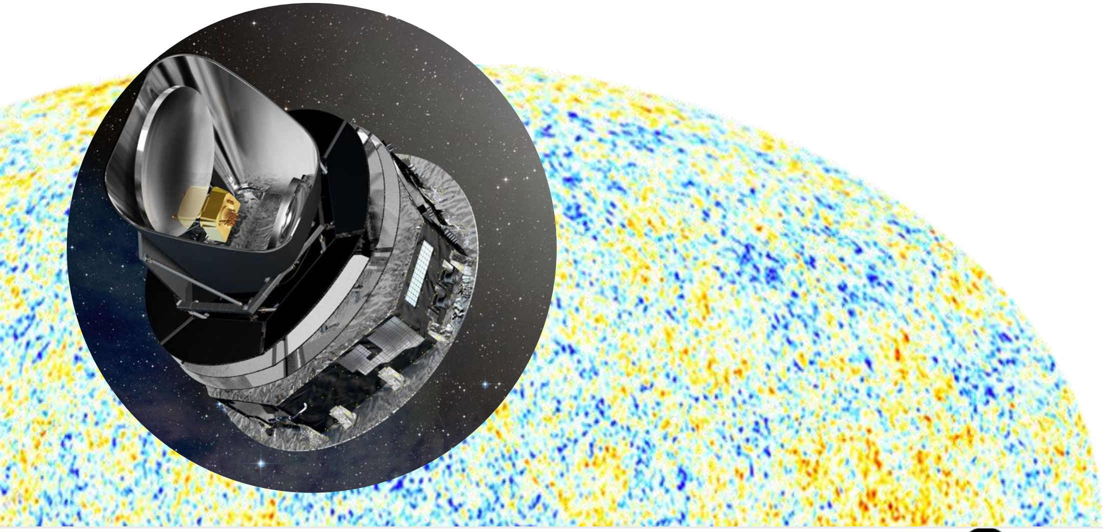
Angular Power Spectra and Planck

$$C_\ell^{TT} = \frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell} |a_{\ell m}|^2$$



$$\frac{\Delta T}{T} \sim 10^{-5}$$

$$\frac{\Delta T}{T}(\hat{n}) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n})$$



Angular Power Spectra and Planck

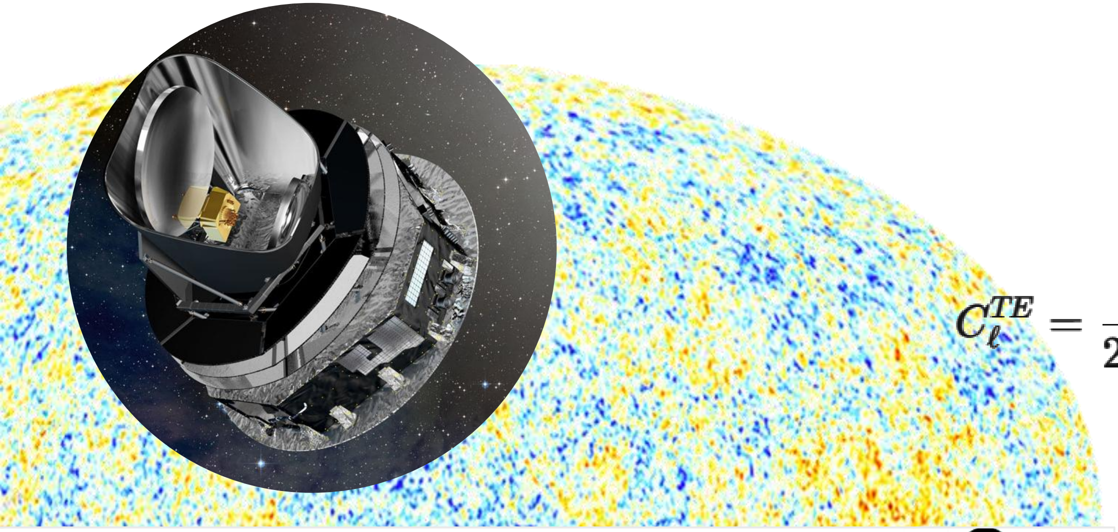
$$\frac{\Delta T}{T} \sim 10^{-5}$$

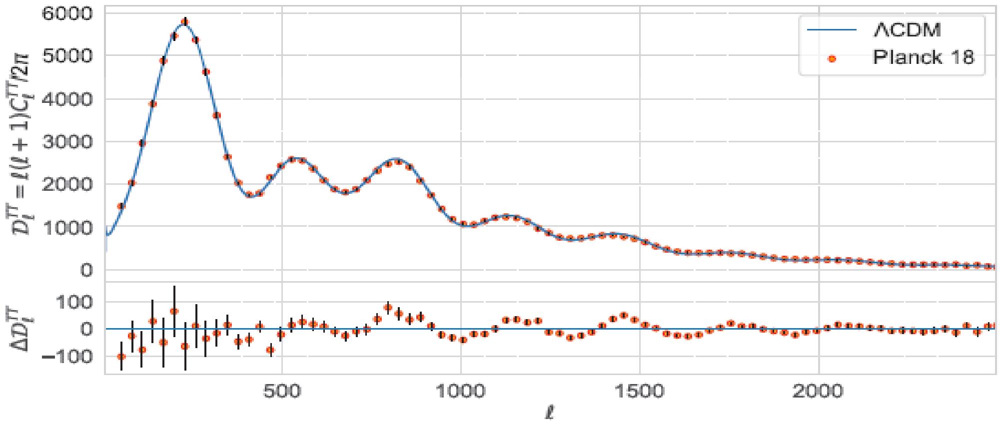
$$\frac{\Delta T}{T}(\hat{n}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n})$$

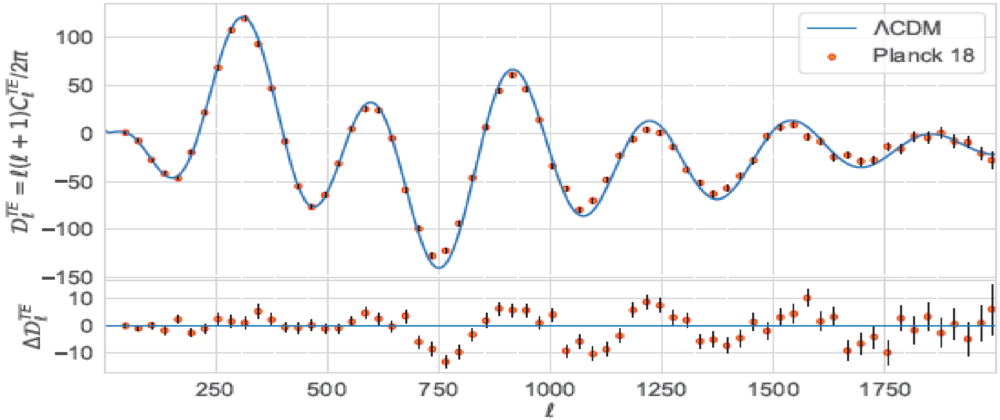
$$C_\ell^{TT} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2$$

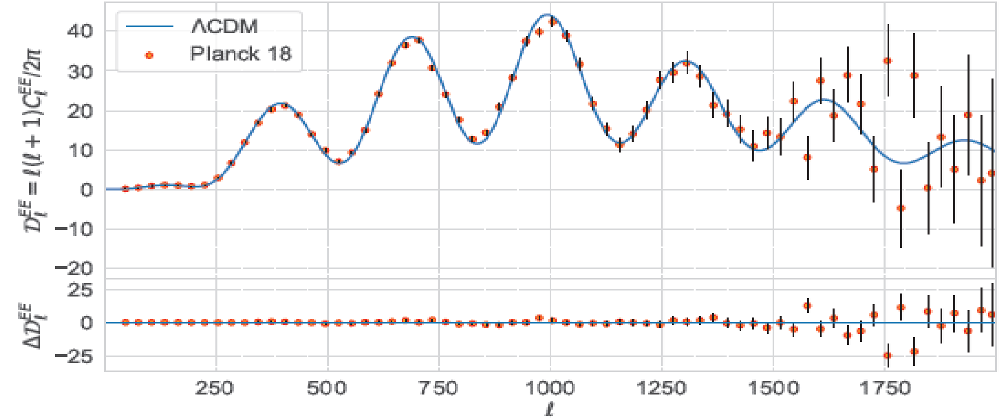
$$C_\ell^{EE} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}^E|^2$$

$$C_\ell^{TE} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m}^T a_{\ell m}^{E*}$$









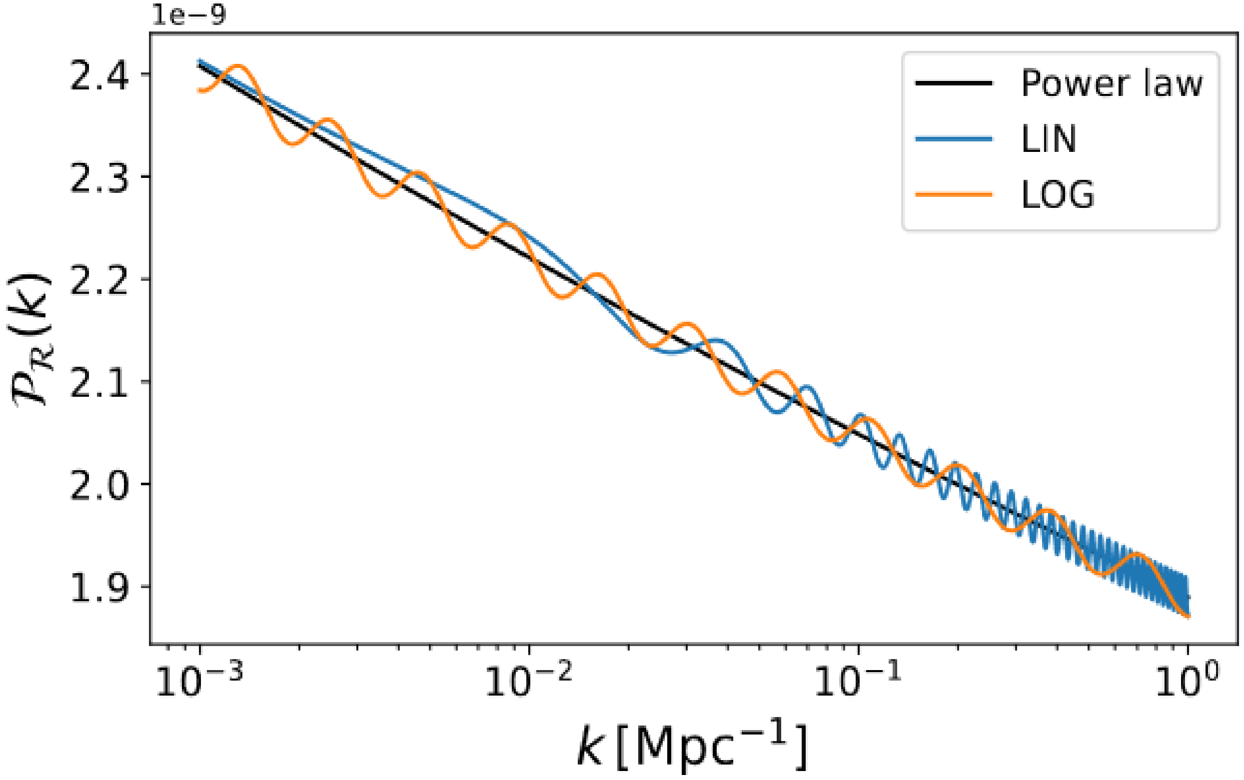
Primordial power spectrum:

Inflation predicts a power law:

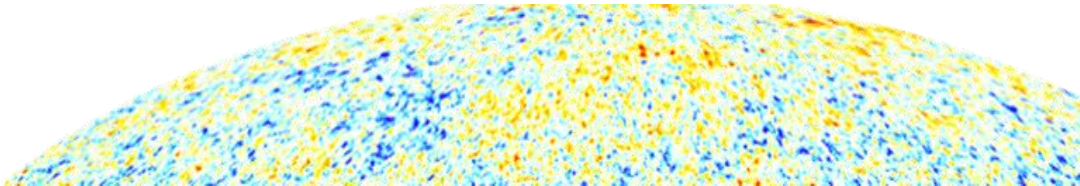
$$P(k) = A_s k^{n_s - 1}$$

As: (from CMB As~2.1×10−9).

ns: (ns = 0.965 ± 0.004).



Source: Euclid consortium 2309.17287

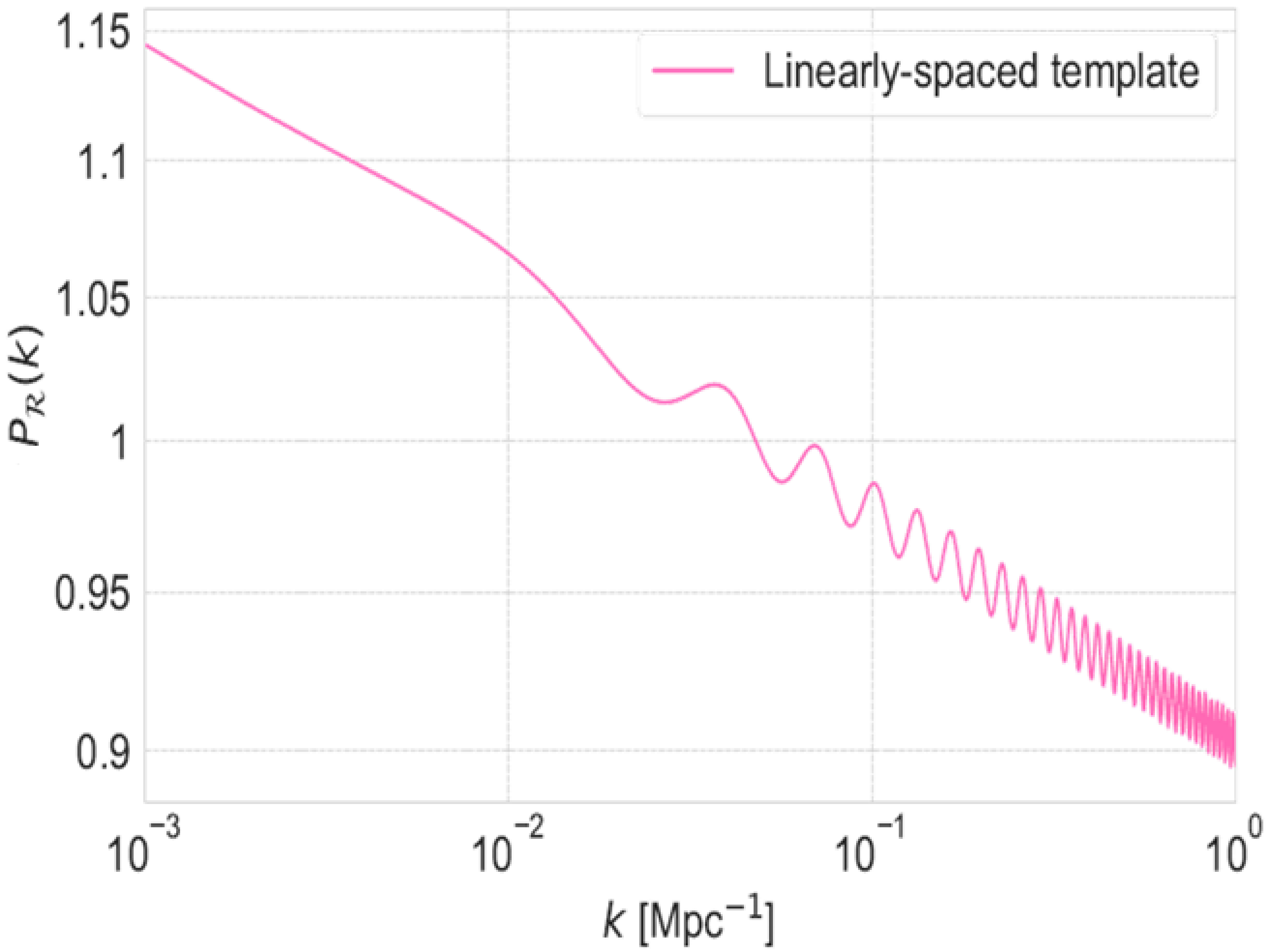


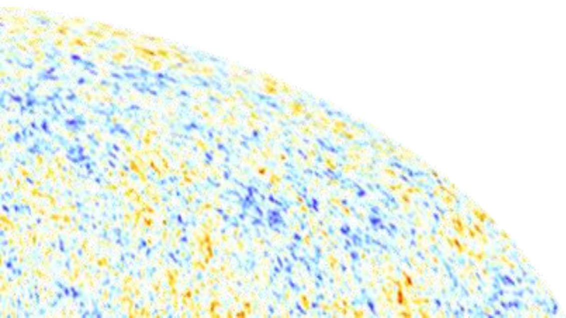
Linearly Spaced feature template:

$$C_\ell^{XY} = \frac{2}{\pi}\int k^2 \mathrm{d}k \underbrace{P(k)}_{\text{Matter power spectrum}} \underbrace{\Delta_{X\ell}(k)\Delta_{Y\ell}(k)}_{\text{Transfer functions}}$$

Primordial features parametrized as small deviations,

$$P_{\mathcal{R}}(k) = P_{\mathcal{R},0}(k)\left[1 + \frac{\Delta P_{\mathcal{R}}}{P_{\mathcal{R},0}}(k)\right],$$





Linearly Spaced feature template:

$$C_\ell^{XY} = \frac{2}{\pi}\int k^2 \mathrm{d}k \underbrace{P(k)}_{\text{Matter power spectrum}} \underbrace{\Delta_{X\ell}(k)\Delta_{Y\ell}(k)}_{\text{Transfer functions}}$$

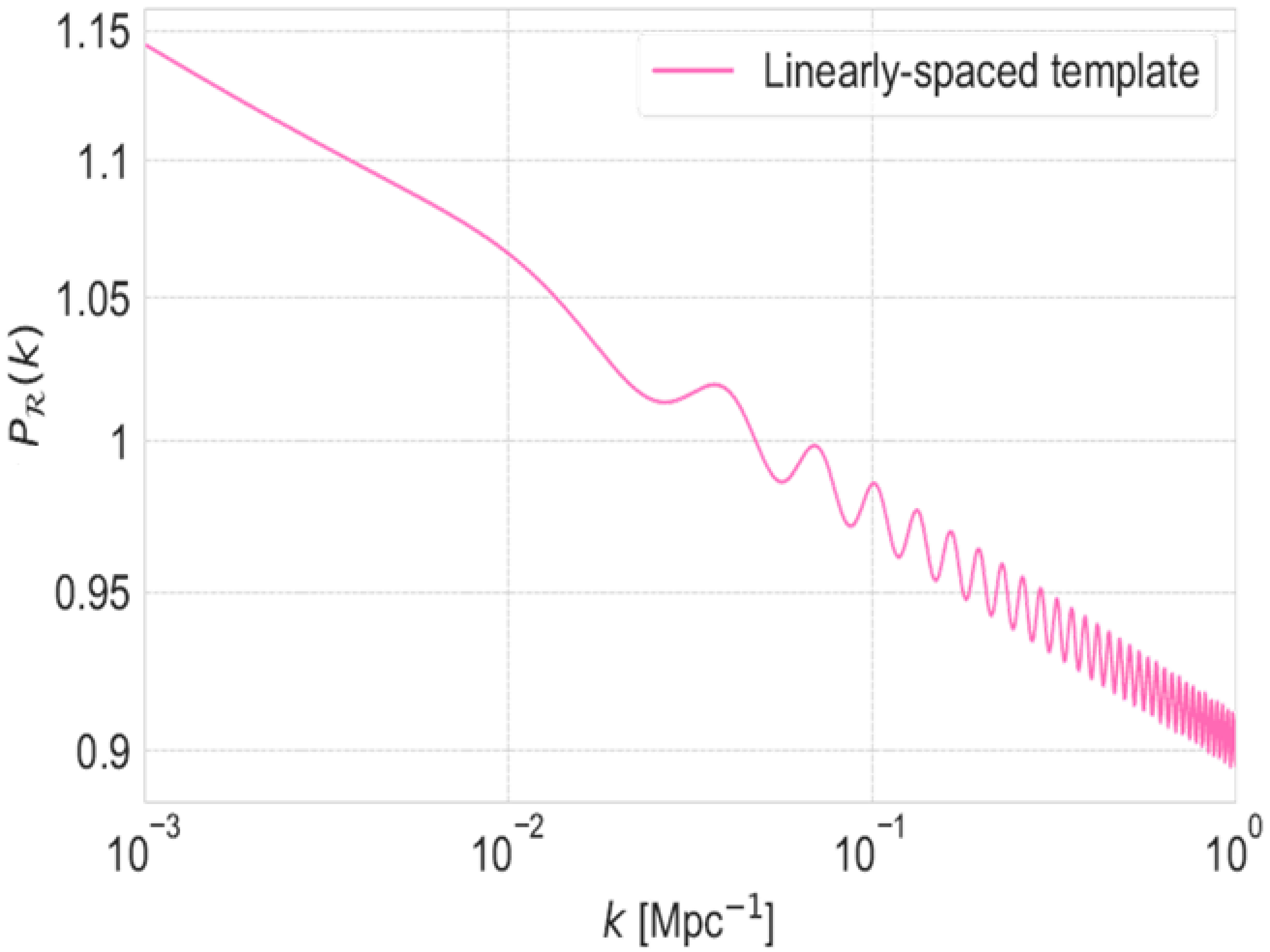
Primordial features parametrized as small deviations,

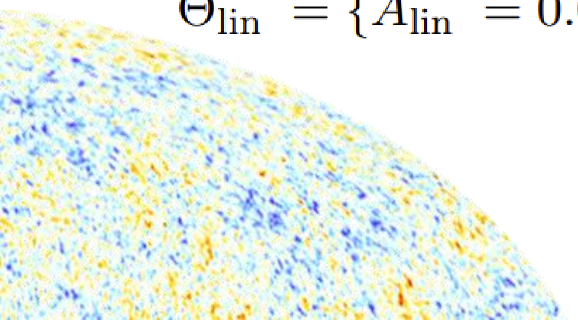
$$P_{\mathcal{R}}(k) = P_{\mathcal{R},0}(k)\left[1 + \frac{\Delta P_{\mathcal{R}}}{P_{\mathcal{R},0}}(k)\right],$$

Feature template with oscillations:

$$\rightarrow \quad \frac{\Delta P_{\mathcal{R}}}{P_{\mathcal{R},0}} = A_{\mathrm{lin}} \sin\left(\omega_{\mathrm{lin}} \frac{k}{k_\star} + \phi\right),$$

$$\Theta_{\mathrm{lin}} = \{A_{\mathrm{lin}} = 0.01, \omega_{\mathrm{lin}} = 10, \phi_{\mathrm{lin}} = 0\}.$$

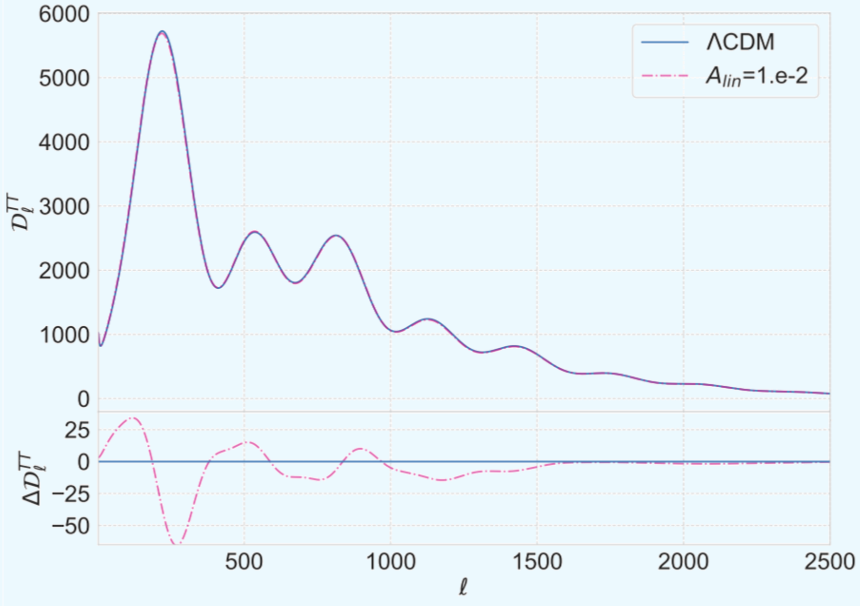


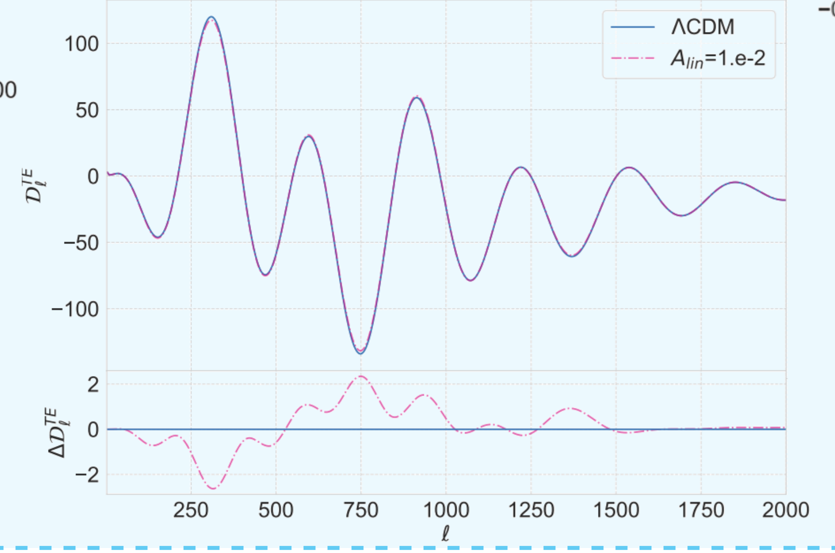


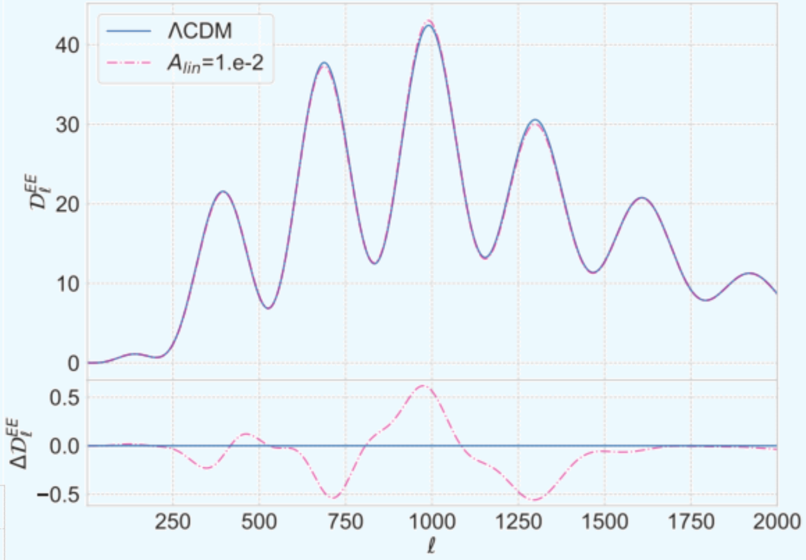
Angular Power Spectra

$$C_\ell^{XY} = \frac{2}{\pi}\int k^2 \mathrm{d}k \underbrace{P(k)}_{\text{Matter power spectrum}} \underbrace{\Delta_{X\ell}(k)\Delta_{Y\ell}(k)}_{\text{Transfer functions}}$$

Linearly-spaced template

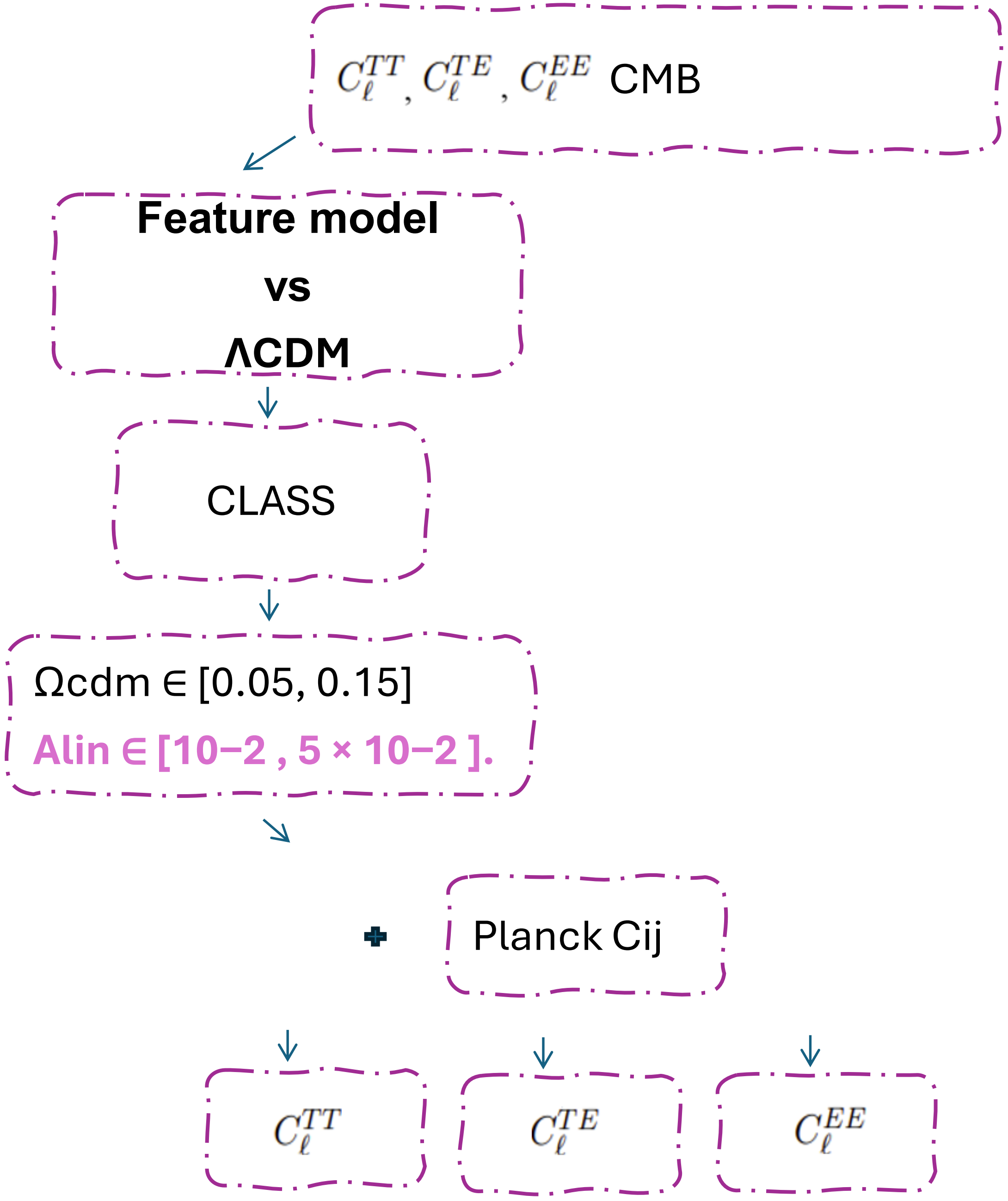


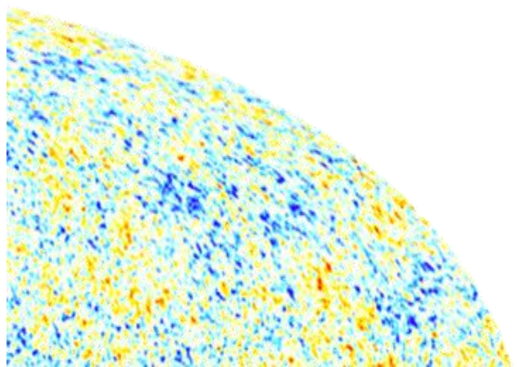




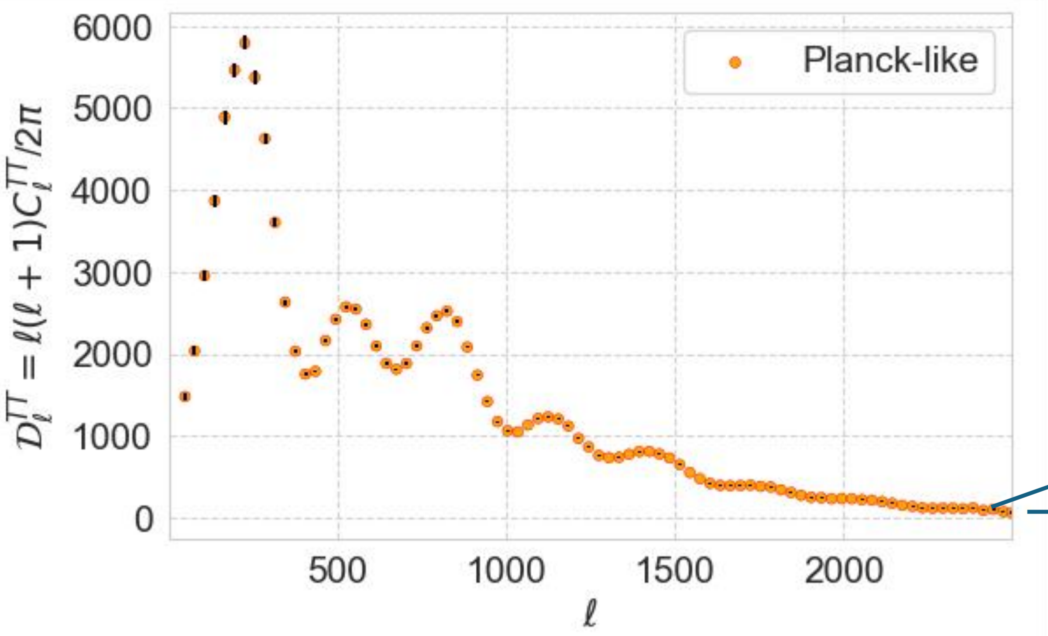
Dataset simulation strategy

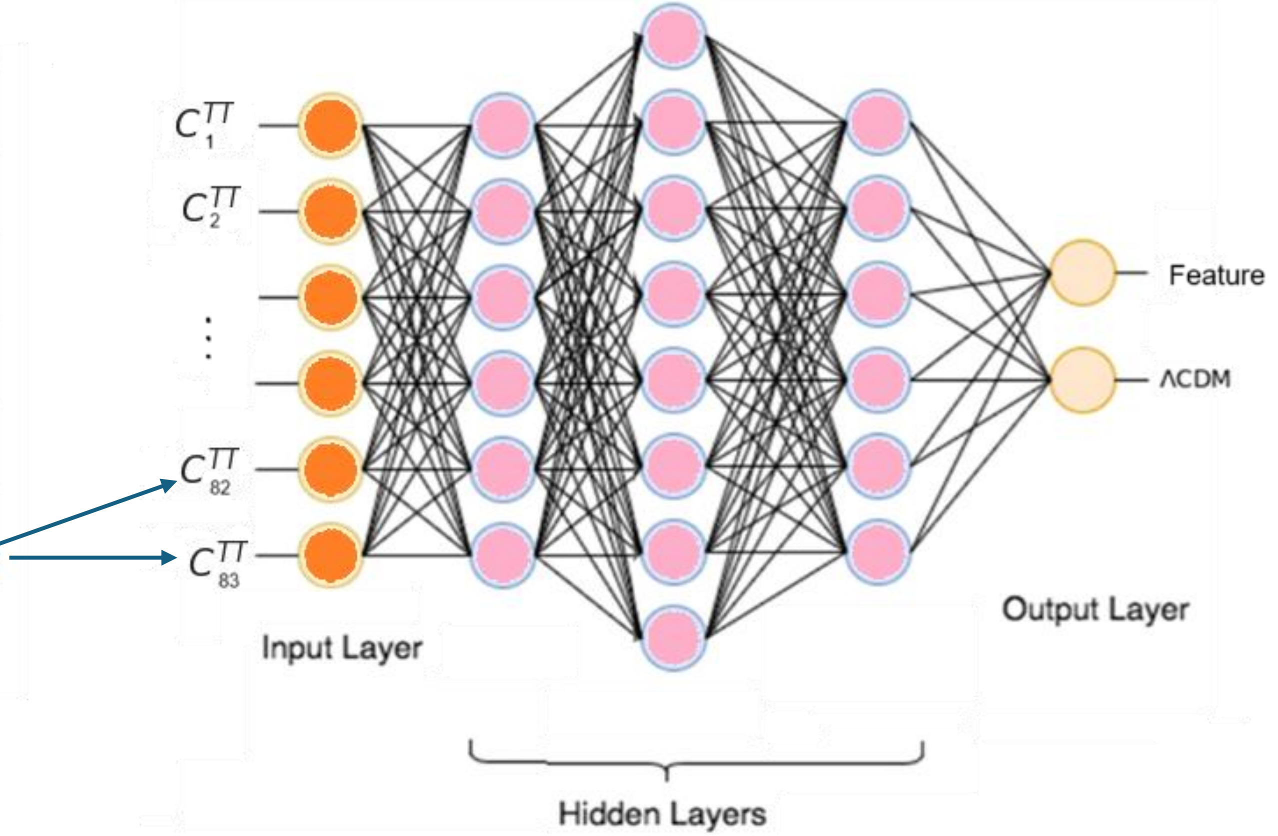
Assumed Planck Cosmological parameters.





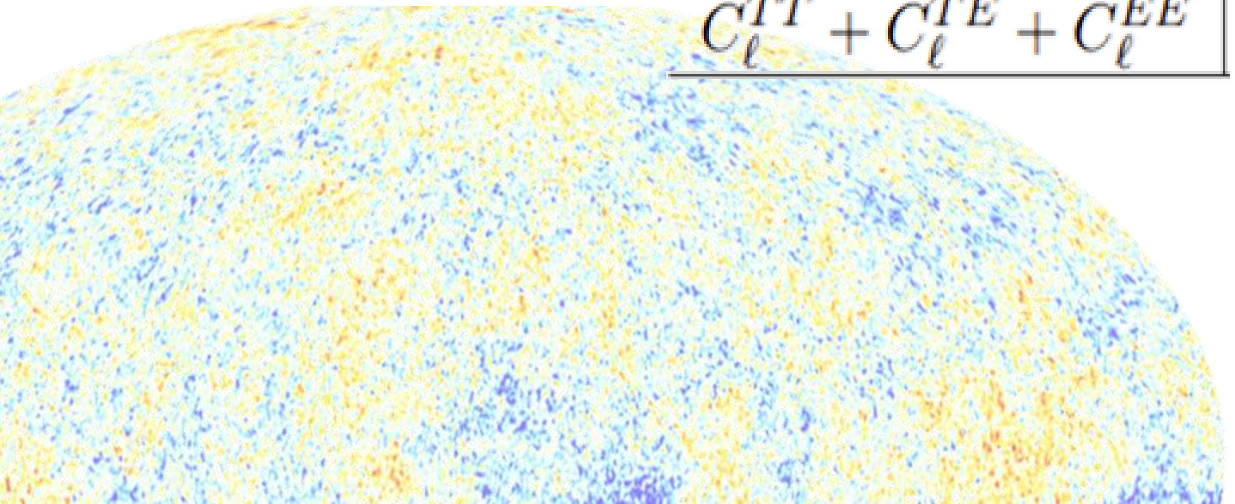
Machine Learning analysis





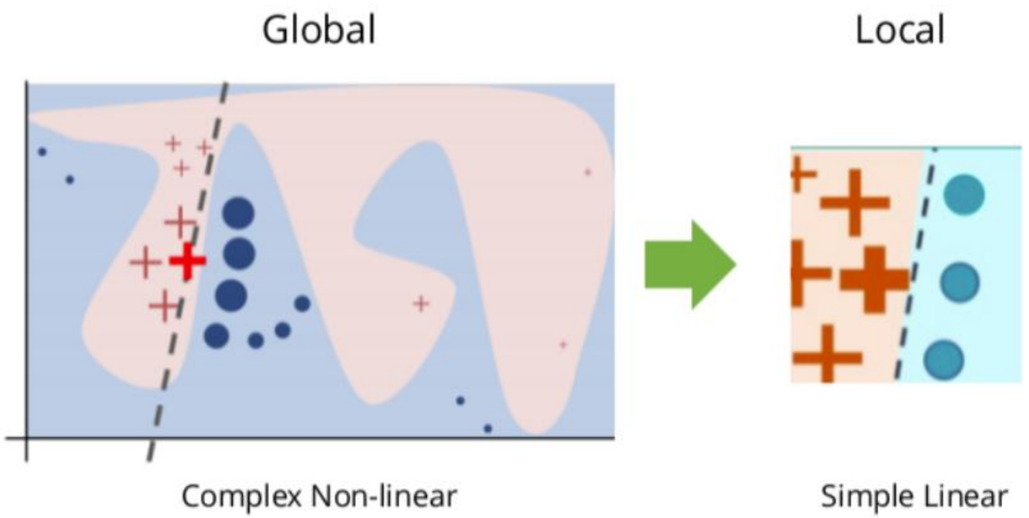
ML architecture and Performance

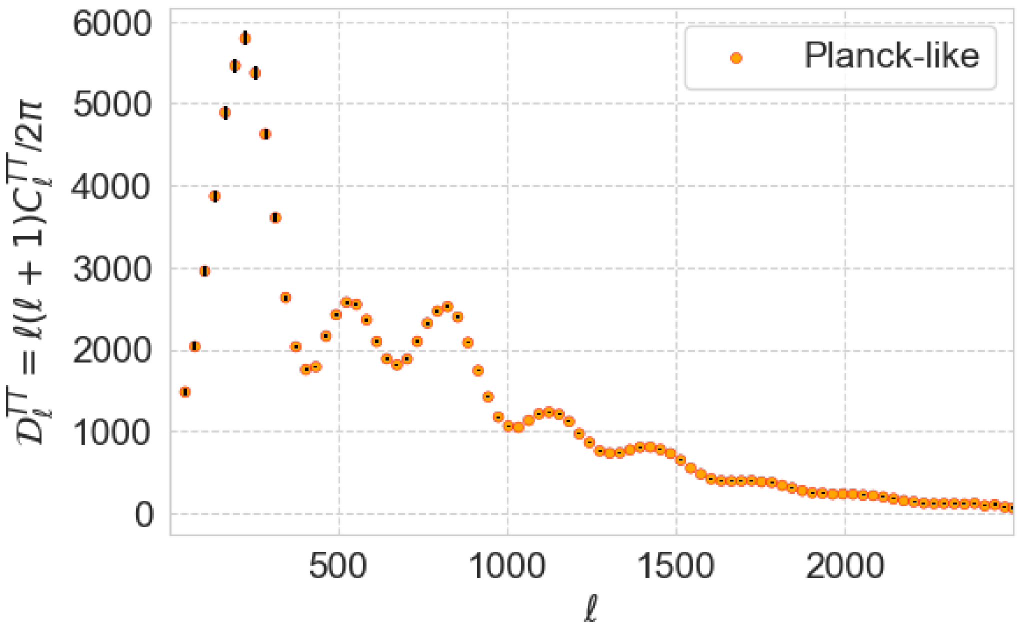
CMB components	linearly-spaced feature	
	Correct	Wrong
C_ℓ^{TT}	1	0
C_ℓ^{TE}	1	0
C_ℓ^{EE}	1	0
$C_\ell^{TT} + C_\ell^{TE} + C_\ell^{EE}$	1	0

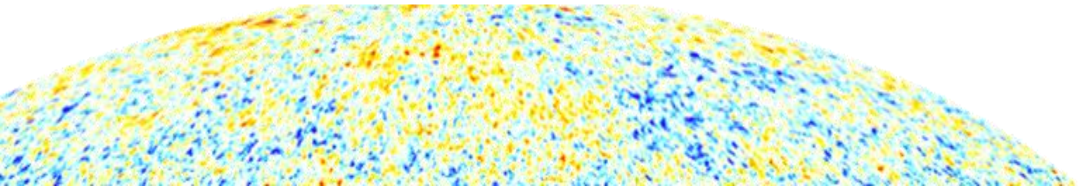


$$C_\ell^{XY} = \frac{2}{\pi}\int k^2 \mathrm{d}k \, \underbrace{P(k)} \, \underbrace{\Delta_{X\ell}(k)\Delta_{Y\ell}(k)}$$

ML interpretability: SHAP (Global)

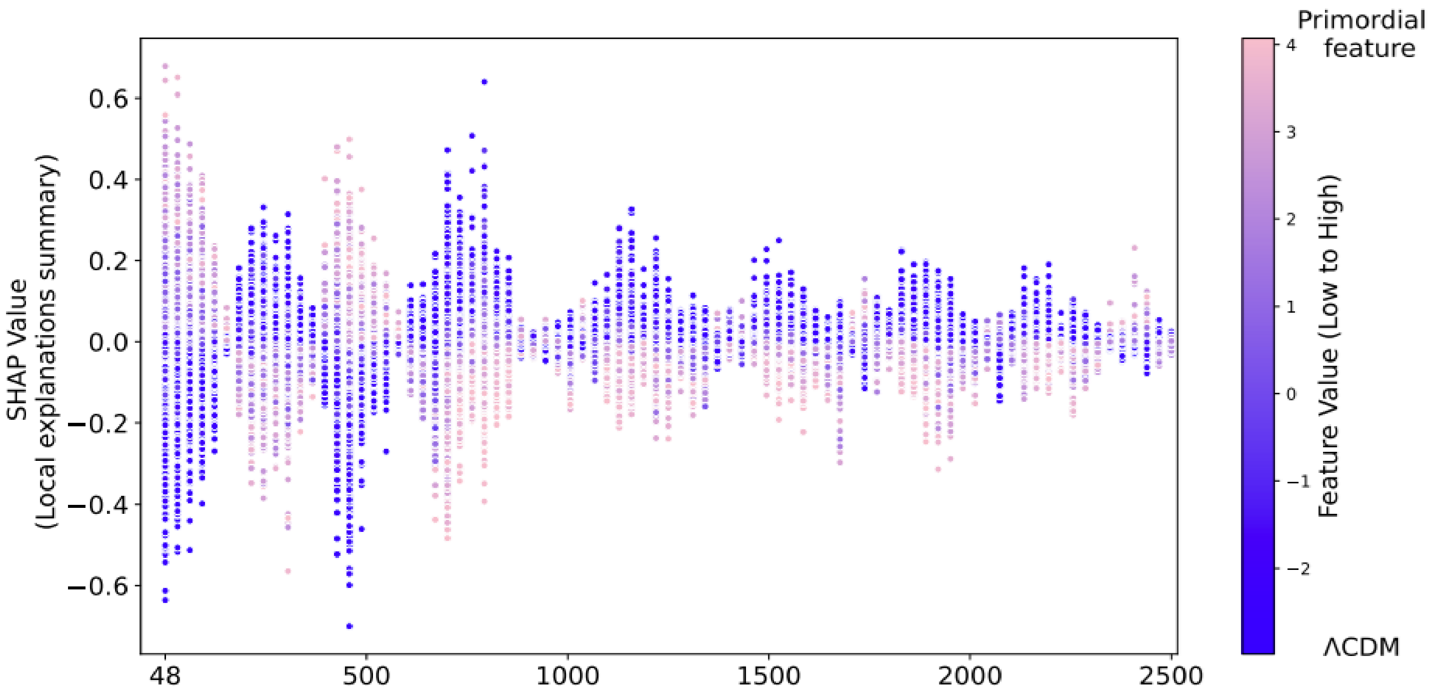


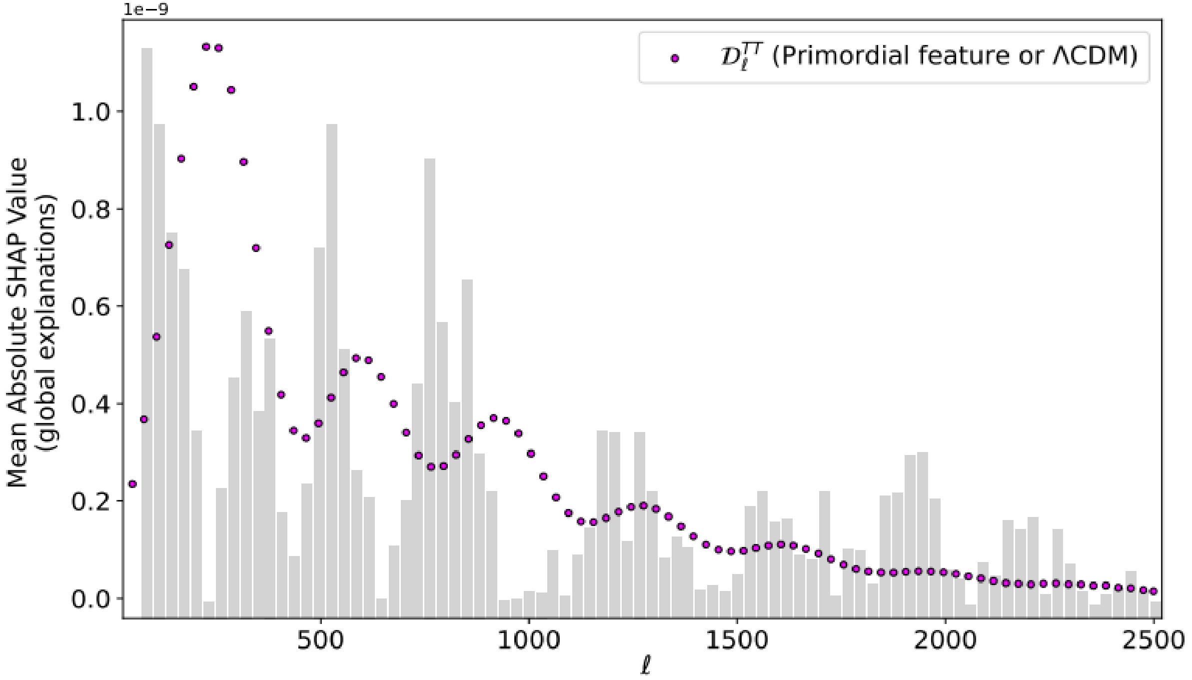




ML Interpretability: Feature template vs ΛCDM

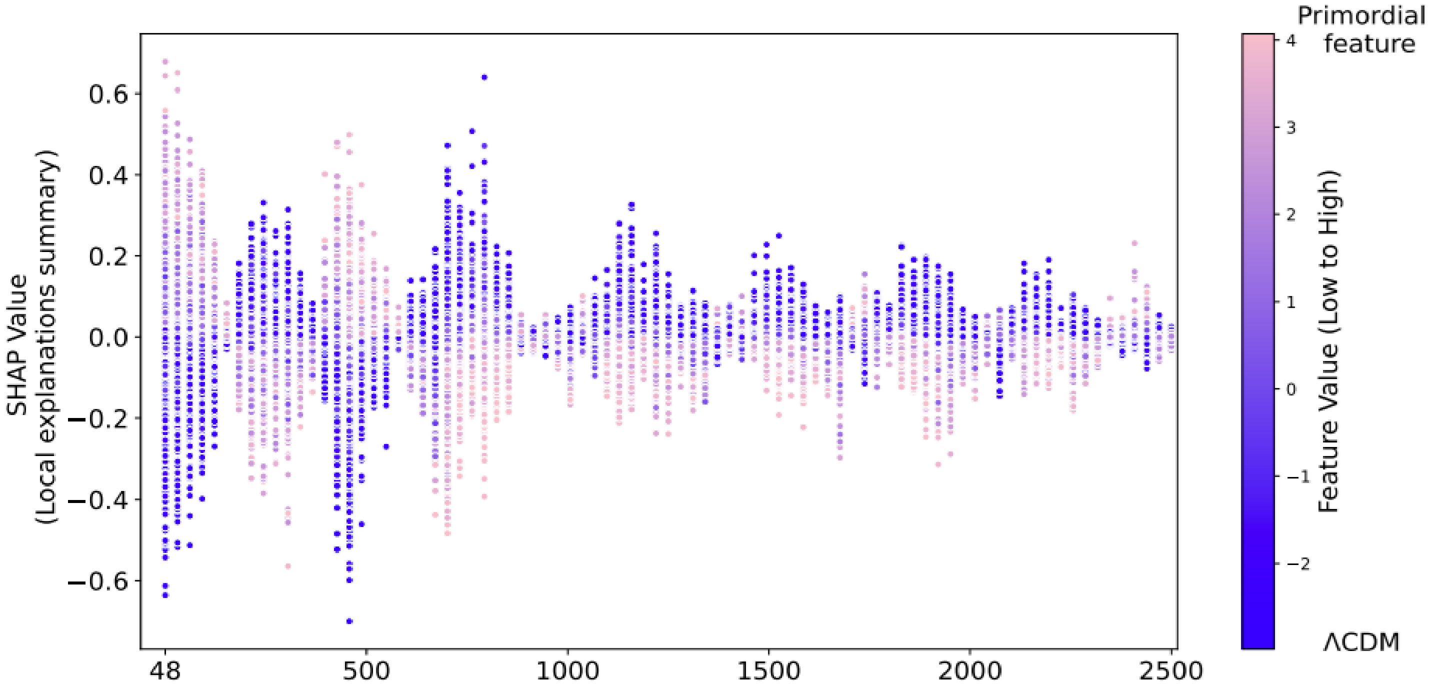
Temperature Angular Power Spectrum



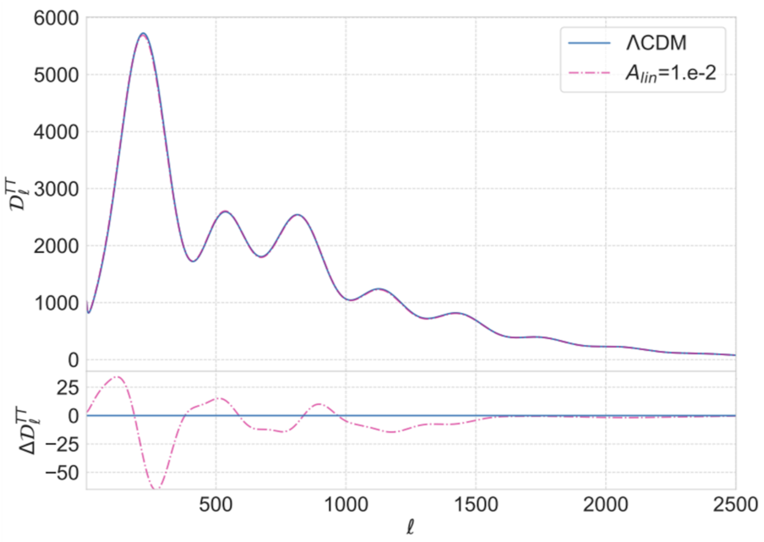


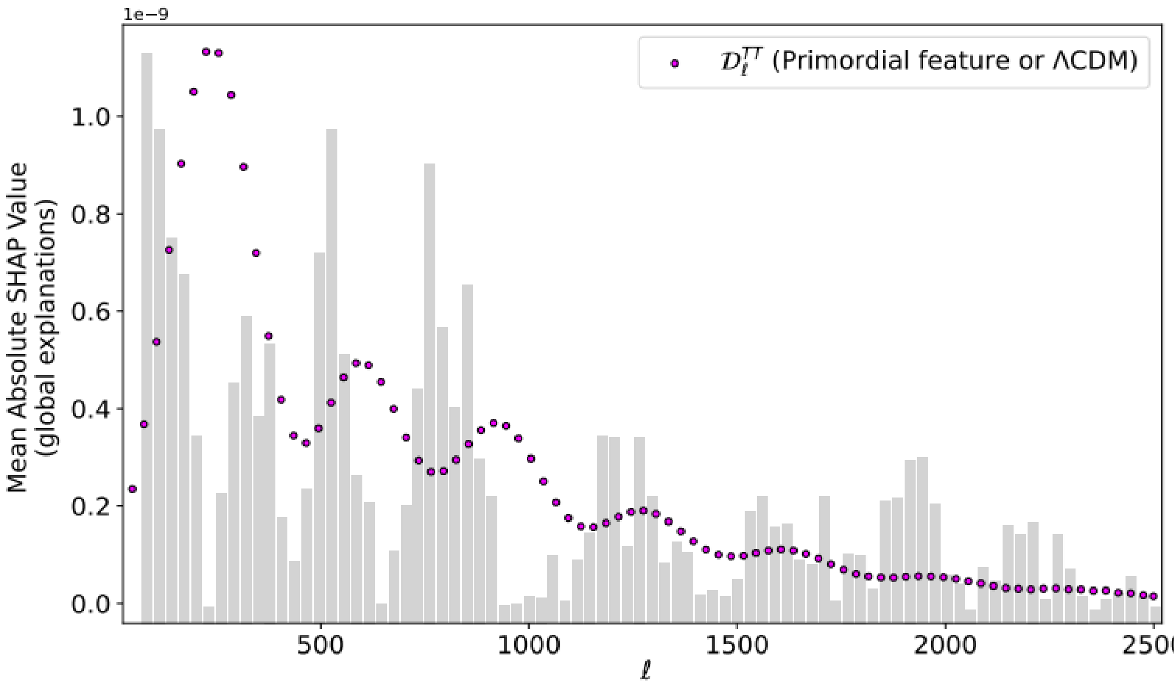
ML Interpretability: Feature template vs ΛCDM

Temperature Angular Power Spectrum



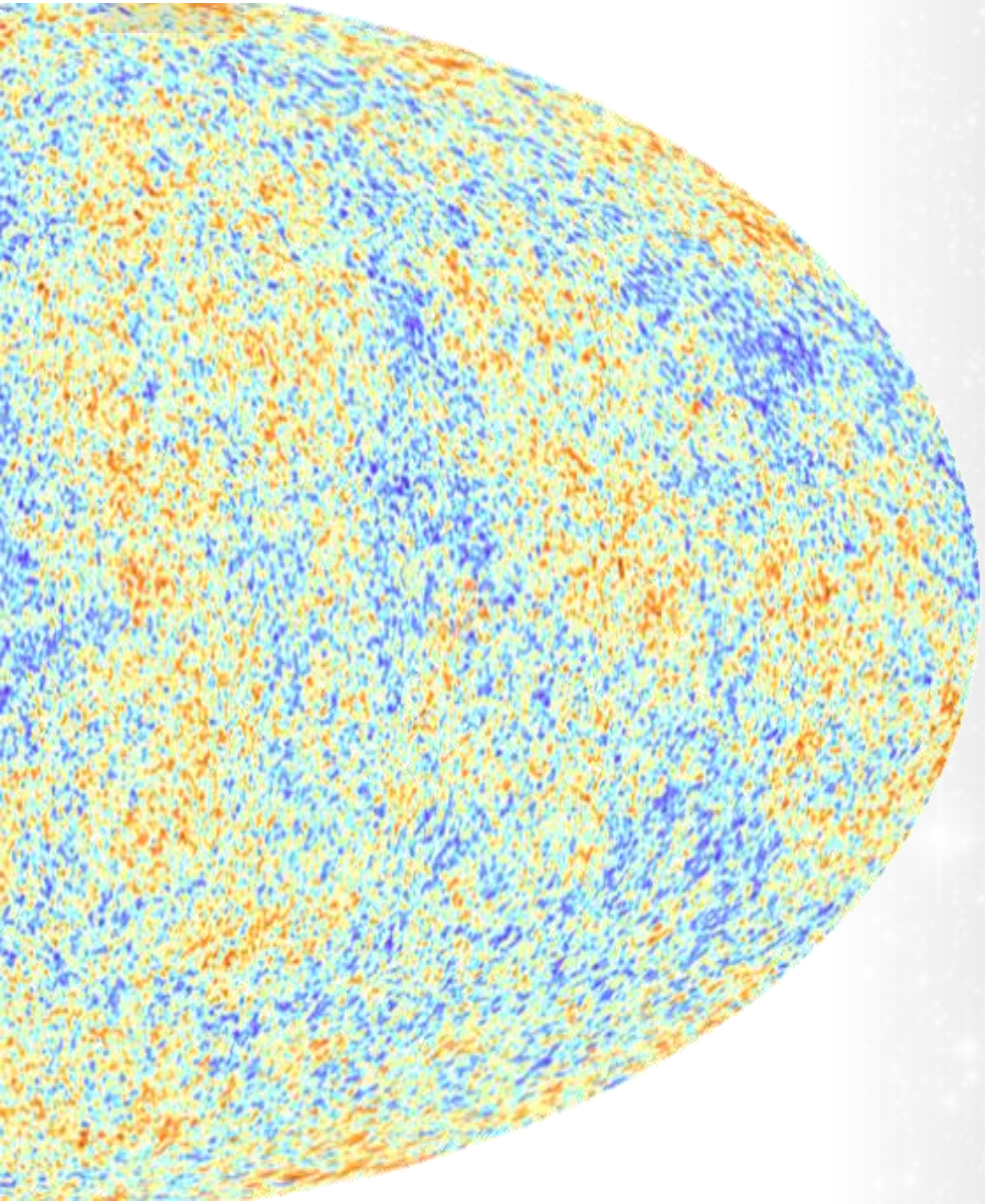
Theoretical Cls





Conclusions

- ML + Interpretability tools are an interesting starting point for verifying that the data is sensitive to some particular model, before doing the full MCMC sampling of the posterior (thousands of chains, computationally expensive).
- In the feature model, when looking at the output of SHAP, the NN is able to extract the introduced feature from the C_ℓ's.
- This methodology can be used to test any other beyond ΛCDM scenario (i.e. w_0 w_a)



Thank you