

A MACHINE-LEARNING APPROACH TO MEASURING THE ESCAPE OF IONIZING RADIATION FROM GALAXIES IN THE REIONIZATION EPOCH

Hannes Jensen, Erik Zackrisson, Kristiaan Pelckmans,
Christian Binggeli, Kristiina Ausmees, Ulrika Lundholm

arXiv:1603.09610

Abstract

Recent observations of galaxies at $z > 7$, along with the low value of the electron scattering optical depth measured by the Planck mission, make galaxies plausible as dominant sources of ionizing photons during the epoch of reionization. However, scenarios of galaxy-driven reionization hinge on the assumption that the average escape fraction of ionizing photons is significantly higher for galaxies in the reionization epoch than in the local Universe. The NIRSpec instrument on the James Webb Space Telescope (JWST) will enable spectroscopic observations of large samples of reionization-epoch galaxies. While the leakage of ionizing photons will not be directly measurable from these spectra, the leakage is predicted to have an indirect effect on the spectral slope and the strength of nebular emission lines in the rest-frame ultraviolet and optical. Here, we apply a machine learning technique known as lasso regression on mock JWST/NIRSpec observations of simulated $z = 7$ galaxies in order to obtain a model that can predict the escape fraction from JWST/NIRSpec data. Barring systematic biases in the simulated spectra, our method is able to retrieve the escape fraction with a mean absolute error of $\Delta f_{\text{esc}} \sim 0.12$ for spectra with $S/N \sim 5$ at a rest-frame wavelength of 1500Å for our fiducial simulation. This prediction accuracy represents a significant improvement over previous similar approaches.

- ▶ 宇宙再電離の解明には銀河からの電離光子のescape fraction f_{esc} が鍵となる
- ▶ JWST/NIRSpecにより、再電離期の銀河を数多く分光することが将来的に可能
- ▶ nebular emissionの強度から f_{esc} を推定することができる
- ▶ Lassoアルゴリズムを用いて、 $z=7$ のシミュレーション銀河のmock sampleから f_{esc} を求めた
- ▶ $S/N \sim 5$ の天体の f_{esc} を、系統誤差を除いて $\Delta f_{\text{esc}} \sim 0.12$ の精度で求めることができる

Introduction

- 近年の $z > 6$ の銀河の観測、PlanckによるCMBの観測などから、宇宙再電離の主要な電離源は星形成銀河からの電離光子 (Lyman continuum, LyC) だと考えられている
- このシナリオが成り立つためには、多くのLyCがISMに吸収されずにIGMに出て行くことが必要
- IGMへ放出されるLyC photonのproduction rateは以下の式で与えられる

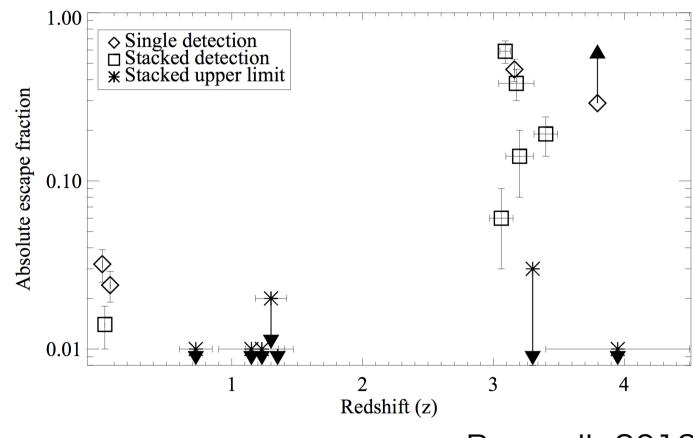
$$\dot{N}_{\text{ion}}(z) = f_{\text{esc}}(z)\xi_{\text{ion}}(z)\rho_{\text{UV}}(z)$$

f_{esc} : LyC escape fraction

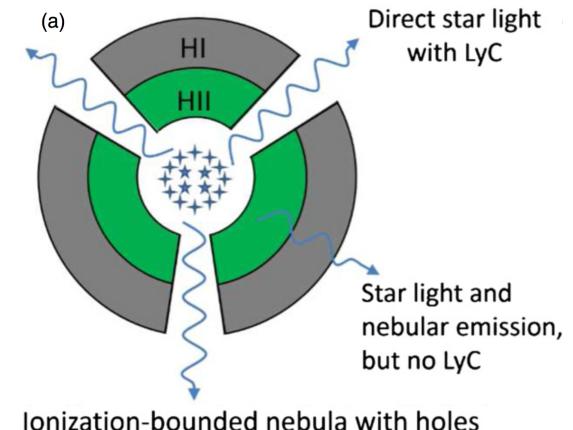
ξ_{ion} : LyC photonのproduction efficiency

ρ_{UV} : 銀河のUV luminosity density

- ρ_{UV} は銀河のUV luminosity functionから、 ξ_{ion} はNIRのSEDなどから制限 (Bouwens+2015a, Finkelstein+2015, Stark+2015, Bouwens+2015b)
- 一方で、 f_{esc} への制限は $z > 6$ では直接的にはついていない (c.f. Ono+2010)
- 星形成銀河だけで再電離を説明するには $f_{\text{esc}} \sim 0.1$ 以上である必要がある (Robertson+2015)
- 直接観測が可能なlow-zでは f_{esc} はもっと小さい
(右図)
→ f_{esc} の急激な進化が必要



- f_{esc} は星とnebular emissionのSEDへの寄与の割合を決める (右図)
- JWST/NIRSpec (Near Infrared Spectrograph) で $z \sim 6-9$ の明るい銀河を分光すれば、 f_{esc} への制限が可能に
- Absorption lineから推定する方法 (Jones+2013) にくらべ観測が容易



Zackrisson+2013

- 最小二乗法によりスペクトルをフィッティングする一般的な方法は今回の研究には不適切
← f_{esc} は rest UV-optical の一部のスペクトル bin のみと相関
- Lasso 回帰アルゴリズムを用いて、 f_{esc} を決める鍵となるスペクトル bin を automatic に選択する
- JWST/NIRSpec 観測のノイズを乗せた mock スペクトルを用いて f_{esc} を計算

Simulations

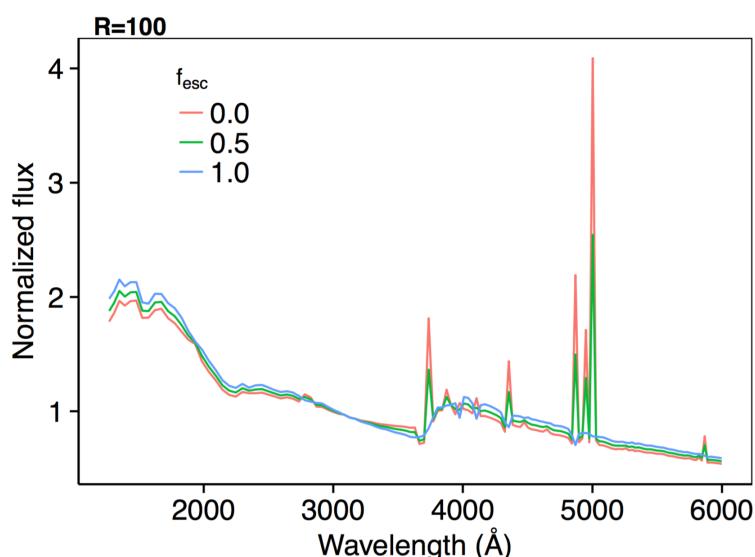
- 3種類のシミュレーションから $z=7$ の銀河を ~1400 個選択

	天体数	Mass limit
Finlator+2013	106	$> 10^7 M_{\odot}$
CROC (Gnedin+2014)	874	$> 10^7 M_{\odot}$
Shimizu+2014	406	$> 5 \times 10^8 M_{\odot}$

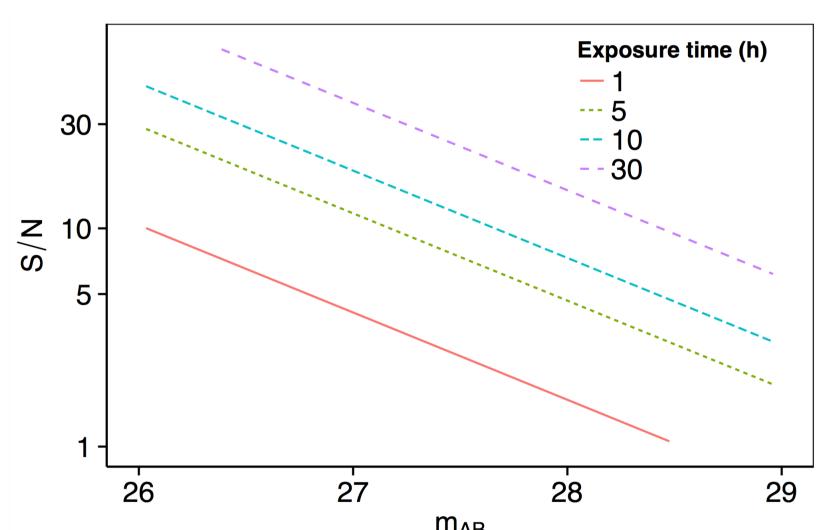
- Yggdrasil spectral evolutionary model (Zackrisson et al. 2011) を使用
 - Starburst99-Geneva, Padova, or BPASS2
- Dust attenuation
 - LMC, SMC, or Calzetti
- 様々な組み合わせで mock sample を作成

Simulated observations

- シミュレーションで得られたスペクトルの解像度を NIRSpec の観測に合わせ、ノイズを乗せる
- $R = 100$, Wavelength range 0.6 - 5.0 μm (750 - 6250 Å at $z=7$)



モデルスペクトルの例（ノイズなし）



R=100, 1500 ÅでのS/N

The Lasso Algorithm

Lasso (Least Absolute Shrinkage and Selection Operator) アルゴリズム

- ▶ 最小二乗法と似ているが、パラメータが多い場合により有効
- ▶ input $\mathbf{x} \in \mathbb{R}^N$ (e.g. 各wavelength binのフラックス) と output $y \in \mathbb{R}$ (e.g. f_{esc}) の m 組の training examples $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)$ から、最も y をよく再現するパラメータの組 β を求める

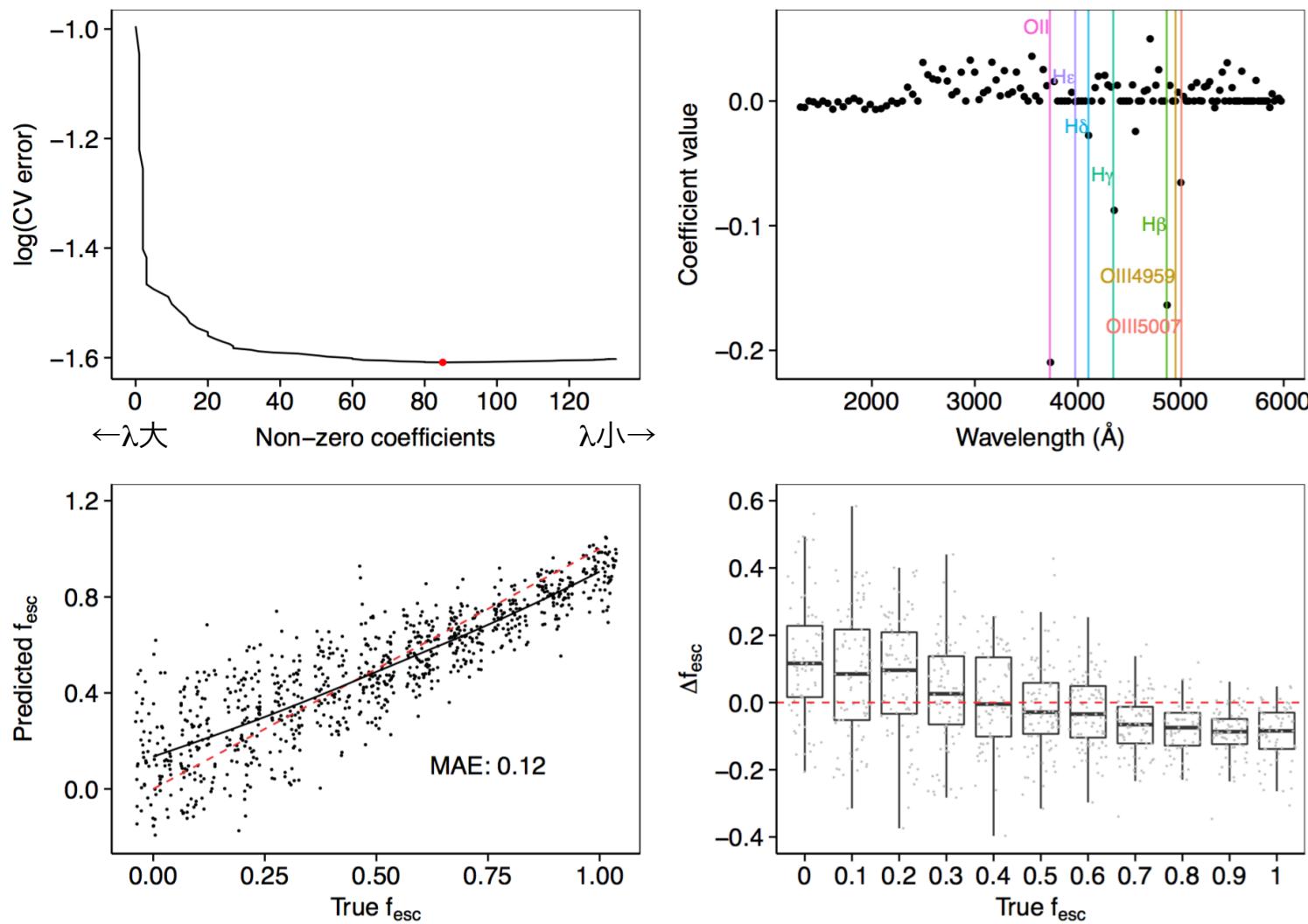
$$\hat{y} = \beta_0 + \sum_{i=1}^N \beta_i x_i$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^m [\hat{y}(\mathbf{x}_i) - y_i]^2 + \lambda \sum_{j=1}^N |\beta_j| \right\}$$

- $\lambda = 0$ のとき：最小二乗法に帰着
- λ を大きくすると、less-importantな係数は0に落ちる
- ▶ inputにoutputと相関の少ない変数が多い場合にoverfittingを防ぐことができる
- ▶ フィッティングの質を上げるには λ を適切な値にすることが必要
 - \mathbf{x} のcomponentの多くが y と相関が強い場合： λ 小
 - \mathbf{x} にノイズが多く含まれている場合： λ 大
- ▶ sampleをtraining setとtest setに分け、 λ のbest valueを求める

Results

- ▶ ShGeLMC (Shimizu, Geneva, LMC), S/N = 5 の場合
- ▶ $f_{\text{esc}} = 0, 0.1, 0.2 \dots 1.0$, 計4466天体



左上図 : non-zeroの係数の数とcross-validationエラー

- ▶ non-zeroの係数は85個がbest
- ▶ 少ない個数でエラーは急激に下がる
→ f_{esc} の情報を含むスペクトルbinは少ない

右上図 : 係数 β のbest-fit value

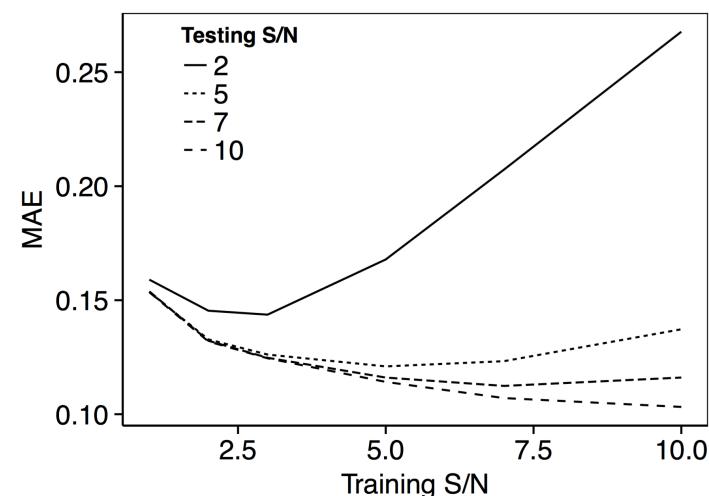
- ▶ 絶対値が大きい β はemission lineの波長にある
- ▶ emission lineの β は全て負 → emissionの強度と f_{esc} は負の相関

下2図 : training setで β を求めた後test setで求めた f_{esc}

- ▶ Δf_{esc} は $f_{\text{esc}} < 0.2$ で大
- ▶ エラーの平均値は 0.12
- ▶ 75%の確率で $f_{\text{esc}} = 0.5$ と $f_{\text{esc}} = 0$ を区別可能

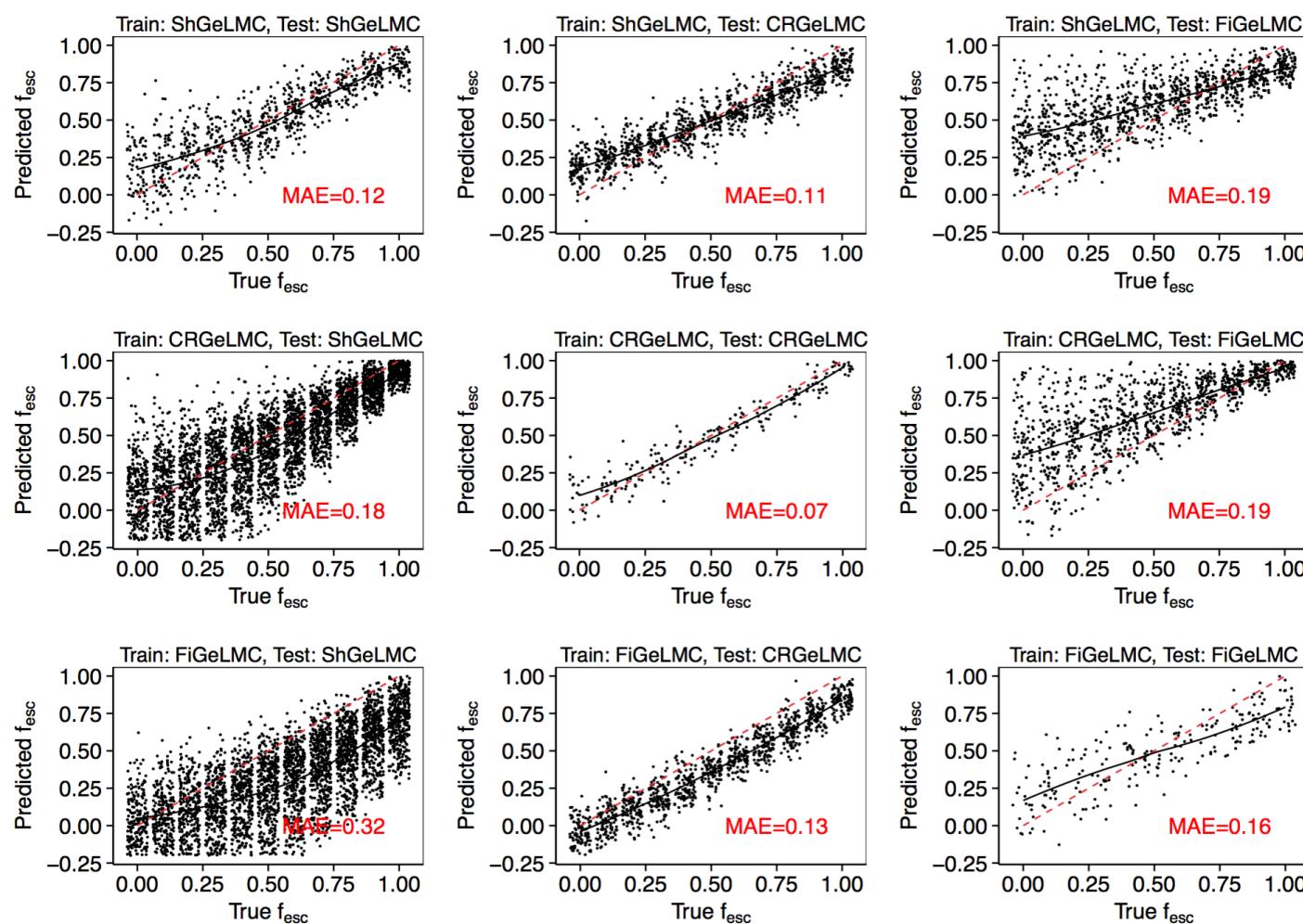
Noise Level and Spectral Resolution

- ノイズレベルを変化
- training setとtest setのノイズレベルが一致したときにエラー最小
 - test setよりlow-noiseでtraining
→ ノイズに埋もれたweak emission lineをフィッティングしてしまう
 - test setよりhigh-noiseでtraining
→ 一部のstrong emission lineしか使えない



Dependence on Simulation Parameters

- モデルの仮定によって結果がどれくらい変わるかを調べる

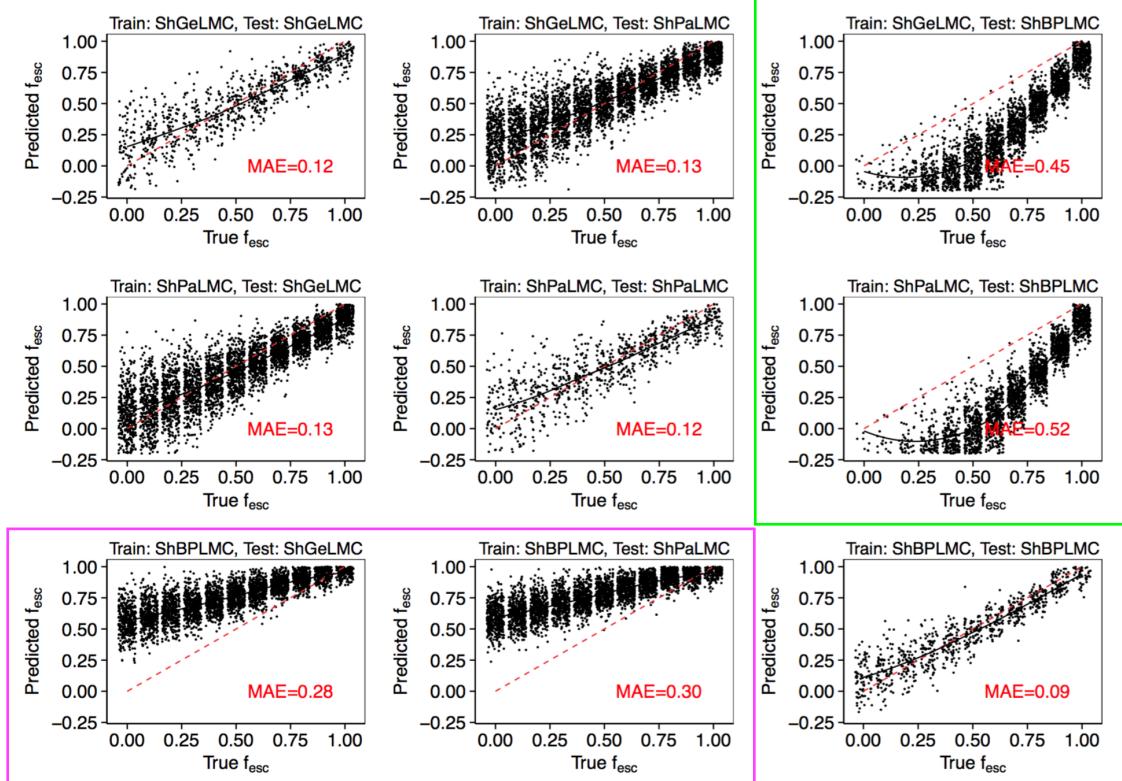


Name	Simulation	Stellar evolution	Dust attenuation
ShGeLMC	Shimizu	Geneva	LMC
FiGeLMC	Finlator	Geneva	LMC
CRGeLMC	CROC	Geneva	LMC
ShPaLMC	Shimizu	Padova	LMC
ShBPLMC	Shimizu	BPASS	LMC
ShGeCal	Shimizu	Geneva	Calzetti
ShGeSMC	Shimizu	Geneva	SMC

- Shimizu or CROC simulationをFinlator simulationに適用した場合にscatterが大きい（右列）
 - Finlator galaxiesのstar formation historyのdiversityが大きいこと・sample数が少ないことが原因

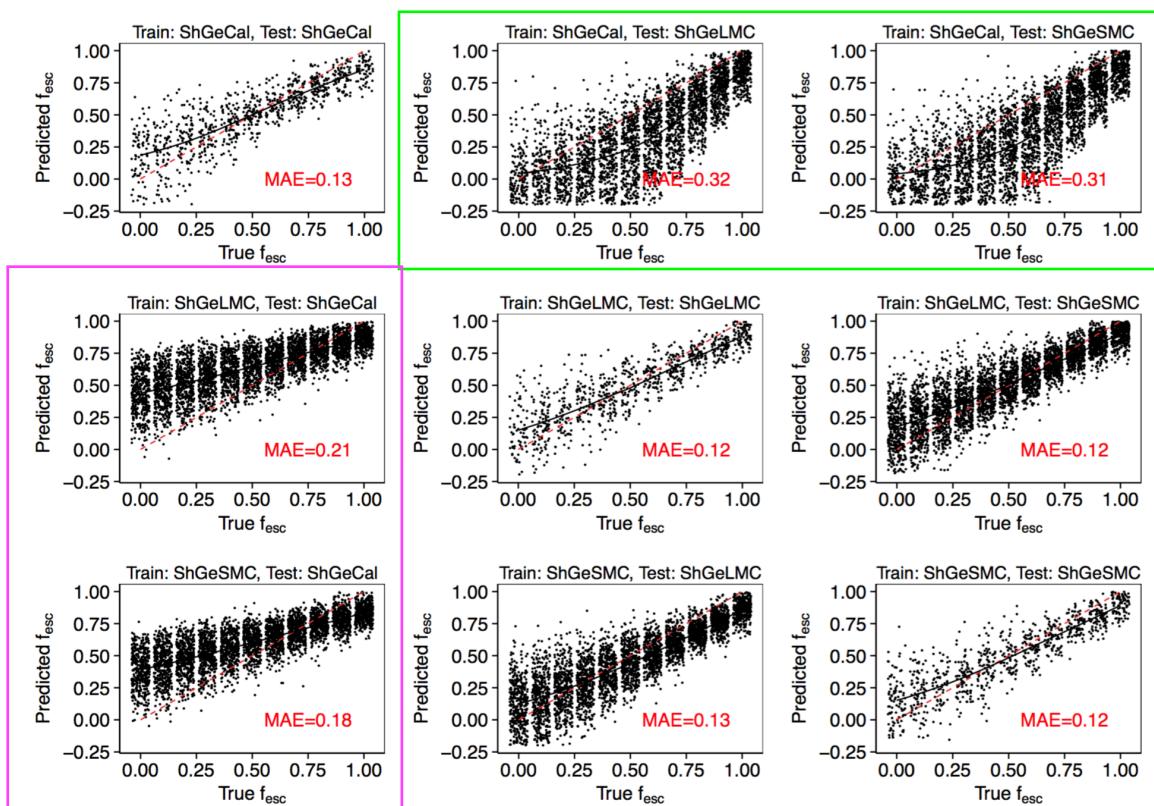
Stellar evolution modelを変えた場合

- BPASS2 → Geneva, Padova:
 f_{esc} をoverestimate
- Geneva, Padova → BPASS2:
 f_{esc} をunderestimate



Dust attenuationを変えた場合

- Calzetti → LMC, SMC:
 f_{esc} をunderestimate
- LMC, SMC → Calzetti:
 f_{esc} をoverestimate



Summary and Discussion

- ▶ Lassoアルゴリズムを用いて実際の観測と同じノイズレベルのスペクトルでtrainingすることで、 f_{esc} を求めることができる
- ▶ $m_{\text{AB}} \sim 29$ 等より明るい天体で $S/N > 5$ (exposure time 10時間)
 - NIRSpec pointing (3.4×3.4 arcmin) でおよそ100天体
(from Bouwens+2015b)
 - それより暗い天体も、population全体の平均的な f_{esc} を求めることが可能
- ▶ これまでのindirectな f_{esc} の決定方法と同様model dependent
 - star formation historyにも依存
 - extreme quenching (star formationの直後にno star formation period) の銀河の場合、青い β を持つがemission lineなし
→ 高い f_{esc} を出してしまう可能性
- ▶ high-z銀河のdust attenuationやstellar evolutionなどをより正しくmodelingすることが必要
→ 今後の観測による制限が重要
- ▶ LyCが直接観測されているlow-zで今回の手法を評価することも必要