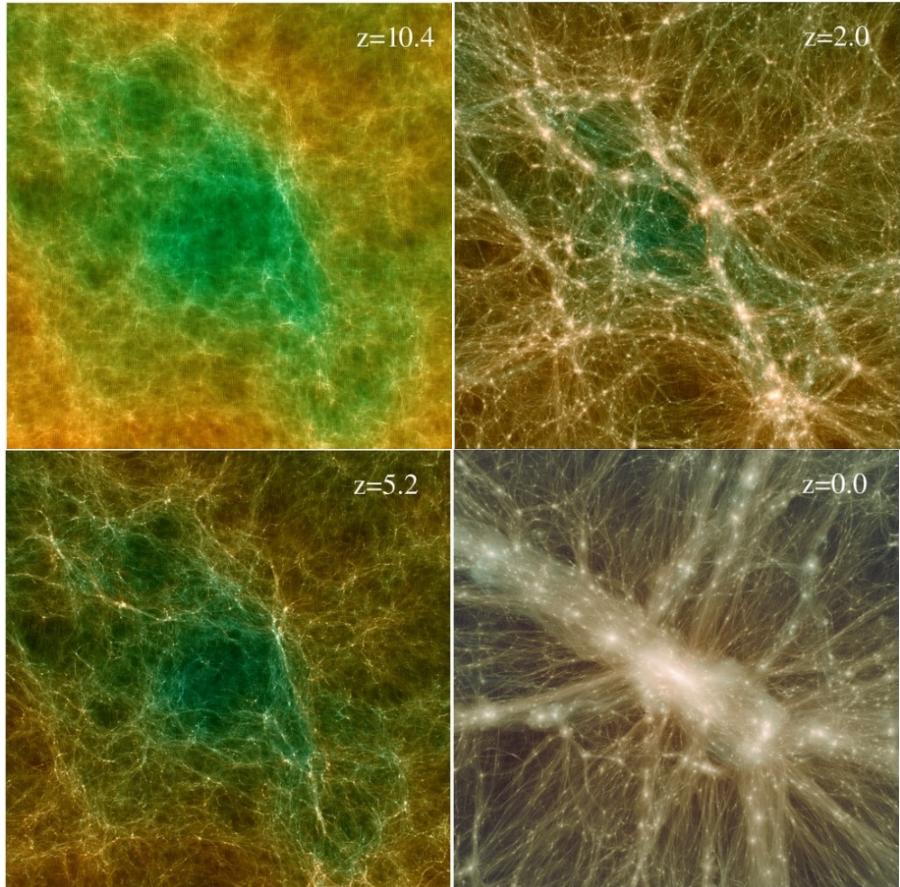


機械学習を用いた 宇宙論的N体シミュレーションの補間

本間慎一朗, 石山智明(千葉大学)

研究背景

- ・宇宙の構造形成の歴史を解明する
- 宇宙論的 N 体シミュレーションで、ダークマター分布の初期密度揺らぎから現在までの進化を追う

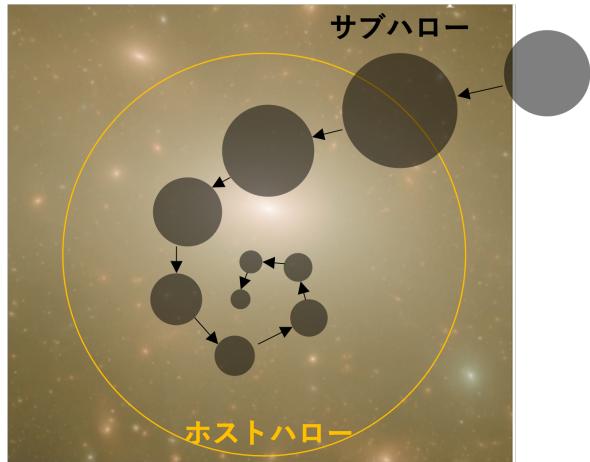


Uchuu の最も大きなハローの構造 (Ishiyama et al. 2021)

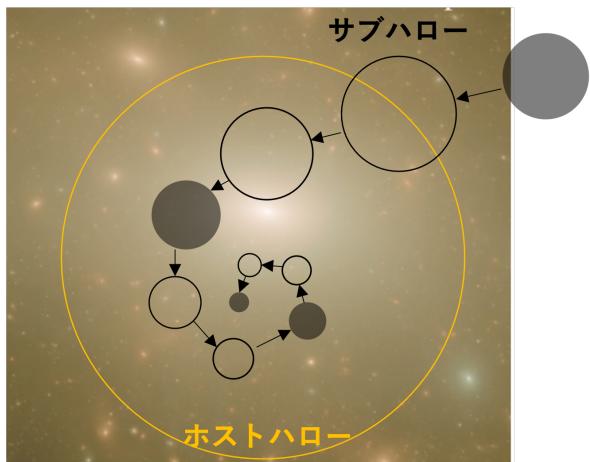
- ・シミュレーションデータを長期間維持することは困難
 - 最先端の宇宙論的 N 体シミュレーションである Uchuu (Ishiyama et al. 2021) の粒子数は 12800^3 個
 - 1 スナップショットあたりのデータサイズは約 60 TB
 - スナップショット数は全体で 50 個
 - 全体のデータサイズは約 3 PB
 - スナップショットの数を減らして時間分解能を落とす
 - スナップショットに含まれる粒子数を減らして質量分解能を落とす
- ↓
- ・機械学習で時間方向の高精度な補間を試みる

本研究の内容

- ・サブハローの軌道情報を補間

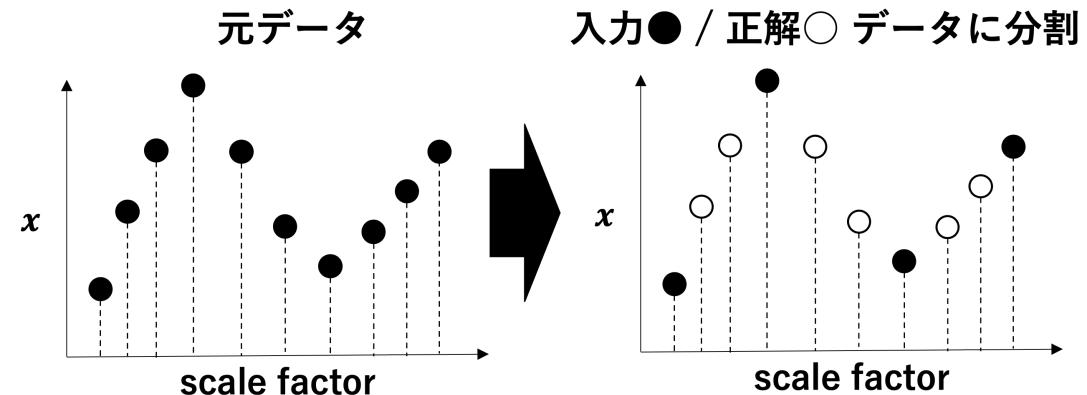


サブハローの運動の様子

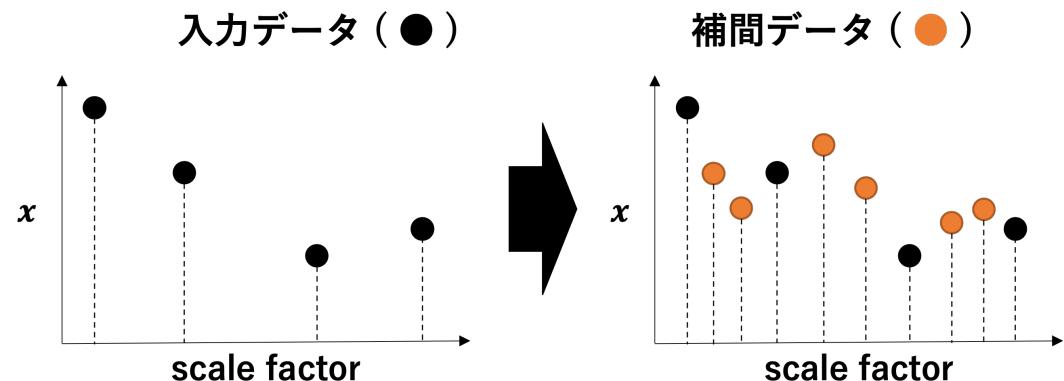


入力● / 正解○ データに分割

- ・低質量分解能シミュレーションから訓練データを作成



- ・高質量分解能シミュレーションを時刻間で補間



シミュレーションデータの詳細

- 既存の高質量分解能シミュレーションの時刻間の情報を、機械学習を用いて補間できるか検証

- Phi - 4096 (Ishiyama et al. 2021)
- シミュレーションのホストハローに付随するサブハローの情報を使用
- 高質量分解能シミュレーションのサブハローは $M_{\text{vir}}(z=0) \geq 1 \times 10^7 (h^{-1} M_{\odot})$
- 低質量分解能シミュレーションのサブハローは $M_{\text{vir}}(z=0) \geq 1 \times 10^8 (h^{-1} M_{\odot})$
- サブハローがホストハローに突入した後 ($z < z_{\text{acc}}$) の merger tree 上の情報を使用
($|x|/R_{\text{host}} > 0.03$)

Phi - 4096 (高質量分解能版)

N	$L(h^{-1}\text{Mpc})$	$\varepsilon(h^{-1}\text{kpc})$	$m_p(h^{-1}M_{\odot})$
4096^3	16.0	0.06	5.13×10^3

Phi - 4096 (低質量分解能版, 初期条件は高質量分解能版と同じ)

N	$L(h^{-1}\text{Mpc})$	$\varepsilon(h^{-1}\text{kpc})$	$m_p(h^{-1}M_{\odot})$
1024^3	16.0	0.24	3.28×10^5

シミュレーションの初期赤方偏移 : 127

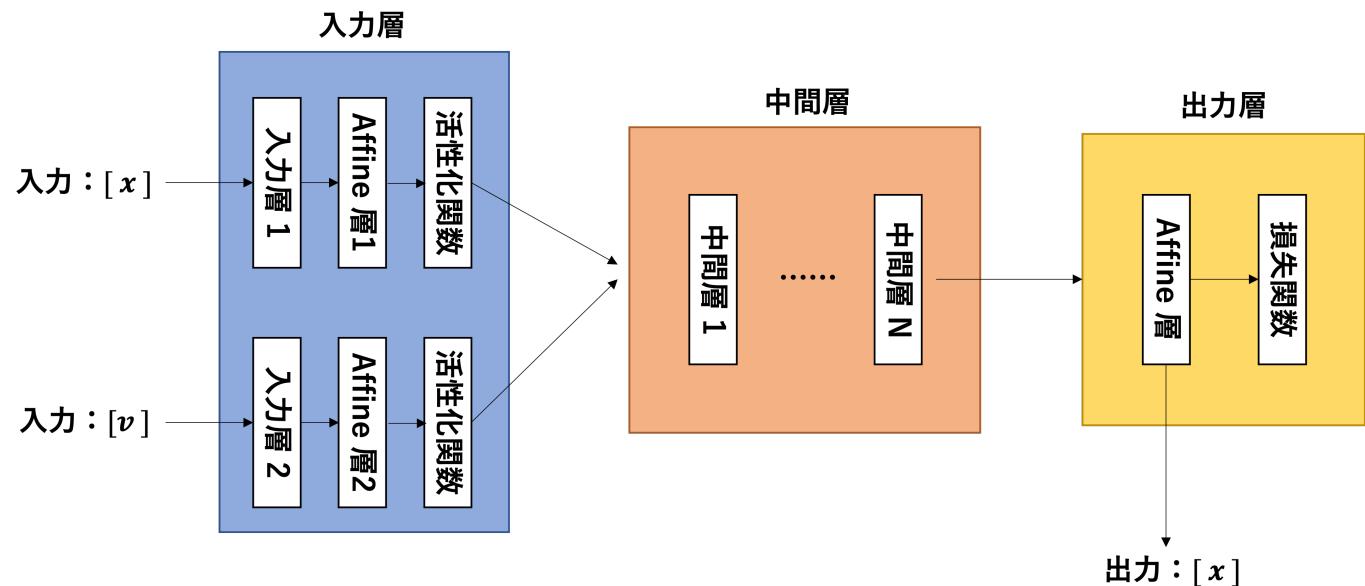
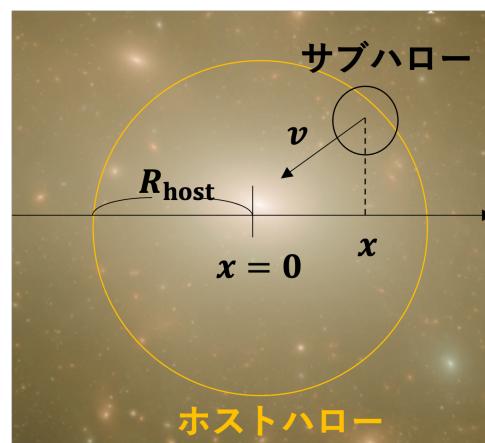
ニューラルネットワークの構成

ニューラルネットワークの詳細

入力 : 出力	中間層	ニューロン数	活性化関数	損失関数	最適化関数	正規化	荷重減衰
2:3	10層	100個	TanhExp	相対誤差	Adam	Z-SCORE	$\lambda = 1.0 \times 10^{-3}$

- ・ネットワークが学習する訓練データ：低質量分解能シミュレーションのデータ
- ・補間精度を評価する検証データ：高質量分解能シミュレーションのデータ
- ・ネットワークと線形補間, 3次スプラインを比較して, ネットワークの補間精度を検証

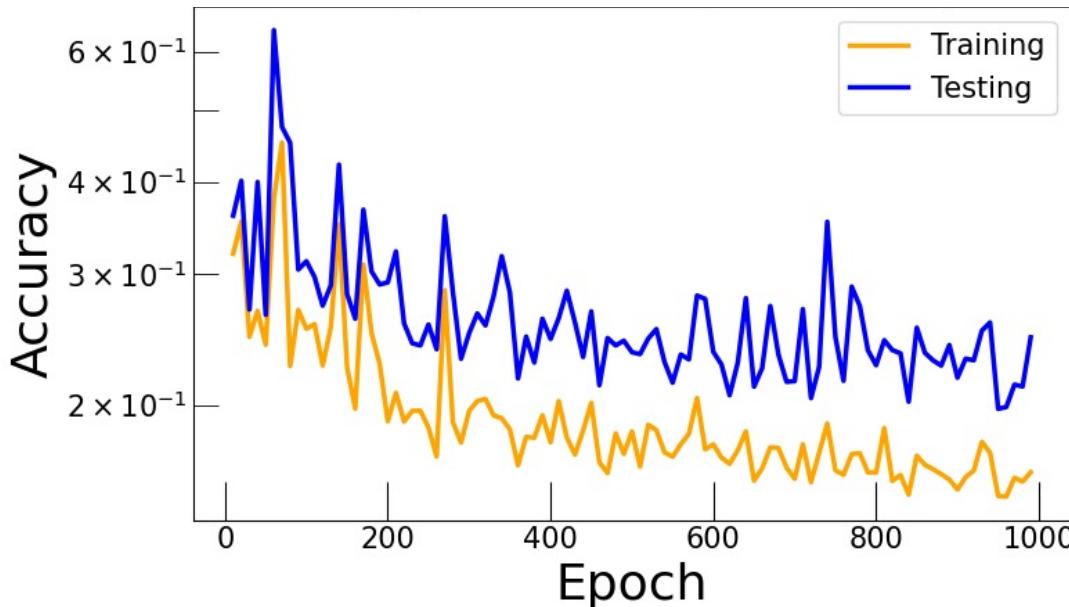
- 入力パラメータ : x, v
- 出力パラメータ : x
- $x = (x_{\text{sub}} - x_{\text{host}})/R_{\text{host}}$
- $v = (v_{\text{xsub}} - v_{\text{xhost}})$



ネットワークの補間精度の評価

- 学習回数と補間精度の推移

- ネットワークの出力： P , 正解データ： T とすると, Accuracy = $\text{mean} \left| \frac{P-T}{T} \right|$

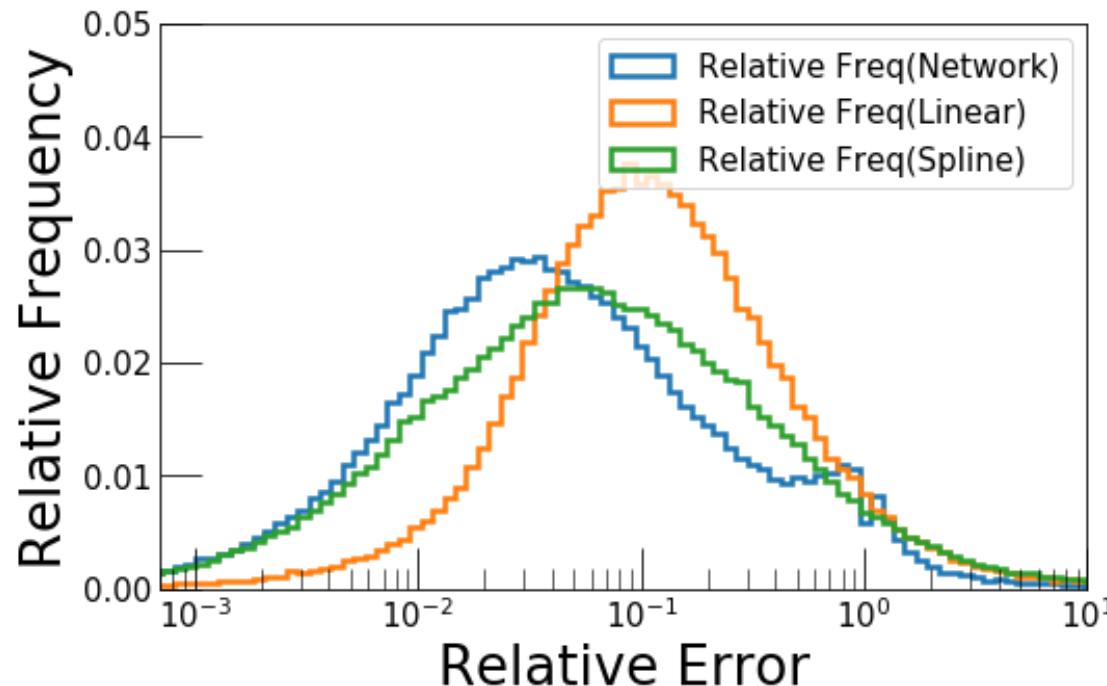


- Training : 訓練（低質量分解能）データに対する補間精度
- Testing : 検証（高質量分解能）データに対する補間精度
- Epochはネットワークの学習回数

- Training と Testing の値が離れるほど, ネットワークの汎用性は低い（過学習）
→過学習の傾向が少し見える

ネットワークと線形補間, 3次スプラインの補間精度を比較

- 検証（高質量分解能）データを各補間手法で補間したときの, 正解データに対する補間データの相対誤差のヒストグラム



➤ 補間データ : $x_{\text{predicted}}$, 正解データ : T

$$\text{Relative Error} = \left| \frac{x_{\text{predicted}} - T}{T} \right|$$

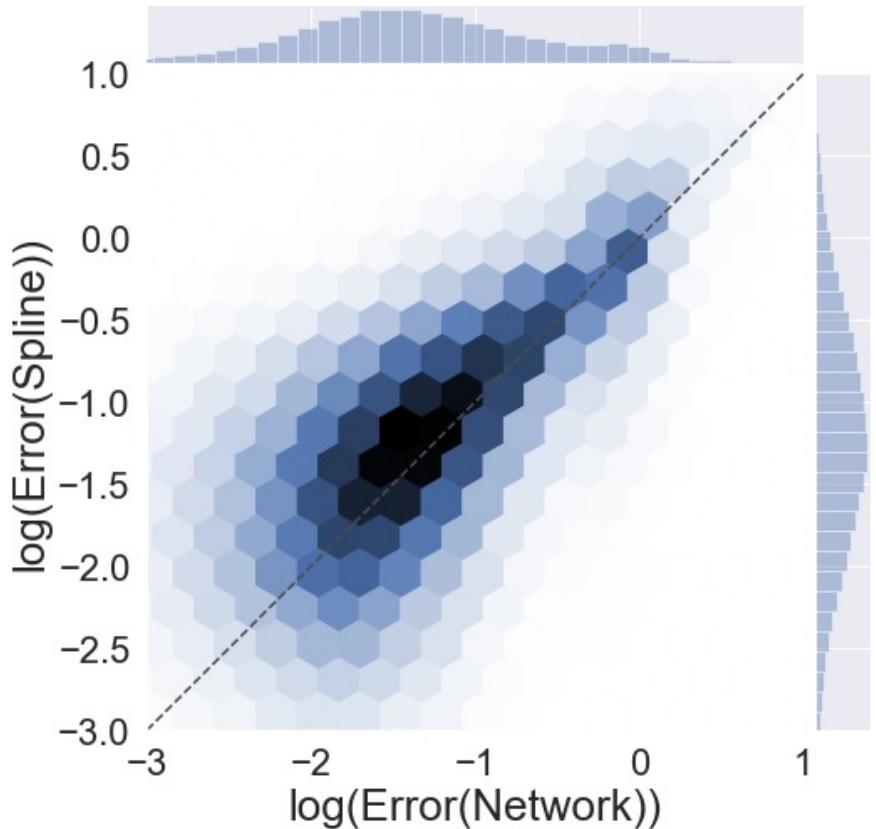
➤ ネットワークの最頻値 : 0.034

➤ 線形補間の最頻値 : 0.088

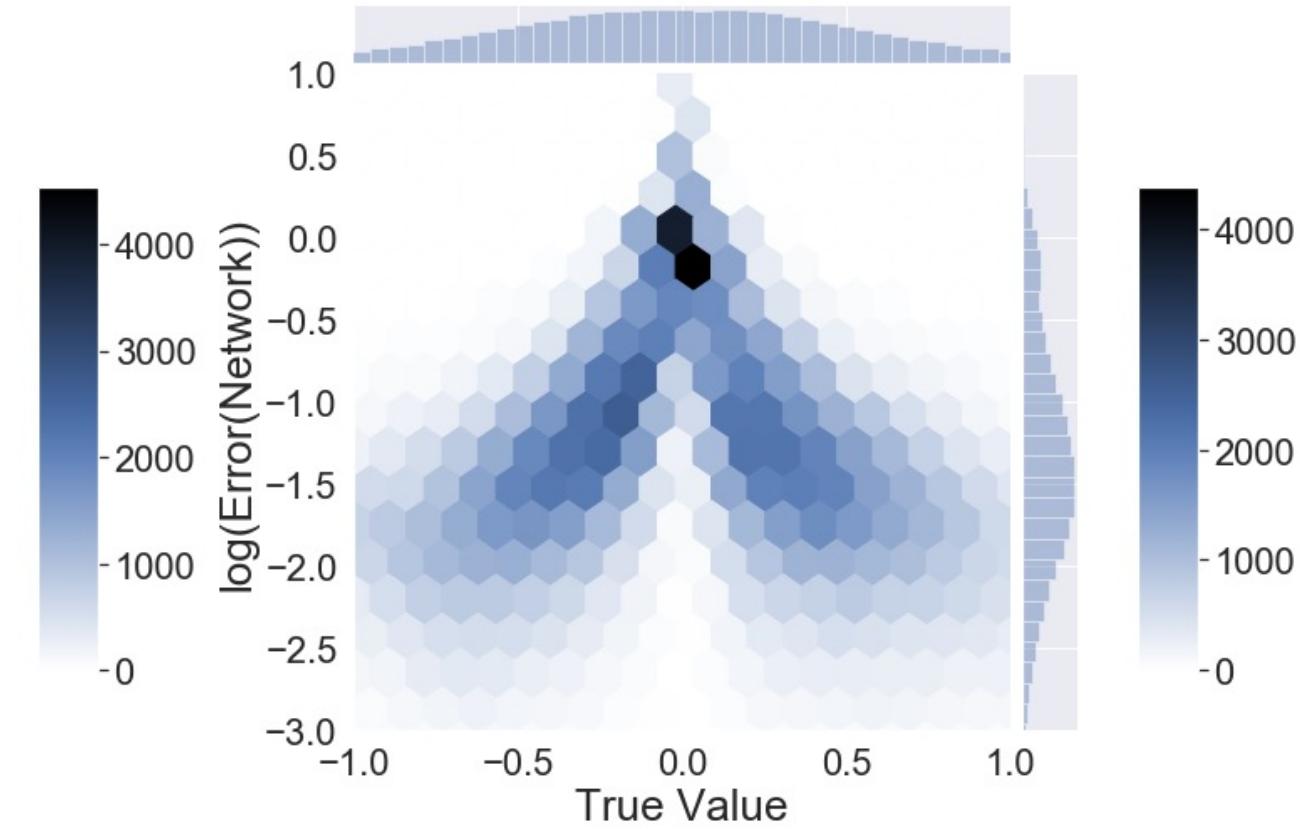
➤ 3次スプラインの最頻値 : 0.044

- ネットワークが平均して最も高い補間精度を示す
- $\text{Relative Error} = 1.0$ でネットワークのヒストグラムにピークが存在する

正解データに対するネットワークと 3 次スプラインの補間精度を比較



同じ正解データに対するネットワークと
3 次スプラインの誤差



正解データに対するネットワークの誤差

- ・3 次スプラインの誤差が大きいデータに対して、ネットワークの誤差が小さい
- ・絶対値が小さいデータに対するネットワークの誤差が大きい

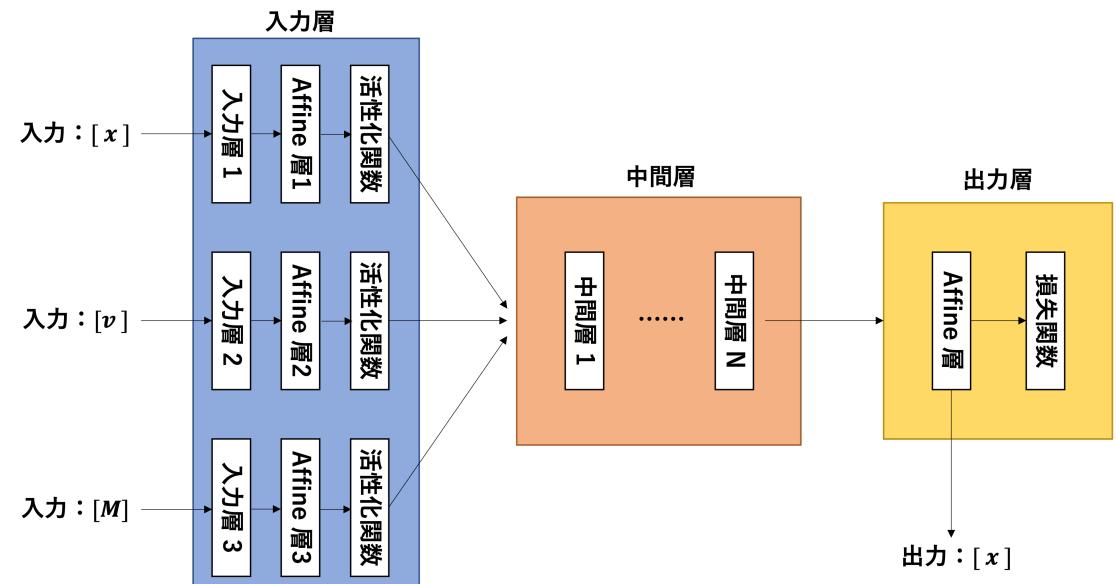
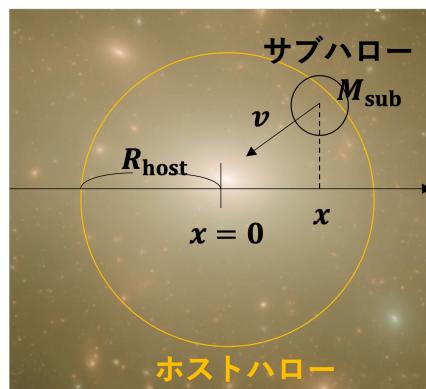
入力パラメータを x, v, M として検証

ニューラルネットワークの詳細

入力 : 出力	中間層	ニューロン数	活性化関数	損失関数	最適化関数	正規化	荷重減衰
2:3	10層	100個	TanhExp	相対誤差	Adam	Z-SCORE	$\lambda = 1.0 \times 10^{-3}$

- ・ネットワークが学習する訓練データ：低質量分解能シミュレーションのデータ
- ・補間精度を評価する検証データ：高質量分解能シミュレーションのデータ
- ・ネットワークと線形補間, 3次スプラインを比較して, ネットワークの補間精度を検証

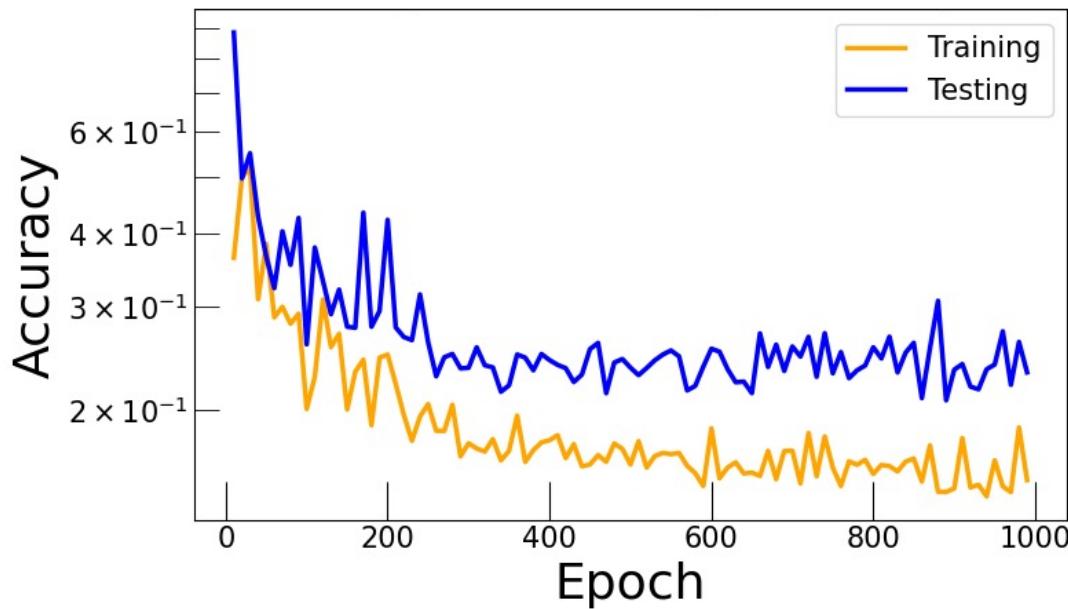
- 入力パラメータ : x, v, M
- 出力パラメータ : x
- $x = (x_{\text{sub}} - x_{\text{host}})/R_{\text{host}}$
- $v = (v_{x\text{sub}} - v_{x\text{host}})$
- $M = \log|M_{\text{sub}}(z \leq z_{\text{acc}})/M_{\text{sub}}(z = z_{\text{acc}})|$



ニューラルネットワークの全体図

ネットワークの補間精度の評価

- ・学習回数と補間精度の推移
- ・ネットワークの出力： P , 正解データ： T とすると, Accuracy = $\text{mean} \left| \frac{P-T}{T} \right|$

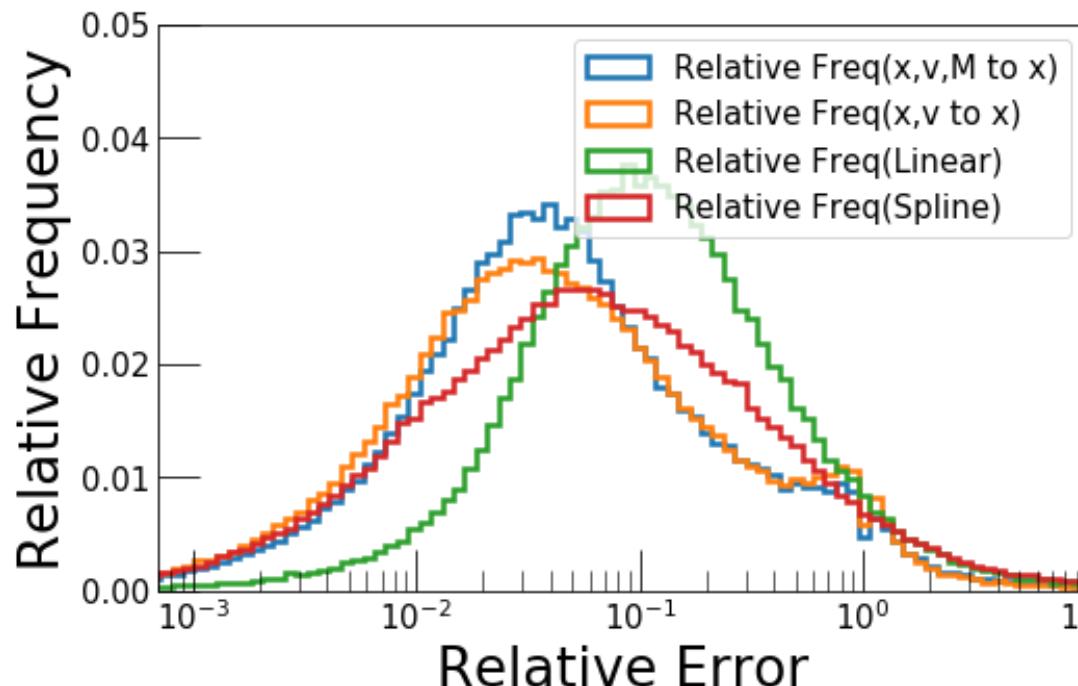


- **Training**：訓練（低質量分解能）データに対する補間精度
- **Testing**：検証（高質量分解能）データに対する補間精度
- Epochはネットワークの学習回数

- ・過学習の傾向がある
- ・入力パラメータに M を追加すると Accuracy の振動が小さくなり安定する

ネットワークと線形補間, 3次スプラインの補間精度を比較

- ・検証（高質量分解能）データを各補間手法で補間したときの, 正解データに対する補間データの相対誤差のヒストグラム



➤ 補間データ : $x_{predicted}$, 正解データ : T

$$\text{Relative Error} = \left| \frac{x_{predicted} - T}{T} \right|$$

➤ $x, v, M \rightarrow x$ の最頻値 : 0.039

➤ $x, v \rightarrow x$ の最頻値 : 0.034

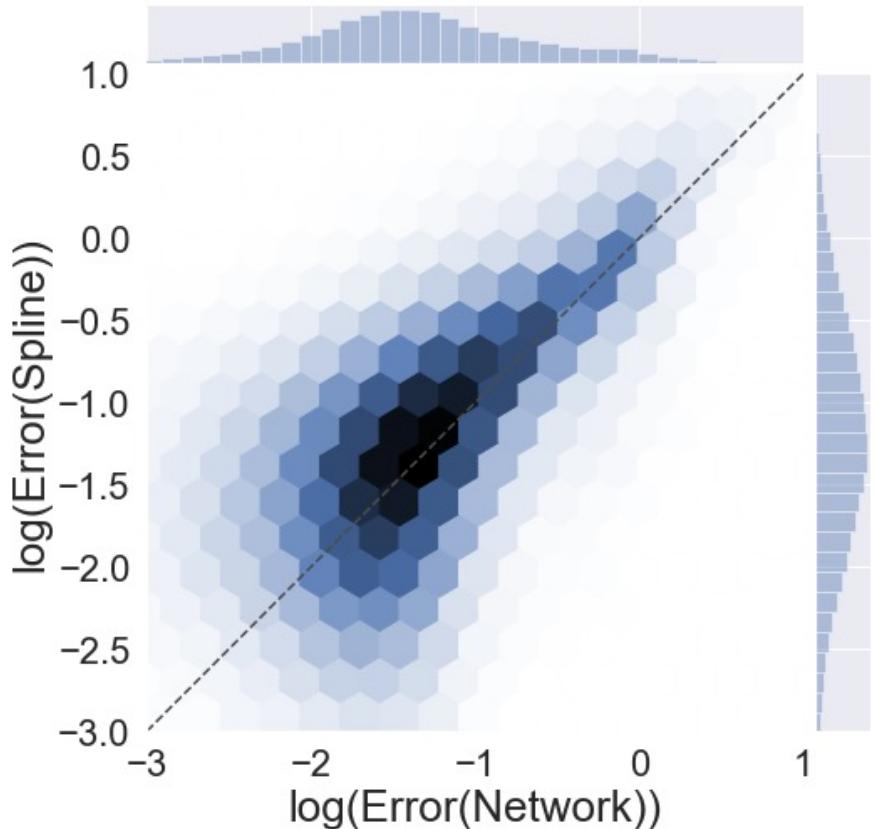
➤ 線形補間の最頻値 : 0.088

➤ 3次スプラインの最頻値 : 0.044

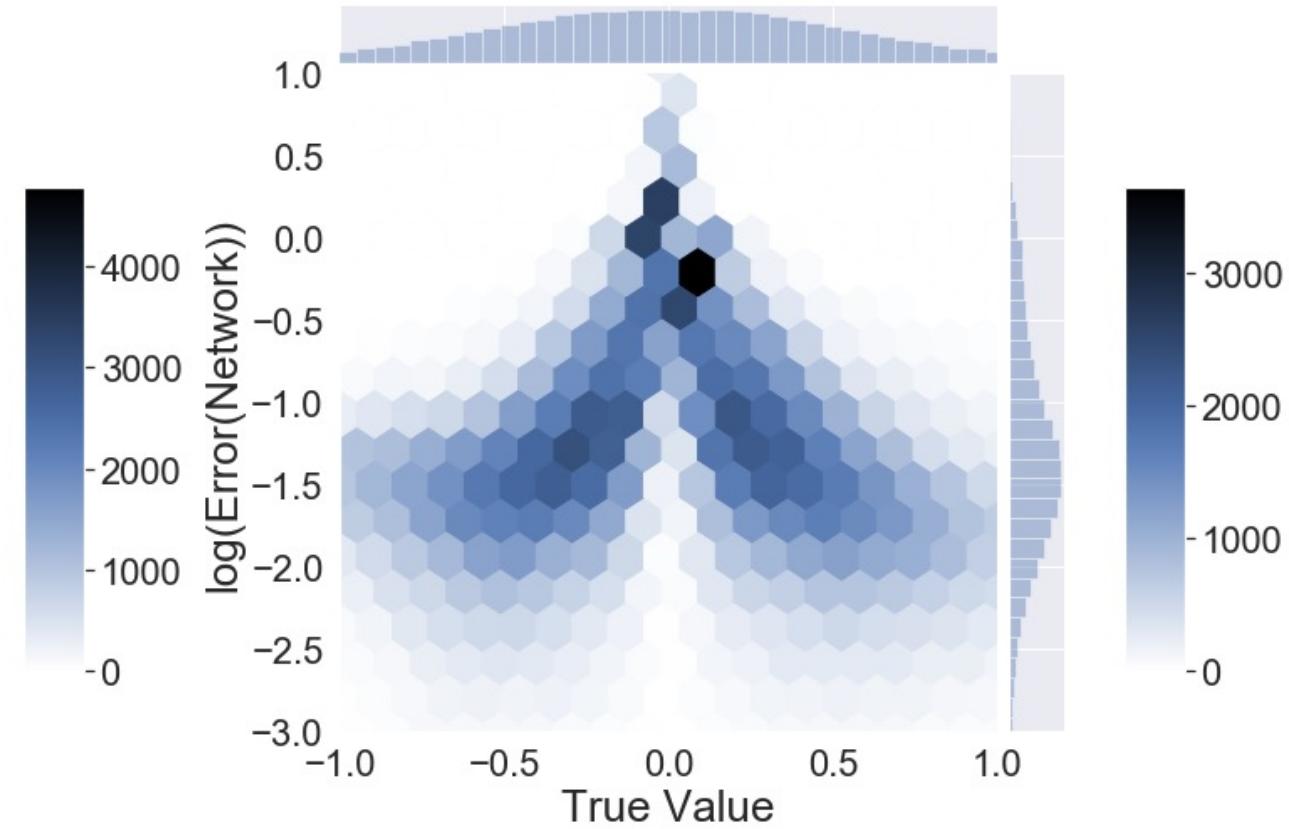
- ・いずれのネットワークも 3 次スプラインより高い補間精度を示す

- ・ $x, v, M \rightarrow x$ のネットワークでは $\text{Relative Error} = 1.0$ のピークが消える

正解データに対するネットワークと 3 次スプラインの補間精度を比較



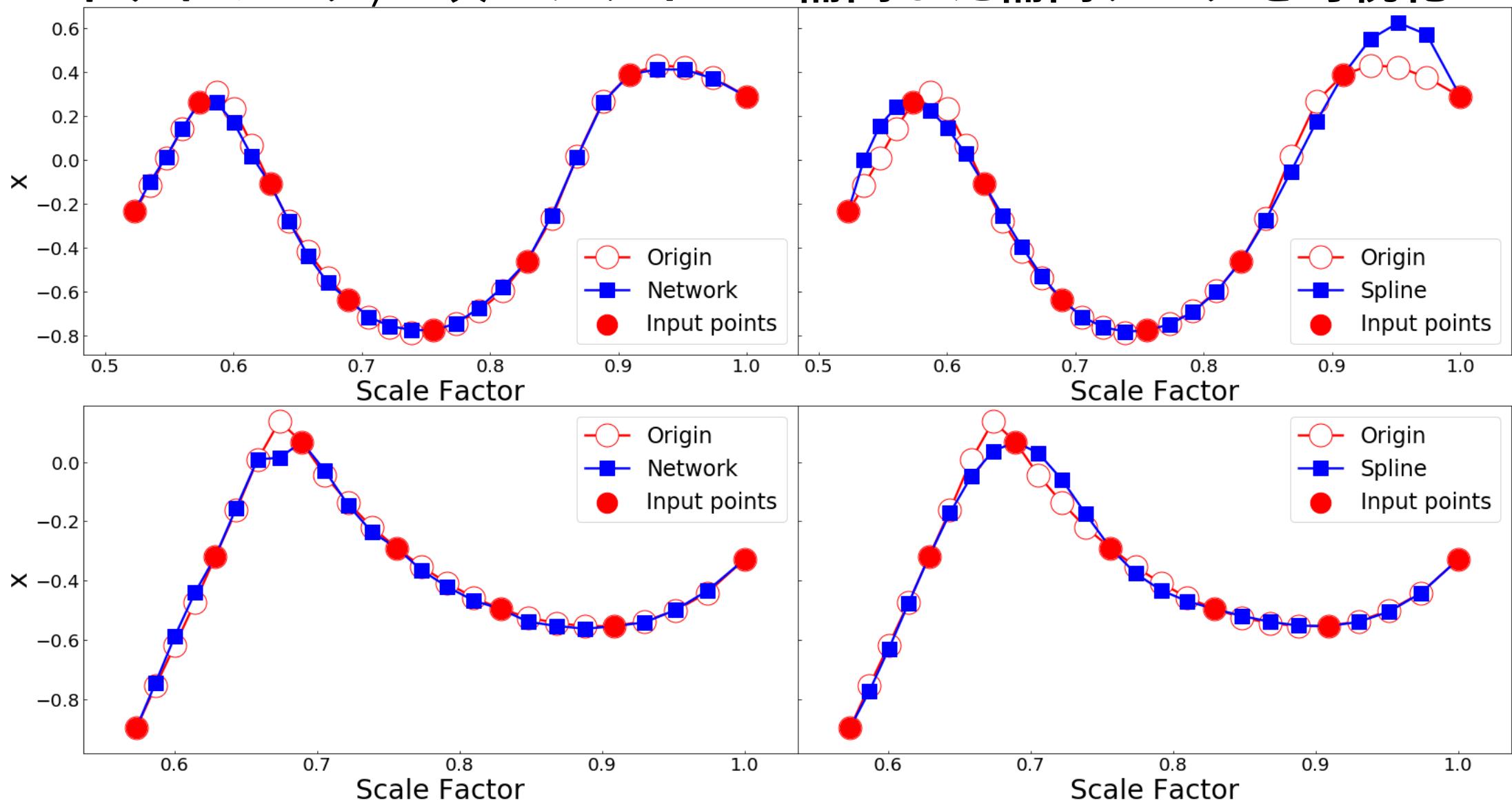
同じ正解データに対するネットワークと
3 次スプラインの誤差



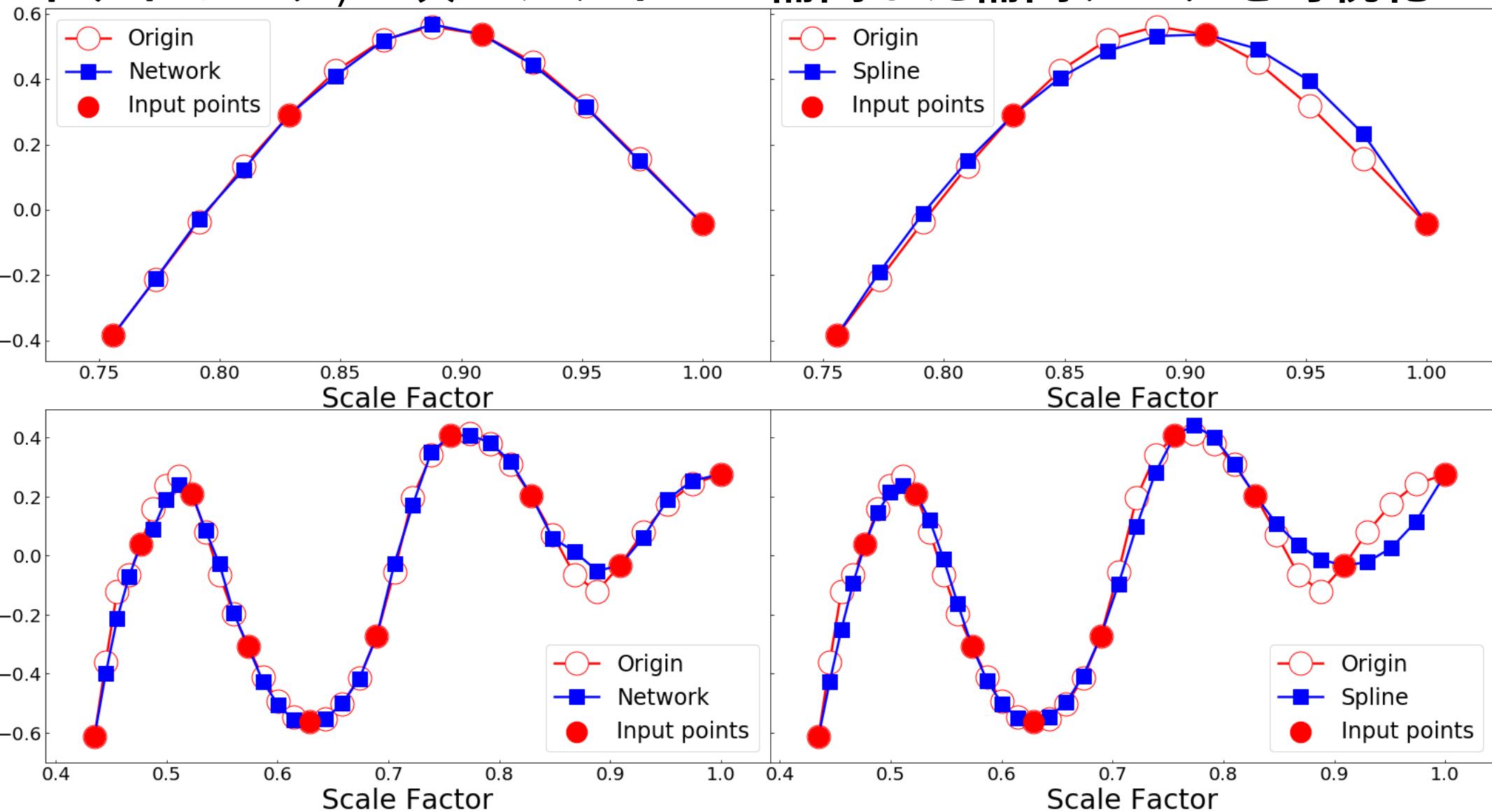
正解データに対するネットワークの誤差

- $x, v \rightarrow x$ のときと比較して、全体の傾向は変わらない
- 絶対値が小さいデータに対して $\text{Relative Error} = 1.0$ のデータ数が減少する

ネットワーク, 3次スプラインで補間した補間データを可視化



ネットワーク, 3 次スプラインで補間した補間データを可視化



まとめ

- Phi - 4096 (Ishiyama et al. 2021) の低質量分解能シミュレーションを学習したネットワークで, 高質量分解能シミュレーションの merger tree 上の時刻間の情報を補間した
- ネットワークは正規化, 規格化された x, v を入力とし, x を補間した
また, M を入力に含むネットワークも作成し, それらを比較した
 M を入力に含むネットワークの学習は安定した補間精度を示した
- ネットワークには過学習の傾向が少し見られた
- ネットワーク, 線形補間, 3 次スプラインの補間精度を比較した結果,
ネットワークの補間精度が最も高かった
- ネットワークと3 次スプラインで補間したデータを可視化した結果,
3 次スプラインで補間精度が低いデータに対して, ネットワークは高い補間精度を
示した