Machine Learning Analysis on the JET Pedestal Database

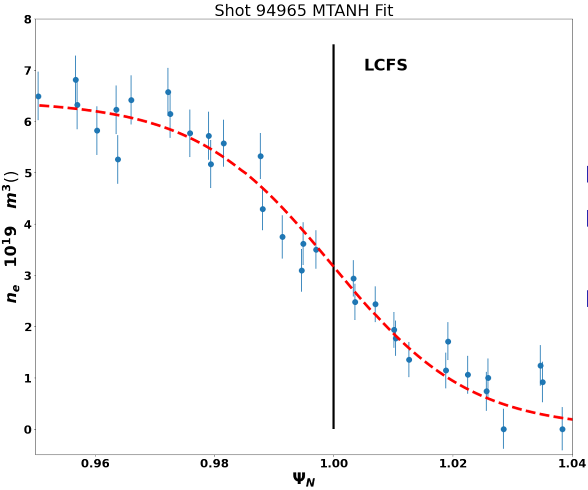
or how I learned to worry about high n_e^{ped}

9 June, 2021

Overview

Table of Contents

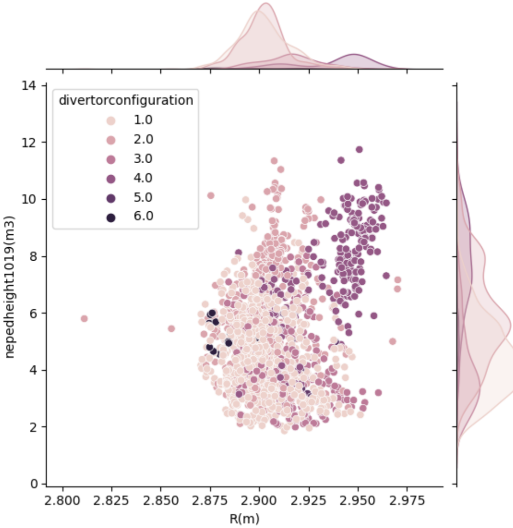
JET Pedestal Database



- H-Mode
- 75-95% ELM Cycle
- Focus on n_e^{ped} height

Input Parameters and Filtering

Eng. Param	Domain
I_P [MA]	$[0.81, 4.48]$
B_T [MW]	$[0.97, 3.68]$
a [m]	$[0.83, 0.97]$
R [m]	$[2.8, 2.975]$
δ [-]	$[0.16, 0.48]$
M_{eff} [-]	$[1.0, 2.18]$
P_{NBI} [MW]	$[10^{-3}, 32.34]$
P_{ICRH} [MW]	$[0, 7.96]$
P_{TOT} [MW]	$[3.4, 38.22]$
V_P [m^3]	$[58.3, 82.19]$
q_{95} [-]	$[2.42, 6.04]$
Γ [10^{22} e/s]	$[0, 15.5]$
H [-]	$[0, 0.18]$
Γ_{SD} [10^{22} e/s]	$[0, 1000]$
DC [-]	$[VV \cdots]$

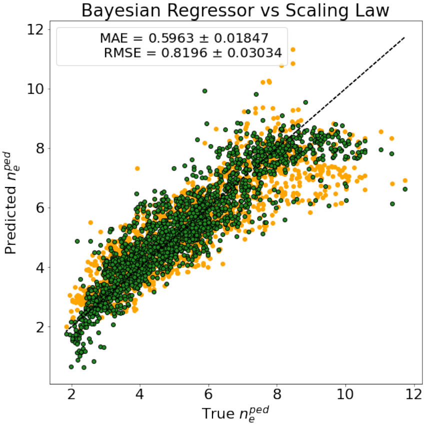


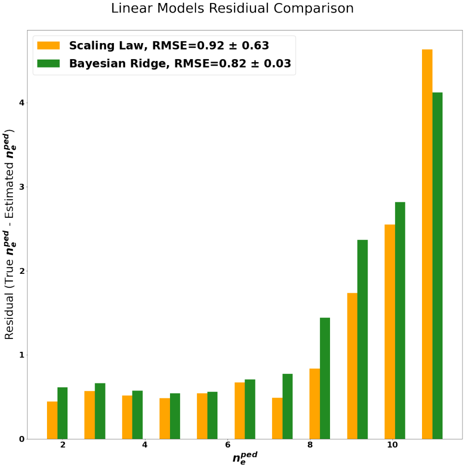
- Shafranov Shift$\rightarrow$ remove R
- Γ_{SD} varies too much.
- Only numerical
- Only Deuterium
- No RMPs, Kicks, or Pellets
- HRTS Validated

Eng. Param	Domain
I_P [MA]	$[0.81, 4.48]$
B_T [MW]	$[0.97, 3.68]$
a [m]	$[0.83, 0.97]$
$\cancel{R\ [\mathrm{m}]}$	$\cancel{[2.8, 2.975]}$
δ [-]	$[0.16, 0.48]$
M_{eff} [-]	$[1.0, 2.18]$
P_{NBI} [MW]	$[10^{-3}, 32.34]$
P_{ICRH} [MW]	$[0, 7.96]$
P_{TOT} [MW]	$[3.4, 38.22]$
V_P [m^3]	$[58.3, 82.19]$
q_{95} [-]	$[2.42, 6.04]$
Γ [10^{22} e/s]	$[0, 15.5]$
H [-]	$[0, 0.18]$
$\cancel{\Gamma_{SD}\ [10^{22} e/s]}$	$\cancel{[0, 1000]}$
$\cancel{DC\ [-]}$	$\cancel{[VV \cdots]}$

Table of Contents

Linear Regression





$$n_e^{ped} = (9.9 \pm 0.3) I_p^{1.24 \pm 0.19} P_{TOT}^{-0.34 \pm 0.11} \delta^{0.62 \pm 0.14} \Gamma^{0.08 \pm 0.04} M_{eff}^{0.2 \pm 0.2} \tag{1}$$

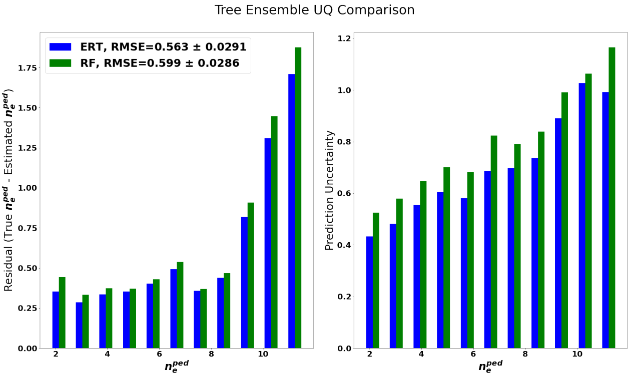
Linear Regression

- Achieve better RMSE using all inputs
- But lose interpretability $\rightarrow$
- Prediction uncertainty normally distributed between 1.6 - 1.8

Feature	μ	σ^2
I_p	0.15	0.06
B_T	0.956	0.072
a	2.966	0.479
δ	12.95	0.154
V_P	-0.05	0.007
q_{95}	-1.064	0.0542
P_{NBI}	-1.911	0.0546
P_{ICRH}	-1.976	0.0561
P_{TOT}	1.926	0.0557
Γ	0.125	0.007
H	-4.016	0.374
M_{eff}	1.369	0.053

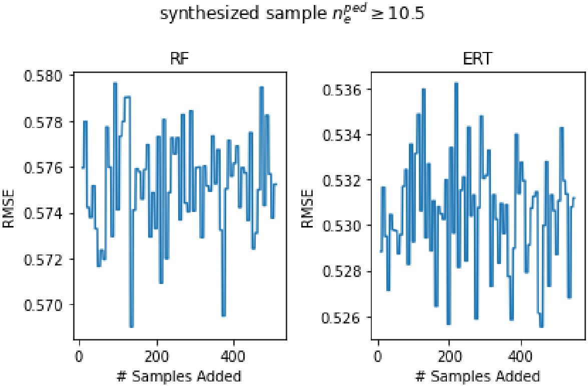
Random Forests & Extreme Randomized Trees

- Uncert. covers residual



Random Forests & Extreme Randomized Trees

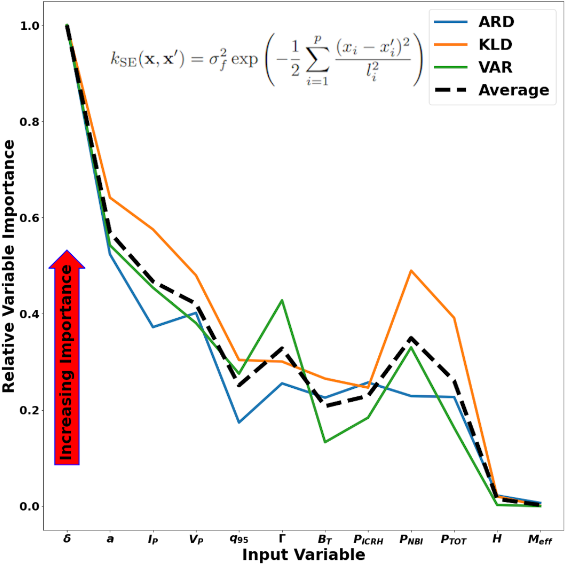
- Uncert. covers residual
- Meta-modeling $\rightarrow$ no effect



Gaussian Processes

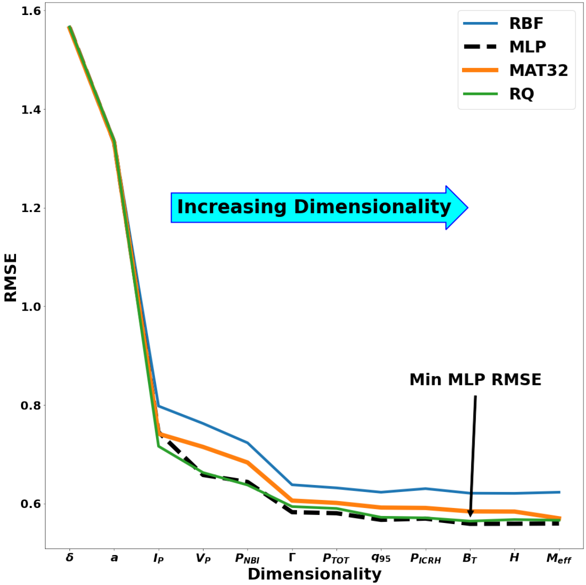
- Sensitivity Analysis

$$\frac{d(p(y_*|\mathrm{x}^{\mathrm{i}},\mathrm{y})||p(y_*|\mathrm{x}^{\mathrm{i}}+\Delta_{\mathrm{j}},\mathrm{y}))}{\Delta}$$



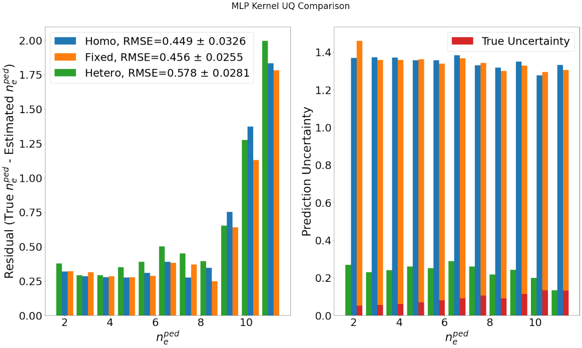
Gaussian Processes

- Sensitivity Analysis



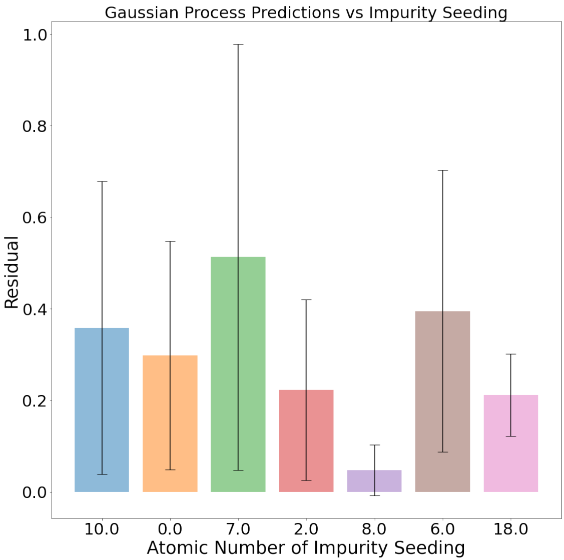
Gaussian Processes

- Sensitivity Analysis
- UQ



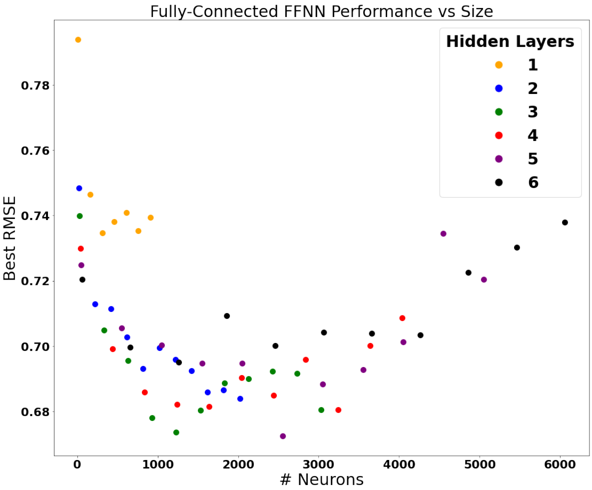
Gaussian Processes

- Sensitivity Analysis
- UQ
- Impurity seeding



ANNs

- Crawling Search Space



ANNs

- Crawling Search Space
- Ensembling

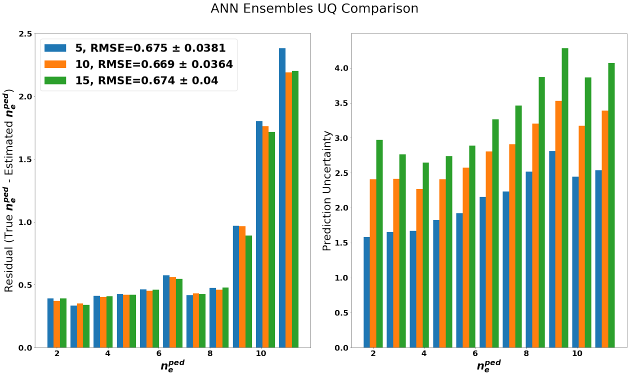


Table of Contents

Conclusion

- Non-linear models outpreform linear models
- RFs and ERTs work well as black box models
- Heteroscedastic GPs can map latent uncertainty
- Shallow ANNs perform best

Future Work

- Fit models on varied subsets of database
  e.g., compare how linear coefficients vary when trained on $q_{95} \leq 3.0$ versus $q_{95} > 3.0$.
- UQ of main engineering parameters
- Ideas?