Using machine learning techniques to study plasma fusion experiments

M. Giselle Fernández-Godino, Julia B. Nakhleh, Michael J. Grosskopf, Brandon M. Wilson, and Gowri Srinivasan

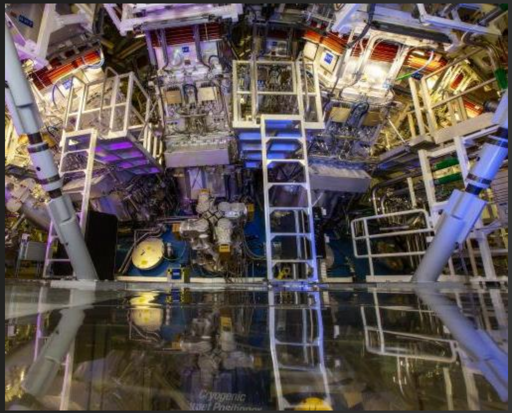
Thursday, July 09, 2020

Los Alamos NATIONAL LABORATORY EST. 1943

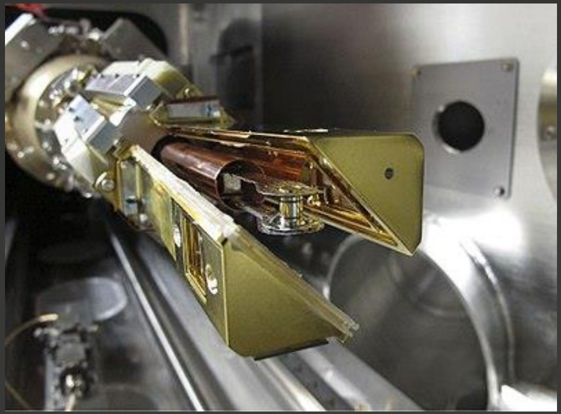
NNSA National Nuclear Security Administration

I. Introduction

ICF at the National Ignition Facility (NIF)



Target chamber

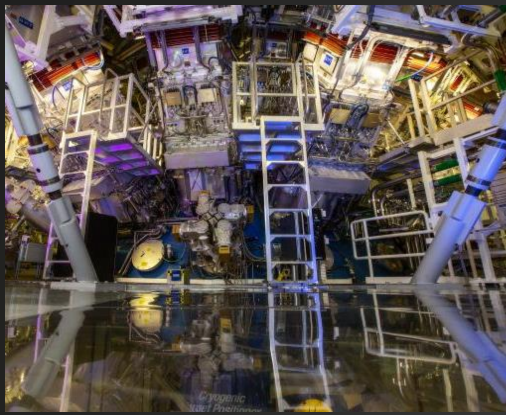


Target assembly

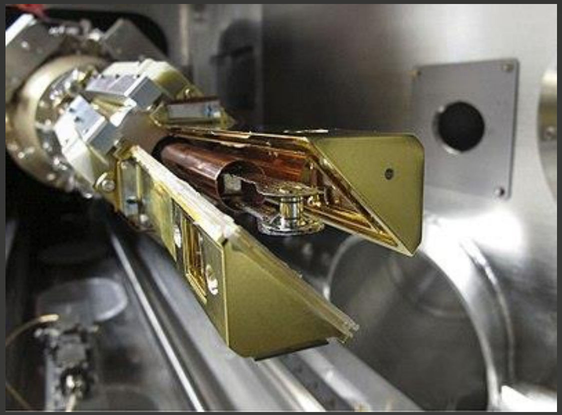
Neutron imaging system

[1] Photo and figures extracted from NIF webpage (https://lasers.llnl.gov/)

ICF at the National Ignition Facility (NIF)



Target chamber



Target assembly

Neutron imaging system

2 Emitted X rays heat the outer surface of the pellet

4 Plasma ignition yielding to many times input energy

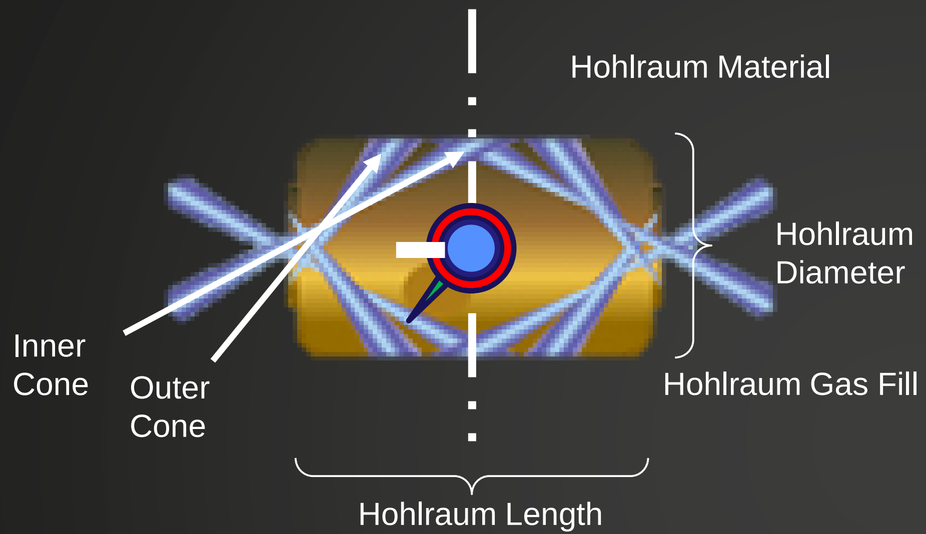
1 Laser beams heat the hohlraum surface

3 Implosion of the fuel capsule (10^2 times the lead density and 10^8 K)

[1] *Photo and figures extracted from NIF webpage (https://lasers.llnl.gov/)*

II. Experimental **Data**

Input Variables: Hohlraum

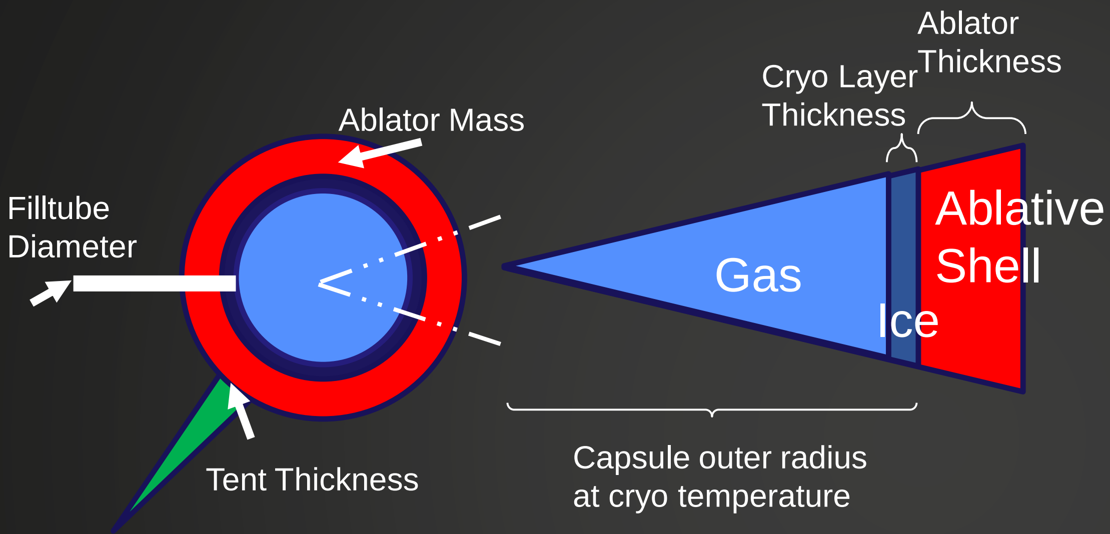


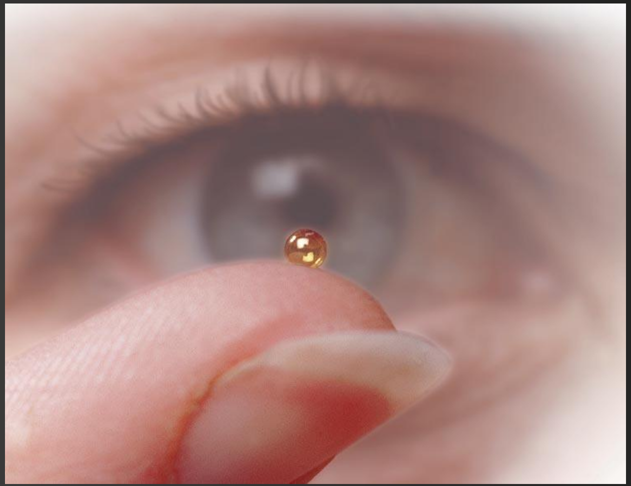
Hohlraum Inputs	Units
Hohlraum length	mm
Hohlraum diameter	mm
Hohlraum gas fill	mg/cc
Hohlraum material	unitless
Dante 1 LEH Diameter	mm

- The lasers heat the inner walls of a gold cavity called **hohlraum** containing the pellet
- 5 input variables are directly related with the **hohlraum**

[2] Photo and figure extracted from NIF webpage (https://lasers.llnl.gov/)

Input Variables: Capsule





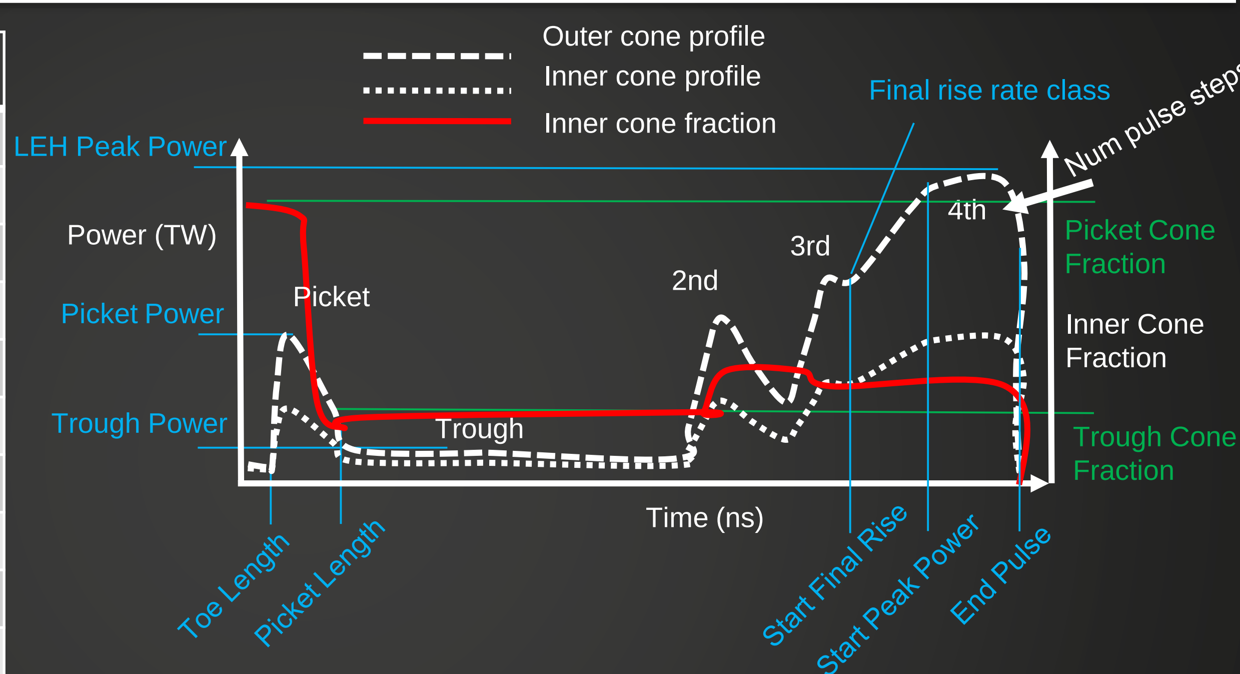
Capsule Inputs	Units
Ablator thickness	μm
Cryo layer thickness	μm
Capsule outer radios at cryo T	μm
Tent thickness	nm
Ablator mass	mg
Filltube diameter	μm

- Often contains a mixture of deuterium and tritium
- About the size of a pinhead and it contains around 10 milligrams of fuel
- 6 input variables are directly related with the **capsule**

[3] Photo extracted from NIF webpage (https://lasers.llnl.gov/)

Input Variables: Laser

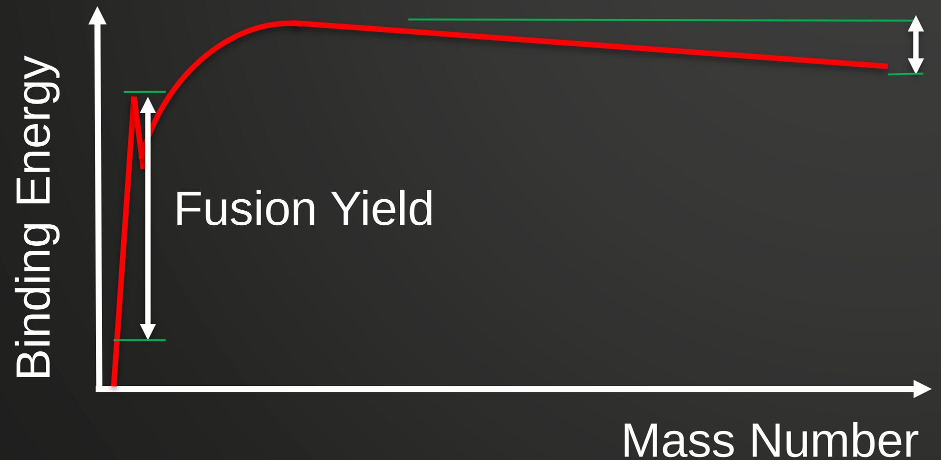
Laser Inputs	Units
LEH laser energy	MJ
Trough cone fraction	unitless
LEH peak power	TW
Picket power	TW
Trough power	TW
Toe length	ns
Picket cone fraction	unitless
Picket length	ns
Number of pulse steps	unitless
Final rise rate class	ns
Start final rise	ns
Start peak power	ns
End pulse	ns
$\Delta\lambda_3$ - $\Delta\lambda_2$	angstrom
$\Delta\lambda_2$	angstrom



- It reaches the target from numerous directions at the same time, within a few picoseconds
- The current design uses 192 beamlines
- 15 input variables are directly related with the **laser**

Output Variable: Total Yield

- Nuclear fusion is very attractive because is a clean energy approximately 4 times more efficient than fission, which is already 8,000 times more efficient than coal.
- However, as a form of energy it is a challenge because it faces technical and practical problems.
- The "Total yield" measures the number of neutrons produced during the entire reaction and it is the output quantity in this work



Tritium, Neutron

$$ {}_{1}^{2}H + {}_{1}^{3}H \rightarrow {}_{2}^{4}He + {}_{0}^{1}n + \boxed{17.6\ MeV} $$

Deuterium, Helium, Net Energy

III. Overview

Presentation Overview

1. **High-dimensional experimental dataset**
   - 26 inputs / 1 output
   - 141 experiments (2011-2019)
2. **Surrogate model selection**
   - Multiple options tested
   - Chosen based on the performance on the test data
3. **ML to interpret complex coupling between inputs**
   - Sparse PCA
4. **Importance of variables**
   - ALE Importance
   - Physical variables and sparse components
5. **Surrogate performance**
   - Physical variables and sparse components
6. **Future work**
   - Combine experiments and simulations for validation purposes
   - Other outputs

IV. Surrogate models

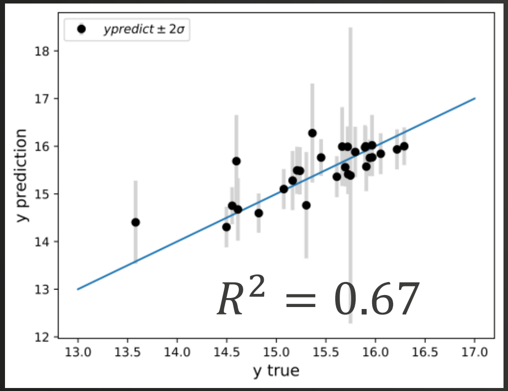
Surrogate Model Selection

- Our first step was to compare multiple **surrogate models** performance to select the best

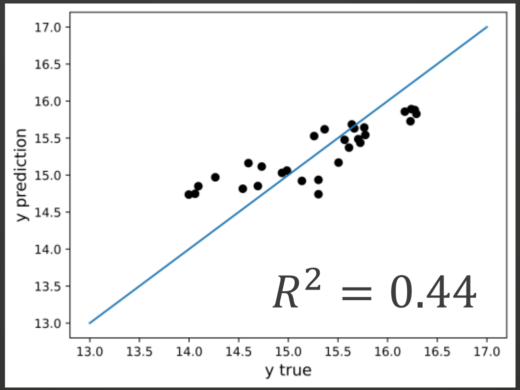
- We tried Gaussian Process, Neural Networks and Random Forest

Surrogate Model Selection

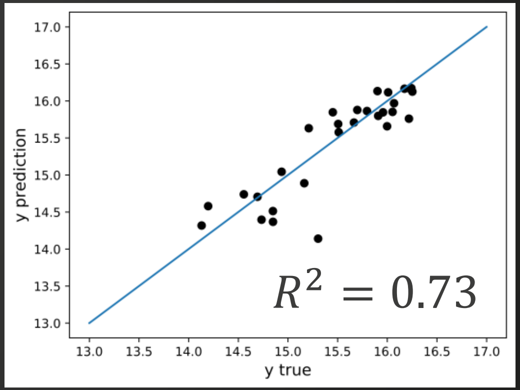
- Our first step was to compare multiple **surrogate models** performance to select the best
- We tried Gaussian Process, Neural Networks and Random Forest
- These figures show the surrogate performance in the test data



Gaussian Process*



Neural Networks*

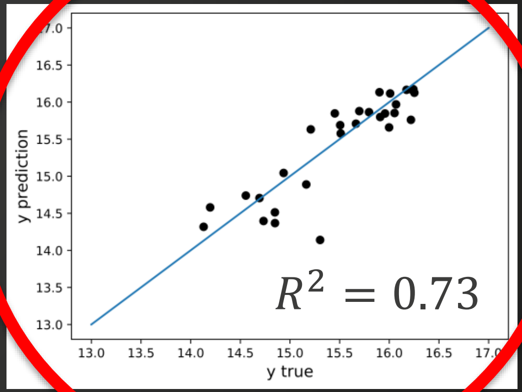


Random Forest*

* R^2 is the mean of 1000 fitting iterations to take into account the randomness in the fitting process

Surrogate Model Selection

- Our first step was to compare multiple **surrogate models** performance to select the best
- We tried Gaussian Process, Neural Networks and Random Forest
- These figures show the surrogate performance in the **test data**

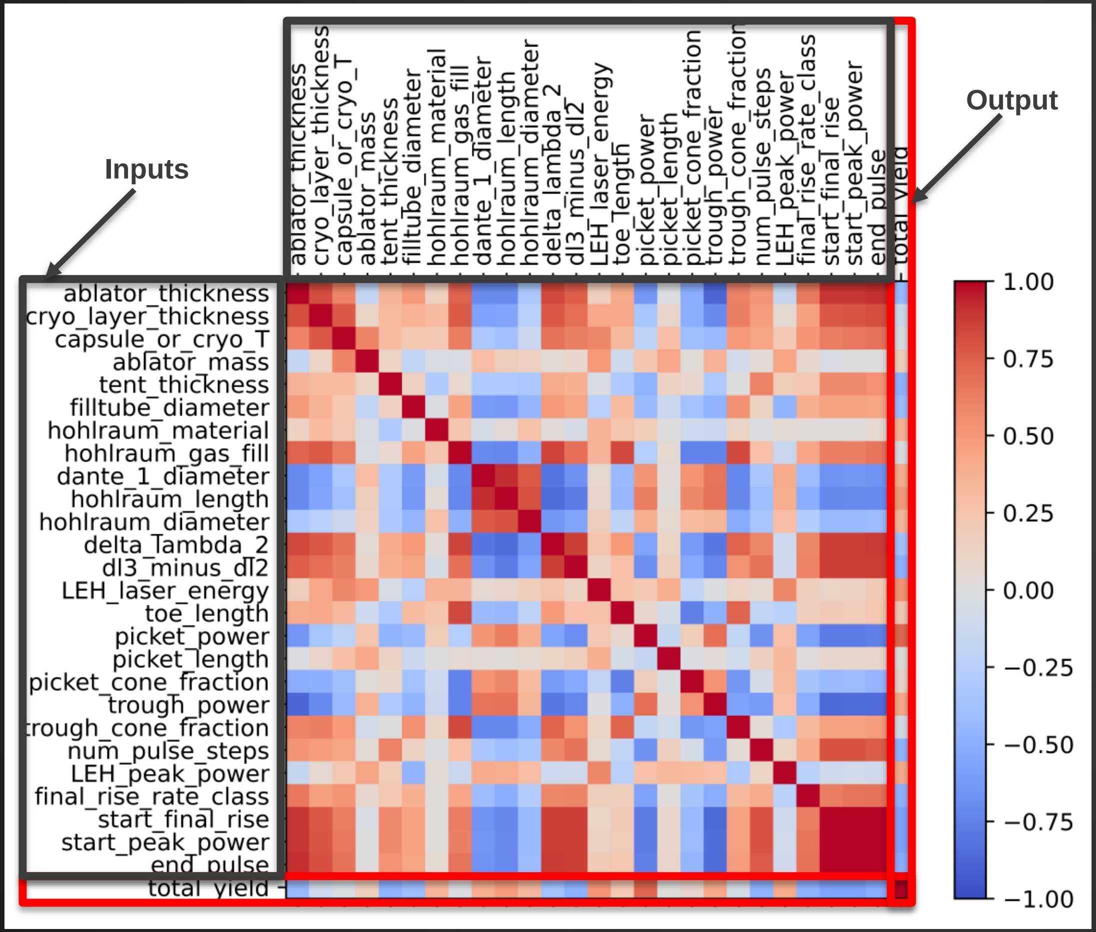


Random Forest*

* R^2 is the mean of 1000 fitting iterations to take into account the randomness in the fitting process

V. Variable **grouping**

Input - Output Correlation Matrix

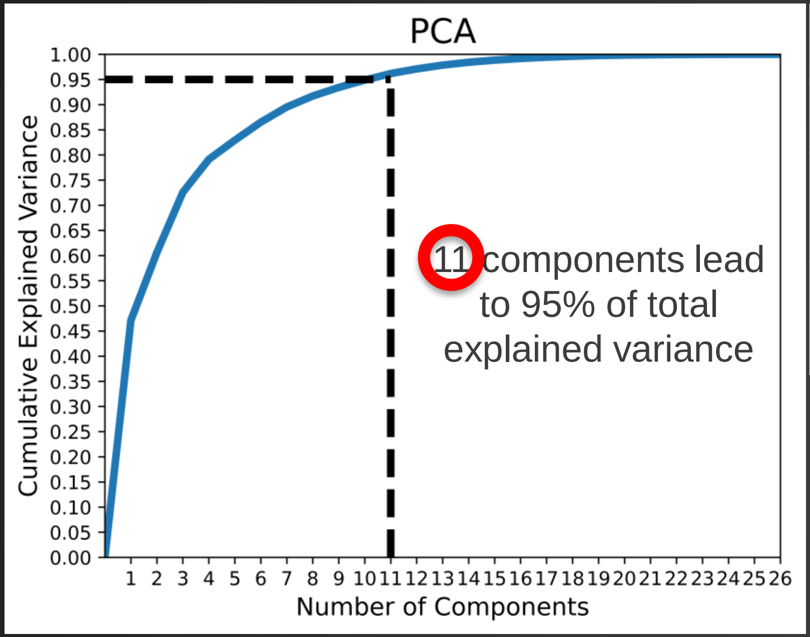


- The matrix shows the correlations between variables
- We found extremely high correlations between “*Start final rise*”, “*Start peak power*” and “*End pulse*”
- We do not see extremely high correlations with the output
- The highest positive correlations with the output are with “*Picket power*” and “*LEH laser energy*”

Sparse PCA for Data-Driven Clustering of Design Inputs

- To reduce the effect of correlated features in assessing relationships between inputs and yield, we utilize Sparse PCA to define clusters of related variables
- To judge meaningfulness of the clusters we use expert assessment
- Relationships between groups and "*Total yield*" can then be assessed, leveraging physics expertise to evaluate the important features within a group

Principal Component Analysis (PCA)



	RF $\pm$ 2σ	RF_PCA $\pm$ 2σ
# Variables	26	11
Rel. RMSE train (%)	0.81±0.05	0.96±0.06
Rel. RMSE test (%)	1.8±0.1	1.7±0.1

$$Rel.\,RMSE\ (\%) = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{y_{pred}^{i} - y_{true}^{i}}{y_{true}^{i}}\right)^2}{N}} \times 100$$

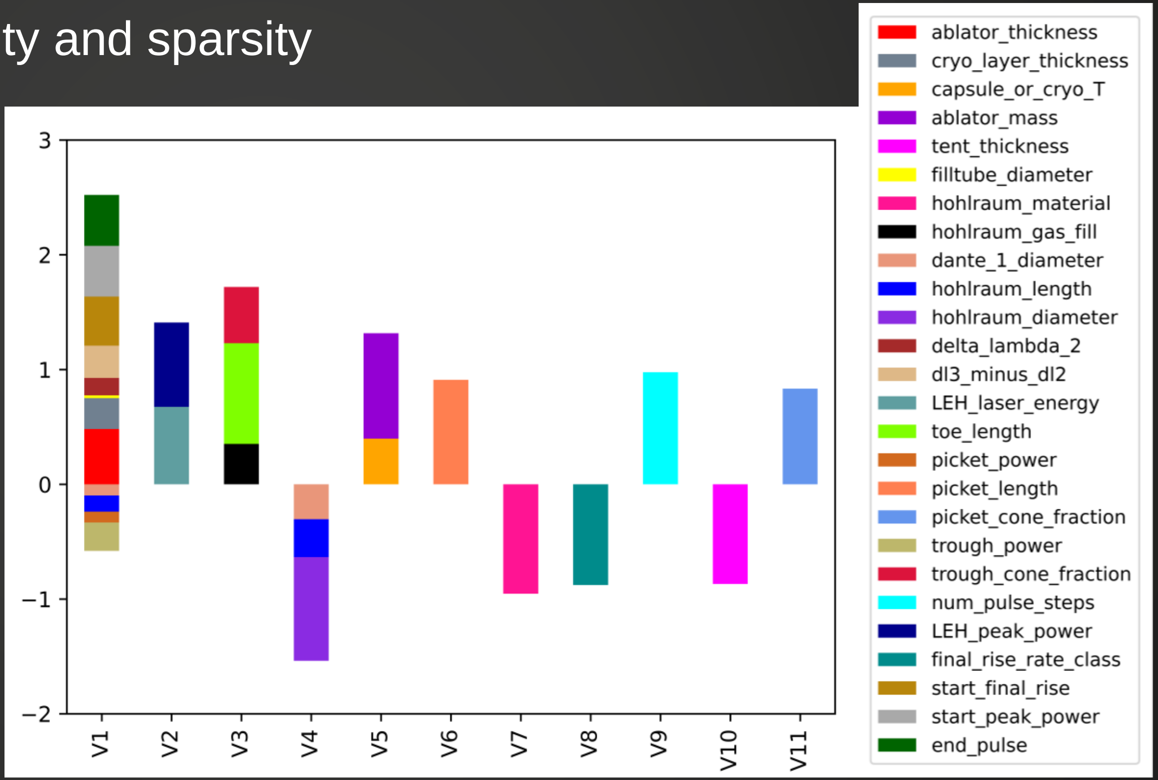
- PCA was used to determined that 11 components explain 95% of the variance within the data
- Errors show that the performance of using only 11 components compares with the performance of using 26

Sparse Principal Component Analysis (SPCA)

- SPCA is used to find meaningful groupings within the physical variables via ML

Input: Hohlraum
Input: Capsule
Input: Laser

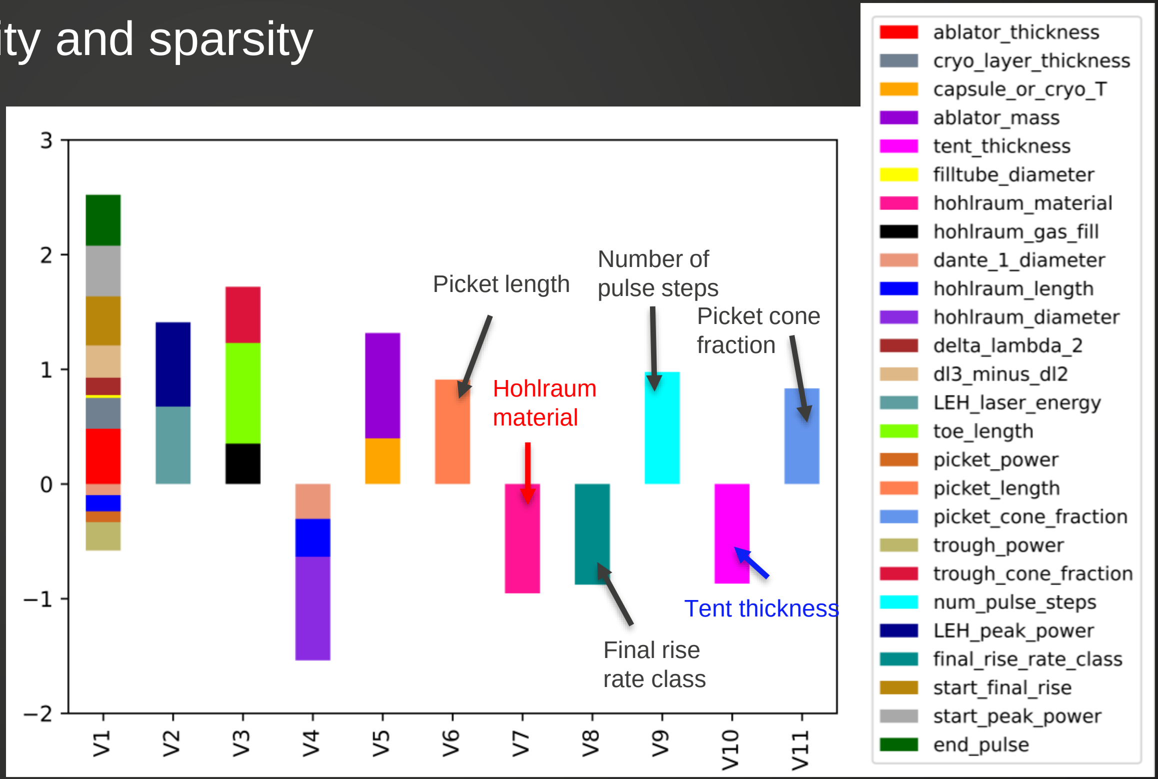
- SPCA balances orthogonality and sparsity



Sparse Principal Component Analysis (SPCA)

- SPCA is used to find meaningful groupings within the physical variables via ML
- SPCA balances orthogonality and sparsity
- 6 single components

Input: Hohlraum
Input: Capsule
Input: Laser



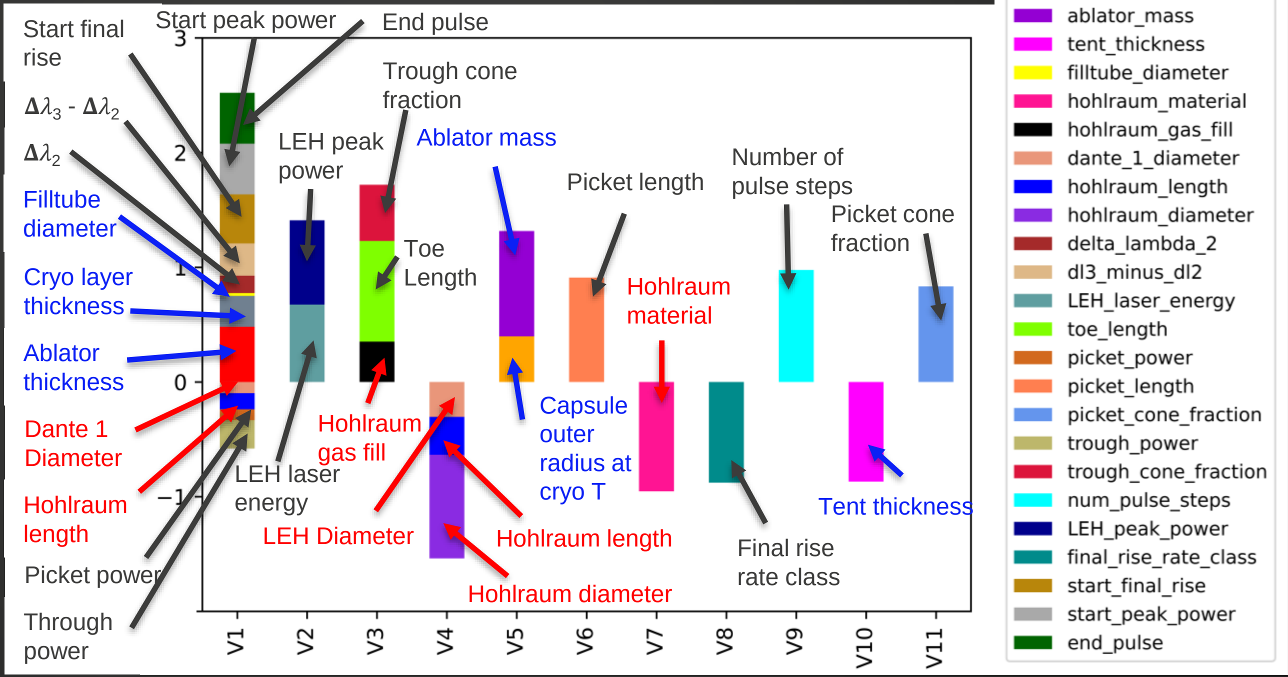
Sparse Principal Component Analysis (SPCA)

- SPCA is used to find meaningful groupings within the physical variables via ML

Input: Hohlraum
Input: Capsule
Input: Laser

- SPCA balances orthogonality and sparsity

- 6 single components
- 5 combined components
- 3 of 5 combined components have physical variables from a single input group



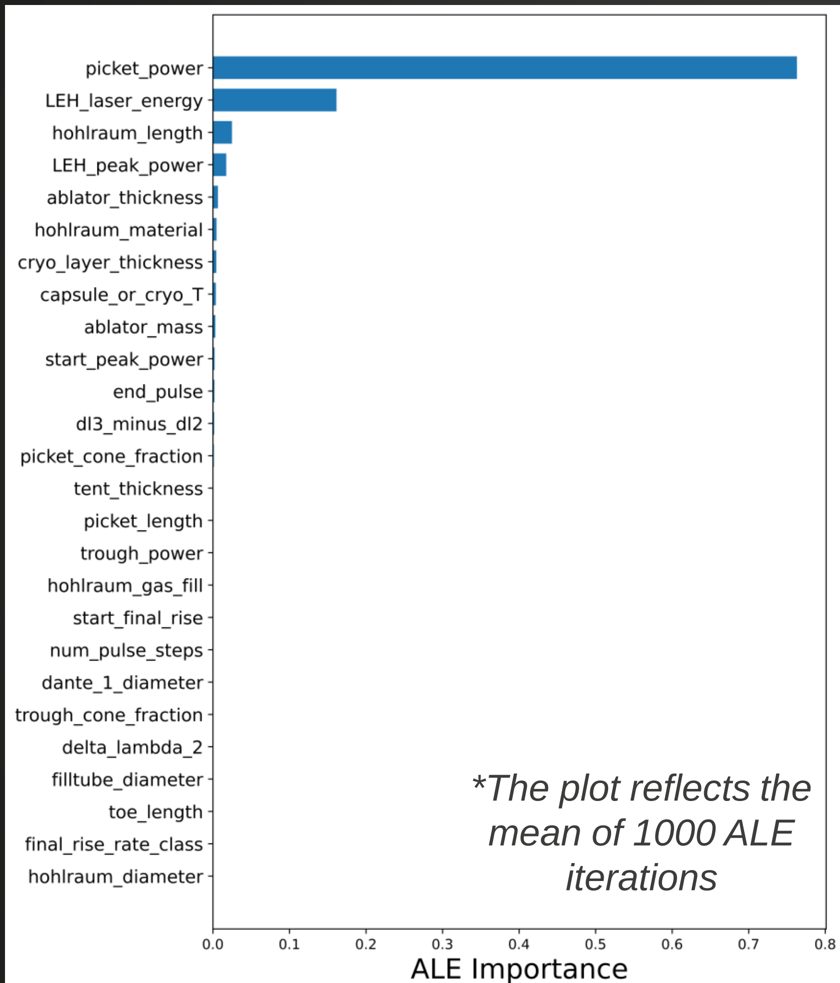
VI. **Importance** metric

Assessing which features strongly contribute to prediction

- Understanding the relationship between both, the design inputs and the newly found groupings, and the output *"Total yield"* can be useful for gaining new physical insights
- We have chosen to measure importance through the accumulated local effects (ALE)[4]
- ALE describes how features or groups of features influence the prediction of an ML model on average
- ALE averages the changes in the predictions and accumulates them over the grid
- ALE is highly recommended when variables are strongly correlated

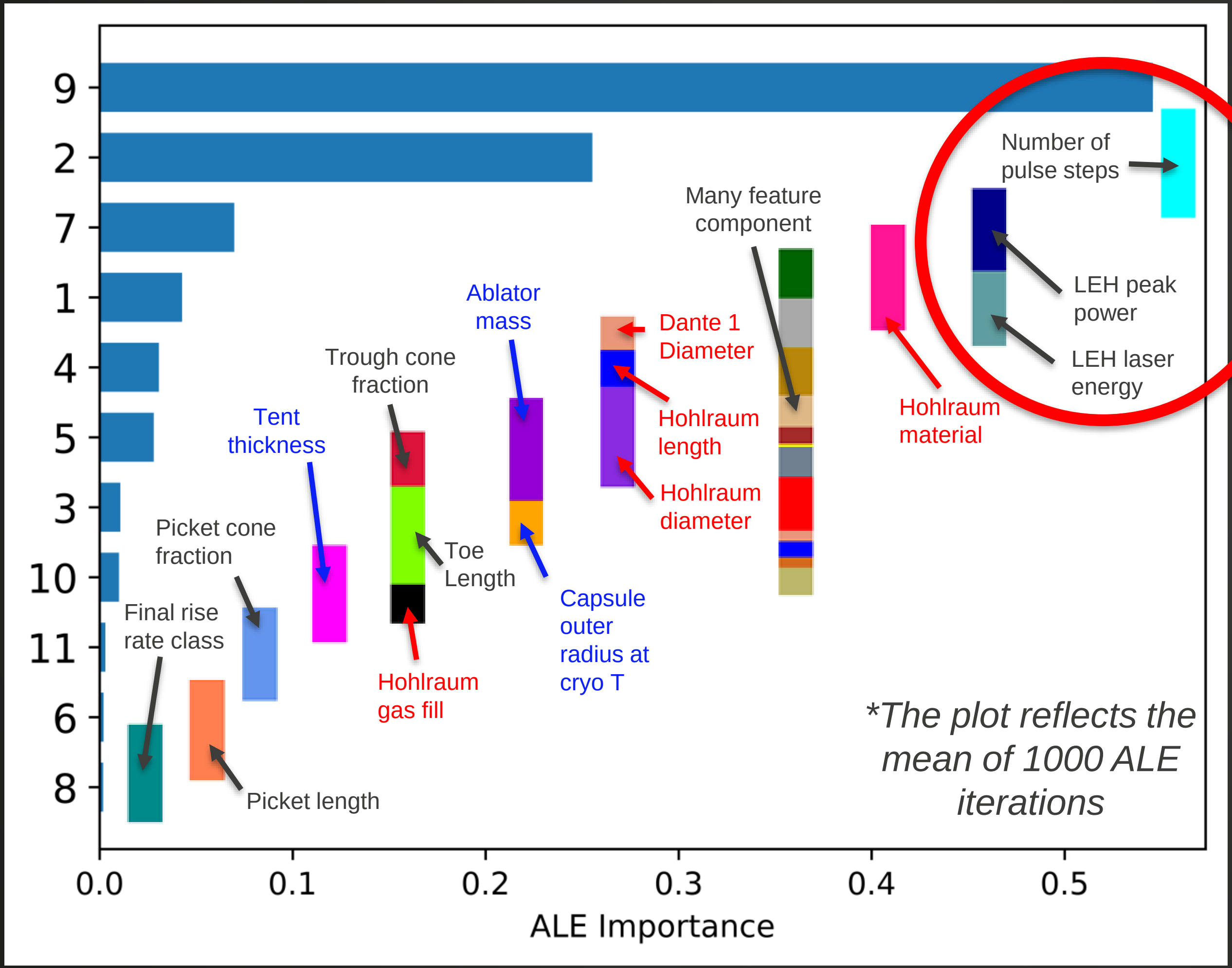
[4] Apley, Daniel W. "Visualizing the effects of predictor variables in black box supervised learning models." arXiv preprint arXiv:1612.08468 (2016)

ALE Plots using Random Forest



- "*Picket power*" and "*LEH laser energy*" are the most important features for predicting "*Total yield*"
- "*Picket power*" and "*LEH laser energy*" are highly positively correlated with "*Total yield*"
- This is consistent with what we found through the correlation matrix
- See J. Nakhleh's poster presentation for more application of ALE importance to this ICF data

ALE Plots using Random Forest trained with SPCs



- The two most important components are PC9 ("*Number of pulse steps*") and PC2 ("*LEH peak power*" and "*LEH laser energy*")
- PC9 ("*Number of pulse steps*") does not seem to be highly correlated with "*Total yield*"
- PC2 is highly positively correlated with "*Total yield*"

Random Forest Performance

- Random forests have been trained using the original dataset and the sparse dataset
- The Rel. RMSE reported is the average of 1000 fitting iterations to take into account the randomness in the fitting process
- The performance on the test data is ≈ 2% for both RF and RF_SPCA

	RF	RF_SPCA
Rel. RMSE train data (%)	0.76±0.05	0.77±0.03
Rel. RMSE test data (%)	1.91±0.07	2.1±0.1

VII. **Conclusions**

Concluding Remarks

- Using 11 principal components we can predict "*Total yield*" as accurately as with the original 26 variables
- SPCA gives groupings that are related with the physical origin of the variables (laser, hohlraum or capsule)
- ALE Importance trained with the original 26 variables found two important variables. Both variables are highly positively correlated with the output, "*Total yield*"
- ALE Importance trained with the 11 sparse components gave a more balanced importance between components. The most important component is not highly correlated with the output, "*Total yield*".

VIII. **Future** Work

Future Work

- Although if we have made progress identifying design input groupings we will continue studying ML based grouping methods and discussing the findings with the experts.

- We will also extend this work to other outputs besides “*Total yield*” such as “*Velocity*” and “*Neutron yield*”.

- We would like to investigate the connection between simulations and experiments

- Simulations are often used when the cost of performing many experiments is prohibitive or not feasible. We will also like to investigate a way to combine experiments and simulations creating a tool that can provide uncertainty estimates where experimental data is not available

Thank you!

Questions?