Learning with Analytical Models

Huda Ibeid, Siping Meng, **Oliver Dobon**, Luke Olson, William Gropp

University of Illinois at Urbana-Champaign

DOE/NNSA/ASC/PSAAPII:
The Center for Exascale Simulation of
Plasma-coupled Combustion

XPACC

Outline

1. Motivation
2. Applications
    1. Stencil Computation
    2. Fast Multipole Method
3. Analytical Models
4. Supervised Machine Learning
5. Approach: Hybrid Model
6. Evaluation
7. Conclusions

Motivation

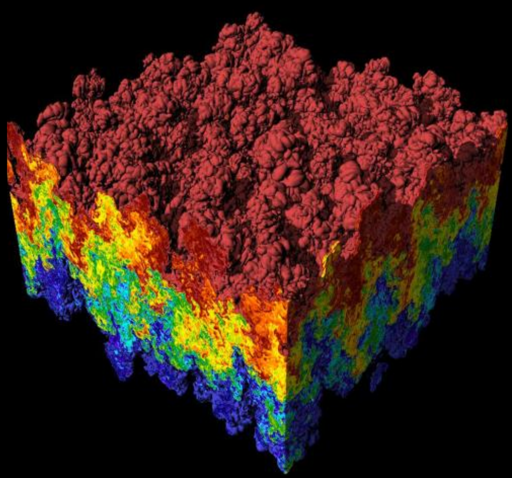
	Analytical Modeling	Machine Learning
Pros	• No or minimal training	• Requires minimum domain expertise
Cons	• Requires domain expertise • Rely on simplifying assumptions • Increasing architecture complexity	• Robustness • Curse of dimensionality

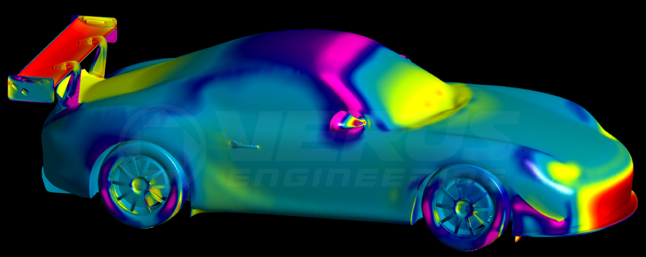
Goals:

- Minimize prediction cost
- Maintain reasonable prediction accuracy

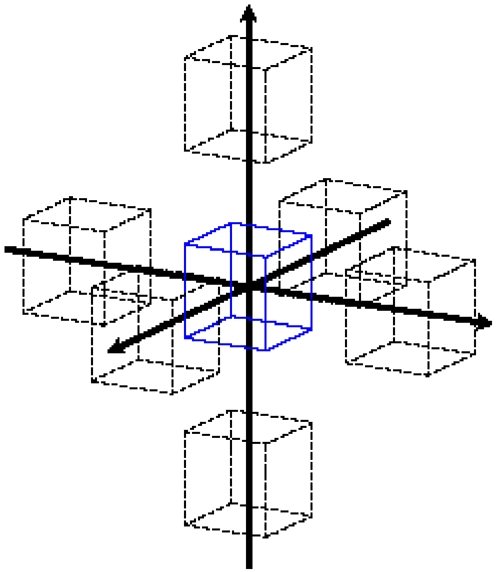
Applications

- Stencil Computation
- Fast Multipole Method





Stencil Computation



for $t \leftarrow 0$ **to** $timesteps$ **do**
  for $k \leftarrow 1$ **to** $KK - 1$ **do**
    for $j \leftarrow 1$ **to** $JJ - 1$ **do**
      for $i \leftarrow 1$ **to** $II - 1$ **do**
        $\chi^{t}_{i,j,k} = C_0 \times \chi^{t-1}_{i,j,k} + C_1 \times (\chi^{t-1}_{i-1,j,k} + \chi^{t-1}_{i+1,j,k} + \chi^{t-1}_{i,j-1,k} + \chi^{t-1}_{i,j+1,k} + \chi^{t-1}_{i,j,k-1} + \chi^{t-1}_{i,j,k+1})$
      end for
    end for
  end for
end for

Stencil Computation

Assumptions

- Arithmetic and memory operations can be overlapped
- Floating point operations negligible

Given a grid size: *N = I* x *J* x *K* elements of order l, total memory requirement to compute an X-Y plane

$$S_{total} = P_{read} \times S_{read} + P_{write} \times S_{write}$$

$$P_{read} = 2 \times l + 1$$

$$S_{read} = II \times JJ$$

$$P_{write} = 1$$

$$S_{write} = I \times J$$

Stencil Computation

On an architecture with a memory hierarchy of *n* cache levels, total time to compute a stencil is

$$T = T_{L1} + T_{Li} + \cdots + T_{Ln} + T_{mem}$$

$$T_{Li} = T_{Li}^{data} \times Hits_{Li}$$

$$T_{Li}^{data} = data * \beta_{\text{mem}_{Li}}$$

$$Hits_{Li} = Misses_{Li-1} - Misses_{Li}$$

$$Misses_{Li} = \lceil II/W \rceil \times JJ \times KK \times nplanes_{Li}$$

$$nplanes_{Li} = \begin{cases} 1, & \text{if } R_1 \\ (1, P_{read} - 1], & \text{if } \neg R_1 \wedge R_2 \\ (P_{read} - 1, P_{read}], & \text{if } \neg R_2 \wedge R_3 \\ (P_{read}, 2 \times P_{read} - 1], & \text{if } \neg R_3 \wedge \neg R_4 \\ 2 \times P_{read} - 1, & \text{if } R_4 \end{cases}$$

$R_1 : ((size_{Li}/W) \times R_{col} \geq S_{total})$,
$R_2 : ((size_{Li}/W) > S_{total})$,
$R_3 : ((size_{Li}/W) \times R_{col} > S_{read})$.
$R_4 : ((size_{Li}/W) \times R_{col} < P_{read} \times II)$

$$R_{col} = P_{read}/(2 \times P_{read} - 1)$$

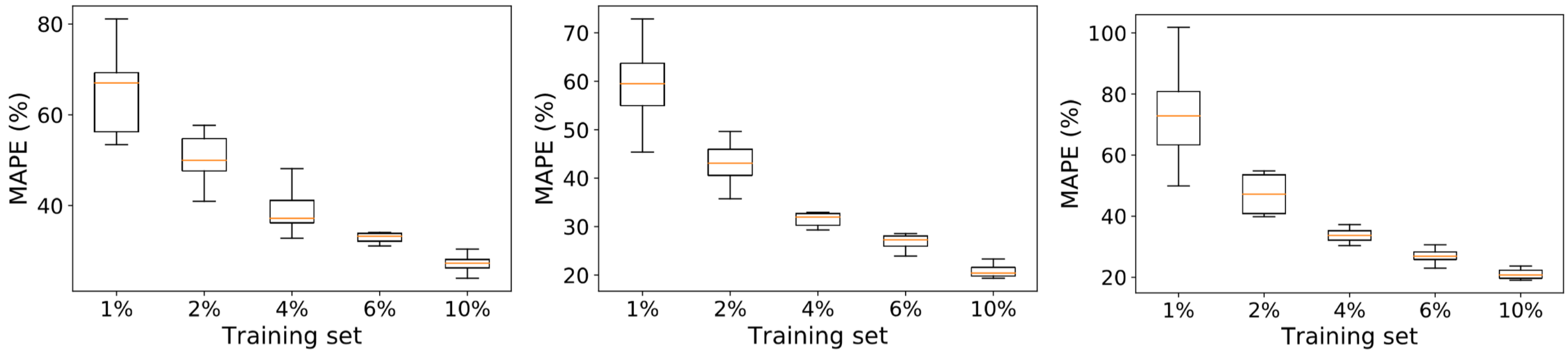
DOE/NNSA/ASC/PSAAPII:
The Center for Exascale Simulation of Plasma-coupled Combustion

XPACC

Supervised Machine Learning

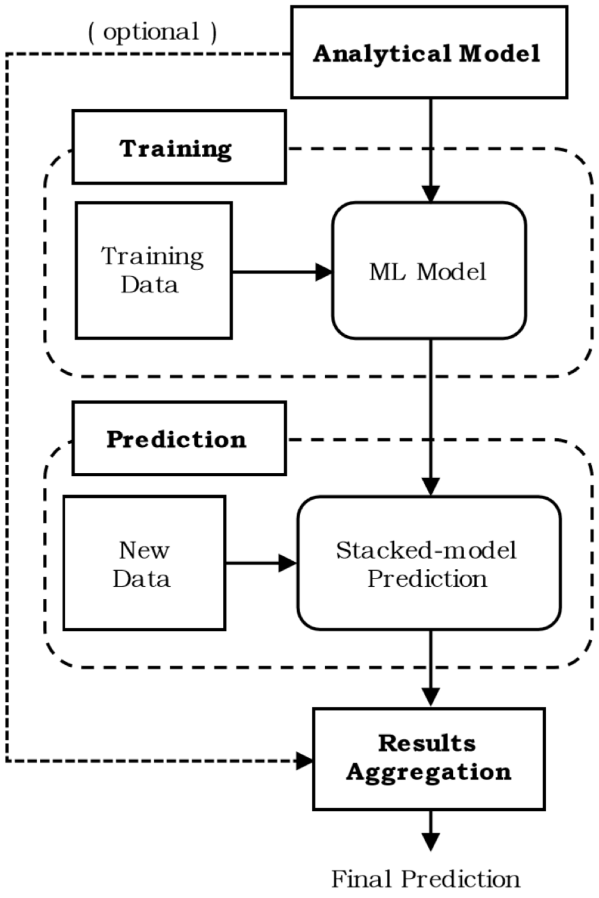
Stencil Computation

- $\boldsymbol{X} = (I, J, K, b_i, b_j, b_k)$ where $I \times J \times K = \{1 \times 16 \times 16 \ldots 1 \times 128 \times 128\}$ with a 16 points stride and $b_i \times b_j \times b_k = \{1 \times 1 \times 1 \ldots I \times J \times K\}$.



(a) Decision Trees (b) Extra Trees (c) Random Forests

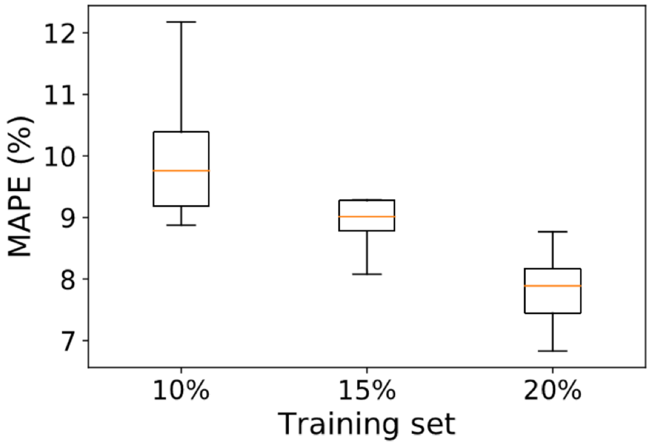
Hybrid Model



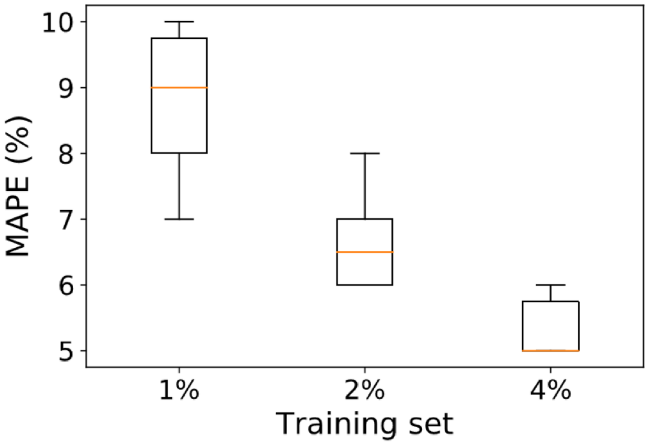
- Analytical model
- Two ensemble methods
- Training algorithm
- Prediction algorithm

Evaluation
Stencil Computation

- First, we evaluate hybrid approach on areas that analytical models cover accurately
- ***X*** = *(I, J, K)* where *I x J x K = {128x128x128 … 256x256x256}* with a 16 points stride



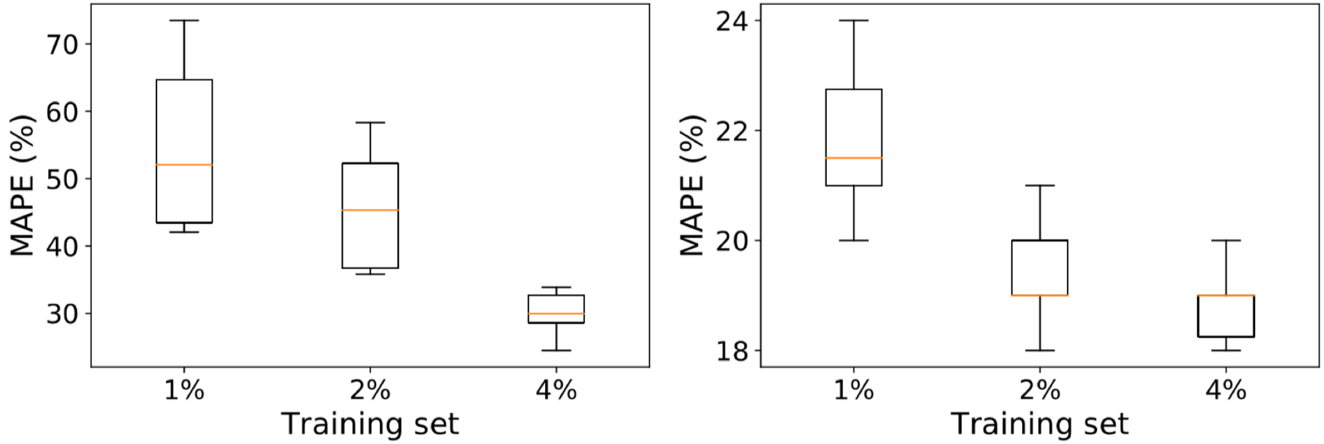
(a) Extra Trees



(b) Hybrid Model

Evaluation
Stencil Computation

- Next, we add loop blocking to the analytical models
- Analytical model *MAPE = 42%*
- $\boldsymbol{X} = (I, J, K, b_i, b_j, b_k)$ where $I \times J \times K = \{1 \times 16 \times 16 \ldots 1 \times 128 \times 128\}$ with a 16 points stride and $b_i \times b_j \times b_k = \{1 \times 1 \times 1 \ldots I \times J \times K\}$

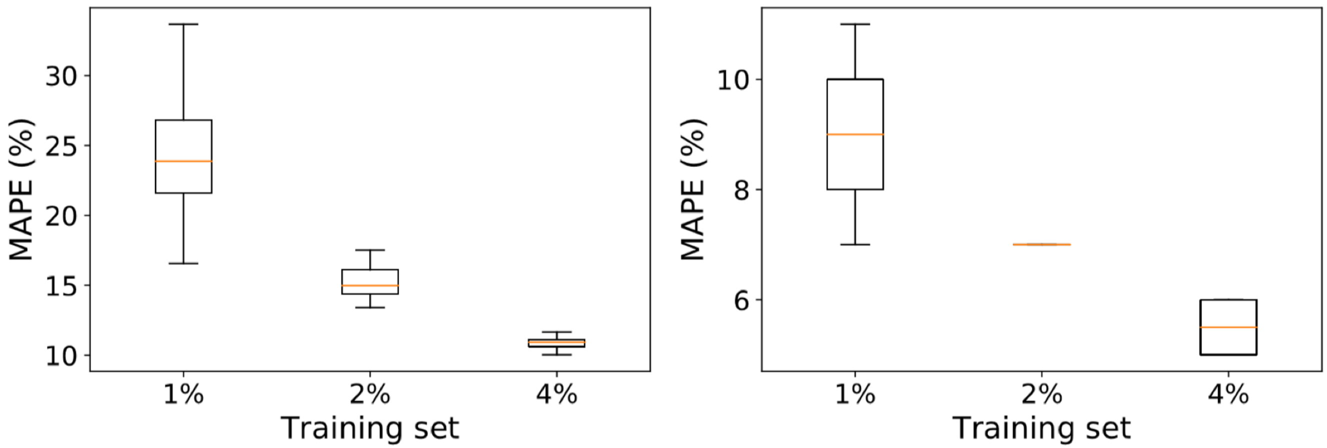


(a) Extra Trees (b) Hybrid Model

Evaluation

Stencil Computation

- Lastly, we evaluate the hybrid model on a region that is not covered by the analytical models
- ***X*** *= (I, J, K, t)* where *I x J x K = {128 x 128 x 1 … 176 x 176 x 1}* with a *16* points stride and the number of threads *t = {1 … 8}*

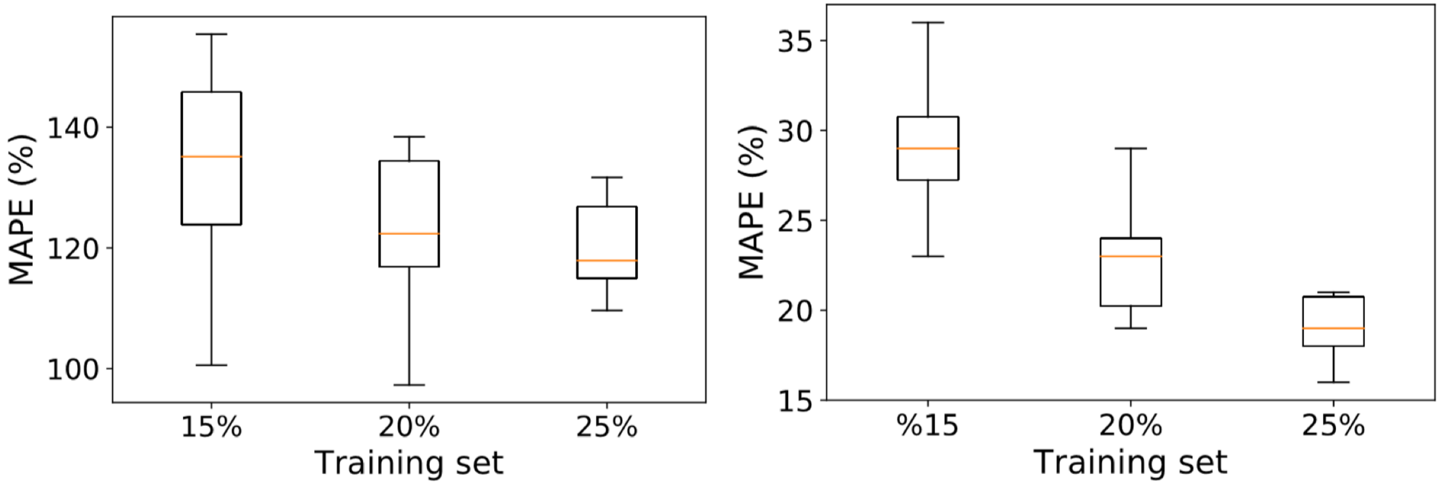


(a) Extra Trees (b) Hybrid Model

Evaluation

Fast Multipole Method

- FMM is a highly complex algorithm with several different phases, a combination of data structures, fast transforms, and irregular data access
- We do not tune the analytical models (*MAPE = 84:5%*)
- ***X*** *= (t, N, q, k)* where *t = {1 … 16}*, *N = {4096, 8192, 16384}*, and *k = {2 … 12}*



(a) Extra Trees (b) Hybrid Model

Conclusions

- The hybrid approach is effective in predicting the execution time by reducing the *MAPE* score of pure machine learning models.
- The hybrid model requires small training dataset to carry out predictions with reasonable accuracy, thus making it suitable for hardware and workload changes.