Machine Learning Panel

S. Girimaji, B. Moser, P. Cinnella, A. Banko, R. Dwight, K. Duraisamy

~~Machine Learning~~ Panel

S. Girimaji, B. Moser, P. Cinnella, A. Banko, R. Dwight, K. Duraisamy

Data-driven Turbulence Modeling Panel

S. Girimaji, B. Moser, P. Cinnella, A. Banko, R. Dwight, K. Duraisamy

Historical perspective of data-driven Turbulence modeling (rough & biased towards supervised learning)

Early 2000s Bayesian Calibration of physics models (e.g. Kennedy O' Hagan)
Mid 2000s Dawn of Uncertainty Quantification in Physical Sciences
2011 *Oliver & Moser* :Gaussian random field (GRF) for Reynolds stress discrepancy
2011 *Dow & Wang* : Augmenting eddy-viscosity with GRF, from multiple cases

UQ Era

==

2013 *Tracey et al.* : Supervised learning of Reynolds stress perturbations; Transformation of model discrepancies from the spatial domain to feature space
2014 *Vollant et al*. Optimal estimation setting for subgrid scale modeling for LES.
2014-2017 *Duraisamy et al.* : Model consistent learning (FIML, etc)
2015-2017 *Xiao et al., Ling et al., Weatheritt et al.* : Detailed work on function representation, invariances
2017 *Symposium 1*

First wave

==

2017 Applications have expanded to LES, multiphase flows, and combustion modeling
2019 APS DFD Number of talks in focus session in ML for Fluid Dynamics
2021 Model consistency widely being adopted + new ideas.
2021 Symposium 2

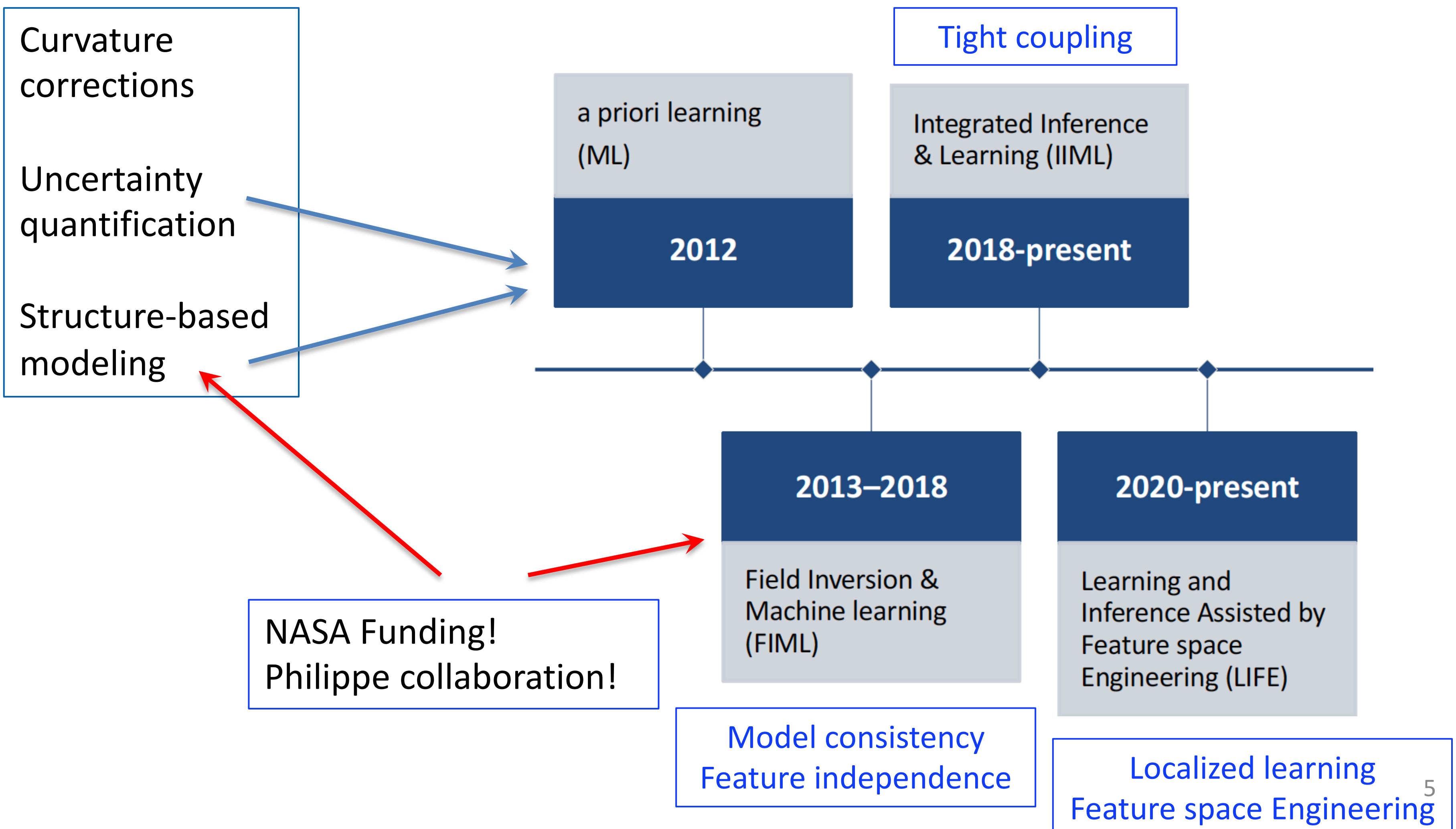
Second wave

==

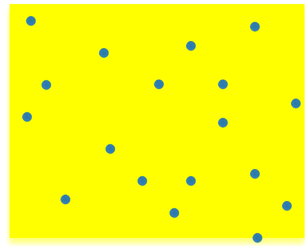
2022 O(100) researchers in the field
2022 Thinking about Do-no-harm/Generalizability/Benchmarking/Reproducability
2022 Symposium 3

Third wave

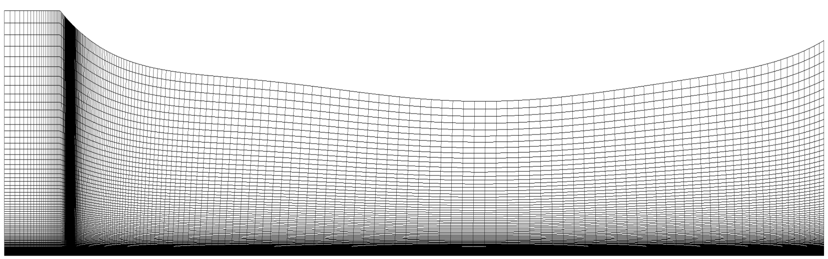
A personal timeline on data-driven turbulence modeling + NASA / Philippe relevance

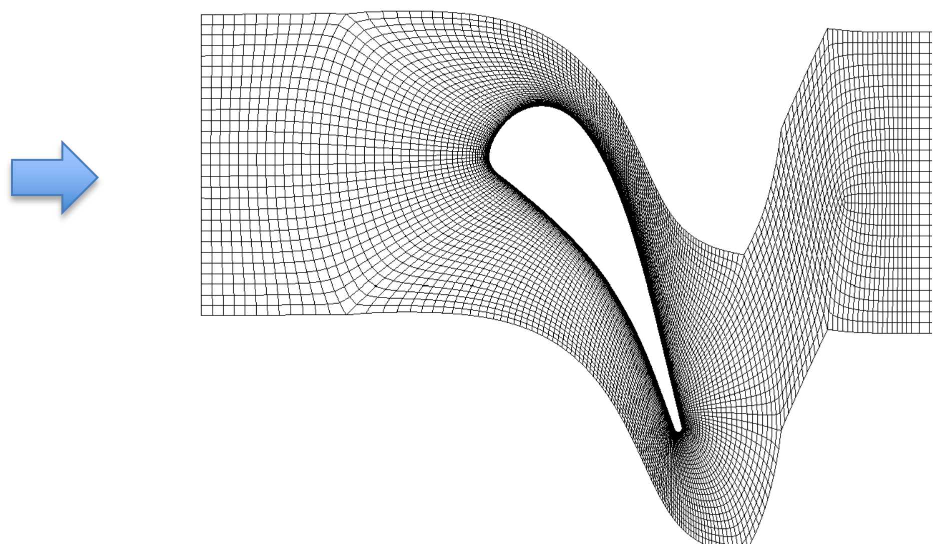


The hope

$R(\bar{u}, \beta($  $)) = 0$

Physics based model +
Interpolation in feature space =
Extrapolation in physical space





Model inadequacies

$$\frac{\partial \overline{u_i' u_j'}}{\partial t} = C_{ij} + P_{ij} + V_{ij} + T_{ij} + \Pi_{ij} + D_{ij}$$

$$\frac{\partial \overline{u_i' u_j'}}{\partial t} = C_{ij} + P_{ij} + V_{ij} + \tilde{T}_{ij} + \delta_T + \tilde{\Pi}_{ij} + \delta_\Pi + \tilde{D}_{ij} + \delta_D$$

$$\frac{\partial \overline{u_i' u_j'}}{\partial t} = C_{ij} + P_{ij} + V_{ij} + \tilde{T}_{ij} + \tilde{\Pi}_{ij} + \tilde{D}_{ij} + \delta_{ij}$$

From our NASA proposal in 2013:

- *"One - seven transport eqns, and up to 30 adjustable constants.*
- *Modeling rests on large amounts of intuition and luck, in spite of starting with a "rigorous" approach*
- *Theories abound for parts of model, but not for output*
- *Model constants calibrated on very limited data*
- *Greater sophistication in RANS models, with mixed degree of success*

  ➔ *More constants to fit , still use canonical problems"*

Model inadequacies

$$\frac{\partial \overline{u_i' u_j'}}{\partial t} = C_{ij} + P_{ij} + V_{ij} + T_{ij} + \Pi_{ij} + D_{ij}$$

$$\frac{\partial \overline{u_i' u_j'}}{\partial t} = C_{ij} + P_{ij} + V_{ij} + \tilde{T}_{ij} + \delta_T + \tilde{\Pi}_{ij} + \delta_\Pi + \tilde{D}_{ij} + \delta_D$$

$$\frac{\partial \overline{u_i' u_j'}}{\partial t} = C_{ij} + P_{ij} + V_{ij} + \tilde{T}_{ij} + \tilde{\Pi}_{ij} + \tilde{D}_{ij} + \delta_{ij}$$

$$\equiv C_{ij} + \beta_{ij} P_{ij} + V_{ij} + \tilde{T}_{ij} + \tilde{\Pi}_{ij} + \tilde{D}_{ij}$$

$$\equiv C_{ij} + P_{ij} + V_{ij} + \tilde{T}_{ij} + \tilde{\Pi}_{ij} + \tilde{D}_{ij} + (\beta_{ij} - 1) P_{ij}$$

$$\delta_{ij} \equiv (\beta_{ij} - 1) P_{ij}$$

Discrepancies we modeled in the original NASA project

$$\frac{D\omega}{Dt} = P_\omega - \beta(x) D_\omega + T_\omega$$

Singh & Duraisamy, PoF 2016

$$g(r) = r + \boldsymbol{\beta}(\mathbf{x}) c_{w2}(r^6 - r)$$

Singh, Pan & Duraisamy, Aviation 2017

$$\frac{DR_{ij}}{Dt} = C_{ij} + P_{ij} + T_{ij} + \Pi_{ij} + D_{ij} + \beta(x)_{ij}\epsilon_{ij}$$

Parish & Duraisamy, Aviation 2014

$$\frac{DR_{ij}}{Dt} = \beta(x)_{ij} a_o \omega(R_{ij,eq} - R_{ij})$$

Singh & Duraisamy, Scitech 2016

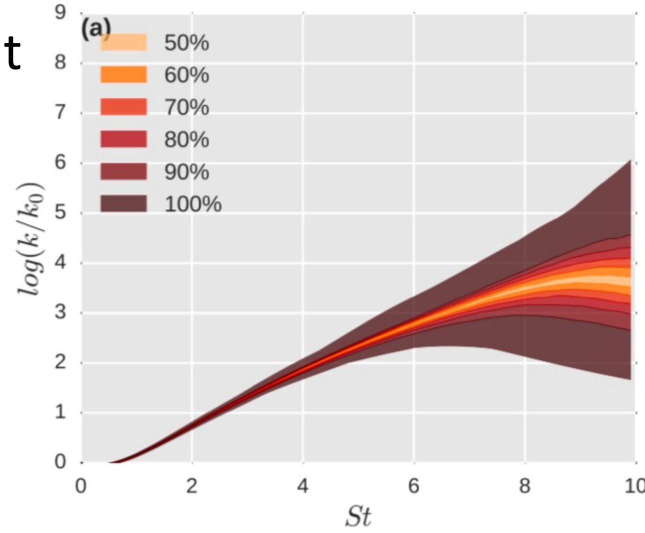
$$\mathbf{R}_p = 2k\left[\frac{\mathbf{I}}{3} + \mathbf{V}(\Lambda + \vec{\beta}(x))\mathbf{V}^T\right]$$

Duraisamy, SIAM 2016

"Levels" of uncertainty in RANS (and LES) models

L1: uncertainties introduced by ensemble averaging that are fundamentally irrecoverable

→ At a given instant in time, there are infinitely many realizations of velocity fields that are compatible with an averaged field ; however, each of these realizations might evolve dynamically in different ways



L2: uncertainties in the functional and operational representation of Reynolds stress

→ Example: Single point closure assumption, Linear eddy viscosity models, algebraic stress models, etc.

L3: uncertainties in functional forms within a model

→ Example: Pressure strain model, rotational correction, etc.

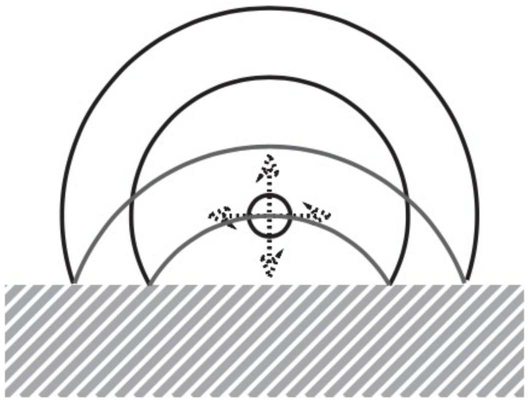
L4: uncertainties in the coefficients within a model

→ Example: C_μ, κ, etc.

Challenges

- Elegant ideas/rigorous theory get obscured in a practical model

- How to make the leap from
    - Homogeneous -> Inhomogeneous
    - Equilibrium -> non-equilibrium
    - Rapid distortion theory -> slower distortion
    - Canonical -> practical



$$p(\boldsymbol{x}) = \frac{1}{2\pi} \iiint_{-\infty}^{\infty} \frac{\rho\, \partial_l U_k\, \partial_k' u_l(x', z', |y'|)}{|\boldsymbol{x} - \boldsymbol{x}'|} \mathrm{d}^3 \boldsymbol{x}'$$

$$L^2 \nabla^2 f_{ij} - f_{ij} = -\frac{\wp_{ij}^{\mathrm{h}} + \varepsilon b_{ij}}{k}$$

$$L = \max\left\{ c_{\mathrm{L}} \frac{k^{3/2}}{\varepsilon},\; c_\eta \left(\frac{\nu^3}{\varepsilon} \right)^{1/4} \right\}$$

Challenges

- Elegant ideas/rigorous theory get obscured in a practical model

- How to make the leap from
  - Homogeneous -> Inhomogeneous
  - Equilibrium -> non-equilibrium
  - Rapid distortion theory -> mixed distortion
  - Canonical -> practical

$$\tau_m = 2k_m \left(\frac{1}{3} I + \left[\sum_{n=1}^{10} \delta_m^{(n)} (\tilde{\eta}_m ; w) T^{(n)} (\tilde{S}_m, \tilde{\Omega}_m) \right] \right)$$

Reasonable when

- $3 < Sk/\varepsilon < 6$

AND

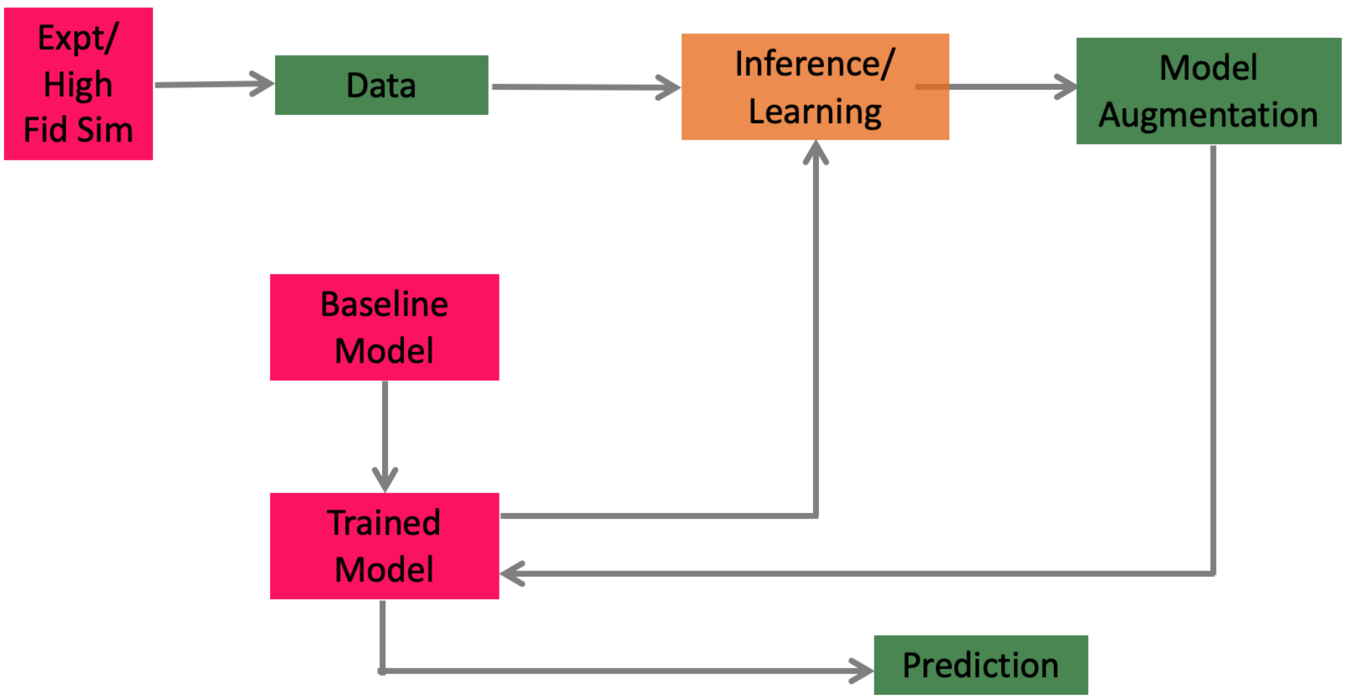
- The state can be completely described by only S, Ω

AND

- Homogeneous turbulence!

Challenges

- Many "seemingly physical" quantities are just operational variables
  - ➔ Use of *apriori* analysis and direct learning is of limited utility
  - ➔ Model consistency will fix some of these issues (AND remove need for fields of DNS data)

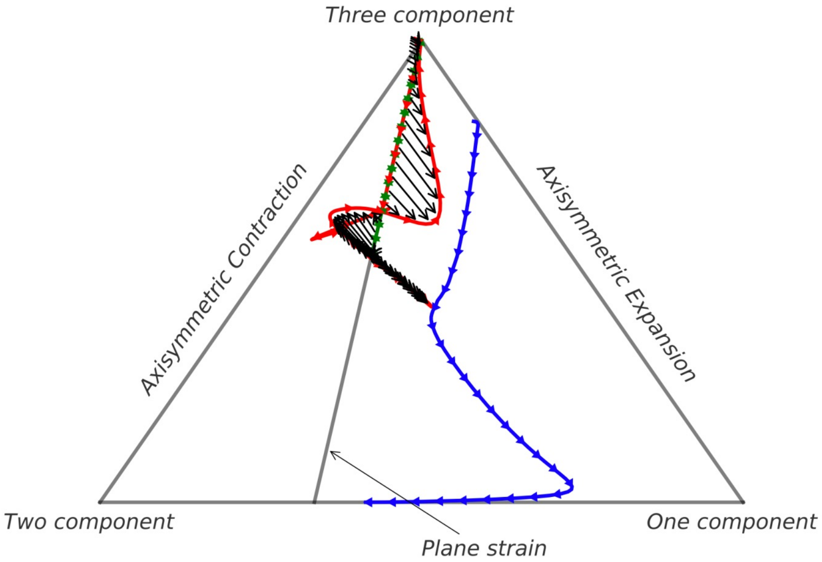


- "More physics" is not necessarily better for modeling
  - ➔ "Right physics" is better

Challenges

“ The central role of creativity and free intuition introduces a danger of proliferation. Any type of new term can be proposed, and many will satisfy the consensus constraints such as Galilean invariance, so that rejecting them becomes a matter of opposing intuition.” - Spalart

- Incomplete Data
- Convergence
- Irrecoverable model discrepancies
- Identifiability
- Generalizability
- Interpretability
- Input & Output constraints



Human input and intuition is irreplaceable

- Elegant ideas/rigorous theory get obscured in a practical model
- Leaps are hard
- "More physics" is not necessarily better for modeling, given the task at hand
- Difficult to separate problem-specific information from "global rules"
- Data is incomplete

What is the goal?

Where to augment?

- What is the intended correction?
- How will the augmentation affect cases where correction is not required?

What are the relevant features?

- Based on Expert Knowledge
- Should not be specific to a problem
- Does the available data span the feature space sufficiently?

What is the function class?

- Non-linear with unknown structure:
  - Neural Networks
  - Gaussian Processes
  - Decision Trees, etc.
- Custom-built based on some knowledge of the structure
- Hybrid – Non-linear with constraints

How to optimize?

- Gradients/no gradients?
- How to weight the data sets?
- Implicit and explicit constraints

"Data driven Modeling" is more appropriate than "Machine learning"

➔ ML is an optional step (and utility is overblown)

Simple things to do

- Dissonance between what researchers in DDTM have actually been doing and what the community thought they have been doing (+ overselling)

    ➔ Set goals and expectations

- Try to use common terminology

- Let’s show bad results (things that didn’t work)

Some topics to discuss

- Do we even have the right descriptors to have a chance at succeeding?
  - ➔ e.g. do we need structure tensors?
- How can we isolate / combine the impact of different phenomena (e.g. separation, secondary flows, pressure gradients, curvature) in model construction?
- How to identify the right set of (hard and soft) constraints that should be satisfied? (Cannot be under- or over-constrained)
  - ➔ Philippe & George gave some good suggestions
- Coordinating high fidelity simulations and experiments with model developments
- Let’s not forget uncertainty quantification !
- How can the community work together more cohesively? How can NASA lead the charge?

Backup

Some discussion topics

- Setting expectations
- How can we isolate / combine the impact of different phenomena (e.g. separation, secondary flows, pressure gradients, curvature) in model construction?
- How to identify the right set of physical constraints that should be satisfied? (Cannot be under- or over-constrained)
- Have we converged on a set of procedures for Data-driven turbulence modeling?
- Should we target model improvements for a class of problems or should we think about more general models?
- Accounting for sparse/incomplete/noisy data
- Quantifying uncertainties
- Importance of model consistency
- Fundamental limits of Machine learning
- How to account for Irrecoverable errors in RANS model forms
- Identifiability, Generalizability, interpretability
- Feature selection
- Impact of Machine learning algorithm
- How can the community work together more cohesively? How can NASA lead the charge?

Opinions

- Model consistency is paramount

  ➔ Also allows for the use of sparse / indirect data (e.g. from experiment. We also don't need fields of DNS data)

  ➔ Well-recognized by now ("field inversion", "integrated inference", "CFD-driven ML", "Iterative ML", etc.)

- "Data driven" is more appropriate than "Machine learning"

  ➔ ML is an optional step (and utility is overblown)

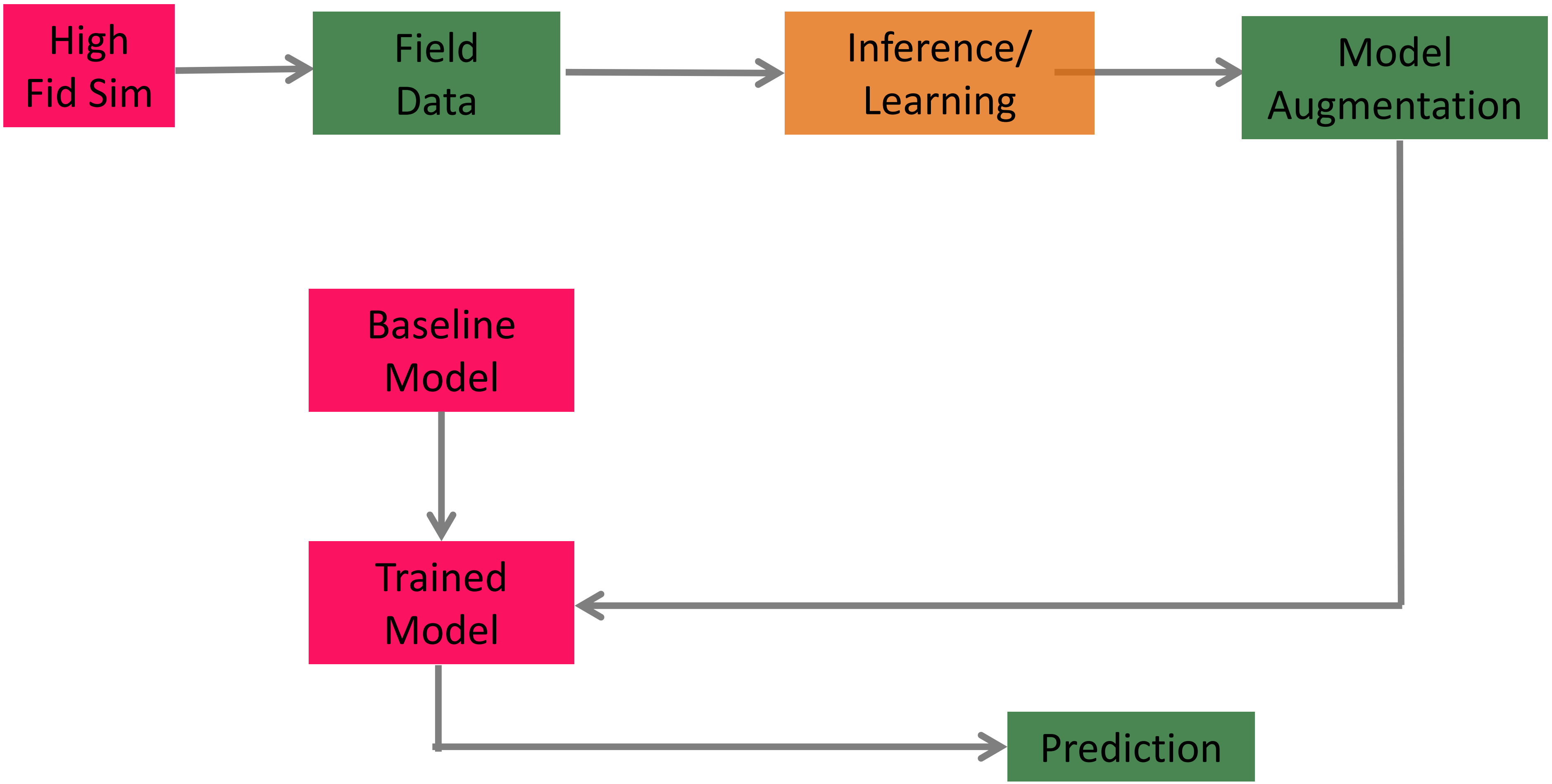
- Very Personal opinion : There is no (and will ever be a) universal *single-point closure model* waiting to be discovered

  ➔ Optimal model, conditional on data and assumptions possible

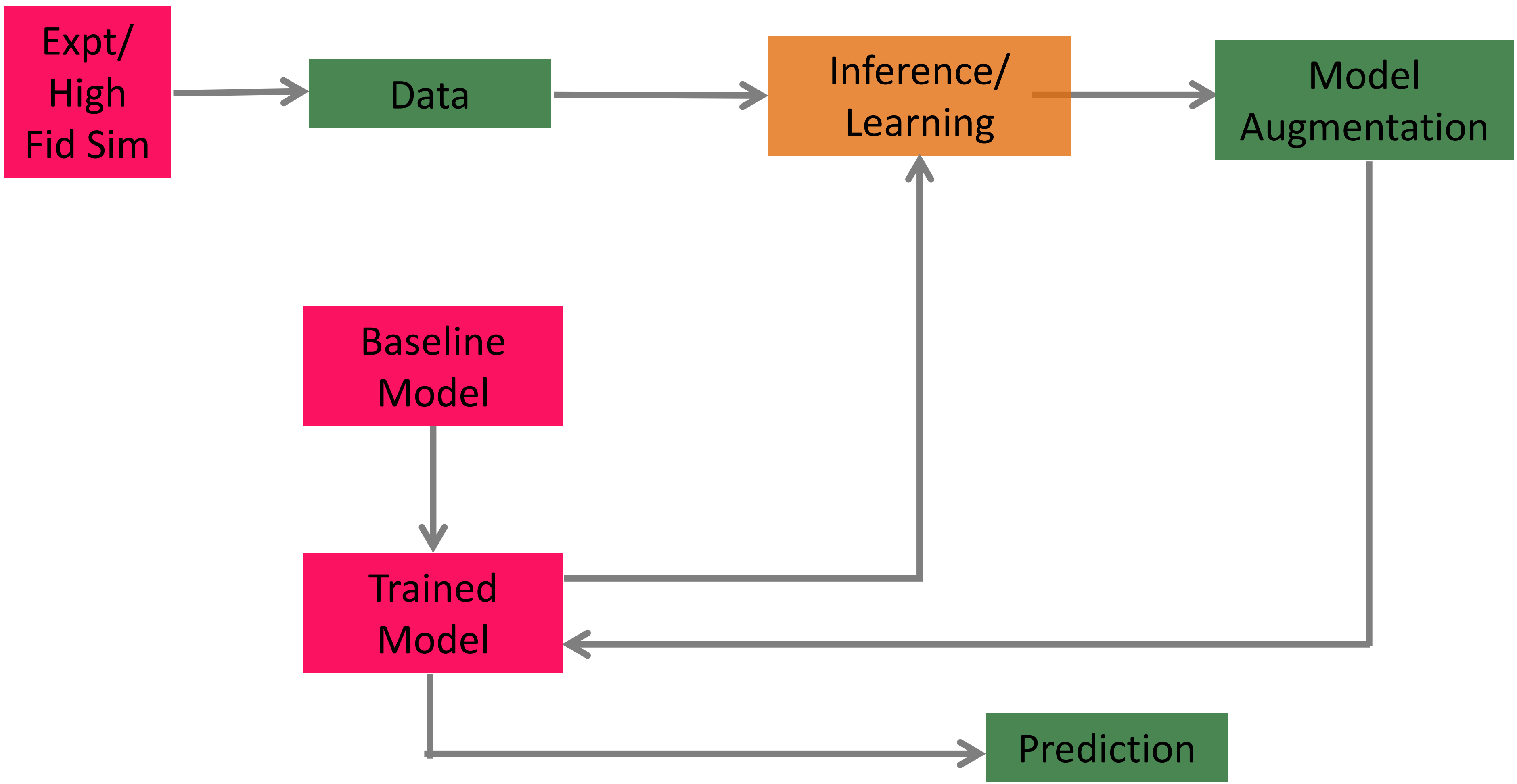
- Data-driven approach is not a substitute to turbulence modeling. It is just a new tool

- True impact in developing generalizable turbulence models requires coordinated/long term research

A priori learning Framework



Model Consistent Framework



A comprehensive approach