

科学机器学习简介

黄政宇

北京大学北京国际数学研究中心
北京大学国际机器学习研究中心



科学计算

➤ 建模仿真

- 理解复杂系统
- 预测复杂系统
-

➤ 工程优化

- 工程系统设计
- 工程系统实时控制
- 量化不确定性、可靠性评估
- 寻找给定问题的最佳解决方案或配置
-

挑战：运算量巨大难以承受

- 多尺度多物理场建模
- 反复求解、迭代
- 需要实时反馈
-



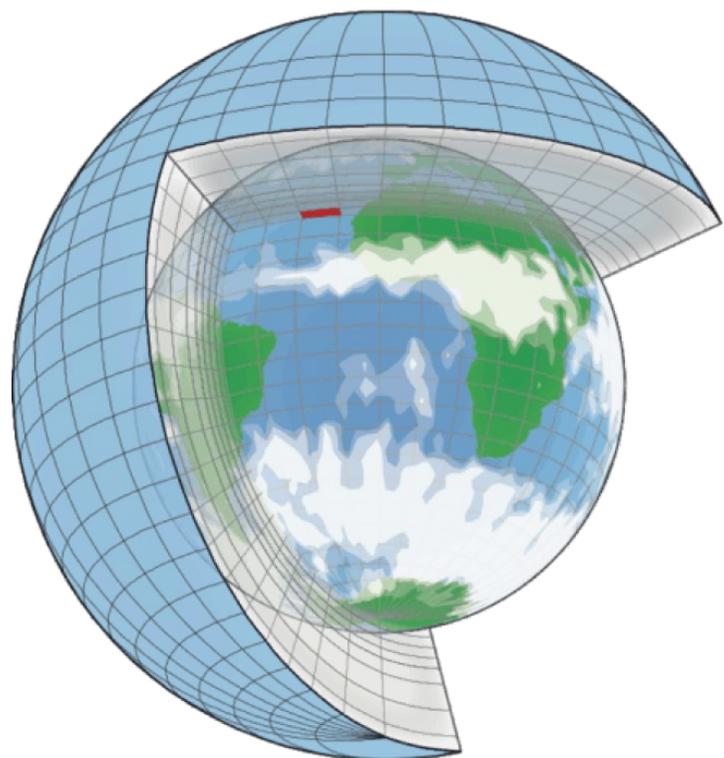
科学计算中的挑战

- 运算量巨大难以承受
 - 多尺度多物理场建模
 - 反复求解、迭代
 - 需要实时反馈
 -



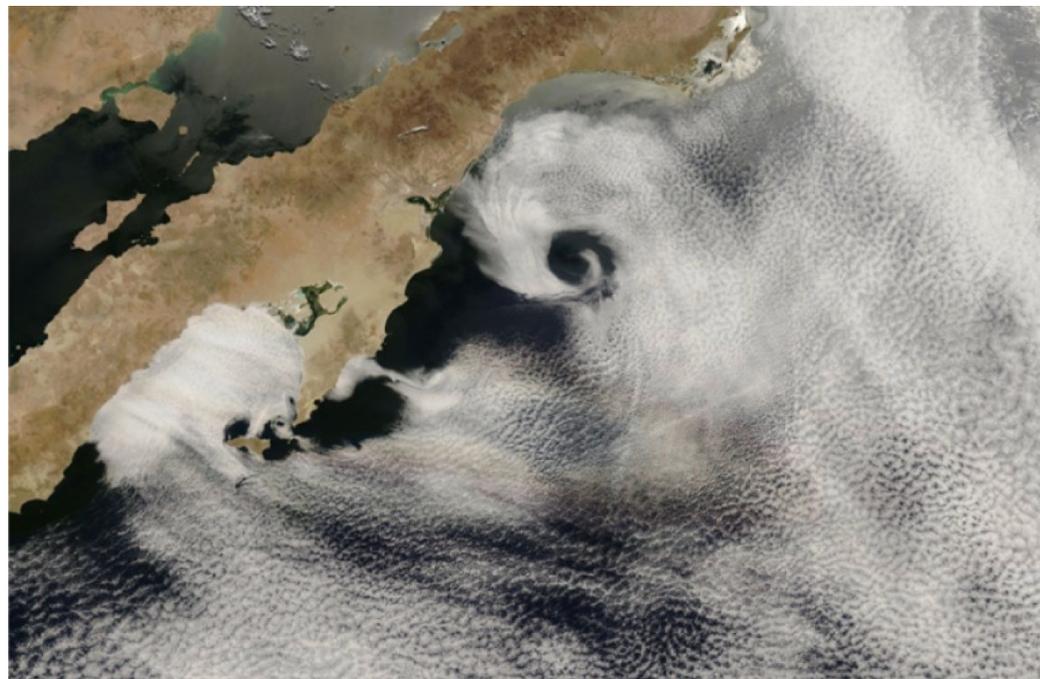
多尺度多物理场建模

➤ 气候模拟



分辨率：10km-50km

云层



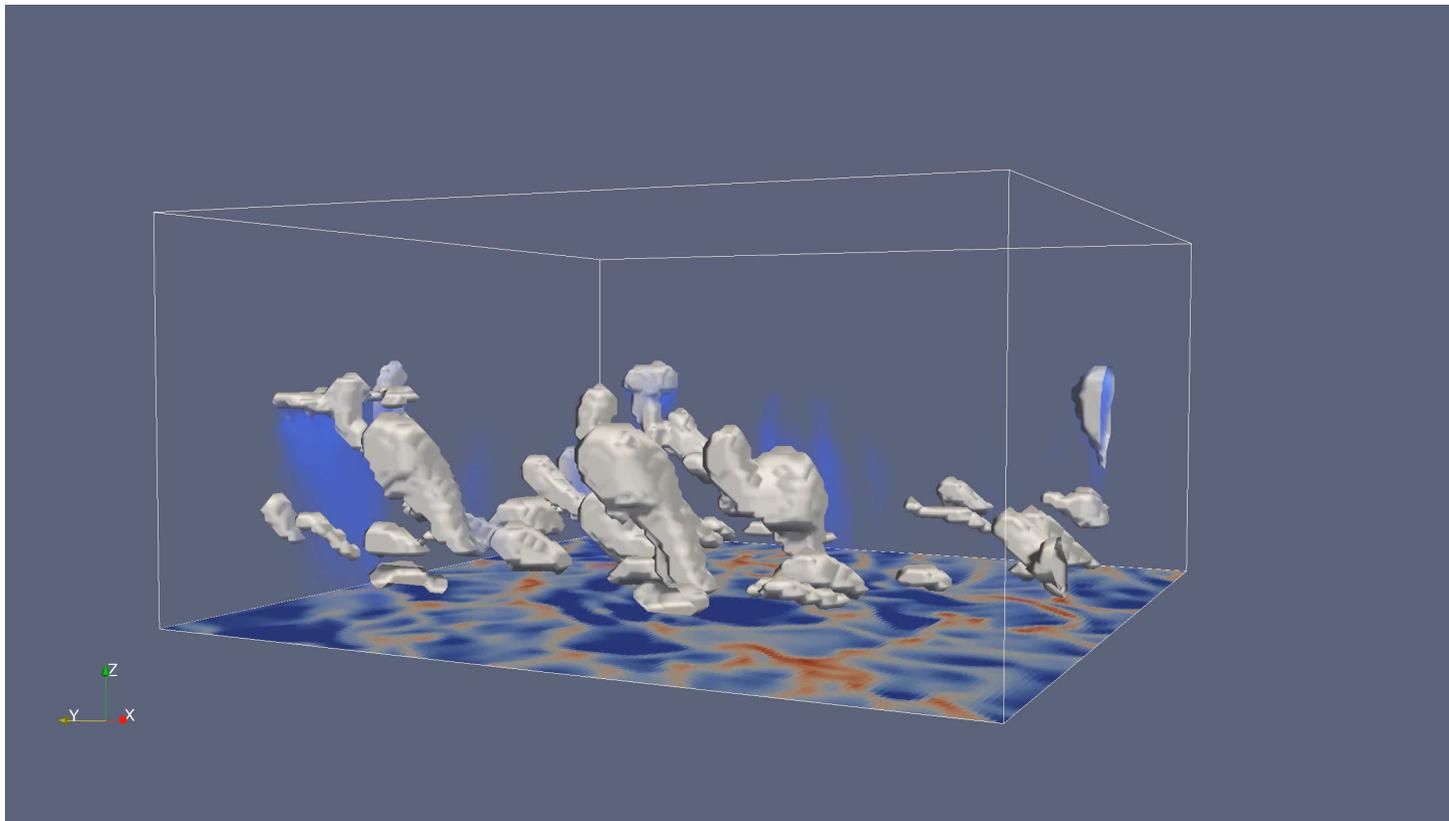
尺度：10m-100m



多尺度多物理场建模

➤ 气候模拟

- 云层





多尺度多物理场建模

➤ 可压缩Navier-Stokes 方程

$$\frac{\partial W}{\partial t} + \nabla F(W) = \nabla R(V, \nabla V) + S$$

$$V = \begin{pmatrix} \rho \\ \mathbf{v} \\ p \end{pmatrix} \quad W = \begin{pmatrix} \rho \\ \rho \mathbf{v} \\ E \end{pmatrix}$$

$$F(W) = \begin{pmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \otimes \mathbf{v} + p \mathbf{I} \\ (E + p) \mathbf{v} \end{pmatrix} \quad R(V, \nabla V) = \begin{pmatrix} 0 \\ \tau \\ \tau \cdot \mathbf{v} - q \end{pmatrix}$$

质量守恒、动量守恒和能量守恒，以及牛顿流体假设、傅里叶热传导定律……



多尺度多物理场建模

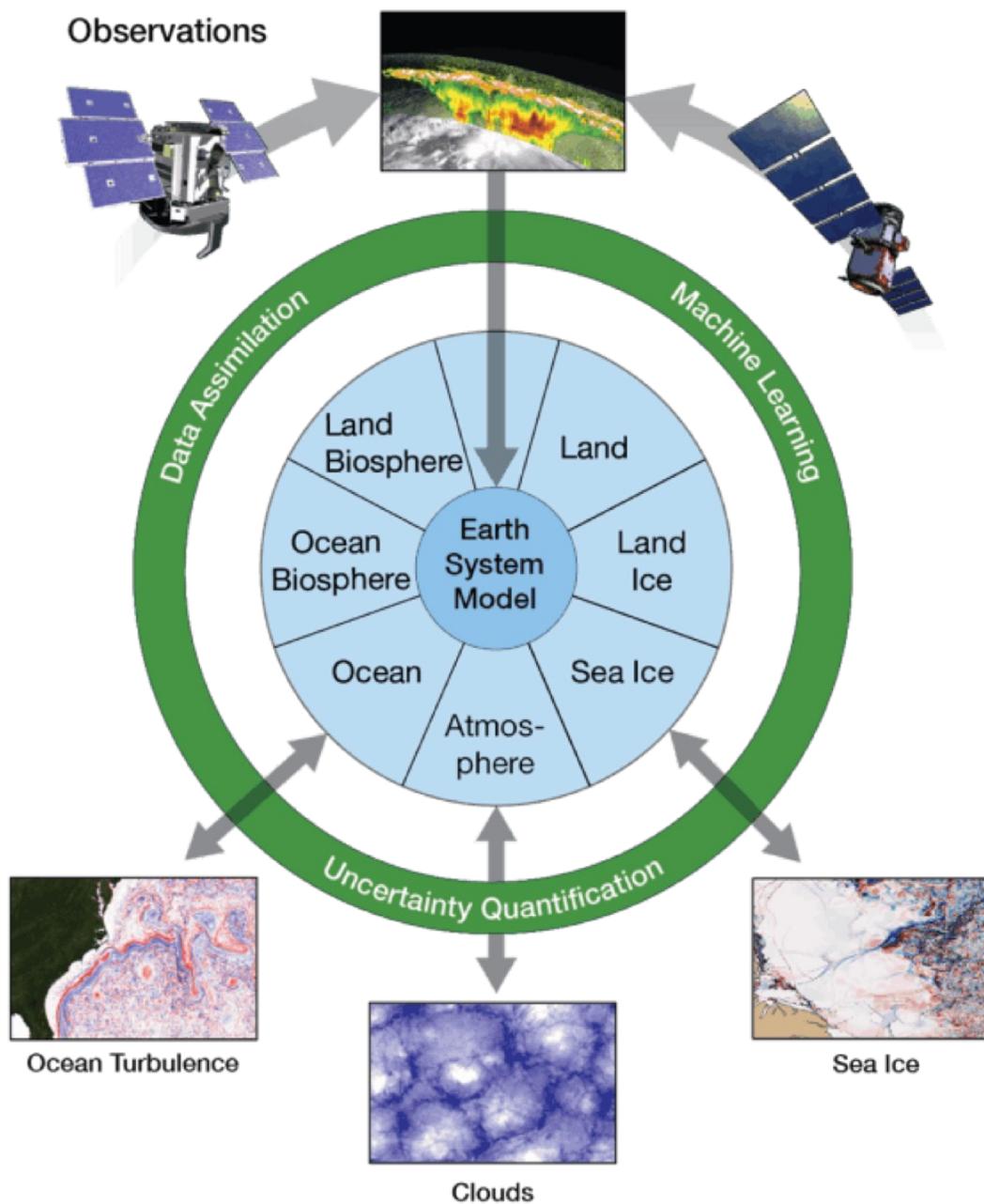




多尺度多物理场建模

➤ 气候模拟

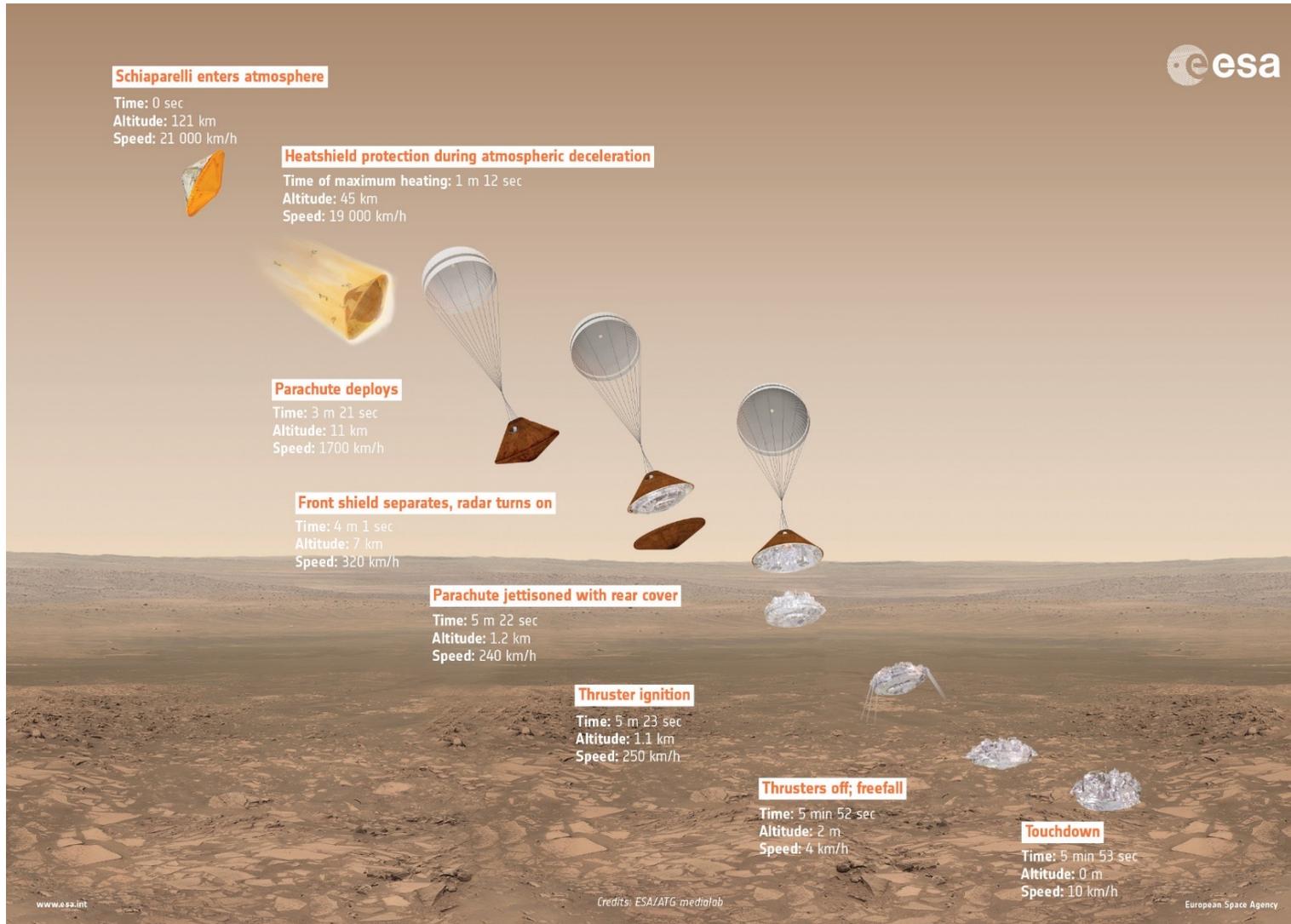
- 云层
- 冰川
- 植被
- 海洋、大气湍流
-





多尺度多物理场建模

➤ 航空航天





多尺度多物理场建模

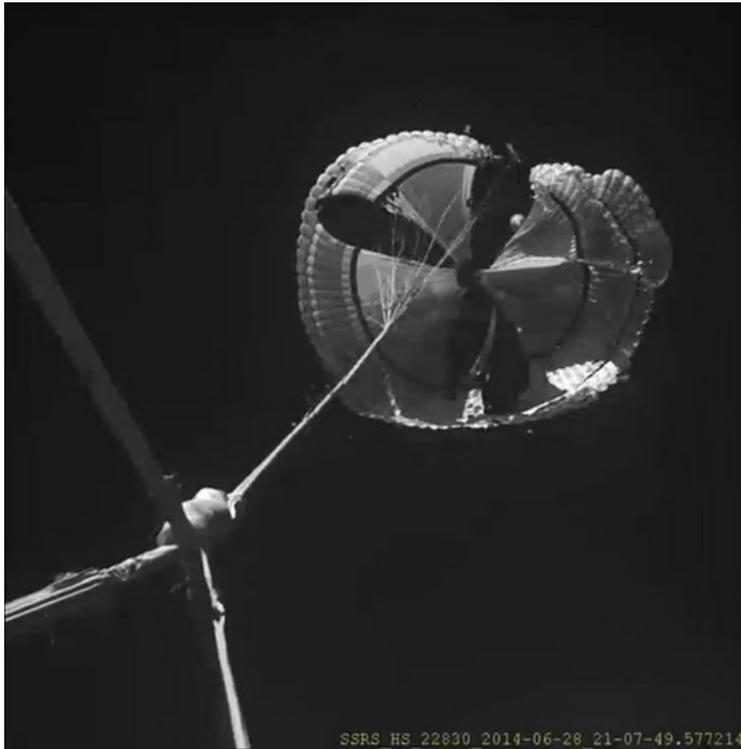
➤ 航空航天



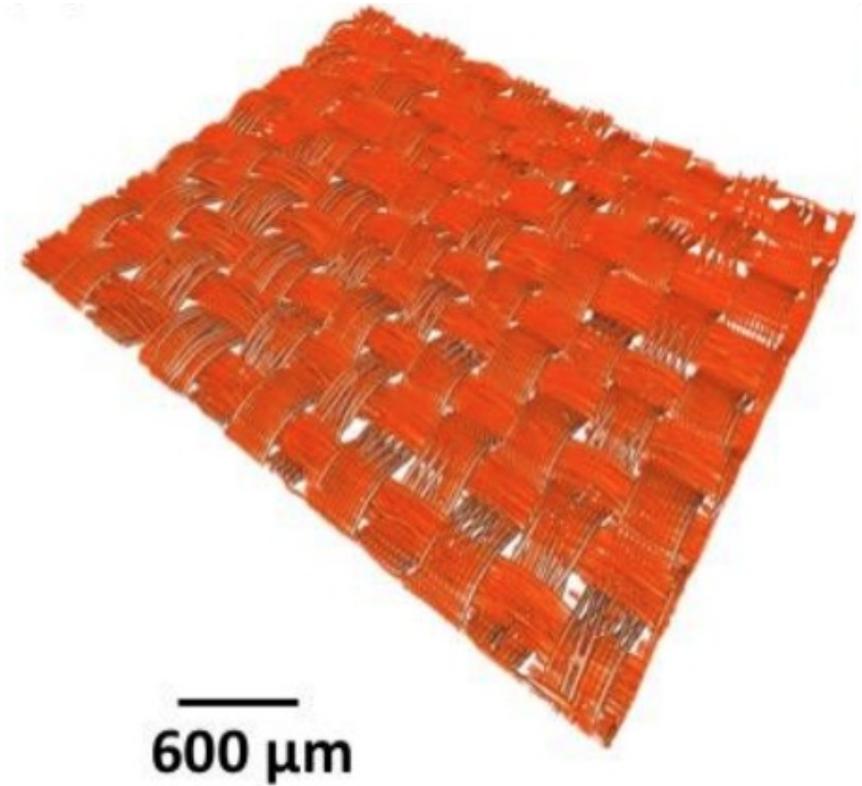


多尺度多物理场建模

➤ 航空航天



尺度：1m-10m



尺度：10 μm -100 μm



多尺度多物理场建模

➤ 固体方程

$$\rho_s \ddot{u}_s - \nabla \cdot \sigma = f_s$$

$$\sigma \cdot n = t \quad \text{on } \Gamma^t$$

$$u_s = u_s^D \quad \text{on } \Gamma^D$$

➤ 本构关系

$$\sigma = \mathcal{M}(\epsilon) \quad \epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$$



多尺度多物理场建模





科学计算中的挑战

- 运算量巨大难以承受
 - 多尺度多物理场建模
 - 反复求解、迭代
 - 需要实时反馈
 -



反复求解、迭代

➤ 航空航天

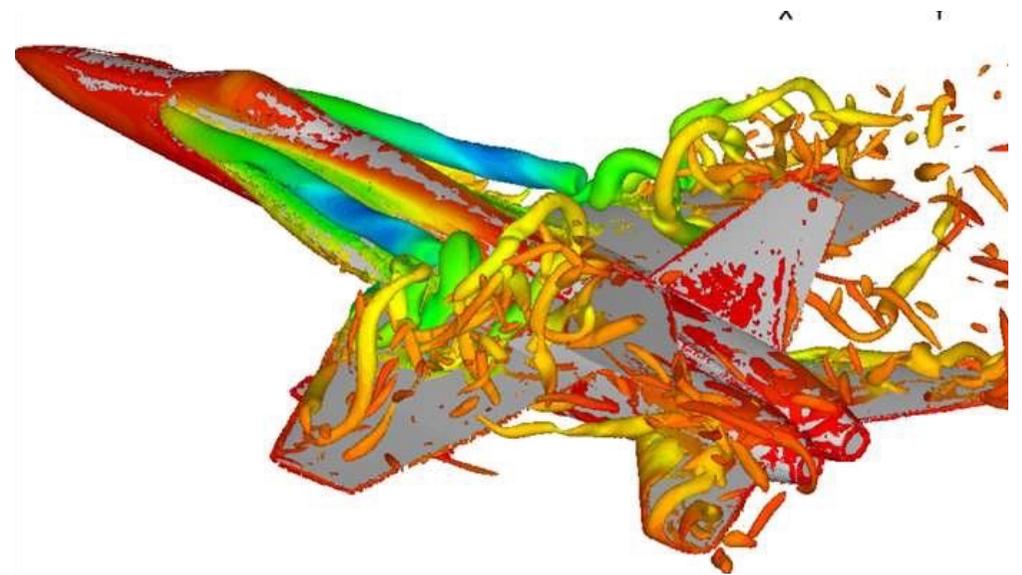


1984，McDonnell Douglas 公司宣布，将承担修正F-18战斗机设计缺陷的费用，大约为2500万美元。该设计缺陷已导致战斗机尾部出现裂缝。



反复求解、迭代

➤ 航空航天





多尺度多物理场建模

➤ 流固耦合方程

$$\frac{\partial W}{\partial t} + \nabla F(W) = \nabla R(V, \nabla V)$$
$$\rho_s \ddot{u}_s - \nabla \cdot \sigma = f_s$$

流固耦合边界：

$$\sigma \cdot n = -pn + \tau \cdot n$$

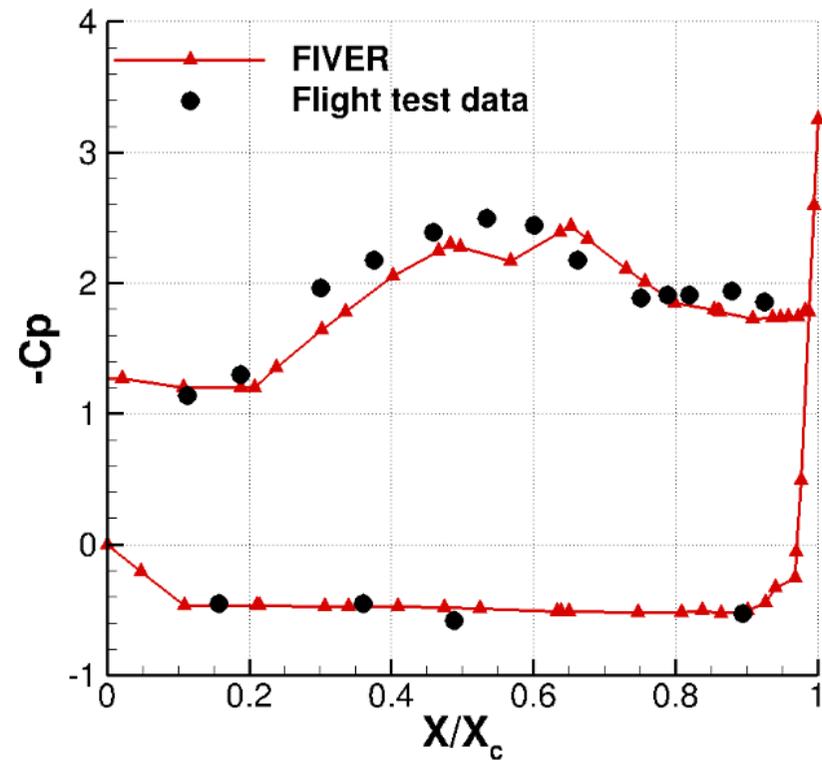
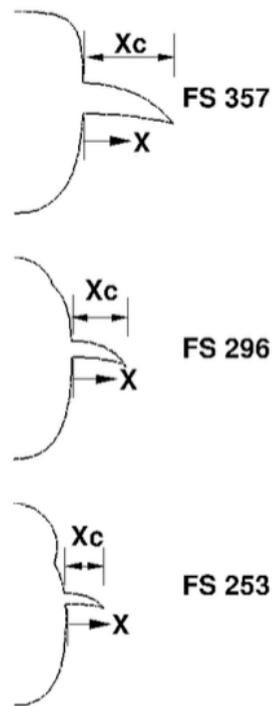
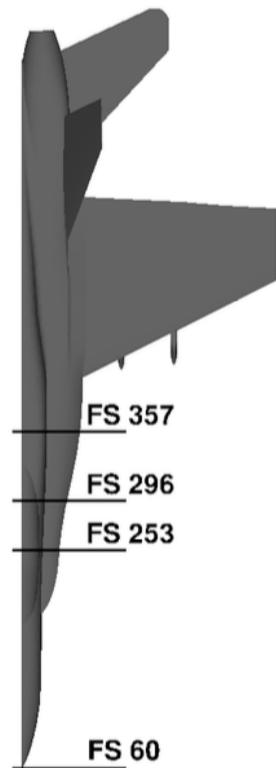
$$\dot{u}_s = v$$



反复求解、迭代

➤ 航空航天

$M_\infty = 0.243, \text{AoA} = 30.3^\circ$



4天 1000CPU_s



反复求解、迭代

➤ 不确定性量化

- 飞行速度、角度
- 飞行高度
- 战机设计参数
-

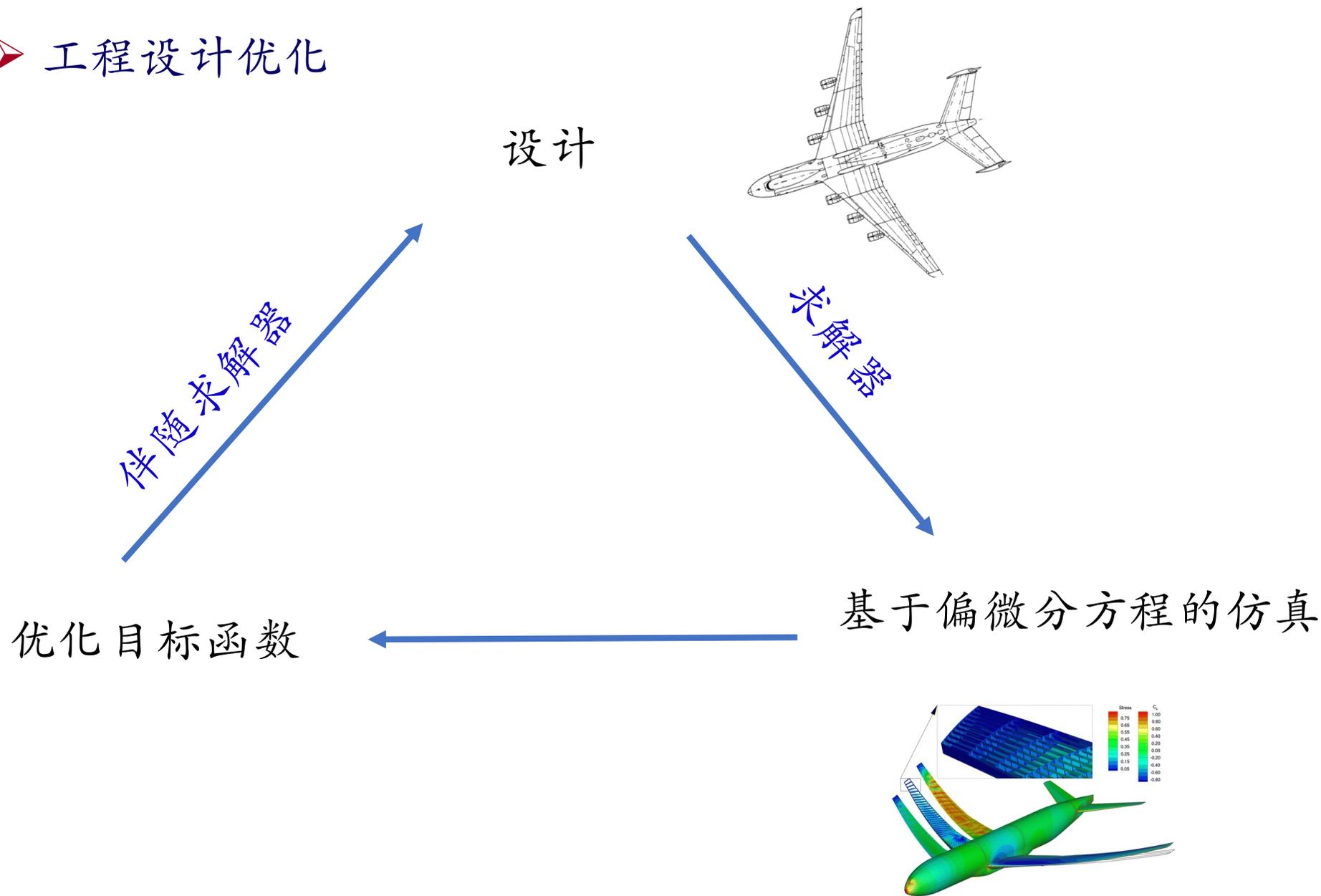


≥1000 仿真



反复求解、迭代

➤ 工程设计优化





多尺度多物理场建模





科学计算中的挑战

- 运算量巨大难以承受
 - 多尺度多物理场建模
 - 反复求解、迭代
 - 需要实时反馈
 -



需要实时反馈

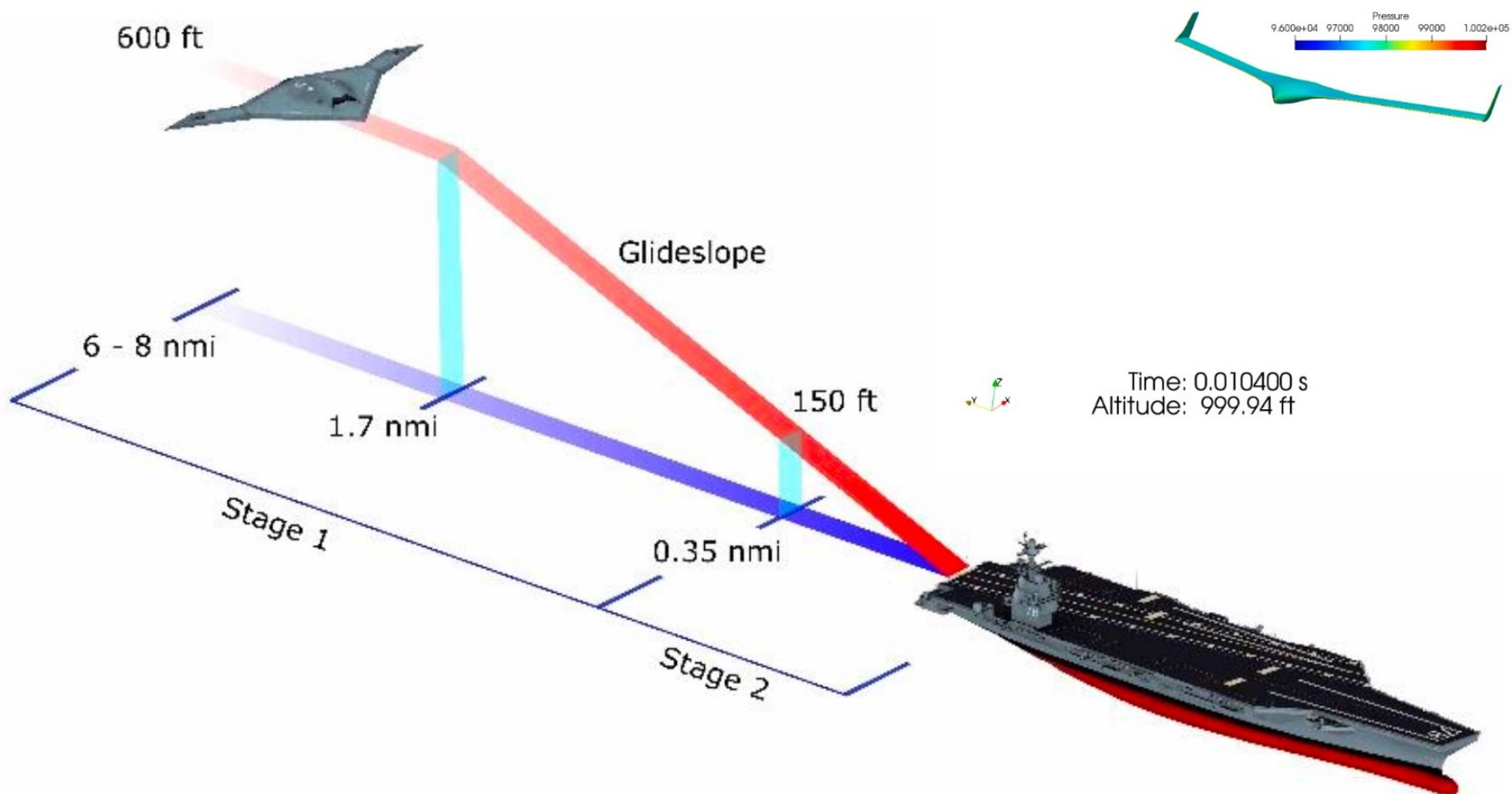
➤ 无人系统的实时预测控制





需要实时反馈

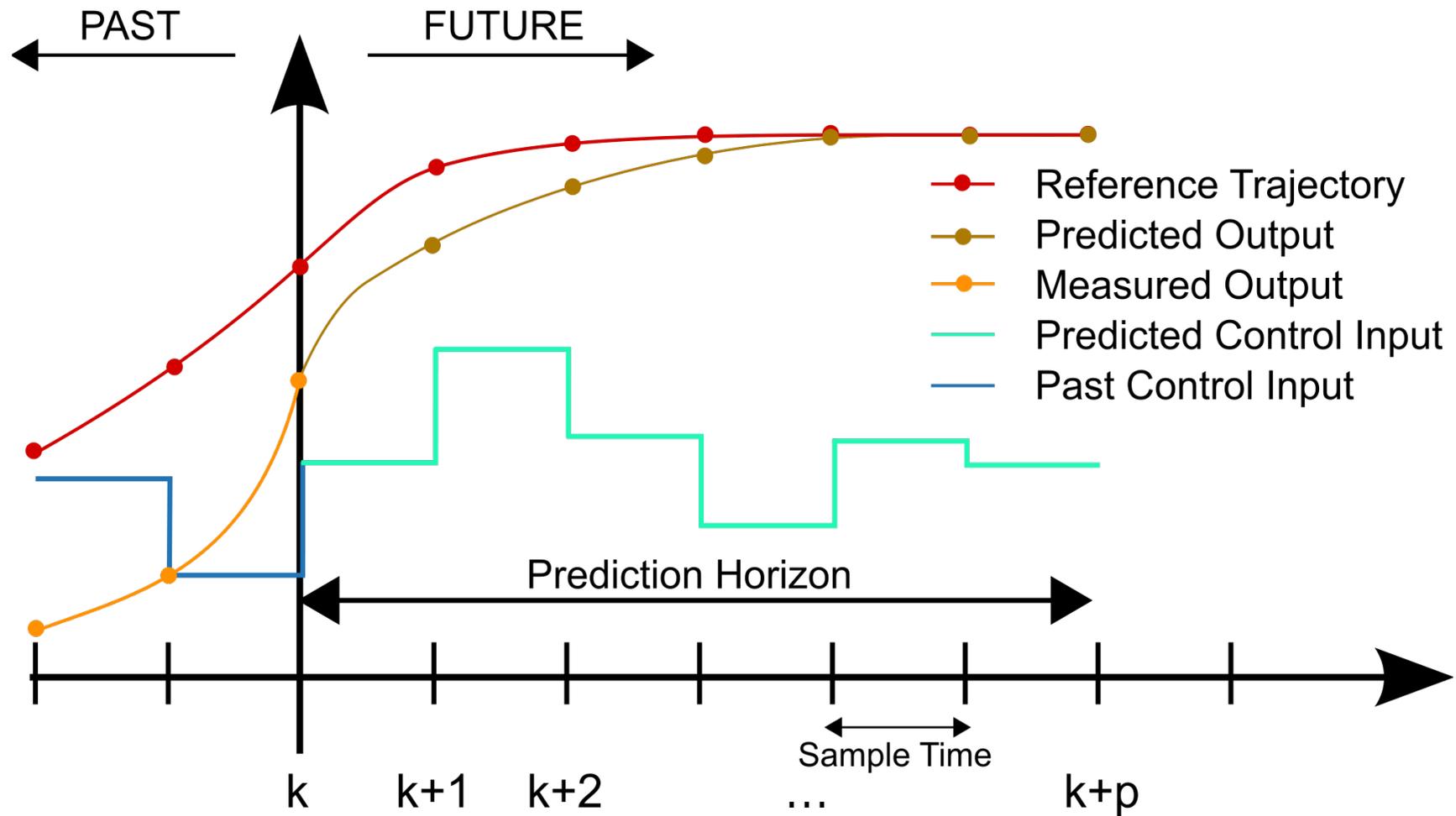
➤ 无人系统的实时预测控制





需要实时反馈

➤ 无人系统的实时预测控制





多尺度多物理场建模





科学计算中的挑战

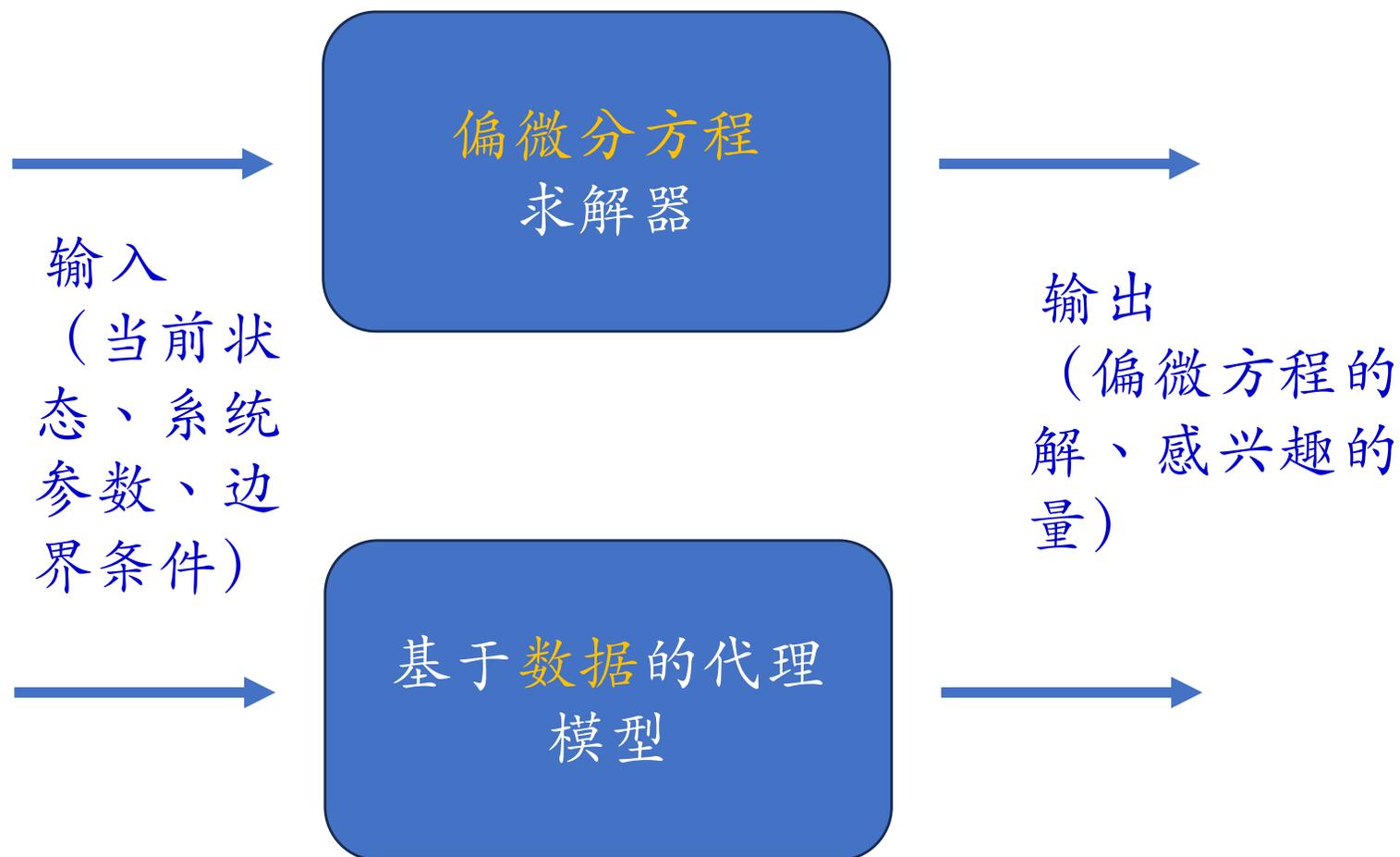
- 运算量巨大难以承受
 - 多尺度多物理场建模
 - 反复求解、迭代
 - 需要实时反馈
 -

简化模型、代理模型



科学机器学习

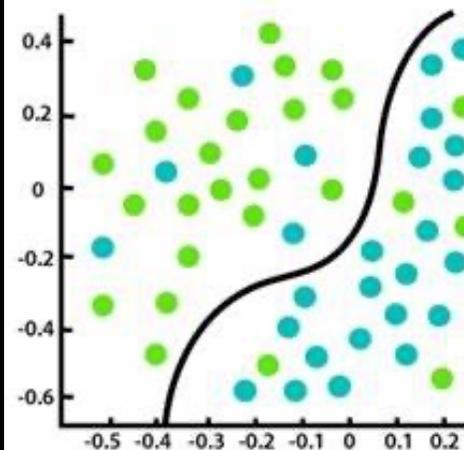
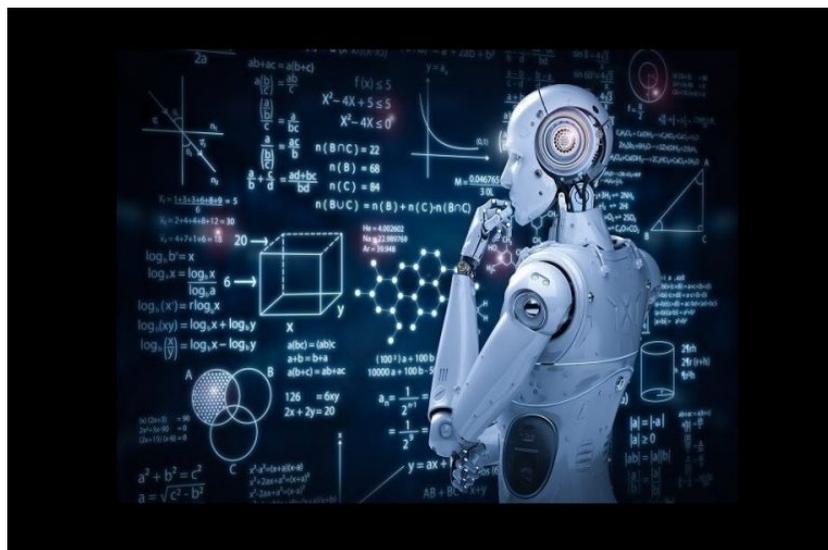
➤ 代理模型(Surrogate models)



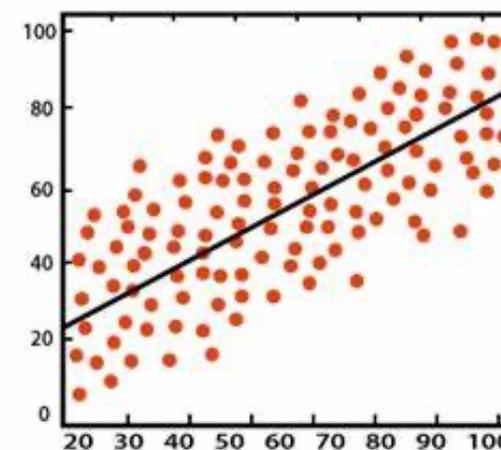


科学机器学习

➤ 机器学习 (Machine learning)



Classification



Regression

- 图像识别、自然语言处理、生成模型
- 蛋白质折叠、分子制药
-

从数据出发，不依赖或并没有基于方程的可靠模型



科学机器学习

➤ 机器学习

- 高效处理高维问题
- 需要大量的数据
- 有一定不确定性
- 对精度、误差等并没有较好的理解
-

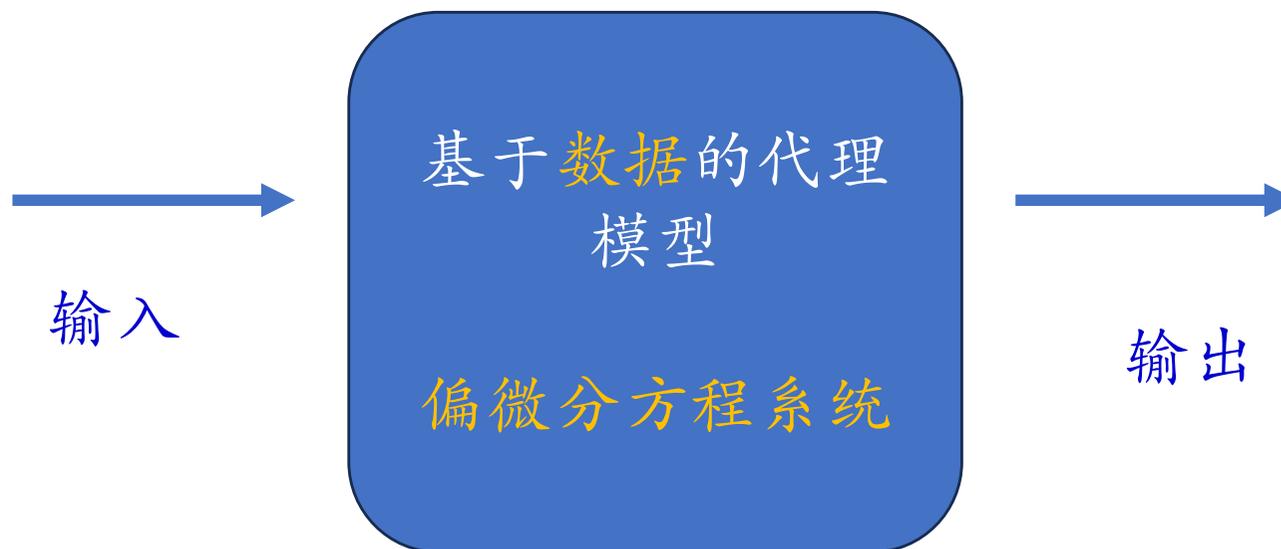
➤ 物理系统

- 遵从物理定律、有基本的方程描述
- 需要定量的预测
- 传统方法对精度、误差有较好的理解
- 数据并不充裕
-



科学机器学习

➤ 代理模型(Surrogate models)



- 建模预测复杂系统
- 工程系统设计
- 工程系统实时控制
- 量化不确定性、可靠性评估
-



科学机器学习

➤ 基于投影的降阶模型

- 一般计算模型都是高维计算模型(High dimensional model, HDM)
- 在利用了目标系统的相关知识后，构造一个低维度模型能够捕捉目标系统的主要行为

高维计算模型: $r(u, \mu) = 0, u \in R^N$

低维模型:

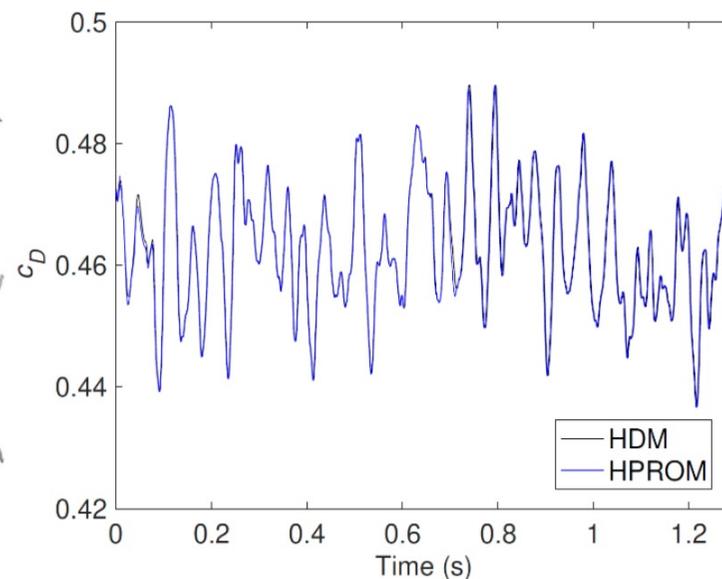
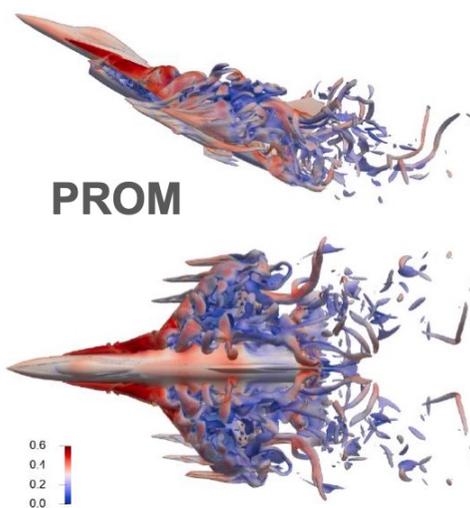
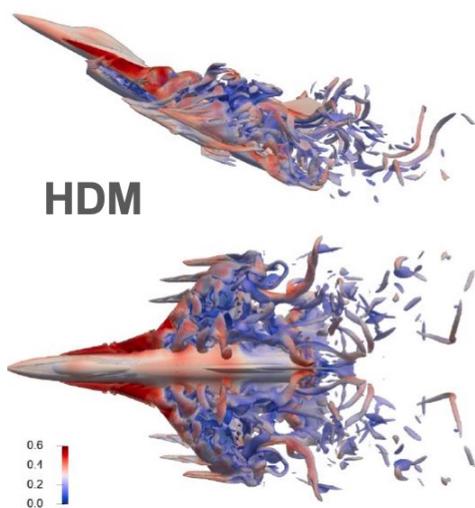
$$u = u_0 + V\Delta y, \quad V \in R^{N \times k}, \quad \Delta y \in R^k, \quad k \ll N$$



科学机器学习

➤ F-16 ($Re=182, 200, 000$)

- 计算流体网格 26, 919, 873 格点
- 模拟仿真时间 $t = 1.29s$



高维计算模型：100h 3584CPU核

基于投影的降阶模型：5.8min, 32CPU核

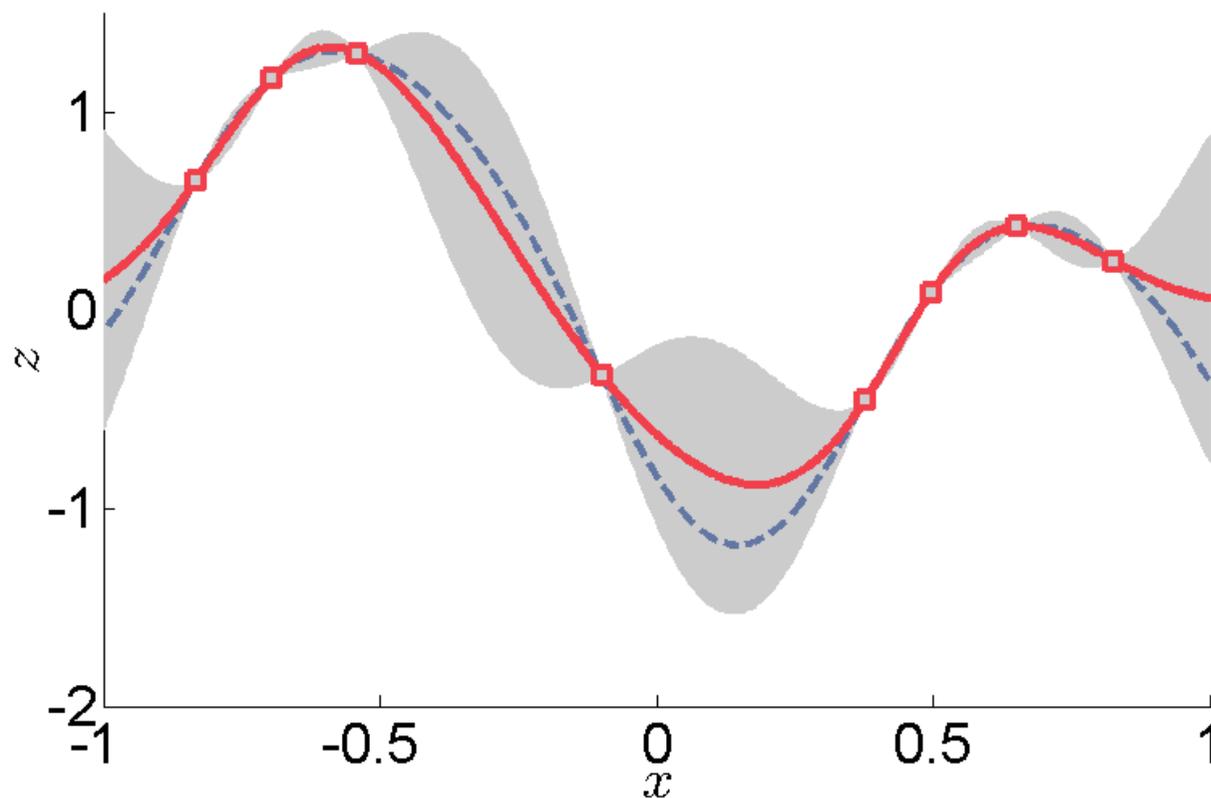
- 10^3 倍提速 (挂钟时间)
- 10^5 倍提速 (CPU时间)



科学机器学习

➤ 高斯过程

- 针对目标系统的关注量(Quantities of Interest, QoI)的插值、不确定性量化
- 基于先验方差高斯过程

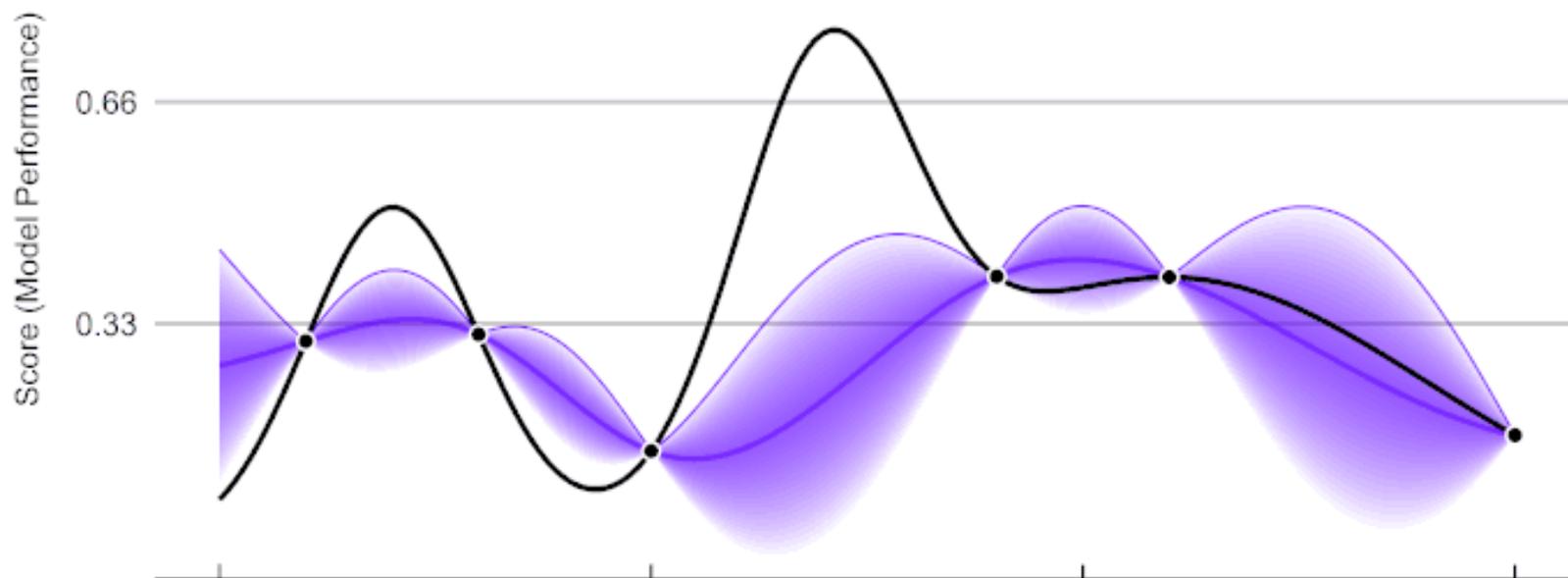




➤ 高斯过程

- 贝叶斯优化

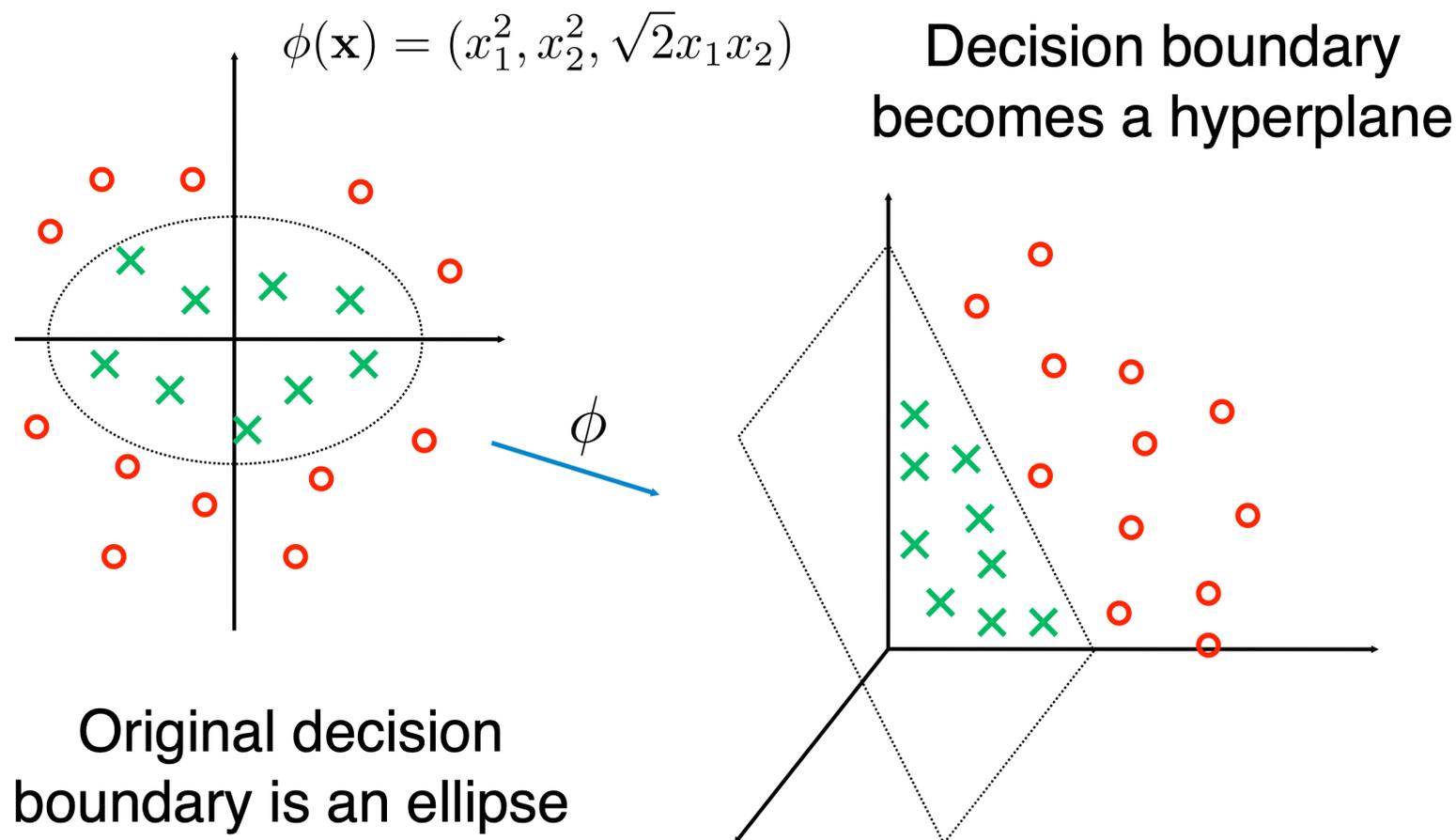
ParBayesianOptimization in Action (Round 1)





➤ 高斯过程

- 核方法(Kernel method)
- 随机特征方法(random feature)





科学机器学习

➤ 基于神经网络的方法

- 针对目标系统的关注量(Quantities of Interest, QoI)的插值
- 构造一个神经网络能够捕捉目标系统的主要行为

函数: $\mathbb{R}^n \mapsto \mathbb{R}^m$

线性函数: $x \mapsto Ax + b$

非线性函数: $\sigma(x) \mapsto \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases}$

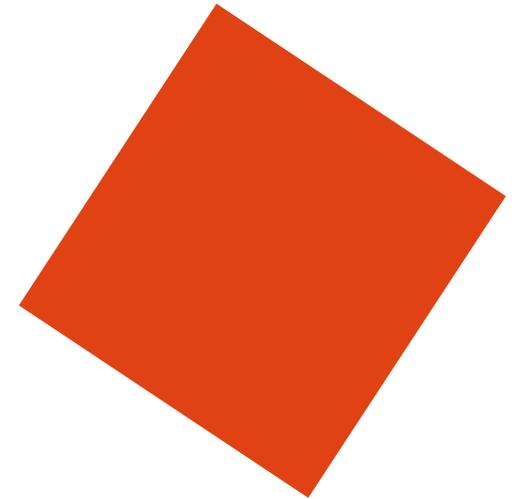
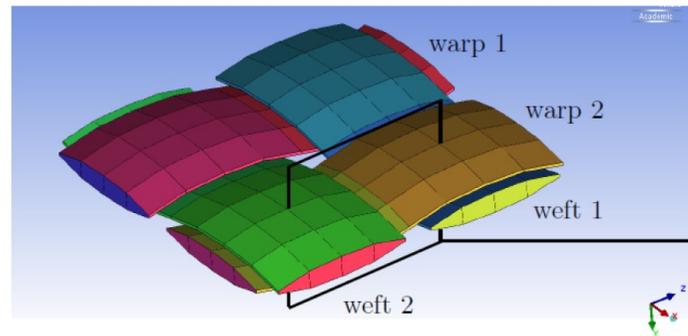
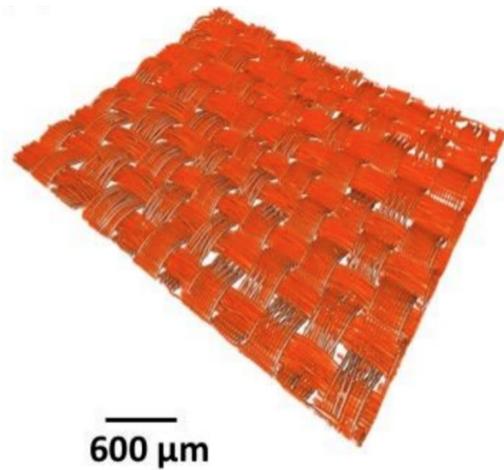
$$\sigma(x) \mapsto \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



科学机器学习

➤ 基于神经网络的方法

- 多尺度建模
- 偏微分方程代理模型



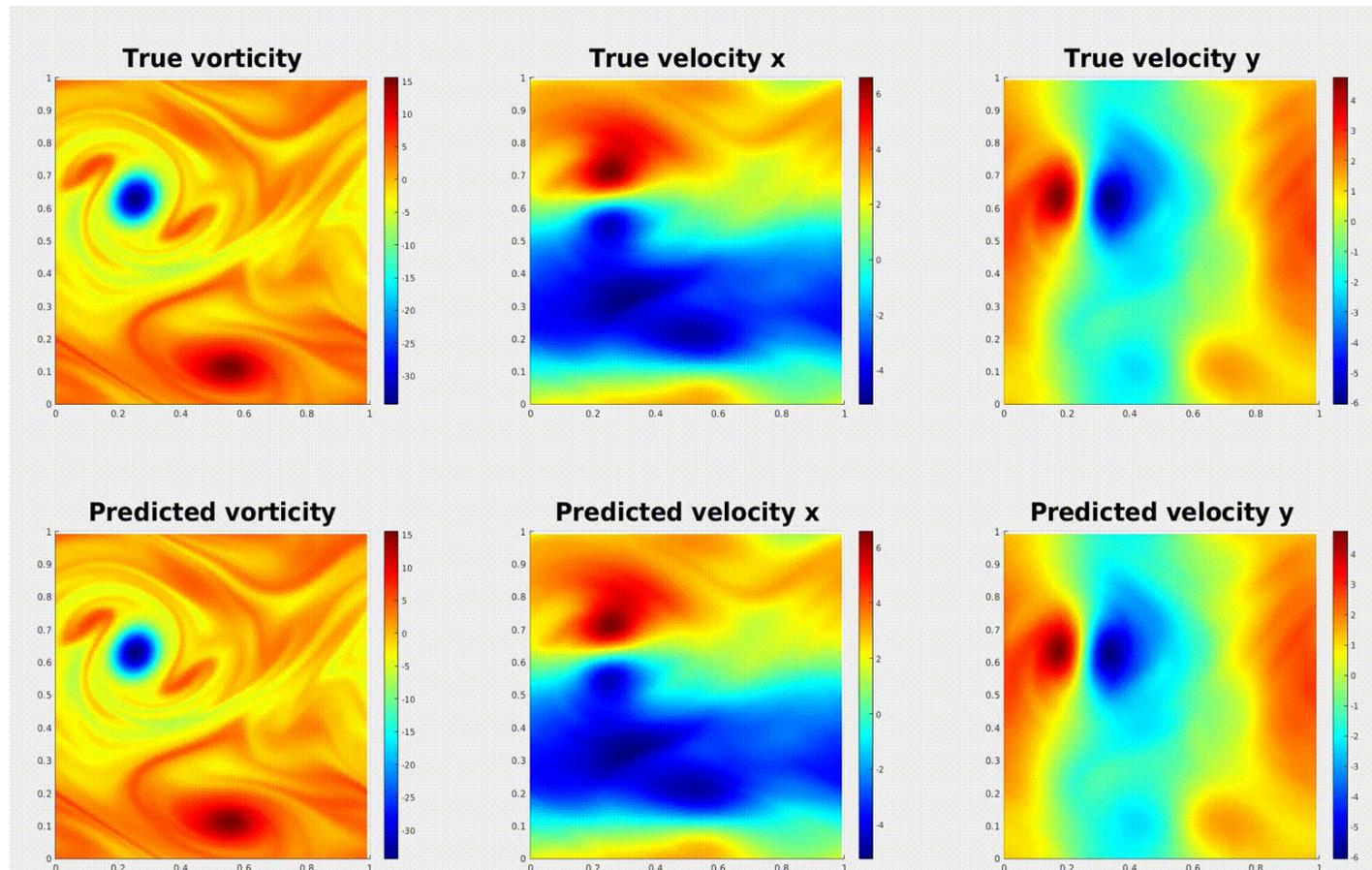
材料本构关系建模： $\sigma = \mathcal{M}_\theta(\epsilon)$



科学机器学习

➤ 基于神经网络的方法

- 多尺度建模
- 偏微分方程代理模型



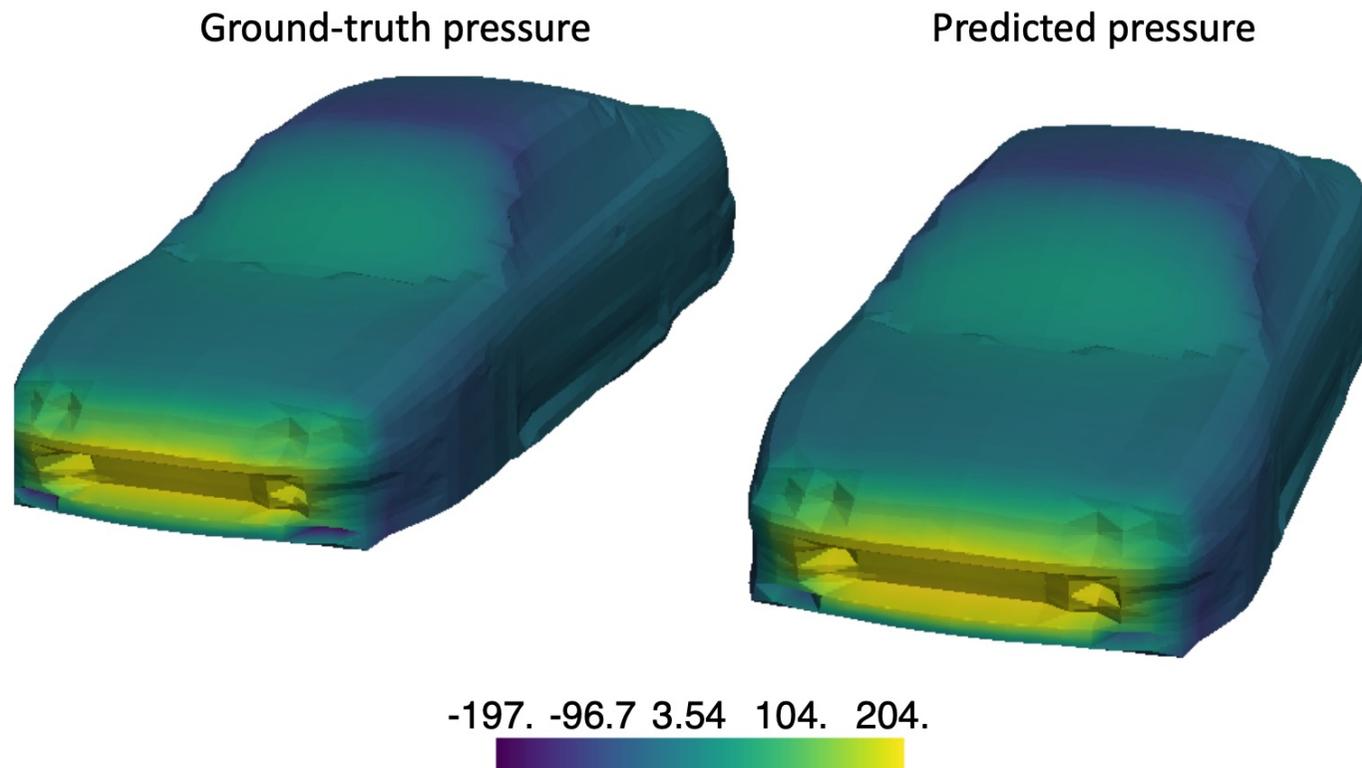
流体模拟



科学机器学习

➤ 基于神经网络的方法

- 多尺度建模
- 偏微分方程代理模型



流体模拟