NREL

Transforming ENERGY

Adaptive Computing and multi-fidelity learning

Juliane Mueller
Computational Science Center

With contributions from Marc Day, Hilary Egan, Ryan King, Kevin Griffin, Jibo Sanyal

Scale-up of complex systems and associated risks

Scale-up: Extending systems and processes that were developed in the laboratory to function in the real world

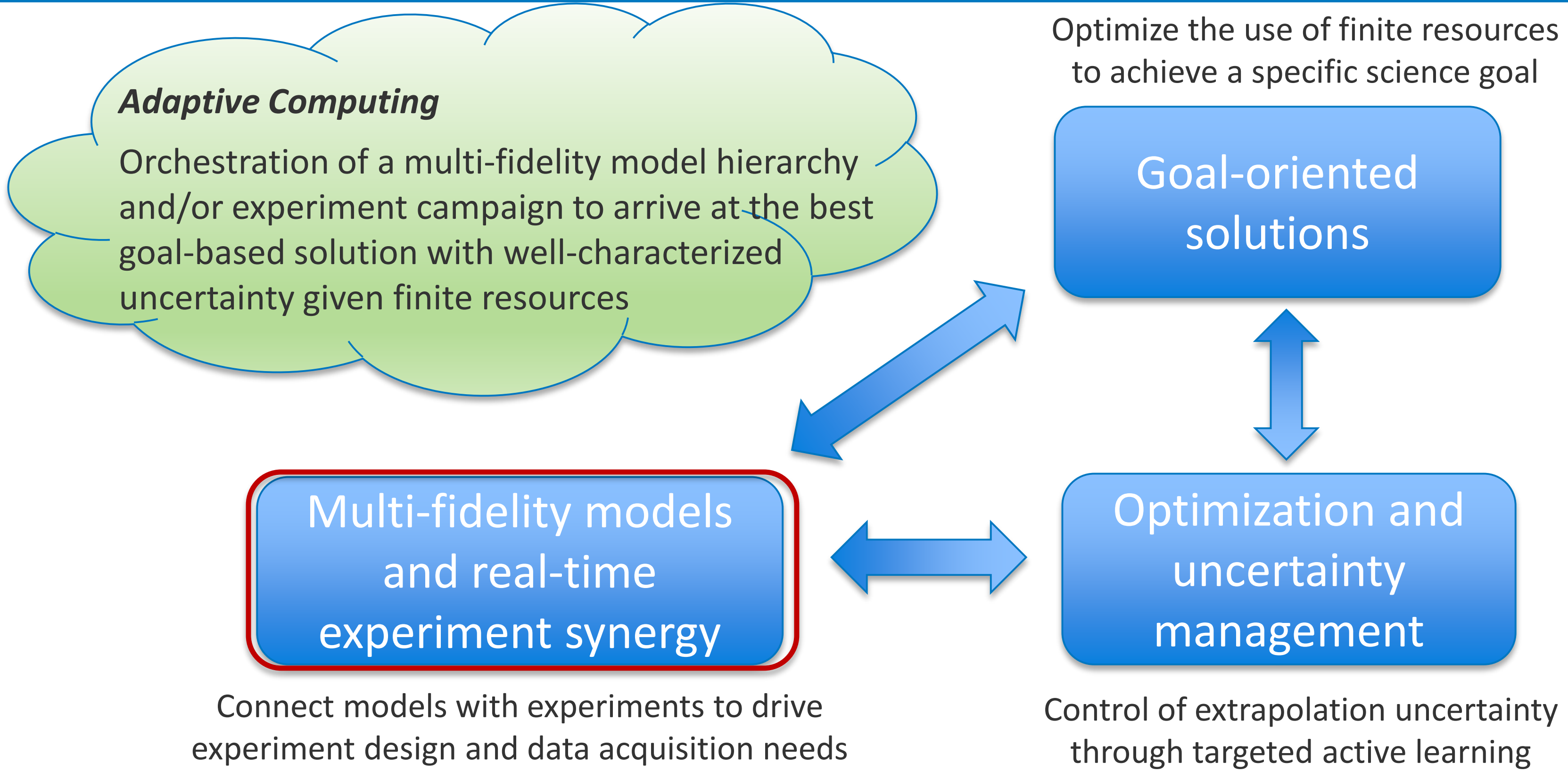
Device and process scale-up comes with significant technical challenges and risk

Typical challenges:

- Data-driven models perform best when interpolating, **extrapolation is inherently uncertain,** and therefore **risky**
- Increasing ranges of scale (spatial, temporal) often lead to new/enriched physics
- High-fidelity physics-based models may capture new physics, but are typically too expensive for design/optimization work
- Operational regimes of existing experiments are limited, and new experiments are expensive

Image Credit: Dennis Schroeder, NREL63958

Goal: reduce scale-up challenges by integrating multi-fidelity modeling and optimal compute resource use



Key capability: multi-fidelity modeling

Most applications feature an assortment of models of widely varying fidelities, developed for different purposes:

- Experiment: "Truth", but limited operational regime
- High-fidelity simulations: Physics-based (PDE/ODE), costly
- Lower fidelity levels: reduced physics, coarser meshes, less costly
- Data-driven surrogates: AI/ML, PINNs, Gaussian Processes (GPs), really cheap

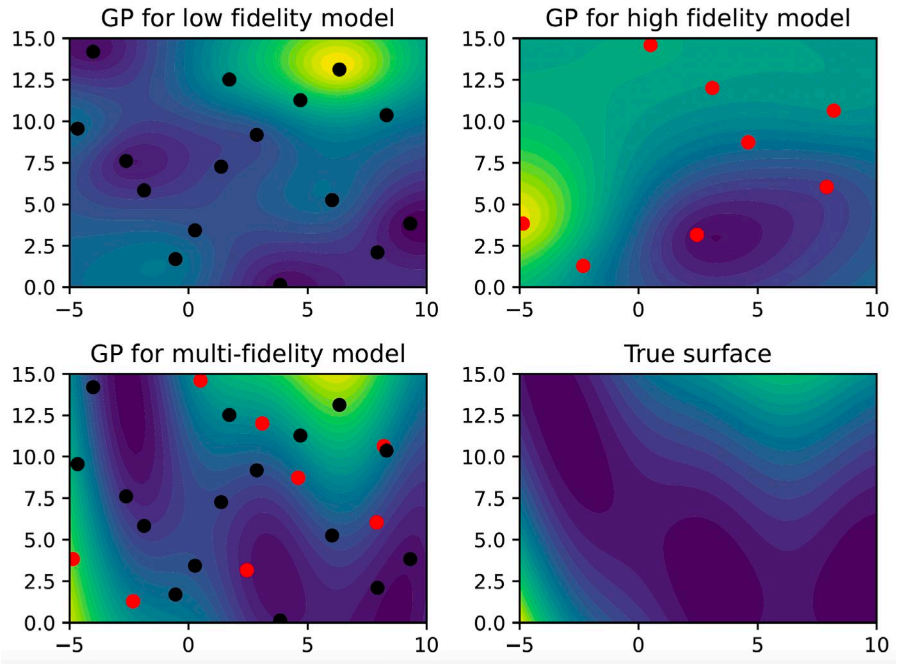


Fig: Exploiting information from multiple fidelity levels can increase surrogate accuracy

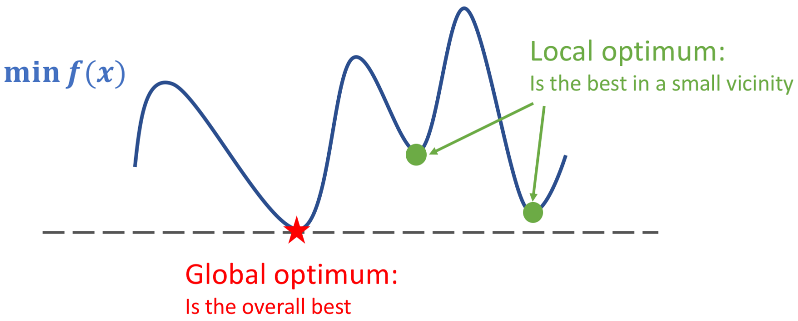
Black-box expensive optimization: high fidelity

$$\min f(x)$$
$$s.t.\ g_i(x) \leq 0, i = 1, \dots, I$$
$$x \in \Omega$$

Objective function to minimize

Constraint functions

Parameter domain



$$\cancel{\frac{df}{dx}}$$

Surrogate models approximate the expensive function

$$f(x) = m_f(x) + e_f(x)$$ Surrogate of the objective function

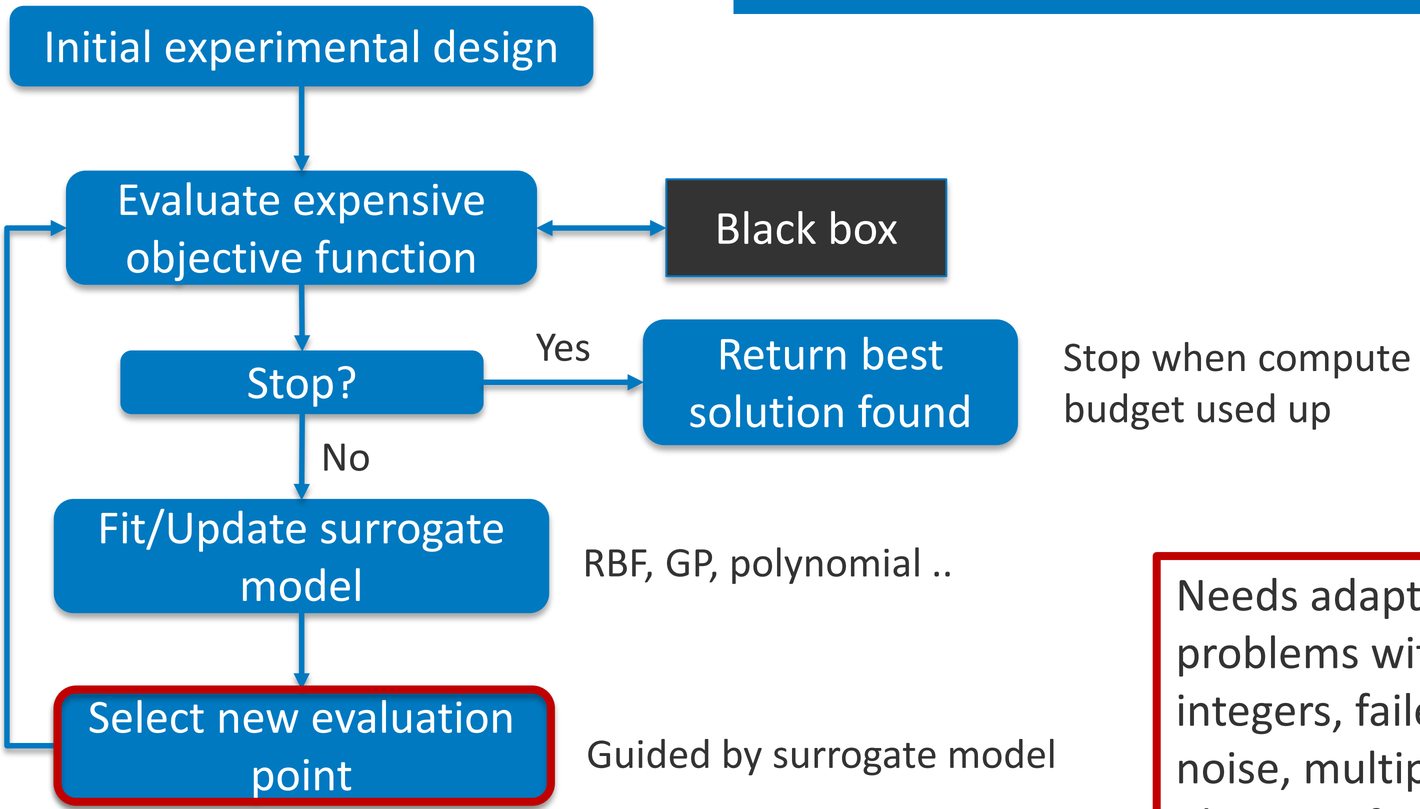
$$g_i(x) = m_{g_i}(x) + e_{g_i}(x)$$ Surrogate of the constraint function

Different types of surrogate models exist:

- Radial basis functions
- Gaussian Process models
- Multivariate adaptive regression splines
- Polynomials
- …..

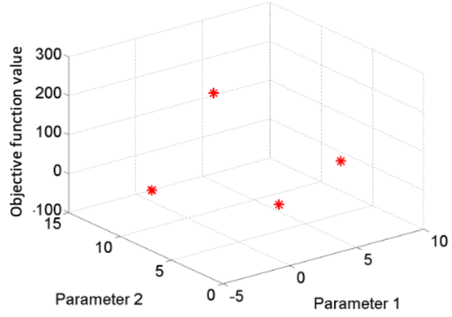
Throughout the optimization, we let the surrogate model guide the search for improved solutions

Surrogate model based optimization loop

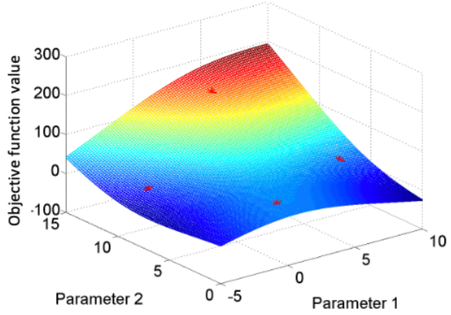


Needs adaptation for problems with constraints, integers, failed evaluations, noise, multiple conflicting objective functions

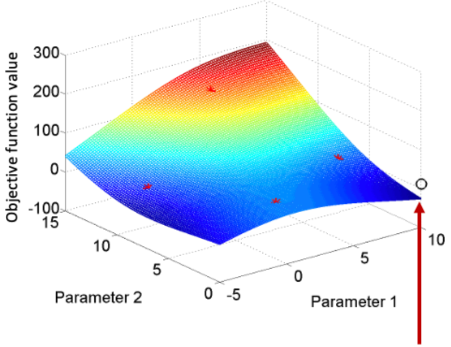
Surrogate model guided sampling



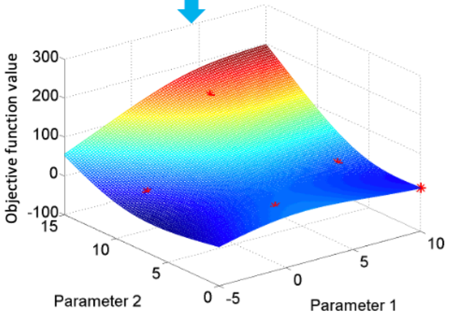
Initial experimental design



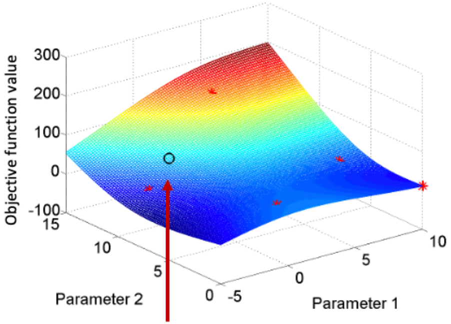
Compute surrogate surface



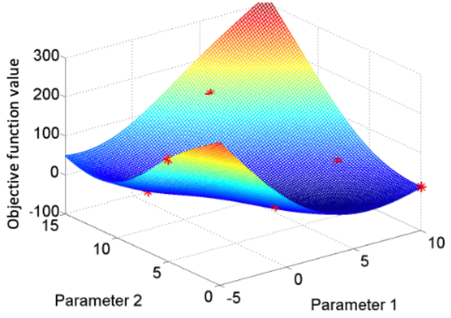
Select and evaluate new point



Update surrogate surface



Select and evaluate new point



Update surrogate surface

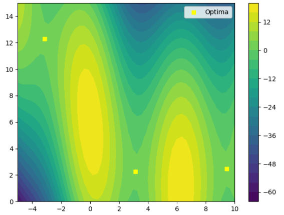
…

Exploiting multiple fidelity levels

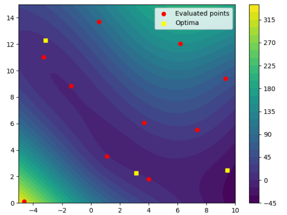
Correcting the low-fidelity model:

- Multiplicative: $\hat{y}_{hf}(x) = \rho(x) * y_{lf}(x)$
- Additive: $\hat{y}_{hf}(x) = y_{lf}(x) + \delta(x)$
- Hybrid:
  - $\hat{y}_{hf}(x) = \rho(x) * y_{lf}(x) + \delta(x)$ (ρ const.)
  - $\hat{y}_{hf}(x) = w(x) * \rho(x) * y_{lf}(x) + \big(1 - w(x)\big) * (y_{lf}(x) + \delta(x)), w \in [0,1]$

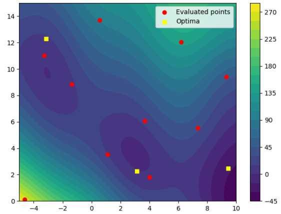
$y_{lf}(x)$

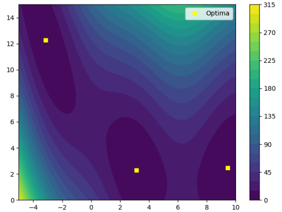


$\delta(x)$



$y_{lf}(x) + \delta(x)$



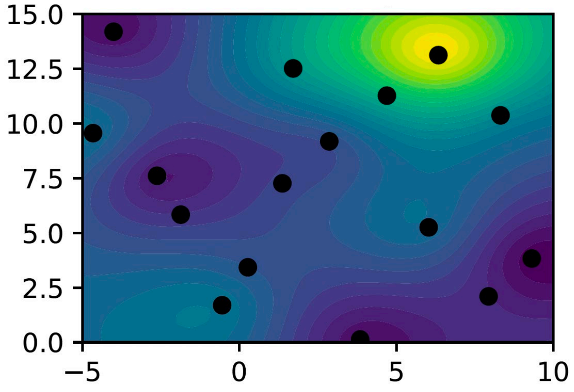


Ground truth

How do we make use of multiple fidelity levels during active learning?

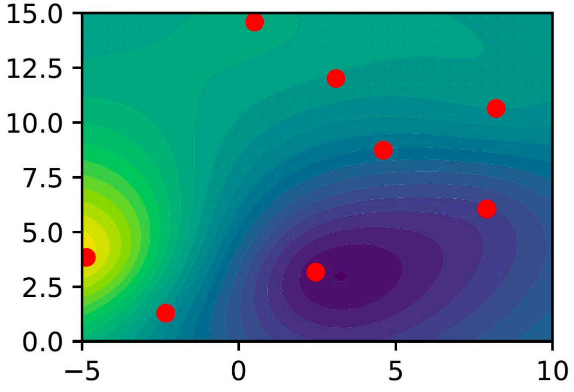
Exploiting multi-fidelity information

Build a surrogate model for the low(er) fidelity function



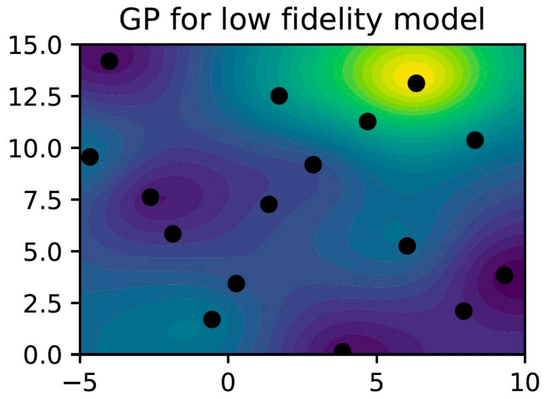
- Allow more samples than for high-fidelity function
- Use this surrogate to decide where to focus the search in the high-fidelity function
- *Low fidelity model does not have to be accurate*

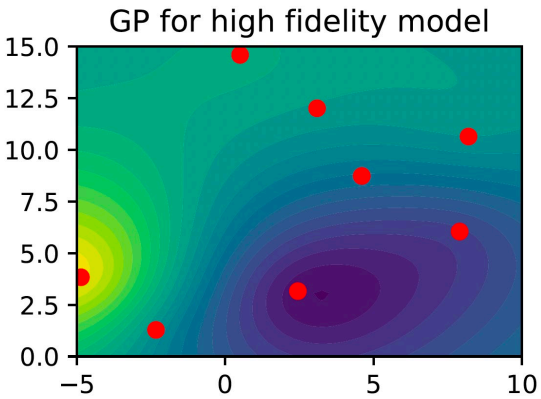
Build a surrogate model for the high-fidelity function



- Fewer samples are affordable
- Surrogate is less accurate (built on less data)
- Surrogate can be used to make (final) sample decisions

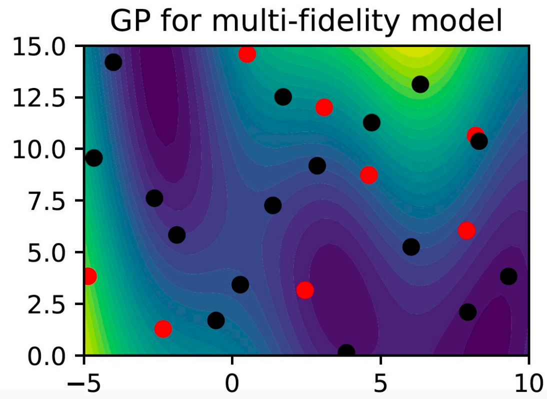
Gaussian Process: Using multiple fidelity information in one model

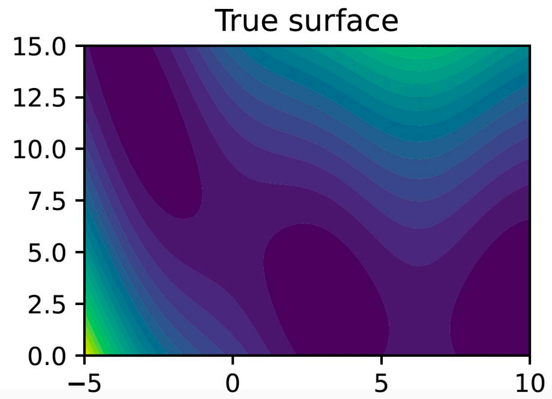




Red = high-fidelity evaluations
Black = Lower fidelity evaluations

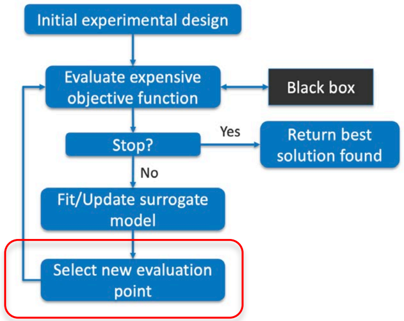
Python package surrogate modeling toolbox (SMT)





Combining high and lower fidelity information can lead to better approximation surface (compare to true contours)

GP-based sampling methods



Expected improvement balances local and global search

$$\mathbb{E}(I)(\mathbf{x}) = s(\mathbf{x})\left(\nu\Phi(\nu) + \phi(\nu)\right), \quad \nu = \frac{f^{\text{best}} - m_{\text{GP}}(\mathbf{x})}{s(\mathbf{x})}$$

Probability of improvement – mostly local search

$$\mathbb{P}(I)(\mathbf{x}) = \Phi\left(\frac{f^{\text{best}} - m_{\text{GP}}(\mathbf{x})}{s(\mathbf{x})}\right)$$

Lower confidence bound

$$LCB(\mathbf{x}) = m_{\text{GP}}(\mathbf{x}) - \kappa s(\mathbf{x})$$

m_{GP} is the GP prediction

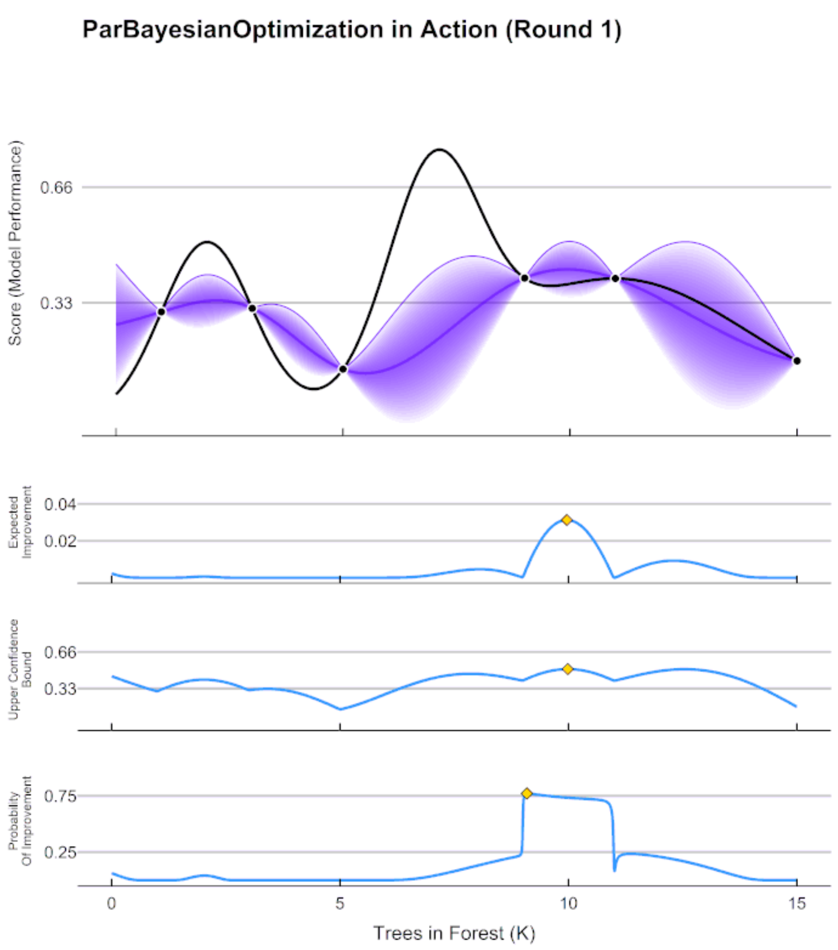
s is the standard deviation of the GP predictions

f^{best} is the best function value found so far

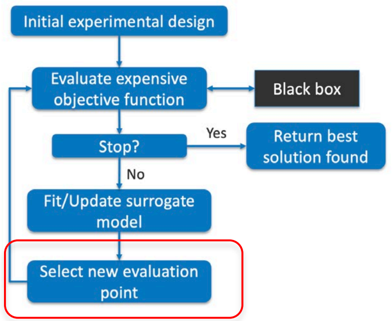
ϕ, Φ are the normal density and cumulative distribution

κ – adjustable parameter

GP based sampling in 1d



Maximization problem

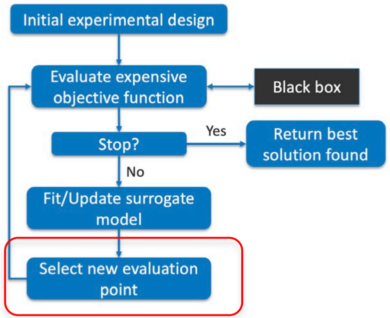


Expected improvement

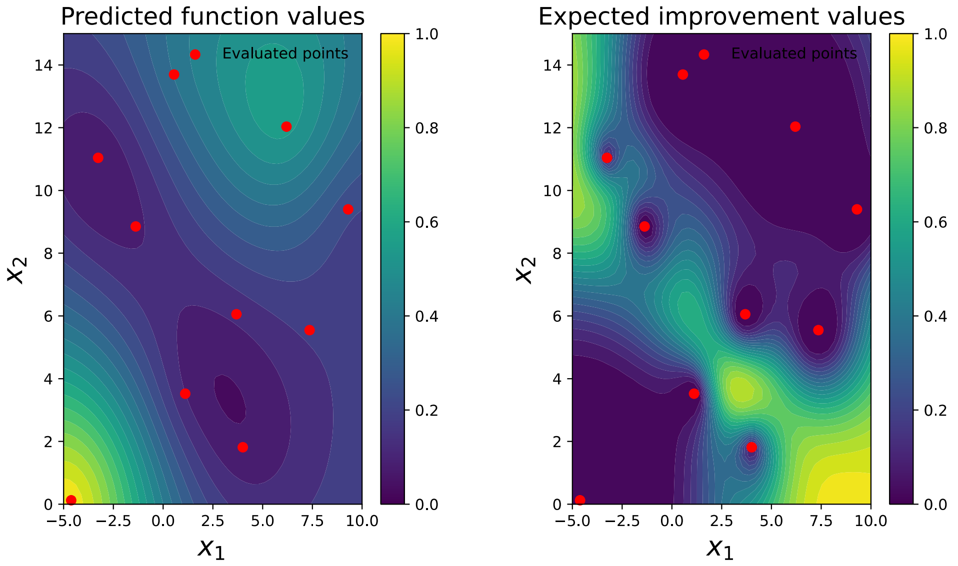
Upper confidence bound

Probability of improvement

Maximize the expected improvement to select a new point

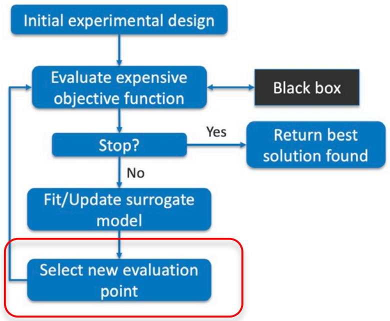


Expected improvement surface is multimodal and can become flat – making it difficult to find the global maximum…

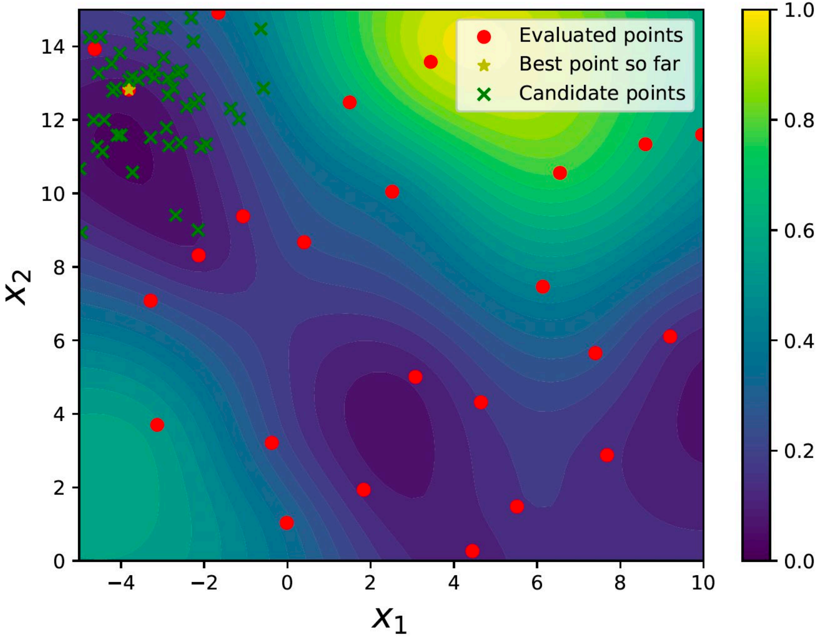


…requires development of other sampling strategies, guided by low-fidelity model

Sampling with candidate points



Add random perturbations to (select) variables of the best point(s) found so far



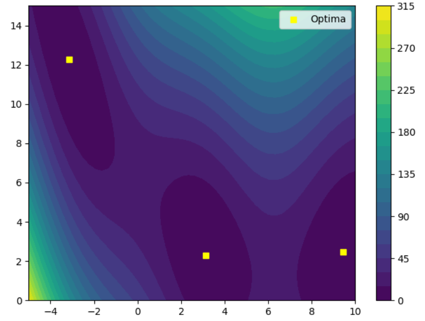
- Maximize a merit function that trades off predicted function value and distance to already evaluated points
  - Low function value -> local search
  - Large distance -> global search
- Select N new points for potential evaluation

Multi-fidelity sampling: when to ignore the low-fidelity model

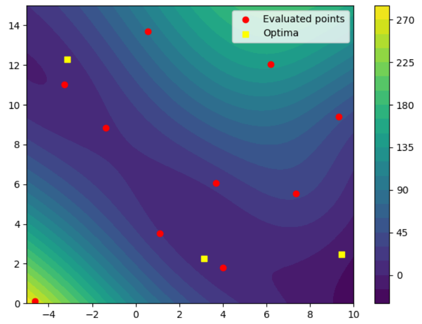
- Make use of as much information as is available
  - Surrogate of high-fidelity model
  - Low-fidelity (cheap) information – what if this one is very inaccurate/uncorrelated?
  - Surrogate of the difference as a selection constraint

1. Define auxiliary function $a(x)$ using the surrogate model predictions
2. Optimize $a(x)$ to find x_{new}
3. If $-\delta \leq d(x_{new}) \leq \delta$ probe with low fidelity model first, otherwise ignore and evaluate high-fidelity model

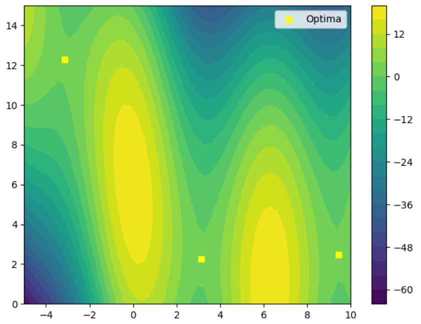
High-fidelity ground truth



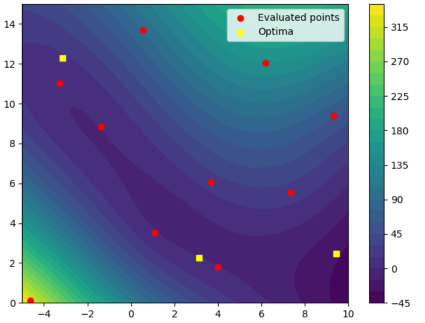
SM of high-fidelity model



Low-fidelity model



Difference between high- and low-fidelity



Goal: reduce scale-up challenges by integrating multi-fidelity modeling and optimal compute resource use

Adaptive Computing

Orchestration of a multi-fidelity model hierarchy and/or experiment campaign to arrive at the best goal-based solution with well-characterized uncertainty given finite resources

Optimize the use of finite resources to achieve a specific science goal

Goal-oriented solutions

Multi-fidelity models and real-time experiment synergy

Connect models with experiments to drive experiment design and data acquisition needs

Optimization and uncertainty management

Control of extrapolation uncertainty through targeted active learning

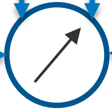
Key capability: diverse compute resources

Experiment

Edge

Cloud

Datacenter



Resource Management

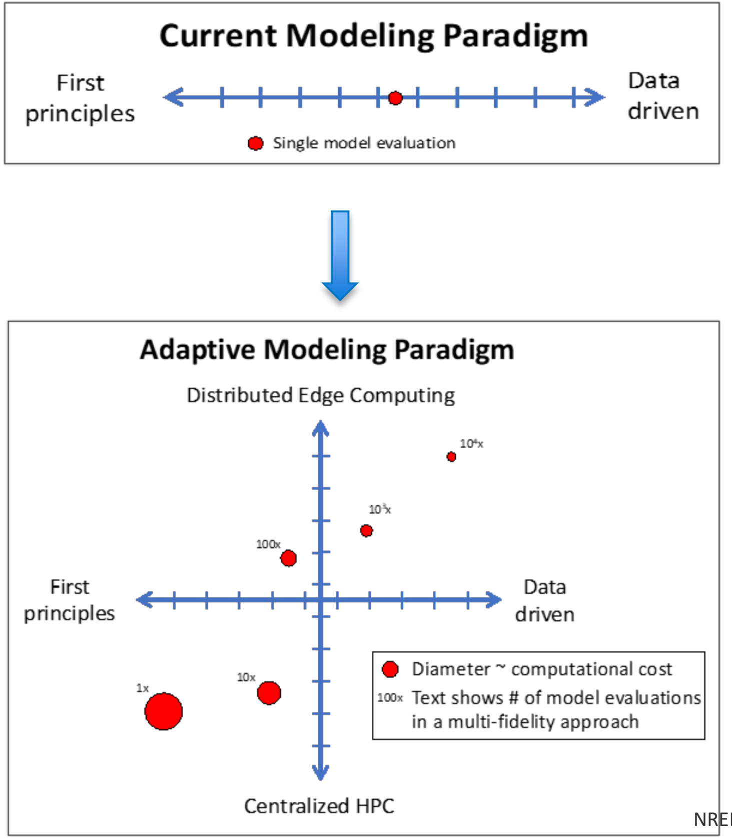
Resource manager:
how much and what kind of resources do I have available?

Solve stochastic discrete optimization problem

Optimal computing strategy driven by specific output quantity of interest

Compute resource optimization problem

- What resources are available when?
- Formulate as optimization problems with stochasticity
- Implement solutions as constraints for multi-fidelity sampling
- Eventually must exploit asynchronous parallel computations



- Enumerate the user-defined simulation types (fidelity levels)
- Possible hardware configurations (# of CPUs, GPUs)
- Corresponding calculation duration
- Measurement noise estimate (aleatoric uncertainty)

Compute resource allocation with stochasticity

1. Select sample points (e.g., maximize EI with multi-start; candidate point approach)

$$\{x_1, x_2, \ldots, x_N\} \in \underset{x \in \mathcal{X}}{\mathit{argmax}}\, a(x)$$

2. Get total resource limit T and per level resource limit T_j and allocate compute resources

$$\max_{\mathrm{b_{ji}} \in \{0,1\}^{\mathrm{k}}} \sum_{j=1}^{J} r_j(x_i) \; * b_{ji}$$

r_j the benefit of evaluating x_i at fidelity level j, e.g., r_j captures accuracy or other QoI

$$b_{ji} = \begin{cases} 1 \text{if } x_i \text{ evaluated with fidelity level } j \\ 0 \text{ else} \end{cases}$$

$J * N$ binaries to optimize

$$\sum_{j=1}^{J} \sum_{i=1}^{N} b_{ji} * t_j(x_i, \zeta_{ji}) \leq \; T$$

Total resource restriction

t_j resource consumption at level j

$\zeta_{ij} \sim ?$

$$\sum_{i=1}^{N} b_{ji} * t_j\left(x_i, \zeta_{ji}\right) \leq \; T_j \quad \forall j$$

Resource restriction on fidelity level j

AC framework

Check resource availability
- Data center
- Edge
- Cloud
- Experiment

Resource mana

Surrogate models

Sampling methods

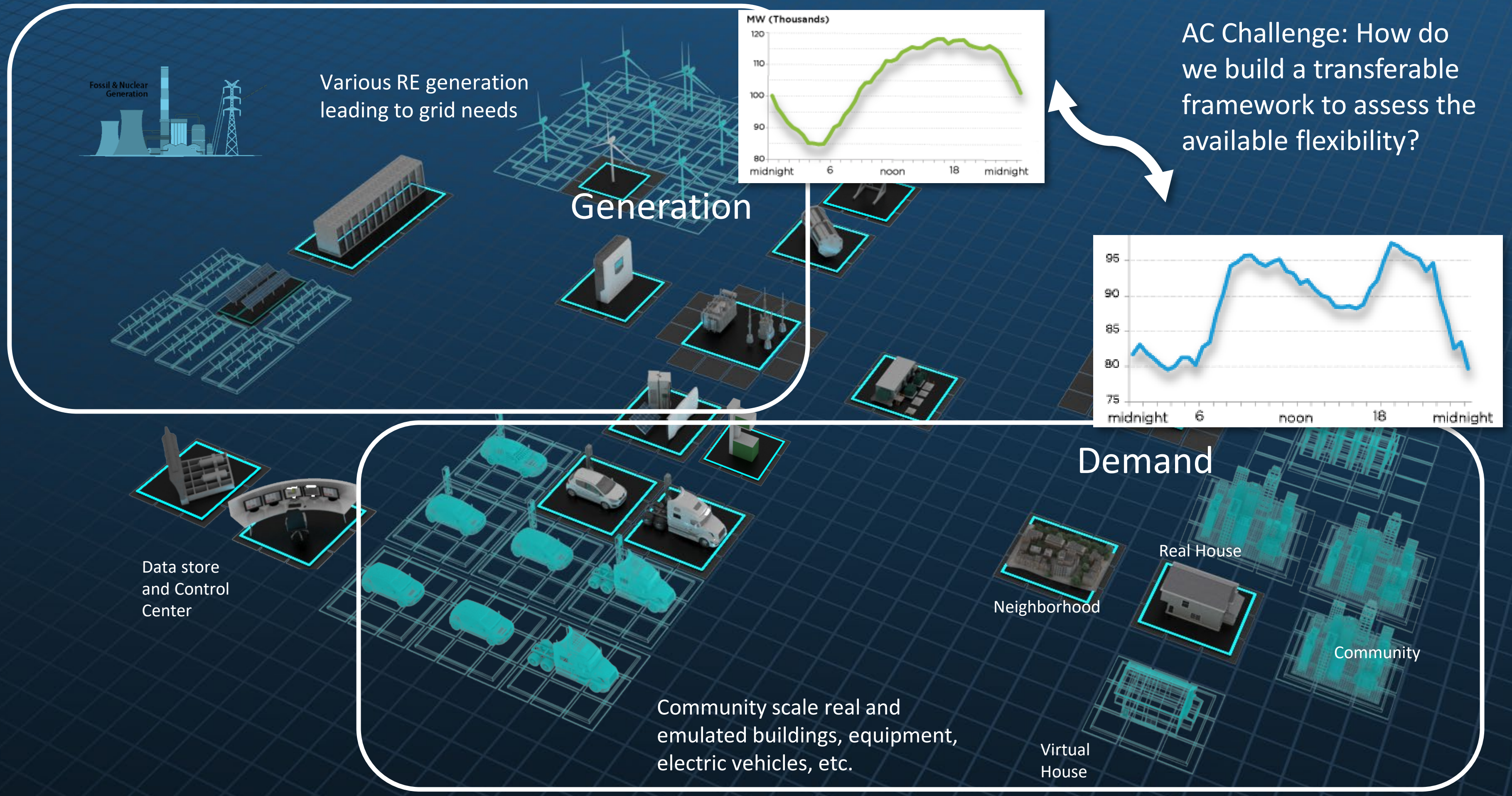
Statistics, analysis, visualization

- Optimization
- Sensitivity analysis

- Fidelity lvl 2
- …
- Fidelity lvl M

Application specific code

ARIES – Advanced Research on Integrated Energy Systems



How much demand flexibility is available to support the grid

Goal: Understand, evaluate, and predict the demand flexibility a device, building, or a community can provide:

- Optimizing or shaping the energy demand to support grid conditions
- Providing a way to tap into extra demand flexibility in extreme situations

Challenge: Computing the anticipated energy demand and the opportunities to exploit optimal control is computationally expensive

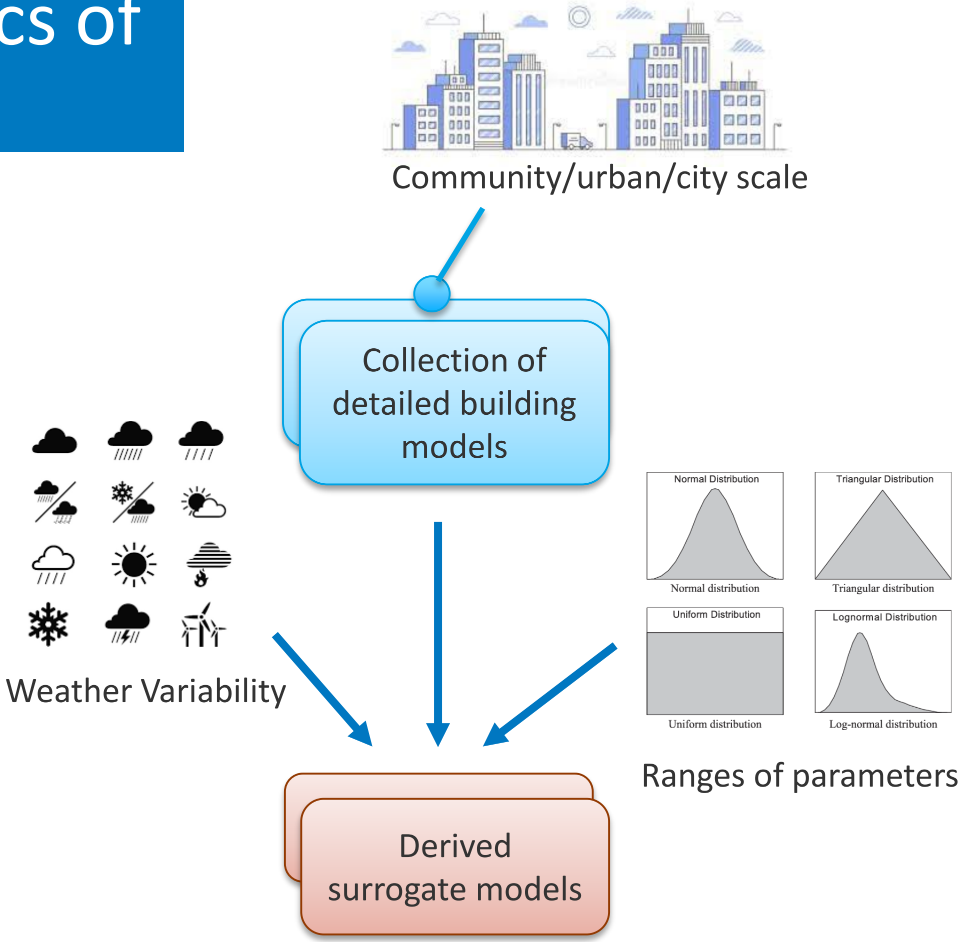
- Data from real-world observations and complex building simulation models to evaluate the available flexibility in the community thru:
  - Control of HVAC, lighting, water heating, etc. individually or in combination
  - Aggregating the demand profile to the distribution network and grid
- Use surrogate models and understand where they are too inaccurate and additional real-world data and/or complex full-scale simulation data must be collected
- Computational cost in deployment settings is substantially more acute than in the lab setting

Sources and characteristics of uncertainty

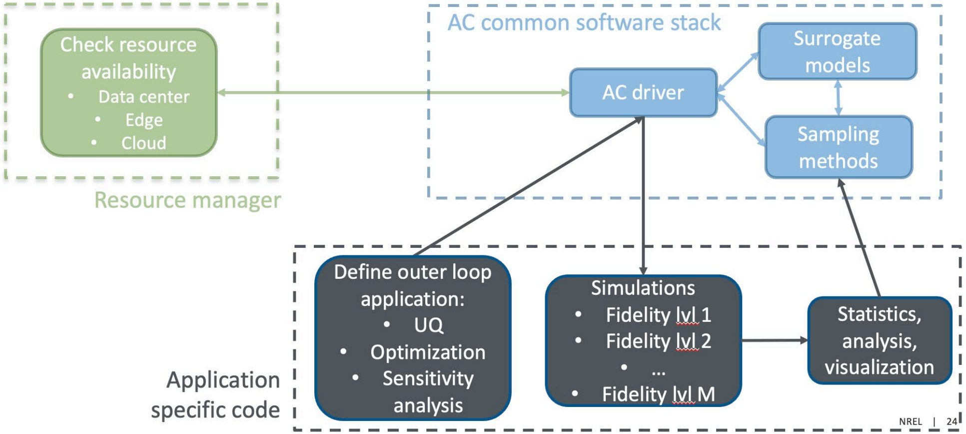
Limitations of surrogate models: surrogate models can only be created for a subset of model parameter distributions, number of dimensions, etc.

- Building energy models have 1000s of parameters with 100s that are important
- Will not represent all possible building types/configurations/equipment classes
- Sampling of only a subset of weather conditions is possible
- Limited transferability across geographic locations

Trust in the surrogate requires uncertainty quantification



ARIES AC framework definitions



Objectives: controls-oriented questions around energy and heating/cooling

- Peak shaving, load shifting, preheating/precooling
- Assess impact of higher efficiency equipment (e.g., LED lighting)
    - Some component changes come in from community questions, renewable energy mandates, prioritize based on carbon footprint
    - ResStock /ComStock libraries for components
- Widespread EV charging

ARIES AC framework definitions

Modeling ecosystem

- *EnergyPlus* is high fidelity model (1-30 minutes per building for a year simulation w/ 1min timestep, heavy on I/O)
- *OCHRE*, struggles in some locations, residential only (E+ does residential and commercial)
- *3R2C* lumped capacitance ROM (4-5 hrs to train) for model predictive control w/ 15min to 1 hour timestep (works for residential and commercial)
- Models are *deterministic and highly nonlinear* (issues with using ROM outside of intended regime)

Parameterizations

- Need to preserve explicit component representation
- Building thermodynamics should be independent of component selection

ARIES AC framework definitions

Challenges

- Scaling simulation and control to larger communities
- Discrete optimization
- ROM accuracy is a function of component choice

Test case going forward

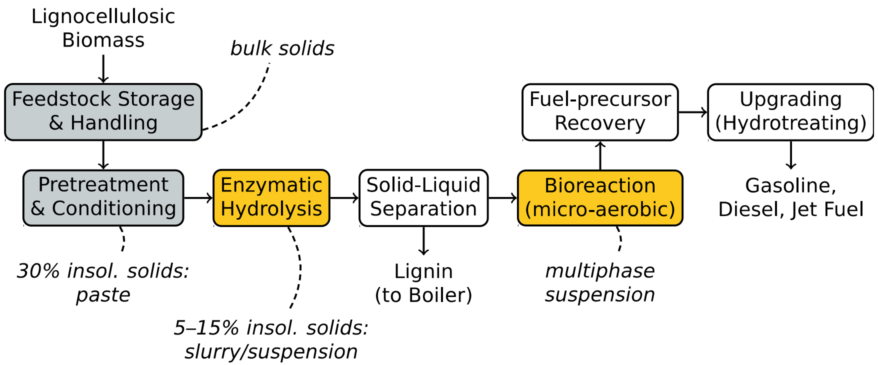
- Pena station with ~150 buildings
    - 50 defined so far
    - Multifamily buildings are challenging, EnergyPlus extra expensive

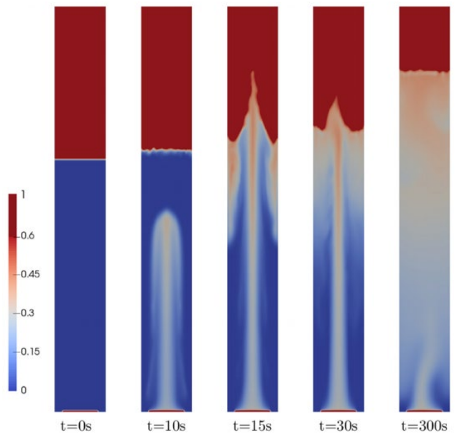
- EnergyPlus and 3R2C ROM already built

- UrbanOpt usually used to generate model inputs, but have a little more info for Pena station

Example application: virtual engineering of biofuels

- Process lignocellulose-rich biomass into biofuel
- 3 step chemical processes
    - Pretreatment: fast simulation
    - Enzymatic hydrolysis: surrogate or CFD calculation
    - Bioreaction: surrogate or CFD calculation





AC specifics for virtual engineering

- **Objective:** maximum reactor-averaged oxygen uptake rate
- **Inputs:** order 10 chemical and processing design parameters

- **Fidelity Lvl 1:** HF simulation (pretreatment, enzymatic hydrolysis, bioreactor)
  - 32 CPU-cores @ 57 hours
- **Fidelity Lvl 2**: HF pretreatment, LF Lignocellulose model, and HF bioreactor
  - {72 CPU-cores @ 4 hours, or 32 CPU-cores @ 9 hours}
- Could add simulation type that varies the time to steady state/grid resolution

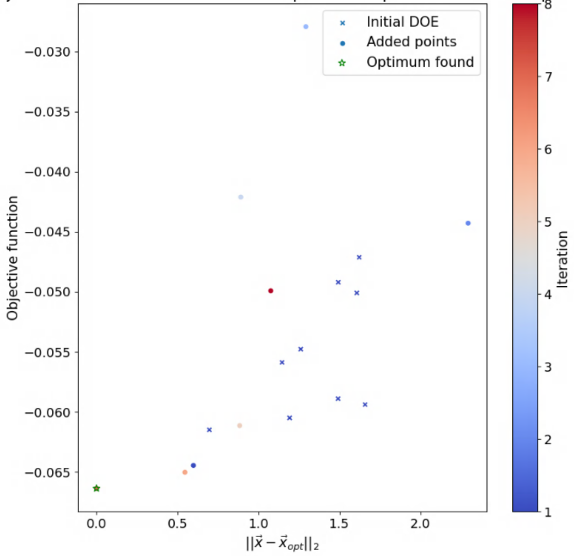
Parameter name	Double or Int*	Default	Min	Max
Fraction of solids that is xylan	Double	0.263	0	1
Fraction of solids that is glucan	Double	0.4	0	1
Porous fraction of the biomass particles	Double	0.8	0	1
Initial concentration of acid	Double	0.0001	0	1
Steam temperature (C)	Double	150	3.8	250.3
Fraction of insoluble solids	Double	0.745	0	1
Enzymatic load	Double	30	0	1000
FIS_0 target	Double	0.05	0	1
Gas velocity (m/s)	Double	0.08	0.01	0.1
Column height (m)	Double	40	10	50
Column diameter (m)	Double	5	1	6
Bubble diameter (m)	Double	0.006	0.003	0.008
OUR_max (mol/m^3/hr)	Double	88.71	5	100

*Integers require adjustment of the MF methods

Preliminary tests on virtual engineering app

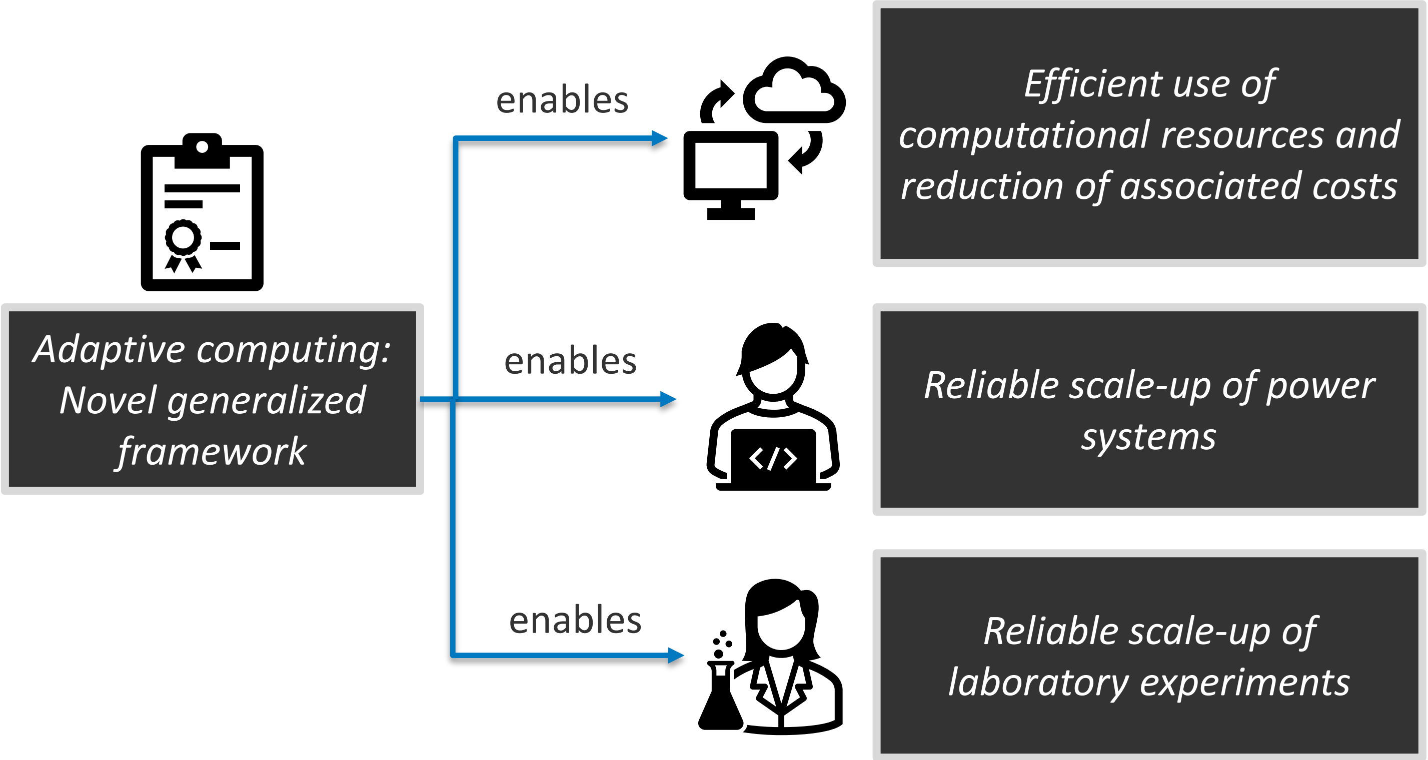
- 8 parameters, 9 random samples from LHS, 10 iterations
- Run for 5 minutes on 1 core, low fidelity models only
- Max Oxygen Uptake Rate = 0.06723323
- For some parameter settings, we obtained NaNs
  - The low fidelity models used may not be valid across the entire parameter space (hidden constraints – we know how to deal with these)

(No multi-fidelity business yet)



Parameter name	VE default	Min	Max	Final
Fraction of solids that is xylan	0.263	0	1	0.32
Fraction of solids that is glucan	0.4	0	1	0.29
Porous fraction of the biomass particles	0.8	0	1	0.64
Initial concentration of acid	1e-4	0	1e-3 (1)	1e-3
Steam temperature (C)	150	3.8	250.3	170
Fraction of insoluble solids	0.745	0	0.99 (1)	0.99
Enzymatic load	30	0	1000	57
FIS_0 target	0.05	0.005 (0)	1	0.005

Adaptive computing: optimizing the use of computational resources to target deficiencies and challenges related to scale-up



Thank You

www.nrel.gov

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by LDRD. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Photo from iStock-627281636

NREL

Transforming ENERGY