This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

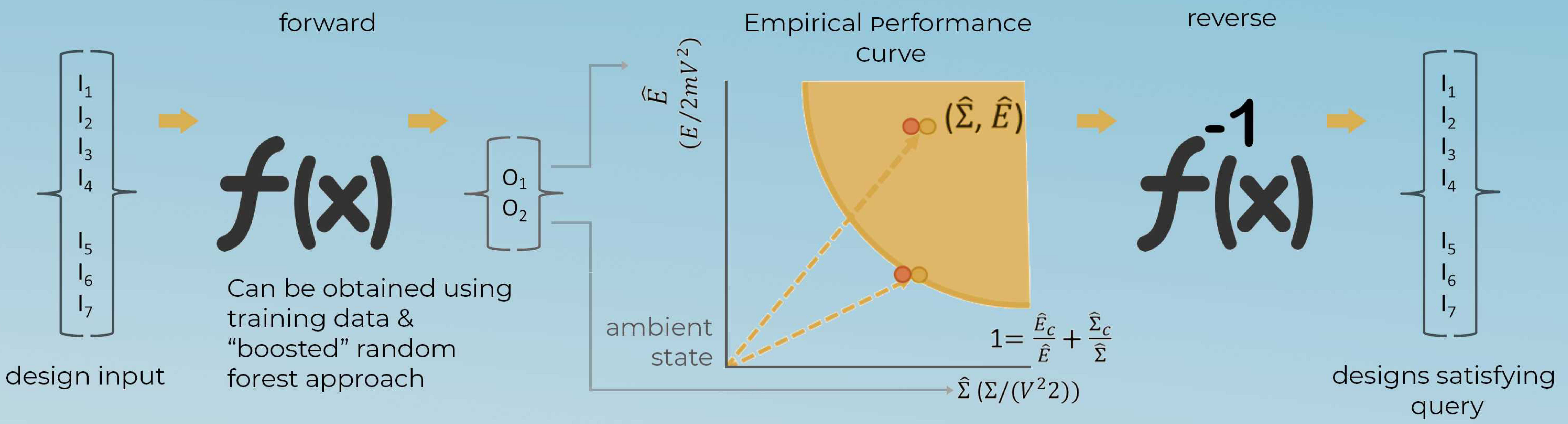
SAND2020-2043C

Leveraging a Machine Learning Approach to Solve an Inverse Problem Performance Query

Christopher J. Garasi, PhD, Sandia National Laboratories* David Dooling, PhD, RS21 Kameron Baumgardner, MS, RS21

This method avoids the inverse analytical problem in high-dimensional input spaces by uniformly sampling the input space and then clustering combinations of inputs based on a predicted proximity to a desired result.

A designers dream


forward
I_1
I_2
I_3
I_4
I_5
I_6
I_7
$f(x)$
O_1
O_2
Empirical Performance curve
$\hat{E}$
$(E/2mV^2)$
$(\hat{\Sigma}, \hat{E})$
ambient state
$1 = \frac{\hat{E}_c}{\hat{E}} + \frac{\hat{\Sigma}_c}{\hat{\Sigma}}$
$\hat{\Sigma}\ (\Sigma/(V^2 2))$
reverse
$f^{-1}(x)$
I_1
I_2
I_3
I_4
I_5
I_6
I_7

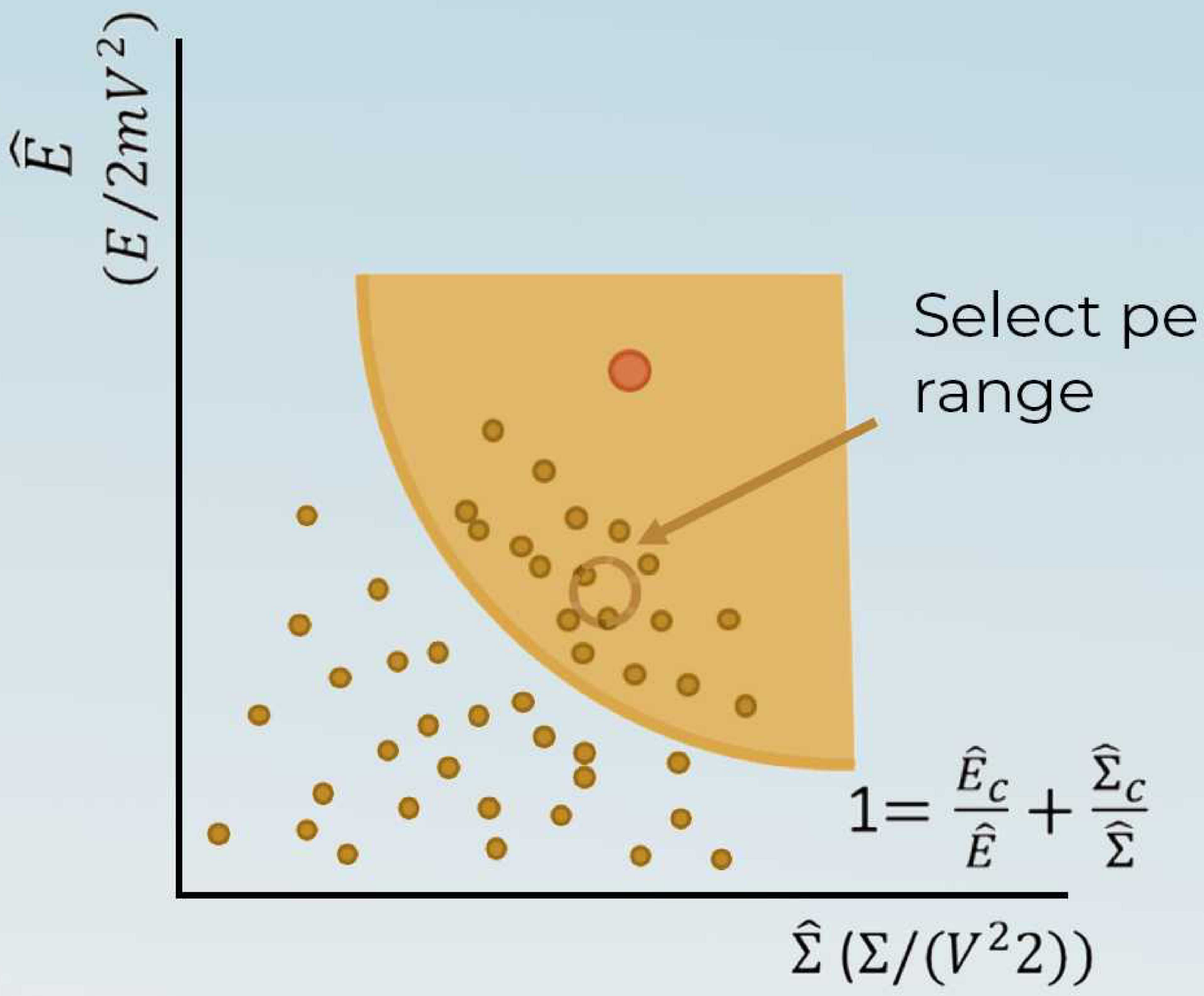
design input

Can be obtained using training data & “boosted” random forest approach

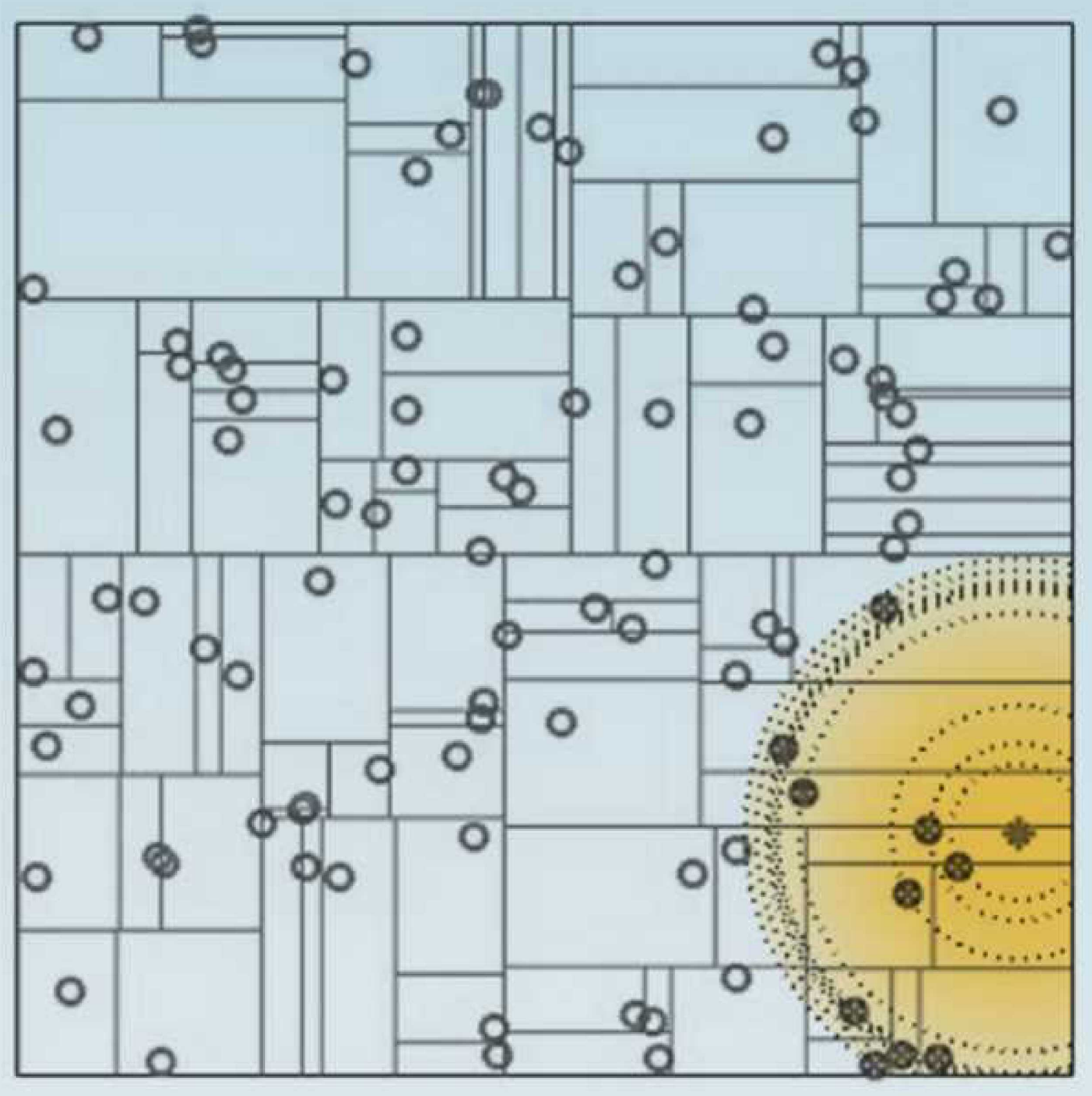
designs satisfying query

How do you solve the reverse problem using $f(x)$?

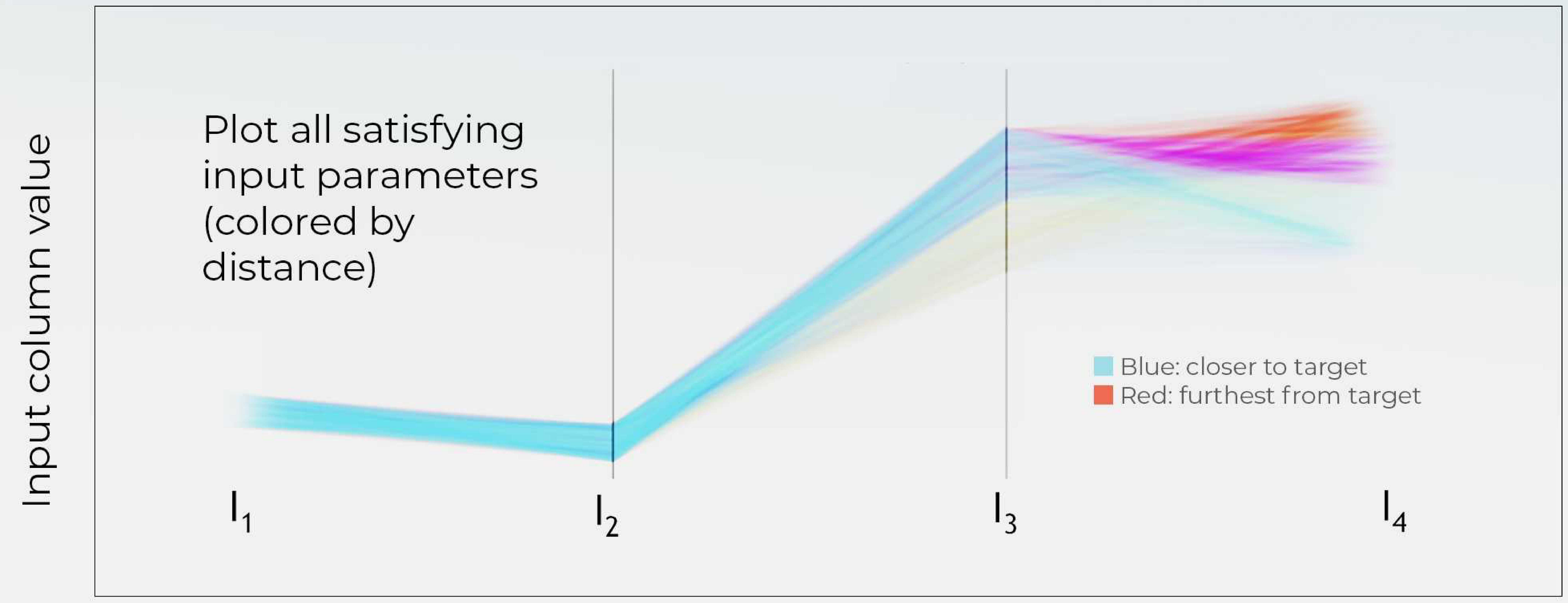
Submit 10M points into $f(x)$



Select performance range



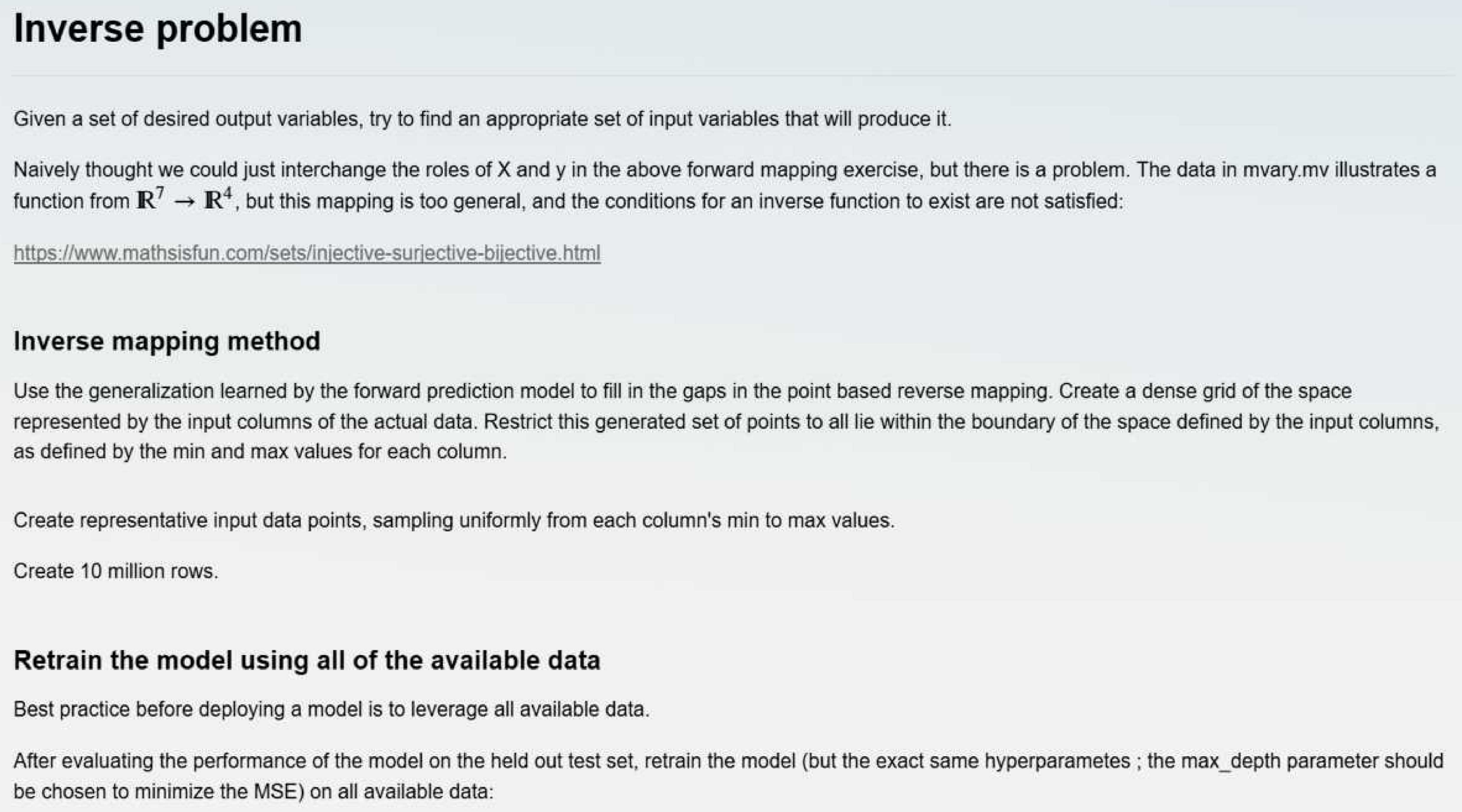
Use a KD tree to quickly find neighbors within the range



Plot all satisfying input parameters (colored by distance)

🟦 Blue: closer to target
🟥 Red: furthest from target

Jupyter Notebook



Inverse problem

Given a set of desired output variables, try to find an appropriate set of input variables that will produce it.

Naively thought we could just interchange the roles of X and y in the above forward mapping exercise, but there is a problem. The data in mvary.mv illustrates a function from $\mathbb{R}^7 \rightarrow \mathbb{R}^4$, but this mapping is too general, and the conditions for an inverse function to exist are not satisfied:

https://www.mathsisfun.com/sets/injective-surjective-bijective.html

Inverse mapping method

Use the generalization learned by the forward prediction model to fill in the gaps in the point based reverse mapping. Create a dense grid of the space represented by the input columns of the actual data. Restrict this generated set of points to all lie within the boundary of the space defined by the input columns, as defined by the min and max values for each column.

Create representative input data points, sampling uniformly from each column's min to max values.

Create 10 million rows.

Retrain the model using all of the available data

Best practice before deploying a model is to leverage all available data.

After evaluating the performance of the model on the held out test set, retrain the model (but the exact same hyperparametes ; the max_depth parameter should be chosen to minimize the MSE) on all available data:

*Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525.